A CLASSIFICATION MODEL FOR THE INFERENCE OF SPATIAL PRECISION OF OPENSTREETMAP BUILDINGS WITH INTRINSIC INDICATORS

Ibrahim Maidaneh Abdi[a,b], Arnaud Le Guilcher[a,]*, Ana-Maria Olteanu-Raimond[a]

[a] Univ.Paris-Est, LASTIG MEIG, IGN, ENSG, F-94160 Saint-Mandé, France - arnaud.le-guilcher@ign.fr
[b] ITU-I, Djibouti University , Djibouti, Djibouti

ABSTRACT:

To evaluate the quality of OSM data, similarities between OSM features and their homologous features represented in a reference database are relevant metrics. However, reference databases do not exist everywhere or are not freely available. Thus, having data quality assessment methods that rely only on intrinsic indicators (i.e. based on data itself without considering external information) would be useful in these cases. This article specifically uses the radial distance as a target quality metric to measure the quality of shapes. Its aim is to build a random-forest based classification method that reconstructs whether this distance is higher or lower than a specified threshold, using only intrinsic indicators as inputs. The classification algorithm is evaluated on a first dataset by computing the ROC (Receiver Operating Characteristic) curve and using the AUC (Area Under Curve) as an evaluation metric. The transferability of the resulting algorithm is then evaluated by measuring its performance on a second, distinct dataset. The experiments show that the algorithm performs reasonably well on both the initial and the second dataset, and that intrinsic indicators give relevant information to infer comparison-based shape quality (i.e. the radial distance).

1. INTRODUCTION

The use of data produced voluntarily or involuntarily by citizens or communities, also known as crowdsourced data, s (See et al., 2016) to update or enrich authoritative databases (Liu et al., 2015, Ivanovic et al., 2020), to make decisions (Westrope et al., 2014), becomes a popular topics in the field of Geographic Information Science (GIS).

More specifically, Volunteered Geographic Information (VGI) has studied since 2007 (Goodchild, 2007), and VGI data and projects have taken a prominent position since then, giving birth to research works aiming at assessing the quality of VGI, or their possible uses. Indeed, different challenges exist when dealing with VGI. The last does not always follow strict sets of specifications, and when specifications exist, their enforcement may be looser to encourage contributor involvement. Being created or edited by contributors of varying expertise, VGI also often has heterogeneous quality. On the other hand, data quality is a crucial issue for both scientist and users of spatial data. For example, incomplete or imprecise data can lead to erroneous findings and inadequate decisions in diverse fields where spatial data are involved, such as urban planning, route optimization, or infrastructure sizing. Thus, VGI quality assessment presents multiple challenges and opportunities becoming a prominent research issue with various data quality assessment research works (Zielstra and Zipf, 2010, Goodchild and Li, 2012, Ivanovic et al., 2019), reviews (Antoniou and Skopeliti, 2015, Senaratne et al., 2017), theoretic frameworks (Barron et al., 2014) or taxonomies (Degrossi et al., 2018).

OpenStreetMap (OSM) is one of the most popular and wide-ranging VGI project. The aim of the OSM project is to create a world wide open geographic database, independent of institutional and commercial databases, based on openness and flexibility to encourage wide participation in the project. Since its creation in 2004, it has rallied a wide community of contributors that constantly edits and improves its database (e.g. adding new features, modifying existing one, importing datasets from open data). OSM has also developed activities of relief mapping after natural catastrophes (Kogan et al., 2016, Poiani et al., 2016). As a consequence, OSM data quality assessment is also a particularly important research subject, since they are are often used and treated as reliable data in many applications. In many countries VGI can be evaluated by comparing them with reference data produced by national mapping agencies (Antoniou and Skopeliti, 2015) using ISO (International Organization for Standardization) ISO 19157-2013 standard (e.g. completeness, consistency, spatial accuracy, temporal accuracy, and thematic accuracy). Data quality assessment becomes even more crucial in the absence of reference data, as OSM data can then be used as a guide for public, commercial or humanitarian actions. In such cases, the usefulness of VGI increases and quality-controlled VGI can be used as a substitute for deriving reference data for specific applications.

Assessing the quality of VGI with only intrinsic characteristics, is still largely an open challenge (Barron et al., 2014, Ivanovic et al., 2019, Truong et al., 2019), and the extent to which quality measurements can be inferred from intrinsic indicators is still an open issue (Maidaneh Abdi et al., 2020, Xu et al., 2017).

The goal of this paper is to propose an approach to assess the shape accuracy of OSM building using only intrinsic indicators. To reach this goal, we propose a machine learning method based on Random Forest approach which uses intrinsic indicators to classify the shape building into two classes: high and low shape accuracy with respect to a threshold. Specifically, we use the radial distance to measure shape accuracy of an OSM building with respect to the shape of its homologous feature issued from authoritative building datasets.

Our proposed approach is based on a twofold process:

---
* Corresponding author

1. We first train and evaluate our classification method on an initial dataset, and identify the most useful intrinsic indicators.

2. We then use a second dataset to evaluate the transferability of the method. We train distinct classifiers on the first and on the second dataset, and evaluate both of them on the first and on the second dataset separately. We compare the classifiers' performances in these four experiences to determine whether a classifier trained with a given dataset remains efficient on another dataset.

2. STATE OF THE ART AND METHOD CHOICES

To evaluate the quality of VGI , and specifically OSM data, a traditional approach is to compare VGI with data from a reference database, in areas where such a reference database exists. It consists in first matching VGI and reference datasets to define homologous features (i.e. features belonging to different databases but representing the same entities from the real world) and then to compute extrinsic indicators between homologous features (Girres and Touya, 2010, Haklay, 2010, Zielstra and Zipf, 2010, Van Damme and Olteanu-Raimond, 2022). Among the many research works on OSM data quality, several have focused on positional accuracy (Haklay, 2010, Kounadi, 2009, Fan et al., 2014, Van Damme and Olteanu-Raimond, 2022), but some of them also tackled other components of spatial data quality, like semantic accuracy (Girres and Touya, 2010, Van Damme and Olteanu-Raimond, 2022), completeness (Zielstra and Zipf, 2010, Fan et al., 2014, Van Damme and Olteanu-Raimond, 2022, Minaei, 2020), or temporality (Minaei, 2020, Schmidl et al., 2021). To go forward from an application point of view and help users to make decision about the usability of VGI, Siebritz (Siebritz, 2014) uses comparison with reference data and proposed to define a threshold of acceptability to select OSM data, and (Van Damme and Olteanu-Raimond, 2022) compute ISO 19157-2013 quality metrics for VGI and described them through machine readable metadata. The extrinsic indicators consists in computing distances between homologous features (e.g. euclidean distance, Hausdorff distance, shape distances) to measure position accuracy or to define confusion matrix to semantic and thematic quality metrics (Antoniou and Skopeliti, 2015). Two limits can be noticed for these comparison approaches. The first is the impossibility to assess the quality of VGI, if reference data are not available. The second is that the metrics are highly dependent on the models of the geographic databases being compared and assumes that the reference database is the perfect representation of the reality.

Other research proposed data quality assessment methods relying on intrinsic characteristics of the data (Barron et al., 2014). These intrinsic characteristics can be of several types (e.g. history, topology, internal consistency), and can be about the data themselves, or about the process of how these data were created. In (Hashemi and Ali Abbaspour, 2015), the authors use the editing history to identify topological inconsistencies. Among approaches using only the data themselves, some use spatial context to analyze the spatial consistency of data (see (Touya and Brando-Escobar, 2013), where the author specifically identify level of detail inconsistencies), while others rely on the geometries of individual objects (Maidaneh Abdi et al., 2020). Some studies concentrate on the history of edition of an object to evaluate its quality (Barron et al., 2014). Several studies analyze quality through the contributors who created and edited the data, and evaluate characteristics of these contributors such as experience, local knowledge, credibility (Flanagin and Metzger, 2008, Van Exel et al., 2010, Bégin et al., 2018). In (Truong et al., 2019), the authors analyze contributor interactions and identify different contributor profiles to refine this knowledge about contributors.

Finally, another category of data quality assessment approaches concerns the mixed approaches combining extrinsic and intrinsic metrics. It is about inferring an extrinsic indicator by using intrinsic indicators based on the research hypothesis that the discrepancies between VGI and reference data can be reconstructed from intrinsic indicators. Machine learning methods are used to establish the relationships between intrinsic features and the target extrinsic indicator (Mohammadi and Malek, 2015, Xu et al., 2017, Maidaneh Abdi et al., 2020). The last showed that individual intrinsic indicators gave relevant information about shape accuracy, but that they did not capture absolute accuracy very well.

Random forests (Breiman, 2001) is a machine learning method that is both flexible and relatively resistant to over-fitting. It can be used in different context both for classification and regression (Criminisi et al., 2012). Thus, we presume that quality assessment of OSM data can be conducted with either a regression or a classification method. Since, one of the uses of quality assessment in an operational context could be to identify data that one can consider as reliable, and data that are too uncertain and warrant additional control, it makes sense to consider the quality assessment such a classification issue by imposing a threshold on the target metric. Moreover, classification also has the merit to be more robust to outliers than regression (Ivanovic et al., 2019).

In this paper, the goal is to study whether or not machine learning classification approaches are able to assess the shape quality of VGI. We focus here on OSM building footprints. Thus, we aim to use a random forest classifier to reconstruct an extrinsic metric capturing the spatial shape accuracy of objects, using intrinsic indicators of the object as inputs. As the extrinsic metric, we choose the radial distance which is computed between an OSM building and its homologous object in a reference database. We make the hypothesis that the radial distance is a satisfactory proxy for the shape accuracy of an OSM building footprint. There is no unique way to define a distance between two polygonal shapes ; several metrics can be used to characterize shape (Basaraner and Cetinkaya, 2017, Zhang and Lu, 2004), and many of these metrics are not very discriminatory. In these cases, two polygon with similar metrics can be very different nonetheless. The radial distance is a pseudo-distance between polygonal shapes. It has been introduced by (Cohen and Guibas, 1997), and used and studied in the context of polygon matching (Vauglin and Bel Hadj Ali, 1998). Contrary to simpler metrics characterizing the shape, it separates polygons relatively well. Being defined as an integral, it is sensitive to perturbations of any point in a polygon. As a result, pertubed or generalized versions of a polygon generally have a low radial distance to the initial polygon (Méneroux et al., 2022). These properties make raidal distance a good indicator of shape accuracy.

3. METHODS

In this section, we describe the three steps of our proposed approach that allow to classify a polygon into high and low shape accuracy.

3.1 Target extrinsic measure and intrinsic indicators

According to (Méneroux et al., 2022), the radial distance is defined based on the radial signature of a polygon. The last is a function that associates to each value of the linear abscissa the distance between the centre of mass of the polygon and the point of the boundary at this linear abscissa. For the computation of the radial distance, we consider the doubly normalized radial signature, which is normalized with respect to both the perimeter of the polygon and the L^2 norm. The radial distance is then defined as the L^2 norm of the difference of the radial signature of the two objects.

Our goal is to build a classifier that distinguishes objects depending on whether their radial distance is higher or lower than a specified threshold, that would separate the more reliable data (for which the inferred radial distance is low enough) and the more imprecise data. It is difficult to find a natural threshold for the radial distance to separate reliable and imprecise data, so we chose to set the radial distance threshold as the median value of the radial distance on our first dataset. This choice has the advantage of testing the discriminating power of the algorithm and avoids balance issues in the training dataset, but it does not enable to capture how the algorithm fares in a setting where objects of dubious quality are rare and can be considered as anomalies. The binary algorithm then labels objects whose inferred radial distance is higher than the threshold as positives, and the other as negatives.

For the input indicators, we use a set of fourteen intrinsic indicators, described in (Maidaneh Abdi et al., 2020). We chose to test a wide range of indicators, because we had little preconceived notion of which indicators would be the most useful, and because random forests methods are robust to the addition of indicators that carry little information. Some of these indicators (such as granularity or compactness) are cited in the literature as relevant indicators for quality (Girres and Touya, 2010), others can be considered as being possible signals of an imprecise or incorrect input of the data. We give a quick textual or mathematical definition for each of them. In the following, for a polygon P we will note $\mathcal{A}$ the area of P, p its perimeter, n its number of vertices, $v \in V$ its vertices and $l \in L$ its edges.

- rec (rectangularity): $rec = \dfrac{\mathcal{A}}{\mathcal{A}_{SBR}}$, where $\mathcal{A}_{SSR}$ is the area of the smallest bounding reactangle of P.

- lme (mean length): $lme = \dfrac{\sum |l|}{|L|} = \dfrac{p}{n}$.

- lmx (max length): $lmx = \dfrac{\max_L |l|}{lme}$.

- lmn (min length): $lmn = \dfrac{\min_L |l|}{lme}$.

- out (outlier):

$$out = \frac{1}{lme} \max_{v \in V} \frac{1}{|V| - 1} \sum_{v' \in V \setminus \{v\}} \|v - v'\|.$$

- cpc (compactness): $cpc = \dfrac{4\pi \mathcal{A}}{p^2}$.

- cvx (convexity): $cvx = \dfrac{\mathcal{A}}{\mathcal{A}_c vx}$, where $\mathcal{A}_c vx$ is the area of the convex envelope of P.

- elg (elongation): $elg = \dfrac{l_{SBR}}{L_{SBR}}$, where l_{SBR} and L_{SBR} are respectively the width and length of the smallest bounding rectangle of P.

- qrc (q-reconstruct) is the proportion of vertices needed to reconstruct 80% of P (for the intersection over union metric).

- $ragl$ (right-angle) is the number of right angles in the polygon (with a small tolerance).

- per (perimeter) is the perimeter of the polygon.

- ori (orientation) is the overall orientation of the SBR of the polygon.

- grn (granularity): $grn = \dfrac{n}{p}$.

3.2 Settings for the initial classification and the study of area transferability

For our study, we use two datasets in two distinct study areas, to study the transferability of our algorithm. To do that, we first train and evaluate a random-forest classifier on the first dataset, and then study transferability using both datasets.

For the initial classification on the first area, the random-forest algorithm is trained and validated through a three-fold cross validation approach, where buildings of the dataset are separated in three subsets of roughly equal sizes with a uniformly random procedure.

When studying the transferability of the algorithm on a second distinct study area, we conduct three new experiments: (1) the classifier is trained on the initial area and validated on the second area; (2) the classifier is trained on the second area and validated on the initial area; (3) the classifier is both trained and tested on the second area. We want all versions of the classifier to be trained with the same number of examples, and as we do cross-validation for the classifiers trained and validated on the same area, we set the number of examples in every training sets a two thirds of the smallest dataset. In our study, the initial dataset has less examples than the second, so we fix T, the size of the training sets, as two thirds of the size of the initial dataset.

Then, for experiments where the classifier is trained and validated on the same area, we perform three-fold cross validation, with folds constituted randomly. When the union of the two training folds is of size bigger than T, we only keep T elements at random to constitute the training set. For experiments where the classifier is trained and validated on different areas, no cross-validation is needed and we randomly pick T examples in the training area to constitute the training set. The classifier is then validated on all examples of the validation area.

For the construction of the random forest classifiers, our implementation uses 500 trees, with a maximum depth of 10, and for each split, $4 \approx \sqrt{13}$ indicators are considered.

3.3 Evaluation of the classifiers

The performances of the different classifiers are computed using the ROC (Receiver Operating Characteristic) curve. For each example in the validation set, the output of the Random forest is the mean of the outputs of the trees of the forest ; it is a real number between 0 and 1, and can be understood as the

inferred probability (for the Random forest classifier) that the example is actually positive. Examples with the output value 1 are certainly positive (i.e., with a radial distance threshold higher than the radial distance threshold chosen), examples with output value 0 are certainly negative, and values between these extremes express the probability for the classifier that the object belongs to the positive class. To label examples as positive or negative, a probability threshold is chosen, and examples with a higher inferred probability are labelled as positive, and the others as negative. For an algorithm trained on a fixed training set, one can chose several probability thresholds and obtain different classifiers. A default setting is to choose $0, 5$ for the threshold, taking a higher value leads to a smaller number of examples labelled as positive, leading to less false positive examples (thus enhancing the sensitivity of the classifier) but also less true positive examples (which lowers the recall). The ROC curve gives a synthetic view of the performances of the algorithm for all possible thresholds, mapping the true positive rate (TPR) (i.e. the ratio of the number of objects correctly labelled as positive, over the total number of objects who really belong to the positive class) as a function of the false positive rate (FPR) (i.e. the ratio of the number of objects really belonging to the negative class, and incorrectly labelled as positive, over the total number of objects really belonging to the negative class) as the probability threshold spans the range $[0, 1]$.

The area under the ROC curve is a real number between 0 and 1 ; it is abbreviated as AUC, and it measures the performance of the classifier over the whole range of possible probability thresholds, instead of focusing on a unique functioning point. A perfect classifier would have an AUC of 1, and a classifier assigning labels completely at random would have an AUC of 0.5 ; high values (near 1) indicate that the classifier has a good discriminating power, while low values (near 0.5 or lower) indicate a poor discriminating power. It is useful to evaluate methods for which the functioning point is not known in advance, and for which users of the classifier may want to penalize false positives and false negatives in a manner specific to the application.

4. RESULTS AND DISCUSSION

4.1 Study areas

The first study area is a rectangular zone in the Val-de-Marne department in France, represented on Figure 1. The Val-de-Marne is a merely urban department south-east of Paris.The length of the study area is 6.83 km, its width is 4.63 km, for an area of 32.84 km^2. North and west of the area are two big cities (Créteil has a population of $92,000$, and Saint-Maur-des-Fossés a population of $75,000$), with a high population density. Figure 2 shows the dense repartition of building footprints in the north part of the area, in the city of Créteil. There, buildings follow strict alignments along streets, and there is little space between buildings, even inside building blocs. The east and south parts of the area are occupied by smaller, less densely populated cities. Figure 3 shows building footprints in the smaller town of Sucy-en-Brie. There, buildings do not follow the strict structures of more densely populated areas, and one observes other kinds of dispositions, with more space between buildings and generally looser alignments.

In this area, we found $29,152$ buildings in the OSM database, among which $22,989$ were successfully matched with an homologous object in the BDTOPO® (the French reference database). We constitute a dataset with $10,530$ buildings, picked randomly among the successfully matched OSM buildings.

Figure 1. Localization of the '94' study area

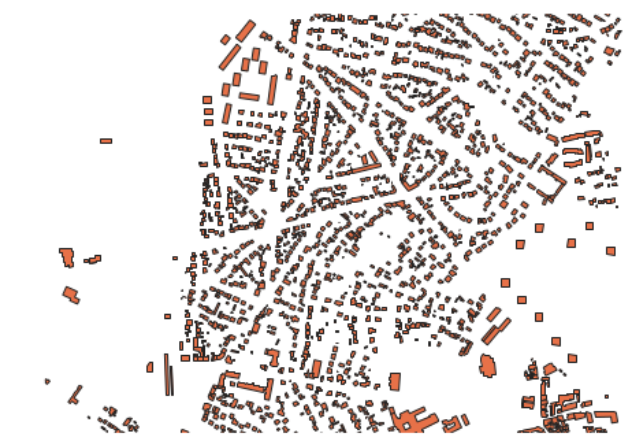Figure 2. Building footprints in the North of the '94' study area

Figure 3. Building footprints in the South-West of the '94 study area

The second study area is the Gers department, which is situated in the South-West of France. It is a rectangle of length 72.4 km and width 42.8 km, with an area of $3,138$ km^2. Figure 4 shows the second study area and the distribution of buildings in this area. There, one observes large areas with very sparse density, and concentrations of buildings in small or middle-sized towns isolated from each other, the largest of them being Auch (with
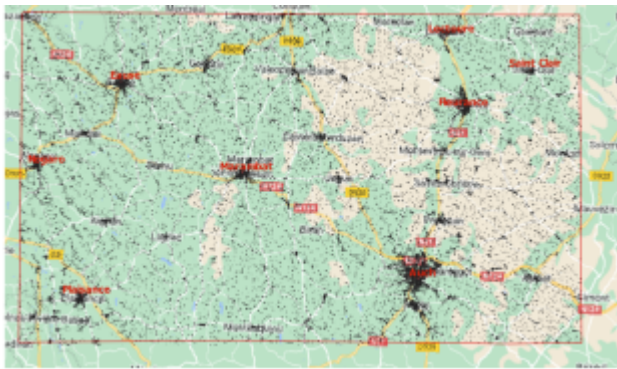
a population of 22,000).

Figure 4. Building footprints in the 'GERS' study area

In this study area, we constitute our second dataset by picking 19,068 OSM buildings among those that were successfully matched with a BDTOPO® counterpart.

4.2 Initial classification

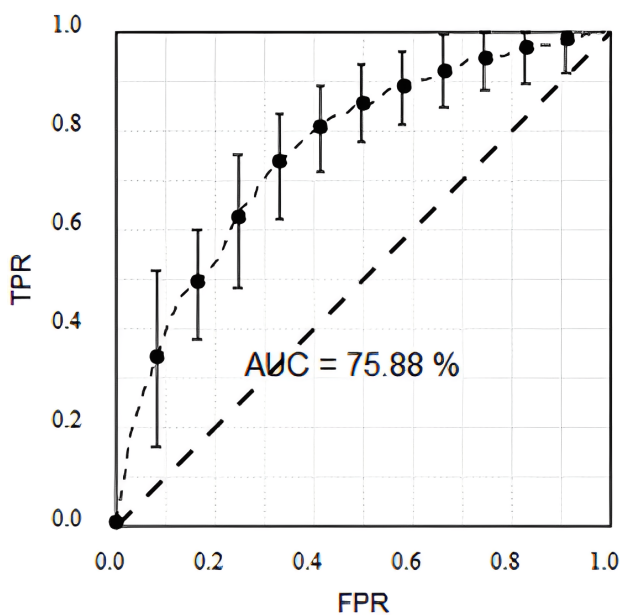Figure 5 shows the ROC curve for the initial classification experiment. The value of the AUC is 75.88%.

Figure 5. ROC curve for classification on the first dataset. TPR is the true positive rate, FPR the false positive rate, defined in subsection 3.3.

In Figure 5, confidence intervals are obtained by applying ten three-fold splits of the dataset and training and evaluating ten random forests with three-fold cross-validation. Confidence intervals correspond to plus or minus one standard deviation over these ten realizations.

According to (Hosmer Jr et al., 2013), a classifier with an AUC in the range [0.7, 0.8] can be considered as providing acceptable discrimination. The value of the AUC could be considered as reasonably good for a classifier that only has access to intrinsic indicators. This result indicate that intrinsic indicators of OSM features truly contain relevant information about data quality (a

classifier labelling completely at random would have an AUC of 50%), and that these indicators can help reconstruct a significant proportion of the precision measure we used. Conversely, the value of the AUC and the regular shape of the ROC curve indicate that such a classifier would hardly be useful in an operational context, where one would need to find a functioning point with either a combination of excellent recall and mediocre sensitivity, or excellent sensitivity and mediocre recall, or relatively high values for both. Here, the ROC curve does not stays near its tangents, which means that no meaningful information can be gained if one wants quasi-perfect recall or sensitivity, and there is no functioning point where both recall and sensitivity are reasonably high (recall and sensitivity reach both 70% for the most balanced functioning point).

4.3 Area transfer results

Figure 6 shows the four ROC curves corresponding to the four experiments decribed in Section 3.
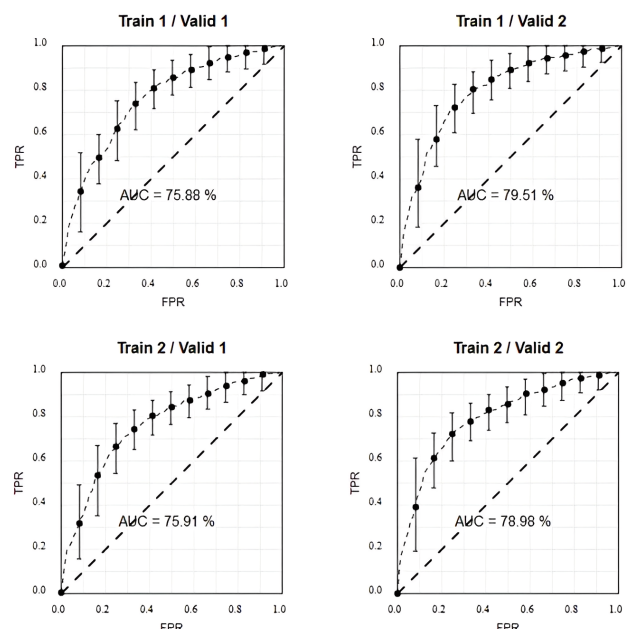

Figure 6. ROC curves for the transferability experiment.
Train i / Valid j is the experiment with the ith dataset used for training and the jth dataset used for validation, $i, j \in \{1, 2\}$

All four ROC curves are visually similar, and the four values of the AUC are near each other. The values of the AUC are 75.88% for the classifier trained and validated on the first dataset (which we note Train 1 / Valid 1), 79.51% for Train 1 / Valid 2, 75.91% for Train 2 / Valid 1, 78.98% for Train 2 / Valid 2. If we consider the pairs of algorithms that used the same validation set, we observe that their performances are similar between the algorithm trained on the same dataset and the algorithm trained on the other dataset. This results shows that the information captured by the algorithm for one dataset is still mostly valid for the other dataset, and thus that the results found in subsection 4.2 can be transferred to new areas. This good transferability could be explained by the fact that the two study areas are not completely different from each other ; both are in the same country, and, more importantly, both span several cities, providing relatively diverse building distributions, which gives algorithms trained on these areas tools to classify buildings in other areas with good performance. The drop in

performance could be more important if the algorithms were tested on a more radically different area, that would be situated in another country, or in with a distinct spatial context, like a seashore or a mountainous area. Yet, there are still meaningful differences between our study areas, the initial area does not feature sparsely populated country areas, and the second one does not feature densely populated and regular suburbs like the initial area. This tends to indicate that area transfer is possible even when the spatial context changes.

We also observe that he AUC is slightly higher for algorithms validated on the second dataset, which could be explained by the fact that this dataset could be slightly easier than the first, in the sense that the link between intrinsic indicators and shape precision is stronger for this dataset.

5. CONCLUSION

The aim of this paper is to assess the quality of polygons features representing building from OSM with respect to their shape. To reach this goal, we considered two matched datasets linking OSM and authoritative homologous buildings and we compute radial distances between homologous buildings to measure the similarity of their shapes. Based on these considerations, we have proposed a random-forest based classifier that distinguish OSM buildings with low shape quality (i.e. with a high radial distance) and OSM buildings with a low radial distance. Although the results, are promising, there are still not sufficient to be used without human intervention in operational context such as deriving a building reference dataset from OSM for countries where the reference data does not exist (e.g. Djibouti). Further research can be considered.

First, in this study, the intrinsic indicators we considered were all related to individual buildings. The performance of the classification model could be improved by adding context indicators that take into account the neighbourhoods of the buildings, either by computing aggregate indicators or by computing indicators that capture how much a building differs from its neighbours. Information on the editing history of buildings and the contributors who edited them could also be added to improve the assessment of the quality of OSM building data. Other measures than shape accuracy should also be investigated to get a more complete picture of data quality

Second, area transfer could also be tested in areas with more distinct spatial context : either mountainous, or seashore areas, or areas in other countries, or anther continent. If the first expansion proposed could be conducted with only OSM and BDTOPO® data, working on other countries could pose more significant problems : many countries do not have freely accessible reference building footprint data created by a national mapping agency, and even when such data exist, the specifications on how buildings are integrated and represented in the database can be very different from those of OSM. When it is the case, it is difficult to assess how much a computed discrepancy is due to data quality and how much is caused by specification differences.

Finally, in the context of deriving a building reference dataset from OSM for countries where authoritative data does not exist, it is necessary to combine both positional accuracy and shape accuracy, knowing that positional accuracy is a relevant trigger, to measure the geometric accuracy of the building. Positional accuracy could be calculated following the same approach, i.e.

using high quality matched datasets and computing similarity measures reflecting the positional accuracy (e.g. overlapping rate, Haussdorf distance). One direction of research to combine positional and shape accuracies could be to analyse the performance of the pre and post-merged classifications. (Joshi et al., 2016).

REFERENCES

Antoniou, V., Skopeliti, A., 2015. MEASURES AND INDICATORS OF VGI QUALITY: AN OVERVIEW. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3/W5, 345–351. https://isprs-annals.copernicus.org/articles/II-3-W5/345/2015/.

Barron, C., Neis, P., Zipf, A., 2014. A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, 18(6), 877–895.

Basaraner, M., Cetinkaya, S., 2017. Performance of shape indices and classification schemes for characterising perceptual shape complexity of building footprints in GIS. *International Journal of Geographical Information Science*, 31(10), 1952-1977. https://doi.org/10.1080/13658816.2017.1346257.

Breiman, L., 2001. Random forests. *Machine learning*, 45, 5–32.

Bégin, D., Devillers, R., Roche, S., 2018. The life cycle of contributors in collaborative online communities -the case of OpenStreetMap. *International Journal of Geographical Information Science*, 32(8), 1611-1630. https://doi.org/10.1080/13658816.2018.1458312.

Cohen, S. D., Guibas, L. J., 1997. Shape-based image retrieval using geometric hashing. *Proceedings of the ARPA Image Understanding Workshop*, ARPA, 669–674.

Criminisi, A., Shotton, J., Konukoglu, E. et al., 2012. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and trends® in computer graphics and vision*, 7(2–3), 81–227.

Degrossi, L. C., Porto de Albuquerque, J., Santos Rocha, R. d., Zipf, A., 2018. A taxonomy of quality assessment methods for volunteered and crowdsourced geographic information. *Transactions in GIS*, 22(2), 542-560. https://onlinelibrary.wiley.com/doi/abs/10.1111/tgis.12329.

Fan, H., Zipf, A., Fu, Q., Neis, P., 2014. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28(4), 700–719.

Flanagin, A. J., Metzger, M. J., 2008. The credibility of volunteered geographic information. *GeoJournal*, 72, 137–148.

Girres, J.-F., Touya, G., 2010. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4), 435–459.

Goodchild, M., 2007. Citizens as Sensors: The World of Volunteered Geography. *GeoJournal*, 69, 211-221.

Goodchild, M. F., Li, L., 2012. Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1, 110-120. https://www.sciencedirect.com/science/article/pii/S2211675312000097.

Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning and design*, 37(4), 682–703.

Hashemi, P., Ali Abbaspour, R., 2015. Assessment of logical consistency in OpenStreetMap based on the spatial similarity concept. *OpenStreetMap in GIScience: experiences, research, and applications*, 19–36.

Hosmer Jr, D. W., Lemeshow, S., Sturdivant, R. X., 2013. *Applied logistic regression*. 398, John Wiley & Sons.

Ivanovic, S., Olteanu-Raimond, A.-M., Mustière, S., Devogele, T., 2019. A Filtering-Based Approach for Improving Crowdsourced GNSS Traces in a Data Update Context. *ISPRS International Journal of Geo-Information*, 8(9), 380. https://www.mdpi.com/2220-9964/8/9/380.

Ivanovic, S. S., Olteanu-Raimond, A.-M., Mustière, S., Devogele, T., 2020. Potential of crowdsourced traces for detecting updates in authoritative geographic data. P. Kyriakidis, D. Hadjimitsis, D. Skarlatos, A. Mansourian (eds), *Geospatial Technologies for Local and Regional Development*, Springer International Publishing, Cham, 205–221.

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M., Kuemmerle, T., Meyfroidt, P., Mitchard, E., Reiche, J., Ryan, C., Waske, B., 2016. A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. *Remote Sensing*, 8(1), 70.

Kogan, M., Anderson, J., Palen, L., Anderson, K. M., Soden, R., 2016. Finding the way to osm mapping practices: Bounding large crisis datasets for qualitative investigation. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2783–2795.

Kounadi, O., 2009. Assessing the quality of OpenStreetMap data. *Msc geographical information science, University College of London Department of Civil, Environmental And Geomatic Engineering*, 19.

Liu, C., Xiong, L., Hu, X., Shan, J., 2015. A Progressive Buffering Method for Road Map Update Using OpenStreetMap Data. *ISPRS International Journal of Geo-Information*, 4(3), 1246–1264. https://www.mdpi.com/2220-9964/4/3/1246.

Maidaneh Abdi, I., Le Guilcher, A., Olteanu-Raimond, A.-M., 2020. A regression model of spatial accuracy precision for OpenStreetMap buildings. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 5(4).

Méneroux, Y., Maidaneh Abdi, I., Le Guilcher, A., Olteanu-Raimond, A.-M., 2022. Is the radial distance really a distance? An analysis of its properties and interest for the matching of polygon features. *International Journal of Geographical Information Science*, 1–38.

Minaei, M., 2020. Evolution, density and completeness of OpenStreetMap road networks in developing countries: The case of Iran. *Applied Geography*, 119, 102246. https://www.sciencedirect.com/science/article/pii/S0143622818312839.

Mohammadi, N., Malek, M., 2015. Artificial intelligence-based solution to estimate the spatial accuracy of volunteered geographic data. *Journal of Spatial Science*, 60(1), 119-135. https://doi.org/10.1080/14498596.2014.927337.

Poiani, T. H., Rocha, R. D. S., Degrossi, L. C., De Albuquerque, J. P., 2016. Potential of collaborative mapping for disaster relief: A case study of openstreetmap in the nepal earthquake 2015. *2016 49th Hawaii International Conference on System Sciences (HICSS)*, IEEE, 188–197.

Schmidl, M., Navratil, G., Giannopoulos, I., 2021. An Approach to Assess the Effect of Currentness of Spatial Data on Routing Quality. *AGILE: GIScience Series*, 2, 13. https://agile-giss.copernicus.org/articles/2/13/2021/.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H.-Y., Milčinski, G., Nikšič, M., Painho, M., Pődör, A., Olteanu-Raimond, A.-M., Rutzinger, M., 2016. Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. *ISPRS International Journal of Geo-Information*, 5(5), 55. http://www.mdpi.com/2220-9964/5/5/55.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., Haklay, M., 2017. A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31(1), 139–167.

Siebritz, L.-A., 2014. Assessing the accuracy of openstreetmap data in south africa for the purpose of integrating it with authoritative data. Master's thesis, University of Cape Town.

Touya, G., Brando-Escobar, C., 2013. Detecting level-of-detail inconsistencies in volunteered geographic information data sets. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 48(2), 134–143.

Truong, Q. T., De Runz, C., Touya, G., 2019. Analysis of collaboration networks in OpenStreetMap through weighted social multigraph mining. *International Journal of Geographical Information Science*, 33(8), 1651–1682.

Van Damme, M.-D., Olteanu-Raimond, A.-M., 2022. A method to produce metadata describing and assessing the quality of spatial landmark datasets in mountain area. *AGILE: GIScience Series*, 3, 17. https://agile-giss.copernicus.org/articles/3/17/2022/.

Van Exel, M., Dias, E., Fruijtier, S., 2010. The impact of crowdsourcing on spatial data quality indicators. *Proceedings of the GIScience 2010 Doctoral Colloquium, Zurich, Switzerland*, 14–17.

Vauglin, F., Bel Hadj Ali, A., 1998. Geometric matching of polygonal surfaces in giss. *Proc. ASPRS Annual Meeting*.

Westrope, C., Banick, R., Levine, M., 2014. Groundtruthing OpenStreetMap Building Damage Assessment. *Procedia Engineering*, 78, 29-39. https://www.sciencedirect.com/science/article/pii/S1877705814010224. Humanitarian Technology: Science, Systems and Global Impact 2014, HumTech2014.

Xu, Y., Chen, Z., Xie, Z., Wu, L., 2017. Quality assessment of building footprint data using a deep autoencoder network. *International Journal of Geographical Information Science*, 31(10), 1929-1951. https://doi.org/10.1080/13658816.2017.1341632.

Zhang, D., Lu, G., 2004. Review of shape represent-
ation and description techniques. *Pattern Recognition*,
37(1), 1-19. https://www.sciencedirect.com/science/article
/pii/S0031320303002759.

Zielstra, D., Zipf, A., 2010. Quantitative studies on the data
quality of openstreetmap in germany. *Proceedings of GIS-
cience*, 2010number 3.