Building and Environment 205 (2021) 108267

Contents lists available at ScienceDirect

Building and Environment

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

Building life-span prediction for life cycle assessment and life cycle cost using machine learning: A big data approach

Sukwon Ji, Bumho Lee, Mun Yong Yi *

Graduate School of Knowledge Service Engineering, Department of Industrial & Systems Engineering, KAIST, Republic of Korea

ARTICLE INFO

Keywords:
Building life span
Life cycle cost
Life cycle assessment
Big data
Machine learning
Deep neural network

ABSTRACT

Life cycle assessment (LCA) and life cycle cost (LCC) are two primary methods used to assess the environmental and economic feasibility of building construction. An estimation of the building's life span is essential to carrying out these methods. However, given the diverse factors that affect the building's life span, it was estimated typically based on its main structural type. However, different buildings have different life spans. Simply assuming that all buildings with the same structural type follow an identical life span can cause serious estimation errors. In this study, we collected 1,812,700 records describing buildings built and demolished in South Korea, analysed the actual life span of each building, and developed a building life-span prediction model using deep-learning and traditional machine learning. The prediction models examined in this study produced root mean square errors of 3.72–4.6 and the coefficients of determination of 0.932–0.955. Among those models, a deep-learning based prediction model was found the most powerful. As anticipated, the conventional method of determining a building's life expectancy using a discrete set of specific factors and associated assumptions of life span did not yield realistic results. This study demonstrates that an application of deep learning to the LCA and LCC of a building is a promising direction, effectively guiding business planning and critical decision making throughout the construction process.

1. Introduction

1.1. Importance of building life span in architectural engineering

From planning to design, construction, maintenance, and disposal, decision making is essential at every stage in the building construction industry. Various economic and environmental feasibility evaluation methods are used to execute major decisions including the selection of structures, materials, and construction methods throughout the design and construction of a building. Among them, life cycle assessment (LCA) and life cycle cost (LCC) analysis are two primary verification methods commonly used across the industry - in particular, LCA for environmental verification and LCC for economic verification [1].

The LCA method evaluates the environmental aspects of a product system at all phases of its life cycle [2]. In the construction sector, LCA assesses the impact of a building on the environment in all phases including material manufacturing, material transportation, construction, operation and maintenance, demolition, and construction waste disposal. According to the 2017 UNEP Global Status Report, the construction sector consumes 35% of global energy use and nearly 40% of energy-related CO_2 emissions [3]. Furthermore, the operation phase accounts for 80–90% of the total energy consumed by the entire life cycle [4]. In the case of apartments made from reinforced concrete (RC), the operation phase contributes to 75.4% of the carbon emissions produced over the entire life cycle [5]. Thus, for energy saving and environment protection, it is highly important that LCA is conducted accurately.

LCC is an economic analysis method that calculates all or part of the total costs of construction, operation, maintenance, and disposal incurred by a project during its useful life. In LCC, the costs incurred in the operating phase, which include expense created from operating a building with its resource consumption such as energy and water, and in the maintenance phase, which include expense from sustaining building condition and quality, account for approximately 40% of all costs over the life cycle of a building [6]. Calculating the operation and maintenance costs requires accurate reference data on the building's life span and the repair cycles of building materials, construction methods, and building equipment.

* Corresponding author.
E-mail address: munyi@kaist.ac.kr (M.Y. Yi).

https://doi.org/10.1016/j.buildenv.2021.108267
Received 14 June 2021; Received in revised form 13 August 2021; Accepted 17 August 2021
Available online 19 August 2021

In the construction industry, an essential component of LCA and LCC is the life span of a building, which must be considered in the calculation of the environmental load and maintenance and operation costs. It serves as a key factor that determines the number of repairs required. Also, building life span has an absolute influence on the energy consumed during the service cycle of the building, total cost and quantity of building repairs, and total environmental load [7]. In addition, as there are few reference data sources describing the standard life span of a building other than those sources assumed from the main structure of buildings, there are many cases in which LCC calculations omit the maintenance cost during the fulfilment of value engineering (VE) in the design phase.

However, although a building's life span strongly influences both the LCA and LCC outcomes, it is unrealistic to consider all the factors that affect it. Hence, the life span is a component typically assumed to be the same for all the buildings with the same main structure type or other similar incidental factors. In LCA, the operation life is often assumed to be 30–50 years, and the guidelines of building operation and maintenance are quite simple. However, some materials do not meet the typical service life period criteria due to mismatched exposure or usage characteristics, making it difficult to use a simplified description to fully express its maintenance conditions [8]. In practice, the life spans of buildings vary considerably due to different influencing factors, and the use of standard assumptions often lead to grossly inaccurate results [9].

1.2. Complexity of building life-span prediction

There are three broad categories of building life span: physical life span, social life span, and functional life span.

1. Physical life span refers to the period a building gets demolished because the major components of the building physically reach their life span and are no longer technically useable. It can be further divided into the structural, facility, interior, and exterior life spans. The structural life span refers to the number of years until the physical strength of the building wanes. It is a systemic description of the performance deteriorations, defects in sections, and poor construction conditions associated with the material types and their characteristics. In addition, the facility life span, which is conceptually related to the idea of physical life span, refers to the number of years until the facility inside a building reaches a state in which it cannot be physically repaired. For example, according to structural standards and specifications, RC-framed apartment buildings in South Korea, which serves as a backdrop for our research, have a structural life span of 60–100 years. In contrast, the facility life spans of these buildings will vary depending on their material characteristics. In general, the facility life span is calculated to be approximately one-third of the structural life span of a building.
2. Social life span refers to the period a building lasts until human desire or legal requirement dictates replacement. It is dependent on changes in social conditions or needs. Apart from the physical or functional life span, the social life span includes the life span for redevelopment, reconstruction, and remodelling in accordance with a national policy or social demand. In addition, a building's social life span may be affected by the adoption of new safety regulations or standards related to asbestos, fire resistance, or seismic resistance.
3. The functional life span refers to the period before which a building loses its functional value due to societal or lifestyle changes. For example, commercial buildings, factories, and hospital buildings are subject to particularly rapid changes in function. Furthermore, housing needs vary over an approximately 10-year cycle due to population changes or resident preference changes. As buildings are directly affected by technological innovation or change in life style, a plan that will help a building withstand changes in function or use is important. In general, the spatial composition, centre (core) location, column spacing, and floor height of a building will affect its functional life span.

In addition, there are various other concepts of life span, including economic, cultural, design, and tax life spans. The economic life span is governed by economics. A building is an assembly constructed of numerous materials using various methods that have diverse life cycles. In addition to the original purpose of a building, which is to shield the inhabitants from the outside environment, buildings are assets that have a strong purpose as a financial investment. Therefore, even if a building structure or component is defective, it may not be demolished; conversely, it may be demolished before reaching its life span limit even if its condition is sufficient for its intended purpose. The cultural life span refers to the cultural relevance of a building. Some buildings are preserved well beyond their life span either by artistry or through continuous renovation as a heritage site. The design life span is related to the relevant design trends. Public buildings, for example, are designed to help them survive for a long time, whereas the design of commercial buildings (and thus their life spans) changes depending on their surroundings, technology, and originality. The tax life span refers to the period the relevant tax law of the land is applicable to a building. In South Korea, for example, the tax law prescribes a tax life span of 40 years for major structures such as steel–RC (reinforced concrete), RC, and steel frame buildings [10]. Thus, the actual life span of a building is not simply the sum of the internal factors of the building itself but also related to the site and surrounding environment (such as weather, humidity, salinity, and disaster), extent of maintenance, building-related regulations and guidelines, re-development and reconstruction standards, asset value, and social demands and implications.

1.3. Limitations of existing building life-span prediction methods

Various researchers have attempted to predict the building life span via building materials and construction methods. Many have only studied the life of materials and construction methods located in a specific region. It has been difficult to analyse all the factors influencing building life span because each material and method used to construct a building has distinct characteristics and are subject to different environmental conditions. Therefore, methods have been developed for predicting the life span of specific buildings according to specific factors by setting the goal and scope of the prediction according to the purpose of the study. Recently, hybrid methods that consider both the internal and external factors of a building to predict its life span have been developed [43,51–54]. However, most of these methods subjectively select important factors from numerous external factors and assign quantified weights based on the opinions of researchers or participating experts. Even in the case of the modeling-based approach [e.g., 52–54], most researchers selected a model simply based on the life span distribution and determined factors deemed to have high weights to the modeling. These methods suffer from inconsistency in assigning the weights across the countries or regions in which the internal and external factors of buildings are different, not to mention the subjectivity of participants. Therefore, the results of different studies are inherently empirical and difficult to standardize, making it impractical to predict the actual life span of any given building using one of those methods.

In summary, the objective of this study is to identify the primary factors that affect the life span of a building by applying machine learning techniques to a large scale real-world dataset, which is the entire set of building permission registry data accumulated since 1950s, the era from which the current form of country-wide building registration system began in South Korea. To examine the comparative advantage of our approach, we employ the current wide-spread method of calculating building life span using the linear regression technique, in addition to deep learning and traditional machine learning techniques. The ability to accurately predict the life span of a building can dramatically improve the output accuracy of LCA and LCC and help building owners, contractors, designers, and other stakeholders make well-informed decisions.

2. Related work

2.1. Expected building life span

Until now, there has been no consensus in the literature on the calculation of building life span [21,22]. The life span of buildings used in prior LCA studies [19,21–40] varies from 40 to 150 years. Most LCA and LCC practitioners do not estimate the actual life span of buildings but only apply default values obtained from the relevant structural calculation codes [e.g., 20]. When wooden or high-strength concrete structures were used to fabricate the main structure of a building, its life expectancy was typically assumed to be 100 years. Additionally, the life spans of 40, 45, and 150 years have been infrequently used in accordance of national corporate tax law, depending on major structural type variation (see Table 1).

2.2. Actual building life span and main influencing factors

O'Connor [41] conducted a demolition survey to identify the age, type, structural material, and reason for demolition of 227 buildings in a major North American city. The life spans of most of the demolished steel and concrete buildings were less than 50 years and no significant relationship was found between structural material and average service life. Moreover, 'local re-development' (34%), 'lack of maintenance' (24%), and 'building no longer suitable for intended use' (22%) were the most important external factors affecting building life. Dias [42]

Table 1
Building life span and basis of estimation in existing LCA and LCC studies.

	Building life span (years)	Basis of estimation	Title	Country	Year	Main frame type
1	50	Simple assumption	Energy use during the life cycle of single-unit dwellings: Examples [21]	Sweden	1997	Wood
2	50	Economic building life span in Sweden	Life cycle assessment of four multi-family buildings [22]	Sweden	2001	Lightweight concrete & concrete, RC, Wood, Steel columns and concrete
3	40	National corporate tax law	Method of economic analysis for remodelling of deteriorated apartments using the life cycle costing [23]	The Republic of Korea	2005	RC
4	50	Buidling life expectancy in midwestern U.S.	Comparison of environmental effects of steel and concrete-framed buildings [24]	US	2005	Steel frame, cast-in-place concrete frame
5	50	Simple assumption	Life cycle assessment of office buildings in Europe and the United States [25]	US & Europe	2006	Steel framed RC
6	100	Simple assumption	Comparison of the life cycle assessments of an insulating concrete form house and a wood frame house [26]	US	2006	Wood, Insulating concrete form
7	50	Simple assumption	Sustainability based on LCM of residential dwellings: a case study in Catalonia, Spain [27]	Spain	2008	Brick
8	50	Simple assumption	Life cycle assessment in buildings: state of the art and simplified LCA methodology as a complement for building certification [28]	Spain	2009	RC
9	50	Simple assumption	The environmental impact of the construction phase: An application to composite walls from a life cycle perspective [29]	Spain	2010	Block
10	50	Simple assumption	Life cycle assessment of a house with alternative exterior walls (Comparison of three impact assessment methods) [30]	Portugal	2011	RC
11	50 (existing building) 100 (high-strength concrete building)	The lower limit of the lifespan of high-strength concrete structure	Life cycle CO_2 evaluation on reinforced concrete structures with high-strength concrete [31]	The Republic of Korea	2011	RC
12	50	Simple assumption	Assessment of the environmental performance of buildings: A critical evaluation of the influence of technical building equipment on residential buildings [32]	Austria	2012	RC, Brick, Timber
13	50	Prior work	Environmental impacts of the UK residential sector Life cycle assessment of houses [19]	UK	2012	Brick and Block
14	50	Simple assumption	Life cycle assessment of the air emissions during building construction process: A case study in Hong Kong [33]	Hong Kong	2013	RC
15	45	A durable period of a build	Sustainability life cycle cost analysis of roof waterproofing methods considering LCCO2 [34]	The Republic of Korea	2014	RC
16	50	Prior work	Life cycle assessment (LCA) of roof waterproofing systems for reinforced concrete building [35]	The Republic of Korea	2014	RC
17	50 (Alt1) 30 (Alt2)	Simple assumption	Life cycle cost optimization within decision making on alternative designs of public buildings [36]	The Czech Republic	2014	RC
18	50–150	Prior work	Life cycle assessment (LCA) of building refurbishment: A literature review [37]	Spain	2017	RC, etc.
19	100	Simple assumption	Life cycle assessment of building materials for a single-family house in Sweden [38]	Sweden	2019	Wood, Concrete
20	100	The expected building lifetime in Czech Republic	LCC Estimation Model: A Construction Material Perspective [39]	The Czech Republic	2019	Masonry
21	40	Simple assumption	Life cycle greenhouse gas emission and cost analysis of prefabricated concrete building façade elements [40]	The Netherlands	2020	Prefabricated concrete

Note: As shown in the table above, there are many cases in the LCC-related studies in which the life span was estimated to be a certain number of years using a simple assumption (without providing any explicit reference). In the LCA-related research, most studies assumed the life span to be 50 years by referring to prior work. These approaches are still widely used today.

surveyed the conditions of several buildings under 125 years old in a humid, tropical environment to find that the change in function of or lack of investment in a building reduced its life span and that new planning regulation or archaeological heritage had an impact on extending the life span. Liu et al. [43] used an improved hedonic model to investigate the factors influencing the life spans of 1732 demolished buildings from 2008 to 2010 in seven communities of Jiangbei District, Chongqing, China. The average service life of the buildings was determined to be 34 years—much shorter than the design life span in China. They also noted that external influences were more important than internal influences, and that internal factors excluding floor area were less important than expected. Grant [9] modelled nine combinations of building envelopes using five service life models. The study results indicated that the life span of a building varied with its conditions, and the annual cumulative life cycle and life cycle effects depended mainly on the predicted life span of the materials and the environmental indicator used.

3. Methodology

3.1. Big data approach

3.1.1. Data sources for building life-span prediction

In many countries, to be in compliance with the building codes, building permits, construction permits, and construction reports are required when constructing a building. Plan checks, which verify the compliance of buildings with the area plan requirement, are also being implemented as part of the approval process [11]. Furthermore, builders must typically procure permits for use upon building completion and apply for a permit prior to building demolition. In the United States, a building permit must be obtained before construction in accordance with the Law on Spatial Planning and Construction. Completed buildings can be occupied and operated once a use permit has been obtained indicating that all the relevant technical documents, standards, regulations, and norms have been observed. Finally, a demolition permit must be obtained to demolish a building. These permits can be obtained from the Department of Town Planning, Housing and Communal Affairs and Ecology. In the United Kingdom, a planning permit or development approval is necessary to construct or extend a building (including major renovations), as well as to demolish a building in some jurisdictions [12, 13]. Under the Building Law of the People's Republic of China, prior to the commencement of construction work, the construction agency must apply for a construction permit with the department-in-charge of the government construction administration of the relevant administrative district. When the construction is completed, a 'Construction Process Completion' certificate is issued following an inspection [14]. In South Africa, major permit stages and obligation items in construction, usage, and demolition are specified in accordance with the National Building Regulations and Building Standards Act (Act No. 103, 1977) [15]. In most countries, the construction permit or declaration is mandatory, and although it may take slightly different names and forms, it involves nearly the same criteria.

In South Korea, the following permit-related records are created with regard to building construction and demolition:

- Authorised Register on Building Permit: A document detailing the 'Building Permit' expressing permission from the Governor of the Special Self-Governing Province or the head of a Si/Gun/Gu, by which a person who intends to construct or repair a building is authorised.
- Authorised Register on Approval of Use: After the completion of building construction, the owner of the building must apply for an 'Approval of Use' in order to occupy the building by attaching a supervision completion report and a construction completion document prepared by the construction supervisor.
- Authorised Register on Demolition Permit: When the owner or manager of a building has lost all or part of the building owing to a disaster or other destruction, they shall prepare an 'Architecture Loss Declaration' and file it with the Governor of the Special Self-Governing Province or the head of the District (called Si, Gun, Gu in Korean) within a week of the scheduled demolition date.

The South Korean construction administration system, named *Seumteo*, has been established to handle almost all construction-related administrative duties. The construction data describing building permits, construction reports, completion (Approval of Use), maintenance, and demolition are provided to the private sector through an open data service named the Architectural Data Private Open System [16]. Even in South Korea, where the information technology industry is relatively advanced, such large-scale data on buildings have not been released to the private sector until recently. Although many other countries have not yet opened administrative building data to the private sector, more and more countries are expected to share the construction- and demolition-related information freely in the form of raw data as time goes on.

The service life (or service period) of a building is the period for which it remains durable, stable, and resilient [17]. The international standard ISO 15686:2014 documents an analysis of this concept [18]. Realistically, the building service life or building life span generally refers to the period between its construction and demolition, as the actual life span does not solely depend on stability and durability. Therefore, the life span of a building can be considered as the period between the date of approval of use and the date of demolition from an administrative perspective. Any country can therefore analyse and predict the actual life span of a building in its territory by utilising its mandatory and practical administrative data, which has been accumulated by law as part of the construction permitting process. The larger the quantity of data and the longer the period of recording, the more accurate a prediction based on this data can be.

3.1.2. Big data characteristics and challenges of the study data

The collection of the Korean dataset described in 3.1.1 involves a government-wide effort on establishing an integrated database from all the records produced in South Korea in relation to construction. Given that various records and data from all over the country have been accumulated over a long time period, the dataset reflected several characteristics of big data and consequently a big-data analysis strategy was required to accomplish the intended purpose of this research. Below are the characteristics and the challenges of the integrated records that we encountered. We addressed the characteristics according to the 7Vs presented in the work of Sivarajah et al. [55].

1. Volume: The Seumteo database we accessed is established with all the construction related documents registered to the local government offices throughout the nation. The partial database of the demolition records that this research mainly utilizes includes 1,812,700 cumulative demolition cases arranged with 78 different building characteristics from the 1950s to the current date. As a full data record of construction events, the data used in this research can be considered as one of the largest datasets used for analysing construction and demolition events.
2. Variety: The database includes data with a different aspect of building characteristics. For example, characteristics such as an address, lot type, and environmental grade are recorded to describe administrational information, while structures, number of floors, and parking space type are recorded to describe the physical aspects of the building. All these different types of information are evaluated and registered from different procedures performed by different groups of people with different professions. The variety of the type and the source of this information cause challenges in processing and

analysing source data, therefore cautious selection, imputation, and integration of the data are necessary.

3. Veracity: The dataset has been collected for decades while the registration procedure, form, input type of raw data have been changed over time. This characteristic produces discrepancy in the data quality and bias depending on the input methodology utilized. Old data get corrupted or lost in physical form, producing data sparsity when they are integrated and transferred to a large electronic database. Therefore, the validation and the control of sparse data is necessary to keep an adequate level of veracity. For the same reason, the database has been set up based on the information submitted to government offices under strict legal restrictions. There also are several audit processes in place to verify the validity and reliability of registered information and data, so that the Seumteo database can be considered one of the most reliable construction information databases in South Korea.
4. Velocity: As the Seumteo database accumulates the registration records submitted and approved by the local government offices, the dataset gets updated on a monthly basis. To provide a more accurate estimation of life span, the utilization of new data is essential so that the model can respond to changes in the building characteristics or new trends in the construction industry. Therefore, methodologies with real-time analytical and evidence-based planning are necessary. The machine learning-based approach in this research is an adequate choice to address such challenges.
5. Variability: In terms of variability, there is a limitation in the database and the dataset we utilize. It can be asserted that the meaning in data changes as the requirements, terms, structures of construction registration change over time. However, the Seumteo database aims to provide a unified set of data for its users so that the variability will be managed according to the legislation.
6. Visualization: Due to the sheet volume of the data and the diverse set of features, it is a big challenge to represent key information in the dataset in a pictorial or graphical layout. The dataset we utilize typically requires the use of a geographic information system to produce real-time, geographical rendering of registered building information.
7. Value: One of our goals in this research is to show the discrepancy between life span calculated by conventional estimation methodology and the actual life span recorded. The Seumteo database can contribute to the comparison with the actual data recorded in the dataset. In general, to estimate or forecast an accurate life span of a building, it is necessary to identify key building characteristics and trace their influences on the life span. The Seumteo database includes a huge amount of data regarding the buildings and construction activities in South Korea that would help to validate and find independent variables that affect the building life span.

3.1.3. *Focus and approach of analysis*

In this research, we chose South Korea as the research backdrop for this study due to the extent and availability of its construction administration records as well as the comprehensibility of data by the authors. The overall flow of the big data analysis approach used to predict a building's life span for use in LCA or LCC is shown in Fig. 1.

The big data analysis approach was implemented in the following steps:

1. An understanding of the characteristics of the target building group was developed.
2. Big data were collected describing as many factors as possible that were likely to affect the life span of the buildings.
3. Predictive models suitable for data preparation and big data characteristics were developed.
4. Several iterations, from the previous stage to the predictive models were run until a model with satisfactory performance was achieved. If the prediction was not acceptable, all steps were repeated starting from Step 1. There was no specific thresholds (e.g., minimum RMSE) but each iteration required adjusting parameters for the model to get a better outcome, which was indicated by the difference between the estimated life span and the actual life span. Thus, if the subsequent trial did not produce a better estimation result than the previous one, the parameters of the trial were disregarded and a new trial was continued. Note that even if the problem is not solved by running the process once from beginning to end, it is not a failure; the results can be used to improve the model.
5. Once an acceptable prediction was achieved, the model could be deployed in LCA and LCC analyses and used to support decision making, and the results could be fed back into the model to better understand the building characteristics.

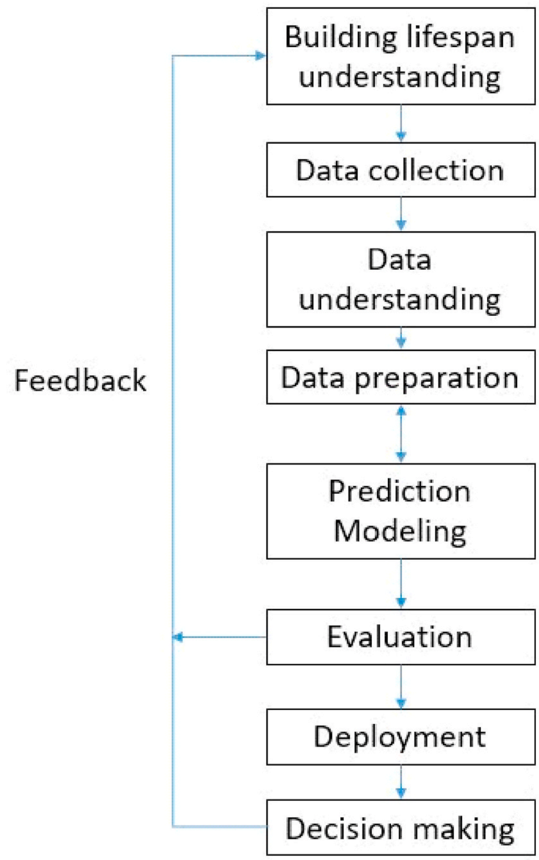

Fig. 1. Flow of building life-span prediction approach using big data.

Indeed, every step of this process explores the data. Thus, the more iterations are run, the more information is generated, resulting in better models [44].

3.2. *Building life span in South Korea*

In South Korea, the life span of a building is considerably shorter than the one assigned to it according to its main structure type. Most buildings in South Korea consist of RC, steel frame, or steel frame–RC structures; among these, 90% are RC structures [35]. Because limited types of structures have been used in South Korea, a large amount of data are available for each structure. Over the period covered in this study, there has been no change in the durability of buildings, and no drastic changes in construction methods, materials, or real estate and building laws have occurred that would significantly affect the life span of buildings in South Korea. In addition, the mandatory legal enforcement of construction practices in South Korea means that there is almost no omission of building data.

3.3. Exploratory study based on building registration records

To identify the characteristics affecting the life spans of buildings, the average life spans of each main frame type in each region were examined using data from an authorised register on demolition permits in the Electronic Architectural Administration Information System (EAIS). These data, described in detail in Appendix, constitute a total of 1,812,700 cumulative demolition records of buildings since the 1950s to December 2019, along with the characteristics of the buildings (i.e. main frame type, site, gross area, etc.). In this study, after removing cases including missing values, misinputted data values and outliers from the database, 971,514 building records were used to analyse building life spans by major structural type and regions. The number of registered records and the average life span from each period are listed in Table 2. The EAIS databased built on the records which were constructed manually from 1950s by local governments personnel. The old handwritten data was transformed and digitized so that the database included a large number of missing and misinputted values. Among 1,812,700 records, 692,894 records with missing values, 148,284 records with misinputted data variables and 8 outliers, which have extremely long life span (>200) were removed. Our primary focus of this research was to propose a better methodology to estimate building life span suitable for LCC and LCA for general buildings, excluding extremely old buildings not typical for ordinary residence purposes.

Interestingly, according to Table 2, fewer than 250 buildings per year were demolished between the 1950s and 1970s. However, the reason for this aberration appears to be that many records from immediately after the end of the Korean War (1953) were lost in the process of digitalisation, as it was often difficult to record data electronically because the existing handwritten records were too old to legitimise their contents. In addition, we have presumed that the number of demolition records would only increase as we approach the modern age because more data would be stored in a computerised format, and there would be less data loss. Thus, as the data are more recent, they are more likely to be complete with no missing values and therefore to be included in the analysis. In this study, the actual life span was calculated as the difference between the year of demolition and the year of approval of use.

3.4. Machine learning-based building life-span estimation using building registration records

The primary objective of this study was to predict the most realistic building life span using various machine learning techniques. Building life span was predicted based on an authorised register of the demolition permit dataset from EAIS, seismic regional classification from Regulation of Building Structure Standards (RBSS), and climate regional classification and insulation requirement information from Energy Saving Design Standards of Buildings (ESDSB) [45,46]. A prediction model containing overall 21 sets of building information as the independent variables was thus developed using 19 sets of building information provided by EAIS and 2 sets of building information on seismic and climatic conditions provided by RBSS and ESDSB. The information describing these independent variables is provided in Table 3. The analyses were performed using Python 3.5 on an Intel i5 8th-Gen CPU computer with 16 GB RAM running Windows 10. The software package was managed using Anaconda 4.8.4, and Pycharm 2020.02 was used as the IDE for the actual programming and various library applications.

Table 2
Demolition records and average life cycle by period within the records to be analysed.

Period	Number of demolition records (Cases)	Average building life span (Years)
1950s	53	12.7
1960s	34	12.8
1970s	237	22.9
1980s	2324	20.3
1990s	11,741	21.8
2000s	357,849	24.9
2010s	599,276	30.0

Table 3
Building-related information for life-span prediction.

		Factors	Details	Source
1	Internal	Building area	Sum of the floor area of all floors in all buildings on a site	EAIS
2	Internal	Building floor area	Floor area of building(s) is the sum of the gross horizontal areas of several floors	EAIS
3	Internal	Building structure type	Main structure types that make up a building	EAIS
4	Internal	Roof structure type	Types of major structures that make up the roof of a building	EAIS
5	Internal	Number of storeys above ground	Number of storeys above ground of a building	EAIS
6	Internal	Number of storeys under ground	Number of storeys underground of a building	EAIS
7	Internal	Number of elevators	Number of elevators in a building	EAIS
8	Internal	Number of outdoor parking spaces	Number of outdoor parking spaces of a building	EAIS
9	Internal	Area of outdoor parking spaces	Area of outdoor parking spaces	EAIS
10	External	Approval date of use	Completion date of construction	EAIS
11	External	Building Type	Detached building; Aggregate building	EAIS
12	External	Building ledger Type	General building; Section for describing the title	EAIS
13	External	Metropolitan city	Metropolitan city where building is located	EAIS
14	External	City or county	City or county where building is located	EAIS
15	External	District	District where building is located	EAIS
16	External	Site Type	Site; Forest land; Road; Rice paddy; Field	EAIS
17	External	Extra parcels	Number of extra parcels	EAIS
18	External	Building subordination type	Main building; Sub-building	EAIS
19	External	Major use classification	Major use of building	EAIS
20	External	Seismic regional classification	Classification of seismic hazard by location of buildings	RBSS
21	External	Climatic regional classification	Classification according to the heat perfusion rate table of buildings by region	ESDSB

3.5. Development of machine-learning-based life-span estimation model

3.5.1. Model development overview

Machine learning commonly involves training a model using data and evaluating the ability of the model to generalise to new test data. In addition to the classification of data, machine learning can be flexibly applied to solve various problems due to its ability to identify key relationships between variables and predict patterns through data learning [47]. Because machine learning makes it easy to analyse data using large volumes of multi-dimensional variables that are otherwise difficult to handle using traditional statistical methods, we considered machine learning a suitable method for developing a prediction model using the data collected in this study.

When applying machine learning, the conditions required to obtain optimal results depend on the problem and the characteristics of the data. Therefore, it is necessary to apply techniques that are specific to the problem. Because the life cycle of a building is a continuous variable, we constructed a prediction model using simple regression method as a baseline and compared it with three other state-of-the-art machine learning models to assess their potential improvements relatively.

3.5.2. Regression model

For the baseline condition, the linear regression method from the Scikit-Learn Python library was used. A linear regression model seeks to describe the relationship between independent variables ($x_1, \cdots, x_n$) and a continuous dependent variable (y) by fitting a linear equation ($\hat{y} = w_1 \times x_1 + \cdots + w_n \times x_n + b$) to the data. The linear regression module provided by Scikit-Learn uses the singular value decomposition (SVD) method to calculate the pseudo-inverse matrix to build a model. Like traditional statistical analysis, a model that minimises the root mean square error (RMSE) and maximizes the coefficient of determination (R^2) is considered to be more accurate.

3.5.3. XGBoost ensemble model

Extreme gradient boosting (XGBoost) is an improved model of the decision tree boosting technique presented by Chen and Guestrin [48]. Boosting is an ensemble technique that combines numerous weak classifiers to create a strong classifier that can compensate for the errors from the previous steps while sequentially executing the weak-tree classifiers. In this manner, XGBoost improves the slow learning speed of boosting through parallelisation and is accordingly gaining popularity by proving its excellence in the data-analysis-competition platform Kaggle (http://www.kaggle.com).

3.5.4. LightGBM ensemble model

The light gradient boosting machine (LightGBM) algorithm decreases the data processing time using the gradient-based one-side sampling and exclusive feature binding techniques. Its speed is thus far greater than that of the XGBoost and quantile gradient boosting regression tree models, which use the existing gradient boosting decision tree [49].

3.5.5. Deep neural network model

As an alternative to XGBoost and LightGBM, we employed an artificial neural network (ANN) based on TensorFlow to implement a deep learning neural network (DNN) model. TensorFlow has become the most used machine learning software library since Google released it as a free and open-source library in November 2015 [50]. The ANN is a data processing system consisting of multiple layers, variable connection strengths, a transition function, and a learning algorithm, and has a structure in which weights are repeatedly adjusted through the input and output data values to eventually reflect the relationship between the two.

In this study, we constructed a hidden layer in the form of an inverted pyramid to output a single dependent variable using a limited number of independent variables. Thus, five hidden layers were used to minimise the mean squared error between the predicted and actual building life spans (Fig. 2). The rectified linear unit (ReLU) was used as the activation function of the hidden layer, and the Adadelta optimiser provided by Keras was used as the gradient boost optimiser.

3.5.6. Pre-processing and model parameter tuning

We pre-processed and optimised the model parameters using a data set to adjust the model performance and training time. To pre-process the data, we used the pre-processing and scaler functions of Scikit-Learn. The training and test data sets were normalised to values between 1 and 0 using the MinMaxScaler method provided with the pre-processing package. Additionally, we sought the optimal model parameter using the GridsearchCV function and Bayesian optimiser packages provided in Scikit-Learn.

3.5.7. Model training and evaluation

Building life spans were trained and predicted by developing the regression, XGBoost, LightGBM, and DNN models using the demolition permit dataset. 971,514 cases, which included data from January 1950 to December 2019 were used to train the prediction models. For test data, we acquired a recently generated dataset separately to test the performance of the model trained with records up to the 2010s. The most recently recorded authorised register on demolition permit dataset (38,676 cases) from January 2020 to June 2020 was used as a test set. The new test dataset was preprocessed in the same manner as training data. Overall, a total of 1,010,190 records from January 1950 to June 2020 were used in this experiment. The resulting RMSE and R^2 were calculated and compared to verify the performance of each model.

4. Results

4.1. Exploratory building life-span analysis

We conducted an exploratory analysis of the data to develop an initial understanding of how building life span changes by main frame type and regional differences.

4.1.1. Analysis of number of buildings by main frame

The number of buildings classified by main frame type and their corresponding percentages of representation in the data are shown in Fig. 3. There were 25 types of main frames in the data. In Fig. 3, the main frame types that accounted for less than 1% of the data, including steel, cement block, and container, are grouped together and indicated as 'Others'. With 36%, brick buildings accounted for the largest proportion of the data, followed by block, RC, general wood, lightweight steel, steel, steel pipe, masonry, and others. Thus, RC buildings, which were assumed to be predominant, accounted for the third largest proportion of buildings in the data set. Although RC has been widely used since the 1950s, brick, block, wood, lightweight steel, etc. are suitable for low-rise buildings, are fireproof, durable, and easy to construct, and as a result were widely used after the Korean War.

Fig. 4 illustrates the percentages of buildings by region in South Korea according to main frame type. Brick buildings constitute the largest share of building frame type in of South Korea overall, and in Seoul, Gyeonggi, Daegu, and Chungbuk. However, block buildings are the most common in Busan and Jeju, whereas wooden buildings are most common in Gangwon, Gyeongnam, Jeonnam, and Jeonbuk, and lightweight steel buildings are most common in Jeonbuk. It is presumed that the preference for a particular frame type varies according to the

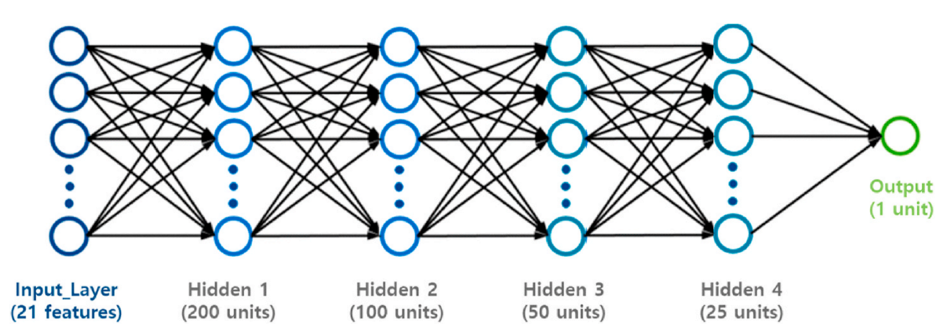

Fig. 2. Visual representation of the structure of a deep neural network.

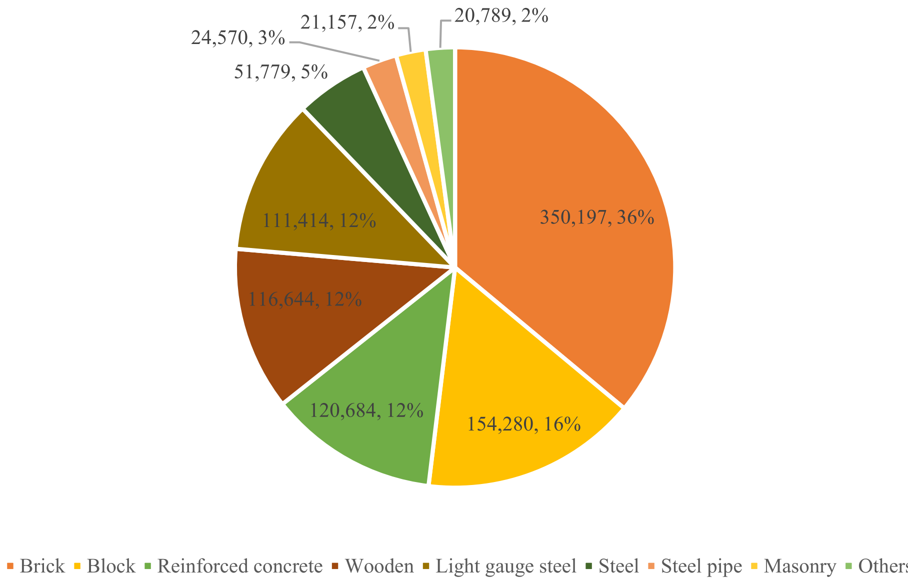

Fig. 3. Number of buildings according to main frame type.

site conditions in each region.

4.1.2. Analysis of building life span by main frame type

To examine how the building life span changes by main frame type, the average of building life span was computed for the ten most used main frame types identified. Fig. 5 illustrates the average building life span for the eight most used main frame types, excluding the two of the top ten that accounted for less than 1% in total. As indicated in the graph, the building life spans exhibited large variations according to main frame type. This overall trend reveals that the differences between the main frame types have a significant influence on the average life span. In addition, it is commonly understood that there can be a

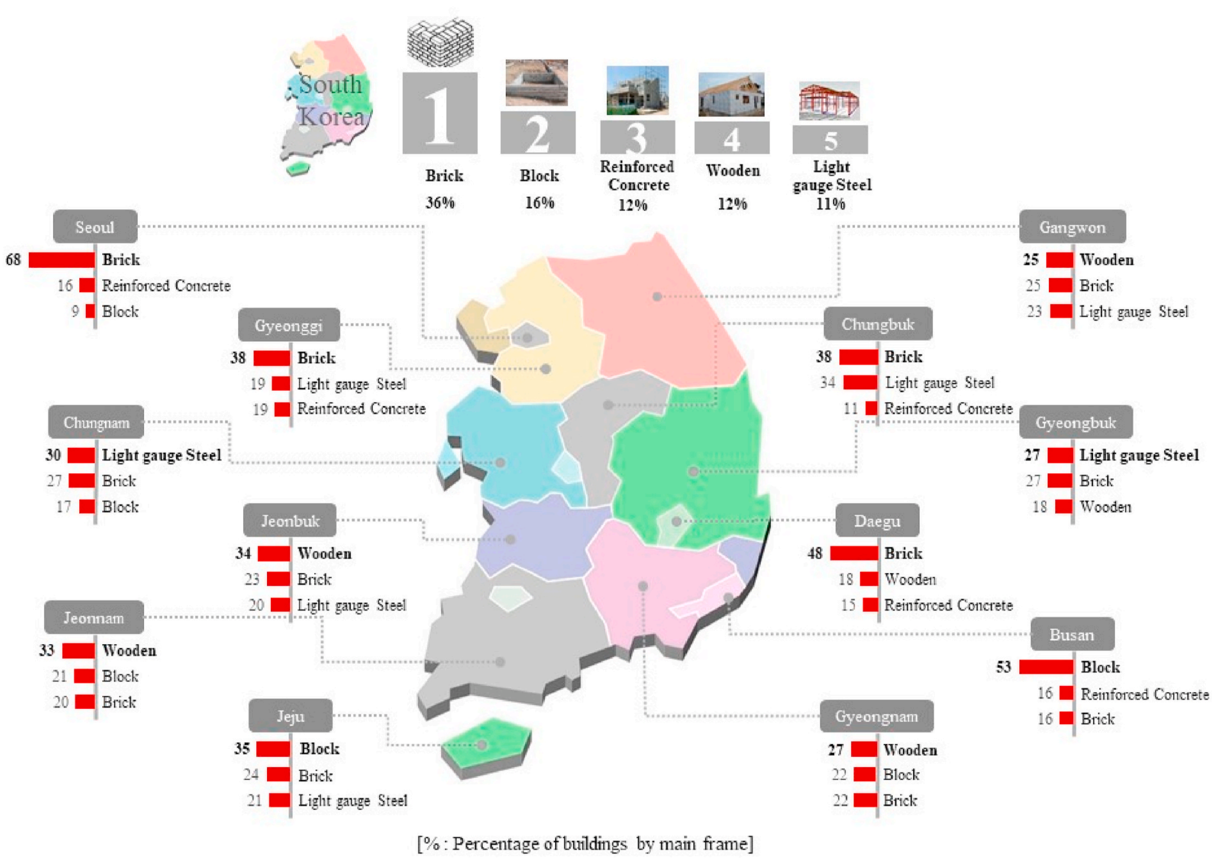

Fig. 4. of buildings in South Korean regions by main frame type.

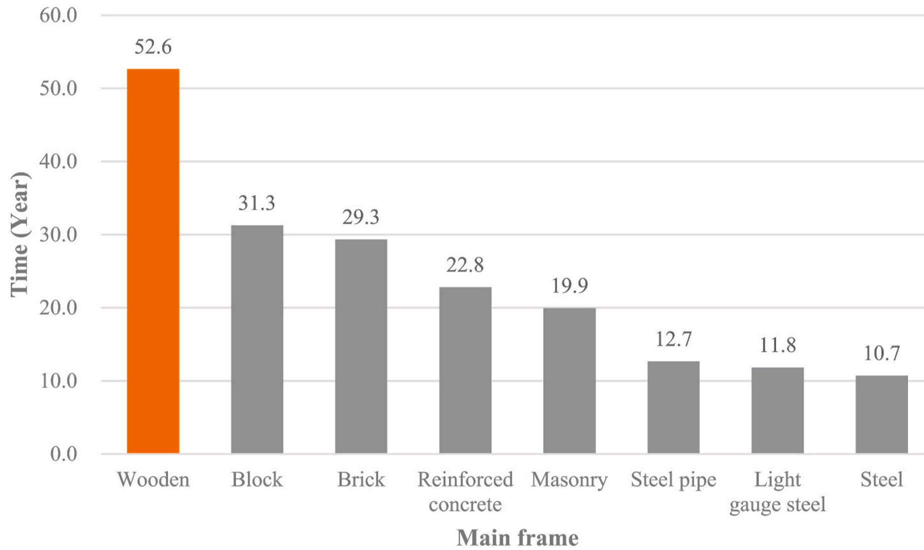

Fig. 5. Distribution of average building life span by main frame type.

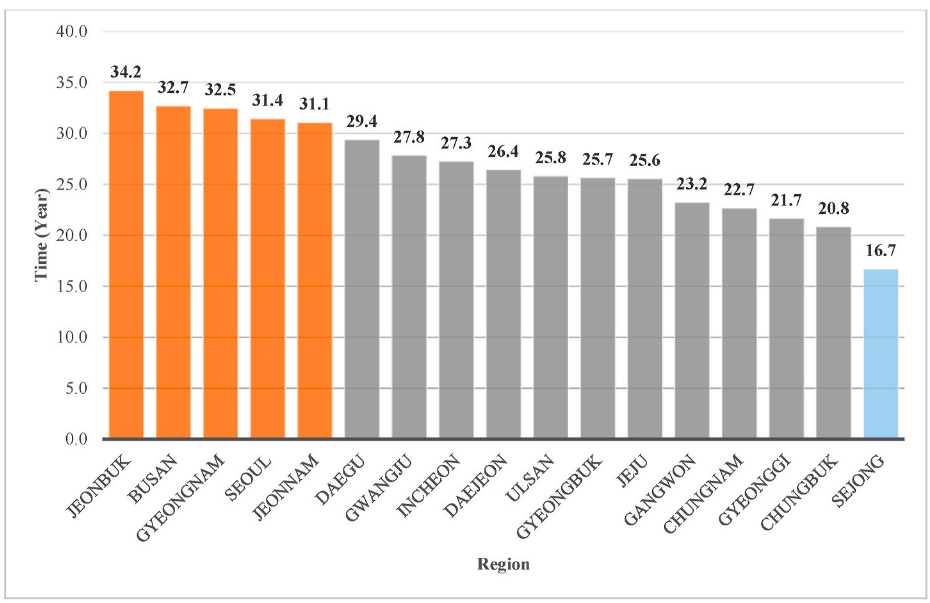

Fig. 6. Average building life span according to region.

significant difference between the durability life span of a building and its actual life span according to the main frame type. For example, a wooden building is generally known to have a life span of 50–100 years but has an actual average building life span of 52.6 years, which is close to the lower limit. The assumed life spans of block and brick buildings are approximately 50 years each, whereas their actual average life spans were found to be 31.3 and 29.3 years, respectively. Buildings made of RC, which is the most used material for new buildings today, have an assumed durability life span of 50–100 years, but an actual average life span of only 22.8 years.

4.1.3. Analysis of building life span by region

The average building life cycle and its distribution were analysed by region according to the classification of metropolitan cities and districts in South Korea. As indicated in Fig. 6, there was little variation between regions in terms of average life cycle. In particular, there were five regions with an average building life of more than 30 years: Jeonbuk (34.2 years), Busan (32.6 years), Gyeongnam (32.4 years), Seoul (31.4 years), and Jeonnam (31 years). Notably, there were no specific factors distinguishing different regions. Therefore, it cannot be concluded that the characteristics of a region (i.e. local vs. city, regional building preference, etc.) are key factors in determining the life span of a building. However, in the case of Sejong City, the average life span was found to be far lower than the other regions, a difference that may reflect the rapid, large-scale demolition of existing buildings occurred recently in accordance with the creation of a newly planned city established by the Korean government policy.

4.1.4. Analysis of building life span according to main frame type by region

Fig. 7 depicts the average life spans of buildings according to the most commonly used main frame types throughout South Korea. In all the regions, wooden buildings were found to have the highest average building life span. Although there were slight differences between regions, brick and block buildings had the second and third longest life spans, respectively. In most areas, RC buildings and gauge steel buildings had the fourth and fifth longest life spans, respectively. There were

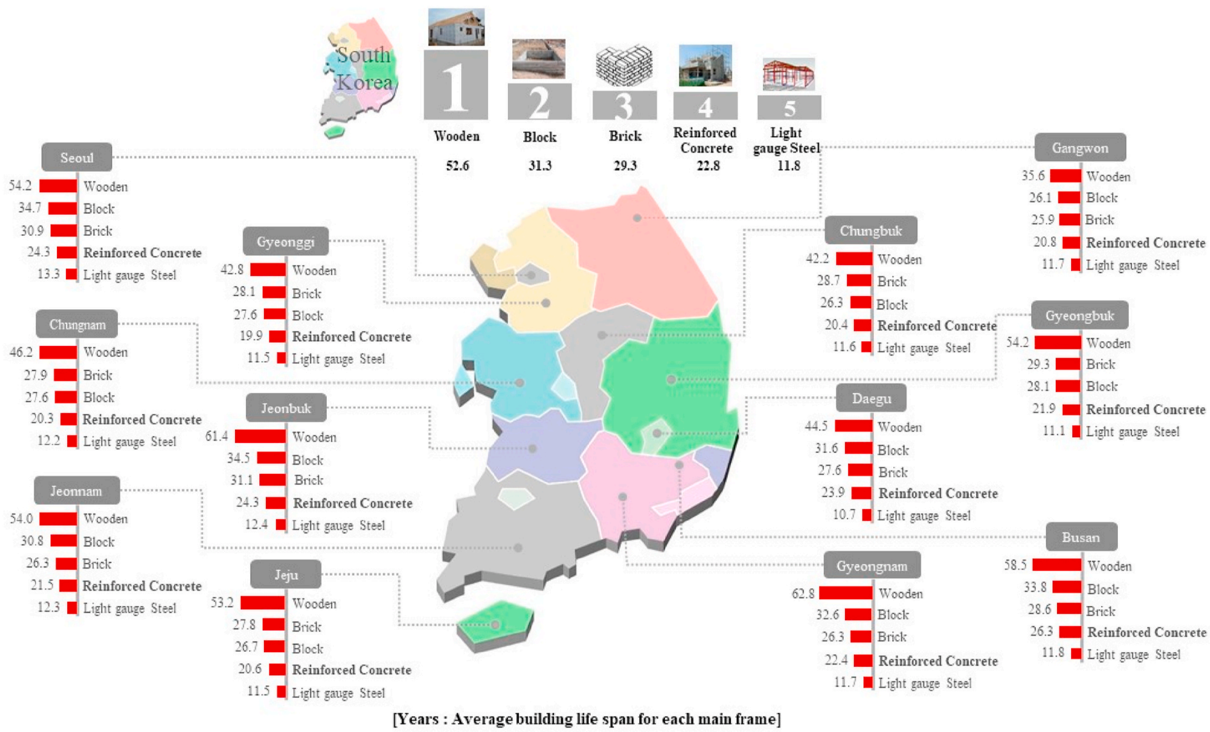

Fig. 7. Average building life span by region for each main frame type.

Table 4
Performance comparison between machine learning models.

Model	RMSE	R^2
Linear regression	4.53	0.934
XGBoost	4.60	0.932
LightGBM	4.33	0.939
DNN	3.72	0.955

no significant differences between the life spans of buildings with the same main frame type according to region. The average RC building life span in Busan was the longest at 26.3 years, followed by Seoul and Jeonbuk, both with an average life span of 24.3 years, and Daegu with the third longest life span of 23.9 years. The region with the shortest average building life span was Gyeonggi Province at 19.9 years. Notably, Seoul is the largest city, with the largest number of residents and buildings in South Korea, while Daegu and Busan are also large cities with relatively high population and building densities. Therefore, the life spans of buildings must be long. However, it is difficult to make a blanket statement based on this simple analysis because Jeonbuk is not a densely populated area.

4.2. Analysis of prediction models and comparison of their performances

The analysis results obtained using the four predictive models are shown in terms of RMSE and R^2 in Table 4. As the RMSE represents an error value, the smaller the number, the smaller the error. In regression model evaluation with machine learning methodology, there is no consensus on a standard metric to assess the results of regression itself. The mean square error (MSE) and its rooted variant (RMSE), or the mean absolute error (MAE) and its percentage variant (MAPE) were often employed to evaluate prediction outcomes [56]. In this research, we used RMSE to evaluate the model outcome because the metric is robust to the influence of outliers and their errors [57]. Even though we standardized values and preprocessed outliers and abnormal values in our dataset, there is the possibility that a few large values may distort the model estimation due to its great volume and large variance from the real data. Therefore, it is necessary to use a metric robust to generalized predictions. Other than prediction precision, we also needed to check if the regression model is well representing the observed value and distribution of life span. Therefore, we present R^2 value (also known as the coefficient of determination). There are arguments in prior studies that SMAPE (symmetric mean absolute percentage error) should be used for such criteria, however, recent comparative research claimed superiority of R^2 over SMAPE [56]. The RMSE of the linear regression model, XGBoost regression model, LightGBM model, and DNN model were distributed between 3.72 and 4.6, and the coefficient of determination was greater than 0.9. The examined models were all able to predict the life span of buildings in South Korea reasonably well and the DNN model was the best among them.

The performance of the XGBoost model was slightly weaker than the regression model. According to prior research, XGBoost is expected to show similar prediction performance as the LightGBM model because they both are based on gradient boosting [49]. It is difficult to explain precise reasons for the weaker performance of XGBoost, but the

Table 5
Key variables for each model sorted in the order of importance.

Regression model	Variables
Linear regression	Date of use permit
	Number of storeys above ground
	Number of storeys underground
	Number of elevators
	Metropolitan city
XGBoost	Date of use permit
	Roof structure type
	Major use
	Building area
	Building floor area
LightGBM	City or county (Si, Gun, Gu)
	Date of use permit
	District (Dong)
	Major use
	Main frame type

implementation details of the two algorithms seem responsible for the subtle difference between the two gradient-boosting models we have utilized in the study. The two algorithms are different in how they manage the nodes in splitting decision trees based on the gradient. LightGBM uses Gradient-based One-Side Sampling (GOSS) to filter out the data instances for finding a split value. This approach applies a multiplier constant to nodes with a small gradient, which allows focusing on under-trained features. However, XGBoost uses a pre-sorted algorithm so that the model can be robust to overfitting but may disregard instances with small gradients.

To further understand the comparative characteristics of the models, we examine the key features of each model we constructed. Except for the DNN model, the importance of each variable to the models can be determined based on the weight or importance of the associated feature in each model. Table 5 lists five main variables with the strongest influence on the predicted values for each model.

Table 5 reveals that each model has a different set of important features. They all agree that the date of use approval is an important variable. However, there are significant differences in the combination of the five variables and their importance levels across the prediction models. Therefore, it shows that each model operates in a distinct way, implying that it is important to consider a diverse set of variables and the influence of those variables together. To more directly examine the problem of conventional life-span estimation, which employ main frame type and region as the independent variables to estimate building life span, we applied those two variables to the same machine learning estimation protocols we used in this research. Table 6 shows the RMSE and R^2 values for each of the predictive models using either or both of the major structural types of buildings and regions, which are most commonly employed features in determining the life span of a building in LCA or LCC studies.

As shown in Table 6, RMSE values are significantly greater than the values presented in Table 4, revealing that there are large discrepancies between real building life span and estimated building life span, which is predicted based on one of, or both of, the structural and regional factors. Therefore, the proposed methodology of this research should be considered superior over the traditional estimation approaches.

5. Discussion

In this study, the prediction of building life span was attempted for the first time based on real-world administrative big data that spans over 65 years of a whole nation. Big data analysis performed in this study showed that the actual life spans of the building differ substantially from those calculated using the mainframe type or region variables, indicating that the current estimation practices simply based on those variables are far from the reality. Despite its widespread use, it is not appropriate to perform an LCA or LCC using the simple estimation of a building's life span that is significantly different from the actual one.

Machine learning methods, especially the DNN, depend less on preclassified key factors (or features) while providing more accurate results than other prediction methods. Because machine learning methods can utilize patterns that exist in the given data set that are otherwise difficult to identify, they represent a suitable approach to solving highly complex problems, including building life span prediction. Thus, as demonstrated in this study, we can build a computational model using machine learning methods that is more accurate than the currently employed life cycle pattern analysis method (which is based on assumptions regarding building characteristics), providing improved predictive power over general regression analysis.

Table 6
Performance comparison between machine learning models estimated with mainframe type and region.

Factor variable	Model	RMSE	R^2
Main frame Type	Linear regression	18.40	−0.080
	XGBoost	20.30	−0.315
	LightGBM	20.31	−0.316
	DNN	19.76	−0.246
Region	Linear regression	18.03	−0.037
	XGBoost	17.90	−0.023
	LightGBM	17.90	−0.023
	DNN	17.97	−0.030
Main frame Type & Region	Linear regression	18.19	−0.005
	XGBoost	18.96	−0.148
	LightGBM	20.13	−0.294
	DNN	19.18	−0.175

However, it should be noted that the methods evaluated in this study have several potential limitations. First, it is difficult to predict the life span of buildings constructed using innovative methods and materials, as there are no actual life span data that reflect the life span of such futuristic buildings. Innovative methods have the potential to drastically increase the general life span of a building. Another limitation may arise from scenarios in which current building laws have been amended or a natural calamity has occurred, suddenly impacting building life spans. However, the current trend of using RC, steel, and wood is not likely to change for a while. Moreover, although a few housing laws and building laws may be enacted, altered, or abolished, no change in the basic construction paradigm is likely to occur. Authorities have endeavoured to prolong the life spans of recently constructed buildings by responding to the predictable functional and environmental changes and improving structural durability. If changes occur in the future that directly affect building life span, such as the introduction of completely different construction methods like building 3D printing technology, popularization of new materials, or revision of re-development and reconstruction laws that affects building profitability, it will be necessary to consider a new hybrid research method that includes a more conventional factor method as well as the big data approach employed in this study.

It is important to note that the proposed methodology and prediction models are estimated based on the prior records of the buildings that are effectively demolished. Therefore, the prediction may not reflect the characteristics of existing buildings with long life spans and buildings with different feature variances which may occur in the future. To mitigate the bias and the discrepancy between the model prediction and such untrained life span cases, we constructed and trained the models with the records from 1950 to 2019, and have the model predict the life span of the buildings demolished in 2020, which are not reflected in the trained models. Even though, there still exists the potential limitation of model performance for cases with different feature variances, our study has demonstrated that the proposed performs better than the conventional approach. Moreover, there is a room for improvement as the construction datasets are continuously being collected and applied to the model training.

The most practical problem that we wanted to address in this research was to introduce an improved methodology that estimates the realistic life span necessary for accurate LCC and LCA results. Thus, we aimed to employ and test a methodology that is universally applicable to the life span estimation modeling for LCC and LCA. As a data-driven approach, our proposed model may not be permanent and have limitations in predicting unobserved cases. However, our study results show that our methology based on the utilization of big data and machine learning is good enough to enhance the outcome of LCC and LCA calculations. The methodology in this study can be employed universally, with only minor adjustments for different datasets.

Building construction requires compulsory administrative processes from building permits to construction reports to authorisation of use to demolition. Therefore, if a computerised administrative registration process for maintenance work such as waterproofing, facilities, and replacement of internal and external equipment is implemented, big data can be collected on the replacement cycle and cost of materials or construction methods during the maintenance stage. Thus, life expectancy could be accurately predicted using big data for construction methods or building materials using the same method presented in this study. In addition, including the required quantity and cost in the record items generated during the maintenance process would help more accurately analyse the economic and environmental impacts of building life cycle characteristics.

6. Conclusion

The objectives of this study are to investigate how much the presumed building life span, which has been commonly utilized to perform LCA or LCC analyses, differs from actual building life span and to find a better approach that can accurately predict building life span, thereby contributing to more accurate practices of LCA or LCC analysis. In prior LCA or LCC studies, the life span was often assigned to be either 50 or 100 years old depending on the structures. However, our analysis results show that the actual life span of the building was very different from that of the 50 years or 100 years that was commonly used in those studies. Therefore, there is a dire need to reflect a realistic building life span in the LCA and LCC, rather than rely on an incorrectly presumed one. Responding to this need, we have explored the possibility of applying prominent machine learning approaches to the data set of South Korean building registration and demolition records (971,514 cases). Our study results clearly show that the actual building life spans are very different from those presumed numbers. In the case of a reinforced concrete structure building, which is currently the most used structure in South Korea, the average life span of the actual buildings was 22.8 years, with a difference of 27.2 years from the standard reference of 50 years. In the case of the brick structure building, which accounts for 36% of the total buildings in the study, the life span of the building was 29.3 years, with a difference of 20.7 years.

In this study, we have also sought an effective way to predict realistic building life spans by applying the latest machine learning prediction methods. Building life spans were predicted using four prediction models including linear regression, XGBoost, LightGBM, and Deep Neural Network (DNN). With this big data approach, even the basic linear regression model showed a fairly powerful prediction accuracy of 93.4%. Moreover, among the alternative machine learning models examined in this study, the prediction accuracy of the latest DNN model was 95.5%, which was 2.2% better than the linear regression model. Overall, the big data approach shows that those machine learning models are far more accurate in predicting actual life spans of buildings, relative to the current practices relying on the presumed numbers, and DNN is the most effective model among those models examined in this study. The findings of this study contribute to the prediction of realistic building life spans, opening up a new horizon for various evaluations and decision making practices in the building construction industry.

Table A1
Data and type by factor

Factors	Data	Type
Building area (m^2)	0–6,316,730	
Building floor area (m^2)	0–204,930,809	
Main frame	'RC', 'Brick', 'Steel frame', 'Light gauge steel', 'Block', 'General wood', 'Masonry', 'General steel', 'Stone', 'Steel pipe', 'Wood', 'Concrete', 'Steel RC', 'Precast concrete', 'Steel concrete', 'Steel house', 'Log', 'Cement block', 'Prefabricated panel', 'Steel pipe', 'Soil brick', 'Structure consisting of beams, columns, and slabs', 'Container', 'Steel RC synthesis', 'Truss wood'	
Roof structure type	'RC', 'Other roof type', 'Slate', 'Tile'	
Number of storeys above ground	1–274	
Number of storeys underground	0–51	
Number of elevators	0–93	
Number of outdoor parking spaces	0–9012	
Outdoor parking area	0–495,420	
Date of approval of use	1800–2018	
Building type	'Detached building', 'multi-unit dwelling'	
Building ledger type	'Section for describing the title', 'General building'	
Metropolitan city	'Seoul', 'Gyeonggi', 'Gangwon', 'Daegu', 'Incheon', 'Gwangju', 'Ulsan', 'Gyeongnam', 'Jeju', 'Daejeon', 'Chungbuk', 'Chungnam', 'Jeonbuk' ', 'Jeonnam', 'Gyeongbuk', 'Busan', 'Sejong'	
City or county code (Si, Gun, Gu)	11,110–50130	
District code (Dong)	00000–47029	
Site type	0,1,2	
Extra parcels	0–4734	
Building subordination type	'Main building', 'Sub building'	
Major use classificatiral accon	'Multi-unit dwelling', 'detached building', 'class 2 neighbourhood living facilities', 'educational research and welfare facilities', 'class 1 neighbourhood living facilities', 'neighbourhood living facilities', 'animals and plants related facilities', 'warehouse facilities', 'factory', 'business facilities', 'offices', 'automotive facilities', 'correctional and military facilities', 'accommodation facilities', 'education and research facilities', 'dangerous goods storage and treatment facilities', 'public facilities', 'elderly people facilities', 'excretion and waste disposal facilities', 'tourist rest facilities', 'cultural and assembly facilities', 'sales and sales facilities', 'religious facilities', 'amusement facilities', 'medical facilities', 'retail stores', 'other warehouse facilities', 'exercise facilities', 'singing practice grounds', 'training facilities', 'sales facilities', 'restaurants', 'multi-family housing', 'oil sales office', 'academy', 'joint house', 'congratulations', 'repair store', 'greenhouse', 'community office', 'dormitory', 'public office', 'market', 'religious assembly hall', 'apartment', 'cemetery related facilities', 'general restaurant', 'manufacturing establishment', 'quarantine hospital', 'transport facility', 'multiple housing', 'shop', 'church', 'police box', 'general business facility', 'multi-family housing', 'gymnasium', 'broadcasting and communication facilities', 'temporary buildings', 'power generation facilities', 'shelters', 'golf driving range', 'power plants', 'kindergarten', 'dangerous goods storage', 'health centre', 'research institute', 'military facility', 'other class 1 neighbourhood facility', 'funeral facility', 'geneommodation facility', 'industrial exhibition centre', 'cosmetic shop', 'wholesale market', 'day-care centre', 'hospital', 'apse chapel', 'tourist accommodation', 'living area training facility', 'garage', 'prison', 'storage shop', 'general factory', 'slaughterhouse', 'liquefied gas handling station', 'parking lot', 'children-related facilities', 'water purification plants', 'waste treatment facilities', 'resource recycling facilities', 'hotels', 'gas stations', 'inns', 'subsidiary facilities', 'living facilities', 'temples', 'social welfare facilities', 'camping facilities', 'bowling alley', 'reading room', 'science hall', 'post office', 'other sales and sales facilities', 'athletic stamp', 'substation', 'airport facilities', 'entertainment tavern', 'excretion treatment facility', 'youth hostel', 'chicken factory', 'newspaper', 'waste recycling facility', 'exhibition hall', 'theatre (movie theatre)', 'drug	

(continued on next column)

(continued on next page)

Table A1 (*continued*)

Factors	Data	Type
	clinic', 'welfare facilities', 'broadcasting stations', 'other public facilities', 'other animals and plants related facilities'	
Seismic regional classification	I, II	
Climatic regional classification	Central region 1, central region 2, southern region	

* Some data outliers were found, but the frequency of data with a building area of 1,000,000 m^2 or more is 1 and that with a total floor area of 3,000,000 m^2 or more is 2, the data frequency of more than 50 storeys above ground level is 2, that of more than 10 basement storeys is 7. It was confirmed that there were only four data frequencies exceeding 1000 other parcels and had little effect on the overall big data analysis.

Table A2
Earthquake area and area coefficient

Earthquake area classification	Administrative district		Earthquake area coefficient
I	City	Seoul, Busan, Incheon, Daegu, Daejeon, Gwangju, Ulsan, Sejong	0.22 g
	Province	Gyeonggi, South Gangwon[1)], Chungbuk, Chungnam, Jeonbuk, Jeonnam, Gyeongbuk, Gyeongnam	
II	Province	North Gangwon[2)], Jeju	0.14 g

Annotation
1) South Gangwon: Gangneung, Donghae, Samcheok, Wonju, Taebaek, Yeongwol, Jeongseon
2) North Gangwon: Sokcho, Chuncheon, Goseong, Yanggu, Yangyang, Inje, Cheorwon, Pyeongchang, Hwacheon, Hongcheon, Hoengseong

Table A3
Classification by region according to climate in Energy Saving Design Standards of Buildings

Region classification	Administrative district
Central region 1	Gangwon (excluding Goseong, Sokcho, Yangyang, Gangneung, Donghae, and Samcheok), Gyeonggi (Yeoncheon, Pocheon, Gapyeong, Namyangju, Uijeongbu, Yangju, Dongducheon, Paju), Chungbuk (Jecheon), Gyeongbuk (Bonghwa, Cheongsong)
Central region 2	Seoul, Daejeon, Sejong, Incheon, Gangwon-do (Goseong, Sokcho, Yangyang, Gangneung, Donghae, Samcheok), Gyeonggi (excluding Yeoncheon, Pocheon, Gapyeong, Namyangju, Uijeongbu, Yangju, Dongducheon, Paju), Chungbuk (excluding Jecheon), Chungnam, Gyeongbuk (except Bonghwa, Cheongsong, Uljin, Yeongdeok, Pohang, Gyeongju, Cheongdo, Gyeongsan), Jeonbuk, Gyeongnam (Geochang, Hamyang)
Southern region	Busan, Daegu, Ulsan, Gwangju, Jeonnam, Gyeongbuk (Uljin, Yeongdeok, Pohang, Gyeongju, Cheongdo, Gyeongsan), Gyeongnam (Excluding Geochang and Hamyang)

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C.M. Mah, T. Fujiwara, C.S. Ho, Life cycle assessment and life cycle costing toward eco-efficiency concrete waste management in Malaysia, J. Clean. Prod. 172 (2018) 3415–3427, https://doi.org/10.1016/j.jclepro.2017.11.200.
[2] A.A. Jensen, J. Elkington, K. Christiansen, L. Hoffmann, B.T. Møller, A. Schmidt, F. van Dijk, Life Cycle Assessment (LCA): A Guide to Approaches, Experiences and Information Sources, European Environment Agency, 1998.
[3] Un Environment and International Energy Agency, Global Status Report 2017: towards a Zero-Emission, Efficient, and Resilient Buildings and Construction Sector, United National Environment Programme, 2017.
[4] B. Unalan, H. Tanrivermis, M. Bulbul, A. Celani, A. Ciaramella, Impact of embodied carbon in the life cycle of buildings on climate change for a sustainable future, Int. J. Hous. Sci. Appl. 40 (1) (2016) 61–71.
[5] A Trend Analysis of Building Life Cycle Assessment (LCA), Ministry of Environment, Korea Environmental Industry Technology Institute, Republic of Korea, 2018.
[6] R. Pernetti, F. Garza, G. Paoletti, D2.2: Spreadsheet with LCCs. A Database for Benchmarking Actual NZEB Lifecycle Costs of the Case Studies, Horizon 2020 Framework, Programme of the European Union, 2018.
[7] Committee on Trade and Investment, Life Cycle Assessment Best Practices of ISO 14040 Series, 2006.
[8] A. Grant, R. Ries, Impact of building service life models on life cycle assessment, Build. Res. Inf. 41 (2) (2013) 168–186, https://doi.org/10.1080/09613218.2012.730735.
[9] A. Grant, R. Ries, C. Kibert, Life cycle assessment and service life prediction: a case study of building envelope materials, J. Ind. Ecol. 18 (2) (2014) 187–200, https://doi.org/10.1111/jiec.12089.
[10] [Table A5] Table of Standard Useful Life and Scope of Useful Life of Buildings, Rules of the Corporate Tax Act, Ministry of Economy and Finance, Republic of Korea, Order No. 792, [Revised] 21 April 2020.
[11] Demolition Plan Check/Permit Application (PDF). Santa Monica. 2017-06-05. [Retrieved] 2020-01-27.
[12] Planning permission—UK government [Retrieved] 27 November 2020, www.gov.uk.
[13] R. Harwood, Planning permission, International Specialized Book Services (24 September 2015). ISBN 9781780434919, [Retrieved] 5 March 2017.
[14] Building Act. People's Republic of China. [Revised] 22 April 2011.
[15] National Building Regulations and Building Standards ACT. Act No. 103, Government Gazette, Republic of South Africa, 1977.
[16] Architectural data private open system, Electronic Architectural administration Information System (EAIS). [Retrieved], https://open.eais.go.kr/. (Accessed 3 November 2020).
[17] Gobierno de Espãna, Código Técnico de la Edificación. DB-SE, Ministerio de la Vivienda, Madrid, 2010.
[18] Iso, Building Construction – Service Life Planning – Part 4: Service Life Planning Using Building Information Modeling, 2014, p. 2014. ISO 156864.
[19] R.M. Cuéllar-Franca, A. Azapagic, Environmental impacts of the UK residential sector: life cycle assessment of houses, Build. Environ. 54 (2012) 86–99, https://doi.org/10.1016/j.buildenv.2012.02.005.
[20] B. Palacios-Munoz, B. Peuportier, L. Gracia-Villa, B. López-Mesa, Sustainability assessment of refurbishment vs. new constructions by means of LCA and durability-based estimations of buildings lifespans: a new approach, Build. Environ. 160 (2019) 106203, https://doi.org/10.1016/j.buildenv.2019.106203.
[21] K. Adalberth, Energy use during the life cycle of single-unit dwellings: Examples, Build, Environ. Times 32 (4) (1997) 321–329, https://doi.org/10.1016/S0360-1323(96)00069-8.
[22] K. Adalberth, A. Almgren, E.H. Petersen, Life cycle assessment of four multi-family buildings, Int. J. Low Energy Sustain. Build. 2 (1) (2001) 1–21.
[23] B. Son, A method of economic analysis for remodeling of deteriorated apartments using the life cycle costing, Architectural Institute of Korea 21 (2001) 73–81.
[24] A.A. Guggemos, A. Horvath, Comparison of environmental effects of steel-and concrete-framed buildings, J. Inf. Syst. 11 (2) (2005) 93–101, https://doi.org/10.1061/(ASCE)1076-0342(2005)11:2(93).
[25] S. Junnila, A. Horvath, A.A. Guggemos, Life-cycle assessment of office buildings in Europe and the United States, J. Inf. Syst. 12 (1) (2006) 10–17, https://doi.org/10.1061/(ASCE)1076-0342(2006)12:1(10).
[26] M.L. Marceau, M.G. VanGeem, Comparison of the life cycle assessments of an insulating concrete form house and a wood frame house, J. ASTM Int. (JAI) 3 (9) (2006) 1–11, https://doi.org/10.1520/JAI13637.
[27] O. Ortiz, C. Bonnet, J.C. Bruno, F. Castells, Sustainability based on LCM of residential dwellings: a case study in Catalonia, Spain, Build. Environ. 44 (3) (2009) 584–594, https://doi.org/10.1016/j.buildenv.2008.05.004.
[28] I.Z. Bribián, A.A. Usón, S. Scarpellini, Life cycle assessment in buildings: state-of-the-art and simplified LCA methodology as a complement for building certification, Build. Environ. 44 (12) (2009) 2510–2520, https://doi.org/10.1016/j.buildenv.2009.05.001.
[29] O. Ortiz, J.C. Pasqualino, G. Díez, F. Castells, The environmental impact of the construction phase: an application to composite walls from a life cycle perspective, Resour. Conserv. Recycl. 54 (11) (2010) 832–840, https://doi.org/10.1016/j.resconrec.2010.01.002.
[30] H. Monteiro, F. Freire, Life-cycle assessment of a house with alternative exterior walls: comparison of three impact assessment methods, Energy Build. 47 (2012) 572–583, https://doi.org/10.1016/j.enbuild.2011.12.032.
[31] S. Tae, C. Baek, S. Shin, Life cycle CO_2 evaluation on reinforced concrete structures with high-strength concrete, Environ. Impact Assess. Rev. 31 (3) (2011) 253–260, https://doi.org/10.1016/j.eiar.2010.07.002.
[32] A. Passer, H. Kreiner, P. Maydl, Assessment of the environmental performance of buildings: a critical evaluation of the influence of technical building equipment on residential buildings, Int. J. Life Cycle Assess. 17 (9) (2012) 1116–1130, https://doi.org/10.1007/s11367-012-0435-6.
[33] X. Zhang, L. Shen, L. Zhang, Life cycle assessment of the air emissions during building construction process: a case study in Hong Kong, Renew. Sustain. Energy Rev. 17 (2013) 160–169, https://doi.org/10.1016/j.rser.2012.09.024.

[34] S. Kim, G.H. Kim, Y.D. Lee, Sustainability life cycle cost analysis of roof waterproofing methods considering LCCO2, Sustainability 6 (1) (2014) 158–174, https://doi.org/10.3390/su6010158.

[35] S. Ji, D. Kyung, W. Lee, Life cycle assessment (LCA) of roof-waterproofing systems for reinforced concrete building, Adv. Environ. Res. 3 (4) (2014) 367–377, https://doi.org/10.12989/aer.2014.3.4.367.

[36] R.S. Heralova, Life cycle cost optimisation within decision making on alternative designs of public buildings, Procedia Eng 85 (2014) 454–463, https://doi.org/10.1016/j.proeng.2014.10.572.

[37] A. Vilches, A. Garcia-Martinez, B. Sanchez-Montanes, Life cycle assessment (LCA) of building refurbishment: a literature review, Energy Build. 135 (2017) 286–301, https://doi.org/10.1016/j.enbuild.2016.11.042.

[38] B. Petrovic, J.A. Myhren, X. Zhang, M. Wallhagen, O. Eriksson, Life cycle assessment of building materials for a single-family house in Sweden, Energy Procedia 158 (2019) 3547–3552, https://doi.org/10.1016/j.egypro.2019.01.913.

[39] V. Biolek, T. Hanák, LCC estimation model: a construction material perspective, Buildings 9 (8) (2019) 182, https://doi.org/10.3390/buildings9080182.

[40] C. Zhang, M. Hu, X. Yang, A. Amati, A. Tukker, Life cycle greenhouse gas emission and cost analysis of prefabricated concrete building façade elements, J. Ind. Ecol. 24 (2020) 1016–1030, https://doi.org/10.1111/jiec.12991.

[41] J. O'Connor, Survey on actual service lives for North American buildings. In: Woodframe Housing Durability and Disaster Issues Conference, October 2004, Las Vegas, US, pp. 1–9.

[42] W.P.S. Dias, Factors influencing the service life of buildings, Engineer 46 (4) (2013) 1–7.

[43] G. Liu, K. Xu, X. Zhang, G. Zhang, Factors influencing the service lifespan of buildings: an improved hedonic model, Habitat Int. 43 (2014) 274–282, https://doi.org/10.1016/j.habitatint.2014.04.009.

[44] F. Provost, T. Fawcett, Data Science for Business: what You Need to Know about Data Mining and Data-Analytic Thinking, O'Reilly Media Inc., California, 2013.

[45] [Table A1] Table of Heat Perfusion Rates for Building Parts by Region, Energy Saving Design Standards of Buildings (ESDSB). Notice No. 2017-881. Ministry of Land, Infrastructure and Transport, Republic of Korea, [Revised] 28 December 2017.

[46] [Table A10] Earthquake Zones and Regional Coefficients, Regulation of Building Structure Standards (RBSS). Order No. 777. Ministry of Land, Infrastructure and Transport, Republic of Korea, [Revised] 9 Nov 2020.

[47] D. Bzdok, N. Altman, M. Krzywinski, Statistics versus machine learning, Nat. Methods 15 (4) (2018) 233–234.

[48] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, In: Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, August 2016, San Francisco, pp. 785–794.

[49] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, T.Y. Liu, Lightgbm: a highly efficient gradient boosting decision tree, in: Advances in Neural Information Processing Systems, 2017, pp. 3146–3154.

[50] A. Geron, Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, second ed., O'Reilly, California, 2019.

[51] A.J.P. Ibáñez, J.M.M. Bernal, M.J.C. de Diego, F.J.A. Sánchez, Expert system for predicting buildings service life under ISO 31000 standard. Application in architectural heritage, J. Cult. Herit. 18 (2016) 209–218.

[52] A. Miatto, H. Schandl, H. Tanikawa, How important are realistic building lifespan assumptions for material stock and demolition waste accounts? Resour. Conserv. Recycl. 122 (2017) 143–154.

[53] C.J. Chen, Y.K. Juan, Y.H. Hsu, Developing a systematic approach to evaluate and predict building service life, J. Civ. Eng. Manag. 23 (7) (2017) 890–901.

[54] W. Zhou, A. Moncaster, D.M. Reiner, P. Guthrie, Estimating lifetimes and stock turnover dynamics of urban residential buildings in China, Sustainability 11 (13) (2019) 3720.

[55] U. Sivarajah, M.M. Kamal, Z. Irani, V. Weerakkody, Critical analysis of Big Data challenges and analytical methods, J. Bus. Res. 70 (2017) 263–286.

[56] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ Computer Science 7 (2021) e623.

[57] T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature, Geosci. Model Dev. (GMD) 7 (3) (2014) 1247–1250.