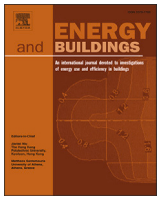
Energy & Buildings 317 (2024) 114360

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild



E-Audit: A “no-touch” energy audit that integrates machine learning and simulation

Lauren E. Excell [a], Abigail Andrews [b], Rishee K. Jain [a,*]

[a] Urban Informatics Lab, Stanford University, 473 Via Ortega, Stanford, 94305, CA, USA
[b] Edward J. Bloustein School of Planning and Public Policy, Rutgers, The State University of New Jersey, 33 Livingston Ave, New Brunswick, 08901, NJ, USA

ARTICLE INFO

Dataset link: E-Audit: A ‘‘no-touch” energy audit that integrates machine learning and simulation (Original Data)

Keywords:
Retrofit analysis tool
Machine learning classification
Building energy modeling
Electricity load matching

ABSTRACT

Identifying potential pathways for enhancing the energy efficiency of our building stock is a clear and compelling pathway to decarbonization. However, doing so requires resource-intensive building energy audits that often require an engineer to make on-site inspections and analyses. In this study, we propose E-Audit, a “no touch” energy audit that combines physics-based simulation methods with data-driven classification methods to identify potential sources of inefficiencies in a building using only hourly electricity data. E-Audit first utilizes a reference building model to create a synthetic library of potential efficient and inefficient hypothetical buildings across 15 building features each with an associated hourly energy usage signature. Next, E-Audit utilizes classification methods to match a given building to the closest time series and therefore identify efficiency opportunities. We tested E-Audit on a data set of 1,323 school buildings of which 325 received retrofits. Results indicate that our E-Audit accurately identifies inefficiencies 91 to 99 percent of the time for plug load, equipment schedules, and non-geometric characteristics such as window construction and insulation. The E-Audit underperforms on boiler efficiency identification as this feature is highly dependent on boiler fuel type and predictions are solely based on electricity data; however, the algorithm effectively distinguishes between electricity and natural gas as fuel sources and therefore was able to predict boiler retrofits with 97 percent accuracy. The method does not predict control system retrofits well, with prediction accuracy below 40 percent for these features across all classification algorithms. We find that a machine learning-based classification method outperformed Euclidean distance matching, with kNN balancing accuracy and efficiency. When applying cost data, we also find that our E-Audit overestimates recommendations for high capital cost retrofits and underestimates inexpensive ones pointing to the need for future work that encompasses cost efficacy into the method. Overall, our E-Audit demonstrates the potential to streamline building decarbonization by improving accessibility, cost-effectiveness, and scalability of building energy efficiency evaluations.

1. Introduction

The built environment is at a transformative moment, where advanced data-driven methods are reshaping energy efficiency analysis. Traditionally, evaluating the energy performance of buildings necessitated meticulous physical inspections such as audits, often resource-intensive and time-consuming [1]. In order to address energy inefficiencies at scale, it is vital to create accessible, rapid methods for scoping energy retrofit projects for cases where there are limited resources, expertise, and time.

To gather information on the constructions of a building, architecture, engineering, and construction (AEC) professionals rely on blueprints or construction plans, which may be misplaced over time or not updated after retrofits or renovations. Architectural site analysis or energy audits are necessary to confirm information or deduce missing information about the constructions and systems in a building before conducting retrofit analysis [2]. Additionally, some governments have energy audit mandates that require on-site evaluations of energy performance to promote energy efficiency measures [3]. This information is necessary when modeling the energy performance of the building, forecasting electricity load, and identifying which retrofits may be necessary to bring the building up to current code standards. While a site visit may not be an unmanageable task for a single building, performing this analysis at scale would require several months of labor.

* Corresponding author.
E-mail address: rishee.jain@stanford.edu (R.K. Jain).

https://doi.org/10.1016/j.enbuild.2024.114360
Received 21 February 2024; Received in revised form 25 April 2024; Accepted 29 May 2024
Available online 4 June 2024

Deciding which retrofits to install in a given building has traditionally been an expensive process that requires an onsite energy audit by an energy expert, costing up to 10% of the annual utility bill per building and taking months to complete [4,5,3]. An energy audit is performed onsite by a consultant, who then recommends which retrofits would be necessary. In a commercial building, a facility manager might do a cost-benefit analysis to make the final decision [1]. In an effort to expedite and automate this process, new methods have been developed to identify which buildings require retrofit, suggest which retrofits to perform, and quantify the energy savings from these retrofits. Benchmarking methods can be used to identify the worst-performing buildings in a portfolio, which would be most in need of retrofits. This method requires building information such as floor area, year built, number of occupants, operational schedules, and annual energy use data by fuel type [6]. While this simplifies the process of identifying which buildings need retrofits, it does not eliminate the need for an energy audit – decision makers would then need to decide which retrofit to install [7]. For decision-makers with limited resources and buildings with limited available data, it can be prohibitively expensive to collect the information required to recommend retrofits. To streamline this task, we propose a simulation-based machine learning approach to classifying building feature information using limited data inputs.

This research defines and tests a methodology for identifying inefficient performance. It also proposes retrofit solutions using a minimal amount of information about the physical characteristics of a building. We propose E-Audit, a “no touch” energy audit methodology that leverages a synthetic library of potential efficient and inefficient building energy models (BEMs) alongside data-driven classification methods to identify inefficiencies of building characteristics. By combining parametric BEM simulation, data-driven electricity load classification, and machine learning, we present a method capable of discerning the energy performance of a building’s physical characteristics using data inputs comprising of electricity load profiles, climate zone, square footage, and building use types. The approximated physical characteristics indicate where the building is inefficient and what retrofits might be necessary. Under this methodology, only the building use type, climate zone, square footage, and electricity meter readings would be required to estimate physical building characteristics using a parametrically-generated database of building energy simulations, as explored by similar methods discussed in Section 2.3. This would allow for faster retrofit recommendations and expedited building archetyping. We demonstrate this methodology using a case study on a data set of primary and secondary school buildings.

This method is useful for applications such as retrofit recommendation, policy analysis, energy auditing, and energy modeling. Understanding the performance of physical features of their buildings allows building managers to focus on the most impactful features to target for retrofits. Similarly, this knowledge can be useful for energy auditors to know which features to focus their efforts on when auditing a building. Energy modelers could use the information from these classifications to set building parameters in urban-scale energy models. Policymakers could use this method to gain insights into a selection of buildings to improve future building energy codes. In particular, this method is useful for identifying inefficiencies in building features, which can be used to recommend retrofits with limited data inputs.

2. Background

In this section, we overview existing methods for evaluating the energy performance of buildings and current tools for analyzing energy efficiency retrofits, demonstrating why the method in this paper is being proposed. We then review previous studies that have pioneered the use of load matching methods to predict information about buildings, which are used as the foundation for the method proposed in this paper.

2.1. Data-driven methods for energy performance prediction

Data-driven methods have been used to predict building types and energy performance using visual data, such as satellite imagery, Google Streetview, OpenStreetMap, and Microsoft Footprints. Atwal et al. use a supervised learning algorithm on OpenStreetMap to classify buildings as residential or non-residential archetypes, which can help approximate building information when unknown [8]. Remote sensing data has been used to predict building energy consumption using computer vision [9]; similarly, remote sensing data and street view data have been used to estimate building energy efficiency using a deep learning model and k-means clustering [10]. These methods help provide fast, non-intrusive estimates of building energy consumption, but the black-box nature of these machine learning methods makes it difficult to identify the physical building characteristics that are driving energy consumption. Purely data-driven approaches neglect the underlying physics of building systems and focus solely on the statistical relationship between inputs and outputs. Therefore, there are limits to the applications for data-driven approaches that neglect building physics. Additionally, these data-driven methods require robust training data to model changes to building systems, which is often infeasible to collect.

2.1.1. Analysis of electricity load data

Increasingly widespread adoption of smart meter technology has increased the availability of electricity meter readings at hourly time intervals. Data at this fine temporal scale allows for load shape pattern identification using machine learning methods, which can be useful for understanding the performance of a building, how it interacts with the grid, and detecting performance anomalies [11]. Current methods that analyze electricity loads focus on predicting or forecasting building energy consumption, predicting building type for archetype assignment, or clustering buildings by performance to either glean customer information, assign building archetypes for energy modeling, or as a first step in retrofit analysis. These methods use load shape analysis, clustering algorithms, and machine learning to provide insights into the building without requiring large amounts of data about the building’s non-geometric characteristics. Some of these methods can predict which appliances are being used within a building based on the electricity load signature of that appliance [12]. Electricity data has also been used to provide insight into occupant behavior, circumventing the need for time-intensive and costly surveys [13].

Load profiling methods can be used to detect performance anomalies throughout the year and help with the ongoing commissioning of building systems. Detecting anomalous energy consumption helps building managers detect issues in building systems and can be an effective ongoing commissioning strategy when used in combination with energy management systems [14] Disaggregating the load data allows for the separation of typical demand and intermittent fluctuations, which can also be used to support power grid planning [11,15]. Li et al. 2021 have used time and frequency domain load profile analysis to enhance energy modeling inputs and calibration, contributing to a positive feedback loop whereby load profiling leads to more accurate retrofit modeling analysis [16].

Load shape analysis can be used to identify electrical appliances present in buildings. Non-intrusive load monitoring can be used to extract features and shapes from the electricity load to match with the electrical appliance’s load [17]. This can be helpful for identifying information about plug loads, but is limited in its application to building features that directly consume electricity. Similarly, electricity data disaggregation can be used to determine the performance of specific home appliances [18]. This can be useful for scheduling loads in a building to optimize energy usage. However, many of the building characteristics that influence energy performance are not direct consumers of energy in the way that plug loads and home appliances are. Therefore, load matching methods need to identify patterns on different time scales to identify the energy performance of physical characteristics. Weekly,

monthly, and yearly time scales exhibit different seasonal patterns that can demonstrate the efficiency of characteristics such as building envelope, equipment, and lighting systems.

Clustering methods can identify groups of customers with similar characteristics, thereby identifying groups of buildings that should be audited for energy efficiency measures [19–24]. These clustering methods are good at identifying whole-building inefficiencies and energy performance, but they do not determine specific characteristics of building systems. An energy audit would still be needed to determine which building systems require retrofits.

These methods are limited to aspects of the building that directly affect the electricity load, such as plug loads and equipment or occupant schedules. While this is helpful for identifying potential changes that can be made to a building to save or shift electricity consumption, this does not cover the full range of potential retrofits. There are building features that do not directly affect the electricity load that need to be included in energy audits and retrofit decision-making because of the high impact that they can have on energy consumption, such as insulation and infiltration.

2.2. Retrofit analysis tools

Several tools have been created to recommend retrofits, or energy efficiency measures (EEMs), to buildings; however, these often require extensive data inputs [25]. Retrofit analysis tools have been created in the public sector, by utilities, and in the private sector, and span from detailed energy models to statistical analysis using regression to user-friendly softwares built on a database of simulations [26]. Using reference buildings, retrofit designs can be modeled on an entire building stock to assess the savings potential [27]. Energy modeling is the standard method of evaluating changes to a building using software such as EnergyPlus, however this method requires expert knowledge to input data and run the simulation. Benchmarking using statistical methods is a more user-friendly approach that does not require expert knowledge to understand the model inputs. The statistical method estimates energy performance using a regression model where variables are the design, operation, and climate of a building. For example, benchmarking with EnergyStar Portfolio Manager uses a regression model to classify energy efficiency amongst peer groups of buildings [6]. Statistical methods have also been created to assess the energy savings potential of large building stocks using multiple linear regression on aggregate data such as building type, year built, floor area, and number of occupants [28,29]. These methods help identify which buildings should be prioritized for retrofits, but the lack of training data limits their ability to estimate savings from specific retrofit types. These models have high potential for impact in data-rich environments, however they cannot easily be generalized to other geographies because they lack the physics-based relationships between building systems inherent in energy models.

In response to the need for an easily accessible retrofit assessment tool, Hong et al. created the Commercial Building Energy Saver (CBES) [30]. CBES is an energy retrofit analysis tool for identifying which retrofits should be installed in a building. This tool is helpful for performing benchmarking, load shape analysis, and retrofit analysis on a building; however, the data inputs for this tool can be extensive, and require knowledge of the building geometry, construction, internal loads, HVAC, schedules, lighting, and equipment. The Building Efficiency Targeting Tool for Energy Retrofits (BETTER) tool was also created to improve public access to energy efficiency strategies by analyzing electricity load data to suggest changes to building operations [25]. Similar softwares have been created for retrofit analysis such as AutoBEM [31], Excel/MATLAB tools [32,33], and machine learning optimization [34–36]. These data-driven approaches face the same limitations due to extensive data inputs. In cases where a building owner does not know this information and does not have a building manager or energy auditor that can aid in collecting this information, it can be very time intensive, difficult, and error prone to input this data [37]. Therefore, a tool that can propose potential retrofits with minimal building information and data inputs can help building owners to get a preliminary estimate of which building systems require a retrofit.

For large-scale retrofit analysis on a portfolio of buildings, urban building energy models (UBEMs) have been used to assess the impact of retrofit strategies on groups of buildings. UBEM tools such as CESAR [38], UMI [39], CityBES [40], URBANopt [41], City Energy Analyst [42], and DUE-S [43] rely on building archetypes to determine the non-geometric information about a building, which is often unavailable at large scales. The energy performance of the actual building features is therefore not reflected in the simulation, which relies on assumed information about certain building types. Although electricity data for individual buildings is not publicly available due to data privacy concerns, government agencies and utilities that do have access to this data would be able to benefit from a method that improves building archetype definition through using electricity data to define non-geometric characteristics.

2.3. Load matching on large-scale simulations

Large simulation databases have been built before to make the benefits of simulation more accessible in cases where expertise and experience with building energy modeling are unavailable. Roth et al. built the DEnCity database of prototype simulations to aid end users in creating an energy simulation for their building leveraging the expertise on inputs, parameters, and outcomes contained in the building database [44]. This simplifies the energy modeling process for users who are not familiar with input requirements and realistic outcomes. We propose leveraging a similar large-scale database to aid end users with identifying inefficiencies in their buildings based on their electricity consumption data.

To perform the matching of actual electricity loads to our simulated building database, we build upon existing load matching methods. Load matching methods are used by utilities to assign consumers to building types by typical usage patterns, which helps predict energy demand; in research, load matching has been proposed to assign building type, which helps identify archetypal physical characteristics needed for BEMs and UBEMs [45,46]. These methods typically require knowing some physical characteristics, however recent research has attempted to perform classification with only the electricity consumption time series. Bass et al. and Miller and Meggers both test load matching methods to classify utility data by use type (e.g., school building, office building, hospital) [45,46]. They take advantage of utility data for which use types are known and match the load of an “unknown” building to determine its use type. Garrison et al. use Energy Plus simulations of building archetypes and match these with square footage normalized utility data using Euclidean distance matching to assign a building archetype to the unknown building [47].

There are two types of load matching methods: direct load matching, and time series features with a machine learning classifier. Direct load matching uses distance methods such as Euclidean distance and dynamic time warping (DTW) to capture similarities in load profiles. Euclidean distance matching computes the distance between each point in the time series at a given time. Dynamic time warping attempts to capture patterns in the time series and accounts for a lag between two time series, thus matching loads more on patterns than the magnitude of consumption. Dynamic time warping accounts for the trends in the data that may be occurring at different points in the time series, and attempts to match points that have similar patterns, hence warping the time [48]. This added complexity makes the calculation of distances much more time intensive.

Machine learning methods for load matching require extracting temporal features with statistics, regression models (e.g., time-of-the-week and temperature, change point model, seasonal and trend decomposition), or temporal patterns. These features can then be characterized

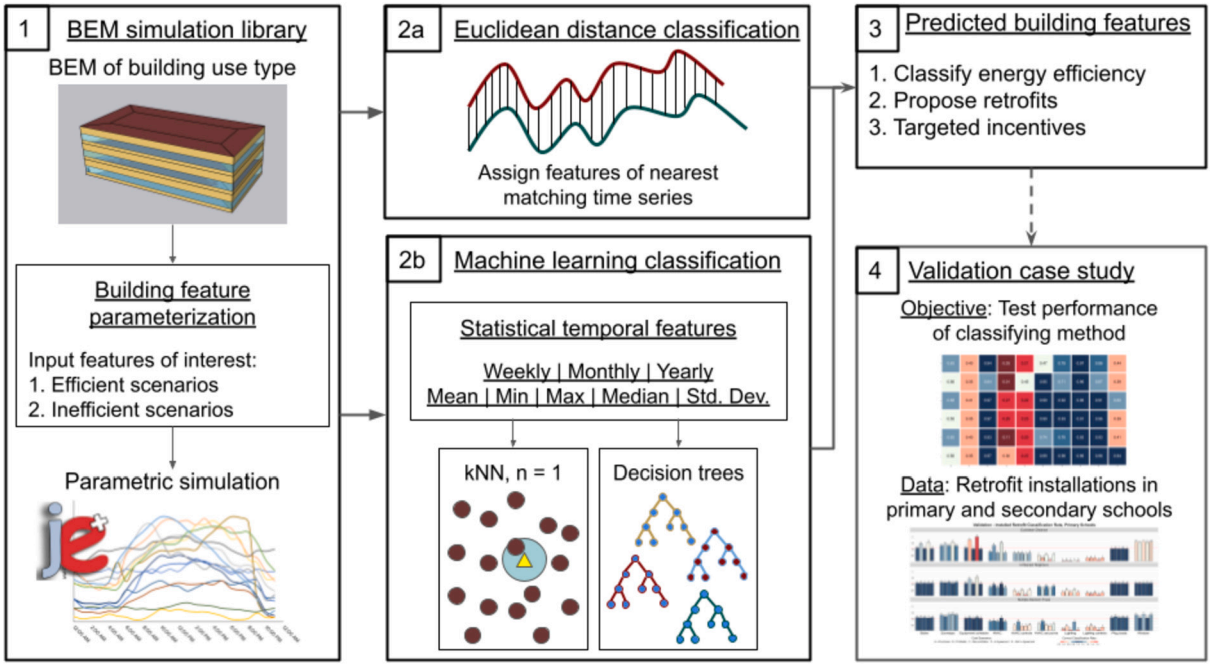


Fig. 1. Overview of the methodology for creating and matching empirical data to the BEM simulation library.

with a supervised machine learning classifier (i.e., random forest, k-nearest neighbors, extreme boosting, etc.). Bass et al. found that Euclidean distance was the fastest and most accurate among these methods, however their data skewed heavily toward one building use type, which was suited to Euclidean distance matching [45]. Miller and Meggers found that, over the baseline model (random assignment of use type), random forest classification improved primary use type prediction by 45.6%, operations type prediction by 63.6%, and building performance class prediction by 24.3%; they also found that pattern-based temporal features were significant indicators of types of behaviors [46].

These studies have laid the groundwork for using machine learning methods to predict building characteristics based on energy consumption time series. They have classified building use type, relative performance, or operational use patterns using data from a utility for which building metadata is known. We further this research by identifying non-geometric physical characteristics (i.e., lighting power density, fuel type, constructions, etc.) of buildings where the use type is known.

The main research gap that this paper fills is identifying non-geometric characteristics of a building using limited data inputs, and recommending retrofits for inefficient building systems without extensive data gathering and input. Previous papers have either been able to identify building type [46,45] and appliance information [18,17,12] using electricity load data or approximate energy efficiency using remote sensing or street view data [9,10]. Our proposed method will be able to identify building characteristics that do not directly influence electricity load and provide insight into the efficiency of those building characteristics. Unlike other methods of retrofit analysis, this method does not require extensive knowledge of building systems or gathering extensive information about building systems and constructions. Models that combine physics-based and data-driven approaches take advantage of the benefits of both models, accounting for the underlying physics of changing building systems through simulation and improving computational speed with machine learning. Through parametric energy simulations, we address the issue of robust training data requirements by creating a comprehensive database of a building under manifold retrofit scenarios. By matching electricity data from an actual building to a building in our database, we can predict the physical characteristics of that building.

3. Methodology

In this section, we overview the load matching methodology that was used to create a library of building energy simulations and match those electricity loads to real buildings to approximate physical characteristics, as seen in Fig. 1. We then describe how this methodology was tested on a simulated data set of 73,728 primary and secondary school buildings, and then how this was validated using a case study of primary and secondary schools in California.

3.1. Load matching methodology

We propose a novel load matching methodology for recommending energy efficiency retrofits based on minimal information on physical aspects of the building. This methodology consists of two parts: simulating inefficiency scenarios for building use types (i.e., restaurant, office building, hospital, etc.), and matching the real building's electricity load to the simulated load. Using parametric energy simulation, key building parameters affecting energy efficiency are varied and simulated to create a library of buildings of varying degrees of energy efficiency. This library is created using standard building energy model archetypes, such as the Department of Energy (DOE) reference buildings. Local weather data or typical meteorological year weather data are used for the relevant ASHRAE climate zone to approximate how a building will perform under the typical weather in a region. This library is then used as a reference for real buildings of the same use type and climate zone. Then, using a load matching method, the electricity use data from the real building is matched to the electricity use data from a building in the simulation library. This match identifies the most likely building features that exist in the real building without extensive energy auditing and data collection. The only data inputs required for this method are the building use type, climate zone, square footage, and electricity use data for a single year, which can be requested from the local utility.

Through initial testing on a subset of the data to compare the time intensity of distance matching (DTW and Euclidean distance) and machine learning methods (kNN, decision trees, random forest), the DTW method proved to be prohibitively time intensive and inaccurate in comparison to other methods. The dynamic time warping distances be-

tween the actual load and simulated load were larger than the Euclidean distance for all of the actual time series, indicating that the Euclidean distances are a closer match. Euclidean distances were computed much more quickly than DTW distances; therefore, we chose to focus on the Euclidean distance matching method for the remainder of the analysis. Additionally, random forests take longer to compute than decision trees; given that this method will be scaled to a large simulation database that will require longer training times, we chose to test the faster decision tree method. When scaling this methodology to many simulations, computational time is an important factor.

In this paper, we compare Euclidean distance and two machine learning methods for accuracy and computational resources required. Euclidean distance is the most straightforward method for matching loads, as the distance between the time series can be directly calculated. Before calculating the Euclidean distance between time series, missing values were imputed using the median value of the time series.

For the machine learning methods, we analyze a k-nearest neighbors approach, which has been used in prior research [46,45], and compare this to a proposed multiple decision trees approach where a tree is created for each physical building parameter rather than for the building as a whole. Under the multiple decision trees method, the optimal match for each parameter would be found, rather than the global optimal match between time series.

To perform the machine learning classification, we calculate time series features for each building. These time series features include the mean, minimum, maximum, median, and standard deviation of the hourly kWh per square foot readings for each year, month, and week in the time series. This allows us to capture monthly and weekly trends in load characteristics while reducing the features to be classified for one year from 8,760 to 325. Doing so helps to reduce the computational time required to perform the classification.

The k-nearest neighbors algorithm, also known as kNN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. The number of neighbors for the kNN algorithm was set to one, as we are looking for the closest matching simulation to identify the corresponding building parameters.

Decision Trees are a non-parametric supervised learning method used for classification and regression. The goal of decision trees is to create a model that learns simple decision rules inferred from input features to predict the value of a target variable. Because a decision tree was created for each building parameter, hyperparameters were tuned for each individual tree using Grid Search cross validation. The tree parameters that were tuned in this study were maximum tree depth, node splitter criterion (i.e., random or best), minimum samples at the split, and minimum samples at the leaf.

3.2. Testing matching methods on simulated data

This section reviews the methodology for testing kNN, decision trees, and Euclidean distance methods on simulated electricity data for primary and secondary schools. The training-testing split of our simulated data was 80-20. To test this methodology, we have completed 73,728 BEM simulations for two building types using Department of Energy (DOE) reference buildings: primary schools and secondary schools. Each simulation produces one year of hourly electricity data, created through parametric analysis in jEPlus [49]. These simulations represent 15 building features that can take on either an efficient or inefficient value. These features represent inefficiency scenarios that might be present in school buildings. For example, parameters might represent inefficient light bulbs or an inefficient boiler. See Appendix Table A.2 for a list of all building parameters and possible values. Due to computational limitations, we simplified our parametric analysis to 15 parameters that may take on two values. In practice, this methodology can be used to analyze more parameters with more values depending on the computational resources available.

To account for the varying sizes of buildings, we normalize the electricity readings by building area. This allows the matches to focus on how efficient the buildings are rather than the magnitude of consumption. We then match electricity load data from the test set to the electricity data from the training set to predict the parameters in the test set of buildings. If the feature in the matching training building belongs to an inefficient scenario (takes on an inefficient value rather than an efficient one), then we predict that the feature is inefficient in the test building. Therefore, we also recommend that this inefficient building feature should be considered for an energy efficiency retrofit.

The algorithms are predicting on the exact parameters that exist in both the training and testing data sets. Whereas in real world applications, the buildings that we will be predicting on do not have information available on the exact parameters that the data is trained on.

For the validation of these methods, we will be approximating the building parameters based on the retrofits that were recommended. Essentially, we will be able to predict whether a building feature was efficient or inefficient, rather than the specific value of a parameter.

3.3. Validation of methods using empirical retrofit data

This section describes the methods and metrics used to evaluate performance on the validation data, and reviews the results of validation against recommended and installed retrofits. To validate this methodology, we collected metered hourly electricity data for primary and secondary schools in California from 2013-2017 through data requests to Pacific Gas and Electric (PG&E). Data on the physical building characteristics (e.g. HVAC system, lighting power density, window U-factor, etc.) for these schools are not collected or readily available to use for validation. To circumvent this data limitation, we obtained information on energy efficiency measures installed in these schools through a state-funded energy efficiency program. This retrofit information is available for schools that received funding through the California Proposition 39 program (Prop 39) to install energy efficiency measures. The hourly electricity data set from PG&E contains 1,323 schools. Of these schools, retrofit information was available for 325 of them. The actual load readings from these schools will be matched to the simulated scenarios to approximate the physical building characteristics of the schools and predict which retrofits were recommended and which were installed. We will then compare the predicted retrofits to those that were recommended and installed through the Prop 39 program. We predicted building features both “before” and “after” retrofit installation using 2014 and 2017 data. A majority of the retrofits were installed in these schools in 2015.

The case study performance was evaluated using accuracy, precision, and recall. These metrics use true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) to assess the performance of machine learning algorithms in classifying positive cases correctly. In this case, a positive case is one in which a retrofit is recommended and installed. A false positive indicates that the algorithm predicted a retrofit but the Prop 39 program did not recommend or install a retrofit. A false negative indicates that the algorithm did not predict a retrofit, but there was one recommended or installed. Accuracy is the proportion of correct predictions to the total number of predictions:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision answers the question, “what proportion of positive cases were correct?” and is calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall answers the question, “what proportion of actual positives were identified correctly?” and is calculated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Table 1
Retrofit recommendations were determined by the value of the building features; if a building feature was predicted to be inefficient, a retrofit was predicted.

Retrofit	Building feature	Value
Boiler	Fuel type OR Boiler efficiency	Natural gas —— 0.6
Envelope	Roof insulation conductivity OR Infiltration rate	0.07 $W/m-k$ —— 0.0015 $m^3/s-m^2$
HVAC	Fuel type	Natural gas
HVAC Controls	HVAC Schedule	Always on
HVAC set points	Cooling set point OR Heating set point	Too low —— Too high
Lighting	Lighting power density	Based on room size: higher of 2 values
Lighting controls	Lighting schedule	On overnight
Plug loads	Heat gains from technology	20 W/m^2
Equipment schedule	Equipment schedule	Always on
Window	U-factor OR Solar heat gain coefficient	2 W/m^2-K —— 0.6

3.3.1. Retrofit recommendations

Using the 2014 validation data, we predict which retrofits are needed to improve energy efficiency, which we refer to as retrofit recommendations. These predictions are compared to the retrofits that were approved for installation. This allows us to determine the correct classification rate for our retrofit recommendations. We recommend a retrofit based on our predictions of the building features. See Table 1 for how the building features were used to determine which retrofits were installed. A binary value is given for a retrofit based on whether it is recommended (1) or not (0). A retrofit recommendation is considered a positive case for the purposes of calculating accuracy, precision, and recall.

3.3.2. Retrofit installations

We take the difference between our "before" and "after" retrofit predictions to determine what we predict was installed in the building. This is then validated against what retrofits were installed in the building. The difference between the "before" and "after" predictions takes three possible values. If the value is 0, then we predict that no retrofit was installed. If the value is 1, then we predict that a retrofit was installed, because energy performance improved over time. If the value is -1, then the building got more inefficient over time, and we predict that a retrofit was not installed. We then calculate accuracy, precision, and recall using this predicted value and the actual installations.

To account for how costs affect retrofit decision making, we integrate a cost heuristic to compare validation cases. The Prop 39 program used the savings-to-investment ratio (SIR) to determine whether the project was profitable and should be invested in. Using the cost data from the approved projects of the Prop 39 program, we determine the median costs and median net present value for each energy efficiency measure (EEM) category. Net present value (NPV) and cost are used to calculate the SIR. A predicted SIR for each school is calculated based on the retrofits that we would recommend.

$$SIR_{school} = \frac{\sum NPV_{EEM}}{\sum Cost_{EEM}} \quad (4)$$

This predicted SIR is used to determine whether the project would be *profitable* or *non-profitable*. Profitable projects have an SIR ≥ 1.01; non-profitable projects have an SIR < 1.01. The predicted SIR is then compared with the actual SIR of the school site from the Prop 39 data to determine whether our predicted SIR is *in agreement* or *not in agreement* with the actual SIR. These four metrics (*profitable*, *non-profitable*, *in agreement*, and *not in agreement*) are used to segment the validation data to gain further insights on the algorithm's predictive abilities. We refer to the five scenarios created using these metrics by using the letters A, B, C, D, and E. Scenario A represents all schools, B represents *profitable* schools, C represents *non-profitable* schools, D represents *in agreement* schools, and E represent *not in agreement* schools. The decision to install retrofits is complex and often accounts for more than the energy efficiency of the building feature. Because cost is a major factor in decision-making, we account for the influence of cost using these four metrics.

4. Results and discussion

4.1. Performance on test data

For all three classification algorithms tested, the test classification rates for 13 out of 15 building features are over 0.75, with 6 or more of these being very close to one for all three algorithms. The test classification rates for each classification algorithm, separated by building type, can be seen in Fig. 2. The decision trees algorithm performed the best across all building features, with 14 out of 15 features having a correct classification rate of 0.90 or above. The boiler efficiency test rate using decision trees was 0.703 for secondary schools and 0.706 for primary schools, performing 20% better than kNN and Euclidean distance for these features. The kNN algorithm had a test classification rate over 0.90 for 11 features in primary schools and 8 features in secondary schools. The lowest test classification rates using the kNN algorithm were for boiler efficiency, lighting in small areas, infiltration rate, and equipment schedules, as seen in Fig. 2. In secondary schools, the kNN algorithm also had lower test correct classification rates for window U-factor (0.77), classroom lighting (0.87), and roof insulation (0.87). The Euclidean distance algorithm had a test classification rate of over 0.90 for 12 features in primary schools and 9 features in secondary schools. The lowest test classification rates using Euclidean distance matching were for boiler efficiency, lighting in small areas, and infiltration rate (see Fig. 2). Similarly to kNN, in secondary schools, the Euclidean distance algorithm also had lower test classification rates for window U-factor (0.83), roof insulation (0.86), and classroom and small area lighting (0.89). In a similar study, Michalakopoulos et al. use a physics-informed DNN to predict envelope performance based on general building information and monthly heating energy consumption data [50]. Their model struggles to predict roof and window U-factors, with R-squared values of 0.05-0.09. The thermodynamics of heat gains and heat losses through the building envelope are complex and dependent on outdoor weather such as air temperature and wind speed, therefore making them difficult to predict without ground-truth weather data inputs.

All three algorithms are able to correctly classify most of the building features that were specified. This is because we are predicting on a very large test data set and these algorithms are able to be trained on nearly all possible combinations of building features of interest. Notably, all algorithms performed relatively poorly on predicting boiler efficiency compared to other metrics. This may be due to the fact that boiler efficiency is tied to the type of fuel consumed by the boiler. Natural gas boilers have more heat loss than electric boilers or air source heat pumps; conventional natural gas boilers have thermal efficiencies around 75% whereas electric boilers have thermal efficiencies around

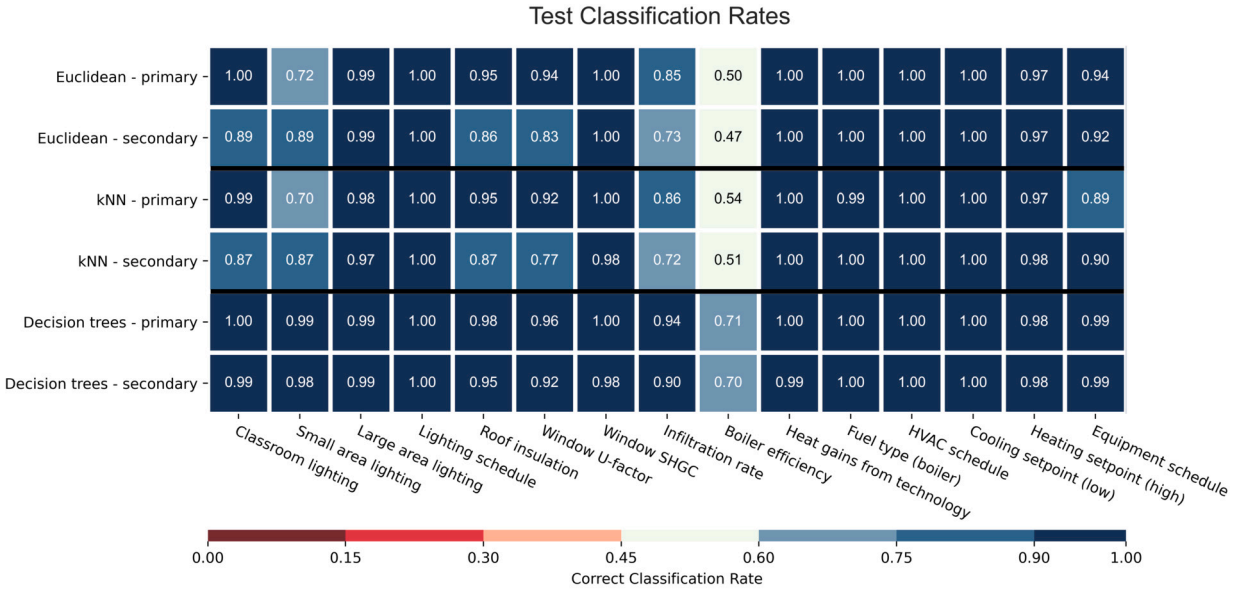


Fig. 2. Test performance of matching methods on primary schools and secondary schools. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

99% [51]. Predicting the efficiency of a natural gas boiler using the electricity data is difficult because we are not using natural gas consumption data to predict performance. Westermann et al. used clustering algorithms to predict the heating system type (i.e. heat pump, gas furnace) and found that each cluster has around 75% of buildings with the same heating system type [12]. In our case study, predicting the boiler efficiency is akin to predicting the heating system type, given that the boiler efficiencies vary depending on the heating system type. Therefore, the predictive performance for heating systems in Westermann et al. is similar to our boiler efficiency prediction rates of around 70%.

The infiltration rate and the lighting power density in small rooms had a 10-20% lower test classification rate for the Euclidean distance and kNN algorithms than decision trees. The difficulty with predicting the infiltration rate is likely due to the indirect effect of infiltration on electricity consumption. Infiltration rate affects the thermodynamic balance of a room, thus affecting the internal heat gain loads placed on the HVAC system. It's possible that the effect of these parameters was masked or picked up on through the identification of related building parameters. The lighting power density in smaller rooms is likely masked in the overall electricity load profile of the building by the lighting consumption in medium and large rooms, therefore the algorithms likely struggle with disentangling the efficiency of those rooms from the others.

The algorithms are able to correctly identify the performance of plug loads (i.e., technology) and equipment schedules, which are also identifiable through previously established methods such as non-intrusive load monitoring. Therefore, this method of load matching using a simulated building database is able to meet and surpass current capabilities of identifying building information available through previously established methods.

4.2. Performance on validation case study

The performance on the validation data set is worse than the test set for most metrics (see Fig. 3), with exceptions for plug loads and equipment schedule predictions, which perform relatively similarly in testing and validation. A discrepancy in performance between test and validation is to be expected because the test and validation cases are based on different data sets. Test performance is based on whether the algorithms are able to classify the building features to agree with the building features data - the training and testing data are both subsets of the simulation database. In the validation case, the algorithms classify the building features as inefficient or efficient. An inefficient classification is used to predict a retrofit, which is then compared to the retrofit data. This is not a direct comparison, which can result in error due to complex factors affecting retrofit decision making.

The algorithms all show poor performance when predicting most of the retrofits that were recommended by the Prop 39 program (see Fig. 3), except plug loads and equipment schedules. This suggests that other factors played into the Prop 39 retrofit decisions other than energy performance. Other studies that are designed to predict retrofit strategies use energy audit information as input features, such as the status of the HVAC system and status of the envelope, and are able to predict the retrofit strategy with up to 75% accuracy, because information about these building features are inputs into the machine learning model [52]. Our machine learning algorithms do not take information about the performance status of the building features as an input but rather predict the performance of features based on electricity data alone and use the performance as a direct indicator for retrofit strategy. Fig. 3 shows the prediction accuracy of all three matching algorithms on the two school types. The algorithms were able to accurately predict whether equipment schedule and plug load retrofits would be recommended, with all of the algorithms showing similar accuracy for plug loads and the machine learning algorithms showing 30% higher accuracy for equipment schedules in secondary schools. The equipment schedule and plug load recommendation accuracy rates were over 0.90 using the kNN algorithm for both building types. The Euclidean distance algorithm had an accuracy of over 0.91 for plug loads in primary and secondary schools, and an equipment schedule accuracy of 0.86 in primary schools. The decision trees algorithm also had high accuracy rates for equipment schedules and plug loads; the decision trees had an accuracy of over 0.94 for equipment schedules and plug loads in secondary schools and over 0.83 for primary schools.

Lighting and controls retrofit recommendations are consistently inaccurately predicted by all classification algorithms. When considering precision (Fig. A.6 in the Appendix), kNN and Euclidean algorithms had high precision for lighting (≥ 0.99) and lighting controls (≥ 80) retrofit recommendations, meaning that the positive cases (in which a retrofit was recommended) were correctly identified. However, the recall for these two building features were relatively low (Fig. A.7 in the Appendix). For example, lighting recall was less than or equal to 0.62 for both machine learning algorithms and less than 0.05 for Euclidean.

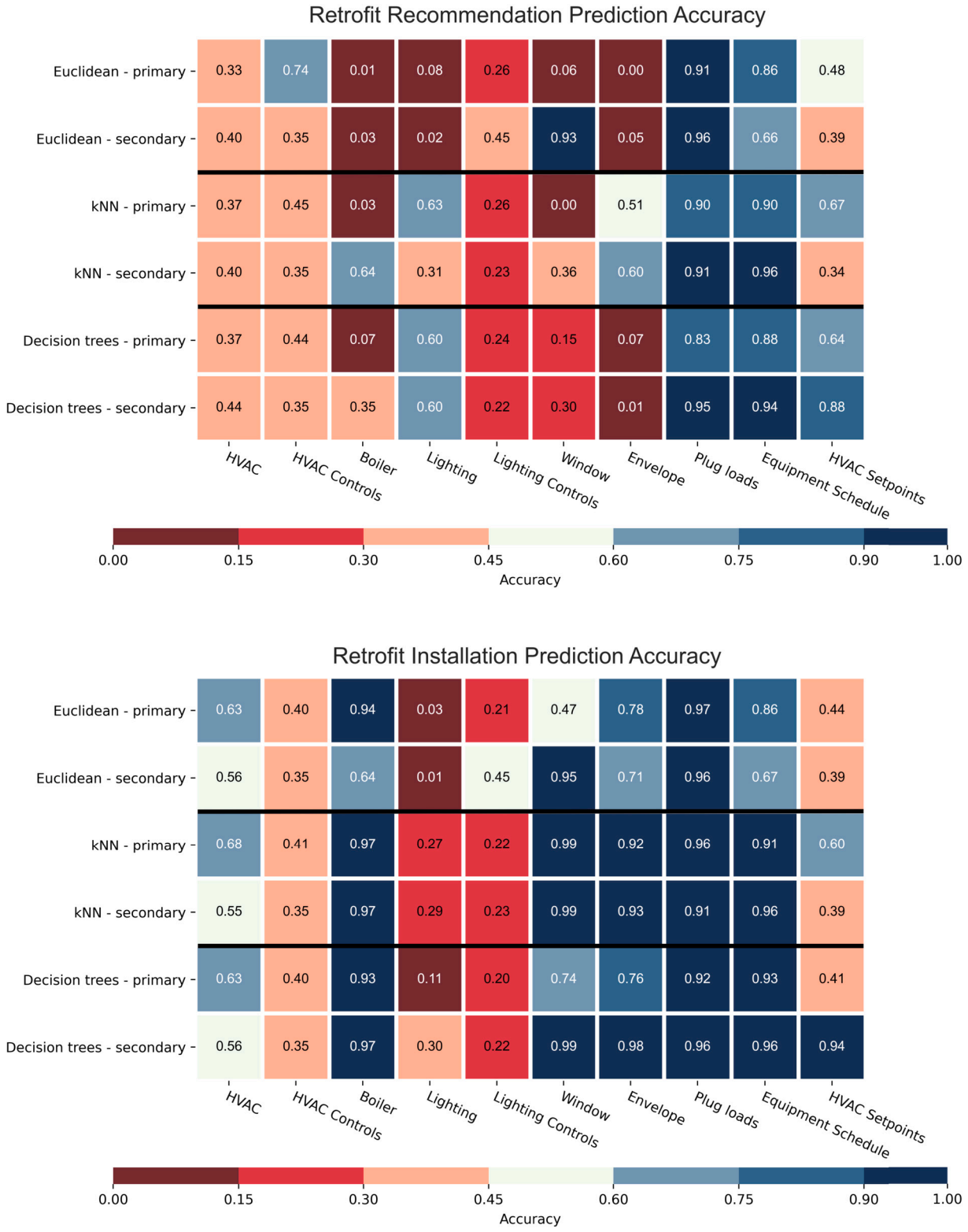


Fig. 3. Prediction accuracy of retrofit recommendations (top) and installation (bottom) for all primary schools and secondary schools. Retrofit installations demonstrated improved performance over retrofit recommendation predictions.

This means that there were not a high proportion of actual positives that were identified correctly, therefore the classification algorithms are not recommending these retrofit types as often as they were recommended in reality. These retrofit types are cheap and easy to install, and the decision to install them in projects is more likely to be influenced by external factors. For example, a lighting retrofit might be installed in a project regardless of whether the lighting is efficient because the funding is available to update the existing fixtures to the latest standard.

Similarly, boiler, envelope, and window retrofit recommendations are inaccurately predicted in most cases; however, kNN had 50-60% higher prediction accuracy for envelope retrofits and some boiler retrofits than both other algorithms, and Euclidean distance had 93% accuracy for window retrofits in secondary schools. Unlike lighting and lighting controls retrofits, the recall was high for boiler, envelope, and window retrofits across all algorithms, which means that a high proportion of the actual positives were identified correctly. Because there was low precision for these features, we can conclude that there were a substantial number of false positive cases in which the algorithms predicted that a retrofit was necessary yet one was not recommended in the Prop 39 program. These retrofit types require a high capital investment, are difficult and time-intensive to install, and require expertise to select the correct system. These external factors may have prevented schools in the case study from installing these retrofits even in cases where that building feature was inefficient.

Despite the poor performance of predicting retrofit recommendations, the algorithms exhibited much higher accuracy for predicting which retrofits were installed. The installed retrofit prediction accuracy of equipment schedule and plug load predictions were very high, showing little change between retrofit recommendations and installations. The equipment schedule and plug load installation accuracy rates were over 0.91 using both machine learning algorithms, in some cases as high as 0.96 (Fig. 3). These rates are only slightly less accurate than load monitoring approaches used for identifying plug loads from appliances, which can have precision and recall over 97% [17]. The Euclidean distance algorithm had accuracy over 0.96 for plug loads, and equipment schedule installation accuracy of 0.86 in primary schools and 0.67 in secondary schools.

For other building features, the classification algorithms are better able to identify what retrofits were installed rather than which retrofits were recommended, meaning that they are good at identifying changes in load patterns due to changes in building features. Although the algorithms do not account for external factors that impact retrofit decision making, by accounting for the change in building performance over time, we can still see how well the algorithms identify building characteristics. Fig. 3 demonstrates the effectiveness of the algorithms at identifying the performance of several building features that have a stronger signal in the energy consumption data: the boiler, envelope, windows, plug loads, and equipment schedule. The accuracy rates greatly improved for boiler, envelope, and window retrofits when accounting for the change in electricity load over time, as can be seen by comparing the retrofit recommendations and installations in Fig. 3. The machine learning algorithms are particularly good at identifying the trends in energy performance of these features over time. For example, window retrofit installation predictions improved over recommendations, with accuracy rates of 0.99 using the kNN algorithm. The accuracy of boiler predictions using the decision trees algorithm on primary schools went from 0.07 for retrofit recommendations to 0.93 for retrofit installations (Fig. 3). Overall, the kNN algorithm performs the best for window, boiler, and envelope retrofits, in some cases having 20-30% higher accuracy than decision trees and Euclidean distance across both building types. Changes to these building features have a large impact in how the electricity load changes over time, therefore the algorithms are better able to identify changes in the building parameters over time, thus improving the accuracy rate. The algorithms don't have the data on other factors that impact retrofit decisions such as age, costs, or subjective opinions of building managers, therefore their ability to predict retrofit recommendations is limited to energy performance.

Although there were improved predictions for several high capital cost retrofits, the algorithms are still inaccurately predicting lighting and controls system retrofit installations, with accuracy ranging from 0.01 using Euclidean distance to 0.30 using decision trees (Fig. 3). The precision and recall performance for lighting and lighting controls retrofits were similar for retrofit installation and recommendation predictions; figures for the precision and recall of retrofit installation predictions can be found in the Appendix Figs. A.6 and A.7. The performance of these retrofit types is dependent on the user's behavior: the control systems must be programmed to reduce electricity consumption when the building is not in use, and the lights need to be turned off to save electricity. Electrical lighting performance is highly dependent on occupant behavior; new occupant behavior models are continuously being developed to predict better the effect of occupants on lighting end-use consumption [53]. If these retrofits were installed but not being used correctly, the impact on energy consumption would be minimal and the classification algorithms are not likely to pick up on the change.

The machine learning algorithms outperform Euclidean distance matching in identifying which retrofits were installed for over half of the retrofit types. The machine learning methods are intended to capture patterns at different time scales, and therefore may be better able to identify changes to the electricity load over time. Euclidean distance, because it is capturing the difference between the overall time series, may be sensitive to extreme peaks in the electricity load. Because the inputs into the machine learning models are time series features that describe more detail than the magnitude of the load, these models may be less sensitive to outlying data. The machine learning algorithms are better at predicting retrofit types that have a substantial effect on energy efficiency, such as boiler, envelope, equipment schedule, plug loads, and windows retrofit installations. They perform more poorly when predicting controls system retrofits, which are more likely to shift when the electricity load is consumed. Adjustments to the machine learning algorithms will need to be made in future work in order to identify the effect of controls systems. Being able to reliably predict the energy efficiency and demand shifting ability of building features will help to narrow the focus on which systems should be targeted for retrofits. This can help reduce the time needed for audits by eliminating the need for a full-scale energy audit.

4.2.1. *Impact of cost metrics on predictive performance*

Accounting for cost metrics shows improved predictive performance in some cases, depending on the cost metric. As seen in Figs. 4 and 5, *profitable* and *in agreement* cost scenarios tended to have worse accuracy compared to *non-profitable* and *not in agreement*; however, this trend is not consistent across retrofit types. In general, there was extremely varied accuracy among cost scenarios across all retrofit types and classification methods. Validation performance varies based on which cost metric is used to segment the schools in the case study. This calls into question how the validation performance should be evaluated, and whether costs should be taken into account in the prediction of retrofit recommendation and installation.

There is no consistent trend in profitability for improving accuracy, nor is there a clear relationship between accuracy and the predicted SIR being in agreement with the actual SIR. This shows the elaborate nature in which costs play into decision making; cost plays a factor in decision making, but it is not the deciding factor. Studies have found that many external factors play into retrofit investment decisions, including policies and regulations, technological capabilities, building-specific information (size, age, occupancy, etc.), and human factors such as comfort requirements, maintenance, and occupancy schedules [2]. This creates an energy efficiency gap, where EEMs are not installed despite high potential returns on investment. A study on commercial buildings found that more frequent energy audits led to higher EEM adoption

Fig. 4. Validation accuracy of retrofit recommendations for primary schools (top) and secondary schools (bottom). Prediction accuracy varies depending on the profitability of the project, and accounting for cost estimates was not able to improve prediction accuracy.

rates [54]. Therefore being able to quickly audit the performance of certain building features over time may improve EEM adoption rates.

4.3. Time-accuracy trade-off

The Euclidean distance and kNN algorithms were able to perform the classification task on the entire dataset in 2-3 hours, whereas the decision trees algorithm took around 10 days to perform the classification. Euclidean distance and kNN are the most efficient algorithms and exhibit comparable performance, with kNN outperforming Euclidean distance for several retrofit types (boiler, envelope, equipment schedule, and window installations). These two methods capture different aspects of the load patterns; Euclidean distance is lower dimensional and compares the entirety of the time series, whereas kNN has more features and is better at clustering based on patterns in energy consumption.

Despite only having inefficient and efficient scenarios in the parametric analysis, these methods are still fairly accurate at predicting building retrofits. With further parameterization, the model's predictions would only improve. Additional parameters were not included in this study due to the time and cost of including additional values for each parameter and additional parameters in the simulation.

These algorithms have all the benefit of a decreased time-cost compared to traditional auditing methods. Conventional energy audits can take from six weeks to four months to complete and require an energy

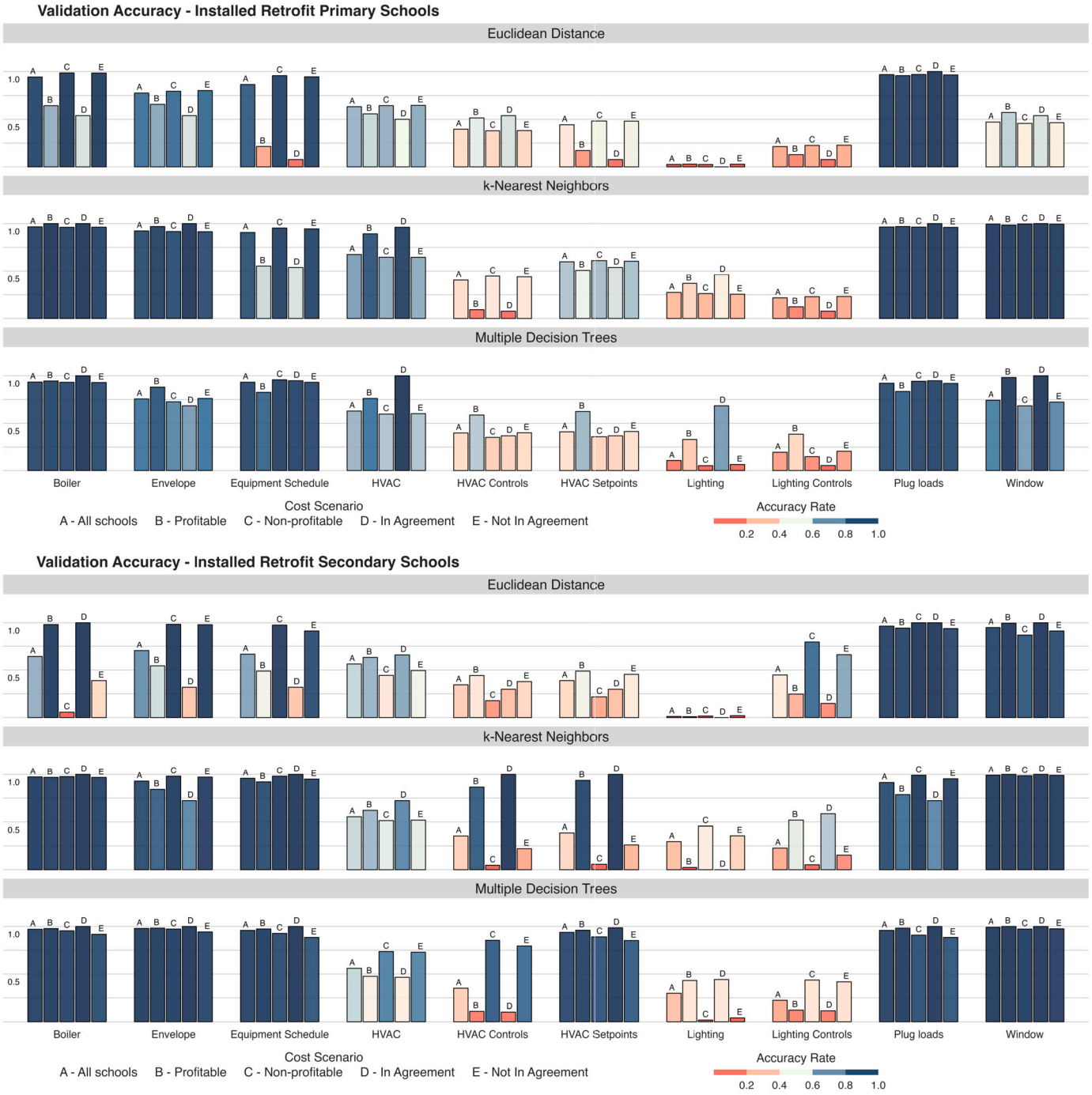


Fig. 5. Validation accuracy of retrofit installations in primary schools (top) and secondary schools (bottom). Accounting for profitability and cost estimates had a varied effect on improving prediction accuracy.

auditor to be onsite to perform the task [4,3,5]. This process provides a very accurate analysis of the building performance with detailed information about specific building systems. While the E-Audit method may not be able to replace an ASHRAE Level 3 energy audit, it can provide a streamlined approach to Level 1 or 2 audits when applied to a large portfolio of buildings. There is an upfront time-cost to run the EnergyPlus models once to create the simulation library on the order of days to weeks; however, once the simulated building database is created, the E-Audit of the entire portfolio can take less than a day. In both the conventional and proposed methods, expert knowledge of building science is required to perform the audit. This limits the accessibility of those from outside the field in applying this method, and this limitation can be addressed in future work by streamlining the building simulation process or by improved sharing of existing simulated building energy model databases.

5. Limitations and future work

The validation case study in this paper is limited by the availability of data on the fuel type, building use type, location, and retrofit installations for the buildings in our study area. We did not have access to natural gas consumption data for these buildings; if this method were applied to natural gas data, we predict that it would be able to correctly classify natural gas boiler efficiency. The validation of this methodology is limited to primary and secondary school buildings in Northern Cali-

fornia, US (ASHRAE climate zone 3C). Future work aims to validate this methodology for additional building use types in various geographic regions. Lastly, the validation data was for retrofits installed in buildings, which does not give direct information on the building features besides that they are inefficient. Because we were trying to predict which retrofits were installed in a building, and the algorithms are trained to predict the performance of building features, we were not directly predicting retrofit recommendations, which can be impacted by myriad factors. While this study is limited due to the validation data available, this work is novel and lays the groundwork for future studies to validate this method on more extensive data.

Across the United States, there's a notable uptick in requirements for large building audits, retuning, and retrocommissioning policies. Take, for instance, New York City's Local Law 87 or Seattle's Building Tune Ups. Local Law 87 mandates buildings larger than 50,000 square feet undergo audits once a decade. In 2023, 14,723 buildings were mandated to perform audits, incurring an estimated local cost of $0.15 per square foot [3]. While policy packages such as New York City's Greener, Greater Buildings Plan and Urban Climate Mobilization Act have succeeded in curbing energy consumption in large buildings, Local Law 87 has seen only marginal savings, primarily due to subpar audit quality coupled with inadequate economic incentives [3]. There's an emerging consensus that algorithmic auditing policy design may begin to relieve some of these barriers [55]. Future work may use this as an opportunity to design a pilot program to assess the integration of E-Audit in auditing policies. Establishing an evaluation pilot has the potential to enhance auditing procedures and performance by standardizing reporting and recommendations, thereby facilitating a more efficient allocation of funds for Energy Conservation Measures.

6. Conclusions

The E-Audit methodology presented in this study can correctly identify several building features that both directly and indirectly impact electricity consumption. It performs as well as existing methods when predictive plug load and equipment schedule performance in a building, with prediction accuracy consistently above 91%. In addition, E-Audit can classify non-geometric characteristics of a building that do not directly consume electricity, such as window construction and insulation/building envelope, showing up to 99% accuracy for these features. It struggled to predict boiler efficiency but was able to correctly identify whether the fuel type was electricity or natural gas, leading to 97% prediction accuracy for boiler retrofits. Controls retrofits were the most difficult to predict, as most accuracy rates for these features were below 40%. For all features, prediction accuracy improved when accounting for changes in energy performance over time.

Among the three classification algorithms tested, the kNN classification had the best performance when considering the trade-off between accuracy and time. The machine learning algorithms outperformed the Euclidean distance matching, yet the decision trees algorithm had the highest time cost, taking over 50 times longer than kNN or Euclidean distance to perform the classification.

Generally, this method overpredicted recommendations for expensive retrofits, and underpredicted cheap retrofits, compared to the ground truth of what was recommended to be installed. This case study demonstrated that the decision to install retrofits is complex and impacted by external factors such as cost, age, regulations, and human factors. Using this method, building managers and energy auditors can determine which building characteristics are performing inefficiently and prioritize energy efficiency retrofits. In research it can be used for building archetyping to identify non-geometric building features; in policy analysis, it can generate insights on how to design retrofits in future building codes. Overall, E-Audit facilitates the energy auditing process and has the potential to improve the accessibility of building energy modeling methods by scaling retrofit analysis for consolidated building archetypes.

CRediT authorship contribution statement

Lauren E. Excell: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Abigail Andrews:** Writing – review & editing, Methodology, Formal analysis, Conceptualization. **Rishee K. Jain:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rishee Jain reports financial support was provided by National Science Foundation under Award #1941695. Lauren Excell reports financial support was provided by National Science Foundation under Award #DGE-1656518. Abigail Andrews reports financial support was provided by National Science Foundation under Award #DGE-1656518. Rishee Jain has patent #S23-531/PROV pending to The Board of Trustees of the Leland Stanford Junior University. Lauren Excell has patent #S23-531/PROV pending to The Board of Trustees of the Leland Stanford Junior University. Abigail Andrews has patent #S23-531/PROV pending to The Board of Trustees of the Leland Stanford Junior University. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Future work aims to validate this method on other building use types and climate zones. To improve the accessibility of this method, a Python repository has been developed that applies the algorithms to a user-defined simulation database. This can be accessed at the following DOI:

E-Audit: A "no-touch" energy audit that integrates machine learning and simulation (Original Data) (Zenodo)

https://doi.org/10.5281/zenodo.10651931.

The data used in the validation case study are subject to a Non-Disclosure Agreement and must be requested directly via the California Public Utilities Commission (CPUC) to comply with privacy controls. Sample data for this method are available through the E-Audit Python repository. The included sample data can be used to verify that the functions are working correctly before running them on a larger simulated database. Instructions for how to create a BEM simulation database like the one used in this paper are included in the README file. To recreate the simulated database from this paper, the parameters from Table A.2 can be used to modify the primary and secondary school models from the DOE commercial reference building dataset; alternatively, please contact the corresponding author to request access to this database.

Acknowledgements

This work was supported by National Science Foundation Graduate Research Fellowships under Award #DGE-1656518 and the National Science Foundation under Award #1941695. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors would also like to thank Kopal Nihar for her contribution to the development of the machine learning classifying code, and Dipashreya Sur for her contribution to the development of the Python package.

Appendix A

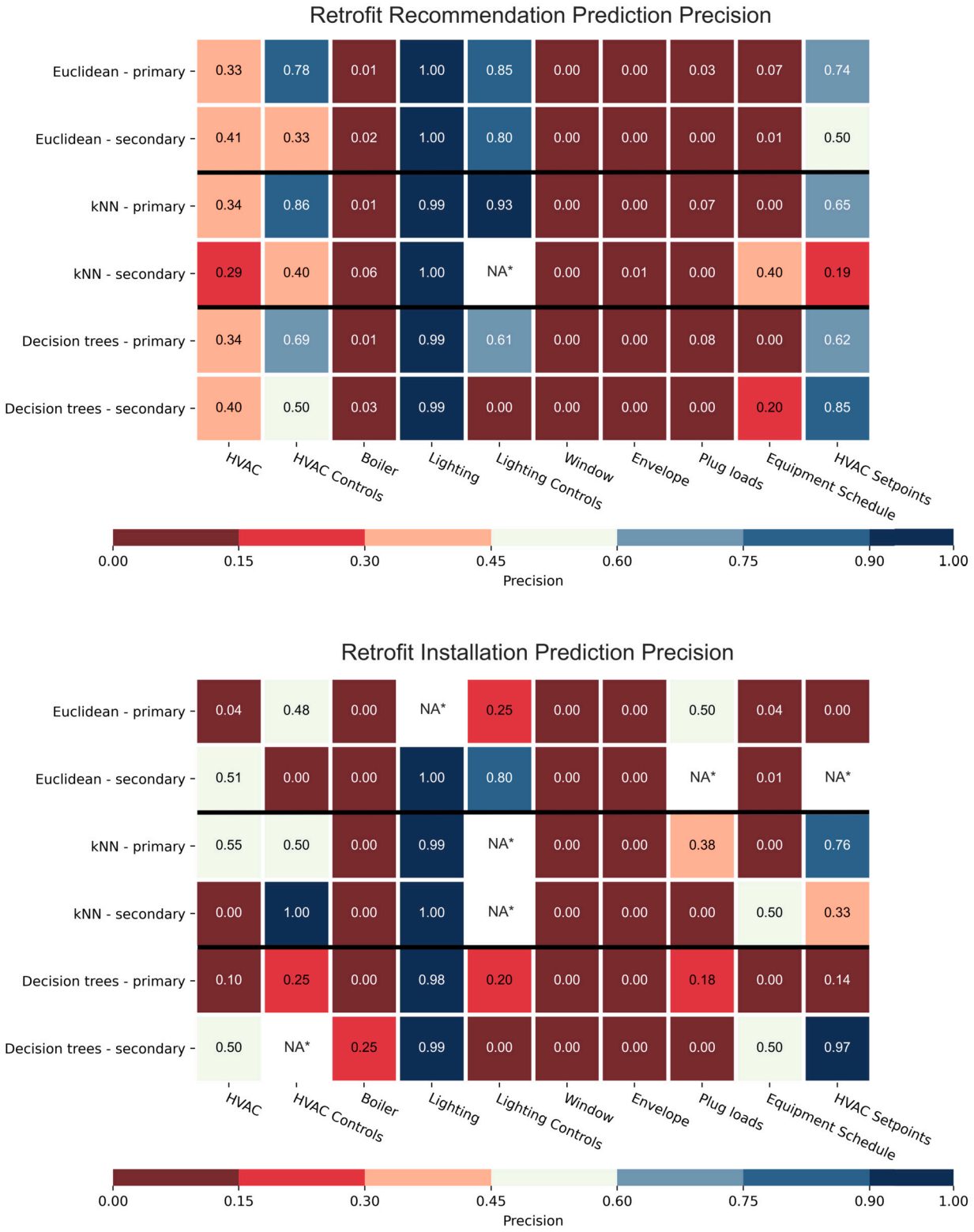


Fig. A.6. Precision of matching algorithms for predicting whether retrofits were recommended (top) or installed (bottom) for schools. *No True Positives or False Positives.

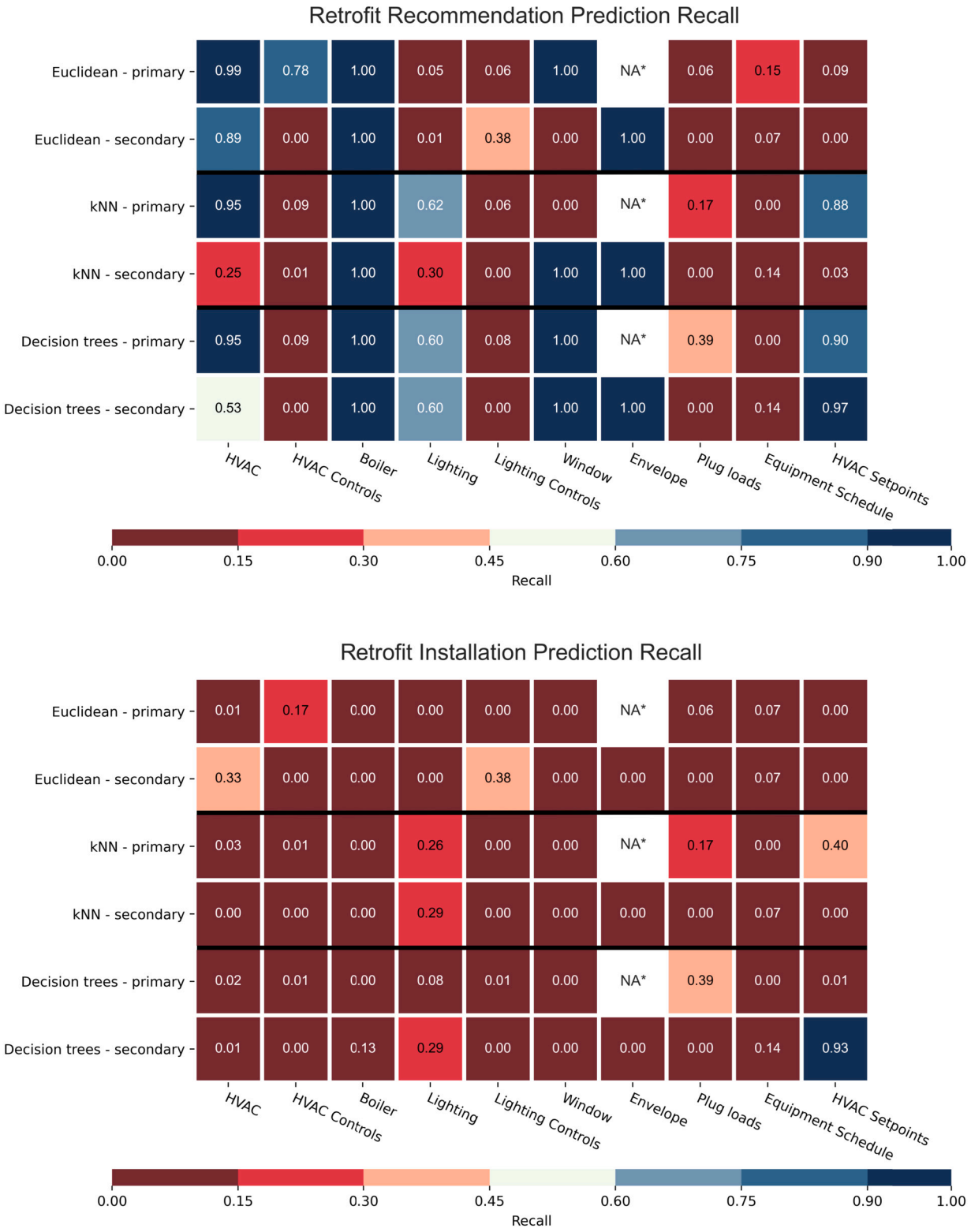


Fig. A.7. Recall of matching algorithms for predicting whether retrofits were recommended (top) or installed (bottom) for schools. *No True Positives or False Negatives.

Table A.2
Building feature inputs for parametric simulation represent 15 inefficiency scenarios. Combinations of these scenarios are represented in the building simulation library.

Building feature	Efficient value	Inefficient value
Classroom lighting power density	15 W/m^2	21 W/m^2
Small area lighting power density	8 W/m^2	16 W/m^2
Large area lighting power density	12 W/m^2	23 W/m^2
Lighting schedule overnight (fraction of lights on)	0.17	0.9
Roof insulation conductivity	0.039 $W/m-K$	0.07 $W/m-K$
Window U-factor	2 W/m^2-K	6 W/m^2-K
Window Solar Heat Gain Coefficient	0.2	0.6
Infiltration (exterior walls)	0.001 $m^3/s-m^2$	0.0015 $m^3/s-m^2$
Boiler efficiency	0.9	0.6
Heat gains from classroom, library, and office technology	5 W/m^2	20 W/m^2
Fuel type for boiler	Electricity	Natural Gas
HVAC schedule	Always off On during the day	Always on
Cooling set point (low)	24 °C	22 °C
Heating set point (high)	17 °C	22 °C
Equipment schedule	Reduce usage overnight Reduce throughout summer	Leave equipment on overnight and throughout summer

References

[1] Pacific Northwest National Lab, Advanced Energy Retrofit Guide: Practical Ways to Improve Energy Performance, Tech. rep., Pacific Northwest National Lab. (PNNL), Richland, WA (United States), Sep. 2011.

[2] Z. Ma, P. Cooper, D. Daly, L. Ledo, Existing building retrofits: methodology and state-of-the-art, Energy Build. 55 (2012) 889–902, https://doi.org/10.1016/j.enbuild.2012.08.018, https://www.sciencedirect.com/science/article/pii/S0378778812004227.

[3] C.E. Kontokosta, D. Spiegel-Feld, S. Papadopoulos, The impact of mandatory energy audits on building energy use, Nat. Energy 5 (4) (2020) 309–316, https://doi.org/10.1038/s41560-020-0589-6, number: 4 Publisher: Nature Publishing Group.

[4] Pacific Northwest National Lab, A Guide to Energy Audits, Tech. rep., Pacific Northwest National Lab. (PNNL), Richland, WA (United States), Sep. 2011.

[5] D. Hsu, How much information disclosure of building energy performance is necessary?, Energy Policy 64 (2014) 263–272, https://doi.org/10.1016/j.enpol.2013.08.094.

[6] Energy Star, How the 1–100 ENERGY STAR score is calculated, https://www.energystar.gov/buildings/benchmark/understand_metrics/how_score_calculated.

[7] J. Roth, B. Lim, R.K. Jain, D. Grueneich, Examining the feasibility of using open data to benchmark building energy usage in cities: a data science and policy perspective, Energy Policy 139 (2020) 111327, https://doi.org/10.1016/j.enpol.2020.111327.

[8] K.S. Atwal, T. Anderson, D. Pfoser, A. Züfle, Predicting building types using OpenStreetMap, Sci. Rep. 12 (2022) 19976, https://doi.org/10.1038/s41598-022-24263-w.

[9] T.R. Dougherty, R.K. Jain, TOM.D: taking advantage of microclimate data for urban building energy modeling, Adv. Appl. Energy 10 (2023) 100138, https://doi.org/10.1016/j.adapen.2023.100138.

[10] K. Mayer, L. Haas, T. Huang, J. Bernabé-Moreno, R. Rajagopal, M. Fischer, Estimating building energy efficiency from street view imagery, aerial imagery, and land surface temperature data, Appl. Energy 333 (2023) 120542, https://doi.org/10.1016/j.apenergy.2022.120542.

[11] J.Y. Park, E. Wilson, A. Parker, Z. Nagy, The good, the bad, and the ugly: data-driven load profile discord identification in a large building portfolio, Energy Build. 215 (2020) 109892, https://doi.org/10.1016/j.enbuild.2020.109892.

[12] P. Westermann, C. Deb, A. Schlueter, R. Evins, Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data, Appl. Energy 264 (2020) 114715, https://doi.org/10.1016/j.apenergy.2020.114715.

[13] S. Yan, K. Li, F. Wang, X. Ge, X. Lu, Z. Mi, H. Chen, S. Chang, Time–frequency feature combination based household characteristic identification approach using smart meter data, IEEE Trans. Ind. Appl. 56 (3) (2020) 2251–2262, https://doi.org/10.1109/TIA.2020.2981916, conference Name: IEEE Transactions on Industry Applications.

[14] J. Zhu, Y. Shen, Z. Song, D. Zhou, Z. Zhang, A. Kusiak, Data-driven building load profiling and energy management, Sustain. Cities Soc. 49 (2019) 101587, https://doi.org/10.1016/j.scs.2019.101587.

[15] M. Anvari, E. Proedrou, B. Schäfer, C. Beck, H. Kantz, M. Timme, Data-driven load profiles and the dynamics of residential electricity consumption, Nat. Commun. 13 (1) (2022) 4593, https://doi.org/10.1038/s41467-022-31942-9, publisher: Nature Publishing Group.

[16] H. Li, Z. Wang, T. Hong, A. Parker, M. Neukomm, Characterizing patterns and variability of building electric load profiles in time and frequency domains, Appl. Energy 291 (2021) 116721, https://doi.org/10.1016/j.apenergy.2021.116721.

[17] M. Hamdi, H. Messaoud, N. Bouguila, A new approach of electrical appliance identification in residential buildings, Electr. Power Syst. Res. 178 (2020) 106037, https://doi.org/10.1016/j.epsr.2019.106037.

[18] A. Zoha, A. Gluhak, M.A. Imran, S. Rajasegarar, Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey, Sensors (Basel, Switzerland) 12 (12) (2012) 16838–16866, https://doi.org/10.3390/s121216838.

[19] G. Chicco, R. Napoli, F. Piglione, Comparisons among clustering techniques for electricity customer classification, IEEE Trans. Power Syst. 21 (2) (2006) 933–940, https://doi.org/10.1109/TPWRS.2006.873122.

[20] I.P. Panapakidis, T.A. Papadopoulos, G.C. Christoforidis, G.K. Papagiannis, Pattern recognition algorithms for electricity load curve analysis of buildings, Energy Build. 73 (2014) 137–145, https://doi.org/10.1016/j.enbuild.2014.01.002.

[21] M.S. Piscitelli, S. Brandi, A. Capozzoli, Recognition and classification of typical load profiles in buildings with non-intrusive learning approach, Appl. Energy 255 (2019) 113727, https://doi.org/10.1016/j.apenergy.2019.113727.

[22] P. Schäfer, Scalable Time Series Similarity Search for Data Analytics, Publisher: Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, Oct. 2015, Accepted: 2017-06-18T14:24:45Z.

[23] G. Tsekouras, N. Hatziargyriou, E. Dialynas, Two-stage pattern recognition of load curves for classification of electricity customers, IEEE Trans. Power Syst. 22 (3) (2007) 1120–1128, https://doi.org/10.1109/TPWRS.2007.901287.

[24] A. Andrews, R.K. Jain, Beyond Energy Efficiency: a clustering approach to embed demand flexibility into building energy benchmarking, Appl. Energy 327 (2022) 119989, https://doi.org/10.1016/j.apenergy.2022.119989.

[25] C. Ding, C. Szum, H. Li, N. Zhou, C. Nesler, Data-driven analysis tool plays critical role in climate neutral buildings, Adv. Appl. Energy 2 (2021) 100014, https://doi.org/10.1016/j.adapen.2021.100014.

[26] S. Lee, T. Hong, M. Piette, Review of Existing Energy Retrofit Tools, Tech. Rep. LBNL-6774E, 1163656, Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States), Jul. 2014.

[27] I. Ballarini, S.P. Corgnati, V. Corrado, Use of reference buildings to assess the energy saving potentials of the residential building stock: the experience of TABULA project, Energy Policy 68 (2014) 273–284, https://doi.org/10.1016/j.enpol.2014.01.027.

[28] A. Mastrucci, O. Baume, F. Stazi, U. Leopold, Estimating energy savings for the residential building stock of an entire city: a GIS-based statistical downscaling approach applied to Rotterdam, Energy Build. 75 (2014) 358–367, https://doi.org/10.1016/j.enbuild.2014.02.032.

[29] U. Ali, M.H. Shamsi, M. Bohacek, C. Hoare, K. Purcell, E. Mangina, J. O'Donnell, A data-driven approach to optimize urban scale energy retrofit decisions for residential buildings, Appl. Energy 267 (2020) 114861, https://doi.org/10.1016/j.apenergy.2020.114861.

[30] T. Hong, M.A. Piette, Y. Chen, S.H. Lee, S.C. Taylor-Lange, R. Zhang, K. Sun, P. Price, Commercial Building Energy Saver: an energy retrofit analysis toolkit, Appl. Energy 159 (2015) 298–309, https://doi.org/10.1016/j.apenergy.2015.09.002.

[31] J. New, M. Adams, A. Berres, B. Bass, N. Clinton, Model America – data and models of every U.S. building, Tech. rep., Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States). Oak Ridge Leadership Computing Facility (OLCF); Argonne National Laboratory (ANL) Leadership Computing Facility (ALCF), Apr. 2021.

[32] P. Farese, R. Gelman, R. Hendron, A tool to prioritize energy efficiency investments, Renew. Energy (2012).

[33] M. Karmellos, A. Kiprakis, G. Mavrotas, A multi-objective approach for optimal prioritization of energy efficiency measures in buildings: model, software and case studies, Appl. Energy 139 (2015) 131–150, https://doi.org/10.1016/j.apenergy.2014.11.023.

[34] L. Magnier, F. Haghighat, Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network, Build. Environ. 45 (3) (2010) 739–746, https://doi.org/10.1016/j.buildenv.2009.08.016.

[35] E. Asadi, M.G.d. Silva, C.H. Antunes, L. Dias, L. Glicksman, Multi-objective optimization for building retrofit: a model using genetic algorithm and artificial neural network and an application, Energy Build. 81 (2014) 444–456, https://doi.org/10.1016/j.enbuild.2014.06.009.

[36] E. Thrampoulidis, G. Mavromatidis, A. Lucchi, K. Orehounig, A machine learning-based surrogate model to approximate optimal building retrofit solutions, Appl. Energy 281 (2021) 116024, https://doi.org/10.1016/j.apenergy.2020.116024.

[37] U. D. of Energy, O. of State and Community Energy Programs, Blueprint 2A: energy efficiency: energy audits, building upgrades, https://www.energy.gov/scep/blueprint-2a-energy-efficiency-energy-audits-building-upgrades.

[38] D. Wang, J. Landolt, G. Mavromatidis, K. Orehounig, J. Carmeliet, CESAR: a bottom-up building stock modelling tool for Switzerland to address sustainable energy transformation strategies, Energy Build. 169 (2018) 9–26, https://doi.org/10.1016/j.enbuild.2018.03.020.

[39] C. Cerezo Davila, C.F. Reinhart, J.L. Bemis, Modeling Boston: a workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets, Energy 117 (2016) 237–250, https://doi.org/10.1016/j.energy.2016.10.057.

[40] Y. Chen, Z. Deng, T. Hong, Automatic and rapid calibration of urban building energy models by learning from energy performance database, Appl. Energy 277 (2020) 115584, https://doi.org/10.1016/j.apenergy.2020.115584.
[41] R.E. Kontar, B. Polly, T. Charan, K. Fleming, N. Moore, N. Long, D. Goldwasser, URBANopt: an open-source software development kit for community and urban district energy modeling, in: Building Performance Modeling Conference and SimBuild, 2020.
[42] J.A. Fonseca, T.-A. Nguyen, A. Schlueter, F. Marechal, City Energy Analyst (CEA): integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts, Energy Build. 113 (2016) 202–226, https://doi.org/10.1016/j.enbuild.2015.11.055.
[43] A. Nutkiewicz, Z. Yang, R.K. Jain, Data-driven Urban Energy Simulation (DUE-S): a framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow, Appl. Energy 225 (2018) 1176–1189, https://doi.org/10.1016/j.apenergy.2018.05.023.
[44] A. Roth, M. Brook, E.T. Hale, B.L. Ball, K. Fleming, N. Long, DEnCity: an open multi-purpose building energy simulation database, in: 2012 ACEEE Summer Study on Energy Efficiency in Buildings, 2012.
[45] B. Bass, J.R. New, E. Ezell, P. Im, E. Garrison, W. Copeland, Utility-scale Building Type Assignment Using Smart Meter Data, Tech. rep., Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States), Sep. 2021, https://www.osti.gov/biblio/1820853-utility-scale-building-type-assignment-using-smart-meter-data.
[46] C. Miller, F. Meggers, Mining electrical meter data to predict principal building use, performance class, and operations strategy for hundreds of non-residential buildings, Energy Build. 156 (2017) 360–373, https://doi.org/10.1016/j.enbuild.2017.09.056.
[47] E. Garrison, J. New, M. Adams, Accuracy of a Crude Approach to Urban Multi-Scale Building Energy Models Compared to 15-min Electricity Use, Tech. rep., Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States), Jan. 2019, https://www.osti.gov/biblio/1510590.
[48] W. Choi, J. Cho, S. Lee, Y. Jung, Fast constrained dynamic time warping for similarity measure of time series data, IEEE Access 8 (2020) 222841–222858, https://doi.org/10.1109/ACCESS.2020.3043839, conference Name: IEEE Access.
[49] Y. Zhang, "Parallel" energyplus and the development of a parametric analysis tool, in: Building Simulation 2009: Eleventh International IBPSA Conference, 2009.
[50] V. Michalakopoulos, S. Pelekis, G. Kormpakis, V. Karakolis, S. Mouzakitis, D. Askounis, Data-driven building energy efficiency prediction based on envelope heat losses using physics-informed neural networks, arXiv:2311.08035 [cs], Nov. 2023.
[51] C. Schoeneberger, J. Zhang, C. McMillan, J.B. Dunn, E. Masanet, Electrification potential of U.S. industrial boilers and assessment of the GHG emissions impact, Adv. Appl. Energy 5 (2022) 100089, https://doi.org/10.1016/j.adapen.2022.100089.
[52] D. Ma, X. Li, B. Lin, Y. Zhu, S. Yue, A dynamic intelligent building retrofit decision-making model in response to climate change, Energy Build. 284 (2023) 112832, https://doi.org/10.1016/j.enbuild.2023.112832.
[53] P. Zhu, M. Gilbride, D. Yan, H. Sun, C. Meek, Lighting energy consumption in ultra-low energy buildings: using a simulation and measurement methodology to model occupant behavior and lighting controls, Build. Simul. 10 (6) (2017) 799–810, https://doi.org/10.1007/s12273-017-0408-6.
[54] C. Schützenhofer, Overcoming the efficiency gap: energy management as a means for overcoming barriers to energy efficiency, empirical support in the case of Austrian large firms, Energy Effic. 14 (5) (2021) 45, https://doi.org/10.1007/s12053-021-09954-z.
[55] D.E. Marasco, C.E. Kontokosta, Applications of machine learning methods to identifying and predicting building retrofit opportunities, Energy Build. 128 (2016) 431–441, https://doi.org/10.1016/j.enbuild.2016.06.092.