International Journal of Applied Geospatial Research
Volume 14 • Issue 1

Evaluating the VGI Users' Level of Expertise:

An Application of Statistical and Artificial Neural Network Approaches

Elaheh Azariasgari, Islamic Azad University, South Tehran, Iran
Farhad Hosseinali, Shahid Rajaee Teacher Training University, Iran*

ABSTRACT

Currently, online maps are of the most innovative and significant sources of information in people's daily life. However, quality assessment of volunteered geographic information (VGI) data raises some challenges. This research aims to analyze the VGI participants' level of expertise through evaluation of their background information. Towards this goal, an android application was developed to test users' knowledge and cognition about some selected regions of city as well as their background information. In order to evaluate the quality of information expressed by participants, some changes were made in Tehran's online map, and users were asked to identify the changes and to guess the vanished attributes. Statistical and ANN approaches were applied for analysis. The results demonstrated that the ANN was able to predict the percentage of correct answers of a new volunteer with mean squared error of 0.2. This research suggests that users' age and familiarity with the specific region in the city play more significant roles in their expertise in using online maps and in probable participating in VGI.

KEYWORDS

Artificial Neural Network, Data Quality, GIS, Statistical Analysis, Tehran, VGI

INTRODUCTION

In recent years, there has been a tendency to use the Web to create, assemble, and disseminate geographic information provided voluntarily by individuals. The users are also interested in contributing to the field of geographic information called volunteered geographic information (VGI) (Goodchild, 2007). VGI has the potential to be a significant source of geographic information at the Earth's surface. However, there are many challenges with VGI (Kessler, 2011). VGI is based on geographic maps and also is a progressive field of study in the former science and technology of geographic information systems (GIS) (Gómez-Barrón et al., 2019). Urban rapid expansion and changes also make it difficult to provide formal up to date maps for everyday works (Hosseinali et al., 2014; Lin, 2018). Websites such as Wikimapia and OpenStreetMap are samples of VGI-based systems that often provide the cheapest source and sometimes the only source of geographic information (Ali

DOI: 10.4018/IJAGR.316770
*Corresponding Author

& Schmid, 2014; Da Costa, 2016). However, VGI data can be even obtained from the websites whose activities are in other areas and allow users to add geographic information as well (Astaburuaga et al., 2022; Feng & Sester, 2018).

Nevertheless, integration of VGI with valid data, Spatial Data Infrastructure (SDI), and other data has emerged as an issue, which is due to the fact that it is difficult to accurately determine the quality of VGI (Fogliaroni et al., 2018; Severinsen et al., 2019). There are two main methods to assess VGI data quality. In the first approach, it is assumed that formal authoritative data is quality data and can be used for VGI data validation. The most important disadvantage of this method is that it requires access to formal authoritative data, which might not be always possible. Also, correctness and consistency of authoritative data cannot be guaranteed, which is due to the slow update rate of authoritative data-sets, resulting in discrepancies between data and reality that arise over time. In the second approach, the evolution of data, namely its history or progress is analysed. The advantage of this approach compared with the first one is that it does not require external data sources and considers the timeliness dimension of data. However, this method does not measure data quality accurately, but instead provides an approximation (Antoniou & Skopeliti, 2015; Dai et al., 2008; Severinsen et al., 2019).

Research has shown that three indicators of practice, skill and motivation can affect the quality of the data produced (Yang et al., 2016). Also, the concept of informational trust has been introduced to the established notion of interpersonal trust, and it has been suggested to use informational trust and reputation as proxy measures for information quality. The question that is raised here is that whether users' background information including occupation, age, etc. can be an indicator of people's level of expertise in producing volunteered geographic information? And if this information plays a role in validity of produced data? Therefore, in this paper, we aim to evaluate the quality of information expressed by VGI participants (in an urban environment) based on their background information, by using Artificial Neural Network (ANN) as well as statistical approaches. The next section discusses related work and illustrates the necessity of our work. Then, the methodology is explained, and the results are expressed and briefly discussed. Finally, the last section summarizes the article and provides recommendations for future works.

BACKGROUND

This part of the study reviews different methods and approaches that have been used by previous researchers to assess the VGI data quality.

Goodchild and Li (2012) described three approaches to quality assurance, which were termed the crowd-sourcing, social, and geographic approaches. Crowd-sourcing approaches "refers to the ability of a group to validate and correct the errors that an individual might make" while social approaches "relies on a hierarchy of trusted individuals who act as moderators or gate-keepers". The geographic approach is powered by certain rules of syntax that reveals what can and cannot occur at a given location (Goodchild & Li, 2012). In another work, Antoniou and Skopeliti (2015) addressed the various aspects of measuring VGI's quality (Antoniou & Skopeliti, 2015). Senaratne et al. (2017) reviewed the literature of VGI data quality assessment and classified them into three categories of: map-based, image-based and text-based (Senaratne et al., 2017). Bordogna et al. (2014) addressed the ways of increasing VGI data quality (Bordogna et al., 2014). Moreri et al. (2018) proposed a Trust and Reputation Modelling (TRM) methodology to measure the quality of VGI data in Mochudi, Botswana without the typical reference to ground truth. TRM measurements involved: Thematic accuracy measure, Semantic accuracy measure, Volunteer credibility determination and Positional accuracy determination. They underlined that "metadata about contributed data sets can be created by the development of rating applications for the public to police themselves in assessing and subjectively rating the accuracy of other volunteer contributions" (Moreri et al., 2018).

If formal data is available in a region, an appropriate approach for data evaluation is to compare formal data and collected data through VGI. This approach was applied by Forghani and Delavar in Tehran (Forghani & Delavar, 2014). Also, Teimoory et al. used the formal data to evaluate the reliability of OSM data regarding their history (Teimoory et al., 2021). However, the problem is that in many cases, appropriate and updated formal data is not available. In these cases, one of the most effective ways to evaluate the validity of data is to rely on the quality of data producers; therefore, some researchers used this approach. For instance, Coleman et al. (2009) discussed that contributors of volunteered information can be divided into five categories of "Neophyte", "Interested Amateur", "Expert Amateur", "Expert Professional" and "Expert Authority", based on their characteristics and motivations (Coleman et al., 2009). Yang et al. (2016) assessed the contributors of OSM in Germany, France and U.K. by investigating whether the contributors were professional. They concluded that the most contributors of OSM "are hardly amateurs, but are professionals instead" (Yang et al., 2016). Rajaram and Manjula (2019) focused on determining user proficiency on OSM data based on contribution behaviour of users. They approximated that in India 17% of users were key contributors and 27% were inexperienced local users (Rajaram & Manjula, 2019). Zhang et al. (2021) exploited evaluation-based weighted PageRank on OSM data for ranking of VGI contributor reputation. The study revealed that more active VGI contributors are more reputed (Zhang et al., 2021). Mohammadi and Sedaghat applied an ANN approach to achieve a quality index of data from VGI users' need (Mohammadi & Sedaghat, 2021). Severinsen et al. (2019) developed a model called VGTrust, which "assesses information about a data author, and the spatial and temporal trust associated with the data they create to produce an overall VGTrust rating metric". VGTrust was a weighted linear combination of six criteria. The authors assumed determining the quality of data author as a major principle of determining the quality of VGI data (Severinsen et al., 2019).

Another issue is that collecting VGI data from sources other than famous VGI sources such as OSM is a difficult task. But, when specific information is needed, collecting this kind of data becomes necessary. For example, Can et al. (2019) used VGI to extract landslides from the photos taken by the volunteers. They developed a convolutional ANN for this purpose and found VGI a good resource for collecting this type of data (Can et al., 2019). Martella et al. (2019) developed an android application to collect spatial information via gaming. People participated in the study used the gaming application whose environment was based on the target map. People (especially students of the university) were gathering information of the target (University of Münster, Germany). The efficiency of the method was verified by the questionnaires filled by the users (Martella et al., 2019).

In addition to the factors mentioned above, the concept of spatial (locational/geographical) knowledge of participants may affect the quality of VGI. Spatial knowledge is a fundamental factor for a VGI contributor's level of expertise. The level of people's spatial knowledge varies from individual to individual and depends on many variables such as people's experiences, age, job, etc (van der Ham & Claessen, 2020). Mullen et al (2015) argued that the quality of data obtained from the participants have a complex relationship with the demographic properties (Mullen et al., 2015). However, some studies aimed to investigate the spatial knowledge of people. Klonner et al (2021) established a field-based study to collect data about the opinion of people about the risk of flood in Eberbach, Germany. Half of the participants of that study were residents of that area and the other half were passengers. The results showed that the residents believe a greater area is disposed to flood (Klonner et al., 2021). Xu et al (2013) evaluated geographic awareness of people based on exploratory analysis of Twitter data. They demonstrated how it is possible to use social media to evaluate the spatial knowledge of a society (Xu et al., 2013). Olszewski and Wendland (2021) developed a platform called Digital Agora which enables the acquisition of spatial knowledge in order to develop a geoinformation society. Digital Agora may be used to collect spatial data and transform them into useful spatial knowledge. The platform allows to understand different attitudes of various groups of participants (such as people of different genders and age groups) toward studied locations (Olszewski & Wendland, 2021).

The review on literatures showed that validating VGI data is an extensive and still novel field of study. In this research, the level of expertise and spatial knowledge of volunteers is assessed through evaluation of participants' background information. The results may be interpreted as a combination of participants' spatial knowledge and participants' trustworthiness. In order to do so, some predefined information of users is entered into a developed VGI application as input data. Therefore, based on the accuracy of geographic information assessed by users, the background information of participants is analysed through statistical and ANN approaches. Finally, the quality (reliability) of the information provided by the users in VGI application can be predicted using their background information. The method proposed in this research can be used to evaluate the VGI users, and it may be applied as an index for the reliability of participants' contribution.

MATERIALS AND METHODS

In this study, ANN and statistical approaches are applied as tools to determine the VGI users' level of expertise by taking into account the participants' background information. In this method, users enter the requested information in VGI app upon registration and start answering predefined tests extracted from the maps. The percentage of users' true answers may indicate their level of expertise. After participation of hundreds of users, the answers are analysed. As a result, immediately after a user enter a VGI app, the level of his/her expertise can be investigated by the system which may be interpreted as the trustworthiness of his/her participation.

In this method, it is required to have primary information about VGI users' background. Since this information is not available, first, it is needed to design an app that has features of a VGI app to provide volunteered geographic information and receive users' primary information. In order to do so, an android application is designed including a questionnaire on primary information of users as well as the map of Tehran (the capital of Iran). It is provided for users through designed app to voluntarily respond the questions. The collected information of users is then analysed through both ANN and statistical approaches. In this research, the App Inventor tool has been used to create an android-based application. Also, statistical analysis and Neural Network programming have been performed using SPSS and MATLAB software, respectively. Figure 1 shows the algorithm considered in this research.

Questionnaire Design

A group of multiple-choice questions was designed to provide participants with multiple answer options, which provide this opportunity to compare the participants' response with each other and prevent from producing wrong and incomplete information. The questionnaire includes 10 variables summarized in Table 1.

The questions about the map of Tehran were designed based on the map provided at https://map.tehran.ir/ by the municipality of Tehran. This map is a formal data. The questions are related to 17 different busy regions scattered all around the city. In each selected region, three changes were made including removing the street or highway names, creating new streets, removing the name of parks or hospitals, etc. The changes were labelled with the question number on the map so that the users can make a relationship between the questions and the map and respond the questions. Three answer options were considered for each question; one correct option, one incorrect option, and one "Don't know" option. To have consistency among users' answers, they were asked to choose their answer from options provided. For instance, three changes were made in Vanak Square according to Figure 2. In the first and third change, the street names were removed, and in the second change, a new street was added to the map, and the users were asked whether this street really exists. They were also asked about the way of acquaintance with that region of the city.

Figure 1. Proposed Research Method Flowchart

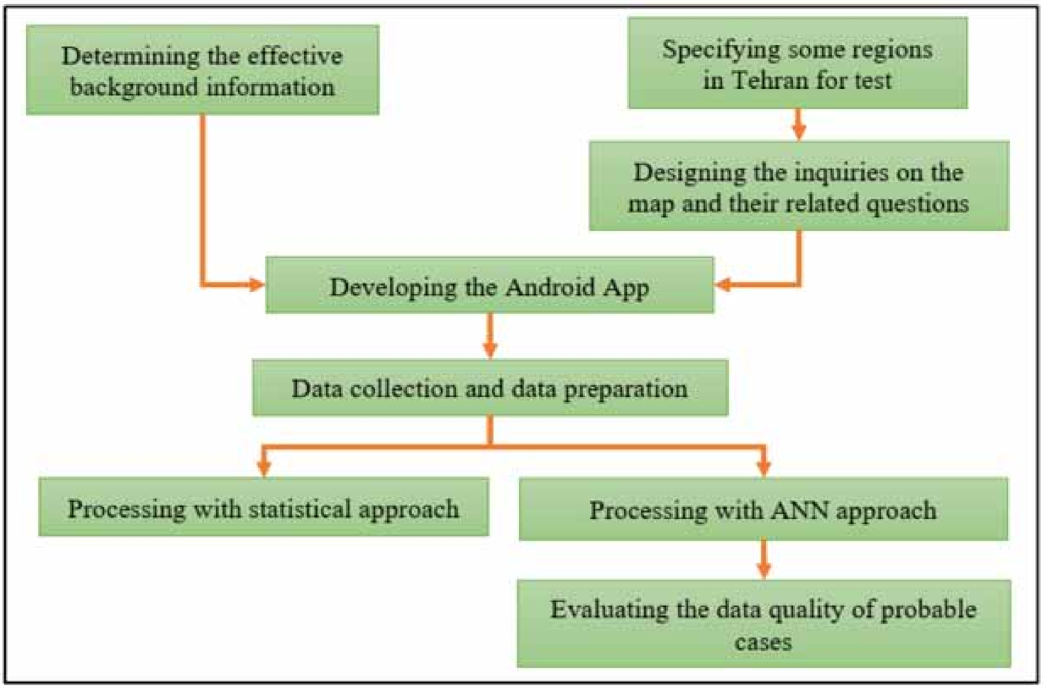


Table 1. The background questions asked from the contributors

Number	Question	Options
1	Age	Less than 18, Between 18 and 30, Between 30 and 40, Between 40 and 50, More than 50
2	Gender	Male/Female
3	Occupation	University professor, teacher, self-employed, retired, un-employed, physician, nurse, housewife, student, university student, legal jobs, employee, engineer, taxi driver, etc.
4	Education	Less than diploma, diploma, associate's degree, bachelor's degree, Master's degree, PhD
5	Major	Engineering, humanities, agricultural sciences, basic science, medicine, literature and language, and art
6	Familiarity with mapping	Yes/No
7	Being a mapping engineer	Yes/No
8	Familiarity with GIS	Yes/No
9	Familiarity with GPS	Yes/No
10	The way of acquaintance	Residential neighborhood, workplace neighborhood, coming and going route, other cases

Figure 2. Modified Map of Vanak Square

Developing Android Application

App Inventor is an open-source web-based programming environment that can be installed and run independent from other programs. In order to use App Inventor programming environment, only a Google account is needed. So, App Inventor was used to develop the android application in this study. The developed app first receives a username from each user that is designed to be unique and avoid saving repetitive usernames. The username does not impose any limitations and allows differentiating between users when data is stored. After the username is set, users are asked to respond questions in the questionnaire by choosing their answers from options provided (completing their profile). The users are then asked to select a region they are more familiar with to answer the relevant questions of that region. Once the questions are answered, the users are again asked to select another region. Each user is required to respond the questions of at least three regions. After the questions of each region are responded, the application removes the name of that region from the list. When users answer the questions of at least three regions, the exit button becomes active, and the users are allowed whether to exit from the application or to continue answering questions of more regions. Figures 3 and 4, respectively, illustrate the question page and answer sheet for Vanak region.

The users' answers must be stored in a virtual database. To store data in a virtual database in the App Inventor, a web-based virtual database must be created in Google Cloud Platform whose address can be used in the application. The created database and the procedure of storing data are demonstrated in Figure 5. Figure 6 presents the overall scheme of the application.

As mentioned before, collecting VGI data through ways other than famous websites such as OSM is difficult. Also, in this research we need the reliable information of participants. Therefore, the application designers asked the people they know to install and run it. It is worth mentioning that all of the participants are the residents of Tehran. Also, they had an option to choose the region they wanted to answer questions about, and they were not given the regions randomly. Thus, the study aims to evaluate their spatial knowledge not the level of their familiarity with Tehran. Although by this approach the volume and variety of data might not be too large, the most important advantage of it is that the background information collected from those individuals are trustworthy because they are known.

Figure 3. (a) Questionnaire answer page, (b) Selecting an option in registration page

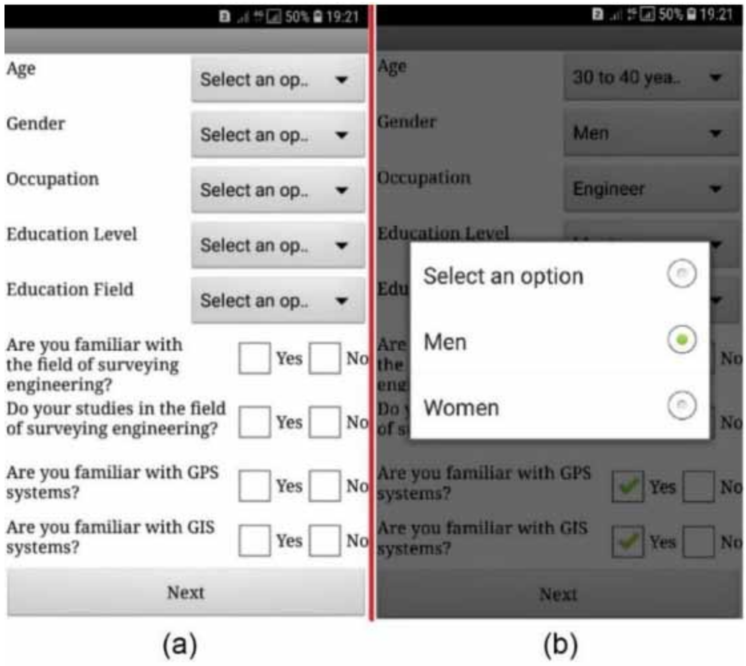
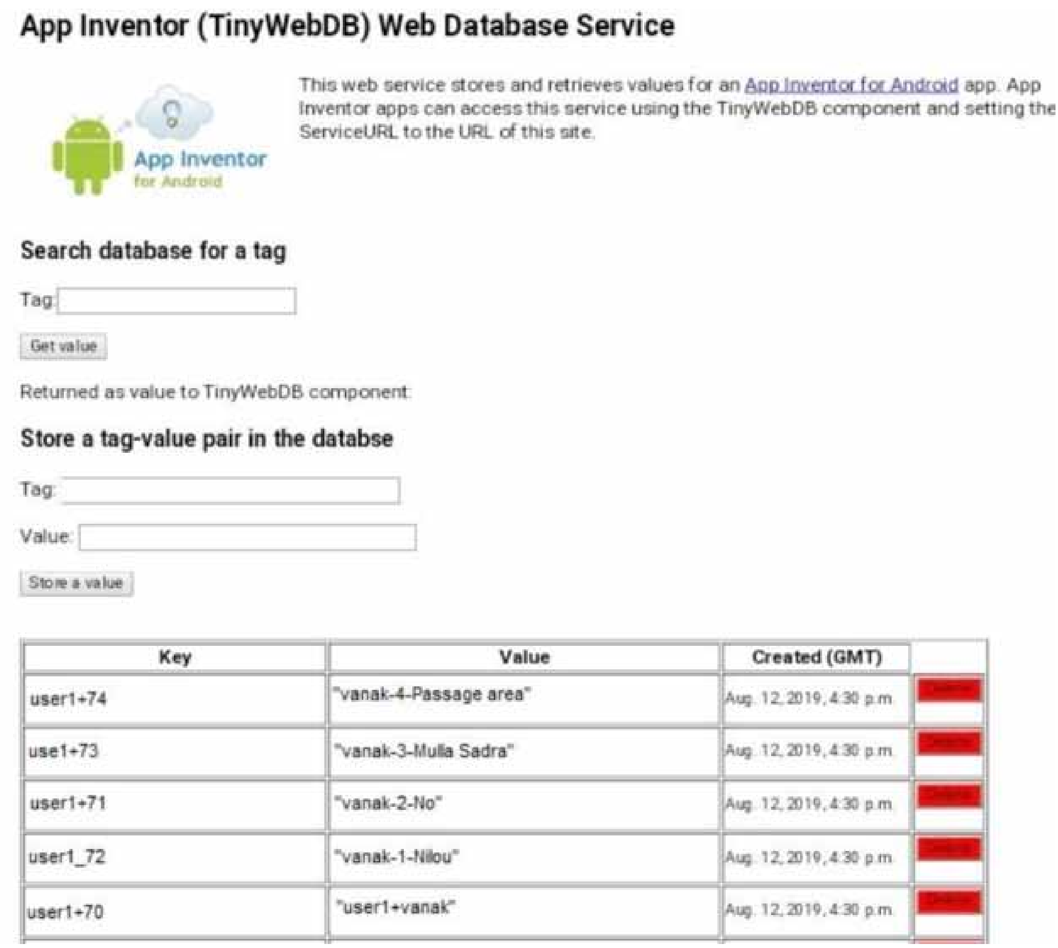


Figure 4. Answer Sheet for Vanak Region

Figure 5. Database of storing application data



Data Preparation

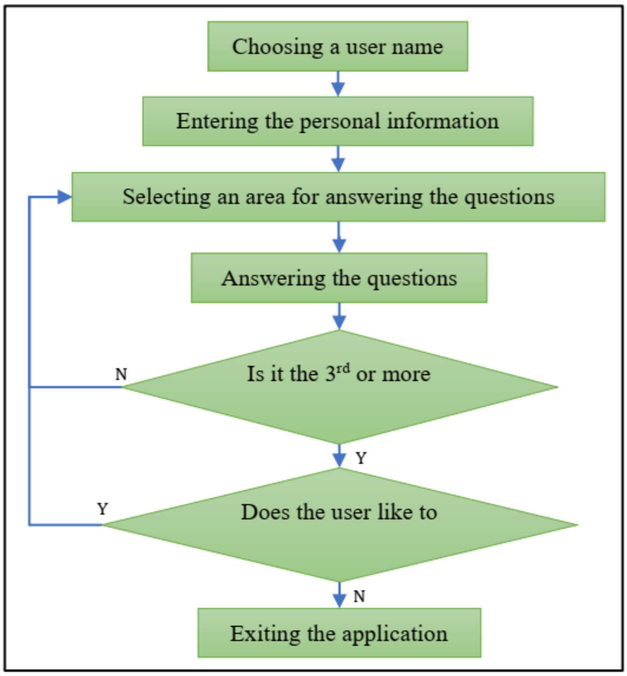
After being collected, the data must be prepared to be used in SPSS software and MATLAB Neural Network Toolbox. Therefore, data was first entered in Excel software, and sorted based on the usernames and selected regions. The users were then scored according to the number of questions they answered correctly. A correct answer was scored 1, an incorrect answer was scored -1, and the answer of “Don’t know” received the score of zero. The average of the scores were calculated and stored as the percentage of correct answers in a new column. Therefore, the scores could range from -100% to 100%, with 100% indicating the highest possible score, and -100% indicating the lowest possible score.

275 individuals participated in this project. Considering that each individual answered the questions of three regions on average, and since they had different reasons for being familiar with that region, which could affect users’ answers, the answers to the questions of each region can be considered as a separated sample. As a result, 1102 samples were collected, sorted, and processed.

Statistical Analyses

Once being prepared, the data entered into SPSS software. Before entering data into the software, the variables must be defined as well. Ten independent variables and one dependent variable of

Figure 6. Outline of the program



"the percentage of correct answers' were defined. After the data was analysed by SPSS software, the skewness and kurtosis values of the dependent variable were calculated to be -0.646 and -0.278, respectively, which fall within the range from -1 to 1, indicating that the data has normal distribution.

Internal consistency and reliability of the questions about "users" background in mapping" must be checked. Therefore, Cronbach's alpha test (de Vet et al., 2017) was performed for the following questions, "familiarity with mapping", "being a surveying Engineer", "familiarity with GIS", and "familiarity with GPS". The value of Cronbach's alpha coefficient for these four questions was 0.799, which is more than 0.7, indicating the reliability of the questions about "familiarity of users with mapping". Since the data is parametric and normally distributed, the Pearson correlation coefficients (Afyouni et al., 2019) between the dependent variable of "the percentage of correct answers" and other variables were calculated. Table 2 summarizes the results of Pearson correlation coefficients calculation. The Pearson correlation test shows that there is a relationship between "the percentage of correct answers" and "age" ($p=0.03$, $n=1102$, $r=0.1$), and the direction of the relationship is positive. Also, there is a relationship between "the percentage of correct answers" and "way of acquaintance with the region" ($p=0.00$, $n=1102$, $r=0.2$), and the direction of the relationship is negative. The relationship between these two variables is stronger.

In order to analyse the effect of "users' background in mapping" on "the percentage of correct answers", an independent t-test (Fişek & Barlas, 2013) was performed on the following variables, "familiarity with mapping", "being a surveying Engineer", "familiarity with GIS", and "familiarity with GPS". According to the results, the significance level for the relationship between "familiarity with mapping" and "the percentage of correct answers", the relationship between "being a surveying Engineer" and "the percentage of correct answers", the relationship between "familiarity with GIS" and "the percentage of correct answers", and the relationship between "familiarity with GPS" and "the percentage of correct answers" were 0.275, 0.141, 0.847 and 0.179, respectively, which are more than 0.05. As a result, it can be said that there is no significant relationship between these variables,

Table 2. Pearson correlation coefficient and significance level of the variable of correct answers percentage with other variables

Variable	The percentage of correct answers	
	Significance level	Pearson correlation coefficient
Age	0.03	0.1
Gender	0.45	0.02
Occupation	0.3	-0.03
Education	0.83	0.01
Major	0.13	0.05
Familiarity with mapping	0.04	0.03
Being a mapping engineer	0.54	0.02
Familiarity with GIS	0.19	0.04
Familiarity with GPS	0.54	0.02
The way of acquaintance	0.00	-0.2

and the hypothesis of relationship between "users' background in mapping" and "the percentage of correct answers" is rejected.

In addition, a multi-variable regression was conducted for research data, and the linear regression assumptions were taken into account. By using enter method, a significant model was obtained that can predict "the percentage of correct answers" with variance of 35%. Two variables of "age" and "way of acquaintance with the region" significantly predict "the percentage of correct answers", and other variables do not have any significant effects on "the percentage of correct answers". The regression model can be described by Equation 1 as follows:

$$\begin{aligned}&\text{The percentage of correct answers} = 49.516 + 17.758\ (\text{Age}) - 0.644\ (\text{Gender})\\&-2.748\ (\text{Occupation}) - 2.236\ (\text{Education}) + 11.909\ (\text{Major}) - 0.888\ (\text{Familiarity with mapping})\\&+ 0.759\ (\text{Being a surveying engineer}) + 3.065\ (\text{Familiarity with GIS})\\&+ 2.516\ (\text{Familiarity with GPS}) - 30.145\ (\text{The way of acquaintance})\end{aligned} \quad (1)$$

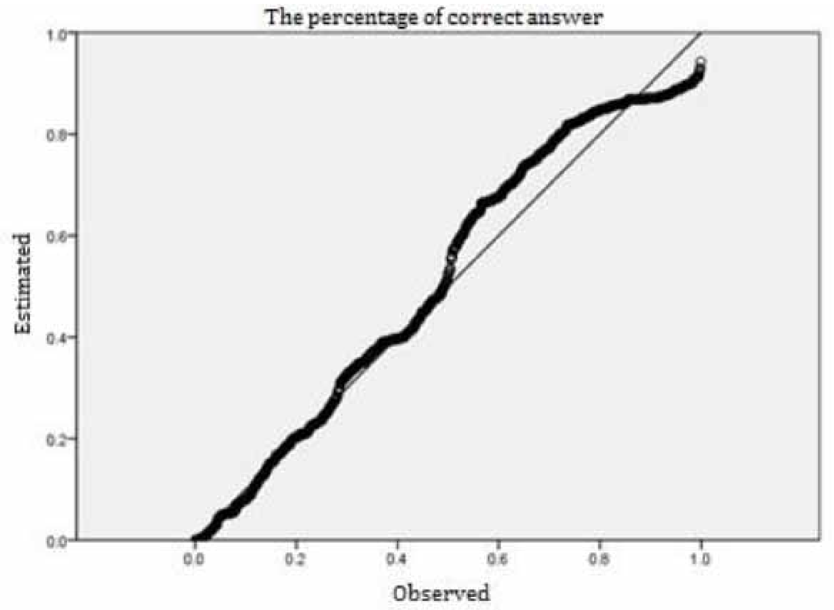
The graph of probability-probability of normal distribution of standard error is presented in Figure 7. As can be seen, the distribution of error is normal.

Implementation of ANN

Considering the type of data and previous researches, multi-layered feed-forward back-propagation Neural Networks were designed for this study. In these networks, in the first phase of training, the information moves forward. Then, the output values of the network are compared with real values to calculate the value of error and bias. In the next phase, the movement is backward in the network to updates weights and biases. This process is repeated until the errors are minimized and an optimized network is created. Once the network is trained, it can estimate the output value of new data.

In order to train the network, Batch Gradient Descent with Momentum training method was applied. Levenberg-Marquardt (LM) algorithm, Mean Squared Error (MSE), and Tangent Sigmoid function were chosen for training function, performance function, and activation function, respectively. Neural networks were designed for two cases: with one hidden layer and with two hidden layers. The number of neurons for each layer was 2 to 9, and the number of training cycles was set to 10. In other words, 80 and 640 training cycles were considered for the first (with one hidden layer) and second

Figure 7. The graph of probability-probability of normal distribution of standard error for the percentage of correct answers



(with two hidden layers) cases, respectively, which means the total number of training cycles were 720. Also, 70% of data were randomly used for network training, 15% for network validation, and 15% for network testing. In every training cycle, the composition of data used for network training, validation and testing is changed randomly. Also, the data was categorized in two groups of input data (10 variables of age, gender, etc) and target data (Percentage of correct answers).

In each training cycle for the network with specific architecture, the mean squared error (MSE) of the best validation performance was stored, and the average value of MSE was calculated for 2700 training cycles to find the best network architecture. Figure 8 shows the best validation performance graph for MSE calculation in one of the network training cases.

After 720 training cycles for different cases, the average value of MSE of the best validation performance was listed in Table 3. Then, according to the results, the best network architecture was selected.

RESULTS AND DISCUSSION

Once the best network architecture was obtained, a matrix with 322560 rows containing all possible permutations of input data were formed, and introduced to the best network architecture as an input. and the target output was obtained. Finally, the output and input data were sorted in order of the highest percentage of correct answers, and unreasonable cases were eliminated. For example, a mapping engineer aged less than 18 who has a diploma was considered an unreasonable case and eliminated. The best predicted case by the network is presented in Table 4.

Having these tools (trained ANN and statistical equation), when a new user wants to start entering or correcting/validating geographic information, if the background information of the user is available, the quality (mostly attribute quality) of his/her volunteered information can be estimated.

One of the important achievements of this research is discovering the relationships between the percent of the correct answers of the contributors and their specific backgrounds. The use of a mobile software in this study resulted in evaluating not only people's spatial knowledge but also their level of

Figure 8. The best validation performance graph for MSE calculation in one of the network training cases

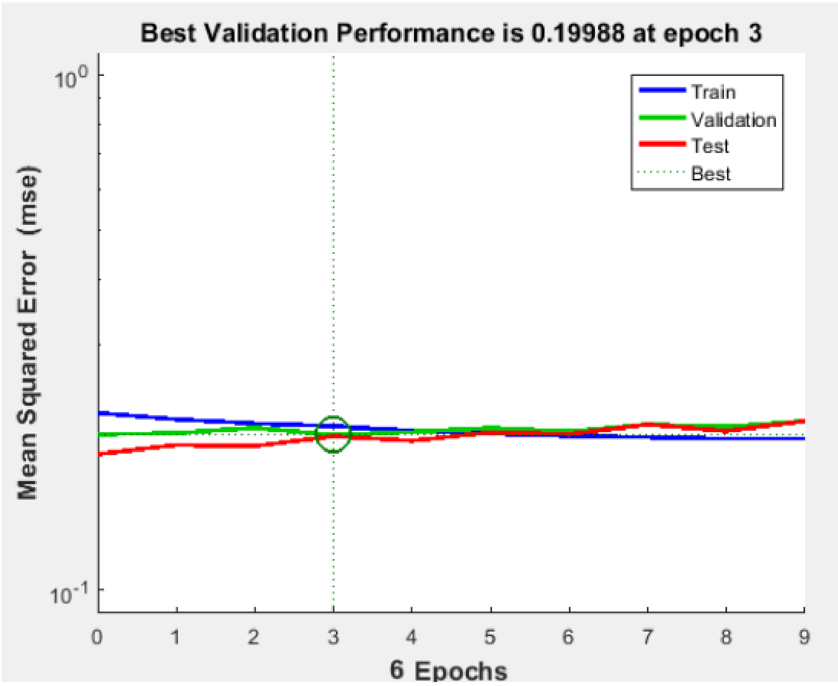


Table 3. The average value of MSE of the best validation in two different performance network architecture

Network architecture	One hidden layer	Two hidden layers
Number of neurons	8	7 - 4
Average network MSE in 10 training cycles	0.20932	0.19980

Table 4. The best predicted case by the best network and with the generated data

Variable	The best Value
Age	40-50 years old
Gender	Men
Occupation	Teacher
Education	Undergraduate
Major	Technical and engineering group
Familiarity whit mapping	Yes
Mapping engineer	No
Familiarity whit GIS	Yes
Familiarity whit GPS	No
The way of acquaintance	Workplace area
The percentage of correct answers	99.79%

expertise in applying online maps (both of which are essential skills for VGI contributors). Two of ten background factors showed to have major effects on contributors' correctness of answers, which are age and the way of acquaintance. Ages between 40-50 seems to be an appropriate age to be informed of the geographical information of surrounding environment. In addition, according to the results of the best predicted case of ANN, people who participated in this study paid the most attention to the geography of their workplace rather than even their place of residence. This may be odd; however, complimentary information could be useful such as how long they have lived or worked in a certain region. This should be investigated in further studies. The homogeneous frequency of the statistical society is another issue. In this study around 14%, 14%, 58% and 14% of contributors stated that the selected region was their workplace neighborhood, residential neighborhood, coming and going route and other cases, respectively. The equality between frequency of workplace neighborhood and frequency of residential neighborhood strengthens the reliability of the results in this case. Another discovery could be about the coming and going route. The results reveal that the coming and going route adds less information to the locational knowledge of people.

However, these are the results of a mixed process of the responses obtained from ANN rather than the direct observation of the statistics. The statistics of each input variable may also be considered separately but it is not a better criterion for evaluating the spatial knowledge of a single person. In other word, to estimate the spatial knowledge of an individual, his/her background data should be checked together.

According to the results, it can be argued that users' background played an important role in the percentage of correct answers given by the users to the questions related to the changes made in the map. What distinguishes this research from other similar researches is that in this study, the following items were considered simultaneously: developing an application to simulate VGI conditions; collecting different valid data about users' background; having correct answers to the map questions and comparing the participants' answers with them in order to obtain the percentage of correct answers; estimating the reliability of users' answers using both statistical and ANN approaches. The aim of this estimation is to have an evaluation of the reliability (quality) of the geographic data created by the volunteers. As a result, the map obtained from VGI will be more reliable or at least it may have one estimated index of reliability. In other words, through evaluation of VGI users, the obtained map can be validated. However, acquiring the trustable background information of participants is still a challenge. Although one cannot say that the ability to identify locations on a map with no labels means a person is a more accurate VGI contributor, it can be claimed that lack of this ability means that the person is likely an inappropriate contributor of VGI.

CONCLUSION AND RECOMMENDATIONS

In this research, level of expertise of VGI participants was evaluated through assessment of VGI participants' background information. Since the background information of users was not available, an android application was designed. To provide the geographic information of the application, some changes were made to the map of 17 regions of Tehran, and the users were asked to answer the questions about these changes. The application was installed by known participants and they respond the questions. In addition to these questions, users were asked to provide information about their background by answering ten questions about their age, gender, occupation, education, major, familiarity with mapping, being a surveying Engineer, familiarity with GIS, familiarity with GPS, and way of acquaintance with that region of the city. To have consistency among users' answers, multiple choice questions were designed. After being collected, the data was sorted and scored, and the percentage of correct answers to the map questions were calculated for each user. Overall, 1102 rows of data were formed. Ten background questions were considered as independent or input variables and "the percentage of correct answers' was set as dependent variable. Statistical and ANN approaches were applied to examine the correctness of answers based on the users' background information.

According to the findings of the statistical approach, there is a relationship between "the percentage of correct answers" and variables of "age" and "way of acquaintance with the region". Also, the results of t-test showed that there is no significant relationship between "the percentage of correct answers" and "users' background in mapping". In addition, by performing a multi-variable regression for all the data, a significant model was obtained that can predict "the percentage of correct answers" with variance of 35%. In the ANN approach, the collected data were used to train the multi-layered feed-forward back-propagation Artificial Neural Networks and to determine the best network architecture. Then, a matrix containing all possible permutations of input data were formed, and introduced to the best network architecture as an input. The input and output data were sorted in order of the highest percentage of correct answers, and unreasonable cases were eliminated.

Therefore, when a new user wants to start entering or correcting/validating geographic information, an estimation of the correctness of his/her volunteered information will be available. This research revealed the impact of VGI user's background information on the quality of their expressed information and as a result, their trustworthiness. Therefore, it is a good idea for VGI websites to capture background information of participants (if possible) before allowing them to enter/change any data.

Considering the previous studies and the methodology applied in this research as well as the obtained results, it is suggested that other variables including users' level of familiarity with the region, duration of familiarity and visit frequency to the region be taken into account in the future works. For questionnaire design, it is also recommended that answers be provided in the form of fuzzy membership.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENT

The authors would like to thank all of the VGI participants who helped us by installing the provided App on their cell phones and answering the questions.

REFERENCES

Afyouni, S., Smith, S. M., & Nichols, T. E. (2019). Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation. *NeuroImage*, *199*, 609–625. doi:10.1016/j.neuroimage.2019.05.011 PMID:31158478

Ali, A. L., & Schmid, F. (2014). Data quality assurance for volunteered geographic information. *International Conference on Geographic Information Science*. Springer. doi:10.1007/978-3-319-11593-1_9

Antoniou, V., & Skopeliti, A. (2015). Measures and indicators of VGI quality: An overview. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences, 2*, 345. 10.1007/978-3-319-11593-1_9

Astaburuaga, J., Martin, M. E., Leszczynski, A., & Gaillard, J. C. (2022). Maps, volunteered geographic information (VGI) and the spatio-discursive construction of nature. *Digital Geography and Society, 3*, 100029.

Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., & Rampini, A. (2014). On predicting and improving the quality of Volunteer Geographic Information projects. *International Journal of Digital Earth*, *9*(2), 134–155. doi:10.1080/17538947.2014.976774

Can, R., Kocaman, S., & Gokceoglu, C. (2019). A convolutional neural network architecture for auto-detection of landslide photographs to assess citizen science and volunteered geographic information data quality. *ISPRS International Journal of Geo-Information*, *8*(7), 300. doi:10.3390/ijgi8070300

Coleman, D. J., Georgiadou, Y., & Labonte, J. (2009). Volunteered Geographic Information: The Nature and Motivation of Produsers. *International Journal of Spatial Data Infrastructures Research*, *4*, 332–358.

Da Costa, J. N. (2016). Novel tool for examination of data completeness based on a comparative study of VGI data and official building datasets. *Geodetski Vestnik*, *60*(3), 495–508. doi:10.15292/geodetski-vestnik.2016.03.495-508

Dai, C., Lin, D., Bertino, E., & Kantarcioglu, M. (2008). An approach to evaluate data trustworthiness based on data provenance. Workshop on Secure Data Management, de Vet, H. C. W., Mokkink, L. B., Mosmuller, D. G., & Terwee, C. B. (2017). Spearman–Brown prophecy formula and Cronbach's alpha: Different faces of reliability and opportunities for new applications. *Journal of Clinical Epidemiology*, *85*, 45–49.

Feng, Y., & Sester, M. (2018). Extraction of pluvial flood relevant volunteered geographic information (VGI) by deep learning from user generated texts and photos. *ISPRS International Journal of Geo-Information*, *7*(2), 39. doi:10.3390/ijgi7020039

Fişek, M. H., & Barlas, Z. (2013). Permutation tests for goodness-of-fit testing of mathematical models to experimental data. *Social Science Research*, *42*(2), 482–495. doi:10.1016/j.ssresearch.2012.09.010 PMID:23347489

Fogliaroni, P., D'Antonio, F., & Clementini, E. (2018). Data trustworthiness and user reputation as indicators of VGI quality. *Geo-Spatial Information Science*, *21*(3), 213–233. doi:10.1080/10095020.2018.1496556

Forghani, M., & Delavar, M. R. (2014). A quality study of the OpenStreetMap dataset for Tehran. *ISPRS International Journal of Geo-Information*, *3*(2), 750–763. doi:10.3390/ijgi3020750

Gómez-Barrón, J. P., Manso-Callejo, M. Á., & Alcarria, R. (2019). Volunteered geographic information systems: Technological design patterns. *Transactions in GIS*, *23*(5), 976–1007. doi:10.1111/tgis.12544

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, *69*(4), 211–221. doi:10.1007/s10708-007-9111-y

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, *1*, 110–120. doi:10.1016/j.spasta.2012.03.002

Hosseinali, F., Alesheikh, A. A., & Nourian, F. (2014). Rapid Urban Growth in the Qazvin Region and Its Environmental Hazards: Implementing an Agent-Based Model. *Polish Journal of Environmental Studies*, *23*(3), 727–735.

Kessler, F. (2011). Volunteered Geographic Information: A Bicycling Enthusiast Perspective. *Cartography and Geographic Information Science*, *38*(3), 258–268. doi:10.1559/15230406382258

Klonner, C., Usón, T. J., Aeschbach, N., & Höfle, B. (2021). Participatory Mapping and Visualization of Local Knowledge: An Example from Eberbach, Germany. *International Journal of Disaster Risk Science*, *12*(1), 56–71. doi:10.1007/s13753-020-00312-8

Lin, W. (2018). Volunteered Geographic Information constructions in a contested terrain: A case of OpenStreetMap in China. *Geoforum*, *89*, 73–82. doi:10.1016/j.geoforum.2018.01.005

Martella, R., Clementini, E., & Kray, C. (2019). Crowdsourcing geographic information with a gamification approach. *Geodetski Vestnik*, *63*(2), 213–233. doi:10.15292/geodetski-vestnik.2019.02.213-233

Mohammadi, N., & Sedaghat, A. (2021). A framework for classification of volunteered geographic data based on user's need. *Geocarto International*, *36*(11), 1276–1291. doi:10.1080/10106049.2019.1641562

Moreri, K. K., Fairbairn, D., & James, P. (2018). Volunteered geographic information quality assessment using trust and reputation modelling in land administration systems in developing countries. *International Journal of Geographical Information Science*, *32*(5), 931–959. doi:10.1080/13658816.2017.1409353

Mullen, W. F., Jackson, S. P., Croitoru, A., Crooks, A., Stefanidis, A., & Agouris, P. (2015). Assessing the impact of demographic characteristics on spatial error in volunteered geographic information features. *GeoJournal*, *80*(4), 587–605. doi:10.1007/s10708-014-9564-8

Olszewski, R., & Wendland, A. (2021). Digital Agora – Knowledge acquisition from spatial databases, geoinformation society VGI and social media data. *Land Use Policy*, *109*, 105614. doi:10.1016/j.landusepol.2021.105614

Rajaram, G., & Manjula, K. (2019). Exploiting the Potential of VGI Metadata to Develop A Data-Driven Framework for Predicting User's Proficiency in OpenStreetMap Context. *ISPRS International Journal of Geo-Information*, *8*(11), 492. doi:10.3390/ijgi8110492

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. doi:10.1080/13658816.2016.1189556

Severinsen, J., de Roiste, M., Reitsma, F., & Hartato, E. (2019). VGTrust: Measuring trust for volunteered geographic information. *International Journal of Geographical Information Science*, *33*(8), 1683–1701. doi:10.1080/13658816.2019.1572893

Teimoory, N., Ali Abbaspour, R., & Chehreghan, A. (2021). Reliability extracted from the history file as an intrinsic indicator for assessing the quality of OpenStreetMap. *Earth Science Informatics*, *14*(3), 1413–1432. doi:10.1007/s12145-021-00675-6

van der Ham, I. J. M., & Claessen, M. H. G. (2020). How age relates to spatial navigation performance: Functional and methodological considerations. *Ageing Research Reviews*, *58*, 101020. doi:10.1016/j.arr.2020.101020 PMID:31954190

Xu, C., Wong, D. W., & Yang, C. (2013). Evaluating the "geographical awareness" of individuals: An exploratory analysis of twitter data. *Cartography and Geographic Information Science*, *40*(2), 103–115. doi:10.1080/15230406.2013.776212

Yang, A., Fan, H., & Jing, N. (2016). Amateur or professional: Assessing the expertise of major contributors in OpenStreetMap based on contributing behaviors. *ISPRS International Journal of Geo-Information*, *5*(2), 21. doi:10.3390/ijgi5020021

Zhang, D., Ge, Y., Stein, A., & Zhang, W.-B. (2021). Ranking of VGI contributor reputation using an evaluation-based weighted pagerank. *Transactions in GIS*, *25*(3), 1439–1459. doi:10.1111/tgis.12735