Al-Azhar Engineering 16th International Conference

Vol. 19, No. 72, July 2024, 102 - 118

DEVELOPING A LOCAL GEOID MODEL FOR EGYPT USING SOME OF MACHINE LEARNING ALGORITHMS

Salah S. Alsadany*, Essam M. Fawaz, Mohammed A. Elshewy, Ahmed M. Hamdy

Civil Engineering Department, Faculty of Engineering, Al-Azhar University, Nasr City, Cairo, Egypt

***Correspondence:** salahelsadany05@gmail.com

Citation:

S.S. Alsadany, E.M. Fawaz, M.A. Elshewy and A.M. Hamdy, "Developing a local geoid model for Egypt using machine learning algorithms", Journal of Al-Azhar University Engineering Sector, vol. 19, pp. 102 - 118, 2024.

Received: 01 December 2023

Revised: 04 January 2024

Accepted: 11 January 2024

DoI: 10.21608/auej.2024.254838.1517

ABSTRACT

This research aims at making use of advanced Machine Learning Algorithms (MLAs) with a view to developning a precise geoid model for Egypt. Being an equipotential surface of the earth's gravity field, geoid plays a crucial role in various geodetic applications. Throughout this study, state-of-the-art Machine Learning Algorithms are employed to address the limitations of conventional geoid modeling approaches. The research methodology involves evaluating the performance of eight Global Geopotential Models(GGMs), namely EGM2008, EIGEN-6C, EIGEN-6C2, EIGEN-6C4, EIGEN-6C3stat, SGG-UGM-1, XGM2019e_2159 and SGG-UGM-2 to choose the suitable GGM that for the study area, i.e. Egypt. MLAs, such as Linear Regression, Support Vector Machine, Random Forest, and Extra Trees, are then applied to train a model capable of determinig the intricate relationships between the input features and the geoid undulations. The study findings conclude that XGM2019e_2159 emerges as the optimal GGM for Egyptian territories, since it has yielded a standard deviation of 0.36 m. Notable enhancements in the local geoid model are observed with the application of the Extra Trees algorithm, which has yielded a standard deviation of 0.11 m.

KEYWORDS: Machine Learning Algorithms, Random Forest; equipotential surface, geoid undulations, local geoid.

تطوير نموذج الجويد المحلى لمصر بإستخدام بعض خوارزميات التعلم الالى

صلاح شكرى السعدنى[1]*،عصام محمد فواز[2]، محمد عبدالودود الشيوى[3] ، أحمد محمد حمدى[4]

قسم الهندسة المدنية، كلية الهندسة، جامعة الأزهر، مدينة نصر، القاهرة، مصر

***البريد الاليكتروني للباحث الرئيسي:** salahelsadany05@gmail.com

الملخص

يهدف هذا البحث إلى الاستفادة من خوارزميات التعلم الآلي المتقدمة بهدف تطوير نموذج جيود دقيق لمصر. كونه سطحًا متساوي الجهد في مجال الجاذبية الأرضية، يلعب الجيود دورًا حاسمًا في التطبيقات الجيوديسية المختلفة. خلال هذه الدراسة، يتم استخدام خوارزميات التعلم الآلي الحديثة لمعالجة القيود المفروضة على أساليب النمذجة الجيود التقليدية. تتضمن منهجية البحث تقييم أداء ثمانية نماذج جويد عالمية من أجل اختيار النموذج المناسب لمنطقة الدراسة. يتم بعد ذلك تطبيق خوارزميات التعلم الآلي، مثل الانحدار الخطى ودعم الآلات الناقلة والغابات العشوائية والاشجار الاضافية، لتدريب نموذج قادرعلى تعلم العلاقات المعقدة بين

ميزات الإدخال وحيود الجويد. تشير النتائج إلى أن (XGM2019e_2159) يظهر باعتباره النموذج الأمثل للأراضي المصرية، حيث كان انحرافه المعياري 0.36 متر. علاوة على ذلك، تم ملاحظة تحسينات في نموذج الجيود المحلي مع تطبيق خوارزمية الاشجار الاضافية، حيث كان الانحراف المعياري 0.11 م.

الكلمات المفتاحية : التعلم الالى، الغابات العشوائية، الأسطح متساوية الجهد، حيود الجويد، الجويد المحلى.

1. INTRODUCTION

Geoid is one of the numerous surfaces of Earth's gravity field and is considered the actual equipotential surface of Earth's gravity field [1]. Civil engineers deal with three basic earth surfaces: the terrain surface, which is the real physical surface of the earth; the geoid, which is the irregular equipotential surface of the earth; and the regular surface (ellipsoid), which is the mathematical surface that is closest to the geoid [2]. However, the most instrumental surface is the geoid, due to the fact that most engineering work and scientific applications require orthometric heights that are obtained through the surface of the geoid [3] The geoid is greatly essential for geographic approaches, surveying, and mapping applications for any country [4]. In addition to its role in converting ellipsoid heights measured from the ellipsoid surface into useful orthometric heights related to the geiod surface, the geoid plays a crucial role in integrating GPS measurements with leveling to investigate the movements of vertical crust over extended periods of time [5]. Geoid can be used in different sciences, such as geophysical interpretations, which help to understand the distribution of the planet's interior mass and solid earth [6].

Since the 1980s, developing reliable geoid models has been considered one of the greatest geodesy-related research issues. Over the past four decades, reliable national geoid models have been produced in many countries all over the world.

Geoid models can be classified into three types according to the data used. The first model is the gravimetric geoid, which is based on using different types of gravity observations, such as terrestrial gravimetry, satellite gravimetry, and airborne gravimetry. The second model is a geometric geoid, which is based on using GPS/leveling observations. The third is a hybrid geoid, which is based on a combination of gravity and GPS/leveling observations. Geoid models can also be classified according to the covered area into local geoid models, which are appropriate for small areas, and global geoid models, which are suitable for large areas [7]. In all geoid types, the basic objective of any model is to calculate the geoid undulation values.

Geoid undulation, i.e., geoid height, is the vertical separation between two surfaces. See **Fig.1.** This undulation can be determined by the fundamental relationship, which consists of the orthometric height (H) obtained through the leveling process and the ellipsoidal height (h) obtained through GNSS measurements. This undulation is given by the following equation [3].

$$N = h - H \qquad \text{Eq. (1)}$$

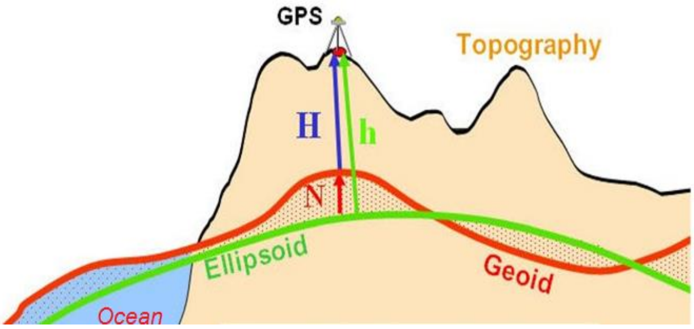

Fig.1. The relationship between ellipsoidal and orthometric height [7]

In reality, the implementation of the above-mentioned equation can be hindered by some factors that cause a great deal of discrepancy, making the equation highly complicated [8]. These factors may include: 1- Random errors in both heights (H and H). 2- Distortions and systematic effects on heights. 3- Various geodynamic effects (land subsidence, monument instabilities, plate deformation at subduction zones, and MSL rise). 4- Instability of reference station monuments over time [9].

Actually, the precision of the geoid model is affected by the distribution, number, and accuracy of utilized stations (leveling/GPS points). Therefore, the distribution of these stations must be conducted in a homogeneous manner throughout the entire study area [10].

1.1 Global Geopotential Models (GGM)

It should be noted that the first step towards accurate geoid modeling is the evaluation of global geopotential models to choose the most appropriate GGM for the area studied. Consequently, several studies have investigated the precision of GGMs, which have been created since the 1960s [11]. GGMs can be considered a spectrum-domain mathematical form used to compute gravitational potential via spherical harmonic expansion [7]. Depending on a suitable GGM, the gravity potential value related to the earth can be determined through using fully normalized Stokes' coefficients for each degree n and order m (c_{nm} and s_{nm}), GGM models can also provide much information about the earth's vertical deflection, gravity anomaly, normal gravity, and geoid undulation [12]. Based on the GGM models as the essential tool utilized for developing geoid in both regional and local forms, the International Center for Global Earth Models (http://icgem.gfz potsdam.d /ICGEM/ICGEM.html) has more than 160 GGM available [13].

Typically, GGMs are generated from satellite gravity data or from a combination of satellite models, terrestrial gravity, altimeters derived from oceanic gravity data, and airborne gravity [1]. So, the maximum degree and accuracy of GGMs vary significantly due to the use of various types of data sources in creating GGMs. GGMs are classified into three main groups. Firstly, so-named satellites only depend on the observations of artificial earth satellites. Secondly, it is called a combined model, which is created from the combination of different sources such as terrestrial, land, airborne gravity, and marine gravity anomalies obtained from satellite altimetry. Third, so-called tailored models, which established from adjustment, improvement, and enhancement the coefficients of spherical harmonics related to GGMs for satellitelite, and combined models to increase the degree of the model [7, 14].

This paper evaluates the accuracy of eight recent GGMs to select the suitable model between them to represent the Egyptian surface. The characteristics of GGMs utilized in this study are illustrated in the following:

1.1.1 EGM2008: The National Geospatial Intelligence Agency of the United States established this model in 2008, using data from altimetry, satellite tracking, and terrestrial gravity. The dgree and order of EGM2008 are 2159 and 2190 respectively. This model enables the estimation of quasigeoid height, gravity anomaly, and vertical deflection [15].

1.1.2 EIGEN-6C: It was produced in 2011 and developed up to degree 1420 by combining the observations of (GRACE), (GOCE), and (LAGEOS).

1.1.3 EIGEN-6C2: It was launched in 2012 and developed to a degree up to 1949 by combining the observations of (GRACE), (GOCE), and (LAGEOS) [16].

1.1.4 EIGEN-6C4: It was produced in 2014 and developed by both the French research center and the German research center (CNES & GFZ) to degrees up to 2190. EIGEN-6C4 utilizes the data of satellite tracking obtained from both GRACE, LAGEOS, and GOCE with a grid of gravity anomalies of a global surface [17].

1.1.5 EIGEN-6C3stat: It was produced in 2014 and developed to a degree since 1949 by combining the observations of GRACE, the Gravity Field and GOCE and data on marine gravity obtained from satellite altimetry data [18].

1.1.6 SGG-UGM-1: It was developed in 2018 by Wuhan University in China by incorporating the signal of GOCE gravity obtained during the period of November 2009 and May 2012 with EGM2008 [19].

1.1.7 XGM2019e_2159: It is considered one of the latest GGMs. It was produced in 2019 and developed to degree 2190. It combines 1' mean sea surface data sets throughout the oceans, 15' terrestrial gravity anomalies database on land, the EARTH2014 topographic model and 5' GOCO06s GGM [17].

1.1.8 SGG-UGM-2: It is considered one of the modern GGMs and was released in 2020. the degree and order of this model, 2190 and 2159. by merging the observations of the (GRACE), EGM2008, gravity field (GOCE), and marine gravity obtained from satellite altimetry data [20].

1.2 Machine learning

Machine Learning Algorithms (MLAs) is a part of Artificial intelligence that is focused on the development of statistical models and algorithms that can be able to learn from and make predictions on data. Through the previous few decades, several studies proved that MLAs are more accurate than traditional statistical techniques like logistic regression or discriminant analysis, especially when the input database is expected to have various statistical distributions or feature space is very complex [21]. Therefore, a great number of mechanisms for regression and classification have been developed. So, in this research some of these regression techniques are investigated for geoid modeling.

1.2.1 Linear Regression

Regression means that predicting the continuous output variables affected by independent input variable. Linear regression is considered one of the important supervised MLAs that can be used for predictions on new data sets, through learning from the labeled data sets and maps the data points to the most optimized linear functions. Finding the ideal linear equation for predicting the values of the dependent variable based on the values of the independent variables is the aim of the linear regression [22]. The equation gives a straight line representing the relation between the value of independent and dependent variables.

In a regression set of observations are present with X and Y values, these data are utilized to learn a function for prediction Y through an unknown X. Therefore, a function is required for predicting continuous Y in the expression case given X even as independent features.

Assume that our independent feature is the location of the reference station (X) and the height difference between the $N_{GNSS/levelling}$ and the N_{GGM} (Y) is the dependent variable. Let's assume there is a linear relation between X and Y, then the height difference can be predicted using the following equation.

$$Y_i = \theta_1 + \theta_2 X_i \qquad \text{Eq.(2)}$$

Y_i $(i = 1, 2, 3, \dots\dots)$ are the predicted values.

X_i $(i = 1, 2, 3, \dots\dots)$ input independent data.

θ_1: intercept

θ_2: coefficient of x

1.2.2 Support Vector Machines (SVM)

Support Vector Machines (SVM) developed by Cortes and Vapnik, (SVM) are supervised learning techniques related to connected learning algorithms which have the ability to analyze data utilize for regression and classification analysis [23]. SVMs are suitable for being applied in many different fields: image detection, recognition and verification prediction, remote sensing, image analysis, time series forecasting and chemical sciences, this algorithms would be suitable when the problem might not be linearly separable [24].

1.2.3 Random Forest

Random forest is considered one of the most distinguished ensemble techniques because it's a substantial advancement on simple decision trees. It is an ensemble training algorithm which constructs numerous decision trees throughout the training. Random forest has the ability to forecast the mode of the groups for forests process and mean prediction of trees for the regression process. It is utilizing random subspace techniques and bagging during tree construction. It has built-in feature importance [25]. As shown in **Fig.2**, an RF comprises of several groups of trees, where each tree is grown depending on randomization distribution [26]. The number of trees required increases with the number of predictors in order to get good performanc. Comparing predictions from a forest against those from a subset of a forest is the most effective way to figure out the appropriate amount of trees needed [25].

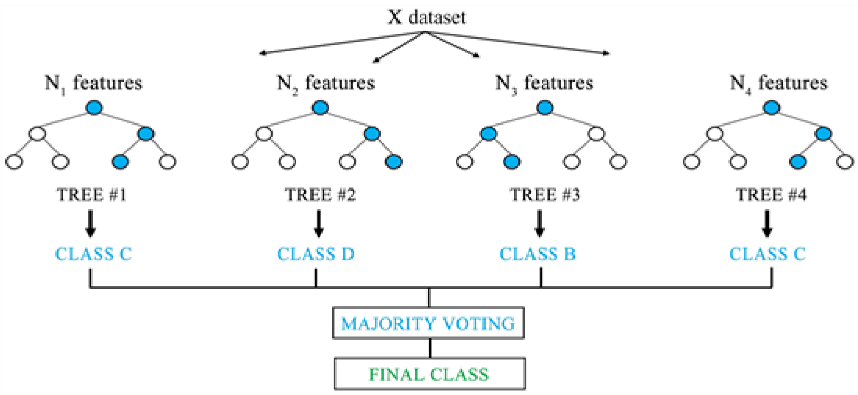

Fig. 2. Random forest mechanism [26]

1.2.4 Extra Trees

One of the very important algorithms is the Extra tree, which the ensemble learning algorithms. It builds a group of decision trees. During tree establishment, the decision rule is randomly chosen. This algorithm is very analogous to Random Forest except for random chosen of split values [27]. The main objective of the Extra tree technique is creating further unpredictable tree establishment in the conditions of numerical insert features, where the selection of the favorable cut point is responsible for a great amount of the variance of the induced tree. The extra trees algorithm is done by generating a great number of unpruned decision trees in the training database. In the regression, predictions are created by averaging the prediction of the decision trees.

The primary objective of this research is to develop a local geoid model for Egypt using different MLAs in Python libraries such as Linear Regression, Support Vector Machine (SVM), Random Forest (RF) and Extra Tress.

2. STUDY AREA

The study area is located in Eygpt. It was divided into three regions; the first region, which stretches for about 600 km along the Mediterranean Sea, started from reference station (31º 29' 15" N, 26º 36' 19" E) located at the west of Egypt to reference station (31º 2' 49" N, 33º 0' 22" E) located at the east of Egypt, is notable for its diverse natural and environmental features as well as its important cities, including Alexandria, Port Said, Al-Almain, and Marsa Matrouh. The second region extended along the coast of the Red Sea for about 700 km from the reference station (24º 12' 16" N, 35º 25' 51" E), which is located in Egypt's south, to the reference station (31º 13' 5" N, 32º 22' 29" E), which is located in Egypt's north. This region is highly valued because it contributes significantly to the country's GDP due to the presence of important tourist destinations like Dahab, Nuweiba, Hurghada, Safaga, and Sharm El-Sheikh. The third region is distinguished by the Nile River, which is Egypt's lifeblood. The majority of the population in this region works in agriculture in addition to various other industries. This region extended nearly 824 km along the River Nile from the reference station (31º 26' 16" N, 25º 23' 55" E) at the north of Egypt to the reference station (23º 56' 26" N, 35º 23' 50" E) at the upper Egypt. **Fig. 3** shows the distribution of all GNSS/leveling utilized in this study.

Fig. 3. Distribution of the GNSS/levelling points

3. METHODOLOGY

The methodology of this research involves developing a local geoid model for Egypt, encompassing the Red Sea and Mediterranean coasts as well as the Delta area and the Nile River. The methodology of this research includes the following steps is shown in **Fig.4**.

1- $N_{GNSS/levelling}$ of all GNSS leveling points with known latitude (Φ), longitude (λ), orthometric height (H) and ellipsodial height (h) was calculated from the equation (1).

2- Depending on $N_{GNSS/levelling}$ obtained from the previous step, eight GGMs as EGM2008, EIGEN-6C, EIGEN-6C2, EIGEN-6C4, EIGEN-6C3stat, SGG-UGM-1, XGM2019e_2159 and SGG-UGM-2 with high degree and order were evaluated in order to choose the suitable model for the study area through determine the difference between N_{GNSS} of GNSS leveling points and N_{GMM} of GGM obtained from the website of the (ICGEM). The difference (Δ_N) calculated from the following equation.

$$\Delta_N = N_{GNSS/levelling} - N_{GMM} \qquad \text{Eq.(3)}$$

3- The difference between $N_{GNSS/level}$ and $N_{best\ GGM}$ was calculated according to the following equation.

$$\Delta_{N\ best\ model} = N_{GNSS/levelling} - N_{best\ GMM} \qquad \text{Eq.(4)}$$

4- The GNSS leveling points are spilt into two groups: a training group and a testing group.

5- Some of Machine Learning Algorithms for example Linear Regression, Random Forest, Support Vector Machine and Extra Tree using python libraries was applied in order to regression the $\Delta_{N\ best\ model}$ of the training points.

6- Comparison between the previous algorithms by performing some statistical operations, the optimum algorithm was chosen.

7- Height difference ($\Delta_{N\ best\ model}$) of points every 10 km of all Egypt is calculated through the code produced by optimum machine learning algorithm.

8- From ICGEM website, ($N_{best\ model}$) of points every 10 km of all Egypt the best GGM is obtained.

9- Geoid model of Egypt is determined from the following equation

$$N_{geoid} = \text{height diff } (\Delta_{N\ best\ model}) + N_{GGM\ best\ model} \qquad \text{Eq.(5)}$$

10- A map of the geoid model obtained from the previous steps was drawn using the surfer program.

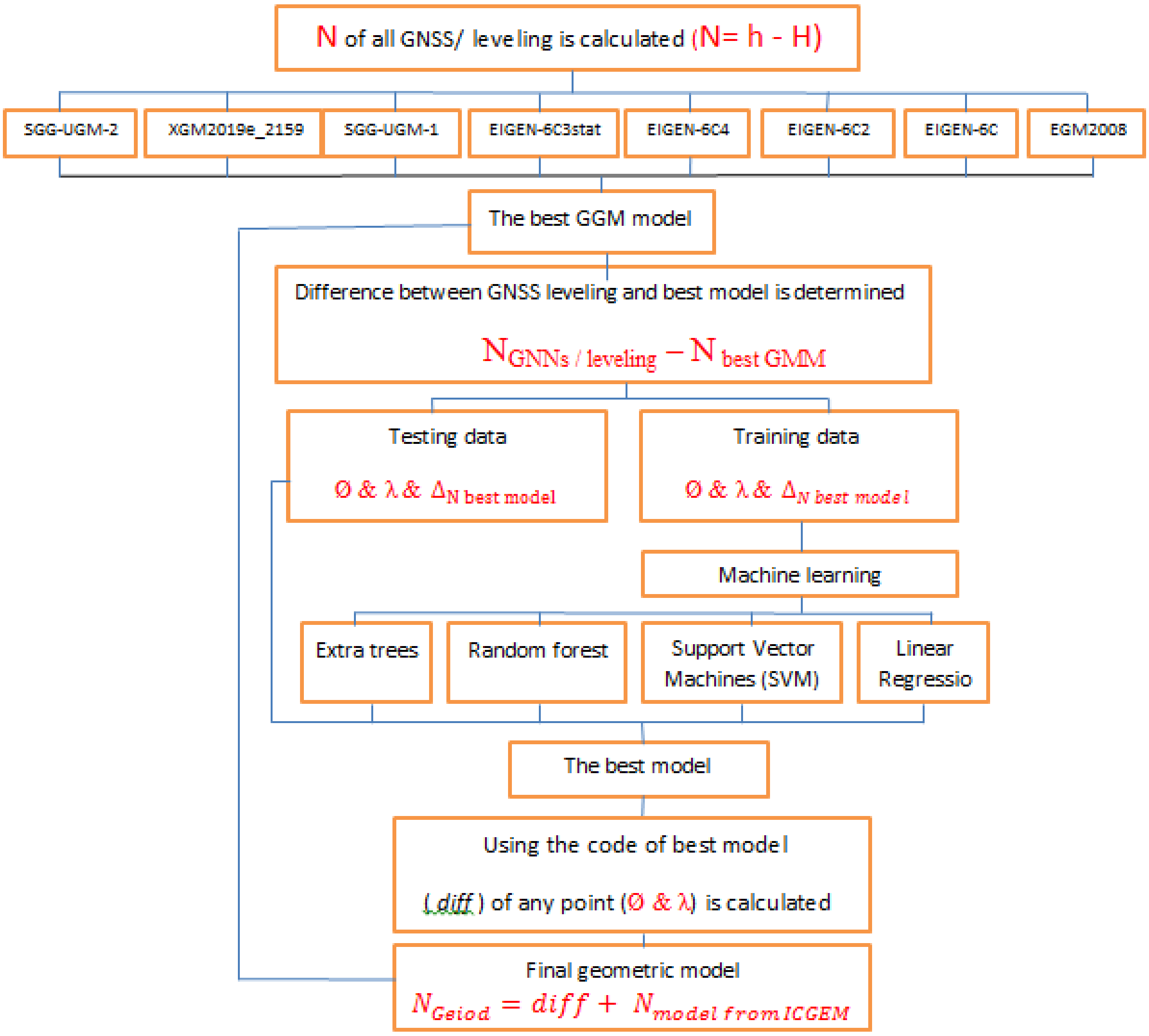

Fig. 4. Flow chart of the methodology

4. RESULTS AND DISCUSSIONS

At first, geoid height values for $N_{GNSS/levelling}$ were calculated from equation (1) for all 514 GNSS/levelling points. These points were separated randomly by Python into two groups where both groups must cover the studied area: the first group is called the training group, which consists of 462 points with known latitude (Φ), longitude (λ) and height difference. This group was used in the modeling process. While the second group is called the testing group and consisted of the remaining points (52 points) with known latitude (Φ), longitude (λ) and height difference, this group was used as (check points) to evaluate the results. **Fig.5** show the distribution of training points, while **Fig. 6** show the testing points' distribution.

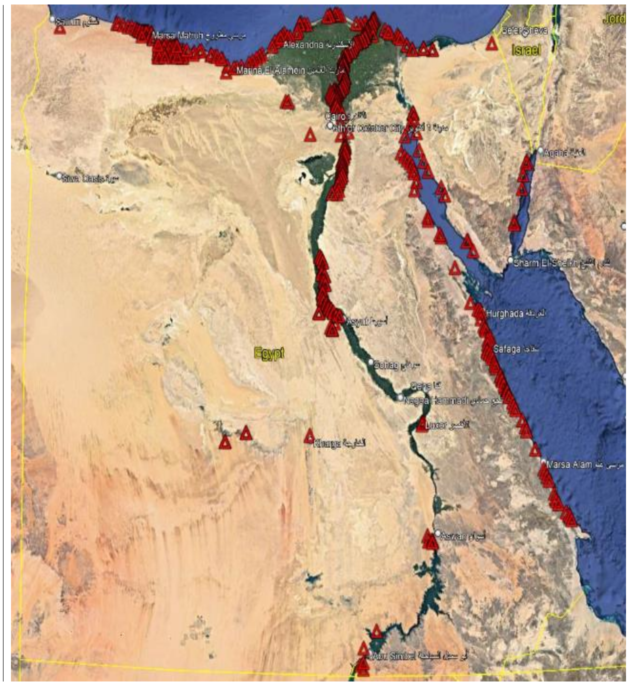

Fig. 5. The distribution of training points

Fig. 6. The distribution of testing points

4.1 Comparison between GGMs

This step is very important; it is considered the initial step for geoid modeling. In order to choose the suitable GGM for the area study, a comparison process was made between eight GGMs with high degree and order: EGM2008, EIGEN-6C, EIGEN-6C2, EIGEN-6C4, EIGEN-6C3stat, SGG-UGM-1, XGM2019e_2159, and SGG-UGM-2. The performance analysis of these GGMs consists of the following procedures: calculating the height difference (Δ_N) between GNSS/leveling points and the GGM obtained from the website of the (ICGEM) using equation (3). Then, by performing some statistical operations on this height difference (Δ_N) for each model to calculate the mean, minimum, maximum, and standard deviation [28], the GGM that achives the minimum standard deviation is considered as the best GGM. The accomplished results are shown in **Table 1**.

Table. 1. Comparison between global Geopotential model

GGM models	Difference between N of GNSS/levelling and N of GGM					
	year	degree	Minimum (m)	Maximum (m)	Mean (m)	Standard deviation (m)
EGM 2008	2008	2190	-1.71	1.5	-0.61	0.38
EIGEN-6C	2011	1420	-1.25	0.72	-0.56	0.37
EIGEN-6C2	2012	1949	-1.23	0.85	-0.58	0.38
EIGEN-6C4	2014	2190	-1.24	0.79	-0.55	0.38
EIGEN-6C3stat	2014	1949	-1.19	0.77	-0.54	0.37
SGG-UGM-1	2018	2159	-1.28	0.76	-0.58	0.37
XGM2019e_2159	2019	2190	-1.25	0.31	-0.57	0.36
SGG-UGM-2	2020	2190	-1.34	0.75	-0.58	0.38

Through the preceding table and from **Fig. 7**, which shows the height difference (Δ_N) between N of GNSS/ levelling and N of each GGM used in this study, all points were presented on the horizontal axis, while the height difference (Δ_N) was presented on the vertical axis. It can be recognized that the results are very similar. This may be attributed to the number and quality of the terrestrial gravity observations, altimetry data, and satellite tracking observations that were utilized in the development of these models. However, there is a very slight superiority to the XGM2019e_2159 model, which achieved a standard deviation of 0.36 m.

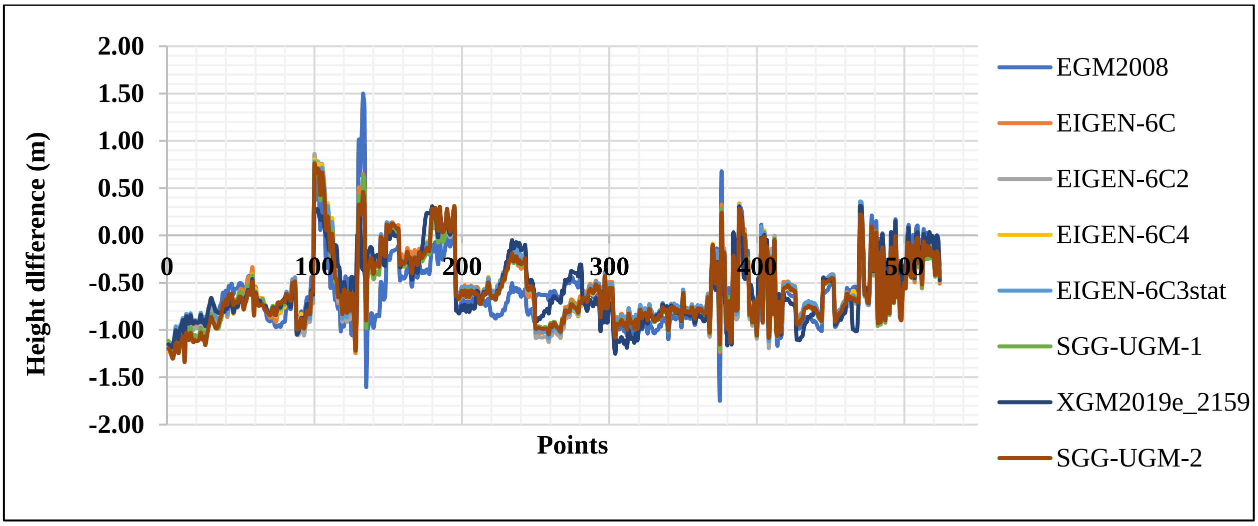

Fig. 7. Height difference (Δ_N) between $N_{GNSS/levelling}$ and N_{GGM}

In order to test whether or not there are significant differences between the geoid heights of the global models, ΔN from the XGM2019e_2159 was chosen to compare ΔN from the other seven GGMs since the model XGM2019e_2159 provided the lowest standard deviation of height differences. Table 2 displays the results of the F test for the global geoid models when the sample size (number of points) is 512 and the F criticle = 1.154 at a 95% confidence level [28].

Table. 2. Results of F test for the global geoid model

Teased model	F value of	F criticle	result
XGM2019e_2159 and EGM2008	1.083	1.154	Insignificant difference
XGM2019e_2159 and EIGEN-6C	1.064	1.154	Insignificant difference
XGM2019e_2159 and EIGEN-6C2	1.127	1.154	Insignificant difference
XGM2019e_2159 and EIGEN-6C4	1.082	1.154	Insignificant difference
XGM2019e_2159 and EIGEN-6C3stat	1.058	1.154	Insignificant difference
XGM2019e_2159 and SGG-UGM-1	1.031	1.154	Insignificant difference
XGM2019e_2159 and SGG-UGM-2	1.093	1.154	Insignificant difference

Table. 2 demonstrates that F value < F criticle, so there is no significant difference between the geoid heights produced from the XGM2019e 2159 model and the EGM2008, EIGEN-6C, EIGEN-6C2, EIGEN-6C4, EIGEN-6C3stat, SGG-UGM-1 and SGG-UGM-2 models.

Based on the above results and according to the previous research [29], it can be concluded that the XGM2019e_2159 is the suitable GGM model for Egyptian lands. As a result, the height difference ($\Delta_{N\ \text{best model}}$) between $\mathbf{N}_{\mathbf{GNSS/levelling}}$ of GNSS/leveling points and $N_{\mathbf{XGM2019e\text{-}2159}}$ was calculated according to equation (4).

For the evaluation of the MLAs used in this study for geoid modeling, the Python libraries have been used for regression of the height difference of the best model ($\Delta_{N\ \text{best model}}$) of the training points to predict the height difference of the check points and compare them with the known height difference of the same points. The MLA that gives the best results is chosen in order to predict the height difference at any point.

4.2 : Modeling utilizing MLAs

4.2.1 Evaluation of Linear regression

Linear regression is considered an important branch of supervised MLAs that can be used for prediction of new data through learning from the labeled datasets. During the regression process, some data set are present with X and Y values. These data are utilized in order to learn a function that can be used for prediction at any X. **Fig. 8** shows the difference between the actual values of the height difference ($\Delta_{\text{N best model}}$) of the testing point and the predicted values of height difference at the same point, where the horizontal axis represents the points of the testing group (known latitude and longitude) and the vertical axis represents the height difference (m). It is worth mentioning that linear regression achieved a standard deviation of 0.30 m. The issue of the regression process is that of recognition, a function that is close to mapping from an insert domain to real numbers depending on training data.

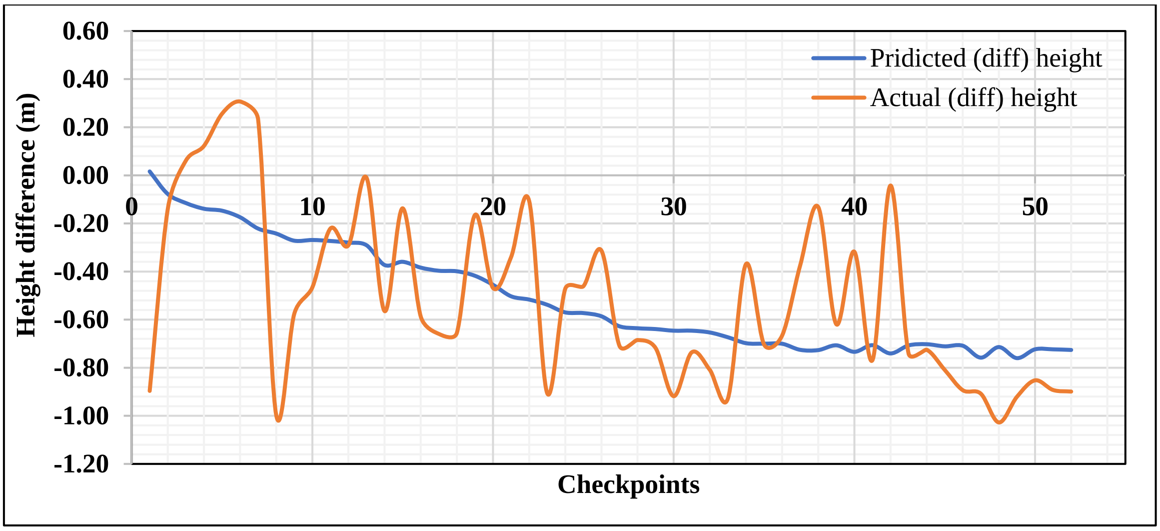

Fig. 8. Actual height difference and predicted height difference by linear regression

4.2.2 Evaluation of Support Vector Machine

The algorithm of SVM is supervised learning models with correlated learning algorithms that are utilized for regression and classification analysis. During the training step, SVM creates a model, maps the decision boundary of the data, and identifies the hyperplanes that separate the various classes. As the hyperplane increases, the prediction accuracy increases. **Fig. 9** shows the difference between the actual values of the height difference ($\Delta_{\text{N best model}}$) at the testing point and the predicted values of the height difference at the same point. The accuracy of the support vector machine reached 0.19 m.

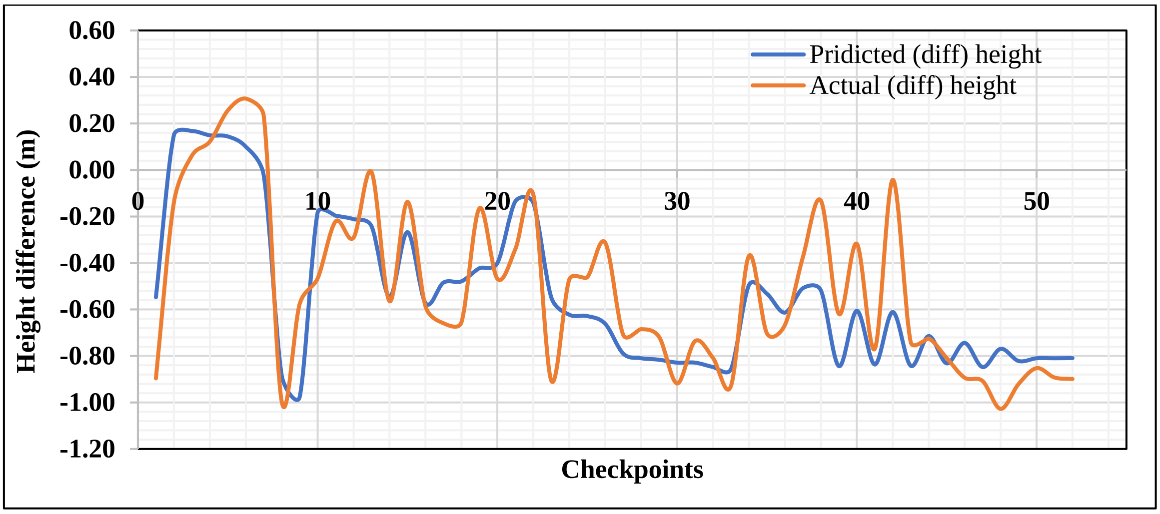

Fig. 9. Actual height difference and predicted height difference by SVM

4.2.3 Evaluation of Random Forest

A Random Forest prediction consists of a number of trees, where all trees grow through some form of randomization. The number of trees needed for ideal performance builds up with the number of predictors. The optimum approach to estimating how many trees are needed is a comparison between predictions made by a forest and predictions made by a subset of a forest. When the subset woraks as well as the full forest, this indicates that there are enough trees. Fig. 10 shows the difference between the actual values of the height difference ($\Delta_{\text{N best model}}$) at the testing point and the corresponding values of the height difference predicted by the random forest algorithm, which achieved a standard deviation of 0.12 m at the same points.

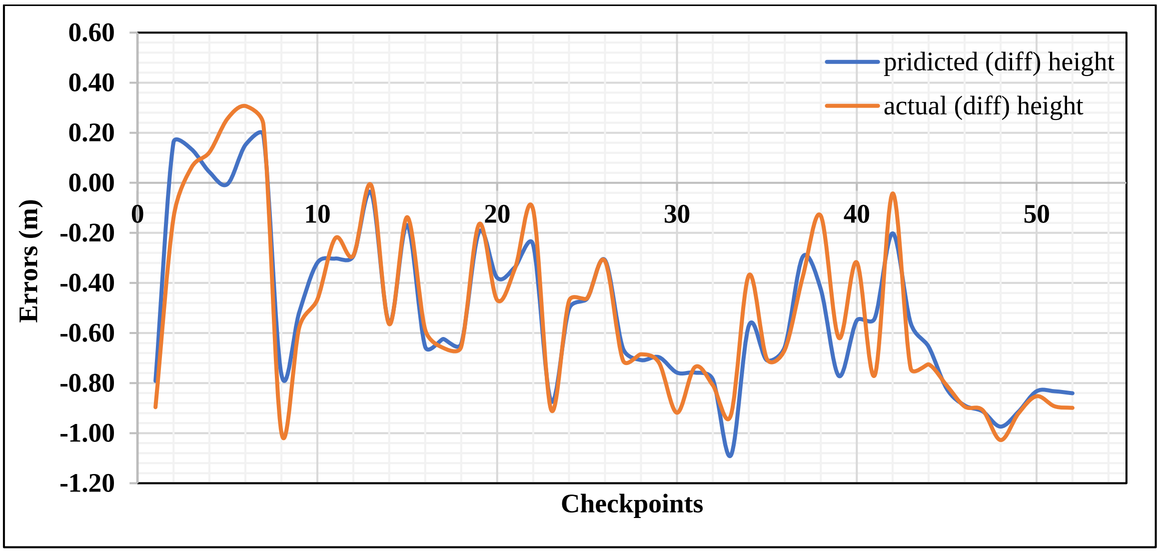

Fig.10. Actual height difference and predicted height difference by Random Forest

4.2.4 Evaluation of Extra Tree

Recently, interest in the Extra Tree technique as one of the important MLAs used for prediction has increased. It works by creating a huge number of decision trees from the training data set. Predictions are made by averaging the predictions of the decision trees in the regression. **Fig. 11** shows the difference between the actual values of the height difference ($\Delta_{\text{N best model}}$) of the testing point and corresponding values of the height difference predicted by the Extra Tree algorithm, which achieved higher accuracy than the previous algorithms where the achieved standard deviation equaled 0.11 m at the same points.

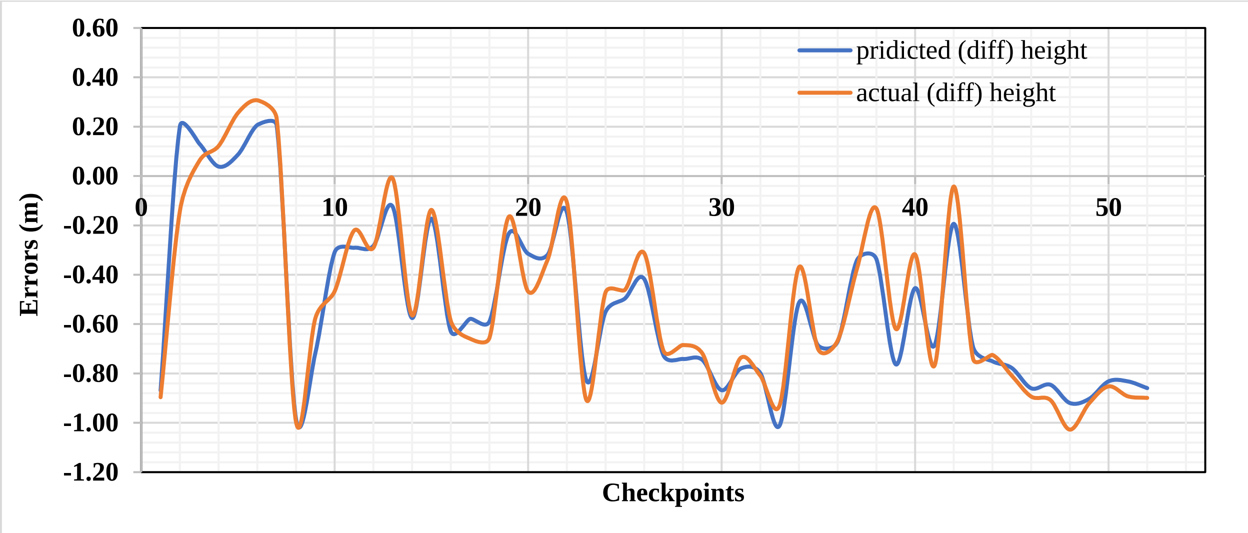

Fig.11. Actual height difference and predicted height difference by Extra Tree

The errors between the actual height difference and predicted height difference predicted by the previuos machine learning algorithms are shown in **Fig. 12**.

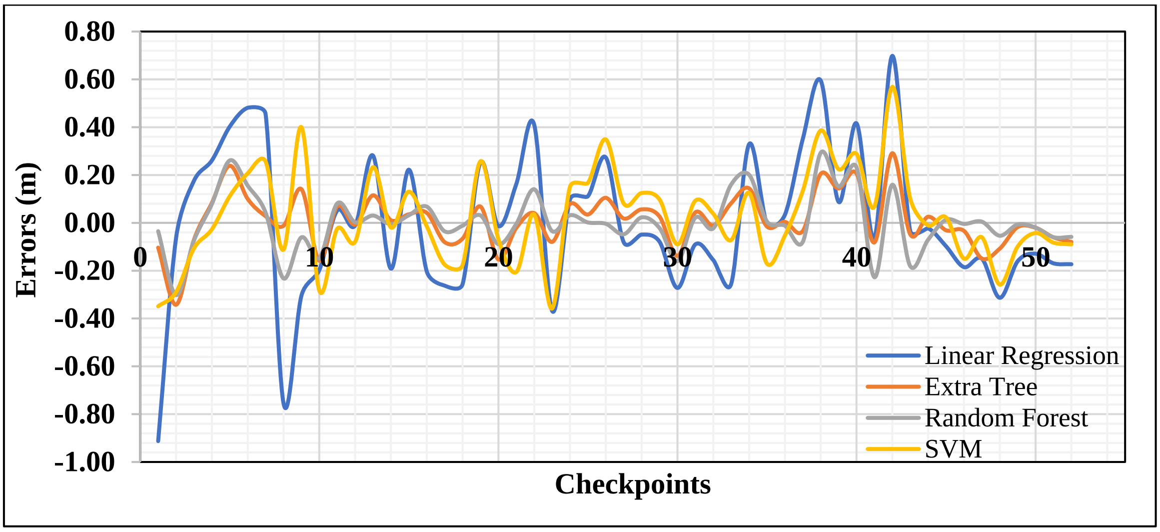

Fig.12. Errors between actual height difference and predicted height difference

According to [28], the statistics of the estimated height differences ($\Delta_{\mathrm{N\ best\ model}}$) by different MLAs are presented in **Table 3**.

Table. 3. Comparison between different MLAs

MLAs	Minimum (m)	Maximum (m)	Mean (m)	Standard deviation (m)
Linear Regression	-0.69	0.91	0.02	0.30
Support Vector Machine (SVM)	-0.57	0.36	-0.02	0.19
Random Forest (RF)	-0.29	0.31	-0.01	0.12
Extra Tree	-0.29	0.34	0.01	0.11

From the previous table, it can be noticed that Extra Trees and Random Forest are two very similar algorithms, but Extra achieved better results and outperformed other algorithms, especially when data was scarce, where the Extra Tree algorithm produced the smallest standard deviation of 0.11 m, while Random Forest came in second with a standard deviation of 0.12 m, while the biggest standard deviation was achieved with linear regression.

As a result of all of the above, Extra Tree is the optimum MLA for modeling the height difference of a geoid, so this algorithm is used to predict the height difference at any point by introducing the latitude and longitude of this point to the code produced by the Extra Tree algorithm. In addition, depending on the latitude and longitude of the same point, the value of N of the best model can be obtained from the ICGEM website, and then, by applying equation (5), the value of the geoid at this point can be calculated.

Based on the above, a perpendicular grid was made for Egypt with horizontal and vertical distances every 10 km; every point has known latitude and longitude; the height difference was determined at all estimated points from the Extra Tree code; and $N_{XGM2019e\text{-}2159}$ at the same points was obtained from the ICGEM website; then the value of the geoid was calculated for the same point. After that, using a surfer program, a map of the geoid was drawn as shown in **Fig. 13**.

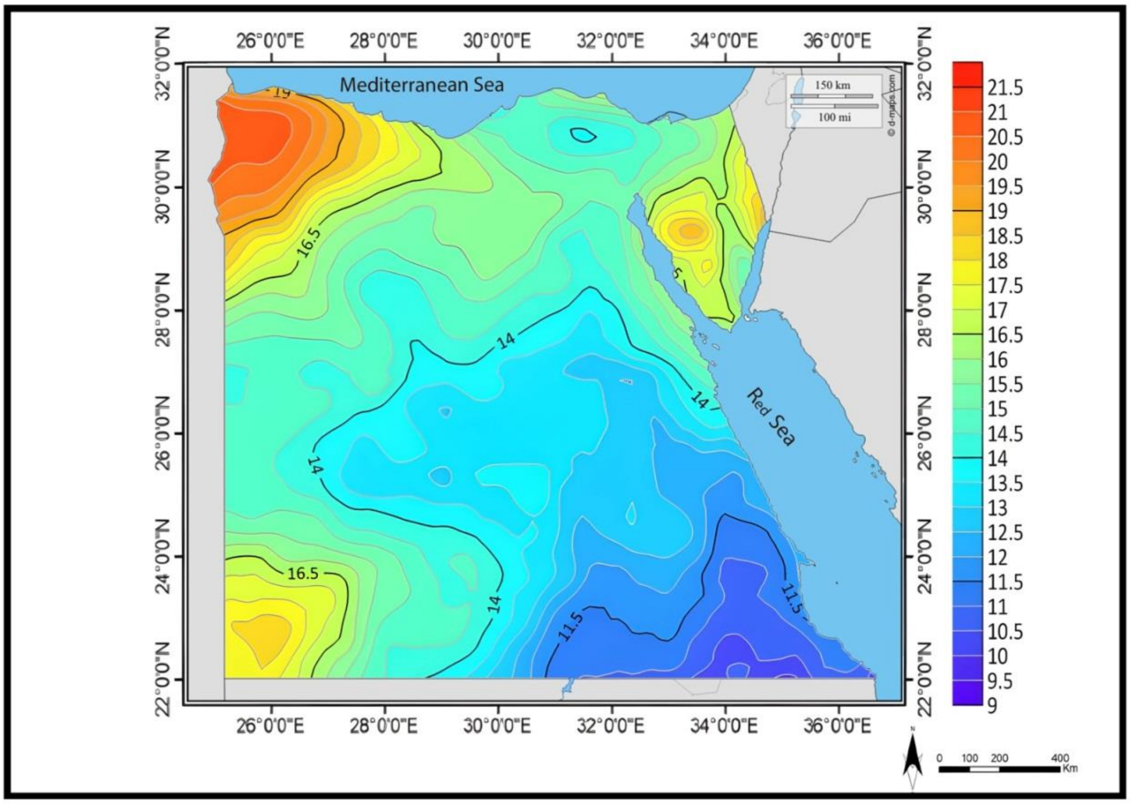

Fig.13. Geoid model for Egypt.

From the preceding figure, it can be noticed that the value of the geiod ranges from 9 to 21.5 m. The lowest value of geiod is in the southeastern zone near the Red Sea coast, where the value of geiod in this region ranges from 9 to 13 m. While the highest value of geiod is in the northwestern zone of the country, where these values range from 18 to 21.5. m. It is also noted that the change in the value of the geiod is not a random change, i.e., the value of the change from one area to another is very small. This is due to the nature of the Egyptian land, which is considered somewhat flat.

5. CONCLUSION

The GNSS has simplified the determination of ellipsoidal height in relation to the ellipsoidal surface. However, this height lacks acceptance and physical significance in many civil engineering applications, including construction, mapping, engineering design, and planning. These fields necessitate the use of a desirable height known as orthometric height, established above the geoid surface. Therefore, achieving precise modeling of the local geoid is imperative to enable the accurate conversion of ellipsoidal height to orthometric height. The primary objective of this research is to develop a local geoid model for Egypt using different MLAs in Python libraries, such as linear regression, support vector machines, random forests, and extra trees. Moreover, the performance of EGM2008, EIGEN-6C, EIGEN-6C2, EIGEN-6C4, EIGEN-6C3stat, SGG-UGM-1, XGM2019e_2159 and SGG-UGM-2 were evaluated to choose the suitable GGM for area study as the initial step in geoid modeling. This research marks a significant stride in the field of geodetics by proposing and implementing a local geoid model tailored specifically for the diverse and intricate geological features of Egypt. The combination of advanced MLAs and GGMs has proven to be a potent methodology, showcasing its effectiveness in achieving a high degree of precision and accuracy. The standout model, XGM2019e_2159, emerged as the most suitable GGM for the Egyptian lands, boasting a commendable standard deviation of 0.36 meters. Notably, the application of the Extra Trees algorithm yielded even more promising results, showcasing substantial enhancements in the local geoid model with a reduced standard deviation of 0.11 meters. This underscores the capacity of machine learning techniques to discern and model intricate relationships within geophysical data, thereby refining the accuracy of geoid predictions.

REFERENCES

[1] Abd-Elrahim Ruby, 2018 Enhancement of global geopotential harmonic models for Egypt. M.Sc. Degree in Surveying and Geodesy, Benha University Faculty of Engineering at Shoubra.
[2] El Shouny A, Al-Karagy EM, Mohamed HF, Dawod GM. Gis-based accuracy assessment of global geopotential models: A case study of Egypt. Am J Geographic Inf Syst. 2018.
[3] T. Kavzoglu, M. H. Saka, Modelling local GPS/levelling geoid undulations using artificial neural networks Journal of Geodesy (2005) 78: 520–52 DOI 10.1007/s00190-004-0420-3.
[4] E. El-Sayed, Development of a national geoid for Egypt using recent surveying data, PH. D. of Philosophy in Engineering, Faculty of Engineering - Shibeen Elkoom University, 2016.
[5] M. El-Ashquer, An Improved Hybrid Local Geoid Model for Egypt. PH. D of Philosophy in Engineering, Faculty of Engineering, Zagazig University, 2017 .
[6] Knudsen, P., Bingham, R. J., Andersen, O., and Rio M. H.: A global mean dynamic topography and ocean circulation estimation using a preliminary GOCE gravity model, J. Geodesy, 85, 861–879, https://doi.org/10.1007/s00190-011-0485-8, 2011.
[7] Ahmed Al Shouny, Ragab Khalil, Abdullah Kamel, and Yehia Miky. Assessments of recent Global Geopotential Models based on GPS/levelling and gravity data along coastal zones of Egypt. received August 12, 2022; accepted December 13, 2022. https://doi.org/10.1515/geo-2022-0450.

[8] Rummel R, Teunissen P (1989) Height datum definition, height datum connection and the role of the geodetic boundary value problem. Bull G´eod 62:477–498.
[9] G. Fotopoulos, 2005 Calibration of geoid error models via a combined adjustment of ellipsoidal, orthometric and gravimetric geoid height data. Received: 13 May 2004 / Accepted: 22 February 2004 / Published online: 4 May 2005 J Geod (2005) 79: 111–123 DOI 10.1007/s00190-005-0449-y.
[10] Al-Karagy, E., Doma, M., and Dawod, G. (2014) Towards an accurate discrimination of the local geoid model in Egypt using GPS/levelling data: A case study at Rosetta area, The International Journal of Innovative Science and Modern Engineering, V. 2, No. 11, pp.10-15.
[11] Dawod, G.,(2008). Towards the redefinition of the Egyptian geoid: Performance analysis of recent global geoid and digital terrain models, Journal of Spatial Science, 53, 1, 31-42.
[12] Aman S, Sulaiman H, Naim N, Talib KH, Wazir MA, Yusof OM. Deriving Orthometric Height Using Global Geopotential Models (GGMs). 7th IEEE Control and System Graduate Research Colloquium (ICSGRC); 2016. p. 193–6. doi: 10.1109/ ICSGRC.2016.7813326
[13] Essam M. Al-Krargy, Magdy M. Hosny, and Gomaa M. Dawod 2015, Investigating the Precision of Recent Global Geoid Models and Global Digital Elevation Models for Geoid Modelling in Egypt, Survey Research Institute National Water Research Center, Egypt. Regional Conference on Surveying & Development Sharm El-Sheikh, Egypt, 3-6 October 2015.
[14] Botai CM, Combrinck L. Global geopotential models from satellite laser ranging data with geophysical applications: A review. South Afr J Sci. 2012;108:1–10.
[15] Hamdy M. Ahmed, Elshewy A. Mohamed, Shaheen A. Bahaa. Evaluating two numerical methods for developing a local geoid model and a local digital elevation model for the Red Sea Coast, Egypt. Accepted 21 April 2021.Journal of King Saud University – Engineering Science.
[16] Bruinsma, S. L., Forste, C., Abrikosov, O., Marty, J. C., Rio, M. H., Mulet, S., Bonvalot, S.; The new ESA satellite-only gravity field model via the direct approach; Geophysical Research Letters, Vol 40, No. 14, p. 3607-3612, doi: 10.1002/grl.50716, 2013.
[17] Essam Mohamed Al-Karargy, Gomaa Mohamed Dawod, 2021 Optimum Combinations of GGM and GDEM Models for Precise National Geoid Modelling, Proceedings of Engineering and Technology Innovation, vol. 18, 2021, pp. 15-24.
[18] S. Doganalp: An Evaluation of Recent Global Geopotential Models for Strip Area Project in Turkey. Earth Sci. Res. J. Vol. 20, No. 3 (September, 2016) : C1 - C10.
[19] Liang, W., Xu, X., Li, J., and Zhu, G. (2018), The determination of an ultra-high gravity field model SGG-UGM-1 by combining EGM2008 gravity anomaly and GOCE observation data, Acta Geodaetica et Cartographica. Sinica, Vol. 47, No. 4, pp. 425-434.
[20] Wei Liang a, Jiancheng Li a b, Xinyu Xu a b, Shengjun Zhang c and Yongqi Zhao a . Research Geodesy and Survey Engineering—Articl A High-Resolution Earth's Gravity Field Model SGG-UGM-2 from GOCE, GRACE, Satellite Altimetry, and EGM2008. Volume 6, Issue 8, August 2020, Pages 860-878.
[21] Abedi, M., Norouzi, G.H. and Bahroudi, A. (2012) Support Vector Machine for Multi-Classification of Mineral Prospectivity Areas. Computers & Geosciences, 46, 272-283. https://doi.org/10.1016/j.cageo.2011.12.014.
[22] Sonny Rosenthal. Regression Analysis, Linear. November 2017. DOI:10.1002/9781118901731.iecrm0208

[23] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. Machine Learning, 20, 273-297. https://doi.org/10.1007/BF00994018.

[24] Kim, K.J. (2003) Financial Time Series Forecasting Using Support Vector Machines. Neurocomputing, 55, 307-319. https://doi.org/10.1016/S0925-2312(03)00372-2.

[25] Breiman, L. (2001) Random Forests. Machine Learning, 45, 5-32. https://doi.org/10.1023/A:1010933404324.

[26] Chen, W., Xie, X.S., Wang, J.L., Pradhan, B., Hong, H.Y., Bui, D.T., Duan, Z. and Ma, J.Q. (2017) A Comparative Study of Logistic Model tree, Random Forest, and Classification and Regression Tree Models for Spatial Prediction of Landslide Susceptibility

[27] P. Geurts, D. Ernst., and L. Wehenkel, Extremely randomized trees, Machine Learning, vol.63, pp.3-42, 2006.

[28] Essam M Fawaz , Engineering statistics Lectures, Al-Azhar University 2023. https://www.youtube.com/playlist?list=PL3MWTRyjjhqWSVgrUJliDapuyIRRaY-0s.

[29] Hamdy M. Ahmed, Elshewy A. Mohamed, Vasily I, Nilipovskiy. A Local Geoid for Egypt's Mediterranean Coast: A Model Based on Artificial Neural Networks.June 2022 International Journal of Geoinformatics 18(3):11.DOI:10.52939/ijg.v18i3.2193