Artificial Intelligence-Enhanced Green Building Design for Environmental Sustainability

Saif Mohammed Ali[1], Omar Al-Boridi[2,*]

[1]Department of Computer Science, Dijlah University Collage, Baghdad, Iraq

[2]School of Engineering - RMIT University, Melbourne, Australia

Emails: saif.alameri@duc.edu.iq; omar.alboridi@rmit.edu.au

Abstract

Green buildings are those that use sustainable methods of construction to either maintain or improve the local quality of life. Decisions affecting a project's quality, safety, profitability, and timetable are made using Artificial Intelligence (AI) in Green Construction by analyzing data gathered from monitoring the construction site and using predictive analytics. For instance, increased accuracy in weather predictions might lead to more production, less waste, lower costs, and less greenhouse gas emissions. Green building construction is a significant source of carbon dioxide released through the breakdown of carbonates. Researchers have concluded that integrating industrial wastes is crucial in green concrete making due to its benefits, such as reducing the requirement for cement. When planning with concrete, its compressive strength must be considered. Due to their high predictive power, AI algorithms may be used to determine the compressive strength of concrete mixtures. Existing artificial intelligence (AI) models may be evaluated for their modeling process and accuracy to inform the creation of new models that more accurately represent the comprehensive evaluation of setting parameters on model performance and boost accuracy. Potential sources of conflict in this anthropocentric future include climate change and the availability of renewable energy sources. Scientists think there is a connection between the increased emission of greenhouse gases like carbon dioxide (Co2) from the combustion of fossil fuels and the acceleration of climate change and global warming. Research has demonstrated that the building sector is a significant source of atmospheric carbon dioxide (Co2). Construction, building activities, and subpar energy sources have all significantly increased atmospheric CO2. The proposed research set out to measure how well AI in Green Building Construction (AI-GBC) might reduce carbon emissions and utility bills. Artificial intelligence uses SVM and GA to reduce energy use and carbon dioxide emissions. Several statistical metrics, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Root Mean Squared Log Error (RMSLE), are used to evaluate the AI-GBC's precision. Both Machine Learning (ML) models yielded positive results, with prediction accuracies above 95%. Regarding predicting Co_2, GA models were close to the mark, with an R2 of 0.95. Ninety-six percent will complete a performance analysis, and 97% will conduct a k-fold cross-validation analysis. Cross-validation is used to ensure that the findings of the extended modeling technique are accurate and prevent overfitting.

Keywords: Green Building; Support Vector Machine; Genetic Programming; Energy Consumption; Artificial Intelligence; Sustainable Environment

1. Introduction

Some typical Artificial Intelligence (AI) applications include intelligent assistants, marketing and customer service chatbots, stock market robot advisers, and health diagnosis and prediction systems [1]. The research interacts with AI without realizing it in many aspects of our daily lives, such as spam filters, search engines, ad networks, and streaming service suggestions. Green building describes using environmentally safe and resource-conserving procedures throughout a building's lifespan, from initial planning to final demolition. [2]. Green buildings have several benefits, including their positive impact on the natural world and our local climate. Green building benefits the environment by decreasing waste, increasing biodiversity, and producing renewable energy using less energy [3]. The Green Building methodology maximizes cost-effectiveness, functionality, long-term comfort, and energy efficiency [4]. Sustainability requires that the environment, people, and profit be considered at every stage of

DOI: https://doi.org/10.54216/FPA.170215

Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

production and distribution [5]. If people are serious about making their cities sustainable and green, they need to start with buildings that can keep up with their needs [6]. Existing facilities should also be upgraded with environmentally friendly and sustainable features [7]. A green building is a structure that incorporates eco-friendly practices through its life cycle, from design through deconstruction. Using eco-friendly materials in construction and recycling and reusing existing materials may assist in reducing waste [8]. There is no substitute for a systematic approach to building design and material selection when developing durable structures [9].

Two essential parts of building green are using renewable energy sources and reducing carbon dioxide emissions [10]. Greenhouse gas emissions are enhanced in large part due to the built world. Up to 40% of worldwide greenhouse gas emissions [11] are blamed on the architecture, engineering, and construction (AEC) industry. Most of these emissions may be traced back to traditional design practices. A carbon footprint is the sum of all the carbon dioxide emissions produced by an activity or amassed by a product during its various stages of use. [12] Refers to "Sustainability of construction works- Assessment of the environmental performance of buildings- Calculation technique" and provides detailed instructions for determining a building's total carbon emissions for its useful life. [13]. Kilograms of carbon dioxide equivalent are often used to measure global warming potential.

By improving the design process with digital tools for anticipating, controlling, and monitoring a building's influence on the environment, Building-Information-Modeling (BIM) technology opens up new avenues for creating environmentally friendly structures [14]. Meanwhile, BIM can increase the energy efficiency of sustainable buildings and decrease carbon emissions, providing people with more conducive and healthier living places [15]. According to the life cycle hypothesis [16], the most significant environmental impact from residential buildings' carbon emissions occurs throughout their use, administration, and physical construction. The scope of Life Cycle Assessment (LCA)--based carbon emission monitoring for a building's life cycle is outlined in [17]. Carbon emission inventory studies covering a building's whole life cycle—from its manifest stage (construction) to its operational location (use) and deconstruction (dismantling)—also helped academics understand the core of low-carbon buildings. A methodology for estimating carbon emissions during construction was first proposed [18]. The model used the baseline standard of shift consumption per unit work capacity and the energy consumption per shift cost norm as parameters for building machines. The proposed AI-GBC research aims to accomplish the following primary goals:

- SVM and GA are being used to make predictions about Co_2 emissions in the suggested studies.
- The SVM approach was presented to forecast Co_2 emission costs. Input variables that directly impacted the growth of Co_2 emissions were the use of energy sources like electricity and coal.
- The outcomes of the k-fold validation were inspected using statistical measures, including R2, MAE, RMSE, and RMSLE. All the models displayed positive results using these settings.

The following is a summary of the research by the AI-GBC. Section 1 introduces the topic at hand; section 2 discusses relevant background reading; section 3 details the mechanics of implementing AI-GBC and making. Co_2 predictions, section 4 offers the study's outcomes, and section 5 provides the conclusion.

2. Related Work

Liu et al. [19] examined the digital progress of intelligent green buildings to facilitate the establishment of continuous ecological zones in green environmental cities. The paper organizes the core concepts of Intelligent Green Buildings (IGB) and provides a concise overview of Digital Twins' IGB (DT-IGB) function. The initial focus is on thoroughly examining the meaning and future trajectory of IGB and its existing manifestation and potential applications. Then, in the framework of IGB for DT intelligent cities, the benefits of DTs are explored in further depth. Finally, the opportunities and threats to IGB's future growth are assessed. Based on the findings of this analysis and investigation, IGB has been realized and used but that DTs have yet to be fully incorporated into its design. Therefore, the IGBs must be designed with the future in mind, considering issues like sustainable development, people's livelihoods, and environmentally friendly building materials.

The Internal Energy Efficiency Improvement Design of green construction was shown by Wan et al. [20]. The intelligent Genetic algorithms and Artificial Neural Networks integrated into the Building Information Modeling (GANN-BIM) model form the basis of this unified strategy for optimizing the process of reducing a building's internal energy consumption. BIM is a technology for creating and managing digital models of physical structures. These adaptive simulations may be stored in a file and then exported for use in building. ANNs and GA enhance the BIM model's capabilities as an intelligent technology. The study's overarching goal is to reduce building energy consumption by developing methods for improving environmentally friendly structures' structural and architectural layouts. For assessing the efficacy of the current approach for green buildings, the GANN-BIM

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

model's capacity to manage complicated and conflicting design criteria with low processing effort may prove effective. New green building technologies are constantly being developed; all of them aim to do one thing: safeguard the health of building occupants, reduce waste and pollution, and lessen the structure's negative impact on the environment.

Sharma et al. [21] developed the Internet of Things (IoT), which has given rise to a new generation of eco-friendly technology: the Green IoT. These smart devices are internet-connected and save their data in the cloud. There are already 31 billion IoT devices, expected to rise above the 170 billion limits by 2050. As the number of IoT devices in use grows, so does the percentage of carbon footprint and GHG (greenhouse gas) emissions, which raises the percentage of total pollution on Earth. The existing article has two main aims: one, to assess the environmental impact of smart devices using IoT and propose substantial ways to reduce it, and second, to improve the LCA evaluation model using Deep Learning (DL) and Data Mining Techniques (DMT) with different effect variables for better and more efficient results. Both aims are connected to the "Green IoT" concept as an alternative to IoT for environmental sustainability. The proposed method reduces the manufacturing cycle rate, which benefits the environment. It encourages the reuse and recycling of gadgets. Additionally, from a technical perspective, it decreases the total price, saves time and energy during the first setup of smart devices with fully updated service packs, and aids in estimating GHG emissions and carbon footprints.

One of the primary contributors to environmental pollution, greenhouse gas emissions, and an increase in global temperature is a building's ventilation and air conditioning system. The noise pollution year-round from these facilities is a serious issue. The drawbacks of ventilation and cooling systems are against green design and sustainable development tenets. As a result, contemporary society places a premium on good indoor air quality. It's important to remember that natural air conditioning drastically improves ventilation air quality, reduces noise and pollution, and lowers energy needs. Niroumand et al. [22] achieved natural ventilation by installing wind catchers that, combined with AI and the IoT, may significantly improve the efficiency of the ventilation system. In truth, energy consumption may be reduced by maximizing the effectiveness of natural ventilation with the help of modern technologies. Natural ventilation through wind catchers and solar chimneys can eliminate the need for expensive, energy-hungry air conditioning. The performance of intelligent buildings is also considered, and a thorough overview of relevant research is provided in this chapter, which centers on natural ventilation in buildings. Natural ventilation systems have also been evaluated, with the findings compared to sustainable architecture's green design and development aims.

Using a System Dynamics Model (SDM) to encourage association within a Food-Energy-Water (FEW) at the scale of the complete building, Valencia et al. [23] investigated a green building retrofit method. An indicator-based strategy was implemented to take advantage of the links between different areas of expertise. The carbon dioxide (Co2), resource, environmental footprints, and other variables such as the food security and energy supply dependability ratio were used. The SDM was developed as a proof-of-concept model for stormwater discharge suitable for rooftop farms. The approach incorporates nutrient recycling through green adsorption medium, stormwater reuse for irrigation, and green energy gathering. Green power, stormwater recycling, and rooftop farming have been shown to have positive impacts on the environment, with some studies even showing that they may decrease the amount of Co_2 released into the atmosphere via carbon sequestration on rooftops. The first scenario considers the existing situation, the second scenario introduces rooftop farming and stormwater reuse, and the third scenario implements extra green energy harvesting to aid rooftop farming. Water and carbon footprints were calculated using a LCA for all three systems. Case 3's energy usage fell by 2.24 percent due to using renewable energy sources. Nutrient cycling was encouraged by retaining 82% nitrogen and 42% phosphorus via a green sorption medium throughout crop growth.

A significant tenet of green building principles is finding a medium ground between livability and eco-friendliness. In addition, cutting-edge technology contributes to the viability of metropolitan areas via greener building practices. Elshafei et al. [24] offered a thorough modern study and current courses based on the GA developments to choose the most optimal green layout. Experts advise combining the GA and the Non-dominated Sorting Genetic Algorithm Approach (GA-NSGA-II) for more accurate long-term predictions. The strategy above's widespread applicability stems from its humility, ease of adoption, and exceptional longevity. In addition to the Neural Network and the Simulated Annealing, the GA was also employed. The most widely used methods currently employ an embedded GA-NSGA-II. The benefits of hiring GA-based Multi-Objective Optimization (MOO) for finding the optimal solution for the building decision-making cycle are reviewed at a high level in this study. The GA-coupled approaches may fulfill all the requirements for obtaining optimality in the case of MOO problem-solving.

3. AI-GBC model for predicting Co_2 emission and energy consumption

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

An ML algorithm can improve itself by analyzing previously collected data. The learning procedure for these algorithms often requires a large quantity of data and a modest number of input characteristics. In recent years, many ML methods have been presented in the construction industry to predict cooling and heating loads, energy consumption, and building performance under various conditions. Models in ML may be treated as a black box that can function independently of any prior knowledge of the underlying infrastructure. Analysts use the provided data to determine the connection between the many input characteristics and the desired outcomes (such as energy performance). When given enough training data, ML models can predict targets for unseen samples, even if the connection between the features and the marks is not established. In the ML community, this process is also known as supervised learning. Here, the model is trained using either simulated or actual energy data. Building energy modeling using supervised learning is depicted schematically in Figure 1.

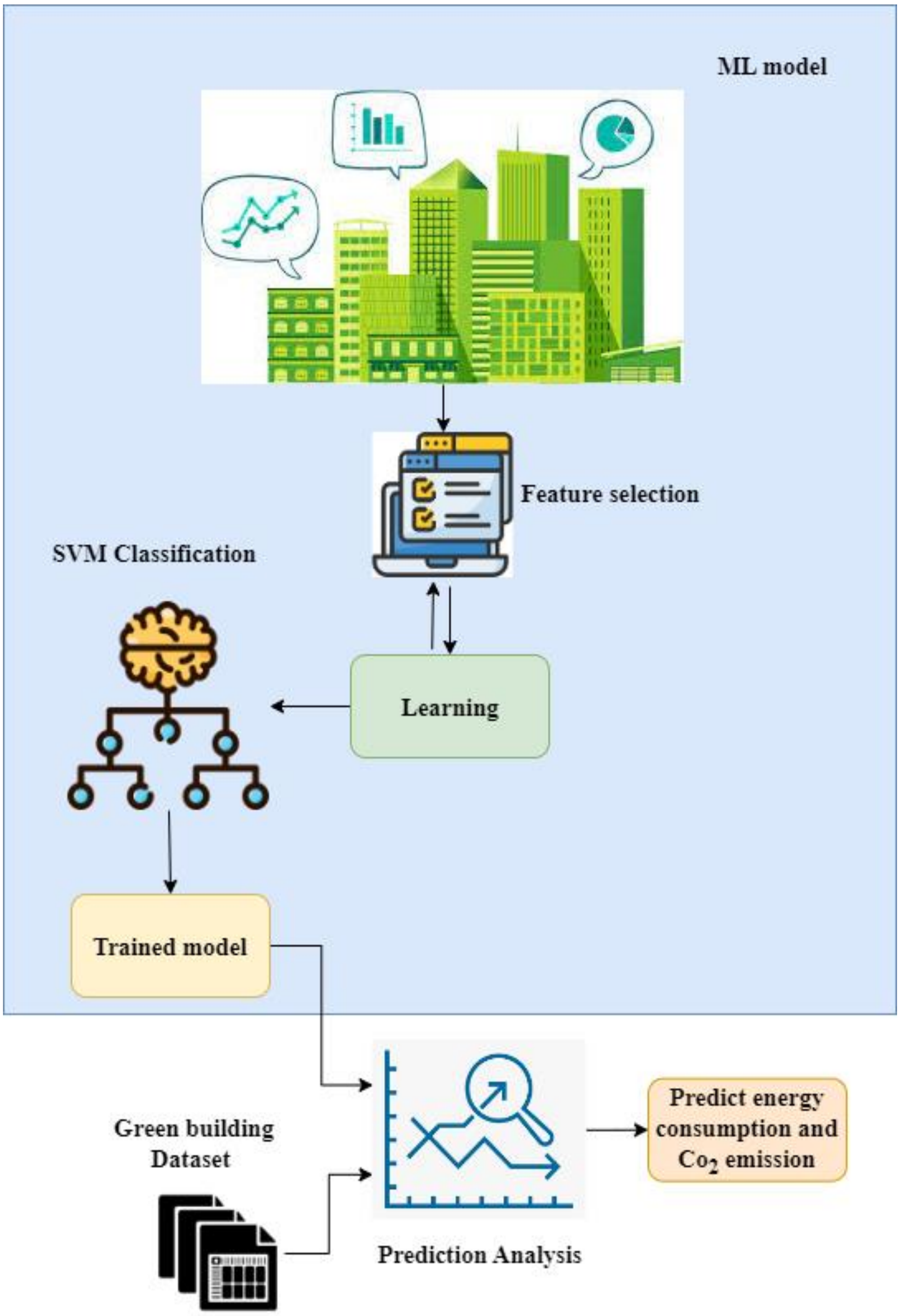

Figure 1. AI-GBC architecture for prediction of Co_2 emission and energy consumption for green building construction.

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

ML is a branch of AI that studies how computers may adapt their actions based on the information provided. Learning from data requires intelligent systems to infer the function that best matches the incoming data. Based on the nature of the available data, machine learning may be classified as either unsupervised or supervised. In ML, the inference process is executed without prior knowledge of the desired outcome by utilizing a training set that is not labeled. The goal is to discover connections by exploring the dataset for parallels. Supervised ML, on the other hand, requires a tagged training set where the predicted outcome has already been decided. Classification is a kind of supervised learning in which an unknown observation is assigned to a known class based on information from a general class's occurrences in a training dataset. A classifier is any algorithm that can map data points from an input to a predetermined category

For optimal performance, classifiers need training based on historical data. N observations are used in the training procedure, all with known labels. This collection of N observations is often split in half to create a training and testing set. First, a classifier that works well with this data is computed with the help of the training dataset. After a classifier has been calculated, its generalization ability is evaluated using the test dataset.

Misclassification rate and test dataset accuracy rate are two standard classifier quality metrics. The misclassification rate is the proportion of times an observation is mislabeled. The success rate may also be expressed in the balance of adequately classified comments and then calculated as follows. When working with limited training and testing data, k-fold cross-validation is superior to other methods of gauging classifier efficacy. A total of l subsamples are drawn randomly from the original set of B observations. Only one of the l subsamples is used for testing, while the other l1 is put to good use in the training process. Each of the l subsamples is used as the test dataset only once throughout the training and testing stages of k-fold cross-validation. Cross-validation accuracy is measured using the mean proportion of correct predictions over all k-test datasets. When l=B, k-fold cross-validation is often referred to as leave-one-out cross-validation.

3.1 Learning

Algorithms used in ML, it has been established, may be trained to become more effective. Three primary methods are currently in use for training ML algorithms. There are three primary types of ML: supervised learning, unsupervised learning, and reinforcement learning. Here, the managed learning method is used. One subset of ML and AI is supervised education, often known as supervised ML. Its distinguishing feature is using labeled datasets for training algorithms to properly perform data classification and prediction. Using a collection of novel, unlabeled input data, supervised learning algorithms attempt to correctly predict a label. Z is the expected result from a mapping function that assigns a class to an input value m: This is the simplest version of a supervised learning strategy.

Genetic algorithms may generate a set of features and SVM parameters simultaneously. This research aims to determine the best settings for the SVM's parameters and feature set such that the classification accuracy is not compromised. The suggested method employs an evolutionary approach to selecting features and tuning parameters. Most of the time, feature set selection methods are split into the filter and wrapper approaches. The precision of the wrapper method makes it the preferred form for selecting elements to use in papers. Additional methods for determining features using GA were also suggested. However, these works ignored optimizing the SVM classifier's parameters instead of concentrating only on feature selection. We have presented a method for feature selection via a genetic algorithm and theoretical constraints on generalization error for SVM. By identifying a group of support vectors that are constituents of the set of training inputs that constitute a hyperplane in the feature space, SVMs may classify data with completely distinct class labels. Using a kernel function, SVMs provide a consistent method analogous to the hyperplane surface for the training data. During training, the SVM picks support vectors along the surface of a kernel function, which may be linear, polynomial, or sigmoid, depending on the user's preferences.

Feature extraction using Principle Component Analysis (PCA)

Based on the principle of conserving primary variance, the PCA [25] technique may successfully decrease the dimensionality of data collection; this is done by generating a new set of indexes, sometimes known as PCs, from the data via an orthogonal transformation that meets the following conditions:

- Each PC may be written as a linear combination of the initial factors.
- There is no relationship between PCs. Each successive PC provides an interpretation for most of the remaining variability in the index that the prior PC did not account for. In this work, PCA was calculated using SPSS v.19.0, and the cumulative explained variation of the chosen PCs should be more than 0.85.

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

Feature selection utilized GA algorithm

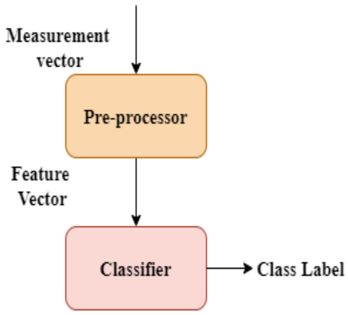

Figure 2. GA working process for feature selection

In other words, GA is an algorithm that uses the concepts of natural selection and genetics to improve its results. In feature selection, it is often used as the search technique to evaluate features in subsets. When applied to feature selection, GA's global search is an improvement over local and greedy search, the two most common search algorithms. As a result, GA is a powerful method since it provides high-quality answers to feature selection challenges. A new population is generated by GA's use of crossover, mutation probability, and survival of the fittest until the maximum conditions are fulfilled in Fig. 2. Chromosomes represent the feature in the GA examine technique, which is then utilized to build a population. As a measure of an individual's hardiness, fitness value plays a crucial role in GA. Equation (1) was used to determine the fitness value F for this research.

$$F = Z_{accuracy} + Z_{Feature} \quad (1)$$

Where, $Z_{accuracy}$ is an accuracy weight and $Z_{Feature}$It is a classification weight based on b features.

When conducting the first experiment, researchers used the available training samples (totaling 3500) and testing samples (1800). In this experiment, GA is concerned with classification accuracy (i.e., the fitness function has 0). The results obtained by the SVM classifier are almost equal to those obtained by the primary classifier alone. Once the fitness function for each chromosome was determined, the population experienced crossover and mutation to improve it. New genes are generated by randomly selecting two parent genes and crossing them together depending on the fitness function score. At the same time, mutation entails producing unknown persons by randomly selecting genes from a chromosome. The fittest individuals have the lowest fitness scores. Due to the computational difficulty in employing wrapper GA for many characteristics, filter GA was used in this investigation. Figure 3 depicts the GA method used to examine the wavelengths and identify the relevant groups in this study. Due to computational challenges in dealing with big data sets that might lead to considerable processing time, the default GA settings were employed for this investigation. With such characteristics, a modest population was used to address the GHG emission problem.

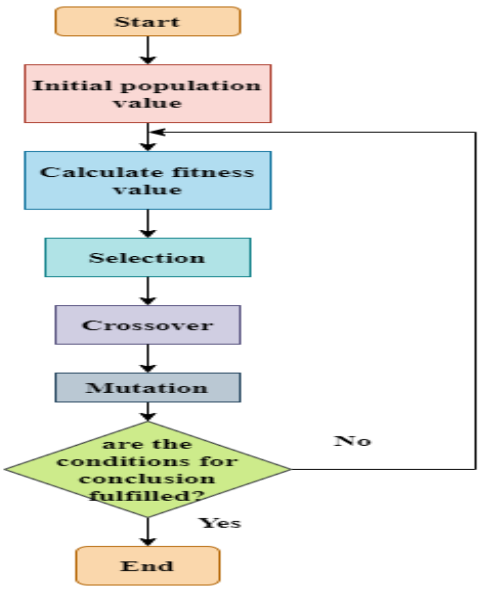

Figure 3. The flowchart for the GA feature selection process

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

3.2 Method for Classification Features

Classifiers and machine learning algorithms are often needed to analyze enormous data sets. Dimensionality reduction strategies applied to the input data help the classifier perform better by reducing the amount of information it has to study. Within statistical machine learning, dimensionality reduction refers to lowering a dataset's number of variables while keeping most of its degrees of freedom. Dimensionality reduction may be achieved by feature extraction and selection.

A c-dimensional data vector is transformed into an a-dimensional output data vector during feature extraction, where $a < c$. For the next phase of classification to be successful, it is essential that the resultant m-dimensional vector, termed a feature vector, retains as much of the original vector's information as possible. Preserving degrees of freedom is a common name for this characteristic. Features may be represented by various data types, including category, numeric, integer, and accurate. Feature selection includes narrowing down a list of candidates to the subset of features (of size b) that most accurately reflects the original pool of elements (of size a) and produces the lowest classification error. Exhaustive Search and Sequential Forward/Backward Floating Search are two approaches that may be used to carry out the feature selection process. Thorough search, on the other hand, investigates every feasible subset, i.e., 2a if b is a constant. Therefore, this strategy ensures that the best subset is chosen despite its excessive use of computer resources. Sequential front or back floating search restricts the search to a more manageable sub-tree by allowing the deletion and insertion of features at each iteration. Thus, this approach has a more manageable computing burden but does not ensure optimum subset selection while being shown to produce near-optimal unsatisfactory outcomes. Performance gains from feature extraction methods are associated with:

- Dimension reduction may help alleviate the curse of dimensionality and lessen the likelihood of over-fitting.
- The simplicity of the resultant classifier requires less storage space and processing power.

3.3 Implementing Classification Using SVM

SVM is a supervised ML technology that may be used for classification and regression based on statistical methods. Some publications classify SVMs as a subset of ANNs despite their principles and derivation distinct from those of ANNs. While SVMs rely primarily on a strict statistic and geometric approach, ANN aims to mimic the functioning of the neural system. SVM was initially conceived as a solution to two classification problems: (i) those with just two possible classes and (ii) those with more than two classes. The best hyperplane decision border, delineating one class from the other, is determined via binary linear SVM classification based on a training dataset. There are two ways to conceptualize optimality, depending on whether or not one can and wants to achieve the perfect classification of the training dataset:

- Hard Margin optimality may be employed if the classes in the training dataset can be perfectly separated. For this scenario, let's choose the decision boundary of the hyperplane such that it is furthest from the closest data point in the training set.
- In cases when exact categorization is either not needed or unattainable, Soft Margin optimality may be used. Here, the choice of the hyperplane is an adjustable compromise between increasing the distance to the next correctly identified training point and decreasing the misclassification rate.

Using an optimization-based learning method, SVM is a system that executes learning bias inside an imaginary space modeled by linear equations in a powerful feature space. The loss of the -incentive function is crucial to the SVM notion. The nonlinear function SVR is a generalization of SVM. The idea of SVM may be characterized as using a single hyperplane across. The elimination of a µ-incentive function is fundamental to the SVM idea. Nonlinear functions may be analyzed using a generalization of SVM called SVR. Dividing space into smaller subspaces that may be resolved non-linearly is the basis of the SVM idea. The SVM model takes into account a two-class classification problem, where the possible values for $j_x = +1, -1$. SVM will locate the optimal hyperplane to partition each training pattern shared by both classes. The hyperplane normal to z and c has a linear equation in (2).

$$z^P \delta(i) + c = 0 \quad (2)$$

If the training data can be split into two linear halves in equations (3) and (4), then a pair (w, b) will exist.

$$z^P \delta(i) + c = -1 \quad (3)$$

$$z^P \delta(i) + c = +1 \quad (4)$$

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

A non-negative slack vector variable $\mu = (\mu_1, \dots, \mu_a)$ will exist in equations (5) and (6) if the classes cannot be separated linearly.

$$z^P \delta_x + c + \mu_x = -1 \quad (5)$$

$$z^P \delta_x + c + \mu_x = +1 \quad (6)$$

Several studies, including those mentioned above, have shown the reliability of numerous approaches to estimating carbon dioxide output in the future. The carbon dioxide forecasting in the eco-friendly building sector will be the focus of this study. The SVM technique of ML will be utilized to make predictions.

This research uses an SVM prediction model to investigate the issue of Co_2emissions, considering the influence of fluctuating energy consumption on the appearance of Co_2 emissions. The study aimed to quantify the contribution of electricity generation and coal combustion to total annual Co_2 emissions. In the Singapore Green Building [26] dataset, the alcoholic beverage sector provided the energy usage statistics. Electrical energy utilized in manufacturing is measured in kilowatt-hours (kWh), while coal burned during production is measured in kilograms (Kg). However, to get Co_2 emission numbers, one needs to convert data on electrical energy and burning coal consumption ($KgCo_2$). The electricity emission factor value is a measure of the emission factors. As illustrated in Equation 7, the quantity of energy utilized multiplied by the calorific value or emission factor yields the converted electrical power.

$$EN_{cd} = \sum_{k=1}^{l} UE_k \times RE_k \quad (7)$$

Where, EN_{cd} is a Co_2 emission due to the use of electrical energy, UE_k Is a Consumption of Electrical Power (kWh), RE_k It is an energy consumption and carbon dioxide emissions per unit of electricity produced (kg/kWh).

Then, numerous factors, including the particular fossil fuel being burnt, contribute to the resulting quantity of Co_2 emissions. Both the amount of fuel used and the kinds of energy utilized are represented in activity statistics and emission factors, respectively. The usual formula for estimating coal's impact on global Co_2 emissions are given in equation (8):

$$EN_{cd} = D \times FE \quad (8)$$

Where, EN_{cd} is a combustion of fuel that releases carbon dioxide (Kg. Co_2), D is a Burning of coal for power (TJ), FE is an emission coefficient (Kg/TJ).

The entire quantity of Co_2 released due to burning fuels, and electrical power generation is then calculated. There was a need for training and testing data for the SVM prediction model to be developed. Co_2 emissions were employed as an output, while electric energy usage and coals data were used as input in this research. The relationship between the input and outcome variables is shown by the following equation (9):

$$Co_2\ emission = i_x[coal - fired\ electricity]j_x \quad (9)$$

In addition, as shown in Figure 2, several stages are involved in developing an SVM model prediction.

This research used a supervised machine learning method to predict future energy use. The cleaned and sorted data was then sent into the learning algorithm. The algorithm was given various feature combinations to produce potential prediction models. Before the data could be used to design and train the model, it was partitioned to create training and testing sets. An SVM with a Radial Basis Function (RBF) kernel was employed for the analysis in this study. This approach is often referred to as a maximum margin classifier, and it is used to solve classification and regression issues for massive datasets. The SVM strategy supports several different kernel options. As illustrated in equation (10), RBF was selected for this investigation because of the dataset's wide range and nonlinear nature.

$$kernal\ RBF(i, \acute{\imath}) = e^{(-\tau i - \acute{\imath}^2)} \quad (10)$$

Where τ is the Euclidean distance between the two sets and $i - \acute{\imath}$ is a gamma parameter that controls the spread distribution of the kernel. The sigma parameter in equation (10) corresponds to a definition in equation (11).

$$kernal\ RBF(i, \acute{\imath}) = e^{(-\frac{i - \acute{\imath}^2}{2\theta^2})} \quad (11)$$

The kernel parameter θ and the cost parameter (C) of SVM are two tuning parameters modified throughout the iterative training process. The sigma value is crucial for a decent model fit to the data. If the data point is miscategorized or exceeds the specified margin, the cost parameter sets the maximum penalty that will be incurred.

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

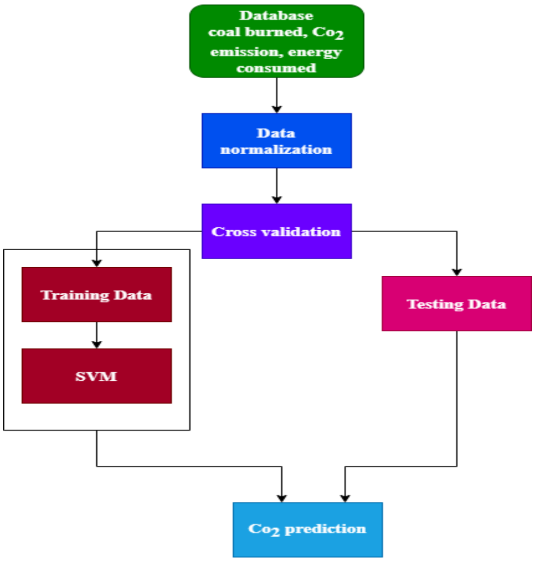

Figure 4. The SVM Design Model for Co_2 prediction

Figure 4 details the data mining approach of preprocessing data required to convert raw data into the understandable and correct format for designing a prediction SVM model. Normalizing the data during preprocessing is necessary to enhance SVM's generalization performance. Subsequently, the dataset is split into training and testing data for cross-model validation's validation purposes. The Co_2 emission prediction model is tested using the SVM technique in this study. SVM, an online learning system, is utilized in Support Vector Regression (SVR). SVR is a subset of the larger SVM that focuses on reducing the dimensionality of data to create regression models. The SVR can resist overfitting based on the proper risk management idea, including estimating a purpose by reducing the limit of simplification error. The training procedure of the SVM model requires finding the ideal values for parameters like the C parameter and Epsilon. The RMSE metric will determine the best value for the SVM model's parameter utilized in the prediction. The RMSE value that was calculated demonstrates the exact forecast performed. The anticipated result can be analyzed.

3.4 Trained model

Any ML technique requires a training model that collects data points. It includes both the sample output data and the input data that contributed to that data. The training model establishes correlations between the inputs and the sample outputs. Computers use machine learning models to recognize specific types of data patterns. When trained, a model is exposed to data and given an algorithm to interpret and learn from. The first stage of ML is model training, which provides a functional model that can be further verified, tested, and ultimately used in action.

4. Results and Discussions

The dataset used in this study are available online and can access by https:// www.kaggle. com/code/kevinnwu/ singapore-green-building (Singapore Green Building) [26].

To encourage, assist, and involve different construction industry sectors in adopting new green buildings, Singapore's Building and Construction Authority (BCA) launched its Green Mark program in 2005, laying the groundwork for the country's first Green Building Masterplan. The Green Building Masterplan has been regularly revised since its introduction in 2006; this is the first book focused on new construction, and its goal is to encourage architects and builders to include sustainable practices in their work. After initial resistance, the Built Environment sector eventually warmed up to sustainable buildings. BCA broadened its focus to include greening the more excellent stock of existing buildings and engaging construction inhabitants to alter energy usage patterns. "At least 80% of buildings (by floor area) in Singapore to be green by 2031" are the objective.

4.1 Co_2 Prediction Analysis

Rapid Miner 5.3 was used in the research to anticipate Co_2 emissions in this investigation. The AI-GBC goal is to track Co_2 output from power use by identifying the forecast with the smallest margin of error. The data in this experiment was divided into a train set of 90% and a testing set of 10% using cross-validation, as was previously mentioned. Then, the 90/10 divides were measured for error, and the process was repeated for each of the ten

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

possible splits. During the training stage, the SVM model's ideal parameters were identified using SVR learning using a trial-and-error strategy. Finding the best possible settings for the SVM model's parameters, such as the C parameters, Epsilon, and the dot function type kernel, is the primary focus of this model. During training, we used a maximum of 10,000 iterations while settling on values for C between -0.1 and 1 and 2 and 50. The C parameter was chosen with the most negligible RMSE value in consideration. Then, after obtaining the C parameter, an Epsilon value between 0.1 and 1 was calculated. The epsilon parameters with the most negligible RMSE value were chosen as an additional criterion. Error minimization during training is achieved by adjusting the C parameters, which, if selected carefully, may also prevent overfitting and underfitting. Underfitting happens when a training method produces a model with low bias and high variance because it fails to correctly detect the underlying trend in the data.

In contrast, overfitting occurs anytime the training process captures noise from the data. Fitting the training data also makes use of Epsilon. The user has control over any or both of these variables. Optimal SVM model parameters for predicting Co_2 are shown in Table 1.

The positioning of the border is the responsibility of the support vectors. The margin is independent of any other data values shown on the graph. The margin will not shift whether I include or exclude non-support-vector data points. However, this does not imply that different data values are meaningless and that only support vectors are helpful. There is a comprehensive evaluation of all data values to determine which data points will be utilized as support vectors. The cost function is identified in the following equation (12).

$$C = \sum_{x=0}^{a} \{ j_x (C_1 (\varepsilon^S i_x)) + (1 - j_x) (C_1 (\varepsilon^S i_x)) \} \quad (12)$$

Where, ε^S is the epsilon of SVM. Figure 5 compares the suggested study AI-GBC to the existing research DT-IGB, GANN-BIM, SDM, and GA-NSGA-II in Co_2 prediction analysis.

Table 1: The Superior Predictive Parameters for Co_2 Emissions Model

$\boldsymbol{C}$	$\boldsymbol{\varepsilon}$	**RMSE**
0.09	0	0.003

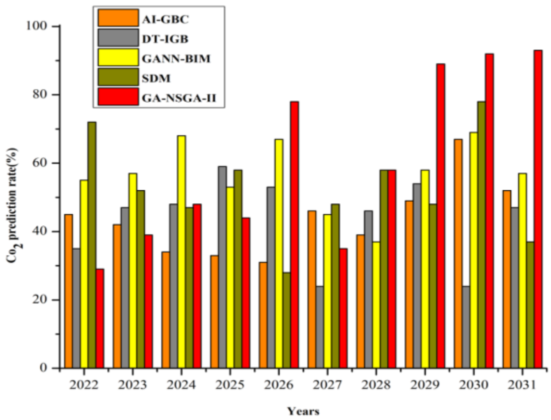

Figure 5. compares the AI-GBCs of the Co_2 prediction rate to that of other models.

4.2 Evaluating performance models

RMSE, MAE, and R2 are the metrics used to assess a model's accuracy. The RMSE is a standard error estimate when forecasting using a specific model and is sensitive to big mistakes. However, outliers might negatively impact the estimated measurements since their square error is more significant than the average. The MAE between anticipated and observed values is thus often utilized in scientific studies. Models with the same unit of error are usually compared using both measurements. It is possible to compare different models using R2 as a standard benchmark. A model's fit quality is evaluated using this index that measures the proportional variation in the response. Greater R2 values indicate superior prediction ability relative to the other two metrics. The finding suggests that feature combination aids in displaying nonlinear connections and may boost a model's prediction accuracy without requiring structural changes. Also, various binary characteristics significantly affect the amount

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

of energy used. Input parameters form a binary feature set shown in Figure 6. In this case, Co_2 levels are linked to the structure of the workday. Specifically, the binary properties of Co_2 concentration and working day, taken together, will impact the energy consumption of buildings. Similarly, Figure 6 shows that several input factors might combine to produce unique binary features.

To express the proportion, consider the following equation in (13). Points on the regression line are denoted by $\widehat{Z_x}$, the predicted value and the mean of the deals are denoted by $\bar{Z}$, and actual values are denoted by Z_x, the observed values.

$$R^2 = \frac{\sum(\widehat{Z_x}-\bar{Z})^2}{\sum(Z_x-\bar{Z})^2} \quad (13)$$

The MAE is the difference between the predicted and actual values, where all input entities are given equal weight. The absolute value is used to get rid of the minus sign. Error magnitudes are measured in the same units used in the calculation. Huge mistakes may occur with any model, including those with MAE values inside a range. The formula (14) for its calculation is:

$$MAE = \frac{1}{b}\sum Z - \widehat{Z_x} \quad (14)$$

Where b is the total amount of data points.

The RMSLE method considers that the estimates and the observed data are from one another. Specifically, it is the difference between values calculated and observed logarithms. Equation (15) determines RMSLE, where x is the expected merit, and Z is the measured merit. Using the log transform to rectify right-skewed results is helpful since it returns the target spread to its original value.

$$RMSLE = \sqrt{\left(\log(Z_x + 1) - \log(\widehat{Z_x} + 1)\right)^2} \quad (15)$$

The RMSE is the standard deviation of the squared differences between the estimated and actual values. The square root of the mistake is a measure of its overall size. That number is the standard deviation of the margin of error. In this method, higher squared differences arise from a stronger emphasis on outliers, whereas more minor squared differences result from a less weighted average. When predicting an output from an input, the root means the square error is a measure of the model's overall accuracy. Lesser RMSE indicates greater model accuracy. An RMSE of 0.5 suggests that the model is not providing accurate predictions. Equation (16) may be used to get the RMSE.

$$RMSE = \sqrt{\sum_{x=1}^{b}\frac{(\widehat{Z_x}-Z_x)^2}{b}} \quad (16)$$

Figure 6 compares the suggested study AI-GBC to the existing research DT-IGB, GANN-BIM, SDM, and GA-NSGA-II in performance analysis.

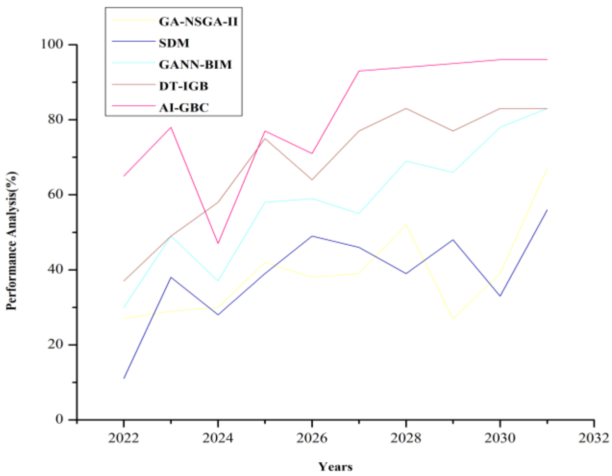

Figure 6. compares the AI-GBC's performance rate to other existing models.

4.3 Analysis of the accuracy of energy-prediction models for green buildings

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

Among the five available prediction models, AI-GBC performance is around average. Interestingly, DT-IGB outperforms LSTM in terms of performance, with a lower RMSE (15.33%) and MAE (51.15%), respectively. To some extent, DT-IGB may be likened to the human brain because it has innate and superior feedback mapping capabilities and learns in a manner reminiscent of how the brain works. The DT-IGB's predictive accuracy is enhanced to some degree by this function. Choosing the correct input variables for a DT-IGB model is crucial when developing a prediction. All the models used in this analysis have similar input parameters. Because of this, DT-IGB's prediction performance may be somewhere in the center, which may also explain why SDM's performance may suffer. GA-NSGA-II has the best prediction accuracy of any SVR model. Possible explanations for this result include that, unlike quadratic programming, the solution of a set of linear equations is all required for GA-NSGA-II, which significantly simplifies the computing burden. To get optimal results, SVR models must use several kernel functions. Unfortunately, there is no assurance that the high-dimensional feature space that the kernel functions map is optimum.

Table 2: Energy consumption predictions for green buildings using SVM

Model	**R2**	**RMSE**	**MAE**
AI-GBC	98.7%	3.27	2.41
DT-IGB	90.01%	28.28	25.27
GANN-BIM	89.17%	30.21	21.98
SDM	85.64%	31.02	19.28
GA-NSGA-II	80.71%	25.29	17.28

Table 2 compares the prediction performance of the model created in the present research to that of prediction models utilized in prior studies to analyze time series data on energy use. There is variation in the accuracy with which these models make predictions, which may be explained by focusing on one of three factors. As a first point, input data might vary from study to study, impacting the accuracy of any given prediction. Second, the goals of prediction might vary between models. Some models, for instance, concentrate on the energy used by HVAC systems, while others look at the point used by buildings. Third, there is room for flexibility between the results of various models. Figure 7 compares the suggested study AI-GBC to the existing DT-IGB, GANN-BIM, SDM, and GA-NSGA-II in energy consumption analysis.

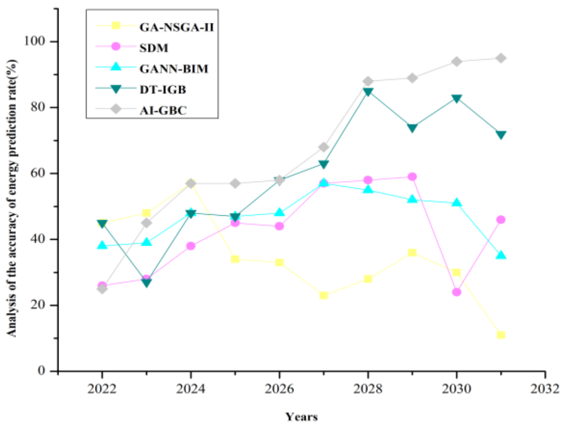

Figure 7. compares the AI-GBCs of energy consumption prediction rate to that of other models.

4.4 Analysis of K folds Cross-Validation

The actual performance of machine learning models may be estimated using cross-validation, a statistical method. It's critical for making sense of the chosen models' objective function. A validation strategy is needed to determine how the model represents the data. The k-fold validation test requires the dataset to be partitioned into l equal-sized sections. The presented analysis separates the experimental samples into ten equal subgroups. Nine ten subsets are utilized, with the tenth reserved for model validation. Then, ten repetitions of the same process are calculated to calculate an average accuracy. The ten-fold cross-validation method is generally acknowledged as a reliable representation of the model's accuracy and conclusion. K-fold cross-validation may be used to examine the test set for signs of bias and reduced variance. Figure 8 shows the evaluation of cross-validation results

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

for energy consumption and Co_2 prediction, respectively, using R2, MAE, RMSLE, and RMSE. The GA model exhibits lower error rates and higher R2 than supervised machine learning methods. Figure 8 displays the range of values for GA modeling, showing an average of 0.83 for a tenfold increase in Co_2 prediction. Figure 8 shows similar results for energy consumption, with an average R2= 0.91, a maximum of 0.99, and a low of 0.55. Validation mistakes are decreasing with each new model. The justification of the result reveals that the average merits of MAE, RMSE, and RMSLE for the SVM model are 4.82, 5.99, and 0.02 and that the corresponding values for the GA model are 0.39, 0.71, and 0.024. The ensemble models also follow this pattern, with considerably higher error rates. Figure 8 compares the suggested study AI-GBC to the existing research DT-IGB, GANN-BIM, SDM, and GA-NSGA-II in Co_2 prediction analysis.

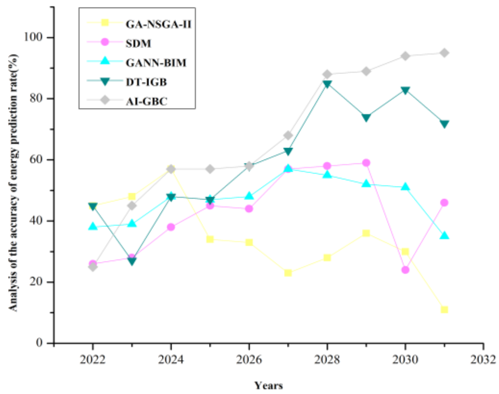

Figure 8. compares the AI-GBCs of the k-fold cross-validation rate to that of other models.

5. Conclusion

Predictions of the green building's Co_2 emissions and energy usage were made using SVM and GA in this research. After reviewing the existing literature, a significant and trustworthy database was built using the collected analysis. Statistical metrics assessed the models, including R2, MAE, RMSE, and RMSLE. The statistical parameter values showed that all models could successfully estimate the green building's Co_2 emissions and energy consumption. Results from the SVM and GA models are compared to those from the existing model. AI-GBC was achieved by the use of both accuracy and sensitivity analyses. Results from ML models were satisfactory, with both achieving high levels of prediction accuracy as 95%. With the R2 value of 0.95 for Co_2 prediction, GA models accurately predicted values congruent with experimental data. Performance analysis was achieved by 96%, and analysis of k-fold cross-validation was performed by 97%. Implementing LSTM in deep learning for prediction is the future of AI-GBC's feature selection and classification.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] KETSEAS, D. (2024). Stochastic Response of an Airfoil and Its Effects on Lco's Behavior under Stall Flutter Regime. International Journal of Mathematics, Statistics, and Computer Science, 2, 168–172. https://doi.org/10.59543/ijmscs.v2i.8663

[2] Debrah, Caleb, Albert PC Chan, and Amos Darko. "Artificial intelligence in green building." *Automation in Construction* 137 (2022): 104192.

[3] Kuthadi, V. M., Selvaraj, R., Baskar, S., & Shakeel, P. M. (2022). Data security tolerance and portable based energy-efficient framework in sensor networks for smart grid environments. Sustainable Energy Technologies and Assessments, 52, 102184.

[4] Nguyen, Hung Duy, and Laura Macchion. "Risk management in green building: a review of the current state of research and future directions." *Environment, Development and Sustainability* (2022): 1-37.

[5] Ur Rehman, Hafiz Saeed, Muhammad Aamer Raza, Rehan Masood, Muhammad Arif Khan, Sabahat Alamgir, Muhammad Afzal Javed, Krishanu Roy, and James BP Lim. "A multi-facet BIM based approach for Green Building design of new multi-family residential building using LEED system." *International Journal of Construction Management* (2022): 1-15.

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024

[6] Debnath, B., Kumar, R., & Shakeel, P. M. (2020). Meandering-trapezoidal shaped MEMS structure for low frequency vibration based energy harvesting applications. Sustainable Energy Technologies and Assessments, 42, 100881.

[7] Durdyev, Serdar, and Serik Tokbolat. "A quantified model for assessment of drivers of acquiring green buildings by potential clients." *Environment, Development and Sustainability* 24, no. 3 (2022): 3815-3831.

[8] Olawumi, Timothy O., and Daniel WM Chan. "Cloud-based sustainability assessment (CSA) system for automating the sustainability decision-making process of built assets." *Expert Systems with Applications* 188 (2022): 116020.

[9] Filho, Marcus VAPM, Bruno BF da Costa, Mohammad Najjar, Karoline V. Figueiredo, Marcos Barreto de Mendonça, and Assed N. Haddad. "Sustainability Assessment of a Low-Income Building: A BIM-LCSA-FAHP-Based Analysis." *Buildings* 12, no. 2 (2022): 181.

[10] Zhang, T., Xie, W., Wei, M., & Xie, X. (2023). Multi-objective sustainable supply chain network optimization based on chaotic particle—Ant colony algorithm. *Plos one*, *18*(7), e0278814.

[11] Xu, A., Zhang, R., Yu, J., & Dong, Y. (2023). Energy Saving Optimization of Commercial Complex Atrium Roof with Resilient Ventilation Using Machine Learning. *Smart Cities*, *6*(5), 2367-2396.

[12] Røpke, Inge. "Trends in the development of ecological economics from the late 1980s to the early 2000s." *Ecological economics* 55, no. 2 (2005): 262-290.

[13] Zavadskas, Edmundas Kazimieras, Jonas Šaparauskas, and Jurgita Antucheviciene. "Sustainability in construction engineering." *Sustainability* 10, no. 7 (2018): 2236.

[14] Khahro, Shabir Hussain, Danish Kumar, Fida Hussain Siddiqui, Tauha Hussain Ali, Muhammad Saleem Raza, and Ali Raza Khoso. "Optimizing energy use, cost and carbon emission through building information modelling and a sustainability approach: A case-study of a hospital building." *Sustainability* 13, no. 7 (2021): 3675.

[15] Khudhair, Ali, Haijiang Li, Guoqian Ren, and Song Liu. "Towards future BIM technology innovations: a bibliometric analysis of the literature." *Applied Sciences* 11, no. 3 (2021): 1232.

[16] Zhou, Shuangxi, Zhenzhen Guo, Yang Ding, Jingliang Dong, Jianming Le, and Jie Fu. "Effect of green construction on a building's carbon emission and its price at materialization." *Sustainability* 13, no. 2 (2021): 642.

[17] Ximenes, Fabiano A., and Tim Grant. "Quantifying the greenhouse benefits of the use of wood products in two popular house designs in Sydney, Australia." *The International Journal of Life Cycle Assessment* 18, no. 4 (2013): 891-908.

[18] Zhang, Xiaoling, Liyin Shen, and Lei Zhang. "Life cycle assessment of the air emissions during building construction process: A case study in Hong Kong." *Renewable and Sustainable Energy Reviews* 17 (2013): 160-169.

[19] Liu, Zhen, Peixuan Li, Fenghong Wang, Mohamed Osmani, and Peter Demian. "Building Information Modeling (BIM) driven carbon emission reduction research: A 14-year bibliometric analysis." *International Journal of Environmental Research and Public Health* 19, no. 19 (2022): 12820.

[20] Wan, Yan, Yujia Zhai, Xiaoxiao Wang, and Can Cui. "Evaluation of Indoor Energy-Saving Optimization Design of Green Buildings Based on the Intelligent GANN-BIM Model." *Mathematical Problems in Engineering* 2022 (2022).

[21] Sharma, Neha, and Deepak Panwar. "Green IoT: Advancements and Sustainability with Environment by 2050." In *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, pp. 1127-1132. IEEE, 2020.

[22] Niroumand, Hamed, Charles J. Kibert, Somayeh Asadi, Fatemeh Mahdavi, and Hadi Arabi. "Wind Catchers as Earth Building: Digital Twins vs Green Sustainable Architecture." In *Digital Twin Technology*, pp. 105-133. CRC Press, 2021.

[23] Valencia, Andrea, Wei Zhang, Lixing Gu, Ni-Bin Chang, and Martin P. Wanielista. "Synergies of green building retrofit strategies for improving sustainability and resilience via a building-scale food-energy-water nexus." *Resources, Conservation and Recycling* 176 (2022): 105939.

[24] Elshafei, Ghada, Silvia Vilčeková, Martina Zeleňáková, and Abdelazim M. Negm. "An Extensive Study for a Wide Utilization of Green Architecture Parameters in Built Environment Based on Genetic Schemes." *Buildings* 11, no. 11 (2021): 507.

[25] Sun, Wei, and Jingyi Sun. "Prediction of carbon dioxide emissions based on principal component analysis with regularized extreme learning machine: The case of China." *Environmental Engineering Research* 22, no. 3 (2017): 302-311.

[26] https://www.kaggle.com/code/kevinnwu/singapore-green-building (Singapore Green Building)

DOI: https://doi.org/10.54216/FPA.170215
Received: January 28, 2024 Revised: April 28, 2024 Accepted: September 29, 2024