APS490 Project Proposal

Intelligent Predictive Building System Maintenance Integration Platform

November 7, 2021

Elizabeth Chelmecki (1004187131)

Anoja Muthucumaru (995825353)

Chi Zhang (1004051250)

Shirley Zhang (1003906213)

Sherry Zuo (1003881711)

Supervisors:
Markus Bussmann
Professor and Chair of the Department of Mechanical and Industrial Engineering (MIE)

Seungjae Lee
Assistant Professor in Department of Civil & Mineral Engineering

Executive Summary

Modern Niagara Group (MNG) is a mechanical and electrical contractor that provides building maintenance services in heating, ventilation, and air conditioning (HVAC) systems. Currently, maintenance is conducted manually through building operator calls and scheduled servicing. MNG would like to automate their building servicing process with AI-driven decision making based on a multiple of sensor data they collect from the HVAC systems. Specifically, a proof of concept that uses the sensor information to autonomously identify building system anomalies in the CPPIB building is required.

After performing literature review and examining the available datasets provided by MNG the team selected the Air Handling Central Unit (CU) and Variable Air Volume (VAV) system for demonstration of proof-of-concept. By taking an in-depth look of the system components as well as their relationships, input and output parameters were identified and served as the foundation for the establishment of detailed functional requirements. The functional requirements were classified into five individual sub-tasks each representing a prediction target, namely zone temperature, damper and valve position, static pressure, supply fan speed, and mixed air temperature. A predicted value that deviates from normal operational conditions acts as an indicator of potential system faults.

The chosen solution approach is to create a digital twin, which is a data-driven representation of the VAV and CU operational behavior. After the digital twin generates predictions of how future system parameters are intended to behave, the predictions will be compared against the actual values of the physical system to identify any anomalies. Alternative designs to creating the digital twin were generated as follows:

- **Divergence:** A literature review on possible approaches to predictive building maintenance
- **Analysis:** We refined our ideas by analyzing the VAV and CU sensor information and grouping the data fields in functional task groups using domain knowledge from the client, literature, and our advisor.
- **Convergence:** Proposed three divergent designs of varying model complexity interconnectedness between system parameters that all achieve the intended functionality

Comparing three alternative designs under a decision-making matrix, we decided on the singular multi-task learning approach that shares information between tasks and best satisfies the client's needs. After training the multi-task learning algorithm, a series of experiments will be conducted to measure its success in autonomy, reliability, efficiency, generalizability, and modularity.

The design team believes we will be able to produce a valuable tool for MNG. A preliminary iterative design work plan has been developed to address the needs and objectives of developing a digital twin and testing said design. The majority of the team's resources will be spent creating an autonomous, reliable, and efficient ML infrastructure to drive additional value for MNG. Following a review of this design proposal in January, further development of the design will be showcased during the final report and presentation.

Table of Contents

1.0 Introduction

Modern Niagara Group (MNG) is the largest mechanical contractor in Canada with a strong presence in commercial real estate as well as government, healthcare, and public infrastructure projects. As a contractor for premium projects, MNG aims to maintain a competitive advantage through offering advanced building solutions. Particularly, the firm is expanding its integrated building technology (IBT) services, which aid in automating the control of HVAC systems. MNG hopes to drive additional value and elevate the customer experience through intelligence predictive maintenance (IPM). Thus, MNG has consulted our multidisciplinary capstone team to leverage data-driven decision making and build an in-house IPM pipeline.

2.0 Problem Statement

MNG provides regular maintenance to their customers via a scheduled service approach. For other maintenance issues (i.e. abrupt faults and degradation faults), MNG is reliant on the customer's local building operator to recognize issues and place a call to MNG's 24/7 dispatch for reactive service [Appendix A]. Manual fault diagnosis is a process that requires equipment downtime, which can negatively affect building occupants' comfort. This may result in delays in regular building activities inside office towers, hospitals, and manufacturing facilities. As such, maintenance events are costly to MNG and detrimental to the satisfaction of their customers.

Currently, MNG employs an analytics software (Figure 1) to track metrics from the various systems in their customer's buildings. As such, there are a multitude of system parameters being tracked and stored by MNG that are not being utilized for any function other than monitoring. These system parameters provide valuable information about the status of the building system in real-time. Before a fault occurs, the system parameters often exhibit abnormal relationships relative to each other [4], indicating an impending service need. Predicting these emergency maintenance events ahead of time, while the fault is in the early stage, would allow MNG to preemptively address the issue, save money through optimized repair scheduling, and improve service.

There are currently no available solutions on the market that achieve predictive maintenance. Building systems are not homogenous, and HVAC systems are complex (especially in large commercial buildings) requiring extensive engineering effort to develop a digital representation of the interconnected physical systems [7]. Therefore, MNG has proposed a capstone project to establish an in-house, data-driven predictive building-maintenance pipeline for HVAC systems using. Specifically, the client requires the team to develop a proof-of-concept that the current sensor information can be applied to identifying potential failures before they occur, without manual intervention. As a result, the team has narrowed the scope to two interconnected systems: the Air Handling Central Unit (CU) and Variable Air Volume (VAV) system. Failure of the CU and VAV can cause the indoor temperature to fluctuate and negatively impact the thermal comfort of building occupants. Thus, being able to detect anomalies in the operational state of the mechanical components as well as system outputs such as zone air temperature provides valuable information on system performance. Representing the data behaviour and relationships in two complex and interconnected systems effectively demonstrates the feasibility of data-driven predictive maintenance.

3.0 Service Environment

As discussed in section 2.0, the team has narrowed the scope of the proposal to CU and VAV boxes that service the seventeenth floor of the Canadian Pension Plan Investment Board (CPPIB) building. The following section will detail this HVAC system, the physical and digital environments in which the system operates, and how they interact.

3.1 The Physical Environment

The building the team is focusing on is One Queen Street East, pictured in Figure 1, which houses CPPIB offices on the 8-27th floors.

Figure 1: One Queen Street East, Toronto, ON M5C 2W5

High occupancy times are likely between the hours of 8 AM and 5 PM. This might require the building HVAC to respond to more frequent changes in temperature and humanity due to heat generated from electronics and people.

Additional information about the physical space that the HVAC system services, such as the layout and use cases of different office zones on the seventeenth floor, is unavailable to the team. However, the team is confident that this will not pose any problems to the design. Any insights obtained from this information can also be obtained from mapping the data provided by MNG (such as how temperature varies throughout the day).

3.2 The Digital Environment

The physical components of the HVAC system are integrated into a digital environment through a building automation system (BAS).

The BAS utilizes a control system to:

1. Automate the control of building climate and lighting based on occupancy schedules
2. Monitor the performance of the devices in all the system
3. Provide alarms in case of malfunctions [1]

To automate the control of building climate and monitor the performance of the physical devices in the HVAC system, the BAS collects data from multiple sensors, as well as tracks setpoints and commands which are prescribed by either the building manager or MNG [1].

The team will utilize this data in order to create a predictive model to meet the functions and requirements of the project. Specifically, the team will be using data collected from the CU and 34 VAV boxes on the seventeenth floor. The control schematics for the seventeenth floor are shown in Figure 2.

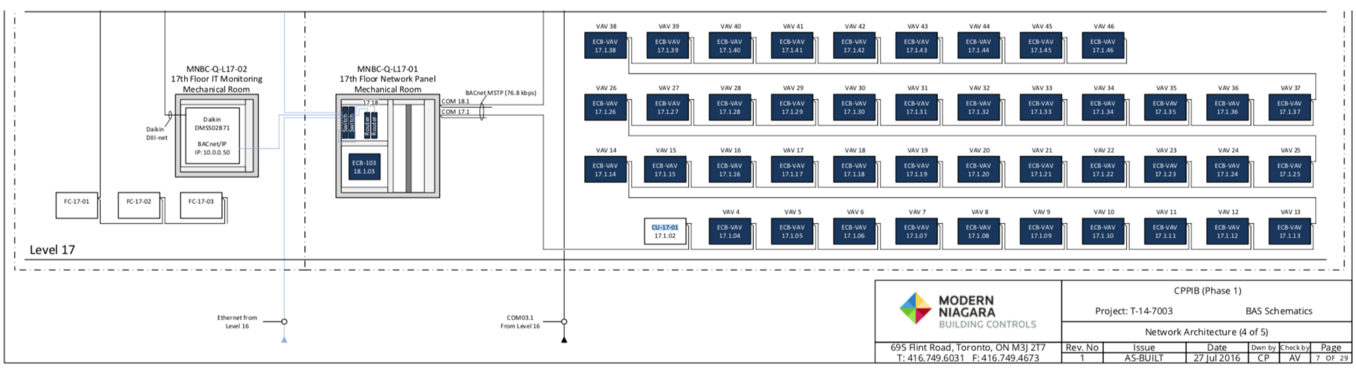


Figure 2: BAS Control Schematics for the seventeenth floor of the CPPIB building

The dataset's variables and meanings are detailed in Table 3. The variables are also classified based on if they are measurements, commands, or setpoints. Measurements are actual values for variables measured by sensors, setpoints are desired or target values for variables, and commands are values that physical components are set to in order to achieve setpoints.

To exemplify the differences between these types, consider the damper command variable. The damper command is a command for the physical position of the damper. It is changed to achieve the zone temperature setpoint, and is determined by the BAS by comparing the actual measured zone temperature with the zone temperature setpoint.

Variable	Definition	Measured/Command/Setpoint
DateTime	Date and time of the variables recorded	Measured
Occupancy	Boolean indicating if space is occupied	Measured
Unit State	The state of the system	Command
Direct Expansion Cooling. Condenser Water Temperature	Temperature of the condenser water leaving the system (after absorbing heat)	Measured
Direct Expansion Cooling.Cooling Capacity	The measure of how much heat the system can remove	Command
Direct Expansion Cooling.Enable	Enables direct expansion cooling (i.e. air handling unit will operate)	Command
Discharge Air.Static Pressure	The static pressure inside the duct system	Measurement
Discharge Air.Static Pressure Setpoint	The target static pressure inside the duct system	Setpoint
Discharge Air.Temperature	The temperature of air leaving the air handling unit	Measurement
Discharge Air.Temperature Cooling Setpoint	The set temperature of air inside the air handling unit	Setpoint
Discharge Air.Temperature Setpoint	The set temperature of air inside the air handling unit	Setpoint
Mixed Air. Temperature	Temperature of the air in the return loop of the air handling unit after mixing with outdoor air	Measurement
Return Air.Temperature	Temperature of the air in the return loop of the air handling unit before mixing with outdoor air	Measurement

Supply Fan.Speed	Speed of the fan blowing air over the coil	Measurement
Supply Fan.Status	If the supply fan is off or on	Measurement
Fin Tube Radiation. Valve Position	Position of the fin tube valve in each VAV box	Measurement
Primary Air. Air Flow	Airflow out of the damper	Measurement
Primary Air.Air Flow Cooling Maximum Setpoint	The maximum amount of airflow allowed through a damper for cooling	Setpoint
Primary Air.Air Flow Cooling Minimum Setpoint	The minimum amount of airflow allowed through a damper for cooling	Setpoint
Primary Air.Air Flow Heating Maximum Setpoint [cfm]	If the heating required the maximum airflow setpoint is 480.	Setpoint
Primary Air.Air Flow Heating Minimum Setpoint [cfm]	If the heating is kept to a minimum the airflow is set to 144.	Setpoint
Primary Air.Air Flow Setpoint [cfm]	Setpoint for airflow dependant on the temperature required in each zone	Command
Primary Air Damper.Damper Command	Position the damper is set to	Command
Primary Air Damper.Damper Position	Position the damper is actually in	Measurement
Zone Air.Temperature [$\circ$ C]	Actual temperature of zone	Measurement
Zone Air.Temperature Cooling Setpoint [$\circ$ C]	Temperature setpoint in cooling mode	Setpoint

Zone Air.Temperature Heating Setpoint [$\circ\ C$]	Temperature setpoint in heating mode	Setpoint

Table 3: Description of variables provided by MNG

In the following section, each variable and how it interacts in the physical system will be explained.

3.3 The HVAC System - Air Handling Unit and VAV boxes

The function of the HVAC system is to provide thermal comfort to occupants of the space and maintain indoor air quality through ventilation [2]. This function is achieved on the seventeenth floor of the CPPIB building with an air handling unit and 34 VAV boxes. Each VAV box services a zone with potentially different heating and cooling requirements.

In VAV systems, the temperature of the air in the system remains constant and airflow into each zone (controlled by the individual VAV boxes) is adjusted in order to meet the different cooling loads of the individual zones [3].

The supply air taken into the air handling unit is a mixture of the outdoor air and return air (this temperature is tracked by the variable *Return Air.Temperature*) from the system. This supply air has an initial temperature (represented by the variable *Mixed Air.Temperature*), and must be cooled by the air handling unit to the discharge air temperature setpoint (in our dataset, this is the variable *Discharge Air.Temperature Cooling Setpoint*, which is fixed at $15\ \circ\ C$). These relationships are shown in Figure 3.

Please note that only variables available in the dataset were modelled in the figures of this section. For example, even though damper positions and other setpoints affect the mixed air temperature, these variables were not included. Additionally, for simplicity, the feedback in the system was not modelled (i.e. the figure represents one time step). The feedback between different components of the system will be discussed further in the proposal.

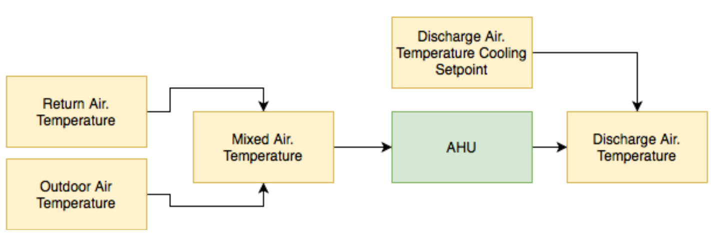


Figure 3: Supply and discharge air temperature variables

The air is cooled to the setpoint temperature when it is blown by the supply fan (which has a speed, measured by the variable *Supply Fan.Speed*) over a coil with refrigerant running through it. Heat transfers from the air to the refrigerant and the temperature of the air decreases. The air temperature leaving the air handling unit and entering the VAV boxes is measured by the variable *Discharge Air.Temperature.* This temperature should be close to the setpoint of $15 \circ C$.

The speed of the fan determines the static pressure inside the duct system. There is a set point for the static pressure (in the dataset, this is represented by the variable *Discharge Air.Static Pressure Setpoint,* which is set to 1.401044325). The actual static pressure inside the duct system is measured by the variable *Discharge Air. Static Pressure*. As the dampers of different VAV boxes open and close, the static pressure inside the duct system changes. In order to maintain the static pressure setpoint, the supply fan speed is adjusted. These relationships are shown in Figure 4.

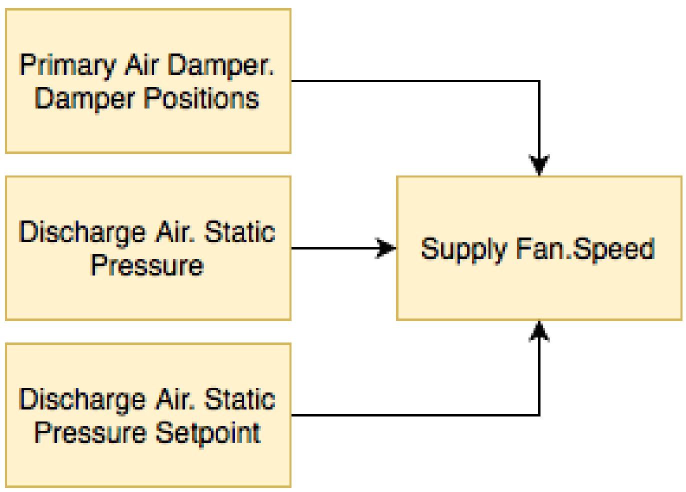


Figure 4: Supply fan speed variables

The air handling unit feeds air of a constant temperature, represented by the variable *Discharge Air.Temperature,* into a duct system with 34 VAV boxes that feed air into different zones on the seventeenth floor. Because the different zones of the seventeenth floor have different zone temperatures, represented by the variable, *Zone Air.Temperature*, the VAV boxes function to vary the airflow entering each zone. They do this via dampers, which open and close depending on how much cooling a zone requires. The position the dampers should be open to is represented by the variable *Primary Air Damper.Damper Command,* and the position they are actually open to in real-time is represented by the variable *Primary Air Damper.Damper Position*.

The position of each damper depends on how much airflow the zone requires, represented by the variable *Primary Air.Primary Air Setpoint.* This airflow setpoint is dependent on the temperature differential of the zone. The primary air setpoint of each zone has a maximum and minimum for heating and cooling, represented by the four variables *Primary Air.Air Flow Cooling Maximum Setpoint*, *Primary Air.Air Flow Cooling Minimum Setpoint, Primary Air.Air Flow Heating Maximum Setpoint*, and *Primary Air.Air Flow Heating Minimum Setpoint*. The relationships between these variables are shown in Figure 5.

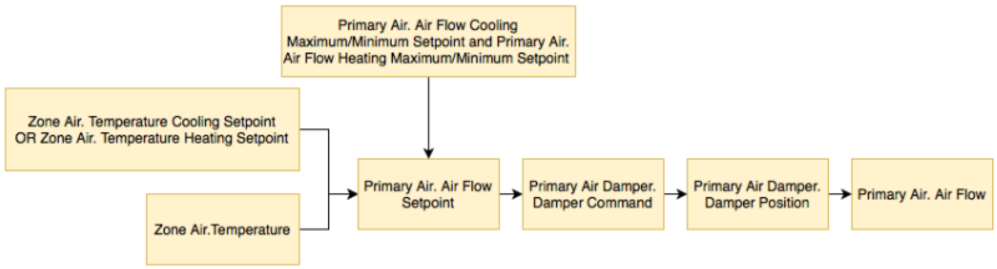


Figure 5: Damper position variables

The setpoint temperatures of each zone are given by two variables, the *Zone Air.Temperature Cooling Setpoint* and the *Zone Air.Temperature Heating Setpoint*. In the cooling mode, the temperature setpoint is the cooling setpoint, and the system will try to match the zone temperature to the cooling setpoint. In heating mode, the system will try to match the zone temperature to the heating setpoint.

If heating in the building is required, this is done via local fins in each VAV box. The operation of these fin tubes is represented by the variable *Fin Tube Radiation.Valve Position.*

3.4 Interactions

The digital BAS system directly interacts with the physical HVAC system. The BAS maintains the various setpoints and commands set by either the building manager (such as the cooling temperature setpoint) or the designer of the system (such as static air pressure inside the AHU). The BAS uses sensors to measure how different variables are from their setpoints. It then controls various physical components in the system to bring these variables to their setpoints.

4.0 Stakeholders

The following table shows the agencies that are influenced by and have influence upon the design for data-driven prediction of potential service requests.

Stakeholder	Impact
Government	

(General Impact: Governing bodies must approve that the design satisfies legal regulations)	
City of Toronto	● If the design requires mechanical components to be added to existing buildings, any updates must comply with City of Toronto regulation. [4]
Province of Ontario	● Collection of building sensor data must comply with guidelines set out by the Information and Privacy Commissioner of Ontario [14][15]
Modern Niagara Group	
IntegratedBuilding Technology (IBT) Team (The team in MNG who is directly in communication with the capstone group)	● The design contributes to their goal of achieving more autonomous buildings, as a result, this team will be supplying all the required data for the project. ● The design could generate additional value / revenue for this team ● The design could modify the current database, software pipeline, and workflow of the team ● The design could save valuable time in the IBT team's day-to-day workflow
Building Service Technicians	● The design aims to optimize the scheduling of repairs for building service technicians ● The design will help technicians narrow down on problems and save valuable time and manpower.
Service Dispatchers	● Instead of receiving phone calls, dispatchers will be alerted of service requests and building failure through the designed system ● The design could increase the efficiency of service dispatching ● Dispatchers will need to learn how to interact with the system
People serviced by MNG	
Clients of MNG (Ex. CIBC, Shopify, Government, etc.)	● The design could provide corporate / institutional clients with an elevated building experience, in which the building is more automated ● The design could reduce costly delays caused by emergency repairs ● The design, after deployment, will collect and monitor the sensor data at client buildings

Occupants of buildings (Ex. Office workers, hospital staff, etc.)	● The design could make the buildings more comfortable for occupants through preventing the building parameters from entering extreme ranges by predicting them ahead of time
Capstone Team	
Capstone Team	● The team, consisting of five students from different faculties and disciplines, are responsible for designing the project according to the project requirements ● Each student is expected to exercise knowledge from their respective area of study, which is outlined in section 10.3 ● The role of each student in the team is in section 10.3 ● The design can help further the students' industry knowledge and develop their technical skills
Faculty Advisors: Professor Markus Bussmann Professor Seungjae Lee	● Faculty advisors will be consulted for guidance throughout the design process and course deliverables ● Professor Lee specializes in machine learning applications to energy systems, and will be the expert advisor for technical knowledge ● The design can help advisors understand the workflow of companies like MNG and as a result enhance the usability of their research findings

Table 4: Stakeholders affected by IPM design

5.0 Detailed Requirements

5.1 Functions

Representing the physical systems is possible, but since each building is different, solutions are not easily transferred from building to building [5]. Additionally, as a building ages, the systems within that building will degrade and the baseline operating conditions will change [5]. Thus, machine learning methods are appropriate to the predictive maintenance task, as the models are able to dynamically learn component relationships and patterns over time.

Ideally, the output of the predictive maintenance model would be failure classification. However, time-series data on previous system failures that aligns with sensor data is unavailable, and directly training the model to detect and classify failure is not feasible. Thus, the team has opted to achieve anomaly detection through creating a digital representation of the building VAV and CU system. Essentially, a digital twin [13] takes sensor information from core operational components of the building system and learns the system behaviour. It can then be used to generate predictions of how the systems *should* function. Thus, MNG and the building operators can use the predicted system parameters as guidelines to assess whether the physical system may have impending faults. Beyond anomaly detection, digital twins can generate additional future value - the MNG could use the model to run simulations and identify areas of optimization.

5.1.1 Functional Requirements

The designed pipeline will process AHU and VAV sensor data from the 17th floor of the CPPIB building to predict system outputs and identify impending system faults. The predicted system outputs from the model will represent what the AHU and VAV system on the 17th floor should look like at a certain time in the future. These predicted outputs will be compared in real-time to the measured system parameters in order to identify any anomalies or inefficiencies.

The model's outputs will create a digital representation of the physical system using previous operating data. If there is a significant discrepancy between the measured and predicted system parameters, the design will send an early warning to the user, which aids in scheduling maintenance. The design is also able to display ongoing metrics.

The above functionalities are illustrated in Figure 6 below:

Functional Basis: Identify impending system faults

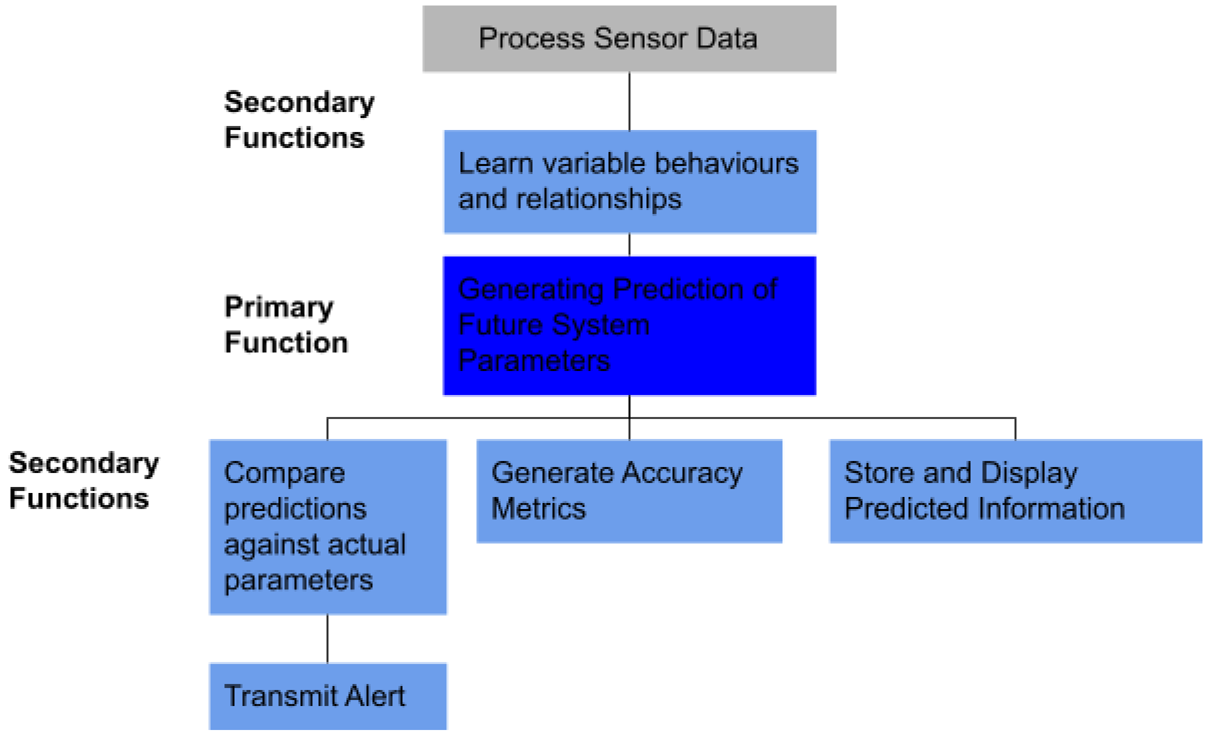


Figure 6: Functional Breakdown of design

5.1.2 Workflow and Task Breakdown

To create a digital representation of the AHU and VAV boxes, the values of five variables will be predicted. The prediction of these variables can be separated into different tasks. The following section details each task and it's inputs and outputs.

The inputs and outputs considered are those that exist in the datasets provided by MNG. Thus, if there is an input that affects an output specified by one of the tasks below, but this input is not tracked by MNG, it is not discussed. The designed solution will have to account for these missing variables.

5.1.2.1 Zone temperature prediction

The zone air temperature is the temperature of the air in each zone corresponding to an individual VAV box. The inputs the model will take to predict zone air temperature are listed in Figure 7. In Figures 7-11, green boxes indicate an output, yellow boxes indicate an input, and orange boxes indicate setpoints (which will also function as inputs).

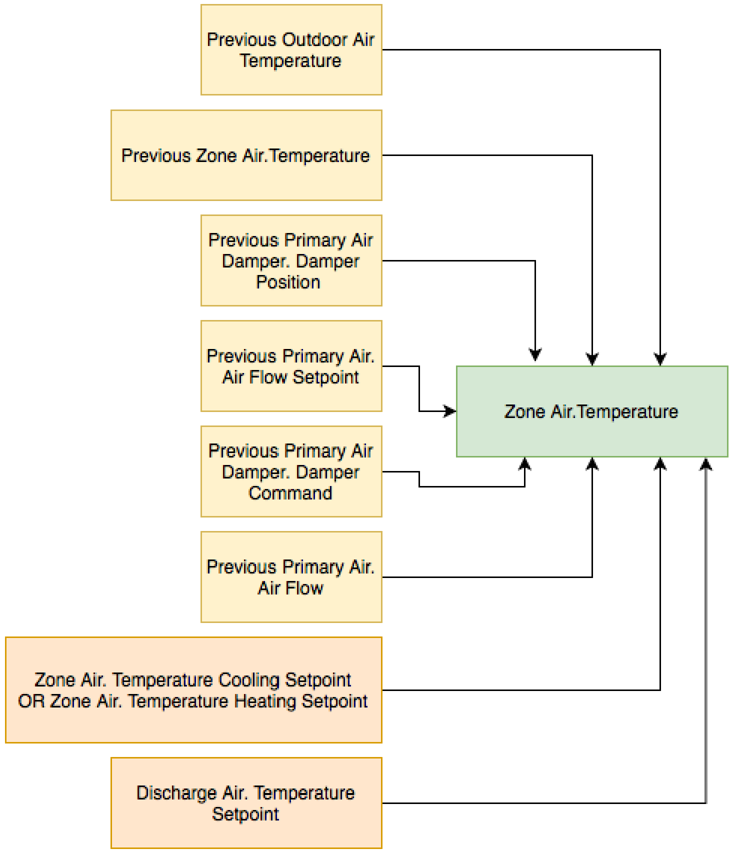


Figure 7: Zone air temperature prediction inputs

5.1.2.2 Damper position and damper air flow prediction

The damper position is the position the dampers are open to. The damper position is determined by how much air must flow through each damper to bring the actual zone temperature to the zone temperature setpoint. The inputs the model will use to predict the damper position and damper air flow are listed in Figure 8.

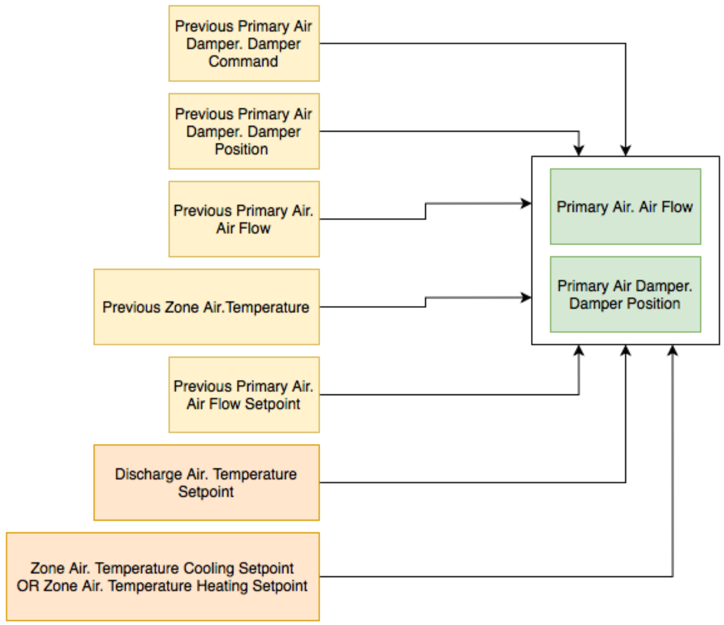


Figure 8: Air flow and damper position prediction inputs

5.2.2.3 Air handling unit static pressure prediction

The air handling unit static pressure is the static pressure of the ducts of the air handling central unit. This pressure is affected by the position of the dampers and the supply fan speed. As the dampers of each VAV box open and close to different positions, the supply fan speed must change in order to maintain the static pressure setpoint. The inputs the model will use to predict the air handling unit static pressure are listed in Figure 9.

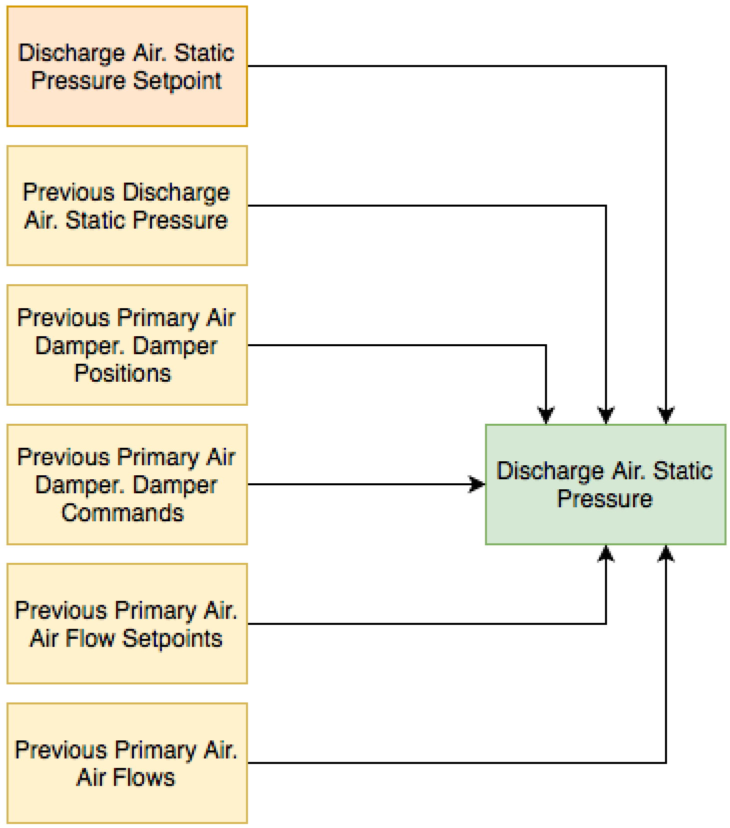


Figure 9: Air handling unit static pressure prediction inputs

5.2.2.4 Supply fan speed prediction

As discussed in 5.2.2.3, the supply fan speed is a function of the damper positions and the static pressure setpoint. The inputs the model will use to predict the supply fan speed are listed in Figure 10.

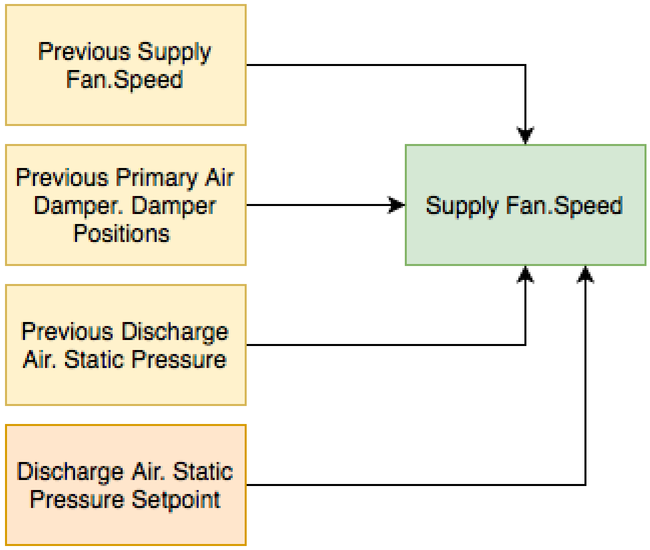


Figure 10: Supply fan speed prediction inputs

5.2.2.5 Mixed air temperature prediction

The temperature of the air to be cooled to the discharge air temperature setpoint by the air handling unit is a function of the mixed air temperature and the outdoor air temperature. The inputs the model will use to predict this mixed air temperature are listed in Figure 11.

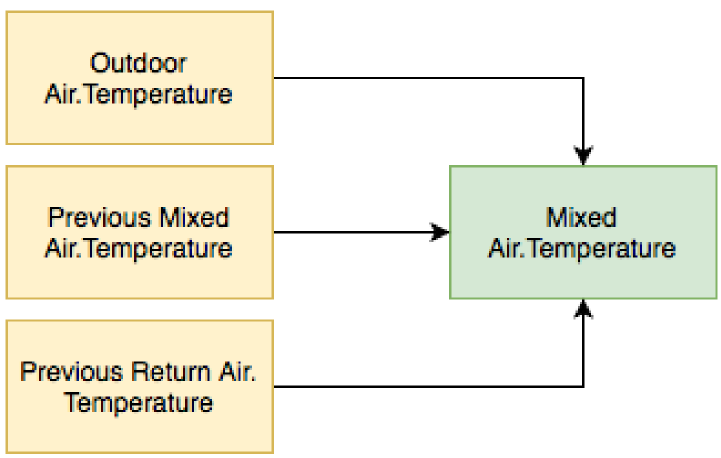


Figure 11: Mixed air temperature prediction inputs

5.2 Objectives

The following objectives were determined based on the client's needs. MNG considers autonomy as their top priority in order to reduce manual work performed by building operators and dispatchers. The client also values reliability which directly impacts the quality of service they will deliver to their customer. In addition, the client wants an efficient, generalizable, and modular algorithm that can be trained and deployed in a short time frame as well as adapt to different customer sites.

5.2.1 Autonomous

- The design will generate predictions based on input data without supervision.

5.2.2 Reliable

- The design will generate real-time predictions that are reasonably close to the actual operational state of the physical system.
    - The design will achieve 90% training accuracy at the minimum [7].

5.2.3 Computationally efficient

- The design will consume the minimum required computational resources and within available computational resources.
    - The design will require at most 16GB RAM [8] to train and make predictions which can be achieved on standard personal computers.
- The design will generate predictions at the same frequency as data input.
    - The design will make predictions at 5-minute intervals (same as input data).
- The design will be able to be re-calibrated as new historical data becomes available.
    - The design will complete training during company off-hours (overnight or weekend).

5.2.4 Generalizable

- Given the sensor data and specific building characteristics, the design will have the capacity to model different buildings with similar performance accuracy.

5.2.5 Modularity

- The design can be isolated into separate components that are tested individually. The failure of one component will not impact other components.

5.3 Constraints

As the design operates across residential and commercial buildings, it must adhere to all relevant bylaws and building codes enforced federally and city-wide in Toronto. It must also follow specific policies imposed by building management agencies and occupants.

5.3.1 Legal Regulations

The design shall comply with all building maintenance standards enforced by the City of Toronto. For example, the design should be compatible with the State of Good Repair (Capital) Plan which requires landlords to provide a five (5) year forecast of major building repairs. The design shall also adhere to technical requirements for the energy-efficient design and construction of high-rise towers set out by The National Energy Code of Canada for Buildings (NECB) [9].

5.3.2 Security

The design shall follow all security protocols of the existing systems it will be integrated with. The design shall not introduce any security vulnerabilities in the network systems it is mounted on [10].

5.3.3 Data Collection

MNG should be informed of and consent to any data collected as a part of the design. The data collected from the client shall only be used for the purpose of the design and made available to MNG upon request. Any intended release of data to third parties shall be approved by MNG.

In addition, the design must respect the privacy of unit occupants, under the guidelines of The Personal Information Protection and Electronic Documents Act (PIPEDA) [11]. Any data collected from open-source databases should also comply with the Freedom of Information and Protection of Privacy Act [6].

5.3.4 Affordable

The cost of the implementing, deploying, and maintaining the design shall not exceed the cost of the current platform and service request process (i.e. the project will generate non-negative return on investment).

5.3.5 Compatible

The design shall be compatible with the existing data acquisition system. The format of the given data shall be used as inputs to the design without the need for alterations.

6.0 Alternative Designs

6.1 Step 1: Predicting Output Variables

6.1.1 Alternative Design 1 for Step 1 - Single Target Prediction

In previous sections, the team identified five different tasks (groups of inputs and outputs) that are necessary in creating a digital representation of the VAV and CU systems. Notably, predicting each selected output only requires a subset of the input features, and therefore each sub-problem can be solved separately, learning separate smaller sets of model parameters. Since each subset of inputs and outputs has implicit biases and behaviours, the team can select the best-suited machine learning model for each task, which potentially improves accuracy. In the single target model, we treat each output variable as independent of each other, assuming no underlying relationships, or sequential order, between the target variables [22]. Essentially, singular outputs are directly predicted using the historical dataset of the subset of input features that it depends on.

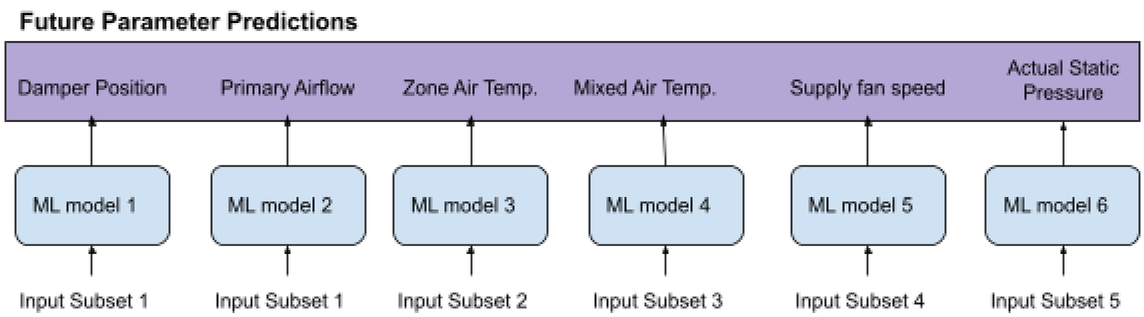


Figure 12: Single Target Model Architecture

Objectives Evaluation

Objective	Specifications
	Single Target Prediction
Autonomous	● The design is fully autonomous and generates predictions based on input data without supervision.
Reliable	● The design is based on the assumption that the output parameters are independent from each other, which is not true. This will decrease the accuracy achieved by the model. ● The accuracy of the model decreases as the horizon prediction time increases. This is because the individual models are not updated with more recent information from other models.
Computationally Efficient	● Variable relationships are not exploited to make the design more computationally efficient. There is a high redundancy because many targets have shared inputs.
Generalizable	● Additional targets and inputs that may be required in other systems can be easily added without affecting the current model.
Modularity	● Because the model for each task is separate, the flaws and inaccuracies in one model will not affect the predictions for the other tasks. Thus, the design is highly modular, with separate components that can be tested individually.

Table ##: Objectives evaluation for alternative design 1 - single target prediction

6.1.2 Alternative Design 2 for Step 1 - Deep Regressor Chains

The machine learning model must predict multiple continuous output parameters from multiple inputs. This type of problem is classified as a Multi Task Regression (MTR) [20].

Since the goal is to create a digital twin, omitting the relationships between the tasks and target variables is insufficient to represent the VAV and CU systems. Thus, the team proposes capturing sequential relationships through using Deep Regressor Chains (DRC)[19] for MTR. In the DRC model, the first model is trained independently of the other target parameters, and the predicted values are added to the training set of the next target parameter that depends on it. Thus, redundancy is minimized because shared input parameters are only passed into the chain once.

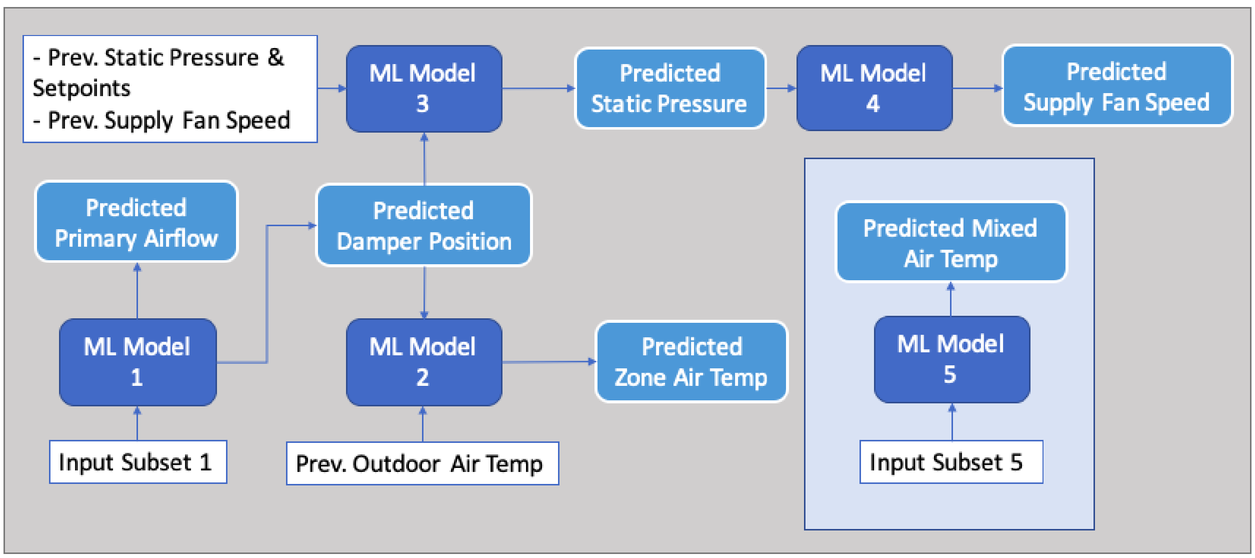


Figure 13: Deep Regressor Chain Model Architecture

Objective	**Specifications**
	Deep Regressor Chains
Autonomous	• The design is fully autonomous and generates predictions based on input data without supervision.
Reliable	• Because tasks are separated, the best model for each task can be chosen, which may increase accuracy. • Flaws in one model's output can create a domino effect in the chain of tasks, leading to large-scale error and lowered accuracy.

Computationally Efficient	● The design is computationally more efficient because the input dimensions are reduced by exploiting the relationships. This results in less model parameters. ● The processes are sequential and thus the computation for one task must finish before the next process can commence.
Generalizable	● Additional targets and inputs that may need to be added for other systems require some knowledge to add to the model and retain the proper relationships between variables.
Modularity	● The system is modular in that it separates the process into five separate tasks. The outputs are produced in sequence, and errors can be identified along the way. ● The system has decreased modularity as a result of the relationships defined between the parameters, which need to be re-defined to add or remove a task.

Table ##: Objectives evaluation for alternative design 2 - deep regressor chains

6.1.3 Alternative Design 3 for Step 1 - Multi-task Learning

To incorporate all inputs in parallel and learn variable relationships autonomously, the team proposes to apply Multi-Task Learning (MTL). MTL aims to find common features between tasks, as sharing features among tasks as a means to learning representations which capture invariant properties to tasks can be highly beneficial [12]. Essentially, all input parameters will be encoded in the shared layers and then the MTL model will process the data into feature-specific layers. In the feature specific layers, there are layers for time series prediction and several linear layers. For each feature, we will choose a model to employ. Finally, we will get the output parameter values.

The MTL model eliminates redundancy in inputs and model parameters and allows for coupling between tasks, so that they send updates to each other.

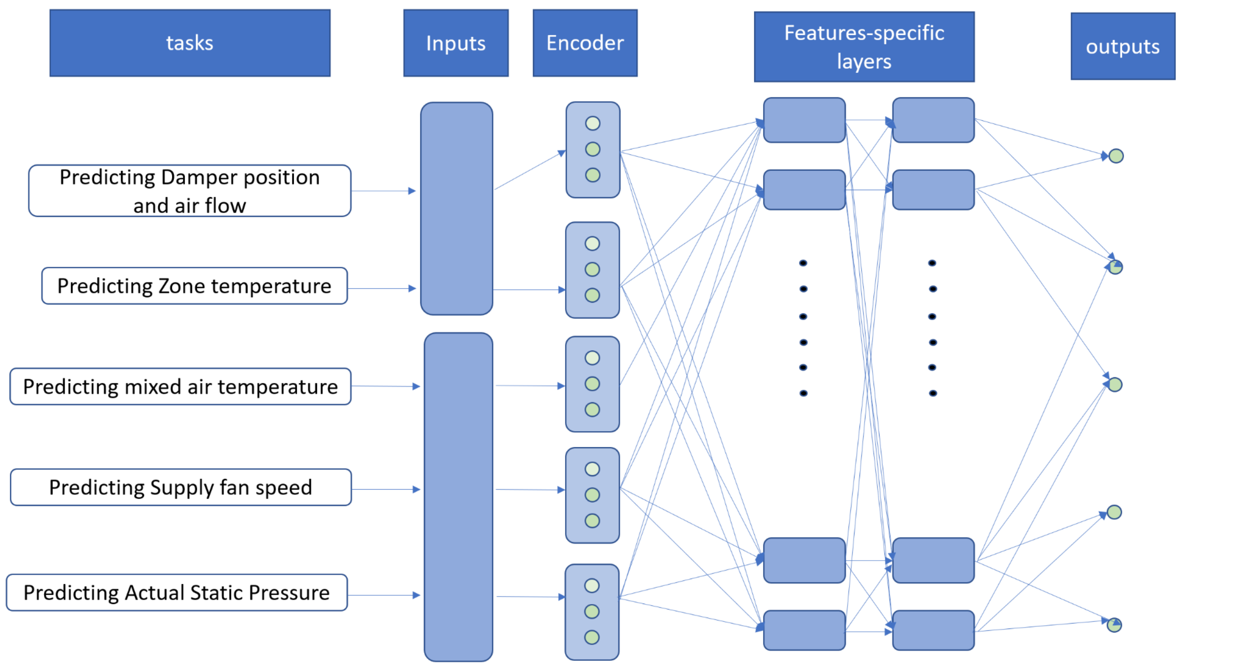


Figure 14: MTL Model Architecture

Objective	**Specifications**
	Multi-task learning
Autonomous	• The design is fully autonomous and generates predictions based on input data without supervision.
Reliable	• The model is complex and requires a large amount of data to train to avoid underfit. • The model has the power to efficiently and dynamically learn the relationship between building parameters
Computationally Efficient	• Has no duplicated input parameters • Different tasks share representations which reduces redundant computation. • Tasks could be conducted parallelly to avoid time delay caused by squeezed tasks in the queue in each round
Generalizable	• Multiple tasks are integrated in a model, which improves the generalization of the model

	● Model is more complex, so greater generalization given sufficient data
Modularity	● The representations of models of different tasks are coupled. The model has the least modularity.

Table ##: Objectives evaluation for alternative design 3 - multi-task learning

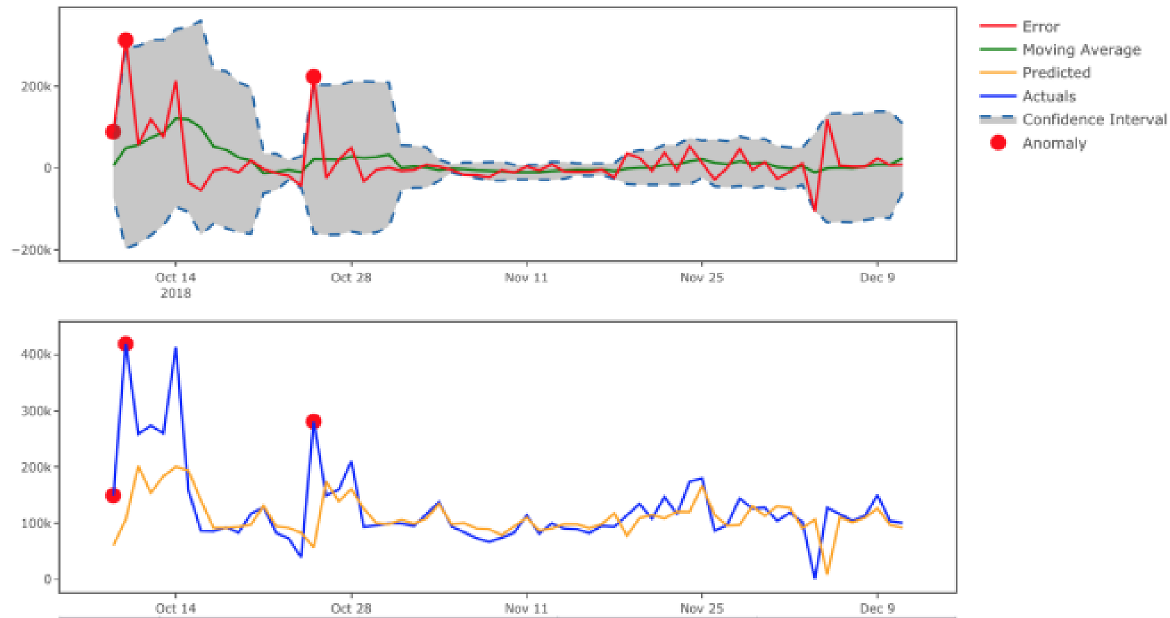
6.2 Step 2: Time Series Anomaly Detection

After the predictions are generated, the team will identify potential anomalies, so that the operator is alerted when an outlier occurs. The team proposes using simplistic methods such as:

- **Threshold-based detection** based on system specifications. An alert will be generated if the predicted parameter values lie outside of a predefined range.
- **Sliding-window confidence interval detection** the predicted values will be compared to the actual values within a predefined sliding window of consideration. If the predicted values are outside of a predefined number of standard deviations from the actual values, an alert will be generated.

Besides the above simplistic anomaly detection methods, the team will also explore more complex statistical methods within the python Anomaly Detection Toolkit (ADTK) [17]

An illustration of a confidence-interval based anomaly detector is shown below:



Date	Actual Values	Predicted	% Difference	Severity (0-3)
2018-11-22	147069	108739.168	26.062%	0
2018-11-21	108126	114978.294	-6.337%	0
2018-11-20	134514	109228.065	18.798%	0
2018-11-19	113630	77791.011	31.54%	2
2018-11-18	94102	117629.725	-25.002%	1
2018-11-17	95410	98640.56	-3.386%	0

Figure 15: Illustration of anomaly detection between actual and predicted values using sliding-window confidence interval approach [16]

7.0 Proposed Design

7.1 Weighted Decision Matrix

The five identified objectives were discussed with MNG, who influenced the weights assigned to the objectives. Modularity was emphasized, because the client desires that any problems should be easily diagnosed and additional components should be able to be added to the model to expand the digital representation. The next two heavily-weighted objectives were reliability and autonomy, because they are related to model performance. However, the client has expressed that beyond a reasonable baseline, small improvements in model accuracy give little additional value. Efficiency and generalizability are more related to model optimization and improvement, so they are weighted less than the objectives related to model functionality.

Objectives	Design 1	Design 2	Design 3
Autonomous (20)		5	
Reliable (20)	2	3	4
Computationally Efficient (15)	2	3	4
Generalizable (15)	2	3	4
Modularity (30)	5	4	2
Total Score	250	270	260

*1 = None / Very Little 5 = Fully
** Numbers for reliability have not been implemented and tested, therefore subject to change

7.2 Final Design

Based on the design evaluation in section 7.1, the Deep Regressor Chain (DRC) model best fits the client objectives. Hence, we will create the first iteration of the digital twin using five separate machine learning models with sequential interconnections between inputs and outputs.

After the digital twin produces the predictions of future parameters, we detect anomalies between the predicted values and the actual values, by threshold-based detection as described in 6.2

8 Measures of Success

Through achieving these metrics, we realize the value of our project. The following table describes the metric and design goals corresponding to each measure of success.

Measures of Success	Design Goals	Metric
Autonomy	The design will generate predictions without human intervention.	Project demonstration: number of human interventions required
Reliability	**For prediction:** The design will generate real-time predictions that are reasonably close to the actual operational state of the physical system.	Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) are applied to measure the accuracy of prediction results, because it is scale

		independent
	For outlier detection: Outlier detection of prediction results should be consistent with rule-based alerts auto-generated by building systems	Accuracy of outlier detection
Computational efficiency	The design will consume the minimum required computational resources and within available computational resources.	Memory usage for implementing training
	The design will generate predictions at the same frequency as data input	Computational time required for making predictions. Will the twin require additional time?
Generalizability	The design will be able to generalize to the VAV and CU systems of different buildings	Out-of-sample test performance on data outside of CPPIB
Interpretable	The design will inform the user of the processes that occur during prediction	Availability and frequency of system logs

9.0 Conclusion

Utilizing the current MNG analytics to create a framework for a predictive maintenance tool is achievable. Predictive maintenance will create efficiencies for MNG in reducing service calls and downtime. Our proposed solution of a Digital Twin of the VAV system will improve the operational performance of the target system. The benefit of this approach is that when MNG starts collecting more time-series data and documenting service calls that align with sensor data, they can tune the Proof-of-Concept model for further accuracy and reliability. The next steps are to develop a timetable for the implementation of the design, to train the model on the VAV sensor data provided by MNG and to test the model for accuracy and reliability. Although we will initially only implement one of the proposed models for ML and predictive maintenance for MNG, this model will be subjected to iteration depending on findings from testing. In the following month, we plan to implement the model, make changes and conduct further research on the model approach and VAV system.

10.0 Project Plan

Now that we have a design plan to develop a model for a digital twin. We will amend the work plan to incorporate the clear work that needs to be done leading up to the Design Review and Critique in January 2022.

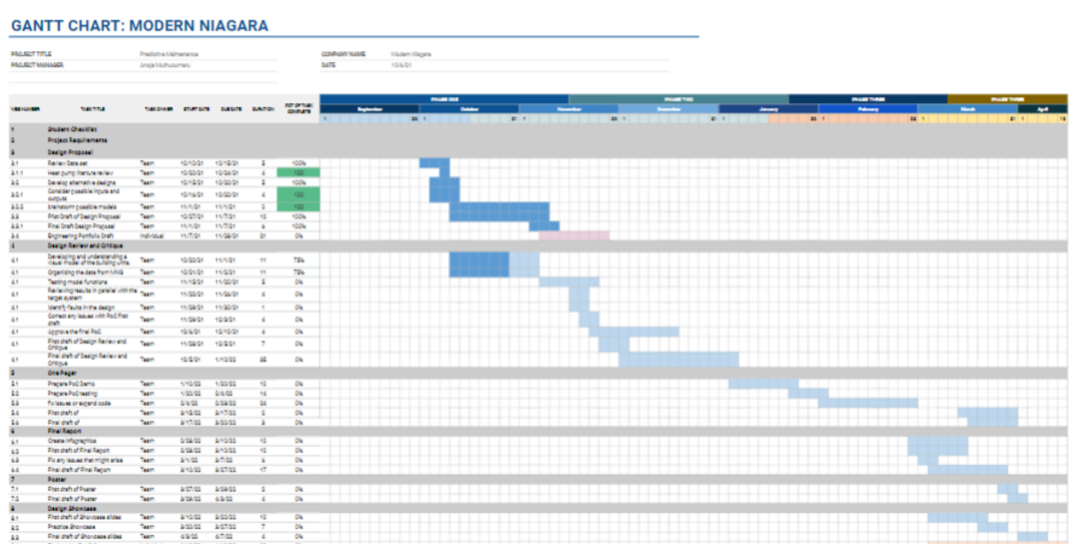
Step 1) Developing and understanding a visual model of the building units.
Step 2) Organizing the data from MNG.
Step 3) Testing model functions
Step 4) Reviewing results in parallel with the target system
Step 5) Identify faults in the design
Step 6) Understanding faults in design
Step 7) Repeat steps 1-6

Following the first iteration of design and the Design Review and Critique in January we will amend our milestones to reflect the testing and changes that need to be undertaken before the final delivery in April 2022.

10.1 Milestones

The milestones we will need to complete before January 2022, are tracked on a Gantt chart linked below. The chart will be updated to reflect the ongoing progress of the work. The chart will be available to MNG and supervisors to continue to maintain and develop transparency and trust between all stakeholders.

Click to view the full chart



10.2 Final deliverables

Our approach to the project will be iterative, requiring trial and error to identify reliable inputs and accurate outputs. MNG will be consulted throughout every phase of the work plan and will be a primary source of information for datasets and for identifying expert knowledge within the building and system. MNG will also work with the team to determine which building is most suitable for the proof of concept given the constraints and available resources. Program supervisors will offer support by providing expert pedagogical knowledge and guidance when requested or when they feel it is warranted.

Deliverable	**Deadline**
~~Student checklist~~	~~09/23/2021~~
~~Safety training complete~~	~~10/03/2021~~
~~Project requirements~~	~~10/10/2021~~
Design Proposal	11/07/2021
Engineering Portfolio Draft	11/28/2021
Design Review and Critique	01/10/2022
One-Pager due	03/20/2022
Final Report	03/27/2022
Poster	04/03/2022
Design Showcase	04/07/2022
Engineering Portfolio due	04/10/2022

Table 1: *The team course deliverables and deadlines.*

10.3 Responsibilities

The team has designated primary and secondary roles for each member, as shown in Table 2. The project managers, Anoja and Elizabeth, will be responsible for ensuring the work is done correctly and on time. The communicators, Shirley and Anoja, will be responsible for setting up meetings, sending out reminders, and communicating with the client and advisors on behalf of the team. The team leaders, Sherry and Chi, will be in charge of ensuring the team has what it needs to succeed. The editors, Elizabeth and Shirley, will provide a style guide for formatting as well as lead the final check for consistency between sections. There are also skill-specific roles for the UX and ML portion of the project. The UX design will be led by Anoja and Chi and the ML design will be led by Chi and Sherry.

Role	Primary	Secondary
Project Manager	Anoja Muthucumaru	Elizabeth Chelmecki
Communications	Shirley Zhang	Anoja Muthuucmaru
Skill Specific	UX: Anoja Muthucumaru ML: Chi Zhang	UX: Chi Zhang ML: Sherry Zuo
Team Leader	Sherry Zuo	Chi Zhang
Editor	Elizabeth Chelmecki	Shirley Zhang

Table 2: Primary and secondary roles of the design team

10.4 Strategies for Risk Assessment/Mitigation

Data preparation will be essential for understanding the design workflow. This process will be essential to identify the variables that are most valuable for identifying data gaps (blank fields, anomalies)[18]. This step will help us consider all possible inputs and outputs. This will be an iterative process where the design will be assessed for possible issues or gaps in knowledge during the modelling, training, and testing of our digital twin.

Rick Assessment Breakdown	
Risks impacting schedule/timeline	To account for the deadline, we will focus on only the VAV and CU systems.
Risk to performance	We conduct weekly scrums: Each member of the group participates in building the final product and updating the group on work completed
Reduction of harm	To reduce harm we rely on our supervisors and clients to confirm our assumptions and claims.
Control	To improve the group’s sense of control in the project we conduct meetings with the client and create transparency.

References

[1] Nejklakmb, “What is building automation,” What is Building Automation - Learning the basics. [Online]. Available: http://www.controlservices.com/learning_automation.htm. [Accessed: 08-Nov-2021].

[2] EPA. [Online]. Available: https://www.epa.gov/iaq-schools/heating-ventilation-and-air-conditioning-systems-part-indoor-air-quality-design-tools. [Accessed: 08-Nov-2021].

[3] “Variable Air Volume (VAV) systems operations and maintenance,” Variable Air Volume (VAV) Systems Operations and Maintenance. [Online]. Available: https://www.pnnl.gov/projects/best-practices/variable-air-volume-systems. [Accessed: 08-Nov-2021].

[4] "RentSafeTO for Building Owners", City of Toronto, 2021. [Online]. Available: https://www.toronto.ca/community-people/housing-shelter/rental-housing-tenant-information/rental-housing-standards/apartment-building-standards/rentsafeto-for-building-owners/?accordion=new-bylaw-requirements-from-april-1. [Accessed: 09- Oct- 2021].

[5] Security of Information Act, R.S.C. 1985, c. O-5. https://laws-lois.justice.gc.ca/eng/acts/o-5/. Retrieved on 2021-10-09

[6] Ontario.ca, "Freedom of Information and Protection of Privacy Act, R.S.O. 1990, c. F.31," 31 August 2021. [Online]. Available: https://www.ontario.ca/laws/statute/90f31/v46. [Accessed 5 October 2021].

[7] Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2015). Machine Learning for Predictive Maintenance: A Multiple Classifiers Approach. IEEE Transactions on Industrial Informatics, 11(3), 812-820 https://doi.org/10.1109/TII.2014.2349359

[8] N. Ben-Zvi, “A 2021-ready deep learning hardware guide,” Medium, 05-Dec-2020. [Online]. Available: https://towardsdatascience.com/another-deep-learning-hardware-guide-73a4c35d3e86. [Accessed: 08-Nov-2021].

[9] National Energy Code of Canada for Buildings, Canadian Commission on Building and Fire Codes, National Research Council of Canada,2017. [Online]. Available: https://publications.gc.ca/collections/collection_2018/cnrc-nrc/NR24-24-2017-eng.pdf

[10] Security of Information Act, R.S.C. 1985, c. O-5. [Online]. Available: https://laws-lois.justice.gc.ca/eng/acts/o-5/. [Accessed: 7 November 2021].

[11] Office of the Privacy Commissioner of Canada. The Personal Information Protection and Electronic Documents Act. [Online]. Available: https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/. [Accessed: 7 November 2021].

[12] Andreas Maurer, Massimiliano Pontil and Bernardino Romera-Paredes. The Benefit of Multitask Representation Learning. Journal of Machine Learning Research 17 (2016) 1-32

[13] IBM, "What is a digital twin?," IBM, 2021. [Online]. Available: https://www.ibm.com/topics/what-is-a-digital-twin. [Accessed 4 11 2021].

[14] Security of Information Act, R.S.C. 1985, c. O-5. https://laws-lois.justice.gc.ca/eng/acts/o-5/. Retrieved on 2021-10-09

[15] Ontario.ca, "Freedom of Information and Protection of Privacy Act, R.S.O. 1990, c. F.31," 31 August 2021. [Online]. Available: https://www.ontario.ca/laws/statute/90f31/v46. [Accessed 5 October 2021].

[16] A. rishnan, "Anomaly Detection with Time Series Forecasting," Towards Data Science, 2019 3 March. [Online]. Available: https://towardsdatascience.com/anomaly-detection-with-time-series-forecasting-c34c6d04b24a. [Accessed 06 11 2021]

[17]Arundo Analytics, Inc., "Anomaly Detection Toolkit (ADTK)," ARUNDO, 2019-2020. [Online]. Available: https://adtk.readthedocs.io/en/stable/index.html. [Accessed 08 11 2021]

[18]N. Sharma, "7 Tools to Build Proof-of-Concept Pipelines for Machine Learning Applications," 19 July 2021. [Online]. Available: https://neptune.ai/blog/tools-proof-of-concept-pipelines-for-machine-learning.

[19] Yi-fan Zhang , Peter Fitch, Peter Thorburn. "Deep Regressor Chain: A Hybrid Approach for Long Term Time Series Prediction" Available: http://www.c3dis.com/wp-content/uploads/2018/05/ivan-zhang-Deep-Regressor-Chain_P10.pdf

[20] Jason Brownlee, "A Gentle Introduction to Multiple-Model Machine Learning" Available: https://machinelearningmastery.com/multiple-model-machine-learning/

[21] Kenan Cem Demirel , Ahmet S ahin and Erinc Albey "Ensemble Learning based on Regressor Chains: A Case on Quality Prediction" Available: https://www.scitepress.org/Papers/2019/79328/79328.pdf