Geodesy and Cartography
ISSN 2029-6991 / eISSN 2029-7009
2020 Volume 46 Issue 2: 55–66
https://doi.org/10.3846/gac.2020.10486

UDC 528.1

FEASIBILITY OF USING GROUP METHOD OF DATA HANDLING (GMDH) APPROACH FOR HORIZONTAL COORDINATE TRANSFORMATION

Bernard KUMI-BOATENG*, Yao Yevenyo ZIGGAH

Department of Geomatic Engineering, Faculty of Mineral Resources Technology, University of Mines and Technology, Tarkwa, Ghana

Received 09 June 2019; accepted 03 June 2020

Abstract. Machine learning algorithms have emerged as a new paradigm shift in geoscience computations and applications. The present study aims to assess the suitability of Group Method of Data Handling (GMDH) in coordinate transformation. The data used for the coordinate transformation constitute the Ghana national triangulation network which is based on the two-horizontal geodetic datums (Accra 1929 and Leigon 1977) utilised for geospatial applications in Ghana. The GMDH result was compared with other standard methods such as Backpropagation Neural Network (BPNN), Radial Basis Function Neural Network (RBFNN), 2D conformal, and 2D affine. It was observed that the proposed GMDH approach is very efficient in transforming coordinates from the Leigon 1977 datum to the official mapping datum of Ghana, i.e. Accra 1929 datum. It was also found that GMDH could produce comparable and satisfactory results just like the widely used BPNN and RBFNN. However, the classical transformation methods (2D affine and 2D conformal) performed poorly when compared with the machine learning models (GMDH, BPNN and RBFNN). The computational strength of the machine learning models' is attributed to its self-adaptive capability to detect patterns in data set without considering the existence of functional relationships between the input and output variables. To this end, the proposed GMDH model could be used as a supplementary computational tool to the existing transformation procedures used in the Ghana geodetic reference network.

Keywords: coordinate transformation, machine learning, geodetic reference system, geodetic datum.

Introduction

Machine learning regression and classification techniques have recently been embraced and applied in various fields of geosciences. These methods have been touted by many scholars as a new paradigm shift in geoscientific computations (Angiuli et al., 2006; El-Assal et al., 2011; Ali et al., 2004). The same phenomenon has also been witnessed in geodesy where it has been implemented to solve different geodetic problems. Often, the researcher's objective is to investigate whether the machine learning methods can serve as a dependable alternative to the traditional procedures for solving geodetic problems. The literature shows that machine learning techniques can indeed be utilised. Some examples of notable geodetic application areas are included but are not limited to deformation studies (Samui, 2012; Li & Kong, 2014), tidal estimation (Okwuashi & Ndehedehe, 2017), change detection (Pal, 2009; Chang et al., 2010), geoid determination (Kavzoglu & Saka, 2005; Veronez et al., 2011), gravity field modelling (Turgut, 2016) and coordinate transformation (Tierra et al., 2008, 2009; Tierra & Romero, 2014). The last, coordinate transformation, is the main interest in the present study.

The application of machine learning as a regression technique to the transformation of coordinates between different geodetic datums is still an on-going research area that has gained wider coverage and attention of mathematicians, geodesists and geospatialists. The Artificial Neural Network (ANN) of backpropagation and radial basis function are the most prominent and widely used machine learning techniques for coordinate transformation. These techniques have regularly served as standard models for performance comparison with other transformation methods. Some notable ANN coordinate transformation studies are (Tierra et al., 2008, 2009; Tierra & Romero, 2014; Barsi, 2001; Zaletnyik, 2004; Lao-Sheng & Yi-Jin, 2006; Turgut, 2010; Gullu, 2010; Gullu et al., 2011; Konakoğlu & Gökalp, 2016; Konakoğlu et al., 2016). In these studies, the ANN results have been compared to the traditional transformation models like the three-parameter, standard

*Corresponding author. E-mail: *kumi@umat.edu.gh*

Copyright © 2020The Author(s). Published by VGTU Press

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Molodensky equation, Bursa-Wolf, Molodensky-Badekas, 2D conformal, and 2D affine. Review of the previous studies have shown that the ANN could transform coordinates with better precision and accuracy compared to the traditional methods. This is because the ANN could learn effectively on the model construction dataset (training) and adapt appropriately to the testing data thereby leading to good generalisation performance.

Other important advantages offered by machine learning over the traditional transformation techniques are the following:

- The machine learning techniques help avoid possible parameter estimation errors because they do not operate on the parameters;
- The machine learning techniques are able to correctly approximate the data set without prior knowledge of the underlying function; and
- The machine learning techniques can overcome ill-posed problems usually encountered by the traditional transformation equations.

The present authors, having noted the qualities of machine learning techniques also observed certain drawbacks, a typical example of which is the black box nature of the ANN. Additionally, the application of ANN requires a lot of human interference in terms of fine tuning the training model parameters to achieve global optimum. Moreover, the obtained accuracy and computational efficiency are deeply affected by the number of hidden layers and number of hidden neurons which are usually determined through sequential trial and error steps in the network (Huang & Babri, 1998; Huang et al., 2006a, 2006b). Therefore, it is important to explore other alternative machine learning procedures.

In view of the foregoing discussions, the Group Method of Data Handling (GMDH), a polynomial neural network offers some mathematical convenience and attractive features. This approach has the ability to overcome the black box system of the ANN by producing a mathematical relationship between the predictor and response variables. In addition, there is less manual task in the GMDH model formulation procedure as the fine tuning of neural network parameters, optimal model structure, number of layers and neurons in the hidden layers are automatically determined due to its self-organising nature. The technique is also highly efficient computationally as it does not require large amounts of training data sets to form the model and predict the outcome. This is by reason of its ability to automatically detect interpretable relationships in data and selects effective input variables. Besides, its computational procedure in the multi-layered architecture makes it suitable to perform multinomial of higher degree and can reveal complex relationships that other modelling techniques find challenging to expose (Malhotra & Chug, 2014). The GMDH method has therefore gained much popularity in computer science, engineering, geosciences, among other disciplines, for solving real life problems. Some important studies can be found in (Farlow, 1984; Sarycheva, 2003; Assaleh et al., 2013; Ayoub et al., 2015). The stated results from these studies have shown the strength and practicability of GMDH application.

The review of literature on the application of GMDH reveals that no study has been conducted to test the potential of GMDH in coordinate transformation. Furthermore, as far as the authors are aware, no work has been done to compare GMDH to the widely used ANN techniques (Backpropagation Neural Network (BPNN) and Radial Basis Function Neural Network (RBFNN)) in coordinate transformation research. Based on the GMDH reported mathematical conveniences, this paper explores the capability of GMDH as a coordinate transformation procedure in Ghana's geodetic reference network. Specifically, it studies the GMDH performance in practice for coordinate transformation and also compares GMDH efficiency to prevailing benchmark models of BPNN, RBFNN, 2D affine, and 2D conformal. This research is a contribution to the body of knowledge on coordinate transformation with GMDH as a novelty.

1. Brief description of the study area

Ghana is a country situated in the Western part of Africa. It shares borders with Togo, Ivory Coast, and Burkina Faso in the East, West and North directions respectively, and it is bordered with the Gulf of Guinea in the southernmost direction. Ghana has a total area coverage of approximately 238 540 km^2 and bounded between latitudes 4° 30' N and 11° N, and longitudes 3° W and 1° E (Fosu et al., 2006; Mugnier, 2000). The Ghana geodetic reference network popularly known as the golden triangle which covers the South most part of the country was used as the case study area (Figure 1).

Currently, two local geodetic reference frames are being utilised for geospatial applications in the country. These are the Accra 1929 datum which is based on the War Office 1926 ellipsoid and Leigon 1977 datum realised on the Clark 1880 (modified) ellipsoid. The War Office 1926 ellipsoid has semi-major axis (a) = 6378299.99899832 m, semi-minor axis (b) = 6356751.68824042 m and flattening (f) = 1/296. The Clark 1880 (modified), on the other hand, has the semi-major axis (a) = 6378249.145 m, semi-minor axis (b) = 6356514.870 m and flattening (f) = 1/293.465006079115. In Ghana, features on topographic maps and plans are presented in two-dimensional (2D) projected grid coordinate system of Easting and Northing based on the Transverse Mercator 1° NW.

2. Methods

This section discuss the basic theoretical concept of the proposed GMDH approach and the other widely used methods (BPNN, RBFNN, 2D affine, and 2D conformal) to perform coordinate transformation. Since the methods are extensively applied in literature, references have been provided for further details on their description.

2.1. Group method of data handling

The GMDH approach was proposed in Ivakhnenko (1966) and can be classified as a technique for identifying and modelling higher order non-linear systems, pattern recognition and function approximation. Ivakhnenko's motivation for the GMDH development was to eliminate the difficulty in identifying system variables which are pre-requisite in other mathematical models and are usually difficult to determine (Ivakhnenko, 1971). In such cases, field practitioners usually resort to guessing the variables which is often time consuming and can lead to unreliable prediction outcomes. Hence, the GMDH was developed to overcome such problems through it self-organising nature (AlBinHassan, & Wang, 2011). The GMDH is made up of feed forward multilayer network (Figure 2) of quadratic neurons that are used to map the input-output variables functional relationship. The technique is acknowledged to have good approximation capability, faster learning speed and converges to the optimal linear or nonlinear regression surface. This phenomenon is achievable because the GMDH uses optimization technique that automatically determines the optimal structure by a pruning process in a layer-by-layer fashion based on the mean squared error (MSE) criterion (Assaleh et al., 2013). The larger the MSE of the proceeding layer than the preceding layer, the system automatically stops adding layers. Here, the system select the lowest MSE component in the highest layer as it final model outcome. The GMDH has not been used to perform coordinate transformation. Therefore, its suitability as a coordinate transformation technique is investigated in this study. Detail mathematical background and theory about the GMDH method can be found in (Ivakhnenko, 1966, 1971).

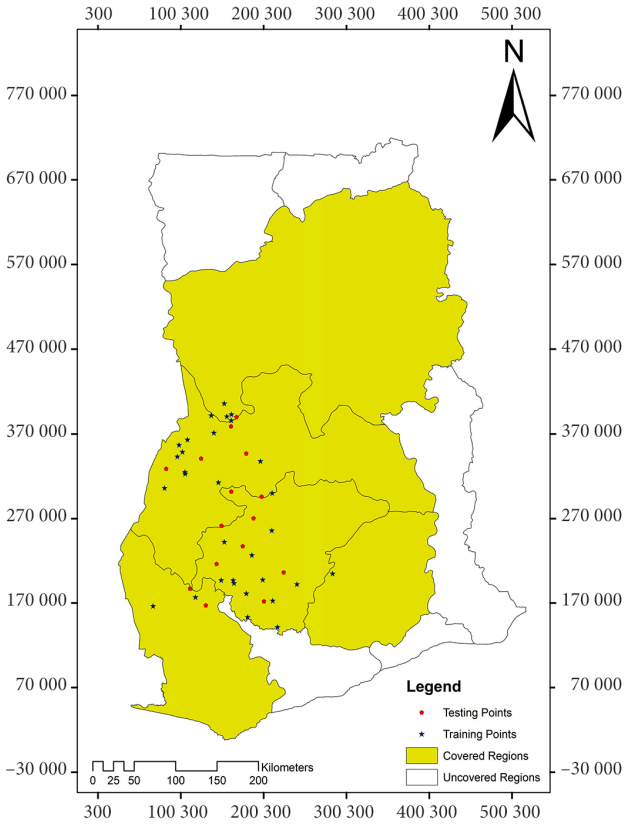


Figure 1. Study area showing distribution of selected training and testing points

2.2. Backpropagation neural network

The BPNN has been regarded as one of the standard and pioneered techniques of artificial neural network. Over the

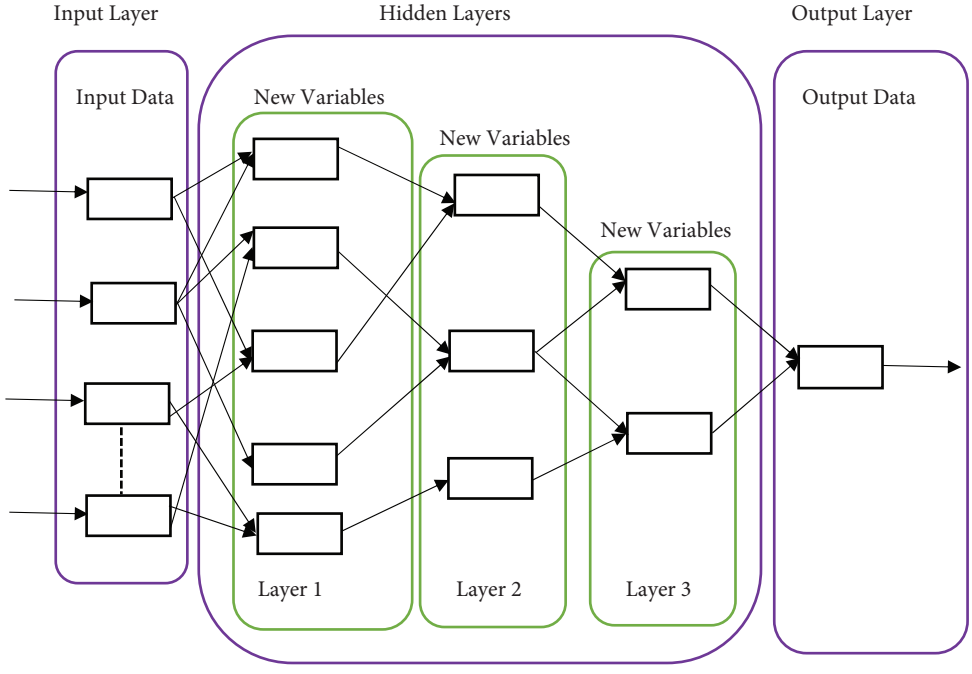


Figure 2. Basic GMDH structure

years, the BPNN capability to perform coordinate transformation has been carried out in several jurisdictions having different geodetic network infrastructure (Turgut, 2010, 2016; Barsi, 2001; Zaletnyik, 2004; Lao-Sheng, & Yi-Jin, 2006; Konakoğlu, & Gökalp, 2016; Konakoğlu et al., 2016; Elshambaky et al., 2018). Generally, the BPNN scheme is made up of input, one or more hidden layers and output layer interconnected together in a feed forward manner (Figure 3). The major role played by the input layer is to accept data from the outside settings. The received data is then directly transferred to the hidden layer where weights are randomly assigned to the data using a training algorithm. As a parallel comparison to the nonlinearity of neurons found in human brains, the hidden layer utilises an activation function (e.g. hyperbolic or log-sigmoid) to process its data. The processed hidden layer results are transmitted to the output layer and randomly assigned weights for further processing. Usually, for function approximation works as witnessed in this study, the linear regression transfer function mostly used in the output layer was adopted. This transfer function will help in converting the results obtained from the hidden layer onto a regression surface. If the differences between the predicted outcomes and the anticipated outputs is not at a minimum, the calculated error will be returned through the network from the output layer to the input layer. Here, the connection weights are changed based on learning rule to minimise the prediction error. The process continues until an acceptable prediction error is achieved. Detailed comprehensive treatment on the BPNN can be found in (Bishop, 2006; Haykin, 1999).

2.3. Radial basis function neural network

The RBFNN has a similar feed forward structure like the BPNN. This technique has also been applied to carry out coordinate transformation (Tierra et al., 2008, 2009; Tierra & Romero, 2014; Gullu, 2010; Gullu et al., 2011). The RBFNN architecture consists of the input, single hidden and output layers connected together (Figure 4). These layers perform similar functions as described for the BPNN. However, there exist some differences in the computational procedures between the BPNN and RBFNN. It is noteworthy that in the RBFNN, the input data sets are transmitted directly to the hidden layer chamber without weighting. Hence, it can be seen from Figure 4 that there is no weight matrix between the input and hidden layer. However, the transmitted data from the input layer to the hidden layer is processed using a radial basis activation function such as the Gaussian type. Here, a two-stage procedure is usually employed as a means of defining an appropriate RBFNN model. The first phase is to determine the centres of the basis function and the width parameter. These centres can be determined automatically using methods such as clustering, gradient descent and least squares insofar as a fixed value of the width parameter of the radial basis function is known. The second phase demands computing the interconnected weights between the hidden and output layers. This is normally accomplished by applying the pseudo-inverse least squares method popularly referred to as Moore-Penrose pseudo inverse. Information on the mathematics and theory governing the RBFNN can be found in Bishop (2006), Haykin (1999), Orr (1996), Chen et al. (1991).

2.4. 2D conformal transformation model

The 2D conformal transformation is noted to be a four-parameter similarity model. The parameters to be determined include two translations of the origin, a rotation and scale factor. The role of the translation parameters is

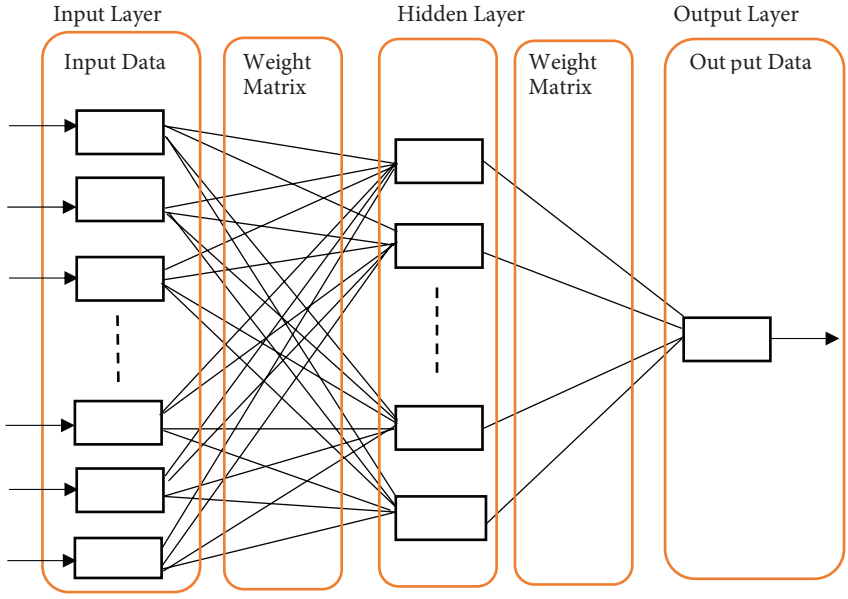


Figure 3. Basic BPNN structure

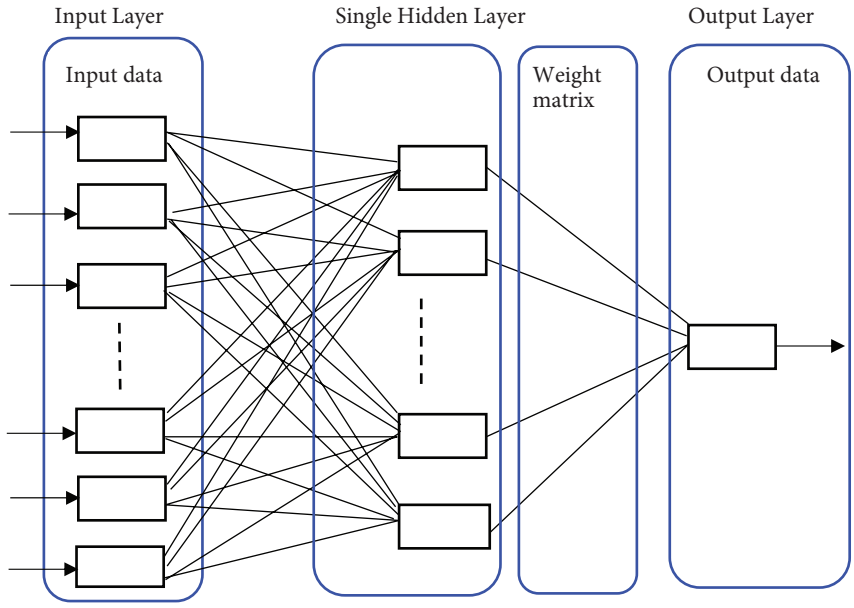


Figure 4. Basic RBFNN structure

to provide the degree of displacement between the origins of the two reference datums. The scale factor, on the other hand, creates equal dimension in the reference axes while, the rotational parameter depicts the parallelism of the reference axes. The 2D affine model has been widely and successfully applied to transform coordinates from satellite image to cadastral map. Others have also applied it to transform coordinates between different geodetic datums. Examples of such works can be found in Tierra and Romeo (2014), Lao-Sheng and Yi-Jin (2006), Konakoğlu and Gökalp (2016), Konakoğlu et al. (2016), Al-Ruzouq and Dimitrova (2006), Deakin (2007), Fotiou and Kaltsikis (2016), Dönmez and Tunc (2016), Zeng (2014) and references therein. To compute the transformation parameters, the least squares method is applied on the observational equations (Ghilani, 2010).

2.5. 2D affine model

The 2D affine model is classified as a six parameter similarity model. The name is analogous to the number of unknown parameters to be determined. The parameters comprise of a rotation about the origin, two translations and two scale factors in the x and y-direction as well as the lack of orthogonality correction between the reference axes. Research works on coordinate transformation have been carried out using the 2D affine model (Lao-Sheng & Yi-Jin, 2006; Sisman, 2014; Zhang et al., 2016). The least squares approach is applied to the set of observation equations to determine the transformation parameters. To compute the transformation parameters, the least squares method is applied on the observational equations (Ghilani, 2010).

3. Model development

3.1. Data acquisition and processing

The data utilised for this work was provided by the Ghana Survey and Mapping Division (SMD) of Lands Commission from the Land Administration Project II financed by the World Bank. A total number of 46 common reference points (Figure 1) which forms part of the Ghana national triangulation network for the Accra 1929 and Leigon 1977 datums have been used in this study. The entire dataset collected are in 2D projected grid coordinate system (Easting and Northing) as shown in Table 1.

The acquired data (Table 1) was used to develop the GMDH model as well as other investigated methods (BPNN, RBFNN, 2D affine, and 2D conformal). Although the official reference frame accepted by the SMD for maps and plan representation is the Accra 1929 datum, the two datums are being used simultaneously. There is therefore the need to perform coordinate transformation from the Leigon 1977 datum to the Accra 1929 datum. Because of that, the input variables chosen for the model building was based on the Leigon 1977 datum designated as (E_{Leigon}, N_{Leigon}). The Accra 1929 datum was selected as the output variables denoted as (E_{Accra}, N_{Accra}).

To apply the methods, the data was partitioned into two different sets. This was done by dividing the entire data into training and testing sets. The training and testing portions were selected to be purposefully distributed in the study area so that the tests carried out are not hindered from statistical meaning. Therefore, visual observation of Figure 1 shows that the selected training data covers a larger area of the study sites considered while the test datasets are found inside the training data area. Hence, the

Table 1. Common points between Leigon 1977 and Accra 1929 datums. The CLARK_E and CLARK_N are the coordinates in the Leigon 1977 datum while the WAR_E and WAR_N are the coordinates in the Accra 1929 datum. The point ID R1 to R31 is the training data and T1 to T15 is the testing data

Point ID	CLARK_E (metres)	CLARK_N (metres)	WAR_E (metres)	WAR_N (metres)
R1	96308.47472	343341.49567	96315.96670	343342.49846
R2	102396.73609	348793.77721	102404.08481	348794.59407
R3	98384.46857	357002.14304	98391.82949	357002.79531
R4	108527.33719	363356.54637	108534.39940	363357.18949
R5	179593.84552	181252.23939	179598.74366	181251.11773
R6	105410.76018	324978.11495	105418.16987	324979.30977
R7	145906.45533	312862.32965	145913.93512	312863.29282
R8	217305.38737	141614.96995	217310.64517	141612.59251
R9	105737.94133	323098.99428	105745.36321	323100.32320
R10	161720.22307	392995.67524	161727.17860	392996.58355
R11	161317.32358	385896.81692	161324.16329	385897.71913
R12	196537.92640	337894.67918	196545.33608	337896.21233
R13	164804.79001	193489.59406	164809.62414	193488.75586
R14	153010.64460	405832.84525	153017.16732	405833.79013
R15	149300.93347	196924.39110	149305.45974	196923.63519
R16	181138.36622	153293.64504	181143.18815	153291.73394
R17	211446.81392	172735.17242	211452.26069	172733.26437
R18	210262.37140	256084.33475	210269.60125	256083.42035
R19	156103.67278	390525.41992	156110.53992	390526.21240
R20	140426.63819	371486.73412	140433.33159	371487.39859
R21	66995.11272	166458.22784	66997.02687	166456.60630
R22	163685.79714	196883.82530	163690.61907	196883.06635
R23	186272.81482	226968.02576	186278.87729	226967.29119
R24	153046.17206	242613.60198	153052.11870	242614.12928
R25	199279.75217	197449.17694	199285.24161	197447.96689
R26	137145.43581	391777.91507	137151.97072	391778.46371
R27	118155.82277	176791.46216	118159.51389	176790.37402
R28	80820.82438	306190.56582	80828.61201	306191.85512
R29	210755.75200	300209.21400	210763.86141	300210.03286
R30	283778.08500	204680.01600	283785.99819	204677.95919
R31	240727.35400	192138.14000	240734.20698	192136.52110
T1	82660.60151	329013.92825	82668.25503	329015.13830
T2	198067.01204	296216.39775	198074.93379	296216.90067
T3	175114.74373	237505.14060	175120.91287	237505.09488
T4	188231.00359	270605.93157	188238.04446	270605.90413
T5	200781.58871	171978.38218	200786.91661	171976.62043
T6	224664.95918	206357.89630	224671.11003	206356.31439
T7	111533.39991	187113.71454	111536.93559	187113.00131
T8	143581.25419	216569.24427	143585.68903	216569.09187
T9	124931.98050	341128.18040	124939.24083	341129.32340
T10	130558.08321	167316.40970	130562.14315	167315.11735
T11	149600.00294	261798.37828	149606.27267	261799.17380
T12	161269.90589	302076.12944	161276.89800	302076.89754
T13	179379.85784	347155.69824	179387.14865	347157.03631
T14	161035.63836	379092.87365	161042.53750	379093.78195
T15	167630.72830	389866.04367	167637.56801	389866.95502

test data selected will aid in ascertaining the strength of the proposed GMDH technique in the different areas of interest that is, five out of ten regions in Ghana (see Figure 1). The partition percentages used for the data division was 67 and 33 corresponding to training and testing sets, respectively. These data division percentages followed the hold-out cross-validation technique whereby the modeller uses a higher portion of the data to build the model and the rest for testing. According to the selected partition percentages, 31 common reference points out of the 46 was set aside for the model training, while 15 data sets were utilised to provide an objective independent assessment of the selected optimum trained model.

Data normalization was then carried out on the training and testing datasets. This step was important since the machine learning algorithms applied are not scale invariant. Therefore, scaling the data set will prevent variables with higher variance to have minimal influence in the transformation process. In so doing, having less effect on the final transformed results produced by the said algorithm implemented. Hence, the machine learning algorithms will be prevented from being skewed in the direction of the dominant variables. Equation (1) (Muller & Hemond, 2013) was used to normalise the data into the range [–1, 1].

$$N_i = N_{\min} + \frac{(N_{\max} - N_{\min}) \times (O_i - O_{\min})}{(O_{\max} - O_{\min})}, \quad (1)$$

where N_i denotes the scaled data, O_i is the observed coordinates, $O_{\min}$ and $O_{\max}$ represent the minimum and maximum range of the observed coordinates with $N_{\max}$ and $N_{\min}$ values set at 1 and –1.

3.2. Proposed GMDH model

This study applied the supervised learning technique to train the GMDH model. The GMDH has a feedforward structure comprising of input, hidden and output layers. The GMDH model construction started by choosing the input variables (E_{Leigon}, N_{Leigon}) to generate the first layer of the network, with each successive layer drawing its input from the previous layer. Here, the interconnectedness between the network neurons are not stationary, but are chosen automatically during training to optimize the network. The number of layers to be used in the network were also selected automatically to avoid overfitting leading to better prediction performance. These listed attributes demonstrate the self-organising nature of the GMDH technique. In the training process, all the input variables (E_{Leigon}, N_{Leigon}) contribution to the successful prediction of the outputs (E_{Accra}, N_{Accra}) were automatically evaluated and the highly significant ones selected. It was noticed in this study that all the two input parameters (E_{Leigon}, N_{Leigon}) were used in the final GMDH models to predict (E_{Accra}, N_{Accra}). This can clearly be seen from the final optimal GMDH models presented in Equations (2) and (3) with their corresponding coefficients. It must be known here that both models developed consisted of single layer and one neuron. MATLAB open source code (Jekabsons, 2010) was modified and used to develop the GMDH model.

$$E\text{------} = a_0 + a_1 N \quad + a_2 E \quad - a_3 E_{Leigon} \times N_{Leigon} - a_4 (N_{Leigon})^2 + a_5 (E_{Leigon})^2, \quad (2)$$

where $a_0, a_1, \ldots, a_5$ are the coefficient of the polynomial function (Equation (2)) given as:

$$a_0 = 175393.555441; \quad a_1 = 1.4703513155;$$
$$a_2 = 108392.543566; \quad a_3 = 2.07085915916;$$
$$a_4 = 2.06485603798; \quad a_5 = 0.479758488881;$$

$$N_{Accra} = b_0 + b_1 N_{Leigon} - b_2 E_{Leigon} + b_3 E_{Leigon} \times N_{Leigon} - b_4 (N_{Leigon})^2 - b_5 (E_{Leigon})^2, \quad (3)$$

where $b_0, b_1, \ldots, b_5$ are the coefficient of the polynomial function (Equation (3)) given as:

$$b_0 = 273724.416306; \quad b_1 = 132110.475472;$$
$$b_2 = 0.604555734093; \quad b_3 = 0.305039778827;$$
$$b_4 = 1.32943688503; \quad b_5 = 0.610729049284.$$

4. Evaluation of model performance

Prediction accuracy analysis of the developed GMDH model against other investigated models (BPNN, RBFNN, 2D affine, and 2D conformal) was conducted to determine whether it is physically correct. This was done by applying the following accuracy measures (Equations (4) to (8)) which have also become de facto standards to check the transformation performance of a model.

Residual Horizontal Distance (RHD):

$$RHD = \sqrt{(E_{O_i} - E_{P_i})^2 + (N_{O_i} - N_{P_i})^2} = \sqrt{\Delta E_i^2 + \Delta N_i^2}. \quad (4)$$

Average Residual Horizontal Distance (A_{RHD}):

$$ARHD = \frac{1}{N} \sum_{i=1}^{N} RHD_i. \quad (5)$$

RMSE Residual Horizontal Distance ($RMSE_{RHD}$):

$$RMSE = \sqrt{\frac{\sum (RHD_i)^2}{N}}. \quad (6)$$

Maximum Residual Horizontal Distance (Max Error):

$$Max\ Error = \max(RHD)_{i=1}^{N}. \quad (7)$$

Standard Deviation (SD) of the Residual Horizontal Distance (SD_{RHD}) :

$$SD_{RHD} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (RHD_i - \overline{RHD})^2}. \quad (8)$$

Here, (E_{O_i} , N_{O_i}) are the measured Easting and Northing coordinates and (E_{P_i} , N_{P_i}) are its corresponding predicted coordinates. $\overline{RHD}$ is the mean horizontal residual distance.

5. Results and analysis

5.1. GMDH model results

The practicality of any developed model is subject to how close its predicted outputs fit well with the measured data. Therefore, employing statistical quantitative methods allow for an objective appraisal of the model prediction capability. In line with this, the present study carried out statistical error analysis as indicated in Table 2 to assess the proposed GMDH model prediction accuracy for both training and testing sets respectively.

Table 2. Summary Statistics of the Proposed GMDH Results (unit: metres)

Statistical Parameter	Training	Testing
ARHD	0.388	0.385
RMSE	0.504	0.468
SD_{RHD}	0.246	0.275
Max Error	0.924	0.971

In Table 2, the A_{RHD} values presented was used as the average standardised measure of the horizontal positional discrepancy between observed coordinates and predicted GMDH coordinates. In line with the results (Table 2), it was observed that the GMDH model shifted averagely about 0.39 m from both training and testing horizontal positional data sets. Furthermore, to ascertain how well the transformations fit the original data, the $RMSE_{RHD}$ statistic was utilised. It was noticed that the GMDH model could fit approximately 0.5 m to the model's data set (training and testing). The SD_{RHD} (Table 2) values signifies a high measure of precision of the GMDH predictions and thus explain the extent they differ from the average residual horizontal distance. The maximum residual values (Table 2) presented describe the size of the residual horizontal distances. The achieved GMDH results agreed comparably with some relevant coordinate transformation studies discussed in literature pertaining to Ghana's geodetic reference network (Ayer, 2008; Ayer & Fosu, 2008; Ayer & Tiennah, 2008; Poku-Gyamfi, 2009; Dzidefo, 2011; Ziggah et al., 2013a, 2013b, 2016a, 2016b, 2017; Annan et al., 2016; Laari et al., 2016). It is therefore concluded that the proposed GMDH model is valid for performing coordinate transformation in the Ghana geodetic reference network. Besides, the GMDH has demonstrated good generalisation capabilities.

5.2. Comparison of GMDH model and other investigated techniques

After the GMDH approach was applied to perform coordinate transformation, the same training and testing dataset was used for the BPNN, RBFNN, 2D conformal and affine. On the classical techniques, the transformation parameters derived based on the least squares principle and their corresponding standard deviations (SD) for the 2D affine, and 2D conformal models are shown in Tables 3 and 4.

Table 3. 2D affine model transformation parameters from Leigon to Accra datum (unit: metres)

Parameters	Values	SD
m	1.00001260	2.38E-06
n	0.00001363	1.36E-06
o	0.5414654	0.6367
p	–5.39E-06	2.38E-06
q	1.0000114	1.36E-06
r	–2.3336951	0.6367

Table 4. 2D conformal model transformation parameters from Leigon to Accra datum (unit: metres)

Parameters	Values	SD
a	1.00001	1.13E-06
b	–0.00001	1.13E-06
c	1.56449	0.37599
d	–1.10457	0.37599

The BPNN model developed contained one hidden layer, input and output layers. This decision to use BPNN with one hidden layer was in line with the conclusion made in Hornik et al. (1989) that such a network could correctly approximate any complex function to the desired results. Due to non-linearity in the data set which was introduced into the network, the BPNN was established using hyperbolic tangent and linear activation functions for the hidden and output layers. Conversely, the RBFNN consisted of input, single hidden and output layers. In the hidden layer, the most generic radial basis function (RBF) known as the Gaussian activation function was applied in this study to process the training data. This type of RBF is highly characterized by a centre position and a width parameter which regulates the amount of decrease of the function during training process. The linear regressor transfer function was used in the RBFNN output layer. The optimum trained models for both BPNN and RBFNN was selected based on the structure that produced the lowest mean squared error. Therefore, the optimum RBFNN structure was [2-18-2] meaning two inputs (E_{clark}, N_{clark}), 18 hidden neurons and two outputs (E_{war}, N_{war}). Accordingly, the BPNN had [2-8-1] for the output E_{war} and [2-11-1] for the output N_{war} . This implies that, the BPNN model for predicting the outputs E_{war} and N_{war} contains two inputs with eight and eleven hidden neurons.

Schematic comparison of GMDH, BPNN and RBFNN (Figures 2, 3 and 4) indicate that the three networks are mostly structured in a feedforward manner with combinations of interconnected neurons. Despite their structural similarity, the neurons found in the GMDH network

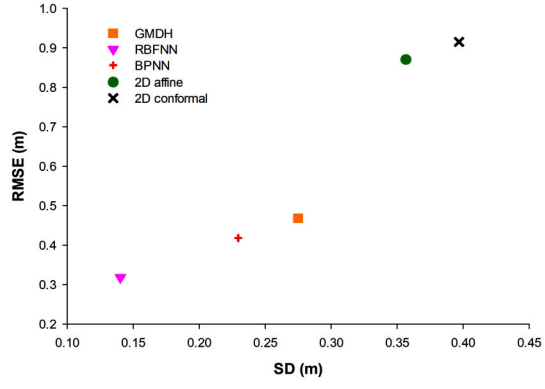
(Figure 2) are not fully connected as compared to the BPNN (Figure 3) and RBFNN (Figure 4) structures. Also, in the GMDH topology, regression equations are formed from each pair of input data. The output results produced by the regression equations becomes the new inputs variables to the next layer. The final GMDH result comprises of selected regression equations from all layers. The GMDH uses quadratic polynomials as the transfer function. By contrast, the RBFNN perform computations using Gaussian and linear regression transfer function in the hidden and output layers. The RBFNN uses only a single hidden layer unlike the GMDH where several layers are considered. The final results from the RBFNN is based on re-computing of the hidden layer outputs in the output layer using the linear regressor function. For the BPNN, two operations of summation and activation is made up of each perceptron. Unlike the GMDH and RBFNN, the BPNN trains the weights between input and hidden layers as well as between the hidden and output layers using a backpropagation algorithm.

For numerical comparison of the proposed GMDH approach to the other investigated methods, the $RMSE_{RHD}$ values (Table 5) have been plotted against the SD_{RHD} (Table 5) for each model, as shown in Figure 5a.

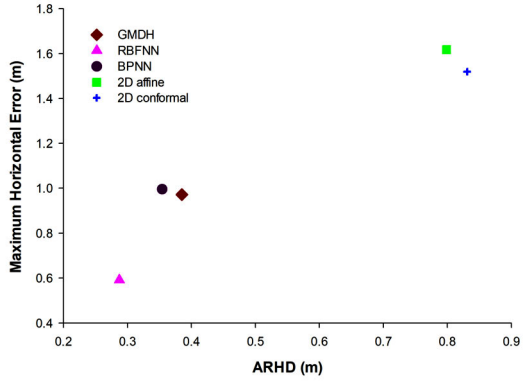
Table 5. Summary statistics of the test results for all the methods (unit: metres)

Model	ARHD	RMSE	SD_{RHD}	Maximum
GMDH	0.385	0.468	0.275	0.971
RBFNN	0.287	0.318	0.140	0.591
BPNN	0.354	0.418	0.230	0.995
2D affine	0.799	0.870	0.357	1.618
2D conformal	0.831	0.915	0.397	1.519

Interpretation of Figure 5a indicates that a very good model will be found at the lower left corner, which will be seen as the intersection between the lower values of the RMSE and SD_{RHD}. The inference made from Figure 5a showed that the proposed GMDH results deviate marginally from the standard methods (RBFNN and BPNN) because they are all concentrated within the same left corner region than the classical approaches (2D affine and 2D conformal). This can be confirmed from the quantitative values presented in Table 5 where GMDH had 0.468 m and 0.275 m for RMSE and SD while RBFNN achieved 0.318 m and 0.14 m. The BPNN produced 0.418 m and 0.230 m in that regard. Furthermore, the ARHD values (Table 5) were plotted against the maximum horizontal residuals, as presented in Figure 5b. Again, the model that could produce satisfactory results will be located at the lower left corner which is indicated by the intersection of the lower values of both statistical parameters considered. Therefore, visual observation of Figure 5b showed that the GMDH approach fall within the same region as the BPNN with the RBFNN slightly better. The 2D affine, and 2D conformal models performed poorly in that regard. This clearly denotes that the computational flexibility lacking in classical transformation equations as a result of its compactness and fixed nature has contributed to their inadequacy to approximate to the underlying "true" functional relationship of the data set. Hence, the self-adaptive learning style of the machine learning techniques (GMDH, RBFNN and BPNN) was able to approximate across the data much better. In comparison, looking at Table 5 and Figure 5a, b the RBFNN was the better among all the models. However, from a practical perspective the tested GMDH model as indicated earlier could provide results that meet cadastral survey accuracies in Ghana. This is because the SD results of 0.275 m produced by GMDH is an acceptable precision adequacy that falls within the 0.354 to 0.446 m reported in Ayer and Fosu (2008) for Ghana geodetic reference network. Moreover, the maximum horizontal positional error tolerance of approximately ±1 metre for cadastral surveying works in Ghana has been met by the GMDH model which produced 0.971 m. Hence, by comparison, the results attest that the tested GMDH can serve as a supplementary computational intelligent tool that can be used to perform coordinate transformation in the Ghana geodetic reference network.



a)



b)

Figure 5. Statistical error analysis for the testing results

Conclusions

The GMDH approach has been applied to formulate a model for transforming coordinates successfully in the Ghana geodetic reference network. The tested GMDH method offered distinctive mathematical benefits in the data processing by evaluating and selecting automatically the most significant input parameters based on their contribution to the final outcome. Hence, there is minimal human involvement in the model building process. In this study, the statistical accuracy measures used have shown that in Ghana the GMDH approach could produce satisfactory fitting error compared with the benchmark machine learning methods (BPNN and RBFNN) investigated. Thus, the achieved GMDH results described in this work corroborated with coordinate transformation studies reported in the literature with respect to Ghana geodetic reference network. On the basis of the findings, the GMDH model has shown encouraging application potential in performing coordinate transformation, and has far-reaching significance of being a supplementary method to the geospatial and built environment practitioners in Ghana.

Acknowledgements

We appreciate the data availability courtesy of the Ghana Survey and Mapping Division of Lands Commission.

Funding

There is no funding information for this research work.

Author contributions

BKB conceived the research concept and design, data analysis and interpretation, and writing of the paper. YYZ collected and compiled the data, developed the machine learning algorithms in MATLAB program and carried out critical revision of the paper from the draft stage to the final output.

Disclosure statement

The authors have no competing financial, professional, or personal interests from other parties.

References

AlBinHassan, N. M., & Wang, Y. (2011). Porosity prediction using the group method of data handling. *Geophysics*, *76*(5), O15–O22. https://doi.org/10.1190/geo2010-0101.1

Ali, M. M., Swain, D., & Weller, R. A. (2004). Estimation of ocean subsurface thermal structure from surface parameters: A neural network approach. *Geophysical Research Letters*, *31*, 1–4. https://doi.org/10.1029/2004GL021192

Al-Ruzouq, R., & Dimitrova, P. (2006, October 8–13). Photogrammetric techniques for cadastral map renewal. In *Shaping the Change XXIII FIG Congress* (pp. 1–15). Munich, Germany.

Angiuli, E., Del Frate, F., & Salvatori, L. (2006). Neural networks for oil spill detection using ERS and ENVISAT Imagery. In *Proceedings of SEASAR, European Space Agency*, (pp. 1–6). Italy.

Annan, R. F., Ziggah, Y. Y., Ayer, J., & Odutola, C. A. (2016). A hybridized centroid technique for 3D Molodensky-Badekas coordinate transformation in the Ghana geodetic reference network using total least squares approach. *South African Journal of Geomatics*, *5*(3), 269–284. https://doi.org/10.4314/sajg.v5i3.1

Assaleh, K., Shanableh, T., & Kheil, Y. A. (2013). Group method of data handling for modeling magnetorheological dampers. *Intelligent Control and Automation*, *4*(1), 70–79. https://doi.org/10.4236/ica.2013.41010

Ayer, J. (2008). Transformation models and procedures for framework integration of the Ghana national geodetic network. *Ghana Surveyor*, *1*(2), 52–58.

Ayer, J., & Fosu, C. (2008). Map coordinates referencing and the use of GPS datasets in Ghana. *Journal of Science and Technology*, *28*(1), 116–127. https://doi.org/10.4314/just.v28i1.33084

Ayer, J., & Tiennah, T. (2008). Datum transformations by iterative solution of the abridging inverse Molodensky formulae. *Ghana Surveyor*, *1*(2), 59–66.

Ayoub, M. A., Negash, B. M., & Saaid, I. M. (2015). Modeling pressure drop in vertical wells using group method of data handling (GMDH) approach. In *Proceedings of International Conference on Integrated Petroleum Engineering and Geosciences* (pp. 69–78). Singapore. https://doi.org/10.1007/978-981-287-368-2_6

Barsi, P. (2001). Performing coordinate transformation by artificial neural network. *Allgemeine Vermessungs Nachrichten*, *4*, 134–137.

Bishop, C. (2006). *Pattern recognition and machine learning* (1st ed.). Springer.

Chang, N. B., Han, M., Yao, W., Chen, L. C., & Xu, S. (2010). Change detection of land use and land cover in an urban region with SPOT-5 images and partial Lanczos extreme learning machine. *Journal of Applied Remote Sensing*, *4*(1), 11–15. https://doi.org/10.1117/1.3518096

Chen, S., Cowan, C. F. N., & Grant, P. M. (1991). Orthogonal least squares learning algorithm for radial basis functions networks. *IEEE Transaction on Neural Networks*, *2*(2), 302–309. https://doi.org/10.1109/72.80341

Deakin, R. E. (2007). *Coordinate transformations for cadastral surveying* (pp. 1–34). School of Mathematical and Geospatial Sciences, RMIT University.

Dönmez, Ş. Ö., & Tunc, A. (2016, July 12–19). Transformation methods for using combination of remotely sensed data and cadastral maps. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B4, 2016 XXIII ISPRS Congress* (pp. 587–589). Prague, Czech Republic. https://doi.org/10.5194/isprs-archives-XLI-B4-587-2016

Dzidefo, A. (2011). *Determination of transformation parameters between the World Geodetic System 1984 and the Ghana Geodetic Network* (Master's Thesis). Department of Civil & Geomatic Engineering, KNUST, Kumasi, Ghana.

El-Assal, A., El-Rabbany, A., & Mesbah, S. (2011). GPS outage recovery using wavelet and neural network models in support of multibeam hydrography. In *Proceedings of OCEANS IEEE* (pp. 1–7). IEEE. https://doi.org/10.1109/Oceans-Spain.2011.6003541

Elshambaky, H. T., Kaloop, M. R., & Hu, J. W. (2018). A novel three-direction datum transformation of geodetic coordinates for Egypt using artificial neural network approach. *Arabian Journal of Geosciences*, *11*(6), 110. https://doi.org/10.1007/s12517-018-3441-6

Farlow, S. J. (1984*). Self-organizing methods in modeling: GMDH type algorithms.* CRC Press, Marcel-Dekker.

Fosu, C., Poku-Gyamfi, Y., & Hein, W. G. (2006). Global Navigation Satellite System (GNSS) – A Utility for Sustainable Development in Africa. In *Proceedings of 5th FIG Regular Conference on Promoting Land Administration and Good Governance, Workshop – AFREF I* (pp. 1–12). Accra, Ghana.

Fotiou, A., & Kaltsikis, C. J. (2016). *Computationally efficient methods and solutions with least squares similarity transformation models*. Retrieved January 2, 2018, from https://www.researchgate.net/profile/Aristeidis_Fotiou/publication/309732142_Computationally_efficient_methods_and_solutions_with_least_squares_similarity_transformation_models/links/58204df808ae12715afbb0c6.pdf

Ghilani, C. D. (2010). *Adjustment computations: Spatial data analysis* (5 ed.). John Wiley and Sons Inc.

Gullu, M. (2010). Coordinate transformation by radial basis function neural network. *Scientific Research and Essays*, *5*, 3141–3146.

Gullu, M., Yilmaz, M., Yilmaz, I., & Turgut, B. (2011). Datum transformation by artificial neural networks for geographic information systems applications. In *International Symposium on Environmental Protection and Planning: Geographic Information Systems (GIS) and Remote Sensing (RS) Applications (ISEPP)* (pp. 13–19). Izmir, Turkey. https://doi.org/10.5053/isepp.2011.1-6

Haykin, S. (1999). *Neural networks: A comprehensive foundation*. Prentice Hall.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feed forward networks are universal approximators. *Neural Networks*, *2*, 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Huang, G. B., & Babri, H. A. (1998). Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. *IEEE Transaction on Neural Networks*, *9*(1), 224–229. https://doi.org/10.1109/72.655045

Huang, G. B., Chen, L., & Siew, C. K. (2006a). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transaction on Neural Networks*, *17*(4), 879–892. https://doi.org/10.1109/TNN.2006.875977

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006b). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*(1), 489–501. https://doi.org/10.1016/j.neucom.2005.12.126

Ivakhnenko, A. G. (1966). Group method of data handling a rival of the method of stochastic approximation. *Soviet Automatic Control*, *13*, 43–71.

Ivakhnenko, A. G. (1971). Polynomial theory of complex systems. *IEEE Transactions on Systems, Man and Cybernetics*, *4*, 364–378. https://doi.org/10.1109/TSMC.1971.4308320

Jekabsons, G. (2010). *GMDH-type polynomial neural networks for MATLAB*. http://www.cs.rtu.lv/jekabsons/

Kavzoglu, T., & Saka, M. H. (2005). Modelling local GPS/Levelling geoid undulations using artificial neural networks. *Journal of Geodesy*, *78*, 520–527. https://doi.org/10.1007/s00190-004-0420-3

Konakoğlu, B., Cakir, L., & Gökalp, E. (2016). 2D coordinates transformation using artificial neural networks. In *Geo Advances 2016: ISPRS Workshop on Multi-dimensional & Multi-scale Spatial Data Modeling. Volume XLII-2/W1: 3rd International GeoAdvances Workshop* (pp. 183–186). Mimar Sinan Fine Arts University, Istanbul. https://doi.org/10.5194/isprs-archives-XLII-2-W1-183-2016

Konakoğlu, B., & Gökalp, E. (2016). A Study on 2D similarity transformation using multilayer perceptron neural networks and a performance comparison with conventional and robust outlier detection methods. *Acta Montanistica Slovaca*, *21*(4), 324–332.

Laari, P. B., Ziggah, Y. Y., & Annan, R. F. (2016). Determination of 3D transformation parameters for the Ghana geodetic reference network using ordinary least squares and total least squares techniques. *International Journal of Geomatics and Geosciences*, *7*(3), 245–261.

Lao-Sheng, L., & Yi-Jin, W. (2006). A study on cadastral coordinate transformation using artificial neural network. In *Proceedings of the 27th Asian Conference on Remote Sensing* (pp. 1–6). Ulaanbaatar, Mongolia.

Li, X. Z., & Kong, J. M. (2014). Application of GA-SVM method with parameter optimization for landslide development prediction. *Natural Hazard Earth System Science*, *14*(3), 525–533. https://doi.org/10.5194/nhess-14-525-2014

Malhotra, R., & Chug, A. (2014). Application of group method of data handling model for software maintainability prediction using object-oriented systems. *International Journal of Systematic Assurance Engineering and Management*, *5*(2), 165–173. https://doi.org/10.1007/s13198-014-0227-4

Mugnier, J. C. (2000). OGP-coordinate conversions and transformations including formulae, COLUMN, Grids and Datums. The Republic of Ghana. In *Photogrammetric Engineering and Remote Sensing* (pp. 695–697). American Society for Photogrammetry and Remote Sensing (ASPRS).

Muller, V. A., & Hemond, F. H. (2013). Extended artificial neural networks: incorporation of a priori chemical knowledge enables use of ion selective electrodes for in-situ measurement of ions at environmentally relevant levels. *Talanta*, *117*, 112–118. https://doi.org/10.1016/j.talanta.2013.08.045

Okwuashi, O., & Ndehedehe, C. (2017). Tide modelling using support vector machine regression. *Journal of Spatial Science*, *62*(1), 29–46. https://doi.org/10.1080/14498596.2016.1215272

Orr, M. J. L. (1996). *Introduction to radial basis function networks*. Center for Cognitive Science, Edinburgh University, Scotland, UK.

Pal, M. (2009). Extreme-learning-machine-based land cover classification. *International Journal of Remote Sensing*, *30*(14), 3835–3841. https://doi.org/10.1080/01431160902788636

Poku-Gyamfi, Y. (2009). *Establishment of GPS reference network in Ghana* (PhD Dissertation). Universitat der Bundeswehr Munchen, Germany.

Samui, P. (2012). Slope stability analysis using multivariate adaptive regression spline. *Metaheuristics in Water, Geotechnical and Transportation Engineering*, *14*, 327–342. https://doi.org/10.1016/B978-0-12-398296-4.00014-3

Sarycheva, L. (2003). Using GMDH in ecological and socio-economical monitoring problems. *Systems Analysis Modelling Simulation*, *43*(10), 1409–1414. https://doi.org/10.1080/02329290290024925

Sisman, Y. (2014). Coordinate transformation of cadastral maps using different adjustment methods. *Journal of Chinese Institute of Engineering*, *37*(7), 869–882. https://doi.org/10.1080/02533839.2014.888800

Tierra, A., Dalazoana, R., & De Freitas, S. (2008). Using an artificial neural network to improve the transformation of coordinates between classical geodetic reference frames. *Computers and Geosciences*, *34*, 181–189. https://doi.org/10.1016/j.cageo.2007.03.011

Tierra, A., & Romero, R. (2014). Planes coordinates transformation between PSAD56 to SIRGAS using a multilayer artifi-

cial neural network. *Geodesy and Cartography*, *63*, 99–209. https://doi.org/10.2478/geocart-2014-0014

Tierra, A. R., De Freitas, S. R. C., & Guevara, P. M. (2009). Using an artificial neural network to transformation of coordinates from PSAD56 to SIRGAS95. *Geodetic Reference Frames, International Association of Geodesy Symposia*, *134*, 173–178. https://doi.org/10.1007/978-3-642-00860-3_27

Turgut, B. (2010). A back-propagation artificial neural network approach for three-dimensional coordinate transformation. *Scientific Research and Essays*, *5*, 3330–3335.

Turgut, B. (2016). Application of back propagation artificial neural networks for gravity field modelling. *Acta Montanistica Slovaca*, *21*(3), 200–207.

Veronez, M. R., Florêncio De Souza, S., Matsuoka, M. T., Reinhardt, A., & Macedônio da Silva, R. (2011). Regional mapping of the geoid using GNSS (GPS) measurements and an artificial neural network. *Remote Sensing*, *3*(4), 668–683. https://doi.org/10.3390/rs3040668

Zaletnyik, P. (2004). Coordinate transformation with neural networks and with polynomials in Hungary. In *International Symposium on Modern Technologies, Education and Professional Practice in Geodesy and Related Fields* (pp. 471–479). Sofia, Bulgaria.

Zeng, H. (2014). Planar coordinate transformation and its parameter estimation in the complex number field. *Acta Geodaetica et Geophysica*, *49*(1), 79–94. https://doi.org/10.1007/s40328-014-0040-1

Zhang, S., Zhang, K., & Liu, P. (2016). Total least-squares estimation for 2D affine coordinate transformation with constraints on physical parameters. *Journal of Surveying Engineering*, *142*(3), 04016009-1-04016009-5. https://doi.org/10.1061/(ASCE)SU.1943-5428.0000180

Ziggah, Y. Y., Akwensi, P. H., & Annan, R. F. (2016a, August). Plane coordinate transformation using general least squares approach – a case study of Ghana geodetic reference network. In *4th UMaT Biennial International Mining and Mineral Conference* (pp. 68–77). Tarkwa, Ghana.

Ziggah, Y. Y., Ayer, J., & Laari, P. B. (2017). Coordinate transformation using Featherstone and Vaníček proposed approach – a case study of Ghana geodetic reference network. *Geoplanning: Journal of Geomatics and Planning*, *4*(1), 19–26. https://doi.org/10.14710/geoplanning.4.1.19-26

Ziggah, Y. Y., Yakubu, I., & Kumi-Boateng, B. (2016b). Analysis of methods for ellipsoidal height estimation – the case of a local geodetic reference network. *Ghana Mining Journal*, *16*(2), 1–9.

Ziggah, Y. Y., Youjian, H., Odutola, C. A., & Fan, D. L. (2013a). Determination of GPS coordinate transformation parameters of geodetic data between reference datums – a case study of Ghana geodetic reference network. *International Journal of Engineering Science and Research Technology*, *2*(4), 2277–9655.

Ziggah, Y. Y., Youjian, H., Odutola, C. A., & Nguyen, T. T. (2013b). Accuracy assessment of centroid computation methods in precise GPS coordinates transformation parameters determination – a case study, Ghana. *European Scientific Journal*, *9*(15), 1857–7431.