# Early Insights into Location-Allocation Decision-Making using Ensemble Learning

Lars De Sloover*, Roel Gevaers**, Wouter Dewulf**, Joris Beckers**, Frank Van Gelder**, Haosheng Huang*, and Nico Van de Weghe*

* Department of Geography, Ghent University, Ghent, Belgium
** Department of Transport and Regional Economics, University of Antwerp, Antwerp, Belgium

**Abstract.** The COVID-19 pandemic underscored the significance of geography in health emergencies, spotlighting the need for optimized spatial decision-making. This paper introduces a novel, data-driven approach to spatial decision-making, leveraging gradient boosting to derive data-informed weights for a Weighted Linear Combination (WLC). The goal is to pinpoint optimal locations for vaccination centres in Flanders, Belgium. Drawing from prior work, we present a foundation for the required number of centres, and then focus on determining the most suitable locations within Flanders. Utilizing a dataset of 91 centres, our ensemble learning technique dynamically determines criteria weights. Criteria that are socio-demographic or mobility oriented are considered. Our methodology, termed Ensemble Analysis for Criteria Trade-offs (ENACT), offers a comprehensive framework, targeting dynamic location-allocation scenarios. Using derived weights, we identify regions with the highest suitability scores for vaccination center placement. The high model performance metrics underline its reliability, with caution on potential overfitting. The study comes with a roadmap for enhancing the methodology's comprehensiveness in future research, suggesting the integration of more criteria and GIS optimization techniques for actionable health infrastructure planning.

**Keywords.** Multi-Criteria Decision Analysis, Gradient Boosting, COVID-19 Vaccination Strategy

## 1. Introduction

The COVID-19 pandemic has emphasized the role of geography in health crises. As the virus spread and responses varied, the importance of geo-

Published in "Proceedings of the 18th International Conference on Location Based Services (LBS 2023)", edited by Haosheng Huang, Nico Van de Weghe and Georg Gartner, LBS 2023, 20-22 November 2023 Ghent, Belgium.

This contribution underwent single-blind peer review based on the paper.
https://doi.org/10.34726/5724 

graphical information science and systems (GIS) became clear (Higgs 2004, Li 2020). Historically used for healthcare planning (Lyseen et al. 2014), the unique challenges of the pandemic spotlighted the limits of traditional spatial decision-making(Parvin *et al.* 2021).

Traditionally, multi-criteria decision analysis (MCDA) has been favoured for spatial decision-making, relying on expert-informed criterion weightings. In contrast, our paper introduces an adaptive approach using the gradient boosting ensemble method. This innovative technique dynamically assigns weights to criteria in a Weighted Linear Combination (WLC), providing data-driven insights for spatial trade-offs. We apply this method to Flanders' vaccination strategy, pinpointing optimal vaccination centre locations based on these data-driven weights.

In the context of Flanders, a region of Belgium hard-hit by COVID-19, this paper builds on two central themes. First, we draw from prior research that determined the optimal number of vaccination centres based on the interplay between vaccine distribution and mobility costs. While these insights set a strategic foundation for the number of centres needed, the specifics of this research are detailed elsewhere (Beckers *et al.* 2021). Our current focus then shifts to spatial modelling, aiming to pinpoint the most suitable locations for these centres within Flanders, given its unique demographic and geographic features.

In response to the pandemic's urgency, the Flemish government swiftly set up 91 vaccination centres throughout Flanders, a number which notably exceeds our previously suggested optimal count in (Beckers *et al.* 2021). Nevertheless, we harness the spatial distribution of these 91 centres, using them as a foundational dataset for our ensemble model. By applying gradient boosting—a renowned supervised ensemble learning method in site suitability analysis (Sahin 2020, Yin *et al.* 2020, Wei *et al.* 2023) —we extract criteria weights for a Weighted Linear Combination (WLC). This approach yields empirically-supported weights for each criterium, offering a more data-driven decision-making framework.

Specific criteria were determined, based on literature (Alemdar *et al.* 2021, Guida and Carpentieri 2021, Song *et al.* 2022) and input from stakeholders, knowledgeable about spatial planning, mobility, and vaccination policy. These were either socio-demographic in nature or mobility-oriented. They include mean population age, node value public transport (Vlaamse Overheid - Departement Omgeving - Afdeling Vlaams Planbureau voor

Omgeving 2021), population density, road proximity, and hospital accessibility[1].

Positioned to offer a robust, adaptable, and comprehensive framework, we present our idea as the Ensemble Analysis for Criteria Trade-offs (ENACT). This method targets the intricacies of location-allocation decision-making in dynamic scenarios. The COVID-19 pandemic, with its vast logistical challenges and the imperative for effective vaccine distribution, presents an ideal backdrop for the implementation of such a methodology. This data-driven approach retroactively tries to depict a nuanced image of Flanders' vaccination strategy, aiming to ascertain the optimal spatial positioning of vaccination centres—retaining only the most strategically located—based on learned criteria weights.

The sections that follow delve into the methodology. In the preliminary results, we aim to illuminate how a gradient boosting ensemble model, yielding dynamically estimated weights for WLC, offers nuanced solutions for location-allocation dilemmas. Our discussion culminates with an assessment of the first outcomes, an exploration of potential pitfalls, and insights into the future potential and evolution of the methodology.

## 2. Methodology

To construct a comprehensive suitability model for the optimal placement of vaccination centres in Flanders, we acquired a collection of socio-demographic and mobility data. All of these criteria were normalised on a scale from zero to one, and processed across Flanders at a spatial resolution of 1x1 kilometres, resulting in a series of raster grids. In parallel, a binary dataset was developed to indicate the presence or absence of existing vaccination centres, which served as our labelled training data for the ensuing ensemble learning phase.

The main objective of this research is to amalgamate the criteria grids in a suitability raster, $S$, using the Weighted Linear Combination (WLC) method. The difficulty is determining the best weights for each grid in this combination. Formally, our targeted suitability score, $S$, for any given cell in the resulting raster map is articulated as $S = \sum_i w_i \times x_i$ where $w_i$ represents the weights attributed to each factor and $x_i$ denotes the scores (values) of the cells within the criteria grids.

---

[1] In the rare case a severe allergic anaphylactic reaction would occur after vaccination (McNeil and DeStefano 2018), it is important that the patient can be taken to the nearest hospital as soon as possible.

The methodology section details our ensemble learning technique using gradient boosting, with current vaccination centres in Flanders as the labelled training data. This approach refines criteria weights for optimal vaccination center placement. In gradient boosting, decision trees iteratively learn, with each tree building upon the errors of its predecessor. For example, while the first tree might focus on population density, subsequent trees might consider factors like travel time to the nearest hospital. After all trees offer their insights, their decisions are combined, with higher weights given to more predictive factors. The resultant feature importances indicate each factor's contribution to the model's prediction capability.

Next, the feature importances derived from gradient boosting are translated into weights for each criterium. This fusion of data-driven insights with the WLC method provides a more empirical and precise avenue for synthesizing decisions.

Beyond the determination of the criterium weights, the gradient boosting algorithm can several other insights regarding model performance. Metrics such as accuracy, and F1-score (a harmonic mean of the precision and recall) were adopted to provide a quantitative understanding of the preliminary model's prediction capability.

The code implementing the ENACT methodology is publicly available and can be accessed at https://github.ugent.be/cartogis/ENACT.

## 3. Preliminary Results

Using the gradient boosting technique, weights were derived for each of our criteria. These weights provide insight into the relative importance of each criterium in determining suitable locations for vaccination centres. The computed criterium weights are: mean population age: 0.096, node value collective transport: 0.203, population density: 0.215, road proximity: 0.440, and hospital accessibility: 0.046.

A composite suitability map was generated by employing the WLC method, utilizing the derived criterium weights. This map reveals regions in Flanders with the highest suitability scores, indicating optimal areas for vaccination centre placement. The visual representation of the composite map can be observed in Figure 1.

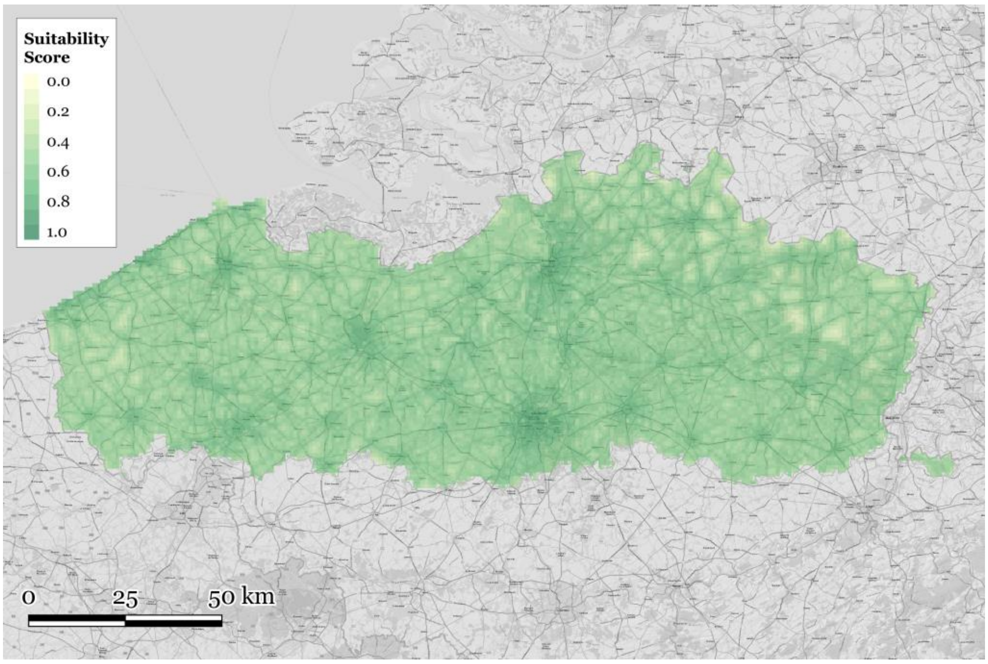


**Figure 1.** Composite map of the normalized suitability score, indicating suitability for vaccination centre placement.

Ultimately, the model performance metrics were calculated, yielding an accuracy of 0.97 and an F1-score of 0.97.

## 4. Discussion

The methodology employed in our study leveraged gradient boosting to derive weights for various criteria, illuminating the relative importance of each factor in the identification of optimal locations for vaccination centres in Flanders. The criteria weights suggest a pronounced importance of road proximity (0.440), which underscores the significance of accessibility via primary transportation networks. Socio-demographic factors like mean population age (0.096) and population density (0.215) also played vital roles, alongside mobility-driven metrics like node value for collective transport (0.203). However, it is notable that hospital accessibility, a seemingly crucial factor, carried a lesser weight (0.046). This counter-intuitive result underlines the importance of using data-driven models to inform our decision-making process, as they can often uncover non-obvious relationships in the data.

Our composite suitability map, constructed using the WLC method and our derived criteria weights, visually encapsulates the regions in Flanders that are most amenable for vaccination center deployment. This visualization serves as a foundational layer in spatial decision-making processes.

The model's accuracy and F1-score, both standing at 0.97, highlight its effectiveness in classifying suitable areas. However, these high-performance metrics also raise potential concerns about overfitting. A model that fits training data too closely might not generalize well to new, unseen data. This underscores the importance of model validation, hyperparameter tuning, and other regularization techniques to ensure a balance between bias and variance and prevent overfitting.

Several considerations can be incorporated in future iterations of this research to enhance its comprehensiveness and applicability. The first one is the integration of hard criteria. Our model is driven by soft criteria focusing on mobility and socio-demographics. Incorporating environmental constraints as hard criteria could be pivotal. For instance, regions with environmental protections might be strictly unsuitable for infrastructural developments, irrespective of their scores on soft criteria. Such constraints can also assist in undersampling the majority class – in this case, all the "0s" in the labelled dataset dominate and can overshadow the minority "1s", potentially leading to model biases. Addressing this imbalance can result in a more balanced dataset for model training. A second consideration is expanding the criteria. A more nuanced picture might emerge by integrating additional socio-demographic and mobility factors such as car ownership, disability rates, poverty levels, etc. These variables can enrich the model by accounting for more segments of the population, ensuring inclusivity in the decision-making process. A third consideration covers the transition from suitability to allocation. The current methodology yields a suitability map, which, while invaluable, doesn't directly translate to optimal spatial location-allocation of the centres. In future work, leveraging this suitability map as an input to optimization techniques like the p-median problem can bridge this gap. Such an approach would meld traditional GIScientific techniques with supervised learning models, producing robust and comprehensive spatial solutions.

In conclusion, our study lays the groundwork for a data-driven, GIS-enabled approach to vaccination center placement in Flanders. By integrating further criteria, optimization techniques, and refining the modelling process, future research can provide actionable insights for health logistics and infrastructure deployment.

## References

Alemdar, K.D., Kaya, Ö., Çodur, M.Y., and Campisi, T., 2021. Accessibility of Vaccination Centers in COVID-19 Outbreak Control : A GIS-Based Multi-Criteria Decision Making Approach. *ISPRS International Journal of Geo-Information*, 10 (708).

Beckers, J., Dewulf, W., De Sloover, L., Gevaers, R., and Van de Weghe, N., 2021. Je kan ook te veel vaccinatiecentra hebben.

Guida, C. and Carpentieri, G., 2021. Quality of life in the urban environment and primary health services for the elderly during the Covid-19 pandemic: An application to the city of Milan (Italy). *Cities*, 110, 103038.

Higgs, G., 2004. A literature review of the use of GIS-based measures of access to health care services. *Health Services and Outcomes Research Methodology*, 5 (2), 119–139.

Li, W., 2020. GeoAI: Where machine learning and big data converge in GIScience. *Journal of Spatial Information Science*, (20), 71–77.

McNeil, M.M. and DeStefano, F., 2018. Vaccine-associated hypersensitivity. *Journal of Allergy and Clinical Immunology*.

Parvin, F., Ali, S.A., Hashmi, S.N.I., and Khatoon, A., 2021. Accessibility and site suitability for healthcare services using GIS-based hybrid decision-making approach: a study in Murshidabad, India. *Spatial Information Research*, 29 (1), 1–18.

Sahin, E.K., 2020. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Sciences*, 2 (7), 1308.

Song, G., He, X., Kong, Y., Li, K., Song, H., Zhai, S., and Luo, J., 2022. Improving the Spatial Accessibility of Community-Level Healthcare Service toward the '15-Minute City' Goal in China. *ISPRS International Journal of Geo-Information*, 11 (8), 436.

Vlaamse Overheid - Departement Omgeving - Afdeling Vlaams Planbureau voor Omgeving, 2021. Knooppuntwaarde per ha - toestand 2019 [online]. Available from: https://www.geopunt.be/catalogus/datasetfolder/f7080d50-937a-403a-90c7-681c94aa36de [Accessed 3 Jun 2022].

Wei, X., Zhang, W., Zhang, Z., Huang, H., and Meng, L., 2023. Urban land use land cover classification based on GF-6 satellite imagery and multi-feature optimization. *Geocarto International*, 38 (1), 2236579.

Yin, S., Li, J., Liang, J., Jia, K., Yang, Z., and Wang, Y., 2020. Optimization of the weighted linear combination method for agricultural land suitability evaluation considering current land use and regional differences. *Sustainability*, 12 (23), 10134.