Journal of King Saud University – Computer and Information Sciences 34 (2022) 5943–5967

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Reviewing the application of machine learning methods to model urban form indicators in planning decision support systems: Potential, issues and challenges

Stephane Cedric Koumetio Tekouabou [a,*], El Bachir Diop [a], Rida Azmi [a], Remi Jaligot [b], Jerome Chenal [a,b]

[a] Center of Urban Systems (CUS), Mohamed VI Polytechnic University (UM6P), Lot 660, Hay Moulay Rachid, Ben Guerir 43150, Morocco
[b] Urban and Regional Planning Community (CEAT), Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 29 April 2021
Revised 20 July 2021
Accepted 8 August 2021
Available online 20 August 2021

Keywords:
Urban form
Urban planning
Artificial intelligence
Machine learning
Deep learning
Urban models
Urban data
Neural networks
Ensemble methods
Urban big data

ABSTRACT

Modern cities dynamically face several challenges including digitalization, sustainability, resilience and economic development. Urban planners and designers must develop urban forms that address these challenges. With the integration of new communication and information technologies (Smartphone, GIS, Drones, IoT, Sensors, etc.), urban activities have generated large volumes of urban data. The rapid growth in terms of collection and big data storage capacities combined with the ever-increasing computational power of modern machines have made possible their efficient treatment using machine (ML) and deep learning (DL) algorithms. The emergence of such groundbreaking methods has in turn helped to address the challenges of modern-day cities in several domains (health, security, mobility, etc). ML algorithms have been proposed to model the urban form's indicators for intelligent urban planning decision making. They have been proven to perform better than the traditional methods. However, the potential of ML has not yet been fully explored in research for urban planning decision support. This paper presents a comprehensive review of ML applications for mitigating the challenges of modern cities planning. First and foremost, an overview of the urban forms, sources of urban data, the ML and DL techniques as well as their potential in solving the aforementioned challenges. For each ML method, we will highlight it working principle, advantages, disadvantages and potential applications using comparative tables. Finally, we will discuss the issues and challenges of ML methods in urban form's modeling while ultimately advocating some future research directions.

Contents

1. Introduction 5944
2. Taxonomy of ML for urban form 5946
   2.1. Urban form elements 5946
   2.2. Urban form indicators 5946
   2.3. Urban data sources 5946
   2.4. Ml models 5946
   2.5. Issues, challenges and future directions 5947
   2.6. Literature search strategy 5947
3. Overview of urban planning DSS 5953

* Corresponding author.
E-mail address: ctekouaboukoumetio@gmail.com (S.C. Koumetio Tekouabou).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2021.08.007
1319-1578/

3.1. urban planning decision making 5953
3.2. Concept of urban form 5953
3.2.1. Urban form features (elements) 5953
3.2.2. urban form indicators. 5954
3.3. Urban data sources 5954
4. Overview of ML methods for Urban form application 5955
4.1. Unsupervised ML models 5955
4.1.1. Clustering methods. 5956
4.1.2. Association rules methods 5956
4.2. Classical Supervised ML models 5956
4.2.1. DT 5956
4.2.2. LR. 5956
4.2.3. KNN 5956
4.2.4. NB 5957
4.2.5. SVM 5957
4.3. Ensemble methods (EM). 5957
4.3.1. Bagging algorithm 5958
4.3.2. Random Forest (RF) Algorithm 5958
4.3.3. Extra-Trees (ET) algorithm. 5959
4.3.4. Adaptif Boosting (Adaboost) algorthm 5959
4.3.5. Gradient Boosting (GB) algorithm 5959
4.4. NN based models 5959
4.4.1. ANN 5959
4.4.2. CNN 5959
4.5. DNN-based models 5960
4.5.1. DCNN. 5960
4.5.2. RNN 5960
4.5.3. Autoencoder (AE) 5961
4.5.4. Transfer Learning (TL) 5961
5. Discussion: potentials, challenges and future directions 5961
5.1. Potentials and issues of ML applications for urban planning DSS 5961
5.1.1. ML for addressing urban sustainability issues 5962
5.1.2. ML for addressing digital and smart city issues 5962
5.1.3. ML for addressing socio-economic issues 5962
5.1.4. ML for addressing urban land use optimization issues 5963
5.2. Challenges of ML for urban applications 5963
5.2.1. ML Implementation and Computational complexity 5963
5.2.2. The challenge of ML for urban data. 5963
5.3. Future research directions 5964
6. Conclusion 5964
Declaration of Competing Interest 5964
References 5964

1. Introduction

Digitization is increasing in almost all domains of daily life. The rapid growth of storage memories and computational capacities of machines have favored the emergence of new algorithms allowing to efficiently process large volumes of collected data and make them useful in various applications (Jordan and Mitchell, 2015; Al-Garadi et al., 2020). These machines are enriched with recent automatic learning methods to meet the needs of artificial intelligence for various applications (Choung and Kim, 2019). Machine learning (ML), which is at the intersection of computer science and statistics, and at the heart of intelligence and data science, is seeing its field of application expand day by day (Jordan and Mitchell, 2015). ML methods are thus finding their way into science, technology and commerce, leading to more evidence-based decision making in many fields, including healthcare, manufacturing, education, financial modeling, law enforcement and marketing (Al-Garadi et al., 2020). More recently, cities are being more and more included in these applications especially to meet the requirements in terms of sustainability, intelligence, economic and financial resilience, etc (Li et al., 2020; Choung and Kim, 2019; Liu et al., 2017). Thus, several approaches are proposed to model the dynamics of urban drivers as a function of the different elements (features) of the urban form. The latest research findings show that ML methods have remarkably transcended the traditional techniques of predictive or prescriptive modeling of urban form indicators to become an essential tool for urban planning decision support (Ma et al., 2020; Hecht et al., 2013). Indeed, in order to meet the challenges of the current complexity of emerging urban big data, the modeling of urban indicators exploits more and more the so-called intelligent automatic methods using ML algorithms which are favored over the less efficient traditional methods. Hence the actual emergence of ML models capable of automatically improving their performance with the experience gained in recent works (Al-Garadi et al., 2020).

Overall, previous studies have demonstrated that ML methods continue to have a huge potential for addressing the problem of modeling modern and intelligent urban forms (Choung and Kim, 2019; Ma et al., 2020). This problem mostly consists of a spatio-temporal analysis (Gómez et al., 2020; Faghmous and Kumar, 2014). Thus, some emerging methods such as convolutional neural networks (CNN) have proven to be effective in extracting features from spatial data, while others such as recurrent neural networks (RNN) deal with temporal data (Geng et al., 2019; Gómez et al., 2020). Ensemble-based methods such as Random Forests (RF), Bagging and boosting methods have proven to be very useful in various studies of smart urban forms (Jochem et al., 2018; Ma et al., 2020; Geiß et al., 2020; Novack et al., 2011; Hecht et al., 2013; Shafizadeh-Moghadam et al., 2017). Other categories of ML methods considered as simple supervised or unsupervised methods have also been widely used for various urban form applications (Abrantes et al., 2019; He et al., 2018; Li et al., 2020; Gao et al., 2020). This enormous potential of ML methods at the crossroads of urban form model complexity opens up a very promising area of research for the years to come (Li et al., 2020). This issue is at the center of modern urban planning challenges which aims to achieve much smarter, digital, sustainable, developed, resilient and inclusive urban forms.

In order to overcome these challenges in the digital era, urban research is now trying to integrate an ever-increasing complexity aiming to qualify the processes that intersect within urban environments and that make up their dynamics. It strives not to reason within an absolute framework, but to think in terms of scenarios, often contrasted, in order to bring together a wider range of possible outcomes. The implementation of such an approach, based on foresight and still relatively new in urban planning and engineering, requires the use of suitable tools to take into account, manage and analyze the complexity and intertwining of urban dynamics, before projecting them into the future. Among these tools, modeling offers today important perspectives for decision support in urban planning.

Thus, the objective of this paper is to provide a critical review of the literature on recent applications of ML methods to urban form modeling problems, the associated challenges, opportunities and future research directions. This review will include an overview of urban forms, urban data sources, and the description of the used ML methods, highlighting their operating principles, advantages, disadvantages and potential applications. Hence, we summarize the key contributions of this review paper as follows:

- Provides a comprehensive review of urban forms and their various data sources for ML applications.
- In-depth review of the ML and recent advances in ML algorithms applied in urban forms.

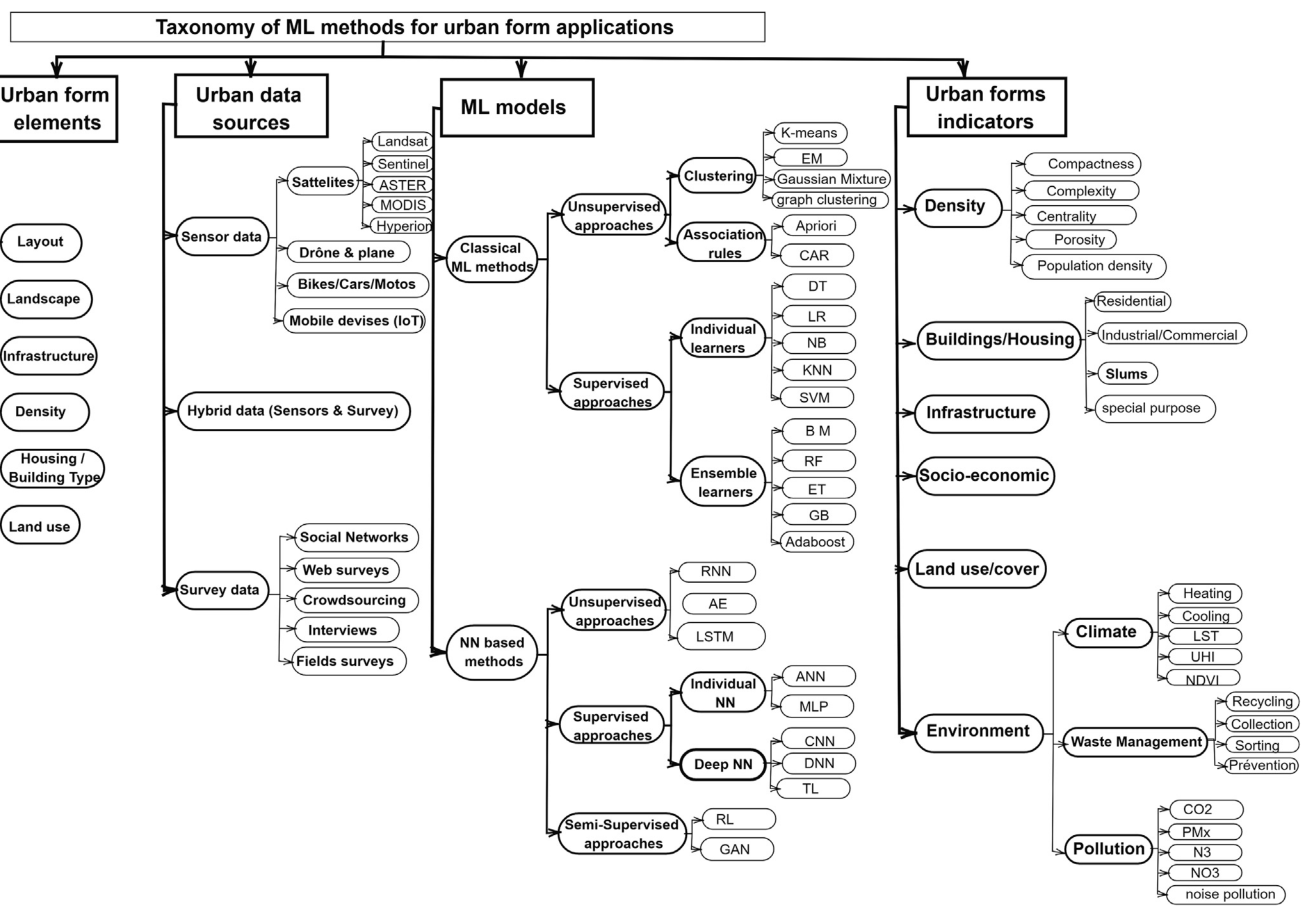

Fig. 1. Thematic taxonomy of ML for urban form applications.

Table 1
Comparison of different types of machine learning.

	supervised learning	unsupervised learning
Input data	uses known and labeled input data	Known input data but unlabeled
Computing complexity	Very complex	Less complex
Real-time	Uses offline analysis	uses a real-time analysis process
Sub-domains	Classification and regression	Clustering and association rules
Precision	Accurate results	Moderate results
Number of classes	Number of classes known	Number of classes is unknown

Table 2
Comparison of unsupervised ML methods for urban form modelling. EMA = Expectation maximisation Algorithms, CAR = Class Association Rules; DBSCAN = Density Based Spatial Clustering of Applications with Noise.

type	Algo	Working Principle	Advantages	Disadvantages	Application in urban modelling
Clustering	K-means (Krishna and Murty, 1999; Li et al., 2020), EM (Moon, 1996; Han et al., 2021), Gaussian Mixture Models (Han et al., 2021; Reynolds, 2009), graph clustering (Schaeffer, 2007), DBSCAN (Khan et al., 2014), Deep cluster (Caron et al., 2018)	The clustering methods partition a dataset X = $\{x^{(1)}, \ldots, x^{(m)}\}$ into k non-overlapping clusters $C^{(1)}, \ldots, C^{(k)}$. Each data point $x^{(i)}$ is assigned to precisely one cluster whose index we denote by $y^{(i)} \in \{1, \ldots, k\}$ (Williams, 1971).	• Easy to understand and implement • Conceptual simplicity and speed • Applicable to any type and size of data	• Number of classes must be fixed at the start • Result depends on the initial draw of the class centers • Clusters built in relation to non-existent objects	Useful for organising urban data into homogeneous clusters according to a common features like Energy (Han et al., 2021), type of urban occupation (Abrantes et al., 2019), urban fabrics (Li et al., 2020), LST (Gao et al., 2020), etc.
Association rules	Apriori (Wei et al., 2009; He et al., 2018), CAR (Nguyen et al., 2013; Lu et al., 2008)	The principle of an association rule $A => B$, is for a set of transactions, some value of itemset A determines the values of itemset B under the condition in which minimum support and confidence are met (Toivonen et al., 1996).	• Ease of understanding, efficiency and simplicity • Unsupervised and general formalism • Possibility of discovering useful knowledge hidden in databases	• The search time for frequent itemsets is enormous. • A number of important association rules, most of which are not interesting • The algorithms used have too many parameters	Used to determine the hidden relationships between features of UBDs such as walking adults (Jack and McCormack, 2014), urban vitality (He et al., 2018), travel behavior (Lu et al., 2008)

- In clear and easy-to-use tables, we summarize the categories of ML methods according to their algorithmic principles, advantages, disadvantages and potentials for intelligent urban planning applications.
- Discussion of the potentials, issues, challenges and future research directions of ML application to meet the next generation urban planning challenges.

The remainder of this paper is organized as follows. Section 1 broadly describes the taxonomy of the review contents we discuss in this paper while Section 3 provides an overview of urban data sources for ML applications. Section 4 provides a in-depth review of the ML and recent advances in ML algorithms applied in urban form modeling. Finally 5 discusses the potentials, opportunities, challenges and future direction of ML application in urban form and finally Section 6 concludes our review.

2. Taxonomy of ML for urban form

Applications of ML algorithms for urban form modeling have evolved enough in recent years to address urban planning challenges. Fig. 1 represents the taxonomy of ML applications involving several domains. We have grouped these areas into four main categories: the urban form elements associated with the chosen study area (city), the reliable data sources and used data type for target task, the potential ML models, the targeted indicators. Additionally, we have added the issues, scope, challenges and future directions. We briefly describe each theme in more detail in the different sections below.

2.1. Urban form elements

In the design and modeling of ML applications for urban form, we define the elements of urban form as the input geo-physical features that will drive the chosen model (Dempsey et al., 2010). The choice of these elements is very important because it allows the best calibration and therefore the best efficiency of the model. In this field of expertise, we include all the parameters that are directly related to the geometric and physical shape of the city. The field of urban indicators is for the most part associated with the shape of the city in the geographical and logical sense. the Section 3 gives more details on these different notions.

2.2. Urban form indicators

The goal of ML applications to city shape is to provide models to simulate indicators from input elements describing the city shape. To be efficient, it is necessary to master the associated indicators. According to the taxonomy of Fig. 1, the study of these indicators is therefore an area of expertise in the process of designing and modeling ML applications for urban form. The Section 3.2.2 gives more details on these different notions.

2.3. Urban data sources

Urban data collection and structuring is an area of high demand, especially for ML-based applications. Without the data, ML models can neither be trained nor calibrated (Al-Garadi et al., 2020; Jordan and Mitchell, 2015). Urban data are characterized by their source, scale, study area, nature (derived or raw, …In the Section 3.3, we make a more detailed analysis on the collection and structuring of urban data for ML applications based on the works of the literature.

2.4. Ml models

Learning algorithms have been widely adopted in many real-world applications due to their unique problem-solving nature (Al-Garadi et al., 2020). Note that the value of these algorithms lies in their ability to build machines that automatically calibrate and progress through experience (Jordan and Mitchell, 2015). With the explosion of processor capacities and the volumes of data collected, learning algorithms have been widely applied in practice in several domains among which urban planning. The development of new algorithms and the emergence of low computational cost algorithms capable of adapting and calibrating new datasets makes the study of ML models and their potentials a central area to the design of intelligent urban form applications. In the Section 4, we study in detail, the ML models and their applications to the urban form.

2.5. Issues, challenges and future directions

In this category, we explore a list of issues, scope, challenges, and future directions for applying ML methods to address current and future challenges in urban form. The Tables 3 and 6 summarize works applying ML to urban form according to data (source, study area, scale, ...), learning types (suppressed (classification, regression), non-suppressed (clustering, association)), year of publication and ML methods used. We discuss a sustainable, resilient and inclusive urban form by including environmental, socio-economic elements and indicators in addition to classical geophysical elements. We also focus on the challenges of ML applications or the integration of ML with other technologies, of ML for an urban form offering interdependent, resilient, intelligent, interconnected and socio-economic sustainable urban environments. Finally, we address the complexity of the ML field in relation to the trade-offs in urban form applications.

2.6. Literature search strategy

In our study, we first conducted a literature search on the SCOPUS database in order to perform the literature review process. The search query performed groups together two categories of key concepts: concept 1 associated with urban includes terms such as "urban planning", "urban form", "urban morphology" and "urban planning" while concept 2 related to machine learning includes the terms "machine learning" and "deep learning". In the search query, the boolean *"AND"* is used to combine the two main concepts, while OR is used to include research articles and finally *"EXCLUDE"* is used to exclude

Table 3
Summary of ML methods for urban form modelling. Data sources have been targeted according to Section 3.3

Ref	Year	ML model												Data source	Study area	Target problem
-		Supervised									Unsup.		PT			
-		Single methods					Ensemb methods									
-		DT	KNN	SVM	NB	LR	B	RF	AB	GB	Clust	Asso				
(Ma et al., 2020)	2020		✓	✓		✓		✓	✓				R	Hybrid	New York	Land values
(Hecht et al., 2013)	2013	✓	✓	✓			✓	✓					C	Sensors	Germany	Urb Struc Types
(Choung and Kim, 2019)	2019			✓									C	Sensors	South Korea	PM_{10}
(Hecht et al., 2015)	2015												C	Hybrid	germany	building footprint
(Wang et al., 2020)	2020							✓					R	Sensors	Texas	Water quality
(Arribas-Bel et al., 2019)	2019										✓		C	Hybrid	Span	urban areas
(Abrantes et al., 2019)	2019										✓		Cl	Hybrid	Portugal	typ urb occup
(Kleine Deters et al., 2017)	2017			✓						✓			C	Sensors	Quito(Ecuador)	Pollution($PM_{2.5}$)
(Chan et al., 2001)	2001	✓		✓									C	Hybrid	Hong Kong	Environment
(Liu et al., 2017)	2017			✓									C	Sensors	Beijing	Environment qlty
(Chen et al., 2013)	2013			✓									R	Hybrid	Guangdong	Energy Consump
(Jochem et al., 2018)	2018		✓					✓					C	Sensors	Afghanistan	Land use areas
(Jack and McCormack, 2014)	2014											✓	As	Surveys	Canada	walking adults
(Duerr et al., 2018)	2018					✓		✓		✓			R	Survey	Florida	water demand
(Kontokosta et al., 2018)	2018									✓			R	Survey	New York	solid waste manag
(Lee, 2019)	2019					✓							R	Hybrid	US	air quality
(Li et al., 2020)	2020										✓		Cl	Sensors	Wuhan	urban fabric.
(Liu et al., 2019)	2019							✓		✓			R	Survey	Nanjing	traffic flow
(Milojevic-Dupont et al., 2020)	2020					✓		✓		✓			R	Hybrid	Europe	building heights
(Reades et al., 2019)	2019							✓					R	Hybrid	London	gentrification
(Novack et al., 2011)	2011	✓		✓			✓	✓					C	Sensors	Sao Paulo	Urb Land Cover
(Sun et al., 2019)	2019					✓		✓					R	Hybrid	China	LST
(Gómez et al., 2020)	2020					✓							C	Sensors	Colombia	urban growth
(Shafizadeh-Moghadam et al., 2017)	2017	✓		✓		✓		✓					C	Sensors	Tehran	Urban growth
(Tran et al., 2017)	2017					✓							R	Sensors	Hanoï	LST
(Hart and Sailor, 2009)	2009		✓										R	Sensors	Portland	UHI
(Yang et al., 2020)	2020										✓		Cl	Hybrid	Wuhan	LST
(Gao et al., 2020)	2020					✓		✓					R	Hybrid	Wuhan	LST
(Okwuashi and Ndehedehe, 2020)	2020			✓									R	Sensors	Lagos (Nigeria)	Land-use
(Chen et al., 2020)	2020				✓			✓					R	Sensors	Singapoure	LST/UHI
(Geiß et al., 2020)	2020								✓				R	Sensors	Germany	Build Height
(Kabano et al., 2021)	2021							✓					R	Sensors	Uganda	UHI
(Yu et al., 2020)	2020							✓					R	Sensors	Singapore	UHI/LST
(Han et al., 2021)	2021										✓		Cl	Sensors	Boston (USA)	Energy
(He et al., 2018)	2018											✓	As	Sensors	China	urban vitality
(Lu et al., 2008)	2008											✓	As	Hybrid	Chicago (USA)	Travel behavior
(Xing and Meng, 2020)	2020			✓	✓			✓					C	Sensors	Shenzhen (China)	urb function
(Kafy et al., 2021)	2021			✓									C	Sensors	Dhaka (Bangladesh)	Land cover

Table 4
Comparison of ML methods for urban form modelling.

Type	Algo	Working Principle	Advantages	Disadvantages	Application in urban modelling
Single ML Methods	DT	DT is a technique for structuring a set of learning data in the form of trees made up of nodes and leaves. Each node represents the test on the given attribute, while the leaf represents the class (Ruggieri, 2002; Bashir et al., 2014).	DT is a simple, easy-to-use and transparent method. DT is the most favourable weak learner algorithm for ensemble combination.	DT requires large storage because of its construction nature. Understanding DTbased methods is easy only if few DTs are involved.	Environment (Chan et al., 2001), urban structures types (Hecht et al., 2013), urban land cover (Novack et al., 2011), urban growth (Shafizadeh-Moghadam et al., 2017), etc
	LR	Given a variable to predict $Y = \{0, 1\}$ and a predictive variable (explanatory variable) $X = (x_1, x_2, \ldots, x_n)$, LR is based on the fundamental assumption of the Eq. (1).	Easily adaptable to both classification and regression problems, easy to regularise, fast to train and resistant to over-fitting	inefficient on separable non-linear data and in multi-class classification	Land values (Ma et al., 2020), Air quality (Lee, 2019), LST (Gao et al., 2020), Urban growth (Gómez et al., 2020).
	KNN	The principle of the KNN algorithm consists in assigning the class or regression value by averaging the k nearest neighboring values, for numerical instances, or by applying the majority vote for k neighbors, if the values of the instances are categorical.	KNN is a popular and effective ML method for urban modelling. It has a simple process which extends its use to several levels of modeling.	The optimal k value usually varies from one dataset to another; therefore, determining the optimal value of k may be a challenging and timeconsuming process.	Land Values (Ma et al., 2020), urban structure type (Hecht et al., 2013), Land use areas (Jochem et al., 2018) and UHI (Hart and Sailor, 2009)
	NB	NB's principle consists in calculating the posterior probability using Bayes' theorem to predict the probability that a particular set of features of unlabelled samples corresponds to a specific label with the hypothesis of independence between the features.	NB is simple, ease of implementation, low training sample requirement and robustness to irrelevant features (Jordan and Mitchell, 2015; Al-Garadi et al., 2020).	NB handles features independently and thus cannot capture useful clues from the relationships and interactions among features (Al-Garadi et al., 2020).	urban function from landscape metrics (Xing and Meng, 2020), LST/UHI (Chen et al., 2020), water demand (Duerr et al., 2018)
	SVM	A SVM is based on nonlinear transformations of the features into a higher-dimensional feature space, where the classification problem becomes linear separable. SVM models are basically binary classifiers. With aggregation techniques, these can be made applicable to multi-class problems.	-Suitable for linear and non-linear separable data, -Widely used model, -High precision and efficiency in large spaces, -Less prone to overlearning and stable, -Noise robustness and the problem of unbalanced classes	-Not suitable for large datasets as the very long training time, -Less effective on noisier data with overlapping classes; -Difficult to choose the right kernel well adapted to the data; -Unable to provide categorical data.	urban type structures (Hecht et al., 2013), land values prediction (Ma et al., 2020), type of urban occupation (Abrantes et al., 2019), Pollution ($PM_{2.5}$ (Kleine Deters et al., 2017), PM_{10} (Choung and Kim, 2019)), uban functions from landscape metrics (Xing and Meng, 2020), land uses (Okwuashi and Ndehedehe, 2020), Land cover (Kafy et al., 2021), etc

Table 5
Comparison of Ensemble ML methods for urban form modelling.

Type	Algo	Working Principle	Advantages	Disadvantages	Application in urban modelling
Ensemble Methods	B	Bagging algorithm (Breiman, 1996) consists of succession of bootstrap samples, basic weak learner model application and aggregation. For classification the results of the weak learners models results are aggregated by voting (see Fig. 9(a)).	-Bagging reduces the variance when predictors are unstable (Breiman, 1996), -Estimate of the prediction error by Bootstrap out of bag: prevents over-fitting	Bagging stable predictors doesn't bring anything and therefore you need classifiers that are sufficiently different from each other to improve performance.	urban structure types (Hecht et al., 2013), urban land cover (Novack et al., 2011).
	RF	RF algorithm(Breiman, 2001) is an EM based on bagging principle where each weak learner in the random forest is trained on a random subset of data according to the principle of "bagging", with a random subset of characteristics (variable data characteristics) according to the principle of "random projections". The predictions are then averaged when the data are quantitative or used for a vote of qualitative data in the case of trees of classification (Breiman, 2001; Bashir et al., 2014).	RF is efficient and robust to over-fit. RF bypasses feature selection and requires few tuning parameters. RF is almost the most widely used algorithm, which guarantees its reliability.	RF is based on the construction of several randomly selected TDs; thus, it may be impractical in specific real-time applications where the required training data set is large (Al-Garadi et al., 2020).	land values (Ma et al., 2020), urban structure types (Hecht et al., 2013), water quality (Wang et al., 2020), land use areas (Jochem et al., 2018), water demand (Duerr et al., 2018), gentrification (Reades et al., 2019), urban land cover (Novack et al., 2011), LST/UHI (Sun et al., 2019; Chen et al., 2020), urban growth (Gómez et al., 2020; Shafizadeh-Moghadam et al., 2017), urban functions (Xing and Meng, 2020) etc.
	ET	ET (Geurts et al., 2006) are a set of decision trees built from bagging as explained above. However, they differ from other EM by 1) the separation that takes place in the internal nodes is random, the attributes and thresholds tested are chosen at random; 2) Extra-Trees use the whole learning set to build the trees, not just a part as in the bagging method.	ET algorithm saves time because it randomly chooses the split point and does not calculate the optimal one compared to RF while they can perform similarly. (Geurts et al., 2006).	-Reducing model interpretability -The design time is high, -difficult to learn and inefficient with such a large volume of data.	suitable for classification and regression problems such as land values (Ma et al., 2020) and build height (Geiß et al., 2020)
	GB	GB algorithm works by sequentially adding predictors to a set, so that each tries to correct the errors of its predecessor (Freund et al., 1999; Freund et al., 1996). However, instead of adjusting the weights of the instances at each iteration, as AdaBoost does, this method tries to adjust the new predictor to the residual errors committed by the previous one (Géron, 2019).	GB is a generic algorithm to find approximate solutions to the additive modeling problem. GB builds trees on previous classifier's residuals thus well capturing variance in data.	GB is sensitive to outliers since every classifier is obliged to fix the errors in the predecessors. GB is almost impossible to scale up because every estimator bases its correctness on the previous predictors.	pollution $PM_{2.8}$ particle in urban area (Kleine Deters et al., 2017), water demand prediction (Duerr et al., 2018), solid waste management (Kontokosta et al., 2018), urban traffic flow (Liu et al., 2019) and building heights (Milojevic-Dupont et al., 2020)
	AB	AB is based on the fact that a new predictor to correct the error of its predecessor simply gives a little more attention to training cases on which the predecessor has adapted less (Mishra et al., 2017; Chengsheng et al., 2017).	AB can be implemented without the reference to gradients by reweighting the training samples based on classifications from previous learners (Chengsheng et al., 2017)	AB is from empirical evidence and particularly sensitive to noise data.. Weak classifiers being too weak can lead to low margins and overfitting (Chengsheng et al., 2017)	Favourable for regression and classification problems, both binary and multi-class. AB has been applied for land values (Ma et al., 2020) and build height (Geiß et al., 2020) predictions.

Table 6
Summary of DL methods for urban form modelling. Data sources have been targeted according to Section 3.3. R = regression and C = Classification

Ref	Year	DL model										PT	Data source	Study erea	Target problem
		Supervised model					Unsup model			RL models					
		ANN	MLP	CNN	DCNN	TL	RNN	AE	LSTM	RL	GAN				
(Ma et al., 2020)	2020		✓									R	Hybrid	New York	land values
(Liu et al., 2017)	2017				✓							C	Sensors	Beijing	Environment Qlty
(Chang et al., 2019)	2019									✓		R	Sensors	Shenshen	Energy
(Nice et al., 2020)	2020	✓										C	Hybrid	1692 cities	Urban typology
(Ibrahim et al., 2019)	2019	✓										C	Sensors	Egypt & Mumbai	Slums
(Moosavi, 2017)	2017							✓				C	Sensors		Urban Structure
(Chan et al., 2001)	2001		✓									C	Hybrid	Hong Kong	change in envmt
(Guo et al., 2020)	2020			✓								C	Sensors	Swi&Portu	Flood prediction
(Kontokosta et al., 2018)	2018	✓										R	Survey	New York	solid waste manag
(Middel et al., 2018)	2018			✓								C	Sensors		Climate
(Middel et al., 2019)	2019			✓								R	Sensors	Philadelphia	Street (Mobility)
(Shen et al., 2018)	2018				✓							R	Hybrid	Wuhan	$PM_{2.5}$
(Novack et al., 2011)	2011										✓	C	Sensors	Sao Paulo	Urban Land Cover
(Shafizadeh-Moghadam et al., 2017)	2017	✓										R	Sensors	Tehran	Urban growth
(Verma et al., 2019)	2019					✓						C	Sensors	Mumbai	Slums detection
(Janarthanan et al., 2021)	2021								✓			R	Sensors	Chennai	Quality of air
(Fu et al., 2016)	2016								✓			R	Sensors	California	Traffic flow
(Koschwitz et al., 2018)	2018						✓					R	Hybrid	Germany	Heating & cooling
(Lu et al., 2021)	2021						✓		✓			R	Hybrid	UK	Traffic flow
(Xayasouk et al., 2020)	2020							✓	✓			R	Sensor	Seoul	Quality of air
(Geng et al., 2019)	2019			✓			✓					R	Hybrid	China	ride-hailing

review type documents. Thus the query performed in SCOPUS was the following: *TITLE-ABS-KEY ((“urban planning” OR “urban form” OR “urban shape” OR “urban morphology”) AND (“machine learning” OR “deep learning”)) AND (EXCLUDE(DOCTYPE, “re”) OR EXCLUDE(DOCTYPE, “cr”) OR EXCLUDE(DOCTYPE, “le”)) AND (LIMIT-TO(LANGUAGE, “English”)).* The query combining these keywords in the SCOPUS database resulted in the retrieval of 551 raw articles, 533 after excluding of reviews, conference reviews and letters, then 521 after limiting to English language. The articles thus retrieved in a *csv* document by this search query went through a filtering process allowing to keep only the most relevant ones for the subject: *“applications of machine learning methods to decision support of sustainable urban planning of urban form”*. In order to obtain targeted documents from the targeted literature, we adapted the search items after several filterings based on data analysis techniques using the ORANGE tool (Demšar et al., 2004; Demšar et al., 2013). The list of search terms finally included in the filtering queries on ORANGE is illustrated by the Fig. 2. However, the articles recognized as relevant and which by chance would not have been retained by the previous phases were added manually in order to end up with 206 the most relevant articles for the mapping

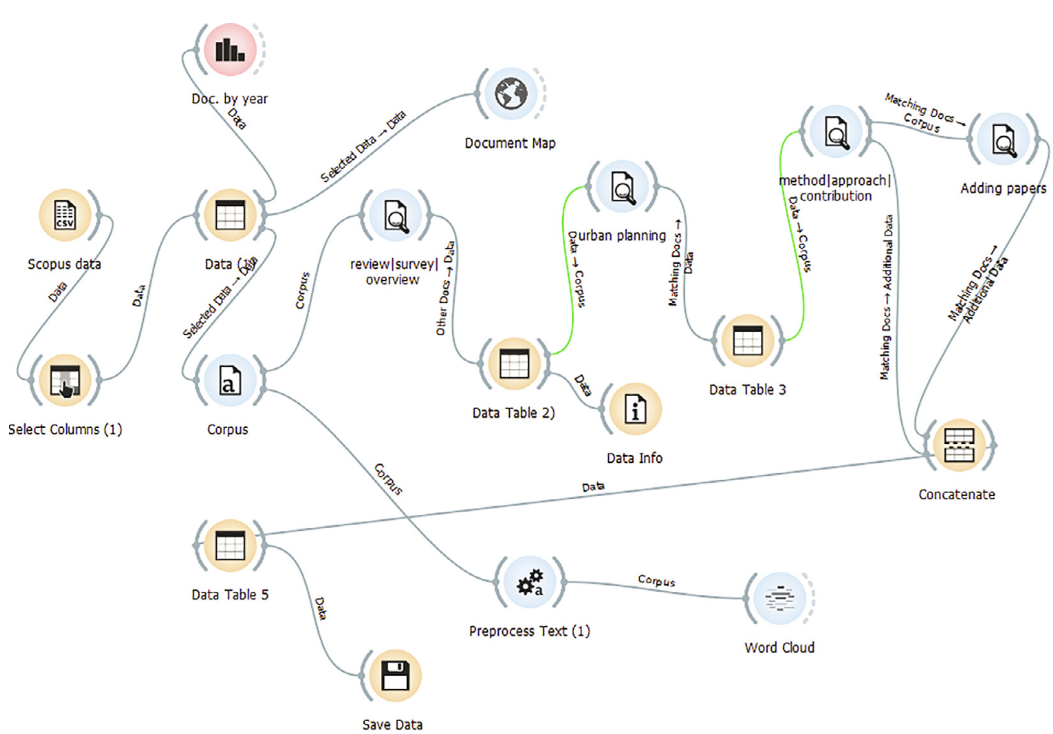

Fig. 2. Filtering process for relevant articles on ORANGE.

Table 7
Comparison of DL methods for urban form modelling.

Type	Algo	Working Principle	Advantages	Disadvantages	Application in urban modelling
Simple NN models	MLP	The operating principle of the artificial neural network (ANN) algorithms represented in Fig. 10 is very schematically inspired by the functioning of the biological neurons to which the method owes its name (Thayse et al., 1988; Wang, 2003).	-High performance and computing power, -Efficiency for high dimensional problems, -Ability to work with complex characteristics, -Parallel processing capability and fault tolerance	-theoretically complex, -requires careful adjustment, -requires a large amount of data to be effective	land values (Ma et al., 2020), Change in urban environment (Chan et al., 2001)
	ANN				Solid waste management (Ma et al., 2020), urban typology (Nice et al., 2020),
CNN models	CNN	DCNNs are based on the receptive field of the brain, which processes input from sensors and is sensitive to certain stimuli, such as the edges of the visual system. They efficiently process large amounts of input data and are therefore widely used in the fields of computer vision (Krizhevsky et al., 2012).	-Efficiency for high dimensional problems, -Requires less data and computing power, -Ability to work with complex features, -Parallel processing capability and tolerance to variations	-theoretically complex, -requires careful adjustment, -requires a large amount of data to be effective	Flood prediction (Guo et al., 2020), Climate (Middel et al., 2018), mobility in street (Middel et al., 2019)
	DCNN				Quality of the urban environment (Liu et al., 2017)
RNN-based models	RNN	RNNs permits continuing information related to past knowledge by utilizing a special kind of looped architecture. They are employed in many areas regarding data with sequences, such as predicting the next word of a sentence (Subasi, 2020).	RNNs and their variants have been found very efficient in many applications with sequential data. Urban features in certain cases can be sequential data (Al-Garadi et al., 2020).	When RNNs are trained over longer sequences than a few elements, they suffer from gradient problems that disappear and explode (Al-Garadi et al., 2020; Pascanu et al., 2013).	Complex time-series prediction such as predictions of traffic flow (Lu et al., 2021), heating & cooling (Koschwitz et al., 2018), ride-hailing (Geng et al., 2019), etc.
	LSTM	A LSTM consists of a cell, an input gate, an output gate and a forget gate. The cell stores values at arbitrary time intervals and the three gates regulate the flow of information in and out of the cell (Hochreiter and Schmidhuber, 1997).	LSTM are well-suited for classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series	LSTM Takes a long time to run, often require more data to train than other models and therefore more suitable for large volumes of data, have lots of input parameters to tune.	Complex time-series prediction such as quality of air (Janarthanan et al., 2021; Xayasouk et al., 2020)
	GRU	GRU is a RNN-based Encoder-Decoder consisting of two RNNs networks. The first one encodes a sequence of symbols into a fixed length vector representation, and the other one decodes the representation into another sequence of symbols. The encoder and decoder are then jointly trained to maximize the conditional probability of a target sequence (Cho et al., 2014; Chung et al., 2014)	GRU networks are very popular and much simpler and require less computational power, so can be used to form really deep networks	They are simpler in their principle of operation but performance can sometimes be limited in the face of certain problems.	Complex time-series prediction such as quality of air (Janarthanan et al., 2021; Xayasouk et al., 2020)

Table 8
Comparison of DL methods for urban form modelling (Suite).

Type	Algo	Working Principle	Advantages	Disadvantages	Application in urban modelling
	AE	AEs allow the input to be coded into a different representation, and reconstruct it from that learned representation. They are therefore trained to reconstruct their original input. Once the model is well trained, the encoder part can be used to extract characteristics from the data (Makhzani et al., 2015).	Effective for feature extraction, learning representation learning to learn features instead of manual features in traditional ML and reduce dimensionality without any prior data knowledge.	AEs are data-specific and therefore very restricted to test data that are similar to train data. The scalability of AEs are reduced.	Aes are used for image denoising and compression; dimensionality reduction and information retrieval as in the case of urban structure (Moosavi, 2017) and Quality of air (Xayasouk et al., 2020).
	GAN	Two networks are placed in competition in a game theory scenario (Goodfellow et al., 2014). The first network is the generator, it generates a sample (e.g. an image), while its opponent, the discriminator, tries to detect whether a sample is real or whether it is the result of the generator (Zhang et al., 2019; Goodfellow et al., 2014). The learning can be modeled as a zero-sum game.	GANs generate data that looks similar to original data to solve problem of data in ML by going into details of data and can easily interpret into different versions.	GANs can be hard to train knowing that you need to provide different types of data continuously to check if it works accurately or not and generating results from text or speech is very complex	Given semantic image or data sample, GANs can be used to generate new plausible examples data like Cityscape, layout, urban design or urban growth (Shafizadeh-Moghadam et al., 2017).
RL models	RL	The RL allow a learning agent to adjust his or her policy based on errors and rewards and to derive an optimal solution through trial and error. Three methods for RL include 1) Value-based 2) Policy-based and Model based learning (Sutton and Barto, 2011).	There is no supervisor, only a real number or reward signal. works on interacting with the environment, whereas the supervised learning method works on given sample data.	RL is computing-heavy and time-consuming in particular when the action space is large (enough data).	RL can be used to optimize large-scale production systems without such how different combinations will affect future energy consumption (Chang et al., 2019)
	TL	TL simply consist to recover the acquired knowledge of a pre-trained model (CNN for example) on a huge dataset (such as Imagenet which contains 1.2 million images (Krizhevsky et al., 2012; Deng et al., 2009)), adapt them to the structure of the model to your own data then re-train it partially instead of training this network from scratch (Tran et al., 2017).	Low resource requirements, easy to implement, short execution time, better performance for some complex problems (Tan et al., 2018)	Transfer learning only works if the initial and target problems are similar enough for the first round of training to be relevant. We can't remove layers with confidence to reduce the number of parameters; very compact model.	TL is commonly used with CNN for computer vision tasks such as in the case of Slums detection (Verma et al., 2019)

processes. After the filtering processes, the remaining articles went through the mapping processes allowing the construction of the Tables 3 and 6 in the following sections.

3. Overview of urban planning DSS

The planning of the urban space or the urban form of a city hosting its population is a central issue that can considerably influence its socio-economic and environmental impact and its sustainability. Hence the need for DSS to master and better control this task. In this section we discuss in turn the stakeholders in the DSS planning process as described in the Fig. 3.

3.1. urban planning decision making

In urban planning, decision support systems (DSS) allow the integration of models in the decision process of urban form planning. These models often consist in simplifying the reality of the world, in order to better understand how decisions and events interact with each other (Maignant, 2005). It then also allows urban planning actors to adjust, reproduce or modify them in vitro, in order to test parametric solutions to influence or direct their consequences, to decide in advance on policies and strategies that can lead to a desirable future. This forms a decision cycle that we have represented in the Fig. 3. Machine learning (ML) algorithms currently offer the best models for challenging urban modeling problems. All the ML approaches and models proposed in the retrieved works take up and adapt this initial idea, applying it specifically to a particular question, a particular case or a specific city. All of them aim to better consider the objectives of sustainable development, and to propose viable solutions, in the short, medium and long term, for controling, monitoring urban forms and moreover predicting future trends in urban form indicators (Kafy et al., 2021; Middel et al., 2018).

In this case, the urban planning DSS process represented in the Fig. 3 particularly integrates the ML models for predictive modeling of urban form indicators according to its elements in order to help in a better decision making. The cyclic process can be restarted and continued until the best desired urban form is obtained. Several ML models have been or can be used in the design of urban planning DSS. But before studying them, it would be important to clarify the concept of urban form, urban form elements and indicators and urban data.

3.2. Concept of urban form

The word *"form"* has a large number of meanings and most of them refer to a notion close to appearance (state, aspect, ...). The form has a subjective connotation linked to the notion of perception which is a function of the experience of each individual (Schwarz, 2010). It is also analytical, there are concrete tools, more or less effective, of characterization of the forms. Dempsey et al. (2010) for example define urban form as "describe a city's physical characteristics" while according to Schwarz (2010), "urban form encompasses both the physical structure and size of the urban fabric as well as the distribution of the population in the area". A broader definition of urban form as in this study could have several variants as well physical (spatial) (Dempsey et al., 2010) as socio-economic (Huang et al., 2007; Schwarz, 2010; Frenkel and Ashkenazi, 2008) and environmental. Thus, the elements and indicators of urban form may differ from one work to another depending on the context and objectives of the study. Mathematicians have been very interested in this notion, with an essentially geometric vision. Physicists, on the other hand, have been more interested in spatial configurations, notably with the arrangement of components in urban space (Huang et al., 2007). The geography or more generally the social sciences are, as for them, leaned towards the analysis of the forms as perceived space, lived, in the analysis of hidden forms and in the representation of the spatial forms. However, there would be a correlation between all these different aspects which would make that spatial approaches (geometrical, physical) or socio-economic or environmental could lead to quasi-similar results. After this conceptual analysis, we have grouped on the Fig. 4 the elements of the broadest possible urban form into six categories including layout, landscape, density, land use, building and infra-structures (Dempsey et al., 2010). Similarly, the modeled urban form indicators of these elements are presented in Fig. 5 and will be discussed in the following section.

3.2.1. Urban form features (elements)

There are large differences between cities in terms of urban form and their ecological footprint. According to (Milder, 2012), there are five key elements of urban form: density, area, land use, road/public transportation infrastructure, and economic relationship with the surrounding environment (Dempsey et al., 2010). However, these elements may vary from context to context and the question of economic relationship as a very physical/geometric element of urban form is quite relative (Schwarz, 2010). Thus, we show on the Fig. 4 six main elements allowing to describe the shape of the city: layaout, Infrastructures, Landscape, Housing/Building type, Land use and Density (Dempsey et al., 2010). The concept of density helps to distinguish or direct the growth of the city towards compactness or sprawl (Li et al., 2020). The compact city is often advocated as a more sustainable urban form, but this is not uncontested (Milder, 2012). In fact, can the urban form automatically influence its environmental impact, or can it only

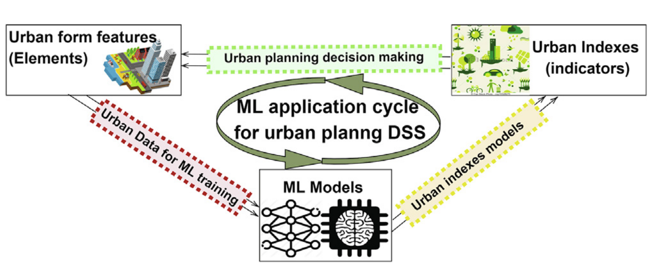

Fig. 3. Cycle of ML application for urban form decision support system (DSS).

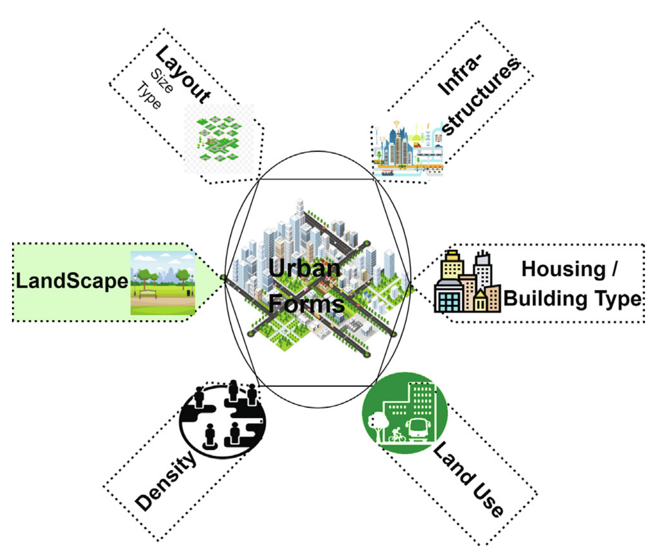

Fig. 4. Features (elements) of urban form. Adapted from (Dempsey et al., 2010).

Fig. 5. Example of key categories of urban forms indicators to handle in urban planning process. These categories of indicators are the most commonly used in urban engineering but may vary from one work to another depending on the objectives, issues and/or field of study. Adapted from (Dempsey et al., 2010; Schwarz, 2010).

facilitate it? In other words, it would be the people and their behavior that would ultimately determine the negative or positive effects on the environment and thus the sustainability of the city. To address the sustainability challenge of urban form, modern urban planning increasingly exploits the large volumes of data collected in cities and the emerging machine learning algorithms to simulate predictive indicators of chosen urban forms.

3.2.2. urban form indicators

Urban form indicators are evaluation and decision-making tools, developed from each of its elements in order to consider the evolution of the urban form in relation to a given moment. The urban forms that surround us can be apprehended by different indicators. By definition, an indicator is a piece of data, selected from a larger statistical set, which has a particular meaning and representativeness (Maignant, 2005; Schwarz, 2010). It is also an instrument capable of conveying information in synthetic form, through different representations (numbers, graphs, thematic maps, etc) (Maignant, 2005).

Here we describe the indicators associated with urban form that have been or can be modeled by ML urban engineering algorithms. As shown in the figure, we have categorized them into six major, non-exhaustive categories: climatic, pollution (environmental), density (land use), socio-economic, energy/water and mobility. The indicators modeled by the different ML algorithms according to the literature are each classified into one of the categories as presented in the Tables 3, 5, 6 and 8. Note that this classification is motivated by the work of (Schwarz, 2010; Dempsey et al., 2010) and the majority of the papers here reviewed. The following sections detail the different indicator types modeled by ML approaches in the literature.

3.3. Urban data sources

ML feeds on the best possible data to better fit the targeted task (Al-Garadi et al., 2020; Jordan and Mitchell, 2015). To infer patterns associated with urban forms by ML techniques, the authors relied on various data sources. These sources range from the most classical ones such as satellite and institutional data to the most recent ones including ubiquitous data, i.e. data from devices massively used by people and present almost everywhere in a city (Niu and Silva, 2020). With all this data, it is possible to provide more intelligent solutions to current challenges of cities, such as "sustainable planning, smart city, digitalization, resilience, "developing better strategies and studying the impact of new urban development projects (Li et al., 2020). Globally the sources of urban data collection are grouped into two main categories including sensor-based technologies and surveys as shown in the figure. These data sources allow us to have an idea of their reliability according to the desired study and the real impact of the obtained results.

Considering the sources of data based on sensors, we can divide them in relation to the technologies of acquisition and transport used that reflect the spatial extent of the territory covered and the scale of accuracy of the data. In this sense, we distinguish:

- **Satellite data** which cover a relatively larger space and with a relatively lower accuracy. They remain among the most used in ML applications due to their high availability. Four constellations are most often used in urban studies, including: Lansat (Choung and Kim, 2019; Aniello et al., 1995; Gómez et al., 2020; Okwuashi and Ndehedehe, 2020), Sentinel (Geiß et al., 2020), MODIS (Kabano et al., 2021), ASTER (EOS, 1999), etc.
- **Large-scale data** which cover a relatively smaller area than satellite data with greater accuracy. They are used more with the emergence of drones and the search for better accuracy (Vergouw et al., 2016).
- **Data collected by terrestrial transporters** like (bikes, motorcycles, cars, ...), is a new era that offers many possibilities in the Open Data sector. Connected vehicles represent one of the most promising solutions, of which several associations as well as the municipality's project have been proposed (Shelton et al., 2019). Data on roads with bicycle use, one of the initiatives proposed by the Open Street Map Foundation - OSM (Luxen and Vetter, 2011). As for the data coming from any crowd-source, their quality is of concern because the volunteers who provide the data do not have sufficient cartographic training and the quality cannot be guaranteed (Basiri et al., 2016).

- **Ubiquitous mobile devices data** that come from sensors embedded in everyday mobile devices (Lee et al., 2008).

As for survey-based data (field, web, or any source), they are paramount as data sources and are generally characterized by superior quality. They are often used alone (Jack and McCormack, 2014; Kontokosta et al., 2018) or in combination with sensor data for more comprehensive and reliable studies (Ma et al., 2020; Abrantes et al., 2019; Lee, 2019). Finally, several studies classify urban data sources according to accessibility, i.e. the storage platform, which distinguishes institutional data from data published in open source.

Considered one of the best geospatial processing platforms, Google Earth Engine (GEE), is a cloud computing platform designed to store and process huge (petabyte scale) datasets for analysis and final decision making. After the Landsat series was made freely available in 2008, Google archived all of the datasets and linked them to the cloud computing engine for open source use. Current data archives include those from other satellites (notably Sentinel I, MODIS), as well as geographic information system (GIS)-based vector datasets, social, demographic, weather, digital elevation model, and climate data layers. (Liu et al., 2017; Sun et al., 2019).

Technically, any urban phenomenon (including urban planning) is described by two main categories of data: Spatial and Attribute.

1. **Attribute data:** gathering all alphanumeric data such as censuses, statistics, surveys, etc).
2. **Spatial data** are any information having a spatial base. They gather two types: (1) vector data (point, line and polygon) such as data on the roads, administrative division, sewerage. . .ect. (2) and raster data (image or raster) such as Satellites images, Radar images, aerial photos etc. This is the most used type of data in urban modelling.

Thus, survey data are most often attributive while sensor data are mostly spatial. There is always this transition between vector and raster (and/or raster and vector). Let's take the example of data from fixed sensors, we will be able to collect these data on several points in the city, and generalize them over the entire study area using the interpolation method. We can therefore go from a discontinuous data point to a continuous or matrix data by applying a spatial interpolation (Tabios and Salas, 1985).

4. Overview of ML methods for Urban form application

The need to know how to build computers that automatically improve through experience has made the field of machine learning one of the most emerging technical fields (Jordan and Mitchell, 2015). Machine learning, which is at the intersection of computer science and statistics, and at the heart of intelligence and data science, is seeing its field of application grow by the day. Thus, several definitions can be attributed to the term Machine Learning (ML). In the literature, it is often attributed the name "Machine Learning" or "Statistical Learning" or "Artificial Learning". Another famous definition of ML is that of T. Mitchell (Mitchell et al., 1997) according to which: *"A computer program would learn from experience E with respect to a certain class of tasks T and performance measure P, if its performance on tasks in T, as measured by P, improves with experience E."* This definition has been criticized by several authors over time including those of A. Munoz (Munoz, 2014), M. Mohri et al. (Mohri et al., 2018), M. Mitzenmacher et al. (Rish et al., 2001) and very recently D. Borchmann et al. (Borchmann et al., 2020). Overall, these authors have thus posed the general problem of ML which alludes to three main factors: (1) A sample of training data; (2) A training model; and (3) One or more metrics to evaluate and validate the model's performance.

Among these three elements, the size of the data and the model used are particularly important to achieve the best performance. That's why the advanced digitalization and the explosion of the volumes of data collected make the models evolve day by day. These models also become more and more complex to be more efficient. This efficiency has made ML inescapable in many and varied fields of application. Urban planning to meet the demands of today's cities is one of the areas where ML algorithms are increasingly used to meet this major challenge. Many applications of ML models in urban planning include the shape of the city. Thus, ML models have already made it possible to better model several elements related to the shape of the application city. The Tables 3 and 6 summarize recent works applying ML models to solve problems related to the shape of the city while the Table 4 draws up a comparative study of the various ML models used in this field. Depending on the target that we want to model by machine learning, ML methods are classified into two main groups: Supervised learning and unsupervised learning. However, there are other models called hybrid or semi-supervised learning that have been widely used in urban planning or learning methods by reinforcement, transfer or recurrent, etc.

The Table 1 draws up a comparison of the types of learning allowing us to classify each algorithmic model in order to detect its potential cases of effective application. Crossing the Table 1 with the Tables 3 and 6 allows to highlight the applications of the supervised and unsupervised learning models for the simulation of the indices associated to the shape of the city. We notice that the supervised learning methods are more widely used compared to the unsupervised methods, which are very little used. This shows that the data collected in most cases already have a precise and known target with the learning process except in some cases where clustering and association rules have been used. Another very important aspect in the automatic learning process concerns the nature of the problem which can be either a classification or a regression. It can be seen that previous works consisted of both classification processes (Novack et al., 2011; Lamb et al., 2019) and regression processes (Gómez et al., 2020; Lee, 2019).

The Fig. 3 details the ML process for simulating urban form indexes for use in urban planning decision support. Several important steps such as data collection and pre-processing precede the choice of ML the model. After training the model, the evaluation, tuning and validation of the performances by appropriate metrics follow. However, the heart of this process lies in the model whose choice must be carefully motivated and need expertise to design and train the best parameters. This is why in the rest of this section we describe the most commonly used automatic learning models in urban form modeling. We group them in three tables: 2, 3 and 6 containing unsupervised, classical suppervised and neural network based learning methods (simple or deep), respectively. Moreover, we discuss their advantages, disadvantages and applications in the field of urban form.

4.1. Unsupervised ML models

A learning problem is said to be unsupervised when only input data is available and no corresponding output variables (Barlow, 1989). Unsupervised learning is particularly used in descriptive analysis and is often used when dealing with urban forms (Han et al., 2021). The objective of unsupervised learning is to model the underlying structure or distribution in the input data set in order to learn more. The name unsupervised learning is due to the fact that unlike supervised learning (see Section 4.2), there is no correct answer or teacher and therefore no a prior known label. The algorithms are left to their own mechanisms to discover and present the interesting structure of the data. Unsupervised learning problems can be grouped into clustering or association. The Table 2 summarizes a comparison between these two groups while giving potential applications to the problems of modeling urban form indexes.

4.1.1. Clustering methods

A clustering problem is a problem where one wants to discover inherent group in the data, such as the clusters of urban occupation types according to urban form indexes. It's a bit similar to multiclass classification, but here we don't provide the labels, the system understands the data itself and aggregates the data (Barlow, 1989). For example, given a set of images, group them into different objects. Several algorithms are proposed for this type of problem. As shown in the Table 2, the most used are K-means (Krishna and Murty, 1999; Li et al., 2020), EM algorithms (Moon, 1996; Han et al., 2021), Gaussian mixture models (Han et al., 2021; Reynolds, 2009), graph clustering (Schaeffer, 2007). Clustering approaches are often used to model indicators of urban forms such as urban tenure types (Abrantes et al., 2019), urban fabrics (Li et al., 2020), LST/UHI (Gao et al., 2020), building energy modelling (Han et al., 2021), ...

4.1.2. Association rules methods

An association rule learning problem is a modeling problem where one wants to discover rules that describe a large part of the data, such that buyers of X also tend to buy Y. Another example is a person who uses one means of transportation to get to work could still use another or could from a given urban area. The most used algorithms for this type of problem are the A-priori algorithm (Wei et al., 2009; He et al., 2018) and the CAR (Class Assoication Algorithm) (Nguyen et al., 2013; Lu et al., 2008). The Apriori algorithm is a very popular solution for associative problems. It allows you to find the most frequently used items together such as walking adults (Jack and McCormack, 2014) or uban vitality (He et al., 2018) based urban form indexe such as urban growth. An alternative to this algorithm for association rule problems is the CAR algorithm that was used to determine work travel behavior as a function of urban form in Chicago by (Lu et al., 2008).

The unsupported learning techniques thus studied are only a fine part of the work applying machine learning techniques to modulate the indices of urban forms. The following Section 4.2 focuses on supervised learning techniques which are widely used.

4.2. Classical Supervised ML models

4.2.1. DT

The decision trees (DT) algorithm is a technique for structuring a set of learning data in the form of trees made up of nodes and leaves. Each node represents the test on the given attribute, while the leaf represents the class (Ruggieri, 2002; Bashir et al., 2014). The basic decision tree induction algorithm is using the top-down recursive method of building a decision tree. The algorithm uses the information gain based on the entropy measurement as heuristic information, selects a sample of classification attributes that can be called (Tekouabou et al., 2021). The attribute then becomes test or decision attribute of the node (Jindal et al., 2016; Xu et al., 2016). Although fast and simple, the DT algorithm becomes very inefficient for large databases but remains widely used in many modeling problems. Moreover, as mentioned in the Table 4, the formed tree can become difficult to understand when the number of branches increases. Because of its instability characterized by high variance and bias, DT is the basis for other, more powerful approaches such as ensemble methods (Tekouabou et al., 2020; Tekouabou et al., 2021). Its relative efficiency and simplicity has earned it several applications in a variety of fields. DT has been used for the modeling of urban patterns associated with the shape of the city such as the environment (Chan et al., 2001), the urban structures types (Hecht et al., 2013), the urban land cover in Sao Paolo (Novack et al., 2011), the urban growth in Tehran (Shafizadeh-Moghadam et al., 2017), etc.

4.2.2. LR

The "logit" or logistic regression (LR) learning model is one of the oldest machine learning models. According to A. de Palma & J.F. Thisse (De and Thisse, 1989), V. Loonis (Loonis, 2006) et J. Hosmer et al. (Hosmer et al., 2013), the technique was first introduced by J. Berkson in 1944 (Berkson, 1944; Berkson, 1951). Given a variable to predict $Y = \{0, 1\}$ and a predictive variable (explanatory variable) $X = (x_1, x_2, \ldots, x_n)$, LR is based on the fundamental assumption of the Eq. (1).

$$ln\frac{p(X|1)}{p(X|0)} = b_0 + b_1.x_1 + b_2.x_2 + \ldots + b_n.x_n \quad (1)$$

The technique is famous because it is easily adaptable to both classification and regression problems. It is very efficient for binary classification problems and when the data are not numerous. But its main disadvantage in classification is that its performance is appropriate in binary classification and much less in multi-class classification and especially when the amount of data increases. Several models of this family have been used to model urban forms among which: BayesianRidge, Lasso Regressor and Ridge regressor by (Gómez et al., 2020) to predict urban growth based on spatio-temporal data. Ordinary Least Square (OLS) regressor which is a type of simple linear regressor is also used in the literature (Chen et al., 2020). Several works have also implemented models based on this algorithm to simulate the patterns of the urban form of which (Ma et al., 2020) for land values prediction, (Lee, 2019) for air quality or (Gao et al., 2020; Chen et al., 2020) to model the LST and many other works that show that the approach is very promising to detect urban pattern from urban form indexes.

4.2.3. KNN

The k nearest neighbor (KNN) algorithm is one of the classical learning algorithms that uses a non-parametric method based on the calculation of similarity or dissimilarity to solve a classification or regression problem (Zhang et al., 2017; Koumétio and Toulni, 2021). The principle of the KNN algorithm (see Fig. 6) consists in assigning the class or regression value by averaging the k nearest neighboring values, for numerical instances, or by applying the majority vote for k neighbors, if the values of the instances are categorical. After selecting the k nearest neighbors, the value can be predicted either by the average of the outputs of the k neighbor (uniform weighting) or by a weighted sum defined by a weighting function (Sinta et al., 2014; Acharya et al., 2017; Koumétio and

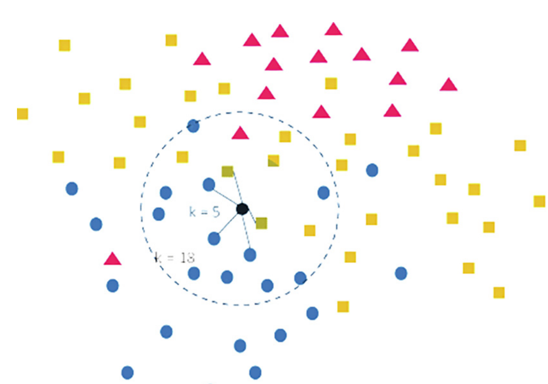

Fig. 6. Illustration of the search strategy for the k nearest neighbors.

Toulni, 2021). The KNN algorithm is also of particular interest for the design of skill regions for dynamic selection algorithms for set classifiers (Roy et al., 2018) or as a weak learner for constructing ensemble models and many other applications.

The great advantage of the KNN model is that it reduces the dimensionality of the features to improve efficiency by using two layers of feature reduction (Cherif, 2018; Al-Garadi et al., 2020). Then the KNN algorithm integrates a classification method based on the choice of distances (Euclidean, Manathan, Jaqqart...). The proposed KNN model has shown good prediction results of the land values (Ma et al., 2020), urban structure type (Hecht et al., 2013), land use ereas (Jochem et al., 2018) and UHI (Hart and Sailor, 2009). This diversity of application shows that the method could have an important scope of application for urban form modeling.

4.2.4. NB

The NB classifier is a type of simple probabilistic Bayesian classification based on the Bayes theorem (Walpole et al., 1993; Mitzenmacher and Upfal, 2017). The basic idea is to calculate the probabilities that the samples belong to their classes. The method is said to be naïve because the model is based on the assumption that all variables are independent. This is a very simple assumption that makes NB a very efficient classifier, especially when the size of the data and the number of variables is small (Lewis, 1998). The NB Classifier has solved many classification problems until it has been outperformed in recent decades by other more efficient classifiers such as the Ensemble Classifiers. NB approach has been applied for modelling LST/UHI (Chen et al., 2020) and urban functions from landscape metrics (Xing and Meng, 2020) according to urban form indicators. Globally, the NB algorithm has been slightly used to model urban patterns because of the complexity of the input data which makes its performance inappropriate. A similar approach was used by (Duerr et al., 2018) as Bayesians Bayesian additive regression trees (BART) model to forecast water demand according to urban form.

4.2.5. SVM

The Support Vector Machine (SVM) algorithm is one of the most widely used supervised machine learning methods for both classification and regression problems. The fundamental principles of SVM are derived from the work of C. Cortes & Vapnik (Cortes and Vapnik, 1995) who introduced in 1995 the principle of support vector network and the first vector support machine concepts (Hearst et al., 1998). This method has been widely used in the scientific community and has proven to be very robust for several general classification and regression problems (Vapnik, 2013).

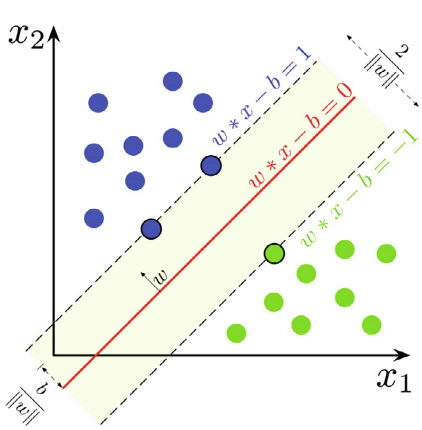

Fig. 7. Operating principle of a conventional SVM (Press et al., 2007).

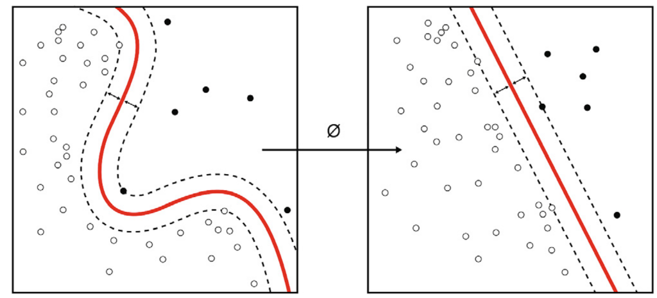

Fig. 8. From an SVM with RBF kernel (left) to an SVM with linear kernel (right) (Press et al., 2007). The choice of the kernel is very important to optimize the performance of the SVM algorithm. This figure shows that from an RBF kernel, we can switch to a linear kernel easier to understand as explained in (Tekouabou et al., 2021; Tekouabou et al., 2020).

SVM algorithms are based on two key ideas which are the concept of maximum margin and the concept of kernel function(Smola and Schölkopf, 2004). In linear classification, the vector support creates a hyperplane that separates the data into two subsets with a maximum margin (Smola and Schölkopf, 2004; Tekouabou et al., 2021) as illustrated in the Fig. 7. In particular, in cases where the data is not linearly separable, they map the data representation space into a larger area where a more appropriate linear separator is likely to be available.

The main objective of using SVM is to accurately classify invisible data by minimizing classification errors through a decision function (Jindal et al., 2016) and the appropriate kernel (Fig. 8).

The SVM algorithm has been particularly used in several prediction works whether in classification or regression. As shown in the Table 3, SVM is one of the most widely used algorithms for modeling patterns associated with the shape of the city. The use of SVM in urban planning has been widely diversified and the algorithm is very promising. Some examples of relevant use cases are the modeling of urban type structures (Hecht et al., 2013), land values prediction (Ma et al., 2020), type of urban occupation (Abrantes et al., 2019), Pollution ($PM_{2.5}$ (Kleine Deters et al., 2017), PM_{10} (Choung and Kim, 2019)), uban functions from landscape metrics (Xing and Meng, 2020), land uses (Okwuashi and Ndehedehe, 2020), etc. However, SVM suffers from some of the problems mentioned in the Table 4 which make several more powerful algorithms outperform its performance facing certain complex problems. This is the case of ensemble methods which he sometimes uses as a basic learner as we will see in the following section.

4.3. Ensemble methods (EM)

In ML modeling processes, one of the most powerful techniques is ensemble learning. It is the combination of weak models into a strong model. (Opitz and Maclin, 1999). Weak learners are in most cases decision trees (DT) or artificial neural networks (ANN) or support vector machines (SVM), nearest neighbors (k-NN) or naive bayes (NB) (Tekouabou et al., 2021). Each of these independent weak learners provides an alternative prediction of the overall problem, and the final prediction results from a combination (usually by weighted or unweighted voting) of these alternative predictions (Bolón-Canedo and Alonso-Betanzos, 2019). The EM approaches generally provide a more stable and accurate prediction because the error is much smaller than that provided by any of the individual base models that make up EM. (Diez-Olivan et al., 2019). EMs have been used a lot in recent years and also in our work. The ensemble technique is based on three main approaches: bagging (Breiman, 1996), boosting (Freund and Schapire, 1995) et stacking. In supervised statistical learning,

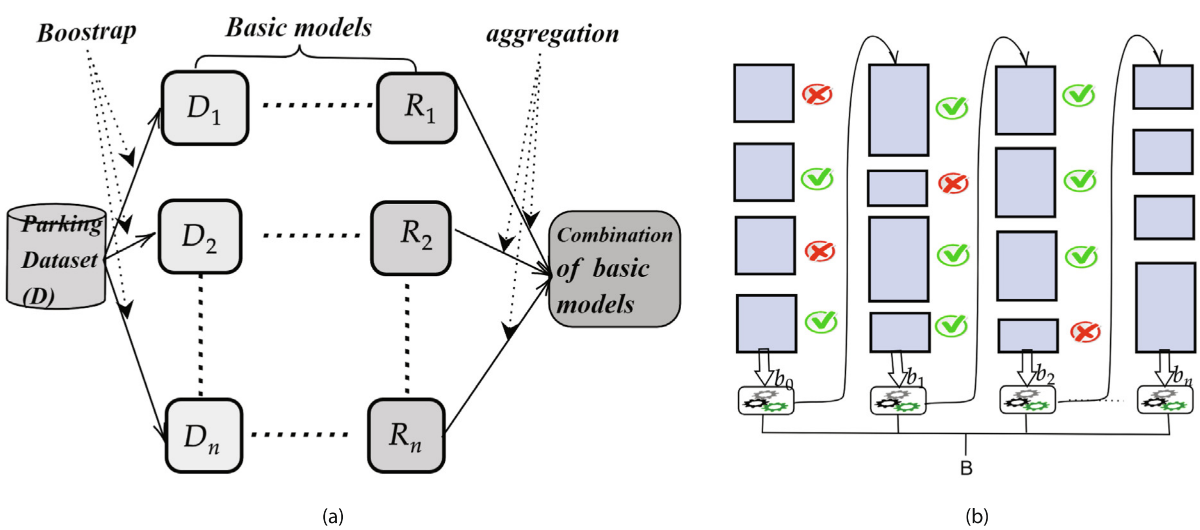

Fig. 9. Two large families of EM combinations (a) Bagging and (b) Boosting (Tekouabou et al., 2020; Tekouabou et al., 2021).

regression and classification, EMs have been particularly developed since 1984 with Breiman's work in statistics for bagging methods. (Breiman, 1996) et de Freund et Schapire in computer science a few years later for boosting methods (Freund and Schapire, 1995). Random forests (RF) (Breiman, 2001) have been very successful and have generated a lot of interest both in their theoretical aspects and in their numerous applications. These methods have been motivated by the fact that some models such as regression and classification trees are "unstable" (Bashir et al., 2014); a slight disturbance in the training sample can lead to a significant change in the structure of the model and improve its performance. Methods based on ensemble technique have been very relevant in the modeling of several urban planning indicators that we will review in turn according to the most used bagging or boosting techniques so far.

4.3.1. Bagging algorithm

The bagging algorithm owes its name to the contraction of "Bootstrap Aggregating" which the principle is depicted by the Fig. 9(a). It is the basis of a family of highly performing MEs for regression problems and supervised classification. We designate by (X, Y) a random vector representing the learning data where X takes its values in R_p and Y in R. We denote $D_n = (X_1, Y_1), \ldots, (X_n, Y_n)$ an independent and equitably distributed sample. And with the same law as (X, Y) and $\hat{m}(x) = E[Y|X = x]$ the classification/regression function. For $x \in R_p$, we consider the mean square error of an estimator m and its bias-variance decomposition as follows (Tekouabou et al., 2020; Bashir et al., 2014):

$$(\hat{m}(x) - m(x))^2 = (E\hat{m}(x) - m(x))^2 + V(\hat{m}(x)) \quad (2)$$

They consist in aggregating a number B of the models $\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_B$ such that:

$$\hat{m}(x) = \frac{1}{B}\sum_{k=1}^{B}(\hat{m}_k(x)) \quad (3)$$

and so we have:

$$E\hat{m}(x) = Em_1(x) \quad (4)$$

and

$$V(\hat{m}(x)) = \frac{1}{B}V(m_1(x)) \quad (5)$$

The bias of the aggregate model is therefore the same as that of the $\hat{m}_k$ but the variance decreases. Of course, in practice it is almost impossible to consider independent models $\hat{m}_k$ insofar as they all depend on the same sample D_n. The bagging approach is to try to mitigate the dependency between the models that are aggregated by building them on bootstrap samples. Almost as efficient as the RF algorithm which owes its basic principle to it, the bagging algorithm has however been widely used to model urban planning indices. Some rare works using this method include (Hecht et al., 2013) for urban structure types modelling and (Novack et al., 2011) for urban land cover.

4.3.2. Random Forest (RF) Algorithm

RF algorithm is nothing more than a particular method of bagging consisting of an aggregation of weak learners based on random variables (Breiman, 2001). Most often, the weak learner consists of trees built with the Classification and Regression Tree Algorithm (CART) whose principle is to recursively partition the space generated by the explanatory variables in a dyadic way (Bashir et al., 2014). More precisely, at each partitioning step, a part of the space is divided into two sub-parts according to a variable (Zhang and Haghani, 2015). In its most classic formula, it performs parallel learning on multiple weak learners randomly constructed and trained on different subsets of data (Breiman, 2001). The ideal number of weak learners is an important parameter that can be several hundred or more. It is highly variable and depends on the problem. Concretely, each weak learner in the random forest is trained on a random subset of data according to the principle of "bagging", with a random subset of characteristics (variable data characteristics) according to the principle of "random projections". The predictions are then averaged when the data are quantitative or used for a vote of qualitative data in the case of trees of classification (Breiman, 2001; Bashir et al., 2014). RF algorithm is known to be one of the most efficient "out-of-the-box" classifiers (i.e. requiring little data preprocessing). RF has been used in many applications, including consumer and complex applications such as real-time image classification. In this trend, it is certainly, as shown in the Table 3, the most widely used algorithm for modeling urban planning indices according to the parameters associated with the shape of the city. Among its many applications where RF has most often defied other algorithms, we can cite the following urban planning pattern modelling: land values (Ma et al., 2020), urban structure types (Hecht et al., 2013), water quality (Wang et al., 2020), land use areas (Jochem et al., 2018), water demand (Duerr et al., 2018), gentrification (Reades et al., 2019), urban land cover (Novack et al., 2011), LST/UHI (Sun et al., 2019; Chen et al., 2020), urban growth (Gómez et al., 2020; Shafizadeh-Moghadam et al., 2017), urban functions (Xing and Meng, 2020) etc.

4.3.3. Extra-Trees (ET) algorithm

The Extra-Trees algorithm (extremely randomized trees) (Geurts et al., 2006) are a set of decision trees built from bagging as explained above. However, they differ from other set methods on several points including: (1) the separation that takes place in the internal nodes is random, the attributes and thresholds tested are chosen at random; (2) Extra-Trees use the whole learning set to build the trees, not just a part as in the bagging method. Relative to the other EM, the ET algorithm remains however little used.

4.3.4. Adaptif Boosting (Adaboost) algorthm

The adaptive boosting (Adaboost) algorithm is part of the family of boosting EMs. AB is based on the fact that a new predictor to correct the error of its predecessor simply gives a little more attention to training cases on which the predecessor has adapted less. The result is new predictors that increasingly focus on difficult cases. For example, to create an AdaBoost classifier, one must consider a first classifier that is nothing more than a decision tree. This trained primitive tree is used to make predictions about the training set. The weight corresponding to misclassified trainings is then increased. A second classifier is then formed on the basis of these updated weights (Mishra et al., 2017). The second classifier again makes predictions about training data. The weights are then updated and so on. Once all the predictors have been trained, the game makes predictions very similar to bagging or gluing operations. The only difference is that the resulting predictors have different weights based on their overall accuracy over the weighted training sets (Géron, 2019). AB algorithm has been used to model the land values (Ma et al., 2020) and build height (Geiß et al., 2020) according to urban form features.

4.3.5. Gradient Boosting (GB) algorithm

GB algorithm works in a similar way to AdaBoost by sequentially adding predictors to a set, so that each tries to correct the errors of its predecessor (Freund et al., 1999; Freund et al., 1996). However, instead of adjusting the weights of the instances at each iteration, as AdaBoost does, this method tries to adjust the new predictor to the residual errors committed by the previous one (Géron, 2019). The steps concerning the process of running these boosting set algorithms as shown in Fig. 9(b) were also detailed in (Tekouabou et al., 2019; Tekouabou et al., 2020). The GB algorithm is another very popular boosting method that has given good performance in several applications. Gb has been applied to detect pollution $PM_{2.8}$ particle in urban area (Kleine Deters et al., 2017), water demand prediction (Duerr et al., 2018), solid waste management (Kontokosta et al., 2018), urban traffic flow (Liu et al., 2019) and building heights (Milojevic-Dupont et al., 2020).

4.4. NN based models

ML based on neural networks and particularly in depth is the fastest growing area of AI that has revolutionized the field of machine learning. The first simulations of human neural networks were the perceptrons created in 1957 by F. Rosenblatt (Rosenblatt, 1958). The algorithm will inspire the current concept of deep learning through multilayer perceptrons, artificial neural networks, deep neural networks, convolutional neural networks and recurrent neural networks, etc. Deep learning is used in tasks such as visual recognition, speech recognition, natural language processing and biomedicine and has had very favorable results. The technique is now very much in demand in urban sciences as in all other fields. To facilitate the simulation, nodes (neurons) are allocated at several levels to simulate a neural network. Generally, there is an input layer and an output layer. In a deep learning network, there are also several hidden layers. Deep learning uses several matrix operations to simulate the functioning of neurons. Calculations in matrix operations are generally simple. However, when many calculations are required, parallel operation is more appropriate. Deep learning methods have been shown to outperform previous state-of-the-art machine learning techniques in several areas, computer vision being one of the most important cases (Voulodimos et al., 2018).

4.4.1. ANN

The operating principle of the artificial neural network (ANN) algorithms represented in Fig. 10 is very schematically inspired by the functioning of the biological neurons to which the method owes its name. Indeed, neurons receive signals (electrical impulses) through highly branched extensions of their cell body (dendrites) and then send information through long extensions (axons) (Thayse et al., 1988). The ANN algorithm learns a particular model using a direct-acting neural network formed by a back-propagation algorithm. In this operation, the neuron is primarily a mathematical operator that performs a weighted sum followed by a non-linear function. The particularity of this function is that it must be limited, continuous and differentiable. The most commonly used are the sigmoid functions (Wang, 2003) because the forms of the derivatives of their inverse functions are extremely simple and easy to calculate, which improves the performance of the algorithm. The working principle of the ANNs is shown in the Fig. 10 and better detailed in (Thayse et al., 1988; Wang, 2003) The oldest neural networks still widely used today are the Multilayers Perceptron (MLP). (Rosenblatt, 1958). Their architecture is shown in the Fig. 11 (on the left side).

For the modeling of urban form indicators, both MLPs and ANNs have proven to be effective in a number of studies. For example, MLP has been used for modelling the land values (Ma et al., 2020), Change in urban environment (Chan et al., 2001) while the ANN algorithm modelled the solid waste management (Ma et al., 2020), urban typology (Nice et al., 2020). However, NN has still been widely used under other architectures including deep and/or convolutional learning.

4.4.2. CNN

One of the drawbacks of traditional ANNs is the parameters used, so CNNs were introduced to overcome this problem by reducing the data parameters (Guo et al., 2017). The principle of parameter reduction is based on three concepts: sparse interaction, parameter sharing and representation. Reducing connections

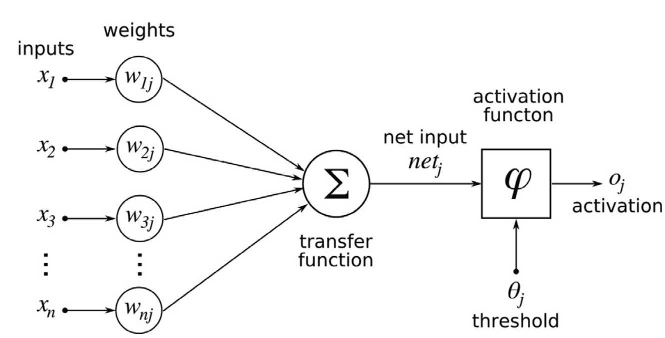

Fig. 10. Operating principle of an artificial neural network.

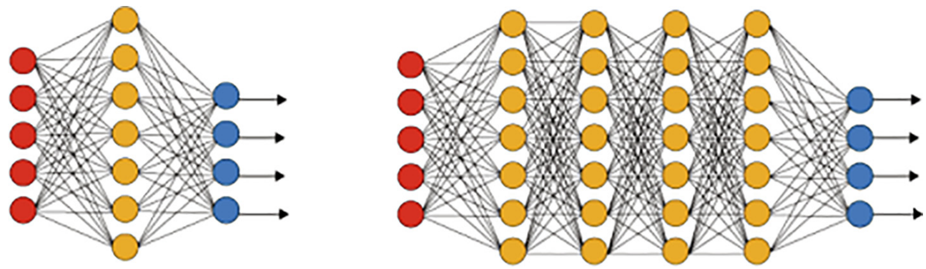

Fig. 11. Comparison between a simple neural network (left) and a deep neural network (right). The input, hidden and output layers are respectively in red, yellow and blue colors.

between layers increases scalability and improves the complexity of NNN training time. The operating principle of CNN is thus inspired by biological processes (Matsugu et al., 2003). They consist of a multilayer stack of perceptrons whose purpose is to pre-process small amounts of information (Kuo, 2016). CNNs have proven to be effective in large applications including image and video recognition, recommendation systems and natural language processing (LeCun et al., 1995; Matsugu et al., 2003). CNNs have been shown to provide better performance in several urban shape indication modeling projects, including the following: flood prediction (Guo et al., 2020), climate (Middel et al., 2018), mobility in street (Middel et al., 2019) etc. CNNs are potentially very promising for urban form modelling applications that are mostly image-based. The method has been at the origin of deep convolutional networks and other approaches more efficient to solve complex problems.

4.5. DNN-based models

The DNN is a kind of NN consisting of several layers of non-linear processing units (hidden layers). Its architecture makes it efficient for feature extraction and transformation. The Fig. 11 shows that a DNN made up of three layers, the input layer, the hidden layers, the output layer. In the input layer, neurons are generalized from features obtained from sensors that perceive the environment (Han et al., 2015). Hidden layers may consist of one or more layers, the neurons in them are called feature representations. The output layer contains the results we want, for example the distribution of all possible actions. Each successive DNN layer uses the output of the previous layer as input. All neurons in the layers are fully activated by weighted connections.

Different activation functions can be used to solve different problems. You can also create your own activation function to solve your own problem. There are also several predefined functions, such as "re-read", "tanh", softmax, sigmoid, etc. After the calculations, the flow undergoes from the input to the output, in the output layer and in each hidden layer, we can calculate the error derivatives backwards, and back propagate the gradients to the input layer, so that the weights can be updated to optimize a certain loss of function. This is the heart of the learning part, to find the right weights and biases. The following deals with the different types of DNNs.

4.5.1. DCNN

The DCNNs come from the deepening of the layers of the NDCs that we presented in the Section 4.4.2. DNNs efficiently process large amounts of input data and are therefore widely used in the fields of computer vision. They consist of a stacking of layers to form different CNN deep network architectures. Thus, a DCNN consists of three main layers: (1) The "convolutional layer" which uses a small patch size to scan over the entire image in giant steps. (2) The "pooling Layer" which pools the layers performs a subsampling operation and reduces the input dimensions. Its function is to progressively reduce the spatial dimensions size of the representation and therefore the number of parameters, the calculation in the network and therefore also to control the overlay. There are many types of pooling layers: Max-pooling, Average-pooling, etc. (3) The "full connected layer" which like the normal layer of the deep neural network is fully connected completely to all activations of the previous layer. There are famous CNN and DCNN based architectures with complex image data sets such as LeNet-5 (LeCun et al., 2015), AlexNet (Wang et al., 2019), Imagenet (Krizhevsky et al., 2012; Deng et al., 2009), Densenet (Huang et al., 2017).

4.5.2. RNN

RNNs are based on the work of (Rumelhart et al., 1986) published in 1986. MLPs can be useful for entries of fixed size, but what about cases where the size of the entries varies, for example natural language sentences? One solution would be to divide the problem into pieces and consider each piece independently using MLPs. However, in naive cases, it is the interaction between the following elements of a sequence that carries the most information. RNN is a type of non-convex model that addresses these concerns. RNN is essentially a sequence of MLPs. When it calculates a hidden vector, it takes into account not only the input but also the value of the hidden vector that precedes it in the sequence (Subasi, 2020). The grating training techniques are the same as for classical gratings (gradient backpropagation), however RNNs face the problem of gradient disappearance to learn how to memorize past events. New architectures have therefore been proposed to overcome this problem, in particular LSTM and GRU networks. Recurrent Neural Network (RNN) based deep learning methods such as LSTM and GRU perform better than auto regressive integrated moving average (ARIMA) model (Fu et al., 2016).

4.5.2.1. Long Short-Term Memory (LSTM). The LSTMs consists of a cell, an input gate, an output gate and a forget gate. The cell stores values at arbitrary time intervals and the three gates regulate the flow of information in and out of the cell (Hochreiter and Schmidhuber, 1997). The LSTM architecture has been proposed to overcome the problem of gradient disappearance posed by RNNs networks.Indeed, RNNs are great tools adapted to a wide variety of tasks. However, they have a serious flaw: when they are trained on sequences longer than a few elements, they suffer from gradient problems that disappear and explode. This was initially a major obstacle to the successful application of RNN to real-world problems. Interestingly enough, the LSTM solution was not really algorithmic, but architectural. By manipulating the set of equations defining the network architecture, (Hochreiter and Schmidhuber, 1997) have succeeded to ensure that there was always a direct path from the network output to any input, thus mitigating the problem of explosive gradients that disappear. LSTM works well for time series prediction. Thus, if data is in a sequential format (Time, Sentence, etc.,) such as many problems in urban modelling, LSTM is good to go for those kind of problems. The LSTM have been well applied to challenge the prediction of quality of air (Janarthanan et al., 2021) and traffic flow (Fu et al., 2016) according to urban form. According to the Tables 7, LSTMs, like their recurrent counterpart in recurrent-based principle GRU, present great potential for the modeling of time series of data associated with urban forms.

4.5.2.2. Gated recurrent unit (GRU). The GRU network is a RNN-based Encoder- Decoder consisting of two RNNs networks. The first one encodes a sequence of symbols into a fixed length vector representation, and the other one decodes the representation into another sequence of symbols. The encoder and decoder are then jointly trained to maximize the conditional probability of a target sequence (Cho et al., 2014; Chung et al., 2014).

4.5.3. Autoencoder (AE)

AEs allow the input to be coded into a different representation, and reconstruct it from that learned representation. Fig. 12 shows the basic architecture of AEs which is divided into two parts: the encoder part and the decoder part. Once the model is well trained, the encoder part can be used to extract characteristics from the data. They are therefore trained to reconstruct their original input. Once the model is well trained, the encoder part can be used to extract characteristics from the data (Makhzani et al., 2015). Indeed, an AE is a form of feature extraction algorithm that can use the features generated by an AE in any other algorithm, for example for classification.

The simplest form of an AE is an MLP with a single hidden layer. But Geoffrey Hinton has developed a pre-training technique for deep auto-encoding (DAE) (Hinton and Salakhutdinov, 2006). This method consists in treating each neighboring set of two layers as a restricted Boltzmann machine so that the pre-training approaches a good solution, then using the backpropagation technique to refine the results (Hinton and Salakhutdinov, 2006). The main objective of DAEs is to make features more robust by presenting, for example, only corrupted images and not real images. These models have proven to be very efficient and are able to learn very powerful representations. Since this does not change the model itself but only the data it is trained on, moreover, it is very simple to apply this principle to any auto-encoder (Hinton et al., 2011).

An alternative to DAEs is to force the model to learn a sparse representation of the input data; these models are called sparse auto-encoders (Ng et al., 2011). Typically, these models have a wider representation than other models, but the units are rarely

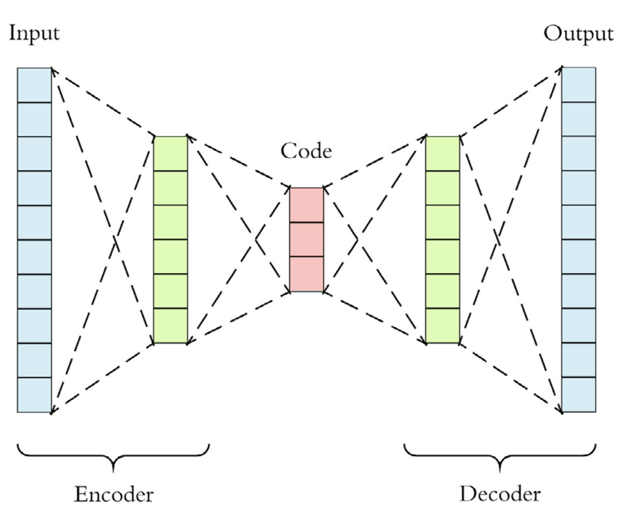

Fig. 12. Architecture of a basic auto-encoder (Hinton and Salakhutdinov, 2006).

activated simultaneously. They are usually trained by adding an additional term to the loss function (Ng et al., 2011; Mienye et al., 2020). urban structure (Moosavi, 2017), Quality of air (Xayasouk et al., 2020).

4.5.4. Transfer Learning (TL)

The TL can be seen as the ability of a system to recognize and apply knowledge and skills, learned from previous tasks, to new tasks or areas sharing similarities (Tan et al., 2018). The question that arises is: how to identify similarities between the target task (s) and the source task(s), and then how to transfer knowledge from the source task(s) to the target task(s)? There are three research challenges in transfer learning, namely "what we transfer", "how to transfer" and "when to transfer". The benefits of TL include the following: low resource requirements, easy to implement, short execution time, better performance for some complex problems (Tan et al., 2018). Research on new original DNN architectures funded by the world's major corporations is rarely oriented towards urban data and is very expensive (Jordan and Mitchell, 2015; Al-Garadi et al., 2020). TL techniques are very important as they allow the transfer of the training parameters of these models and adapt them to urban data with good performances (Verma et al., 2019). In practice, TL is commonly used with CNN for computer vision tasks, as this network is data intensive. Instead of training this network from scratch, a CNN pre-trained on a huge dataset (such as Imagenet which contains 1.2 million images (Krizhevsky et al., 2012; Deng et al., 2009)) can be exploited in three ways. Mamy algorithms are available for transfer learning in Markov logic networks (Mihalkova et al., 2007), Bayesian networks (Niculescu-Mizil and Caruana, 2007) and DNN (Tan et al., 2018). Transfer learning has also been applied to cancer subtype discovery (Hajiramezanali et al., 2018), building utilization (Mihalkova et al., 2007; Arief-Ang et al., 2018), digit recognition (Maitra et al., 2015), etc. They are very potential for urban form indicators modelling application such as urban growth model proposed by (Shafizadeh-Moghadam et al., 2017).

5. Discussion: potentials, challenges and future directions

This section is dedicated to the discussion of ML methods as decision support tools in the urban planning process. The major aspects of this discussion are ML potentials and issues, challenges and finally future research directions. We will refer much more to the data on most relevant papers in the literature that we summarized in Tables 3 for those involving non NN-based methods and 6 for the others. The summery is carried out according to the year of publication, type of ML model (i.e. the type of learning problem), data sources, study areas and target problems. This summary addresses issues related to the implementation of ML strategies and the possibilities of integrating ML with other technologies, issues of computational complexity, and the requirements of trade-offs between urban form, its indicators and urban planning.

5.1. Potentials and issues of ML applications for urban planning DSS

ML models currently offer enormous potential for applications in several areas including urban planning. To better convey this section, we have illustrated the potential of ML in Fig. 13. This figure shows that ML methods are at the center of effective modeling of urban indexes to aid in intelligent planning decisions for urban forms that meet current and future challenges.

From the point of view of urban planning and the uses of urban forms, ML models address several issues that we group here into three major categories including: a) the socio-economic issues, which address challenges such as inclusiveness, development and

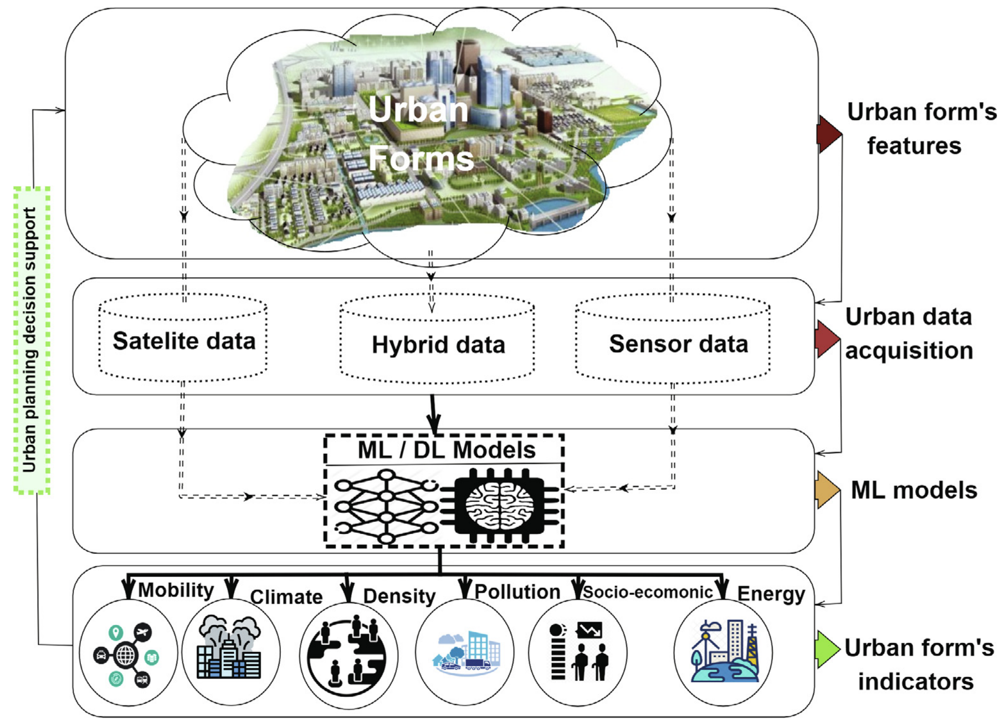

Fig. 13. Illustration of the potential role of ML for urban forms applications.

resilience of the city, etc.; b) the environment and sustainability, which includes indicators to address the challenges of ecological sustainability of the city, and finally c) the digital and smart city, which includes indicators of digital integration to address the challenges of smart and digital city. We will address these different aspects in turn in the following sub-sections.

5.1.1. ML for addressing urban sustainability issues

In general, ML models help predict a target indicator based on several variables. This predictive analysis allows to anticipate some extreme phenomena. Thus, thanks to ML models, urban planning can be assisted by decision support tools to simulate and plan the sustainable urban forms development. The type of building, the types of occupation of urban spaces, the density of the population, etc considerably affect the sustainability of the city. The strength of ML is that it allows to simulate these indicators according to the urban forms and find the most sustainable one by detecting indicators changes (Chan et al., 2001). Ml models can be useful for designing IoT and AI which provide tools to monitor air pollution, UHI, LST (Xu et al., 2019), solid waste and other indicators of urban form in real-time. ML tools can allow the predictive identification of pollution sources in order to simulate sustainable urban form indicators. According to a report by the WHO, 97% of cities in low- and middle- income countries with more than 100,000 inhabitants do not meet WHO air quality guidelines (Osseiran and Chriscaden, 2016). Several works have shown that ML models could be used to operate urban waste management and recycling, energy use according to urban form elements. The power of ML algorithms allows to evaluate environmental indicators on a large scale (Liu et al., 2017). Multiple examples show that the application of smart methods is central to modern sustainable urban form planning.

5.1.2. ML for addressing digital and smart city issues

The creation of smart, digital and connected cities is at the center of modern urban planning. The evolution of urban forms in today's cities is a key issue with the galloping urbanization worldwide (Ma et al., 2020; Middel et al., 2019; He et al., 2018). To become more influential, cities must demonstrate digital transformation and integration initiatives by creating true smart cities. The Internet of Things is at the heart of the smart city and enables the enrichment of collected urban data into Urban Big Data (UBD). These UBD not only aggregate information on all urban activities but are also the ideal raw material for ML algorithms. Indeed, from these UBD, ML will allow the creation of AIs allowing to make the city more intelligent (Jordan and Mitchell, 2015). But also and most importantly, the processing of UBD by the rising power of ML algorithms will allow urban planners to predict the evolutionary trends of the city and to regulate them by orienting the shape of the city towards the most sustainable, intelligent, digital and connected form possible. Thus, urban data enriches the ML to create AI for IoT and intelligent urban planning which then makes the city more and more intelligent, digital and connected which in turn helps enrich UBD. And the cycle repeats over and over. The power of ML algorithms, the engine of AI and therefore of IoT applications, and intelligent urban planning are crucial for addressing the challenges of smart and digital city (Al-Garadi et al., 2020; Jordan and Mitchell, 2015).

5.1.3. ML for addressing socio-economic issues

In the previous section, we have shown that the ML could allow the development of sustainable urban forms. This is achieved through intelligent urban planning decision support tools. Smart planning would help design and mobilize appropriate investments in the development of sustainable, resilient and inclusive urban forms. Indeed, with ML, city planners can simulate and predict the number of people who will continue to suffer the adverse effects not only of climate change, but also of reduced economic growth, lower quality of life and increasing social instability (Osseiran and Chriscaden, 2016). These simulations should make it possible to move towards the most inclusive urban forms that offer better socio-economic indicators (Grant et al., 2010). Some applications of ML for handling socio-economic issues from urban form features include: urban vitality (He et al., 2018), travel behavior (Lu et al., 2008), gentrification (Reades et al., 2019), urban functions (Xing and Meng, 2020), etc.

Some urban economic infrastructure arrangements have proven to be more efficient, resilient, and inclusive than others, offering better quality of life. Thanks to ML methods, we can now better model the patterns of these urban forms in order to refine those of future cities to meet the needs of rapid urbanization. In Africa for example, ML models are at the center of segmentation of slums and predictive modeling of their evolution to better anticipate responses. In Africa, ML also helps better target areas of poverty and future social conflict to guide NGO aid and government strategies (Osseiran and Chriscaden, 2016).

5.1.4. ML for addressing urban land use optimization issues

The use of space is central to urban planning (Schwarz, 2010). Some even define the urban forms that are the result of urban planning as the way in which urban space is used. This is the spatial component of the problem of modeling by ML techniques (Geng et al., 2019; Gómez et al., 2020; Faghmous and Kumar, 2014). This spatial component is most often based on vector image data from sensors. Classical models use a lot of pre-processing techniques to reduce this data before training the models with relatively low accuracies. ML and particularly DL methods allow to process this data faster, easier and more efficiently (Jordan and Mitchell, 2015). Thus, they allow to optimize the accuracy of land use DSS in order to make better long term predictions and anticipate phenomena such as urban sprawl. Several studies have used ML approaches to predict for example urban growth gomez2020spatiotemporal,shafizadeh2017coupling, land values (Ma et al., 2020), urban land cover (Novack et al., 2011), land use areas (Jochem et al., 2018), type of urban occupation (Abrantes et al., 2019), urban function from landscape metrics (Xing and Meng, 2020). Controlling the use of urban space is key to better directing the growth of the city towards the most sustainable forms while ensuring the best uses.

5.2. Challenges of ML for urban applications

The rapid growth of the urban population throughout the world in general and more particularly in Africa. The type of occupation of urban spaces highlight the planning challenges of the cities of the future (Novack et al., 2011; Shafizadeh-Moghadam et al., 2017). The city is more than ever at the center of tomorrow's challenges. One of the greatest challenges of urban planning today is to produce urban forms that meet the challenges of today's cities. With global warming, smart city, digital deployment, political instability, urban sprawl, public health, changes in lifestyles and work habits (He et al., 2018), our world and its cities are at the heart of many transformations. These changes need to be accurately represented in the collected datasets in order to design the most realistic models possible, which is not always the case. Thus ML applications to urban planning must cope with data collection and algorithmic complexity. We discuss these two challenges in the remainder of this section.

5.2.1. ML Implementation and Computational complexity

The ML algorithms are sometimes complex and require adequate skills to build the optimal model in order to obtain the desired results. This complexity could be further increased with the large volume of urban data collected nowadays. As a result, the implementation and deployment of ML and especially DL models requires high computational capability and energy to operate. Energy consumption is a crucial challenge for the adoption of ML for real-time application (Jordan and Mitchell, 2015). One of the current execution and deployment solutions is the use of cloud servers which in turn suffer from a high wireless power overhead. Furthermore, the availability of applications for such solutions is based on network conditions. Therefore, if the network connectivity is low, the offloading of the cloud will be impossible, which results in the unavailability of the applications (Al-Garadi et al., 2020). On the other hand, ML frameworks capable of effectively reducing computational complexity should be developed. Better yet, graphical interfaces to efficiently process urban data for urban planners who do not necessarily have advanced ML skills should be developed. Thus, reducing computational complexity and increasing algorithmic transparency is of practical importance for future research.

Another very important challenge is that many emerging DL model architectures are not trained on urban data. They are then adapted to urban data with relatively lower performance and time gab. Research on new ML models (DNN architecture) specific to urban data can be extremely costly and only a few institutions such as Facebook do it. In this sense, transfer learning could be a very important technique to implement these models more easily to urban data but there will always be this slow development compared to other fields (Chang et al., 2019).

5.2.2. The challenge of ML for urban data

In Section 3.3 we have discussed the different sources and types of urban data. The Section 4 enlightened us on the methods of machine learning and eventually their potential applications.

As mentioned in the Section 3.3, data is the lifeblood of ML applications. They have used various urban data sources which we discuss here in terms of the sources as mentioned in Tables 3 and 6. The objective here is to know the types of data, their sources and suitability for ML applications as grouped in Tables 2, 4 and 7. Globally, these tables show that both simple ML and NN-based methods are used for all types of data. However, we have noticed that the simple ML methods are much more often used when it comes to attribute data or data usually extracted from satellite images which are vector data, most often by deep learning methods (Guo et al., 2020; Middel et al., 2018; Middel et al., 2019). This allows ML to propose insightful solutions to the problem of modeling the spacial aspects of urban forms for various types of data. On the other hand, recent methods based on recurrent neural networks complement these spatial solutions by proposing models (RNN (Koschwitz et al., 2018; Lu et al., 2021; Geng et al., 2019), LSTM (Janarthanan et al., 2021; Fu et al., 2016; Lu et al., 2021), GRU (Janarthanan et al., 2021; Xayasouk et al., 2020)) for efficiently modeling temporal data of complex urban forms. This explains the completeness of ML's current strength in modeling urban form problems. From the point of view of the type of ML problem, temporal components are most often *a regression* while spatial component are association, clustering or mostly classification. Reinforcement learning has experienced a meteoric progress which has a great impact in urban form applications, especially through the simulation of new forms from millions of existing old forms in the world. Knowing that images are often complex to process and require more skills in ML algorithms implementation especially for complex problems, learning techniques by simpler transfer is also more solicited (Verma et al., 2019).

From Tables 3 and 6, we can see that urban data come from satellite image sensors and are therefore often large and complex. ML and particularly DL methods associated with currently emerging microprocessor techniques remain among the best tools of the moment to take advantage of these data. In one case or another, this profit is the result of the intelligent modeling that ML allows us. The question then arises: which models for which type of data and which urban applications? Thus, in addition to the data, in the previous sections, we have highlighted that the urban modeling problem is most often composed of spatial and temporal components. Whether the data is attribute or vector, from sensors, surveys or both, the Tables 7, 5 and 7 compare simple, ensemble and NN-based algorithms respectively by highlighting their applications to different problems. Thus, through ML-based intelligent decision support tools, urban planners can already simulate different indicators of urban form. However, there is an uneven distribution of scientific study areas around the world. Very few works have been conducted in African countries.

5.3. Future research directions

With the advent of UBD, most static models do not fit the evolution of today's cities. More efficient ML methods allow for better processing of these data. Thus, intelligent urban planning could benefit from it both in the design of urban forms and in the reorientation of the planning process according to the evolution of the city. The future direction of the applications of ML methods will consist very much in the diversification of the offered models, their adaptation to the scalability of the collected urban data. The types of urban data and their spatio-temporal characteristics show that DL methods for example will be used more. However, the algorithmic complexity and the explanatory power of the designed models will have to be addressed. The applications of ML to urban planning should better address the current challenges that we have highlighted in the previous sections. This has already been the case in several areas such as medical diagnosis, intelligent and automated marketing, automotive, etc. where research on the applications of ML models is quite advanced. Developing countries are still very little studied even though these areas have the highest urbanization rates in the world. More than 97% of these cities do not meet WHO air quality guidelines (Osseiran and Chriscaden, 2016). Therefore it becomes more than urgent to integrate new elements to the urban planning process (or at least the future urban forms should integrate the following new elements) in order to meet the challenges of the current cities. On the other hand, new emerging NN architectures such as GAN networks should be much more widely adopted to simulate the growth of the city and design innovative future urban forms that meet the current requirements (Shafizadeh-Moghadam et al., 2017). In order to include the processes of urban evolution in these models, nothing less than a paradigm shift of the growing or accelerating urban regions is required. This statement is valid not only for the field of urban simulation but also for urban research in general, which more than ever should be enriched by ML applications to meet modern city requirements. When the processes of urban evolution and mutation are analyzed and quantified empirically, simulation models can be built to support the planning of better future urban forms.

6. Conclusion

Intelligent predictive modeling is a major component of current urban planning to meet the challenges of modern urban forms. With the advent of UBDs, most static models are not well adapted to the evolution of today's cities. More efficient ML methods allow for a better processing of these data for both in the urban form usages. It can help for simulation, monitoring and better yet prediction of the best indicators to support decision making in intelligent planning process according to desired future of the city. Thus, the quite complex ML algorithms, little known in their great emerging diversity, are at the center of intelligent planning to meet the challenges in current and future urban planning. Therefore, in this paper, we have provided a detailed literature review of all the elements that enable the implementation of ML models for smart city planning. We started with an overview of urban planning and urban form by discussing the elements and indicators of modern urban form, followed by an overview of the sources of urban data. Then we reviewed the most popular methods of ML by highlighting their operating principles, advantages, disadvantages and especially their potential applications to the problem of urban modeling. Finally we discussed the potentials, challenges, and the future potential applications of ML models to urban planning. This study introduces a clear and transparent document that can encourage researchers to advance intelligent modeling of urban form indicators to support smart planning and thus meet the challenges of the sustainable, smart, digital, inclusive and resilient city of tomorrow. In addition to these challenges, there is a very uneven distribution in terms of study area. Almost all contributions are concentrated in Western and Asian countries while very few studies have been conducted in African and Latin American countries. Future research on the applications of ML methods to urban planning should strive to address the challenges highlighted in this study.

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

Abrantes, P., Rocha, J., Marques da Costa, E., Gomes, E., Morgado, P., Costa, N., 2019. Modelling urban form: A multidimensional typology of urban occupation for spatial analysis. Environ. Planning B: Urban Analytics City Sci. 46 (1), 47–65.
Acharya, U.R., Bhat, S., Koh, J.E., Bhandary, S.V., Adeli, H., 2017. A novel algorithm to detect glaucoma risk using texton and local configuration pattern features extracted from fundus images. Computers Biology Med. 88, 72–83.
Al-Garadi, M.A., Mohamed, A., Al-Ali, A.K., Du, X., Ali, I., Guizani, M., 2020. A survey of machine and deep learning methods for internet of things (iot) security. IEEE Commun. Surveys Tutorials 22 (3), 1646–1685.
Aniello, C., Morgan, K., Busbey, A., Newland, L., 1995. Mapping micro-urban heat islands using landsat tm and a gis. Computers Geosci. 21 (8), 965–969.
Arief-Ang, I.B., Hamilton, M., Salim, F.D., 2018. A scalable room occupancy prediction with transferable time series decomposition of co2 sensor data. ACM Trans. Sensor Networks (TOSN) 14 (3–4), 1–28.
Arribas-Bel, D., Garcia-López, M.-À., Viladecans-Marsal, E., 2019. Building (s and) cities: Delineating urban areas with a machine learning algorithm. J. Urban Econ. 103217.
Barlow, H.B., 1989. Unsupervised learning. Neural Comput. 1 (3), 295–311.
Bashir, S., Qamar, U., Khan, F.H., Javed, M.Y., 2014. An efficient rule-based classification of diabetes using id3, c4. 5, & cart ensembles. In: 2014 12th International Conference on Frontiers of Information Technology. IEEE, pp. 226–231.
Basiri, A., Amirian, P., Mooney, P., 2016. Using crowdsourced trajectories for automated osm data entry approach. Sensors 16 (9), 1510.
Berkson, J., 1944. Application of the logistic function to bio-assay. J. Am. Stat. Assoc. 39 (227), 357–365.
Berkson, J., 1951. Why i prefer logits to probits. Biometrics 7 (4), 327–339.
Bolón-Canedo, V., Alonso-Betanzos, A., 2019. Ensembles for feature selection: A review and future trends. Inform. Fusion 52, 1–12.
Borchmann, D., Hanika, T., Obiedkov, S., 2020. Probably approximately correct learning of horn envelopes from queries. Discrete Appl. Math. 273, 30–42.
Breiman, L., 1996. Bagging predictors. Mach. Learn. 24 (2), 123–140.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Caron, M., Bojanowski, P., Joulin, A., Douze, M., 2018. Deep clustering for unsupervised learning of visual features. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 132–149.
Chan, J.C.-W., Chan, K.-P., Yeh, A.G.-O., 2001. Detecting the nature of change in an urban environment: A comparison of machine learning algorithms. Photogrammetric Eng. Remote Sensing 67 (2), 213–226.
Chang, S., Saha, N., Castro-Lacouture, D., Yang, P.P.-J., 2019. Generative design and performance modeling for relationships between urban built forms, sky opening, solar radiation and energy. Energy Procedia 158, 3994–4002.
Chen, Y., Li, X., Wang, S., Liu, X., Ai, B., 2013. Simulating urban form and energy consumption in the pearl river delta under different development strategies. Ann. Assoc. Am. Geogr. 103 (6), 1567–1585.
Chen, S., Zhang, W., Wong, N.H., Ignatius, M., 2020. Combining citygml files and data-driven models for microclimate simulations in a tropical city. Build. Environ. 185, 107314.
Chengsheng, T., Huacheng, L., Bing, X., 2017. Adaboost typical algorithm and its application research. In: MATEC Web of Conferences. Vol. 139. EDP Sciences, p. 00222.

Cherif, W., 2018. Optimization of k-nn algorithm by clustering and reliability coefficients: application to breast-cancer diagnosis. Procedia Computer Sci. 127, 293–299.
Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation.
Choung, Y.-J., Kim, J.-M., 2019. Study of the relationship between urban expansion and pm10 concentration using multi-temporal spatial datasets and the machine learning technique: Case study for daegu, south korea. Appl. Sci. 9 (6), 1098.
Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling.
Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.
Dempsey, N., Brown, C., Raman, S., Porta, S., Jenks, M., Jones, C., Bramley, G., 2010. Elements of urban form. In: Dimensions of the sustainable city. Springer, pp. 21–51.
Demšar, J., Zupan, B., Leban, G., Curk, T., 2004. Orange: From experimental machine learning to interactive data mining. In: European conference on principles of data mining and knowledge discovery. Springer, pp. 537–539.
Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., et al., 2013. Orange: data mining toolbox in python. J. Mach. Learn. Res. 14 (1), 2349–2353.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, pp. 248–255.
De Palma, A., Thisse, J.-F., 1989. Les modèles de choix discrets. Annales d'Economie et de Statistique, 151–190.
Diez-Olivan, A., Del Ser, J., Galar, D., Sierra, B., 2019. Data fusion and machine learning for industrial prognosis: Trends and perspectives towards industry 4.0. Inform. Fusion 50, 92–111.
Duerr, I., Merrill, H.R., Wang, C., Bai, R., Boyer, M., Dukes, M.D., Bliznyuk, N., 2018. Forecasting urban household water demand with statistical and machine learning methods using large space-time data: A comparative study. Environ. Modelling Softw. 102, 29–38.
EOS, 1999. Advanced spaceborne thermal emission and reflection radiometer.
Faghmous, J.H., Kumar, V., 2014. Spatio-temporal data mining for climate data: Advances, challenges, and opportunities. In: Data mining and knowledge discovery for big data. Springer, pp. 83–116.
Frenkel, A., Ashkenazi, M., 2008. Measuring urban sprawl: how can we deal with it? Environ. Planning B: Planning Design 35 (1), 56–79.
Freund, Y., Schapire, R.E., 1995. A desicion-theoretic generalization of on-line learning and an application to boosting. In: European conference on computational learning theory. Springer, pp. 23–37.
Freund, Y., Schapire, R.E., et al., 1996. Experiments with a new boosting algorithm. In: icml. Vol. 96. Citeseer, pp. 148–156.
Freund, Y., Schapire, R., Abe, N., 1999. A short introduction to boosting. J.-Jpn. Society Artificial Intell. 14 (771–780), 1612.
Fu, R., Zhang, Z., Li, L., 2016. Using lstm and gru neural network methods for traffic flow prediction. In: 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE, pp. 324–328.
Gao, S., Zhan, Q., Yang, C., Liu, H., 2020. The diversified impacts of urban morphology on land surface temperature among urban functional zones. Int. J. Environ. Res. Public Health 17 (24), 9578.
Geiß, C., Schrade, H., Pelizari, P.A., Taubenböck, H., 2020. Multistrategy ensemble regression for mapping of built-up density and height with sentinel-2 data. ISPRS J. Photogrammetry Remote Sensing 170, 57–71.
Geng, X., Li, Y., Wang, L., Zhang, L., Yang, Q., Ye, J., Liu, Y., 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. Proceedings of the AAAI conference on artificial intelligence, Vol. 33. AAAI Press, pp. 3656–3663.
Géron, A., 2019. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.
Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63 (1), 3–42.
Gómez, J.A., Patiño, J.E., Duque, J.C., Passos, S., 2020. Spatiotemporal modeling of urban growth using machine learning. Remote Sensing 12 (1), 109.
Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. arXiv preprint arXiv:1406.2661.
Grant, T.L., Edwards, N., Sveistrup, H., Andrew, C., Egan, M., 2010. Inequitable walking conditions among older people: examining the interrelationship of neighbourhood socio-economic status and urban form using a comparative case study. BMC Public Health 10 (1), 677.
Guo, T., Dong, J., Li, H., Gao, Y., 2017. Simple convolutional neural network on image classification. In: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). IEEE, pp. 721–724.
Guo, Z., Leitão, J.P., Simões, N.E., Moosavi, V., 2020. Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks. J. Flood Risk Management. e12684.
Hajiramezanali, E., Dadaneh, S.Z., Karbalayghareh, A., Zhou, M., Qian, X., 2018. Bayesian multi-domain learning for cancer subtype discovery from next-generation sequencing count data. arXiv preprint arXiv:1810.09433. .
Han, S., Mao, H., Dally, W.J., 2015. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. arXiv preprint arXiv:1510.00149.
Han, M., Wang, Z., Zhang, X., 2021. An approach to data acquisition for urban building energy modeling using a gaussian mixture model and expectation-maximization algorithm. Buildings 11 (1), 30.
Hart, M.A., Sailor, D.J., 2009. Quantifying the influence of land-use and surface characteristics on spatial variability in the urban heat island. Theor. Appl. Climatol. 95 (3–4), 397–406.
Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B., 1998. Support vector machines. IEEE Intell. Systems Appl. 13 (4), 18–28.
Hecht, R., Herold, H., Meinel, G., Buchroithner, M., 2013. Automatic derivation of urban structure types from topographic maps by means of image analysis and machine learning. In: 26th International Cartographic Conference.
Hecht, R., Meinel, G., Buchroithner, M., 2015. Automatic identification of building types based on topographic databases–a comparison of different data sources. Int. J. Cartography 1 (1), 18–31.
He, Q., He, W., Song, Y., Wu, J., Yin, C., Mou, Y., 2018. The impact of urban growth patterns on urban vitality in newly built-up areas based on an association rules analysis using geographical 'big data'. Land Use Policy 78, 726–738.
Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313 (5786), 504–507.
Hinton, G.E., Krizhevsky, A., Wang, S.D., 2011. Transforming auto-encoders. In: International conference on artificial neural networks. Springer, pp. 44–51.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hosmer Jr, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied logistic regression, Vol. 398. John Wiley & Sons.
Huang, J., Lu, X.X., Sellers, J.M., 2007. A global comparative analysis of urban form: Applying spatial metrics and remote sensing. Landscape Urban Planning 82 (4), 184–197.
Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.
Ibrahim, M.R., Titheridge, H., Cheng, T., Haworth, J., 2019. predictslums: A new model for identifying and predicting informal settlements and slums in cities from street intersections using machine learning. Comput. Environ. Urban Syst. 76, 31–56.
Jack, E., McCormack, G.R., 2014. The associations between objectively-determined and self-reported urban form characteristics and neighborhood-based walking in adults. Int. J. Behavioral Nutrition Phys. Activity 11 (1), 71.
Janarthanan, R., Partheeban, P., Somasundaram, S., Elamparithi, P.N., 2021. A deep learning approach for prediction of air quality index in a metropolitan city. Sustainable Cities Society 102720.
Jindal, A., Dua, A., Kaur, K., Singh, M., Kumar, N., Mishra, S., 2016. Decision tree and svm-based data analytics for theft detection in smart grid. IEEE Trans. Industr. Inf. 12 (3), 1005–1016.
Jochem, W.C., Bird, T.J., Tatem, A.J., 2018. Identifying residential neighbourhood types from settlement points in a machine learning approach. Computers, Environ. Urban Syst. 69, 104–113.
Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. Science 349 (6245), 255–260.
Kabano, P., Lindley, S., Harris, A., 2021. Evidence of urban heat island impacts on the vegetation growing season length in a tropical city. Landscape Urban Planning 206, 103989.
Kafy, A.-A., Naim, M.N.H., Subramanyam, G., Faisal, A.-A., Ahmed, N.U., Al Rakib, A., Kona, M.A., Sattar, G.S., 2021. Cellular automata approach in dynamic modeling of land cover changes using rapideye images in dhaka, bangladesh. Environ. Challenges 100084.
Khan, K., Rehman, S.U., Aziz, K., Fong, S., Sarasvady, S., 2014. Dbscan: Past, present and future. In: The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014). IEEE, pp. 232–238.
Kleine Deters, J., Zalakeviciute, R., Gonzalez, M., Rybarczyk, Y., 2017. Modeling pm2. 5 urban pollution using machine learning and selected meteorological parameters. Journal of Electrical and Computer Engineering 2017.
Kontokosta, C.E., Hong, B., Johnson, N.E., Starobin, D., 2018. Using machine learning and small area estimation to predict building-level municipal solid waste generation in cities. Comput. Environ. Urban Syst. 70, 151–162.
Koschwitz, D., Frisch, J., Van Treeck, C., 2018. Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and narx recurrent neural network: A comparative study on district scale. Energy 165, 134–142.
Koumétio, S.C.T., Toulni, H., 2021. Improving knn model for direct marketing prediction in smart cities. In: Machine Intelligence and Data Analytics for Sustainable Future Smart Cities. Springer, pp. 107–118. .
Krishna, K., Murty, M.N., 1999. Genetic k-means algorithm. IEEE Trans. Syst., Man, Cybern., Part B (Cybern.) 29 (3), 433–439.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inform. Processing Syst. 25, 1097–1105.
Kuo, C.-C.J., 2016. Understanding convolutional neural networks with a mathematical model. J. Vis. Commun. Image Represent. 41, 406–413.
Lamb, W.F., Creutzig, F., Callaghan, M.W., Minx, J.C., 2019. Learning about urban climate solutions from case studies. Nature Climate Change 9 (4), 279–287.
LeCun, Y., Bengio, Y., et al., 1995. Convolutional networks for images, speech, and time series. Handbook Brain Theory Neural Networks 3361 (10), 1995.
LeCun, Y., et al., 2015. Lenet-5, convolutional neural networks. URL: http://yann. lecun. com/exdb/lenet 20 (5), 14.

Lee, C., 2019. Impacts of urban form on air quality: Emissions on the road and concentrations in the us metropolitan areas. J. Environ. Manage. 246, 192–202.
Lee, S.-H., Yigitcanlar, T., Han, J.-H., Leem, Y.-T., 2008. Ubiquitous urban infrastructure: Infrastructure planning and development in korea. Innovation 10 (2–3), 282–292.
Lewis, D.D., 1998. Naive (bayes) at forty: The independence assumption in information retrieval. In: European conference on machine learning. Springer, pp. 4–15.
Li, X., Cheng, S., Lv, Z., Song, H., Jia, T., Lu, N., 2020. Data analytics of urban fabric metrics for smart cities. Future Generation Computer Syst. 107, 871–882.
Liu, L., Silva, E.A., Wu, C., Wang, H., 2017. A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. Comput. Environ. Urban Syst. 65, 113–125.
Liu, Z., Liu, Y., Meng, Q., Cheng, Q., 2019. A tailored machine learning approach for urban transport network flow estimation. Transp. Res. Part C: Emerging Technol. 108, 130–150.
Loonis, V., 2006. Les déterminants de la réélection des députés français de. Histoire & Mesure 21 (XXI-1), 221–254. 1871 à 2002.
Lu, Y., Kawamura, K., Zellner, M.L., 2008. Exploring the influence of urban form on work travel behavior with agent-based modeling. Transp. Res. Record 2082 (1), 132–140.
Lu, S., Zhang, Q., Chen, G., Seng, D., 2021. A combined method for short-term traffic flow prediction based on recurrent neural network. Alexandria Eng. J. 60 (1), 87–94.
Luxen, D., Vetter, C., 2011. Real-time routing with openstreetmap data. In: Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems, pp. 513–516.
Ma, J., Cheng, J.C., Jiang, F., Chen, W., Zhang, J., 2020. Analyzing driving factors of land values in urban scale based on big data and non-linear machine learning techniques. Land Use Policy 94, 104537.
Maignant, G., 2005. Compacité et forme urbaine, une analyse environnementale dans la perspective d'un développement urbain durable. In: Actes du colloque Développement urbain durable, gestion des ressources et gouvernance.
Maitra, D.S., Bhattacharya, U., Parui, S.K., 2015. Cnn based common approach to handwritten character recognition of multiple scripts. In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR). IEEE, pp. 1021–1025.
Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., Frey, B., 2015. Adversarial autoencoders. arXiv preprint arXiv:1511.05644.
Matsugu, M., Mori, K., Mitari, Y., Kaneda, Y., 2003. Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Networks 16 (5–6), 555–559.
Middel, A., Lukasczyk, J., Maciejewski, R., Demuzere, M., Roth, M., 2018. Sky view factor footprints for urban climate modeling. Urban climate 25, 120–134.
Middel, A., Lukasczyk, J., Zakrzewski, S., Arnold, M., Maciejewski, R., 2019. Urban form and composition of street canyons: A human-centric big data and deep learning approach. Landscape Urban Planning 183, 122–132.
Mienye, I.D., Sun, Y., Wang, Z., 2020. Improved sparse autoencoder based artificial neural network approach for prediction of heart disease. Inform. Med. Unlocked 18, 100307.
Mihalkova, L., Huynh, T., Mooney, R.J., 2007. Mapping and revising markov logic networks for transfer learning. Aaai 7, 608–614.
Milder, J., 2012. Sustainable urban form. In: Sustainable Urban Environments. Springer, pp. 263–284.
Milojevic-Dupont, N., Hans, N., Kaack, L.H., Zumwald, M., Andrieux, F., de Barros Soares, D., Lohrey, S., Pichler, P.-P., Creutzig, F., 2020. Learning from urban form to predict building heights. Plos one 15, (12) e0242010.
Mishra, S., Mishra, D., Santra, G.H., 2017. Adaptive boosting of weak regressors for forecasting of crop production considering climatic variability: An empirical assessment. Journal of King Saud University-Computer and Information Sciences.
Mitchell, T.M., et al., 1997. Machine learning.
Mitzenmacher, M., Upfal, E., 2017. Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis. Cambridge University Press.
Mohri, M., Rostamizadeh, A., Talwalkar, A., 2018. Foundations of machine learning. MIT press.
Moon, T.K., 1996. The expectation-maximization algorithm. IEEE Signal Processing Magazine 13 (6), 47–60.
Moosavi, V., 2017. Urban morphology meets deep learning: Exploring urban forms in one million cities, town and villages across the planet. arXiv preprint arXiv:1709.02939. URL:https://arxiv.org/abs/1709.02939v2.
Munoz, A., 2014. Machine learning and optimization. URL: https://www. cims. nyu. edu/ munoz/files/ml_optimization. pdf [accessed 2016-03-02][WebCite Cache ID 6fiLfZvnG].
Ng, A., et al., 2011. Sparse autoencoder. CS294A Lecture Notes 72 (2011), 1–19.
Nguyen, L.T., Vo, B., Hong, T.-P., Thanh, H.C., 2013. Car-miner: An efficient algorithm for mining class-association rules. Expert Syst. Appl. 40 (6), 2305–2311.
Nice, K.A., Thompson, J., Wijnands, J.S., Aschwanden, G.D., Stevenson, M., 2020. The paris-end of town? deriving urban typologies using three imagery types. Urban Science 4 (2), 27.
Niculescu-Mizil, A., Caruana, R., 2007. Inductive transfer for bayesian network structure learning. In: Artificial intelligence and statistics. PMLR, pp. 339–346. .
Niu, H., Silva, E.A., 2020. Crowdsourced data mining for urban activity: Review of data sources, applications, and methods. Journal of Urban Planning and Development 146 (2), 04020007.
Novack, T., Esch, T., Kux, H., Stilla, U., 2011. Machine learning comparison between worldview-2 and quickbird-2-simulated imagery regarding object-based urban land cover classification. Remote Sensing 3 (10), 2263–2282.
Okwuashi, O., Ndehedehe, C.E., 2020. Integrating machine learning with markov chain and cellular automata models for modelling urban land use change. Remote Sensing Applications. Society and Environment, 100461.
Opitz, D., Maclin, R., 1999. Popular ensemble methods: An empirical study. J. Artif. Intell. Res. 11, 169–198.
Osseiran, N., Chriscaden, K., 2016. Air pollution levels rising in many of the world's poorest cities. WHO, Geneva.
Pascanu, R., Mikolov, T., Bengio, Y., 2013. On the difficulty of training recurrent neural networks. In: International conference on machine learning. PMLR, pp. 1310–1318.
Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. Numerical recipes. . 3rd edition. The art of scientific computing 3rd edition. Cambridge University Press.
Reades, J., De Souza, J., Hubbard, P., 2019. Understanding urban gentrification through machine learning. Urban Studies 56 (5), 922–942.
Reynolds, D.A., 2009. Gaussian mixture models. Encyclopedia of biometrics 741, 659–663.
Rish, I., et al., 2001. An empirical study of the naive bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence. Vol. 3. pp. 41–46.
Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Rev. 65 (6), 386.
Roy, A., Cruz, R.M., Sabourin, R., Cavalcanti, G.D., 2018. A study on combining dynamic selection and data preprocessing for imbalance learning. Neurocomputing 286, 179–192.
Ruggieri, S., 2002. Efficient c4.5 [classification algorithm]. IEEE transactions on knowledge and data engineering 14 (2), 438–444.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323 (6088), 533–536.
Schaeffer, S.E., 2007. Graph clustering. Computer Sci. Rev. 1 (1), 27–64.
Schwarz, N., 2010. Urban form revisited–selecting indicators for characterising european cities. Landscape Urban Plann. 96 (1), 29–47.
Shafizadeh-Moghadam, H., Asghari, A., Tayyebi, A., Taleai, M., 2017. Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. Comput. Environ. Urban Syst. 64, 297–308.
Shelton, J., Wagner, J., Samant, S., Goodin, G., Lomax, T., Seymour, E., 2019. Impacts of connected vehicles in a complex, congested urban freeway setting using multi-resolution modeling methods. Int. J. Transp. Sci. Technol. 8 (1), 25–34.
Shen, H., Li, T., Yuan, Q., Zhang, L., 2018. Estimating regional ground-level pm2. 5 directly from satellite top-of-atmosphere reflectance using deep belief networks. J. Geophys. Res.: Atmos. 123 (24), 13–875.
Sinta, D., Wijayanto, H., Sartono, B., 2014. Ensemble k-nearest neighbors method to predict rice price in indonesia. Appl. Math. Sci. 8 (160), 7993–8005.
Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14 (3), 199–222.
Subasi, A., 2020. Practical Machine Learning for Data Analysis Using Python. Academic Press.
Sun, Y., Gao, C., Li, J., Wang, R., Liu, J., 2019. Quantifying the effects of urban form on land surface temperature in subtropical high-density urban areas using machine learning. Remote Sensing 11 (8), 959.
Sutton, R.S., Barto, A.G., 2011. Reinforcement learning: An introduction.
Tabios III, G.Q., Salas, J.D., 1985. A comparative analysis of techniques for spatial interpolation of precipitation 1. JAWRA J. Am. Water Resources Assoc. 21 (3), 365–380.
Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., 2018. A survey on deep transfer learning. In: International conference on artificial neural networks. IEEE, pp. 270–279.
Tekouabou, S.C.K. et al., 2021. Intelligent management of bike sharing in smart cities using machine learning and internet of things. Sustainable Cities Society 67, 102702.
Tekouabou, S.C.K., Cherif, W., Silkan, H., et al., 2019. Prediction of bus waiting time in smart cities using machine learning techniques and the internet of things. In: In: International Conference on Advanced Intelligent Systems for Sustainable Development. Springer, pp. 352–362.
Tekouabou, S.C.K., Alaoui, E.A.A., Chabbar, I., Cherif, W., Silkan, H., 2020. Machine learning apprroach for early detection of glaucoma from visual fields. In: Proceedings of the 3rd International Conference on Networking, Information Systems & Security, pp. 1–5.
Tekouabou, S.C.K., Cherif, W., Silkan, H., et al., 2020. Improving parking availability prediction in smart cities with iot and ensemble-based model. Journal of King Saud University-Computer and Information Sciences.
Tekouabou, S.C.K., Hartini, S., Rustam, Z., Silkan, H., Agoujil, S., et al., 2021. Improvement in automated diagnosis of soft tissues tumors using machine learning. Big Data Mining and Analytics 4 (1), 33–46.
Thayse, A., Gribomont, P., Louis, G., Snyers, D., Wodon, P., Gochet, P., Grégoire, E., Sanchez, E., Delsarte, P., 1988. Approche logique de l'intelligence artificielle. Dunod Paris.
Toivonen, H. et al., 1996. Sampling large databases for association rules. Vldb. 96, 134–145.
Tran, D.X., Pla, F., Latorre-Carmona, P., Myint, S.W., Caetano, M., Kieu, H.V., 2017. Characterizing the relationship between land use land cover change and land surface temperature. ISPRS J. Photogrammetry Remote Sensing 124, 119–132.
Vapnik, V., 2013. The nature of statistical learning theory. Springer science & business media.

Vergouw, B., Nagel, H., Bondt, G., Custers, B., 2016. Drone technology: Types, payloads, applications, frequency spectrum issues and future developments. In: The future of drone use. Springer, pp. 21–45.
Verma, D., Jana, A., Ramamritham, K., 2019. Transfer learning approach to map urban slums using high and medium resolution satellite imagery. Habitat Int. 88, 101981.
Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E., 2018. Deep learning for computer vision: A brief review. In: Computational intelligence and neuroscience 2018.
Walpole, R.E., Myers, R.H., Myers, S.L., Ye, K., 1993. Probability and statistics for engineers and scientists, Vol. 5. Macmillan New York.
Wang, S.-C., 2003. Artificial neural network. In: Interdisciplinary computing in java programming. Springer, pp. 81–100.
Wang, S.-H., Xie, S., Chen, X., Guttery, D.S., Tang, C., Sun, J., Zhang, Y.-D., 2019. Alcoholism identification based on an alexnet transfer learning model. Frontiers in psychiatry 10.
Wang, R., Kim, J.-H., Li, M.-H., 2020. Predicting stream water quality under different urban development pattern scenarios with an interpretable machine learning approach. Sci. Total Environ. 144057.
Wei, Y.-Q., Yang, R.-H., Liu, P.-Y., 2009. An improved apriori algorithm for association rules of mining. In: 2009 IEEE International Symposium on IT in Medicine & Education. Vol. 1. IEEE, pp. 942–946. .
Williams, W., 1971. Principles of clustering. Annu. Rev. Ecol. Syst. 2 (1), 303–326.
Xayasouk, T., Lee, H., Lee, G., 2020. Air pollution prediction using long short-term memory (lstm) and deep autoencoder (dae) models. Sustainability 12 (6), 2570.
Xing, H., Meng, Y., 2020. Measuring urban landscapes for urban function classification using spatial metrics. Ecol. Ind. 108, 105722.
Xu, H., Wang, L., Gan, W., 2016. Application of improved decision tree method based on rough set in building smart medical analysis crm system. Int. J. Smart Home 10 (1), 251–266.
Xu, L., Cui, S., Tang, J., Nguyen, M., Liu, J., Zhao, Y., 2019. Assessing the adaptive capacity of urban form to climate stress: a case study on an urban heat island. Environ. Res. Lett. 14, (4) 044013.
Yang, C., Zhan, Q., Gao, S., Liu, H., 2020. Characterizing the spatial and temporal variation of the land surface temperature hotspots in wuhan from a local scale. Geo-spatial Inform. Sci. 23 (4), 327–340.
Yu, Z., Chen, S., Wong, N.H., Ignatius, M., Deng, J., He, Y., Hii, D.J.C., 2020. Dependence between urban morphology and outdoor air temperature: A tropical campus study using random forests algorithm. Sustainable Cities Society 61, 102200.
Zhang, Y., Haghani, A., 2015. A gradient boosting method to improve travel time prediction. Transp. Res. Part C: Emerging Technologies 58, 308–324.
Zhang, S., Li, X., Zong, M., Zhu, X., Wang, R., 2017. Efficient knn classification with different numbers of nearest neighbors. IEEE Trans. Neural Networks Learning Syst. 29 (5), 1774–1785.
Zhang, H., Goodfellow, I., Metaxas, D., Odena, A., 2019. Self-attention generative adversarial networks. In: International conference on machine learning. PMLR, pp. 7354–7363.