

Article

Leveraging Road Characteristics and Contributor Behaviour for Assessing Road Type Quality in OSM

Amerah Alghanim ^{1,†}, Musfira Jilani ^{2,†}, Michela Bertolotto ^{3,†}  and Gavin McArdle ^{3,*,†} 

¹ School of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland; amerah.alghanim@ucdconnect.ie

² Independent Researcher, Dublin 4, Ireland; musfira.jilani.engg@gmail.com

³ School of Compute Science, University College Dublin, Belfield, Dublin 4, Ireland; michela.bertolotto@ucd.ie

* Correspondence: gavin.mcardle@ucd.ie

† These authors contributed equally to this work.

Abstract: Volunteered Geographic Information (VGI) is often collected by non-expert users. This raises concerns about the quality and veracity of such data. There has been much effort to understand and quantify the quality of VGI. Extrinsic measures which compare VGI to authoritative data sources such as National Mapping Agencies are common but the cost and slow update frequency of such data hinder the task. On the other hand, intrinsic measures which compare the data to heuristics or models built from the VGI data are becoming increasingly popular. Supervised machine learning techniques are particularly suitable for intrinsic measures of quality where they can infer and predict the properties of spatial data. In this article we are interested in assessing the quality of semantic information, such as the road type, associated with data in OpenStreetMap (OSM). We have developed a machine learning approach which utilises new intrinsic input features collected from the VGI dataset. Specifically, using our proposed novel approach we obtained an average classification accuracy of 84.12%. This result outperforms existing techniques on the same semantic inference task. The trustworthiness of the data used for developing and training machine learning models is important. To address this issue we have also developed a new measure for this using direct and indirect characteristics of OSM data such as its edit history along with an assessment of the users who contributed the data. An evaluation of the impact of data determined to be trustworthy within the machine learning model shows that the trusted data collected with the new approach improves the prediction accuracy of our machine learning technique. Specifically, our results demonstrate that the classification accuracy of our developed model is 87.75% when applied to a trusted dataset and 57.98% when applied to an untrusted dataset. Consequently, such results can be used to assess the quality of OSM and suggest improvements to the data set.

Keywords: Volunteer Geographic Information; spatial data quality; semantic inference; machine learning; data trustworthiness; user reputation



Citation: Alghanim, A.; Jilani, M.; Bertolotto, M.; McArdle, G. Leveraging Road Characteristics and Contributor Behaviour for Assessing Road Type Quality in OSM. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 436. <https://doi.org/10.3390/ijgi10070436>

Academic Editor: Wolfgang Kainz

Received: 30 April 2021

Accepted: 23 June 2021

Published: 25 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Many applications rely on the use of spatial data, in particular maps. Traditionally, authoritative maps produced by official/professional agencies were used. However, the use of authoritative map data has legal, technical and financial restrictions that prevent people from using them in many useful ways. In particular, limitations due to the cost of production affect development and updating and result in outdated data; they also make the acquisition of such data not accessible to all.

The need for free and up-to-date geospatial data, combined with the evolution of internet and web services, produced the phenomena of Volunteered Geographic Information (VGI) [1]. VGI projects use tools to create, assemble and disseminate geographic data provided voluntarily by individuals. The most popular VGI project is OpenStreetMap (OSM). Users can freely use the OSM platform to edit maps, potentially exploiting their

in-depth knowledge of the environment, and upload spatial data which becomes available to all.

However, the emergence of VGI has posed new challenges related to the veracity and accuracy of the spatial data which were not prevalent when a high level of control is imposed by cartographers and authoritative institutions [2]. The crowdsourcing nature of VGI implies the potential lack of cartographic skills of contributors. Therefore much attention has been devoted to assessing the usability of this data and identifying any limitations. In particular significant efforts have been made in evaluating the quality of the OSM data, including completeness, positional, topological and semantic accuracy, etc. [3–12].

Many existing OSM data quality assessment methods require referencing to authoritative data. However, such referencing is ineffective as authoritative data are not free and are often not up-to-date. Therefore, in our work, we propose the use of alternative approaches that rely on Machine Learning (ML) techniques and analyse specific characteristics (features) of the data.

One of the most interesting characteristics of VGI data relates to the contributors that edit the data. As there are no restrictions on who can contribute and no official data validation processes, being able to assess the trustworthiness of volunteers and their reputation among the community could provide indicators of data quality. Some research has been carried in this area but the results obtained are limited [13,14].

In our work, we focus on roads, as they represent the most fundamental objects of the OSM database (in this paper the terms road and street are used interchangeably). In particular, we aim to assess the semantic type of roads (i.e., the class a road belongs to, described by a tag in OSM). We have developed an approach to assess the quality of OSM road semantics that applies ML techniques and combines the specific features of a road (including its context in the map), together with the data trustworthiness and reputation of the volunteers that edited such a road. We chose London (UK) as test bed for our approach as OSM data in London is considered to be of very high quality due to an active community of mappers [15]. The main contributions of our work include:

1. The identification of street features and street context (represented by buildings surrounding streets) for application of ML techniques to assess and predict the type of streets in the OSM data. Classification results using data from the city of London (UK) show improvement over existing approaches.
2. The development of methods for calculating OSM data trustworthiness and user reputation based on historical edits. We used these methods to extract subsets of the OSM London data with specific trustworthiness/user reputation values and applied ML techniques to these subsets. Our results show that utilising information on the data trustworthiness and user reputation may contribute to improving the prediction of road types as they provide quality data for ML models.

The remainder of the article is organized as follows. Section 2 discusses related work. Section 3 presents our ML approach for assessing street types based on street features and context. Section 4 describes our data trustworthiness and user reputation model, as well as experiments carried out to validate its effectiveness. Finally, Section 5 provides some conclusions and ideas for future work.

2. Related Work

Given the concerns regarding the ability of novice volunteers to accurately capture and record spatial data [16], there has been much interest in understanding the accuracy and veracity of VGI. The International Cartographic Association (ICA) [17] identified seven measures to assess the veracity of spatial data. These seven measures which include the positional, attribute and temporal accuracy of the dataset as well as understanding the completeness, lineage and logical consistency were extended by Barron et al. [18] to include measuring the semantic and geometric accuracy and usability of the data. Recognising the

importance of volunteers contributions to data veracity, the trustworthiness and reputation of the data and users are also often included in measures to assess VGI [14,19].

Research has mainly focused on assessing semantic and positional accuracy as well as completeness. There are two broad methods used to assess VGI. Firstly, the data can be assessed extrinsically by comparing it to other external sources of spatial data. Secondly, VGI can be assessed intrinsically by identifying measures of accuracy within the data itself.

2.1. Extrinsic Measures

Extrinsic measures are most common [20] and typically involve comparing VGI to some authoritative map such as maps produced by National Mapping Agencies (NMAs) or commercial entities. Indeed, several feature types in OSM such as the road network [3–5], street names [6], POIs [7], educational POIs [8], pedestrian paths [9], road names [10], routing ways [11] and services [12] have all been compared to data from NMAs, Transport Operators, NAVTEQ Maps, TomTom Maps, Google Maps and Bing Maps. Typically studies focus on assessing accuracy in one test area. The results indicate that the accuracy of OSM concerning the completeness and positional accuracy is generally high, however, it can vary between features and locations. The semantic and attribute information is typically not as accurate when considering official map data as the ground truth. The assumption that the external data is accurate can be problematic. Authoritative data is often updated infrequently which can create a lag between the real world conditions and the map data. Indeed, one of the strengths of VGI is to fill this gap and so comparing VGI to NMA data may be misleading. Furthermore, comparing VGI to commercial and authoritative datasets is costly and may require licensing, this makes global comparisons difficult.

2.2. Intrinsic Measures

To address the limitations of extrinsic approaches for assessing accuracy and quality, intrinsic methods have been proposed. Intrinsic approaches do not rely on external or authoritative data sources for validating VGI. For example, the quality of the data can be assessed against predetermined logic rules regarding how features in the real world can be physically positioned [21]. Alternatively, the rules can be developed by examining the VGI data to identify meaningful patterns such as the co-existence of features, the distance between features, common placement of features [22] and common topological and geometrical patterns [23]. The context and urban function in which feature types are positioned can be used to assess if the semantic and attribute data are accurate [24]. This can also be supported by the use of ontologies for interoperability between semantic terms [25]. The identified patterns and rules can be used to produce probabilities regarding the accuracy of spatial data within VGI. However, they are not fully suited for understanding the completeness of VGI. Using a transfer learning approach, the rules generated in regions with rich data can be applied to regions where the data are more sparse. This assumes that there is adequate inter-domain similarities for the transfer approach to be effective [26]. Such approaches result in tools that can be used to find potential errors or to suggest corrections for map features [21] without necessarily producing an error score.

2.3. Machine Learning

Several of the intrinsic methods discussed above rely on ML techniques to detect and learn the rules regarding the relationships between features and space within the data being examined. There are two broad categories of ML algorithms. Supervised learning uses sample data to build and train a mathematical model. The model is then used to make predictions on unseen data. Unsupervised learning identifies patterns directly in the data being analysed and can be used for detecting features such as clusters and outliers. Supervised Learning is predominantly used for spatial data quality assessment.

For example, Sester [27] used a supervised learning approach, a decision tree model, based on geometrical and topological features to discriminate between houses, streets and land parcels. Walter Luo [28], developed a feature vector with different geometrical

measures related to the size and shape of map objects. A Neural Network (NN) was then used to classify map objects such as roads and buildings. Huang et al. [29] used the Markov Random Field (MRF) model to infer building type. The model used several features that shape the footprints of buildings, such as effective width and branching degree to learn and predict building type. Henn et al. [30] used similar features to describe building types but developed a Support Vector Machine (SVM) to classify building types. By learning from previously annotated entities, Giannopoulos et al. [31], also used an SVM algorithm to help recommend geospatial tags to be assigned to specific objects in OSM.

Funke et al. [32] developed a random forest classifier to detect gaps in the road network and propose missing street names by learning the topological and semantic characteristics of road networks in OSM. Jilani et al. [23] utilised the geometrical features of roads, such as the length, the number of dead ends, the number of intersections and linearity, as well as topological information regarding the type of adjacent streets, node degree and betweenness centrality, to predict street type. In the work SVN, NN and Random Forest (RF) approaches were all assessed with RF performing the best.

2.4. Data Trustworthiness and User Reputation

In addition to measures and techniques for assessing spatial data quality, there has also been a focus on determining the trustworthiness of the data along with the reputation of the person who provided the data. This is motivated by studies that have confirmed that the edit history [33] and the number of users editing [34] contribute to the quality of the data. Within this context, several studies have examined the contributors to VGI to determine their reputation and the trustworthiness of the data they contribute.

Sztompka [35] proposed measuring trust using two parameters, the distance of the contributor to the area they are mapping combined with a temporal decay. Keßler et al. [36] introduced vocabulary for assessing data provenance and trust in OSM. The assessment involved examining the history of map features by counting the number of edits, corrections, confirmations, versions and rollbacks associated with them. Keßler and de Groot [37] assessed the quality OSM features in the city of Münster using trust as an indication. They assigned positive indications to the number of contributors, versions and confirmations while the number of corrections and revisions were considered as negative indications for features trustworthiness. D'Antonio et al. [38] introduced a weighted sum of direct and indirect effects and time on the semantic, geometric and qualitative trustworthiness of data. Direct indications compare versions of map features to identify changes. A contributor who edits a feature without changing existing elements of it indirectly confirms the correctness of the existing elements. Time is used as a decay function to account for changes in the physical world that may not be captured in older VGI.

In their work, D'Antonio et al. [38] also propose a reputation score for users as the average calculation of the trustworthiness of all of the feature versions that the user produced. Fogliaroni et al. [39] extended this work and applied it by using the features versions' edits, for example, creation, modification and deletion to score the trustworthiness and, for each author, to score the reputation. The same method was used in a model by Zhou and Zhao [40] to find similar versions of map features in OSM and to calculate the user reputation based on a trust degree. Forati and Karimipour [41] considered social factors, such as previous behaviour as well as gender, age and occupation in determining trustworthiness.

Given the benefit of intrinsic approaches for assessing data quality, in this article, we build on existing literature and propose a novel approach for utilising intrinsic measures to assess the semantic quality of VGI data. Unlike other techniques, we utilise a variety of road features and place an emphasis on the context of a road feature as a means of assessing the veracity of its road type label. ML approaches are applied to learn and predict the patterns of these labels. We also propose a new approach for determining trustworthiness which focuses on data trustworthiness and user reputation. To evaluate the effectiveness of the approach, we extracted trusted data and untrusted data from OSM. The two data sets

were used independently with the ML approach. The results show the ML model performs better predictions with the trusted data than the untrusted data.

3. Street Characteristics for Semantic Type Assessment

Street or road type semantics in OSM, such as a motorway, primary street, residential street, etc. are represented using the ‘highway’ key. In this section, we hypothesize that the type of a street is a function of street restriction rules and the geometrical and contextual properties of a street. Therefore, we develop a supervised machine learning model that learns the types of streets given properties such as geometries and context as well as information about street restriction rules. Such a model can be used for predicting and correcting the semantic types of streets in OSM.

Toward the development of the street semantics model (road type classification), we first identify and prepare a suitable subset of the OSM database. The details of the dataset preparation are described in Section 3.1. Next, we construct and extract suitable features that are representative of the semantic types of streets (Section 3.2). Finally, we develop a machine learning model that learns the associations between the various features of streets (geometrical, contextual, street restriction rules) and their semantic types. We evaluate this model with respect to various standard evaluation metrics (Section 3.3).

3.1. Dataset Preparation

The supervised learning of road types proposed in this paper necessitates the availability of good ground truth data. OSM data in London in the UK is generally considered to be of very high quality [15], hence, in this work, we use the OSM London database for our model development and evaluation. In addition, only drivable streets that form the greatest majority of streets on the OSM street network, available in OSM through 13 distinct values of the ‘highway’ tag are considered in this study. These 13 street types include: ‘motorway’, ‘trunk’, ‘primary’, ‘secondary’, ‘tertiary’, ‘residential’, ‘motorway_link’, ‘trunk_link’, ‘primary_link’, ‘secondary_link’, ‘tertiary_link’, ‘living_street’ and ‘service’. Table 1 shows the distribution of road types in the study area. Specifically, we downloaded the shapefile of OSM London data from geofabrik and later used QGIS and PostgreSQL for viewing, analysing, and organising the street data so as to facilitate the extraction and construction of features efficiently.

Table 1. The frequency of different road types in the London dataset.

Type of Road	Count	Percentage
Motorway	225	0.17
Trunk	5446	4.14
Primary	10,421	7.92
Secondary	3093	2.35
Tertiary	7808	5.93
Residential	63,752	48.43
Motorway_Link	199	0.15
Trunk_Link	1001	0.76
Primary_Link	434	0.33
Secondary_Link	63	0.05
Tertiary_Link	90	0.07
Living_Street	157	0.12
Service	38,938	29.58

3.2. Feature Construction and Extraction

In this section, we describe the construction and extraction of geometric and contextual features, and restriction rules that are used as features in the proposed machine learning framework. The features used in this study fall into the categories of context, restriction rules and geometric characteristics. Context is the set of objects (object tags) adjacent to a road. Restriction rules represents some of the OSM road rules, (for example, the speed limit for a road type) and geometric characteristics are features related to the geometry of a road (for example, street length). While many features may have been selected, the rationale for choosing or disregarding certain tags (criteria) was based on knowledge of the most distinctive features for roads. For example, the feature surface was not selected because, for all drivable streets, it has the value 'Asphalt'. Therefore, it was not a distinctive feature. Similarly, for other tags not selected, their values appear in a small percentage of each type of road and so did not characterise the road type significantly.

3.2.1. Geometric Features

The various geometric characteristics of a street are usually a good indication of the semantic type of the street. For example, a motorway is usually a long, linear street with no dead-ends whereas a residential street is short and may consist of several dead-ends. Specifically, inspired by the work of Jilani et al. [23], this work uses four different geometric features namely, street length, number of nodes, number of intersections, and number of dead-ends. A definition of these features as used in this work is presented in Table 2. Please note that no linearity or shape feature is explicitly calculated in this work but is implicitly available through the features such as the number of nodes, the number of intersections, and the number of dead-ends. The length of an OSM street is the sum of the lengths of the segments called "ways" that constitute the street. This is important as it gives the total length of a given street rather than the distance between intersections on a street.

Table 2. Geometric features with their definitions/computation details.

Feature	Definition/Computation
Length	Multiple nodes are linked to represent a segment of a street; a total street length is computed by summing up the lengths of these segments
Number of Nodes	For each street type, a count of the number of nodes assigned to that street is calculated
Number of Intersections	This is a representation of the connectivity of a given street. If a street crosses another street at a specific node, then these intersected nodes are counted for each type of streets
Number of Dead-ends	This is mainly an in- or out-point and not a through a point on the map, represented by a node of degree one

3.2.2. Context

In addition to a street's geometric features, we hypothesise that the semantic type information of streets is also implicitly a function of its context. This work considers two types of contextual information. First, the type of buildings in the vicinity of a street. This information is available through the 'building' tag in OSM. Second, two aspects relating to the construction of a street, namely the presence of a tunnel and/or a bridge on the street are also considered. These two construction aspects of the streets are available in OSM through tags 'tunnel' and 'bridge' respectively. A buffer function was used to surround each street. The size of the buffer was decided empirically by observing the size which provides the most meaningful set of objects to form the context of a street. For example, a high street may have a buffer of 10 m since various stores are likely in close proximity to the street, while a highway may require a larger buffer to account for the lack of buildings touching the highway.

- **Buildings:** The values of the building tag can be used to describe the function or type of a building, for example, a building can be described as a House or a Hospital. It is expected that the hospital building feature would be adjacent to service streets and, as such, it would be able to identify service streets. Similarly, we expect that the house building feature would be adjacent to residential streets, and can identify residential streets. To understand this feature and support our hypothesis, the OSM database was analysed to find how frequently a given type of building is adjacent to each street type. It was found that each building type could be assigned to more than one street type. For example, 45.60% of houses are adjacent to residential streets and 20.65% of houses are adjacent to service streets. Based on statistical analysis and common knowledge, the building values considered in this work are *house, apartments, commercial, office, retail, university, hotel, hospital, school, outbuilding, shop, supermarket, bridge, industrial and garage*.
- **Bridge:** The “bridge = *” tag in OSM can be used to indicate the presence or absence of bridges on streets. Common knowledge suggests that bridges are not present on all street types but are a characteristic of specific street types. An analysis of the OSM London database confirms this. For example, it was found that the highest frequency of the bridge tag is for the primary streets (26.86%) followed by tertiary streets (17%). In other words, the primary and tertiary streets will more frequently have a bridge tag than other street types. Bridge features can occur on other types of street such as secondary and residential but the frequency of such an occurrence is lower. Furthermore, it has been assumed that the non-existence of a bridge may help to identify other street types. These results may be a factor of the London data used in this study and so care is needed if using these results in other jurisdictions.
- **Tunnel:** The “Tunnel = *” tag represents an underground passage for a street. A feature cannot be tagged as a bridge and a tunnel at the same time. Common knowledge suggests that the presence (or absence) of a tunnel on a street is a function of the semantic type of the street. An analysis of the OSM London database confirmed this. It was observed that the highest appearance of a tunnel is on the service street type (69.42%) (which means the tunnel is an underground passage for a service street). Furthermore, it was found that the primary_link never has a tunnel. Similarly these findings may be specific to the data set used in this study and analysis is required to understand the distribution of features in other regions.

3.2.3. Restriction Rules

Restriction rules are OSM rules applied to several OSM objects and indicate a prohibition of usage. Specifically, the restriction rules considered in this work are those that restrict drivers from using the street in a certain way. For example, the speed limit for a specific type of street should not be violated. Specifically, the two restriction rules considered in this work are the maximum drivable speed limit on streets and the oneway information of a street.

- **Maxspeed:** Common knowledge suggests that the maximum drivable limit of a street is a good indicator of the semantic type of the street. This information is available in OSM through the tag “maxspeed = *”. An analysis of the OSM London database was carried to understand the maxspeed speed values associated with the various street types.
- **Oneway:** The “oneway = *” tag is a restriction on streets where driving is permitted in one direction only. Specifically, “oneway = T” is used to indicate that a given street is oneway only. The rationale for including the oneway tag as a feature comes from common knowledge that certain types of streets such as *motorway* tend to bi-directional whereas street types such as *residential* tend to be one-way only. Table 3 provides a distribution of the oneway streets across the 13 semantic types of streets considered in this work (for our OSM London dataset).

Table 3. Distribution of the Oneway = T tag-value (in percentage rounded to the nearest whole number) across 13 semantic types of streets (C1–C13). Specifically, C1: motorway, C2: trunk, C3: primary, C4: secondary, C5: tertiary, C6: residential, C7: motorway_link, C8: trunk_link, C9: primary_link, C10: secondary_link, C11: tertiary_link, C12: living_street, C13: service.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
Oneway = T	0	4	9	3	6	38	0	0	0	0	0	0	41

3.3. Modelling and Evaluation

The OSM street network database is highly imbalanced in terms of the number of counts of various street types considered in the study. For example, the count of the motorway is much smaller compared to residential streets. In order to mitigate the impact of this imbalance, this work uses a split ratio of 50:50 for training and testing sets, respectively. This allows all 13 street types to be represented in both training and testing sets.

3.3.1. Modelling

Five popular machine learning algorithms namely SVM, Decision Trees, Random Forest, Multi-layer Perceptron Neural Networks and Naive Bayes were used in this work. The Python scikit-learn library (<https://scikit-learn.org/stable/>; accessed on 1 June 2021) was used to run test these approaches.

3.3.2. Evaluation

The machine learning models developed in this work are evaluated in two ways: firstly, in terms of performance with respect to standard evaluation metrics, and secondly, by comparing with previous work. In all cases, ten-fold cross-validation with 90% training and 10% testing sets was used. The average of the ten folds is reported as the result.

Metric Performance: It was observed that the model performance is best when all the features constructed in this work are used. In Table 4, we present a comparison of the performance of various machine learning algorithms, using all the proposed features (geometric, contextual, and restriction rules) for all the 13 street types. Specifically, we evaluate the models with respect to the four commonly used evaluation metrics namely, accuracy, precision, recall, and f1-score, respectively. It can be observed from the table that the performance of a Random Forest model is best in terms of all the considered metrics. In addition, for the given problem of learning 13 street types, an overall average accuracy of 84.12% is a very good result. This good performance of the RF model can be attributed to the fact that it is an ensemble of decision tree based classifiers, which are generally found to work well when the problem involves learning various rules inherent within the data [42].

Table 4. ML model performance in terms of various evaluation metrics (averaged over all 13 street types).

Algorithm	Accuracy (%)	Precision	Recall	F1-Score
Decision Tree	65.10	0.5107	0.6510	0.5663
KNN	68.46	0.6380	0.6846	0.6441
SVM	71.66	71.33	0.7166	0.6693
Neural Network	67.73	0.6249	0.6775	0.6214
Random Forest	84.12	0.8365	0.8406	0.8378
Naive Bayes	63.29	0.5709	0.6329	0.5475

Comparison with previous work: To the best of our knowledge, only a few works exist in the area of automatically learning and predicting semantic types of streets in OSM. Hence, in order to further understand the usefulness of our developed model, we compare it with the only directly comparable work of Jilani et al. [23]. This comparison is provided

in Table 5. As can be seen in the table, except for a slight fall in performance for classifying `living_street`, for the other 12 types of streets considered, the model developed in this paper outperforms the previous work. Even for the street types such as ‘trunk’ and ‘secondary’ roads, where the previous model struggled considerably, good accuracy values of 84% and 75% have been achieved. The improvement in the performance of the proposed model can likely be attributed to the incorporation of a combination of context information into the model such as the building type, availability of tunnel/bridge and the street restriction rules. Assessing which of the features is the most significant would require ablation study which we discuss in Section 5.

Table 5. Comparison of the Accuracy (%) with previous results from Jilani et al. ([23]) where C1–C13 represent 13 different types of roads. Specifically, C1: motorway, C2: trunk, C3: primary, C4: secondary, C5: tertiary, C6: residential, C7: motorway_link, C8: trunk_link, C9: primary_link, C10: secondary_link, C11: tertiary_link, C12: living_street, C13: service.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
Jilani et al. [23]	67	41	53	20	75	65	100	69	60	17	100	77	60
Our work	88	84	82	75	77	88	100	82	69	61	77	73	72

4. Data Trustworthiness and Contributor Behaviour Analysis for Semantic Type Assessment

In the previous section, we demonstrated the effectiveness of using ML to predict street types. The ML approach can be used to learn from existing data to label new unseen data. Its effectiveness relies on the quality and trustworthiness of the data used to build the ML model. Building upon previous studies [37,38,43], in this section, we hypothesise that data trustworthiness and contributor reputation can be used as indicators of semantic type quality in OSM. Toward confirming our hypothesis, we first prepare a historic OSM dataset. Next, we develop methodologies for calculating Data Trustworthiness (Section 4.1) and Contributor Reputation (Section 4.3) scores. Finally, we validate that the trustworthiness and reputation methodologies presented in this work (Section 4.4) are indicative of the data quality in OSM by analysing two samples of data with the ML approach described in the previous section. One sample is considered to have good quality data (trustworthy) and the other sample is deemed to be of poorer quality. We demonstrate that the scores for the validation metrics are influenced by the data trustworthiness and user reputation.

4.1. Data Trustworthiness

Trust of a feature $T(F)$ (in our case, a street or road) can be considered to be affected by several indicators such as Direct Indicator $T_d(F)$, Indirect Indicator $T_i(F)$ as well as Time Indicator $T_{time}(F)$, as described in Equation (1) [38,43]. The individual influence of these indicators in the overall calculation of $T(F)$ is governed by introducing three corresponding parameters, namely Direct Indicator Weight (W_d), Indirect Indicator Weight (W_i), and Time Indicator Weight (W_{time}), respectively. These are assigned weights of 0.5, 0.25 and 0.25, respectively. The values for the weights were selected based on the perceived importance in the validation accuracy of the data. Intuitively, direct indicators are the most significant. These values may be adjusted for other study areas.

$$T(F) = T_d(F).W_d + T_i(F).W_i + T_{time}(F).W_{time} \quad (1)$$

4.1.1. Direct Indicator

A Direct Indicator $T_d(F)$ is the overall trustworthiness that depends on the feature (Roads/Streets) version information [38]. The direct indicator equation used in finding the trustworthiness of the road data information contains measures such as the number of versions, the number of direct confirmations, the number of users, the number of edits,

the number of rollbacks, and the number of tags. Specifically, $T_d(F)$ is calculated as shown in Equation (2).

$$T_d(F) = (W_{num} \cdot V_{num}) + (W_{dir_c} \cdot Dir_C) + (W_{uc} \cdot UC) + (W_{edit} \cdot Tag_edits) + (W_{rollB} \cdot RollBk) + (W_{Tag} \cdot Tag) \quad (2)$$

A description of each measure, along with the rationale for its inclusion in the equation is provided below. A summary of the features, along with their relative weight parameter in the equation is provided in Table 6. The central tendency calculation for each parameter was determined through statistical analysis of the data to create a meaningful cut-off point. In many cases, this was the median value of the measure. Features having a value below the corresponding central tendency measure are assigned a value of *zero* value and those having a value above the corresponding central tendency measure are assigned a value of *one*. The weight parameter of a feature in the equation is related to its significance in determining data trustworthiness. The sum of the weights is equal to one. These parameter values have been manually set based on our domain knowledge of the relative importance of these individual features in the overall calculation of $T_d(F)$. Most indicator weights are set at 0.20 indicating equal importance between them in the overall calculation of quality. Based on our intuition, the indicators *Changes to Tags* and *Rollbacks* are seen as less significant in determining the quality of a feature and so have a weight of 0.10 assigned. It is safe to test these weights in other regions as they are universal.

- *Number of versions (V_num)*: A feature version is a source of information about the history of the feature starting from its creation. In OSM each object has a history recorded in a set of versions which include several attributes, such as name, road type, etc. It was hypothesised that the higher the number of versions of a road, the higher the quality of the road because the road was checked by many users and edited many times [15]. 3 is the median number of versions per road and was chosen as the central tendency measure. 55% of roads reached the central tendency value. The corresponding weight, W_{num} , has been assigned a value of 20%.
- *Number of direct confirmations (Dir_C)*: A confirmation depends on the trust a user has for the previous road version information created by another user in the road history record. For instance, if the version created by user "B" did not change information from the previous version created by user "A" then a direct confirmation will be counted. Only confirmations by different users are counted. Confirmations of the tags which are related to our semantics evaluation were considered, such as, name, highway (road_type), oneway, maxspeed, tunnel, bridge and geom (geometry). It was concluded from the overall statistics that a road should have at least one direct confirmation to state that the road information is trusted. The confirmation is directly related to the changes and edits to the road semantic data which is important to the semantic evaluation. Hence, the corresponding weight, W_{dir_c} , has been assigned a value of 20%.
- *Number of users for each road user count (UC)*: For every road created, a record of the user ID, name and the updates the user edited are recorded in the history versions of that road. The more users edit the road, the higher the trust of the data edited by those users (many eyes principle) [15]. It is believed that several users involved in editing the same road indicate it was checked and possibly corrected by different contributors which increases the trust in that road. The median value of 3 users was utilised as the central tendency measure for this metric. 54% of roads have 3 or more user count records. The UC is an important feature in the overall calculation of $T_d(F)$. Hence, the corresponding weight, W_{uc} , has been assigned a value of 20%.
- *Number of changes to road tags (Tag_edits)*: This factor concerns the edits that occur from one version to the next (independent of the contributor). A change could be adding new information, removing existing tag information, or changing the information of the tag from its previous status. An example is to change a road from one-way

to two-way. A statistical calculation was applied to analyse tags that are related to the semantic evaluation. The following seven tags were evaluated: *name*, *highway* (*road_type*), *oneway*, *tunnel*, *bridge*, *maxspeed*, and *geom*. As any change indicates a potential improvement, a central tendency of 2 was chosen. All of the previous edits are captured in the *Tag_edits* parameter and contribute towards the evaluation of the road semantic data quality. It is assumed that editing these seven tags has an impact on changing the road information. Therefore, the more editing of the tags occurs, the slower the process of improving the streets information. Based on domain knowledge regarding the importance of the feature *Tag_edits*, its associated weight parameter, W_{edit} , has been assigned a value of 10%.

- *Number of rollbacks (RollBk)*: In a road history, a rollback is defined as the deletion of the last version entirely to restore it to a previous state. Statistical analysis was carried out to find the percentage of roads with information deleted from the last version in the history record. It was found that about approximately 9% of the London, roads have no data in their last version, which can indicate that it was rolled back. This relatively high value of rollbacks may be as a result of the active community of OSM mappers in London who are anxious to ensure the quality of the data and so are likely to rollback data if there is any doubt over its quality. It is therefore assumed that the rollback has two possible explanations: either an act of vandalism or the action of correcting inaccurate data. Any rollback is important and 1 was chosen as the central tendency value. The impact of the rollback is not as significant as other measures. Hence the associated weight parameter, W_{rollB} , has been assigned a value of 10%.
- *Number of selected tags (Tag)*: For each road, a version history is recorded which contains several tags, for example, the road name tag and road_type tag. It is assumed that the higher the number of tags that are presented in the latest version of a road feature, the higher its quality. The selected tags are *name*, *road type*, *oneway*, *maxspeed*, *bridge* and *tunnel*. A median value of 2 (2 tags per road) was chosen as the value of the central tendency (i.e., 2 or more tags contribute positively to the quality of a feature). Given the significance and relevance of this feature in the overall calculation of $T_d(F)$, the associated weight parameter, W_{Tag} , has been assigned a value of 20%.

Table 6. Summary of the direct measures used for determining data trustworthiness

Direct Measure	Central Tendency Measure	Central Tendency Value	Weight (%)
# Versions	Median	3	20
# Direct Confirmations	Count	1	20
# Users for each road user count	Median	3	20
# Changes to road tags	Count	2	10
# Rollbacks	Count	1	10
# Selected tags	Median	2	20

4.1.2. Indirect Indicator

The trust indirect indicator $T_i(F)$ is the overall trust of feature information that does not directly depend on the feature version history but considers the editing of neighbouring spatial features. The indirect indicator depends on the context of the surrounding changes near the road being evaluated, such as changing the building information adjacent to the road. It is assumed that during editing of the surrounding area (context), road features in the area may also be checked [15]. For example, a user who is editing the information about another feature (buildings) near a road may check the road is located and named correctly. This can be considered as an indirect confirmation that helps in evaluating road data quality. To assess this, the level of activity in the surrounding area was considered and it was decided to assign the value of one to the indirect indicator of a road feature (F)

if it is a highly active area and zero if the area is not very active. The calculation of $T_i(F)$ is as follows:

$$T_i(F) = \begin{cases} 1 & \text{if } Active\ Area; \\ 0 & \text{if } Non\ Active\ Area \end{cases} \quad (3)$$

To identify the area activity, several aspects have been considered. These include the number of users editing each object (road, building and POIs) in the area in question; the number of objects, such as roads, buildings and POIs edited in that area. Roads, buildings and Point of Interest (POIs) were selected because they are the most edited features in the OSM database. A grid with cells of 1 km² was used to define areas. The objects (roads, building and POIs) were examined as well as the number of users who edited them to determine if a given cell is active.

In our statistical analysis, the central tendency calculation for each element we are considering (the number of objects, the number of users and the last update time for the grid cell) was determined by analysing three measures: average, mode and median and selecting an appropriate value which was representative of the data distribution. If there is a significant difference between the three values, it is necessary to find a single value to effectively describe the central tendency and cut-off value. Analysis of the data distribution found that the average of the three values was suitable and produced good results. The final central tendency values for each element are listed in Table 7. Given there is a difference in the editing frequency for the different types of objects, each type was considered independently.

Table 7. Indirect measure table.

Indirect Measure	Central Tendency Measure	Central Tendency Value
Roads in grid cells	Combination of Mean, Standard Deviation and Median	46
Users editing roads in cells	Visual Analysis of Data	144
Last Editing Time roads in cells	Combination of Mean, Standard Deviation and Median	c 3 years
POIs in grid cells	Combination of Mean, Standard Deviation and Median	29
Users editing POIs in cells	Combination of Mean, Standard Deviation and Median	66
Last Editing Time POIs cells	Combination of Mean, Standard Deviation and Median	c 3 years, 7 months
Buildings in grid cells	Visual Analysis of Data	93
Users editing Buildings in cells	Combination of Mean, Standard Deviation and Median	135
Last Editing Time of Buildings in cells	Combination of Mean, Standard Deviation and Median	c 1 years 8 months

4.2. Trust Time Indicator

The trust time indicator $T_{time}(F)$ includes the number of days from the update of the road's last version (v_{last}) timestamp until the whole database was downloaded. It is assumed that if the road information was not edited for a certain time that the data are trusted and the quality is high; otherwise if the data are still in the process of editing and not trusted, the quality might not be very high. The method used for finding the central tendency value for the Time Indicator was similar to the indirect indication statistical analysis used to find a balance between mean, median and standard deviation. It was found that about 65.22% of the roads were not updated or edited recently (more than 3 years, 5 months, 19 days, 19 h, 37 min, 38 s). This result seems to indicate that the majority of road data is trusted because it was stable. The value of the time indicator, in that case,

will be equal to one as a contribution to improving the data quality. The other roads with less than 3 years, 5 months, 19 days, 19 h, 37 min, 38 s are assigned zero values in the Equation (1). The following is the equation for computing Time Indicator:

$$T_{time}(F) = \begin{cases} 1 & \text{if } \text{downloaded date} - \text{time}(v_{last}) \geq \varepsilon \\ 0 & \text{if } \text{downloaded date} - \text{time}(v_{last}) < \varepsilon \end{cases} \quad (4)$$

Example Data Trust Score Calculation

Here we demonstrate the calculation of the proposed Data Trustworthiness metric for a given road. Toward this, we chose a random road segment with ID 74.

- **Step 1 : Compute Direct Indicator $T_d(F)$:**

For Road ID 74, we observe the following:

$V_{num} = 1$ as the road has 8 versions which is higher than the central tendency measure of 3. $Dir_C = 1$ as the road has 3 direct confirmations which is greater than the central tendency measure of 2. $UC = 1$ as there are 5 users editing the road which is above the central tendency value of 3. $Tag_edits = 0$ as the road has 2 edits which equal to the central tendency value of 2. $RollB = 1$ as the road did not have a rollback. $Tag = 1$ as the road has 3 complete tags (name, road_type and maxspeed) which is above the central tendency value of 2.

Using Equation (2), the above feature values are multiplied with their respective weights to calculate $T_d(74)$ as follows:

$$T_d(74) = (0.20 * 1) + (0.20 * 1) + (0.20 * 1) + (0.10 * 1) + (0.10 * 0) + (0.20 * 1) = 0.9$$

- **Step 2 : Compute Indirect Indicator $T_i(F)$:**

Next, we identify that the road is in an active area and the road area achieved more than two conditions. Hence, using Equation (3) we observe that $T_i(74) = 1$.

- **Step 3 : Compute Time Indicator $T_{time}(F)$:**

The last time Road 74 was edited was approximately 1 month, 10 days which is less than the central tendency value. Hence, using Equation (4) we observe that $T_{time}(74) = 0$.

- **Step 4 : Combine:**

Finally, using Equation (1) these computations are combined with their respective weights as follows indicating that the trust value for Road ID 74 is 70%.

$$T(74) = T_d(74) * 0.50 + T_i(74) * 0.25 + T_{time}(74) * 0.25 = 0.70$$

To ascertain a suitable value to indicate if data are trusted or untrusted, the metric described above was applied to all roads in the data set and the average of the median, standard deviation and mean was computed to determine a central tendency value of 46.23%. Roads with a score higher than this are trusted while a score lower than this indicates an untrusted road. The percentage of roads that have a trustworthiness value greater than or equal to 46.23% is 60.78%. The outcome was satisfying as the majority of roads were trusted. In our example above we can conclude that road 74 has trusted data. It is likely that the central tendency values will be different in different cities and regions but they can be calculated using the approach described here.

4.3. User Reputation

D'Antonio et al. [38] have defined the reputation (R) of a user (u), editing the road data, as the sum of the trusted road versions data $T(vi)$ edited by the user divided by all of the roads versions $V(u, t)$ edited by that user in time (t_{last}) which is the duration of the user life edits. In this paper, we propose a modified and more detailed methodology for computing user reputation scores as given in Equation (5). We develop Equation $T(vi)$ of the user trusted version into a more detailed calculation by dividing the version into tags as elements $T(vi_{tags})$. The sum of the trusted tags was divided by the number of tag

types used. Specifically in this research, we consider six tags related to our interest in the semantic quality of OSM data, namely *name*, *road_type*, *bridge*, *tunnel*, *oneway*, and *maxspeed*.

$$R(u, t_{last}) = \frac{\sum_{v_i \in Tags} T(v_i)_{tags}}{\|Tags\|} \quad (5)$$

where $Tags = \{name_{tag}, road_type_{tag}, bridge_{tag}, tunnel_{tag}, oneway_{tag}, maxspeed_{tag}\}$

Table 8 illustrates the six tag elements and the corresponding descriptions. For each of the six elements, a statistical analysis was carried out to find the central tendency value to indicate if the tags are trusted or not. The calculation uses the average of the mean, standard deviation and median values for the particular tag element description in the dataset. These central tendency values (computed as 0.8 for *name_tag*, 0.5 for *road_type_tag*, 0.8 for *oneway_tag*, 0.7 for *bridge_tag*, 0.7 for *tunnel_tag*, and 0.8 for *maxspeed_tag*, respectively) have been used for deciding the final trusted tag value.

Table 8. Six elements for User Reputation.

Trusted Tags Elements	Description	Final Value
Name tags $T(v_{i_{name_tags}})$	For each user, the number of trusted name tags divided by all of the user name tag contributions.	0 if tag data not trusted < (0.8); 1 if tag data trusted >= (0.8)
Road_type tags $T(v_{i_{road_type_tags}})$	For each user, the number of trusted road_type tags divided by all of the user road_type tag contributions.	0 if tag data not trusted < (0.5); 1 if tag data trusted >= (0.5)
Oneway tags $T(v_{i_{oneway_tags}})$	For each user, the number of trusted oneway tags divided by all of the user oneway tag contributions.	0 if tag data not trusted < (0.8); 1 if tag data trusted >= (0.8)
Bridge tags $T(v_{i_{bridge_tags}})$	For each user, the number of trusted bridge tags divide by all of the user bridge tag contributions.	0 if tag data not trusted < (0.7); 1 if tag data trusted >= (0.7)
Tunnel tags $T(v_{i_{tunnel_tags}})$	For each user, the number of trusted tunnel tags divide by all of the user tunnel tag contributions.	0 if tag data not trusted < (0.7); 1 if tag data trusted >= (0.7)
Maxspeed tags $T(v_{i_{maxspeed_tags}})$	For each user, the number of trusted maxspeed tags divide by all of the user maxspeed tag contributions.	0 if tag data not trusted < (0.8); 1 if tag data trusted >= (0.8)

The scale of the reputation of the user was determined from a statistical analysis of all of the OSM users editing the road database in London (UK). The statistical calculation for the three central tendency measures has different results. The standard deviation was calculated to find the difference and it was very high. Therefore, the central tendency value was calculated using further considerations. It was considered that the user should have at least three reputation elements with value one. The user reputation final central tendency was set to 50%, i.e., if the user achieved three reputation elements. This decision affected about 400 users who had obtained a good contribution for the *name_tag* and *road_type_tag*. It was decided that if the user reputation equals 50% or higher, the user will be considered as a user with a good reputation. Otherwise, if the reputation value for a user is less than 50%, the user will have a low reputation.

Example Contributor Reputation Score Calculation

An OSM user who contributed to the London OSM database was selected. For the user's privacy, user OSM name or OSM ID are not included. The number of edits the user made for the six tag elements were calculated. The user edited the name_tag 115 times, 113 of which were considered as "trusted" based on the analysis described in the previous subsection. This means that 98% of name_tags the user edited were trusted. The final reputation for the user name_tag is equal to 1 because using the central tendency measure, the user has $0.98 \geq 0.8$. In another example, the user edited the road_type_tag 172 times, 152 of which were considered as "trusted". This means that 88% of road_type_tags the user edited were trusted. The final reputation for the user road_type_tag is equal to 1 because using the central tendency value, the user has $0.88 \geq 0.5$. The user did not edit the tunnel_tag and the value is 0. The final reputation for the user tunnel_tag is equal to 0. Similar calculations were applied for the other six tag elements (see Table 9).

Table 9. Example of a user reputation calculation.

Tags	Number of Trusted Tags	Number of All Tags Contributed	Trusted Tags/All Tags Contributed	Compared with Central Tendency	Final Reputation
Name Tag	113	115	0.9826	≥ 0.8	1
Road Type Tag	152	172	0.8837	≥ 0.5	1
Oneway Tag	51	51	1	≥ 0.8	1
Bridge Tag	5	5	1	≥ 0.7	1
Tunnel Tag	0	0	0	< 0.7	0
Maxspeed Tag	4	4	1	≥ 0.8	1
Total Reputation					5
Final Reputation					0.83

The final reputations from each tag were summed and divided by 6. The final reputation for the user is 0.83 which was compared with the reputation central tendency. The result was ≥ 0.5 which concludes that the user has a good reputation. Approximately 76.11% of users have a value greater than or equal to 50% (these users are considered to have a good reputation and trusted data).

4.4. Machine Learning-Based Validation

In this section, we provide a validation of the proposed data trust and user reputation metrics. Specifically, we hypothesise that using the highest reputation users and the highest trusted road data which those users provided, gives a higher accuracy for street type prediction using the machine learning model described in Section 3.

Toward confirming this hypothesis, we extract two different samples of OSM data. The first sample corresponds to data with high data trust and high contributor reputation scores according to the equations described above. Specifically, we obtained data contributed by 925 contributors each having a reputation score of 100%. Next, for these 925 users, we selected only their data that corresponded to a data trust value ranging between 80% and 100%. We refer to this data consisting of 19,004 roads as Sample 1 and trusted roads. Similarly, the other sample obtained corresponds to data with low data trust and low contributor reputation scores. Specifically, we collected data contributed by 162 contributors each having a reputation score of less than 67%. Next, for these 162 users, we selected only that data that corresponded to a data trust value ranging between 0% and 30%. We refer to this data consisting of 1260 roads as Sample 2 and untrusted roads.

Next, we compare the performance of the Random Forest model described in Section 3 on the whole dataset (the entire London dataset), Sample 1 (good data), and Sample 2 (poor data). Table 10 shows that the performance of the model is best on Sample 1 (good data) in terms of all the four metrics of accuracy, precision, recall, and f1-score considered in this paper. The model was trained solely with the trusted data and then solely without trusted data in order to compare the quality metrics of data.

Table 10. Overall results of accuracy, precision, recall and F1-score for the whole data, trusted with high reputation and not trusted with low reputation users.

	All Data	Trusted Roads Data	Untrusted Roads Data
Accuracy (%)	84.12	87.75	57.98
Precision	0.8365	0.8737	0.5787
Recall	0.8406	0.8770	0.5797
F1-score	0.8378	0.8694	0.5483

These results demonstrate that using the data trustworthiness and user reputation can have an impact on improving the evaluation of road types using Machine Learning. Further, by using them as an indication, the data quality evaluation can be significantly improved.

5. Conclusions

The aim of our work is the assessment and prediction of the road types in OSM. We have developed an approach that relies on ML techniques and combines specific features of a road and its context with data trustworthiness and the reputation of the contributors that edited such a road. The main contribution of our work includes an ML approach to predict road types more accurately than the state-of-the-art techniques which demonstrates that context is an important feature that should be taken into account when determining road type.

In our study, the road context is represented by buildings surrounding roads. We applied our approach to OSM data from the city of London (UK). The results we obtained show improvement over existing approaches for predicting the street types and indicates that the measures we selected are appropriate and improve on state-of-the-art. Specifically, our experiments showed that a Random Forest (RF) based learning model can handle a rather complex problem of learning 13 types of streets with a very promising average accuracy of 84.12%. The 13 different semantic types of streets considered in this work are inherently a function of the various restriction rules and features of those streets. Indeed, the good performance of the RF model can be attributed to the fact that it is an ensemble of decision tree based classifiers, which are known to work very well when the problem involves learning various rules inherent within the data [42].

Another important result relates to the development of metrics that take into account not only features of the OSM editing process but also information related to the contributors as potential indicators of data quality. The random forest supervised learning model requires accurately labelled input data in order to produce results. To bootstrap the provision of such accurate data, we developed a metric to calculate the trustworthiness of data and the user reputation based on historical direct and indirect edits and interaction with OSM. To validate the approach, these techniques were applied to the London dataset to extract subsets with specific trustworthiness/user reputation values. We ran the same ML algorithms used for the original experiment and obtained results that indicated that information on the data trustworthiness and user reputation can contribute to improving the prediction of road types.

There are some limitations to the proposed methodology. For example, some of the road types we considered have common features which makes it difficult for the Machine Learning to correctly classify them (e.g., maxspeed = 48 for both primary and trunk roads). To improve the performance of the proposed model and reduce the misclassification of each road type, several possible strategies could prove worthwhile. For example, analyzing in detail how much each feature impacts on the identification of each road type. An ablation study could be an appropriate tool to assess this.

We restricted our experiments to 13 drivable road types; therefore our results are limited to those types. Different features should be identified and experiments should be conducted to see whether we can achieve comparable results for other road types.

In our proposed methodology, one way in which we leverage context is by considering the frequency associations of a given road type with various building types (within a certain buffer/neighbourhood). In future, we aim to make this approach more robust so that it can also cater to cases where a certain building type may not necessarily be the frequent type in the neighbourhood but is still a dominant feature. For example, the presence of building types such as ‘hospital’, ‘school’, etc in a street’s neighbourhood is often a good indicator of road type but is usually not the frequent building type in the considered neighbourhood. Toward incorporating such information, future work involves using the normalised counts of building types with their respective areas instead of relying on the raw count values. Furthermore, other contextual objects, e.g., Points of Interest (POIs), and associated information, could also be used.

Although a satisfying improved model was developed using data trust and user reputation with interesting results, a strong claim cannot be made in terms of usability of contributors data as an indication of quality independently, but rather that it can be used in conjunction with other methods. More research is needed in this area.

Finally, our results are limited to data from the UK (and, in particular, the city of London). This may impact the widespread applicability of some of the metrics which were derived from statistical analysis of the data. In particular the identification of the central tendency and cut-off could be improved where there is a significant variance in the data. In the future, it would be interesting to see how the approach can adapt automatically to the characteristics of road networks in other countries. Indeed, the use of the transfer learning paradigm [26] within VGI may be of particular relevance to this study.

Author Contributions: Amerah Alghanim, Michela Bertolotto and Gavin McArdle conceived and designed the research presented in this paper. Amerah Alghanim carried out the data collection, curation and data analysis; Musfira Jilani, Gavin McArdle and Michela Bertolotto wrote the initial draft while all authors reviewed and edited subsequent drafts and revisions. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Saudi Arabian Bureau (Ministry of Education) the King of Saudi Arabia Scholarship.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goodchild, M.F. Citizens as voluntary sensors: Spatial data infrastructure in the world of Web 2.0. *Int. J. Spat. Data Infrastruct. Res.* **2007**, *2*, 24–32.
2. Bertolotto, M.; McArdle, G.; Schoen-Phelan, B. Volunteered and crowdsourced geographic information: The OpenStreetMap project. *J. Spat. Inf. Sci.* **2020**, *2020*, 65–70. [[CrossRef](#)]
3. Koukoletsos, T.; Haklay, M.; Ellul, C. An automated method to assess data completeness and positional accuracy of OpenStreetMap. In Proceedings of GeoComputation, London, UK, 20–22 July 2011.
4. Brovelli, M.A.; Minghini, M.; Molinari, M.E. An Automated GRASS-Based Procedure to Assess the Geometrical Accuracy of the OpenStreetMap Paris Road Network. In Proceedings of the ISPRS Congress, Prague, Czech Republic, 12–19 July 2016.
5. Brovelli, M.A.; Minghini, M.; Molinari, M.; Mooney, P. Towards an automated comparison of OpenStreetMap with authoritative road datasets. *Trans. GIS* **2017**, *21*, 191–206. [[CrossRef](#)]
6. Jackson, S.P.; Mullen, W.; Agouris, P.; Crooks, A.; Croitoru, A.; Stefanidis, A. Assessing completeness and spatial error of features in volunteered geographic information. *ISPRS Int. J. Geo-Inf.* **2013**, *2*, 507–530. [[CrossRef](#)]
7. Ciepluch, B.; Jacob, R.; Mooney, P.; Winstanley, A.C. Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. In Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Leicester, UK, 20–23 July 2010; p. 337.
8. Siebritz, L.A. 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, Cape Town, South Africa, 2014.
9. Zielstra, D.; Hochmair, H.H. Comparative study of pedestrian accessibility to transit stations using free and proprietary network data. *Transp. Res. Rec.* **2011**, *2217*, 145–152. [[CrossRef](#)]
10. Ludwig, I.; Voss, A.; Krause-Traudes, M. A Comparison of the Street Networks of Navteq and OSM in Germany. In *Advancing Geoinformation Science for a Changing World*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 65–84.
11. Zielstra, D.; Hochmair, H.H. Using free and proprietary data to compare shortest-path lengths for effective pedestrian routing in street networks. *Transp. Res. Rec.* **2012**, *2299*, 41–47. [[CrossRef](#)]

12. Rabiei-Dastjerdi, H.; McArdle, G.; Ballatore, A. Urban Consumption Patterns: OpenStreetMap Quality for Social Science Research. In Proceedings of the 6th International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM), Prague, Czech Republic, 7–9 May 2020.
13. Madubedube, A.; Coetzee, S.; Rautenbach, V. A Contributor-Focused Intrinsic Quality Assessment of OpenStreetMap in Mozambique Using Unsupervised Machine Learning. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 156. [[CrossRef](#)]
14. Schott, M.; Grinberger, A.Y.; Lautenbach, S.; Zipf, A. The Impact of Community Happenings in OpenStreetMap—Establishing a Framework for Online Community Member Activity Analyses. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 164. [[CrossRef](#)]
15. Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ. Plan. B Plan. Des.* **2010**, *37*, 682–703. [[CrossRef](#)]
16. Antoniou, V.; Skopeliti, A. Measures and Indicators of VGI Quality: An Overview. In Proceedings of the ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, La Grande Motte, France, 28 September–3 October 2015; Volume 2.
17. Morrison, J.L. Spatial data quality. *Elem. Spat. Data Qual.* **1995**, *202*, 1–12.
18. Barron, C.; Neis, P.; Zipf, A. A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Trans. GIS* **2014**, *18*, 877–895. [[CrossRef](#)]
19. Senaratne, H.; Mobasheri, A.; Ali, A.L.; Capineri, C.; Haklay, M. A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 139–167. [[CrossRef](#)]
20. Kaur, J.; Singh, J.; Sehra, S.S.; Rai, H.S. Systematic literature review of data quality within openstreetmap. In Proceedings of the 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS), Jammu, India, 11–12 December 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 177–182.
21. Pourabdollah, A.; Morley, J.; Feldman, S.; Jackson, M. Towards an authoritative OpenStreetMap: Conflating OSM and OS OpenData national maps' road network. *ISPRS Int. J. Geo-Inf.* **2013**, *2*, 704–728. [[CrossRef](#)]
22. Kashian, A.; Richter, K.F.; Rajabifard, A.; Chen, Y.; Both, A.; Duckham, M.; Kealy, A. Mining the co-existence of POIs in OpenStreetMap for faulty entry detection. In Proceedings of the 3rd Annual Conference of Research@Locate, the Academic Research Stream at Locate, Melbourne, Australia, 12–14 April 2016.
23. Jilani, M.; Corcoran, P.; Bertolotto, M. Automated highway tag assessment of OpenStreetMap road networks. In Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Dallas, TX, USA, 4–7 November 2014; pp. 449–452.
24. Iddianozie, C.; Bertolotto, M.; McArdle, G. Exploring budgeted learning for data-driven semantic inference via urban functions. *IEEE Access* **2020**, *8*, 32258–32269. [[CrossRef](#)]
25. Baglatzi, A.; Kokla, M.; Kavouras, M. Semantifying OpenStreetMap. In Proceedings of Terra Cognita@ ISWC, Boston, MA, USA, 12 November 2012; pp. 39–50.
26. Iddianozie, C.; McArdle, G. A transfer learning paradigm for spatial networks. In Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, Limassol, Cyprus, 8–12 April 2019; pp. 659–666.
27. Sester, M. Knowledge acquisition for the automatic interpretation of spatial data. *Int. J. Geogr. Inf. Sci.* **2000**, *14*, 1–24. [[CrossRef](#)]
28. Walter, V.; Luo, F. Automatic interpretation of digital maps. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 519–528. [[CrossRef](#)]
29. Huang, C.; Davis, L.S.; Townshend, J.R.G. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* **2002**, *23*, 725–749. [[CrossRef](#)]
30. Henn, A.; Römer, C.; Gröger, G.; Plümer, L. Automatic classification of building types in 3D city models. *GeoInformatica* **2012**, *16*, 281–306. [[CrossRef](#)]
31. Giannopoulos, G.; Karagiannakis, N.; Skoutas, D.; Athanasiou, S. Automatic recommendations of categories for geospatial entities. In Proceedings of LocalRec@ RecSys, Vienna, Austria, 19 September, 2015; pp. 9–12.
32. Funke, S.; Schirrmeister, R.; Storandt, S. Automatic extrapolation of missing road network data in OpenStreetMap. In Proceedings of the 2nd International Conference on Mining Urban Data-Volume 1392, Lille, France, 11 July 2015; pp. 27–35.
33. Mooney, P.; Corcoran, P. How social is OpenStreetMap. In Proceedings of the 15th AGILE Conference on Geographic Information Science, Avignon, France, 24–27 April 2012.
34. Napolitano, M.; Mooney, P. MVP OSM: A tool to identify areas of high quality contributor activity in OpenStreetMap. *Bull. Soc. Cartogr.* **2012**, *45*, 10–18.
35. Sztompka, P. *Trust: A Sociological Theory*; Cambridge University Press: Cambridge, UK, 1999.
36. Keßler, C.; Trame, J.; Kauppinen, T. Tracking editing processes in volunteered geographic information: The case of OpenStreetMap. In Proceedings of the Identifying Objects, Processes and Events in Spatio-Temporally Distributed Data (IOPE), Workshop at Conference on Spatial Information Theory, Belfast, ME, USA, 12–16 September 2011; Volume 12, pp. 6–8.
37. Keßler, C.; De Groot, R.T.A. Trust as a proxy measure for the quality of volunteered geographic information in the case of OpenStreetMap. In *Geographic Information Science at the Heart of Europe*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 21–37.
38. D'Antonio, F.; Fogliaroni, P.; Kauppinen, T. VGI Edit History Reveals data Trustworthiness and User Reputation. In Proceedings of the 17th AGILE International Conference on Geographic Information Science, Castellon, Spain, 3–6 June 2014.
39. Fogliaroni, P.; D'Antonio, F.; Clementini, E. Data trustworthiness and user reputation as indicators of VGI quality. *Geo-Spat. Inf. Sci.* **2018**, *21*, 213–233. [[CrossRef](#)]

40. Zhou, X.; Zhao, Y. A version-similarity based trust degree computation model for crowdsourcing geographic data. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, Prague, Czech Republic, 12–19 July 2016; Volume 41.
41. Forati, A.M.; Karimipour, F. A VGI quality assessment method for VGI based on trustworthiness. *GI_Forum 2016* **2016**, *4*, 3–11. [[CrossRef](#)]
42. Hartshorn, S. *Machine Learning with Random Forests and Decision Trees: A Visual Guide for Beginners*; Kindle ed.; Available online: https://www.amazon.com/Machine-Learning-Random-Forests-Decision-ebook/dp/B01JBL8YVK#detailBullets_feature_div2016 (accessed on 1 June 2021).
43. Groot, R.T.A.D. Evaluation of a Volunteered Geographical Information Trust Measure in the Case of OpenStreetMap. Ph.D. Thesis, Universidade Nova de Lisboa, Lisbon, Portugal, 2012.