



## **La classification et les bâtiments type au service de la rénovation énergétique des parcs immobiliers**

Sergeï Agapoff, Mathieu Rivallain, Yunseok Lee

### **► To cite this version:**

Sergeï Agapoff, Mathieu Rivallain, Yunseok Lee. La classification et les bâtiments type au service de la rénovation énergétique des parcs immobiliers. Conférence IBPSA France, May 2018, Bordeaux, France. <hal-01795050>

**HAL Id: hal-01795050**

**<https://hal.science/hal-01795050v1>**

Submitted on 18 May 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## La classification et les bâtiments type au service de la rénovation énergétique des parcs immobiliers

Sergeï Agapoff\*<sup>1</sup>, Mathieu Rivallain<sup>1</sup>, Yunseok Lee<sup>1</sup>

<sup>1</sup>CSTB (Centre Scientifique et Technique du Bâtiment)

84 avenue Jean Jaurès – Champs-sur-Marne

77447 Marne-la-Vallée cedex,

\*[sergei.agapoff@cstb.fr](mailto:sergei.agapoff@cstb.fr)

---

*RESUME. Dans le cadre de la rénovation énergétique des bâtiments, les bailleurs sociaux, les gestionnaires de parc, les pouvoirs publics territoriaux doivent pouvoir disposer d'une photographie de leur parc immobilier pour prendre des décisions quant au renouvellement des systèmes de chauffage, des ouvrants ou encore de l'isolation des bâtiments. Dans cet article, nous proposons une méthodologie basée sur l'apprentissage supervisé pour compléter la carte d'identité des bâtiments d'un tel parc immobilier. Cette utilisation de ce type d'algorithmes, novatrice dans le domaine du bâtiment, constitue une alternative moins coûteuse mais aussi moins précise aux audits énergétiques réalisés sur le terrain. Les travaux de recherche sont toujours en cours de réalisation et une procédure d'évaluation des performances de la méthode sur le parc de l'Office Public de l'Habitat Montreuillois (Montreuil, 93) est développée.*

*MOTS-CLÉS : rénovation énergétique, bâtiment type, classification.*

---

---

*ABSTRACT. When it comes to building energy retrofit, it becomes necessary for a building stock manager, social housing company or local administration to have a clear picture of their building stock to make the right decisions for HVAC systems, windows or insulation renewal. In this paper, we propose a methodology based on supervised machine-learning algorithms to complete the identity card of such buildings. This innovative use of such algorithms – for buildings – is an alternative, less expensive but also less precise, to building energy audits. This is on-going research and a performance assessment procedure with data from "Office Public de l'Habitat Montreuillois" (Montreuil, France) is exposed.*

*KEYWORDS : energy retrofit, typical buildings, classification.*

---

### 1. INTRODUCTION

Dans le cadre de la rénovation énergétique des bâtiments, les bailleurs sociaux, les gestionnaires de parc, les pouvoirs publics territoriaux doivent pouvoir disposer d'une photographie de leur parc immobilier pour prendre des décisions quant au renouvellement des systèmes de chauffage, des ouvrants ou encore de l'isolation des bâtiments. Cette photographie est habituellement réalisée sur la base des connaissances expertes sur les bâtiments du parc ou, lorsque ces informations ne sont pas suffisantes, sur la base d'audits qui permettent de dresser la carte d'identité de chacun des bâtiments du bailleur. De telles informations sont précieuses pour réaliser des simulations numériques des différents scénarios envisagés afin d'appuyer le processus de décision. Cependant, cette collecte de données par mobilisation d'experts ou audit énergétique est coûteuse tant sur le plan économique que sur le plan temporel. Des outils numériques, certes moins précis mais surtout moins chers et plus rapides à mettre en œuvre, sont alors nécessaires pour permettre à un bailleur – ou tout autre maître d'ouvrage – d'utiliser les

connaissances minimales qu'il détient sur ses bâtiments (surfaces, usages...) afin de compléter leur carte d'identité.

Pour répondre à ce besoin, nous proposons une méthodologie d'identification des bâtiments réels (partiellement caractérisés) à des bâtiments types (totalement caractérisés). Cette identification permet de caractériser les éléments inconnus des bâtiments réels afin d'obtenir une représentation des bâtiments étudiés comme sur la Figure 1 qui présente le processus général dans lequel s'inscrit la méthodologie d'identification. Une fois le bâtiment type identifié, il est nécessaire de développer des méthodes pour représenter le bâtiment et fournir des informations pertinentes pour la rénovation énergétique. La méthodologie proposée se base sur des algorithmes de classification pour identifier le bâtiment type associé à un bâtiment réel et la caractérisation des bâtiments réels est complétée grâce aux informations des bâtiments type.



Figure 1 : Processus général de représentation des bâtiments d'un parc

L'utilisation de méthodes d'apprentissage supervisé dans le domaine du bâtiment étant peu répandue, nous proposons en premier lieu un état l'art des approches par classification. La méthodologie proposée est ensuite détaillée avant de décrire la procédure de d'évaluation des performances sur le cas de l'Office Public de l'Habitat de Montreuil (93).

## 2. ETAT DE L'ART DES APPROCHES PAR CLASSIFICATION

Le concept de classification recouvre un ensemble de méthodes de *supervised learning* (apprentissage supervisé d'une structure de données par un outil informatique) : ces méthodes exploitent une loi de correspondance entre entrées (objets) et sorties (classes) obtenue à partir d'objets pour lesquels la classe est connue. La construction de cette loi de correspondance est communément appelée « étape d'entraînement » dans la description d'algorithmes de classification. Les méthodes les plus répandues sont les suivantes (Lee 2015) :

- **Decision Tree (DT)** : la classe d'un objet est définie par partitions successives selon la valeur de ses attributs (qui peuvent être catégoriques et numériques). La visualisation de l'arbre de décision permet une approche simple et intuitive de la loi de classification ;
- **k-Nearest Neighbors (k-NN)** : la classe d'un objet est déterminée par vote majoritaire des k objets d'apprentissage les plus proches. Puisqu'une distance doit être définie, ce type de méthode ne fonctionne qu'avec des attributs numériques ;
- **Artificial Neural Networks (ANN)** : la classe d'un objet est définie via une « boîte noire » composée de plusieurs couches de noeuds (neurones) qui traitent les données grâce à des fonctions d'activation. Ce type d'algorithme permet de fournir des lois de classification robustes (même pour des jeux de données complexes et volumineux) mais est difficilement interprétable du fait de son caractère boîte noire ;
- **Support Vector Machine (SVM)** : ce type d'algorithme permet de résoudre les problèmes de classification binaires en cherchant la frontière entre les deux classes considérées. Cette

frontière est choisie comme celle qui a la distance la plus grande aux objets les plus proches (marge maximale). La loi résultante d'une telle classification est difficile à interpréter lorsqu'elle n'est pas linéaire.

Dans la littérature étudiée, la méthode SVM est souvent utilisée et lorsqu'elle est comparée à d'autres méthodes (DT et Random Forest dans (Z. Lu et al. 2014) ou Linear Discriminant Analysis et Naïve Bayes dans (J. Lu, Mamun, and Chau 2015)), cette méthode semble donner de meilleurs résultats. Les outils les plus populaires pour l'implémentation de ces méthodes sont RapidMiner, MS-EXCEL et KNIME pour les outils dédiés et R, SQL et Python pour les langages de programmation. Parmi ces outils, R et Python sont de plus en plus utilisés par la communauté scientifique du fait de leur facilité de prise en main et leur adaptabilité. Les langages R et Python disposent de modules dédiés au *machine learning* (caret pour R (Kuhn 2008) et scikit-learn pour Python (Pedregosa et al. 2011)) : la force de ces modules est d'avoir un seul format de données utilisables avec toutes les méthodes disponibles.

Les méthodes de classification sont largement employées dans des applications de traitement du signal pour automatiser la détection et l'identification de motifs dans :

- une image : photographie aérienne/satellite (Diwediga et al. 2015; Z. Lu et al. 2014; Belgiu and Drăguț 2014; Kluckner and Bischof 2010; Matikainen, Hyppä, and Hyppä 2003), photographie 3D urbaine (Serna and Marcotegui 2014; Hernandez 2008), trafic routier (Zangenehpour, Miranda-Moreno, and Saunier 2015), analyse chimique (Milanez and Pontes 2014) ;
- un son : analyse chimique (Morizet et al. 2016) ;
- une impulsion électrique : analyse médicale (Djemili, Bourouba, and Amara Korba 2015), interface cerveau-machine (J. Lu, Mamun, and Chau 2015) ;
- un déplacement : aide à la conduite (Boyalı, Hashimoto, and Matsumoto 2015), diagnostic mécanique (Fedala et al. 2015; Li et al. 2015).

Le traitement du signal nécessite l'extraction de données pour caractériser le signal (*features extraction*) avant de procéder à la classification, tandis que dans d'autres domaines le point de départ peut être une collection d'objets déjà caractérisés (état d'une exploitation minière (Zhou et al. 2013)).

Lorsque des attributs nécessaires à la classification ne sont pas disponibles pour les objets étudiés, il est nécessaire d'adapter la méthodologie (Saar-Tsechansky and Provost 2007) : retirer les objets partiellement décrits, obtenir les valeurs manquantes (faire une campagne de mesures, remplacer un capteur défectueux ...), estimer les valeurs manquantes en fonction d'objets connus (valeur la plus probable, valeur par défaut, distribution probable de valeurs ...) ou encore recalibrer la loi de classification pour ne prendre en compte que les attributs connus pour l'objet étudié (modèles réduits).

Dans la littérature étudiée, les applications dans le domaine du bâtiment (Z. Lu et al. 2014; Belgiu and Drăguț 2014; Kluckner and Bischof 2010; Matikainen, Hyppä, and Hyppä 2003; Serna and Marcotegui 2014; Hernandez 2008) se focalisent sur l'analyse d'image pour détecter et identifier des bâtiments dans une zone donnée. De plus, les méthodes proposées ne permettent en général que la distinction entre les objets « bâtiment » et les autres objets « non-bâtiment » (végétation, voiture, lampadaire) : seuls Lu et al. proposent une distinction entre les bâtiments détectés (Z. Lu et al. 2014).

Une dernière approche qui ne fait pas intervenir de phase d'entraînement, mais peut être assimilée à de la classification, nous paraît importante à mentionner : l'association d'un objet à une classe par minimisation de la distance entre l'objet et un objet fictif représentant la classe. S'il est possible de définir un objet représentatif de chaque classe par des attributs représentatifs de la classe, alors la minimisation de la distance entre les attributs de l'objet étudié et ceux du représentant de la classe constitue une forme de classification.

### 3. METHODOLOGIE D'IDENTIFICATION

#### 3.1. BATIMENTS TYPES, CLASSES ET LOIS DE CLASSIFICATION ASSOCIEES

L'objectif de la méthodologie proposée est d'identifier un bâtiment réel, dont on ne connaît que quelques attributs (surface, nombre d'étages, année de construction, localisation...), à un bâtiment type qui possède des caractéristiques techniques plus précises (nature des parois, taux de vitrage, type de systèmes de chauffage, ventilation, climatisation...). Le sujet des « bâtiments type » est l'un des plus vastes sujets de la recherche sur la caractérisation des bâtiments (Fery and Lahrech 2010) : il existe quantité de bibliothèques de bâtiments types pour différents périmètres et de méthodes pour les définir. Pour la méthodologie que nous proposons nous faisons l'hypothèse qu'une bibliothèque de bâtiments type a été identifiée pour le périmètre de bâtiments étudié. Nous inscrivons notre démarche dans le périmètre des bâtiments résidentiels existants et neufs.

Dans notre méthodologie, chaque bâtiment type constitue une classe de bâtiment : dire qu'un bâtiment réel  $B_x$  est de type  $T_i$  signifie que ce bâtiment peut être assimilé au bâtiment type  $i$  et donc qu'il appartient à la classe  $i$ . Les concepts de classe (au sens du *machine learning*) et de type (au sens de la caractérisation des bâtiments) sont donc confondus dans notre méthodologie. L'assimilation d'un bâtiment réel à bâtiment type est le processus de classification, qu'il s'agisse de classification naturelle lorsqu'il n'y a pas besoin d'intermédiaire pour déterminer la classe, de classification experte lorsqu'elle est réalisée par un humain ou de classification par apprentissage lorsqu'elle est réalisée par un algorithme. Pour que la classification par apprentissage soit efficace il est nécessaire de mettre en évidence des fonctions ou lois qui permettent d'analyser les caractéristiques connues du bâtiment étudié et de déterminer la classe de bâtiment dans lequel le ranger. Dans notre cas, la construction de ces lois, ou « étape d'entraînement », résulte de la classification experte de bâtiments bien caractérisés. Ces bâtiments constituent ce que nous appelons un parc « référence » : c'est par rapport à ce parc référence et les bâtiments types sélectionnés que les lois de classification vont être construites. L'ensemble des attributs retenus pour entraîner l'algorithme de classification est déterminant dans la construction de la loi de classification, modifier cet ensemble résultera en une nouvelle loi de classification. Notons enfin que chaque loi de classification est naturellement liée à un algorithme de classification donné (DT, SVM, ANN, etc.)

#### 3.2. PRINCIPE

La méthodologie que nous proposons se compose de 4 étapes principales résumée dans la Figure 2.

**1 : Construction d'une loi de classification****1a : Choix des bâtiments type, du parc référence, de l'algorithme de classification et des attributs de classification****1b : Association experte entre les bâtiments référence et les bâtiments type****1c : Entraînement de l'algorithme de classification sélectionné****2 : Sélection des bâtiments réels à identifier****3 : Application de la loi de classification sur les attributs choisis à l'étape 1****4 : Pour chaque bâtiment réel, identification du bâtiment type associé grâce à la classe obtenue à l'étape 3***Figure 2 : Méthodologie pour l'identification des bâtiments*

La première étape permet de traduire la connaissance experte sur l'association de bâtiments réels et de bâtiments type au travers de l'apprentissage automatique. Une fois la loi de classification construite, l'étape de sélection des bâtiments réels à identifier doit mettre en évidence, parmi l'ensemble des bâtiments étudiés, ceux qui sont compatibles avec la loi construite : il est impératif que les bâtiments sélectionnés soient au moins caractérisés par les attributs choisis lors de l'étape 1. Si un attribut est manquant, le bâtiment en question ne peut être sélectionné à moins de mettre en œuvre une technique d'adaptation telles que celles proposées par Saar-Tsechansky et Provost. Enfin les étapes 3 et 4 constituent le cœur de la méthodologie à savoir la classification et l'association entre la classe obtenue et le bâtiment type. Une fois le bâtiment type obtenu, les caractéristiques manquantes pour le bâtiment étudiées sont assimilées à celles du bâtiment type.

**4. EVALUATION DES PERFORMANCES****4.1. PROCEDURE**

Pour évaluer la qualité de la méthodologie d'identification des bâtiments, nous proposons de réunir un panel d'experts dont le rôle est d'associer les bâtiments type aux bâtiments réels pour confronter les résultats de la méthodologie à ceux d'une classification experte. Les identifications effectuées par le panel d'expert permettent de rendre compte de la variabilité d'interprétation des caractéristiques des bâtiments. Pour garantir la validité de la procédure, les experts mobilisés pour effectuer l'identification manuelle sont différents de celui ou ceux mobilisés pour l'initialisation de la méthodologie (construction de la loi de classification). Pour permettre une évaluation de la chaîne complète, les attributs de classification doivent être choisis dans un sous-ensemble des attributs connus pour le parc de bâtiments réels étudié. Ainsi, il est possible de comparer les caractéristiques obtenues après identification des bâtiments type et les caractéristiques réelles des bâtiments étudiés.

Ainsi, pour un ensemble de bâtiments réels, les performances suivantes seront calculées :

- $A_b$  : accord algorithme-expert pour chaque bâtiment  $b$  étudié. Il s'agit du pourcentage d'experts ayant classé le bâtiment de la même manière que l'algorithme ;
- $P_b^2$  : précision de la représentation de chaque bâtiment  $b$ . Il s'agit de la somme des carrés des écarts entre les caractéristiques réelles et les caractéristiques estimées.

- $\tau_{alg}$  : temps moyen d'identification par l'algorithme. Il s'agit du temps nécessaire à l'algorithme pour identifier un bâtiment (étapes 3 et 4 de la méthodologie) moyenné sur l'ensemble des bâtiments étudiés ;
- $\tau_{exp}$  : temps moyen d'identification par les experts. Il s'agit du temps nécessaire à un expert pour identifier un bâtiment (étapes 3 et 4 de la méthodologie) moyenné sur l'ensemble des bâtiments étudiés et des experts mobilisés.

#### 4.2. CAS D'ETUDE ENVISAGE

L'Office Public de l'Habitat Montreuillois est le premier propriétaire et gestionnaire de logements sur la ville de Montreuil et le troisième Office Public de l'Habitat en Seine-Saint-Denis (93). Le parc de bâtiments de ce gestionnaire est constitué de plus de 11000 logements présentant des caractéristiques variées. Cette structure dispose de données techniques exhaustives sur ses bâtiments ce qui fait de ce parc un bon terrain d'expérimentation pour la méthodologie proposée.

Un sous-ensemble de ce parc immobilier sera choisi pour réaliser l'évaluation des performances de la méthodologie. Entre 10 et 20 bâtiments seront sélectionnés selon leur caractérisation : nous nous focaliserons sur des bâtiments pour lesquels un grand nombre d'attributs sont caractérisés.

### 5. CONCLUSIONS ET PERSPECTIVES

Dans cet article nous présentons une méthodologie d'identification des bâtiments réels sur la base de classification pour les associer à des bâtiments type. Cette méthodologie représente une alternative à des méthodes plus longues et plus coûteuses de caractérisation des bâtiments telles que l'audit énergétique. La qualité de notre méthodologie n'est pas encore démontrée et nous proposons une procédure d'évaluation de ses performances. Les performances seront évaluées sur une partie du parc de l'OPH Montreuil.

Dans la suite de ces travaux, nous envisageons de proposer une méthodologie avancée pour l'identification des bâtiments. Sur la Figure 2, il est évident que l'étape 1 peut être commune à plusieurs itérations de la méthodologie. En fixant la bibliothèque de bâtiments type et le parc référence, il est possible de créer une base de lois de classification pour différents algorithmes et différents ensembles d'attributs. Sur le principe des « modèles réduits » exposé par Saar-Tsechansky et Provost, le fait d'envisager des lois de classification adaptées à différents ensembles d'attributs permet d'adapter l'identification aux seules caractéristiques connues sur les bâtiments étudiés. Cette base de lois de classification nous permettra donc de proposer une méthodologie avancée (Figure 3) qui s'adapte aux bâtiments étudiés et permet à l'utilisateur de choisir l'algorithme le plus adapté à son cas d'étude.

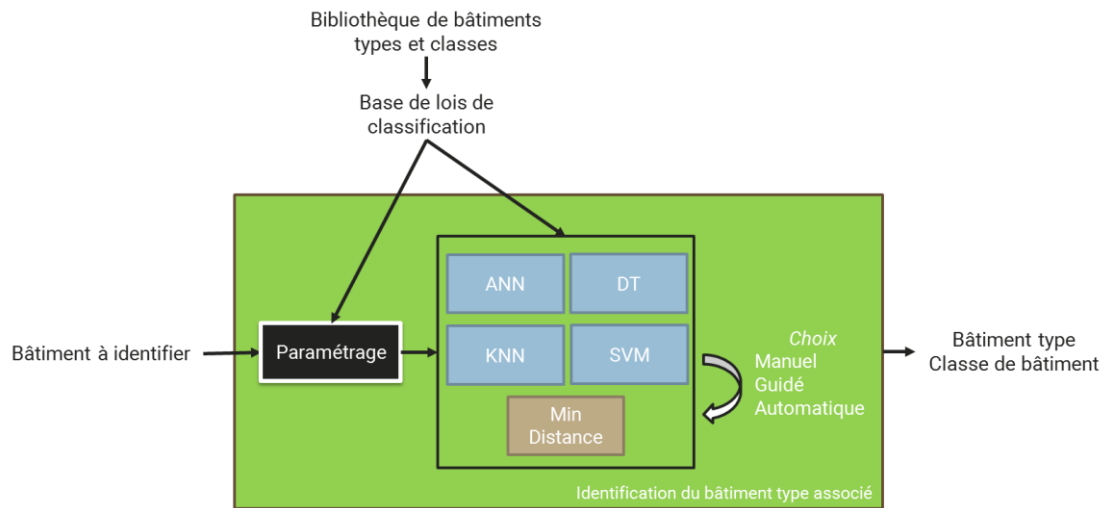


Figure 3 : Méthodologie avancée pour l'identification des bâtiments

La phase de « Paramétrage » proposée dans cette méthodologie avancée correspond à l'analyse des attributs connus pour le bâtiment à identifier afin de sélectionner les lois de classification applicables selon les attributs auxquelles elles sont associées. Cette méthodologie avancée permettra aussi de choisir l'algorithme de classification utilisé selon les performances observées lors de la construction de la base de lois de classification.

## 6. REMERCIEMENTS

Ces travaux ont bénéficié d'une aide de l'Agence Nationale de la Recherche (projet REHA-PARCS, ANR-15-CE22-0011).

## 7. BIBLIOGRAPHIE

- Belgiu, Mariana, and Lucian Drăguț. 2014. 'Comparing Supervised and Unsupervised Multiresolution Segmentation Approaches for Extracting Buildings from Very High Resolution Imagery'. *ISPRS Journal of Photogrammetry and Remote Sensing* 96 (October): 67–75. <https://doi.org/10.1016/j.isprsjprs.2014.07.002>.
- Boyali, Ali, Naohisa Hashimoto, and Osamu Matsumoto. 2015. 'A Signal Pattern Recognition Approach for Mobile Devices and Its Application to Braking State Classification on Robotic Mobility Devices'. *Robotics and Autonomous Systems* 72 (October): 37–47. <https://doi.org/10.1016/j.robot.2015.04.008>.
- Diwediga, Badabate, Kperkouma Wala, Fousseni Folega, Marra Dourma, Yao A. Woegan, Koffi Akpagana, and Quang Bao Le. 2015. 'Biophysical and Anthropogenous Determinants of Landscape Patterns and Degradation of Plant Communities in Mo Hilly Basin (Togo)'. *Ecological Engineering* 85 (December): 132–43. <https://doi.org/10.1016/j.ecoleng.2015.09.059>.
- Djemili, Rafik, Hocine Bourouba, and M. C. Amara Korba. 2015. 'Application of Empirical Mode Decomposition and Artificial Neural Network for the Classification of Normal and Epileptic EEG Signals'. *Biocybernetics and Biomedical Engineering*. <https://doi.org/10.1016/j.bbe.2015.10.006>.
- Fedala, S., D. Rémond, R. Zegadi, and A. Felkaoui. 2015. 'Contribution of Angular Measurements to Intelligent Gear Faults Diagnosis'.
- Fery, Augustine, and Rofaïda Lahrech. 2010. 'Base de Connaissances Batiments Existants, Recensement Batiments Types - Rapport Final'. ESE/DE/PEB-2010.018RR. CSTB.



- Hernandez, Jorge. 2008. 'Segmentation et Interprétation de Nuages de Points Pour La Modélisation d'Environnements Urbains'. *Revue Française de Photogrammetrie et de Télédétection* 191: 28–35.
- Kluckner, Stefan, and Horst Bischof. 2010. 'Image-Based Building Classification and 3d Modeling with Super-Pixels'. In *ISPRS Technical Commission III Symposium on Photogrammetry Computer Vision and Image Analysis*, 233–238.  
<http://www.isprs.org/proceedings/XXXVIII/part3/a/pdf/zwi/41-XXXVIII-part3A.pdf>.
- Kuhn, Max. 2008. 'Building Predictive Models in R Using the Caret Package'. *Journal of Statistical Software* 28 (5): 1–26.
- Lee, Yunseok. 2015. 'Modeling Segmentation of Building Stocks for Energy Management and Retrofit'. Internship Report. Ecole des Ponts - CSTB.
- Li, Chuan, René-Vinicio Sanchez, Grover Zurita, Mariela Cerrada, Diego Cabrera, and Rafael E. Vásquez. 2015. 'Multimodal Deep Support Vector Classification with Homologous Features and Its Application to Gearbox Fault Diagnosis'. *Neurocomputing* 168 (November): 119–27.  
<https://doi.org/10.1016/j.neucom.2015.06.008>.
- Lu, J., K.A. Mamun, and T. Chau. 2015. 'Pattern Classification to Optimize the Performance of Transcranial Doppler Ultrasonography-Based Brain Machine Interface'. *Pattern Recognition Letters* 66: 135–43. <https://doi.org/10.1016/j.patrec.2015.07.020>.
- Lu, Zhenyu, Jungho Im, Jinyoung Rhee, and Michael Hodgson. 2014. 'Building Type Classification Using Spatial and Landscape Attributes Derived from LiDAR Remote Sensing Data'. *Landscape and Urban Planning* 130 (October): 134–48.  
<https://doi.org/10.1016/j.landurbplan.2014.07.005>.
- Matikainen, Leena, Juha Hyypä, and Hannu Hyypä. 2003. 'Automatic Detection of Buildings from Laser Scanner Data for Map Updating'. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 34 (3/W13): 218–224.
- Milanez, Karla Danielle Tavares Melo, and Márcio José Coelho Pontes. 2014. 'Classification of Edible Vegetable Oil Using Digital Image and Pattern Recognition Techniques'. *Microchemical Journal* 113 (March): 10–16. <https://doi.org/10.1016/j.microc.2013.10.011>.
- Morizet, N., N. Godin, J. Tang, E. Maillet, M. Fregonese, and B. Normand. 2016. 'Classification of Acoustic Emission Signals Using Wavelets and Random Forests : Application to Localized Corrosion'. *Mechanical Systems and Signal Processing* 70–71 (March): 1026–37.  
<https://doi.org/10.1016/j.ymssp.2015.09.025>.
- Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. 'Scikit-Learn: Machine Learning in Python'. *The Journal of Machine Learning Research* 12: 2825–2830.
- Saar-Tschekansky, Maytal, and Foster Provost. 2007. 'Handling Missing Values When Applying Classification Models'. <http://archive.nyu.edu/handle/2451/27813>.
- Serna, Andrés, and Beatriz Marcotegui. 2014. 'Detection, Segmentation and Classification of 3D Urban Objects Using Mathematical Morphology and Supervised Learning'. *ISPRS Journal of Photogrammetry and Remote Sensing* 93: 243–255.
- Zangenehpour, Sohail, Luis F. Miranda-Moreno, and Nicolas Saunier. 2015. 'Automated Classification Based on Video Data at Intersections with Heavy Pedestrian and Bicycle Traffic: Methodology and Application'. *Transportation Research Part C: Emerging Technologies* 56 (July): 161–76. <https://doi.org/10.1016/j.trc.2015.04.003>.
- Zhou, Jian, Xibing Li, Mitri Hani S., Shiming Wang, and Wei Wei. 2013. 'Identification of Large-Scale Goaf Instability in Underground Mine Using Particle Swarm Optimization and Support Vector Machine'. *International Journal of Mining Science and Technology* 23 (5): 701–7.  
<https://doi.org/10.1016/j.ijmst.2013.08.014>.