International Journal on Advanced Science Engineering Information Technology

Vol.12 (2022) No. 3
ISSN: 2088-5334

Modeling Orbital Propagation Using Regression Technique and Artificial Neural Network

Nor'asnilawati Salleh [a,b,*], Siti Sophiayati Yuhaniz [a], Nurulhuda Firdaus Mohd Azmi [a]

[a] *Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia*
[b] *Engineering and Space Technology Division, Malaysian Space Agency, 42700 Banting, Selangor, Malaysia*
*Corresponding author: *norasnilawati@gmail.com*

***Abstract*— Orbital propagation models are used to predict the position and velocity of natural and artificial objects orbiting the Earth. It is crucial to get accurate predictions to ensure proper satellite operational planning and early detection of possible disasters. It became critical as the number of space objects grew due to many countries scrambling to explore space for various purposes such as communications, remote sensing, scientific mission, and many more. Physical-based and mathematical expression approaches provide orbital propagation with high accuracy. However, these approaches require substantial expenditure to provide suitable facilities and are complicated for those with no expertise in this field. The orbital propagation model is developed using regression techniques and artificial neural networks in this study. The aim is to have a reliable and precise orbital propagation model with minimal computational and cost savings. The past orbital data is used instead of complicated numerical equations and expensive tools. As a result, the trained orbital propagation model with accuracy up to 99.49% with a distance error of 18.73km per minute is achievable. The trained model can be improved further by modifying the network model and various input data. This model is also expected to provide vital information for organizations and anyone interested. Finally, this research can help organizations with insufficient resources to have their orbit propagation model without special tools or rely on other countries with satellite data at a lower cost.**

***Keywords*— Orbital propagation; prediction technique; time-series data; regression technique; artificial neural network.**

Manuscript received 28 May 2021; revised 10 Sep. 2021; accepted 3 Dec. 2021. Date of publication 30 Jun. 2022.

I. INTRODUCTION

Awareness of the situation in space is essential for future missions of satellites, preventive measures in the event of a collision, and the identification of untraceable space objects [1], [2]. All of these are critical issues affecting the space industry. An incident of a space object occurred in February 2009, involving the U.S. Iridium communications satellite and Russia's Cosmos 2251 communications satellite. One of the leading causes of this phenomenon is the inability orbital propagation models to obtain accurate information about the satellite's position [3], [4]. Therefore, getting a reliable and precise orbital propagation model is very important to avoid such things recurring. As the number of space objects increases, the risk of conflict will indirectly increase [5], [6]. Figure 1 shows the trend of space objects in Earth Orbit according to the types of objects that increase each year, which require attention from all parties involved [7]. It is because this issue not only can cause problems for the space industry but also endanger humankind.

Various approaches are used in the orbital propagation model: physical-based, mathematical expression, analytic solution, machine learning, data-driven, and hybrid method [8]–[12]. Figure 2 describes the orbital propagation approaches used. Each approach has its pro and cons. The physical-based gave an accurate result, but it is very costly and usually used by the country with the expert and financial capability. While the mathematical expression approach also gives high accuracy, it is not effortless and requires the expert to do the task. A data analytic solution is recently preferred as it is accessible and used despite limited expert and financial capabilities. Same with the machine learning, data-driven, and hybrid approach. Therefore, many researchers have conducted studies to find better solutions and make them accessible to everyone.

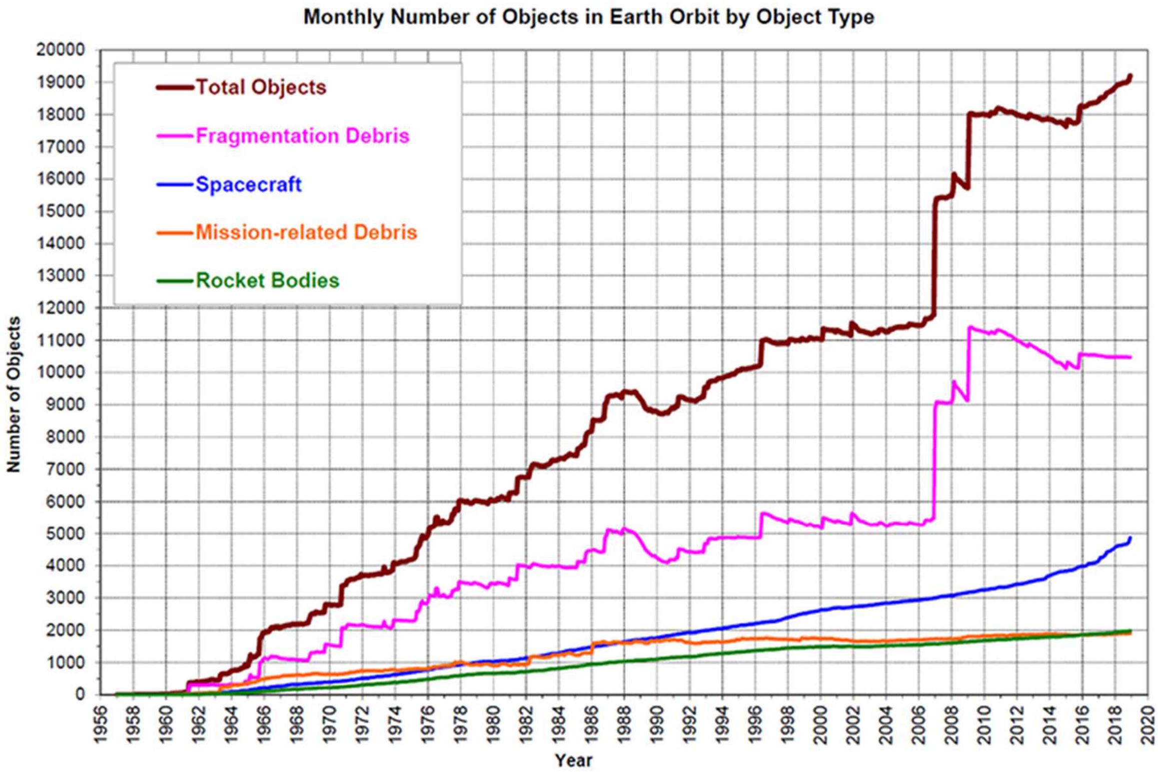

Fig. 1 A Number of Objects in Earth Orbit by Object Type [7]

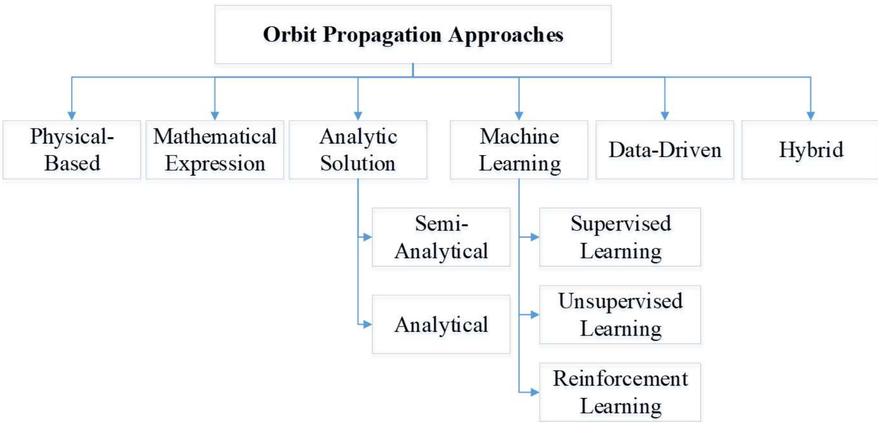

Fig. 2 Orbit Propagation Approaches [8]

The machine learning approach under the supervised learning technique is used for this study. Machine learning has solved a lot of complicated tasks by learning the data [13]. This approach enables the prediction process without explicitly modeling space objects and limited space environment information. Instead, the models are developed based on historical data only. Numerous studies conducted had improved the accuracy of the orbital propagation model using machine learning techniques. Among them are the support vector machine, nonlinear regression, artificial neural network (ANN), etc. [3], [14], [15]. These studies proved that the learning techniques could make predictions while maintaining the orbit propagation model's accuracy using historical data with limited resources. However, each method has its strengths and limitations depending on different data types, sizes, dataset behavior, etc. Therefore, further studies and analysis are indispensable to provide the best possible solution.

In this study, the regression technique and ANN are studied for modeling orbital propagation. Compared with previous research, this study aims to use minimal input data from the historical data to develop a reliable orbital propagation model. Furthermore, the minimal input data is expected to address data uncertainty and limited information that causes various assumptions during the modeling process. Besides that, an analysis is done to check the required input features for creating an optimal trained model network structure while still giving an accurate result. Thus, this study targets the developed orbit propagation with minimal computation and cost-effectiveness, such as processing speed, time, etc. This paper is arranged as follows. Section II describes the prediction technique used for modeling orbital propagation and methodology. Next, section III discusses the findings and limitations of the study. Finally, the last section summarizes the paper and shows the future work of this study.

II. Materials and Method

This section presents an overview of the prediction techniques and elaborates on the methodology used in this study.

A. Prediction Technique

The orbital propagation data is time-series data. Thus, the selected prediction techniques must be appropriate and suitable for this data type. Furthermore, it ensures that it can help improve the current model and the prediction process become more accurate and reliable. For this study, the regression technique and ANN technique are chosen.

The regression technique is used for the input data features selection. This technique is used as it is not complicated and easy to interpret. Also, it is suitable for time-series data [16], [17]. The feature selection is required to identify the most affected input data needed for the trained model. Thus, it ensures only the necessary data is used and speeds up the model development process while maintaining accuracy.

Meanwhile, the ANN is recommended to handle time-series data and nonlinear patterns [18], [19]. It is also proven to be the most common stochastic learning method for predicting [14], [15]. Besides that, the ANN can be modified to help the prediction process become more accurate and reliable. In this study, a nonlinear autoregressive exogenous (NARX) model is a solution used in the neural network for the time series data prediction [20]. It relates the current value of a time series to both: past values of the same series; and current and past values of the driving (exogenous) series — that is, of the externally determined series that influences the sequence of interest. This model is stated as follows.

$$y(t) = f(x(t-1), \dots x(t-d), y(t-1), \dots, y(t-d)) \qquad (1)$$

Where y is the predicted series, d is the past value of $y(t)$, and another series of $x(t)$. At the same time, the function f is a neural network.

B. Methodology

In this study, the first action taken is data collection. The data used in this study is the two-line element (TLE) data provided by North American Aerospace Defense Command (NORAD). The TLE is an open-source data accessible worldwide except for the United States' military data and alliances[21], [22]. NORAD supplies the TLE data that belong to satellites, namely special-interest satellites, weather, Earth resources satellites, communication satellites, navigation satellites, scientific satellites, and various satellites.

For this study, the space object information is extracted from TLE data and processed using the Simplified General Perturbations-4 (SGP4) model to prepare the trained data. The SGP4 model was used to ensure maximum predictive accuracy obtained [15], [23], [24]. In addition, it is to ensure only valid data is prepared and used for the trained model. The SGP4 model covers various elements and values of orbital interference; thus, it is a complete model compared to other available orbital propagation models such as Two-Body, J2, J4, GPS (SEM/YUMA), LOP SGP4, Astrogator, etc. [21], [23].

A specific space object's raw data from the TLE data is then processed to propagate the space object's position. The data sampling can be done for different intervals such as second, minute, hour, etc. For this study, the data sampling is done per minute as it is enough for the model to learn the data pattern. Data with a large interval will make it difficult for the model to understand the data, while the small sampling interval will cause too much data and slow down the modeling process. Special tools may be required to speed up the modeling process if it continues.

Next, the selection for data input features is made. The analysis is done to check the most affected features that need to be used in the trained model. Then, the design and modeling are executed. Finally, the trained model is evaluated to check the model's performance before the prediction can be made. Figure 3 shows the process flow of the study.

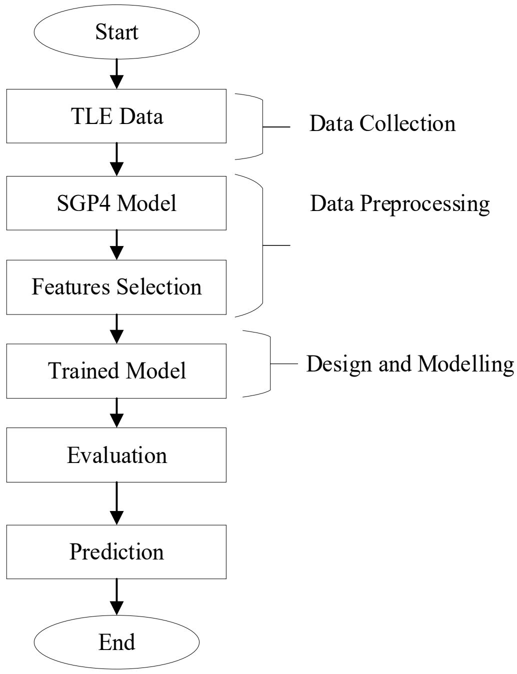

Fig. 3 Process flow

C. Data Preprocessing

The regression technique will process the data to identify which input features are most affected by the response output. The finding will help the model to train well. In this study, neighborhood component analysis (NCA) is used. It is a non-parametric method for selecting features to maximize regression and classification algorithms [25]. For NCA features selection regression, the response values are continuous. The function to perform the NCA features selection for regression is given as follows.

$$S = \{(x_i, y_i), i = 1,2, \dots, N\} \qquad (2)$$

Where N is observations $y_i \in \mathbb{R}$ are continuous, and response y is aimed to predict with the training set S. Figure 4 shows the X Position (ro_x) for satellite 25994 in 30 days of observation used in the NCA process.

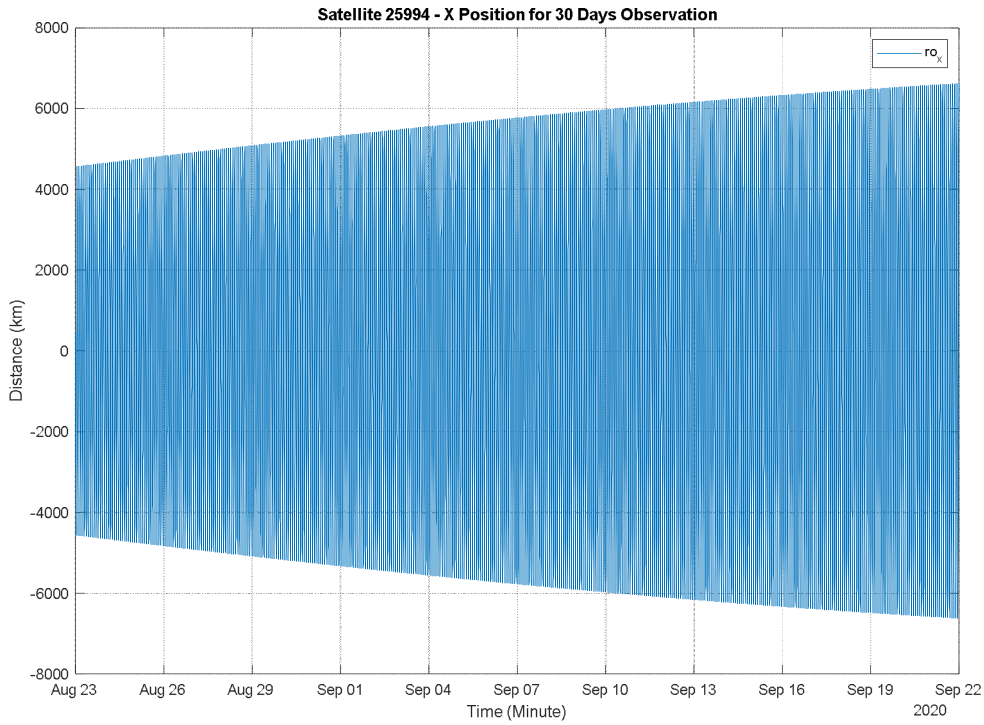

Fig. 4 X Position (ro_x) for 30 days observation

Meanwhile, Figure 5 shows a periodic pattern of X position (ro_x), which only the first 1440 observations plotted as it is difficult to see detailed features for all data samples as in Figure 4.

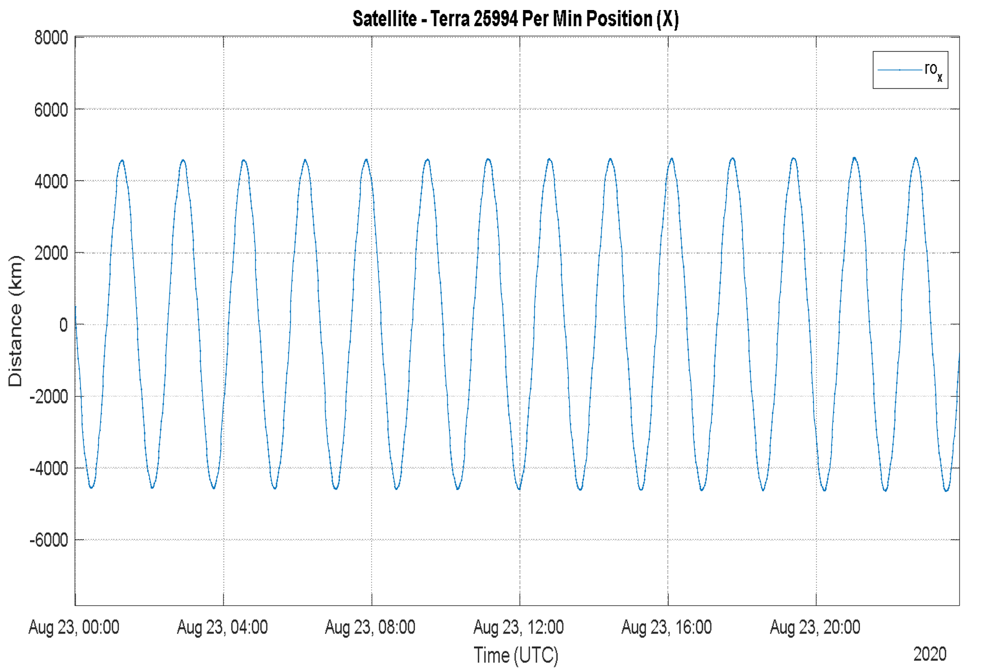

Fig. 5 X Position (ro_x) for one day observation

The X Position (ro_x) is the response y, and eight (8) input data features are related to this response, x_i. These features are (1) drag or radiation pressure coefficient, (2) eccentricity, (3) epoch, (4) an argument of perigee, (5) inclination, (6) mean anomaly, (7) mean motion, and (8) right ascension of the ascending node. These eight (8) features are the space object information extracted from the TLE data. This data information is also known as the Keplerian elements.

Next, the function achieves feature selection by regularizing the feature weights. The weights of the irrelevant features are zero, and relevant features will give a particular weight value. Table 1 lists the result of the weight features to the response X position (ro_x).

TABLE I
FEATURES WEIGHT RESULTS

Features	Features Index	Feature Weight
Drag	1	1
Eccentricity	2	1
Epoch	3	1.7
Argument of Perigee	4	3
Inclination	5	1
Mean Anomaly	6	3
Mean Motion	7	1
Right Ascension of The Ascending Node	8	1.8

From the feature selection result, the weights of all features are more than zero. Therefore, it is shown that all eight (8) features affect the response output. However, the most affected features with the highest value are the argument of perigee and mean anomaly. Thus, further analysis will be done to see whether using two features is enough to train the model or all features required for the training model.

D. Design and Modelling

The training data set typically consists of single-column data frame values for time series prediction. In this study, the values are the positions of a space object in coordinate value (ro_x, ro_y, ro_z). Each value is formulated as sequence such as, ro_x = [ro_x_1, ro_x_2, ro_x_3, ..., ro_x_n]. Figure 6 shows the NARX Neural Network Model architecture used to develop the orbital propagation model of eight (8) input data predictors (features).

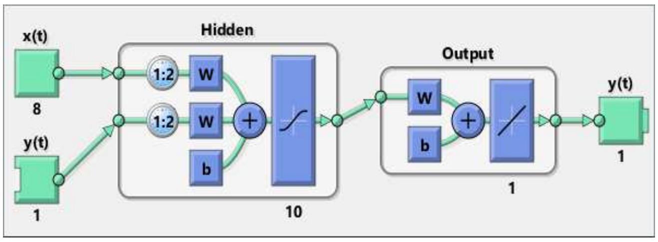

Fig. 6 NARX Neural Network Orbit Propagation Model using Eight (8) Features

The training input data consist of a predictor, x_t which includes eight (8) features (drag or radiation pressure coefficient; eccentricity; epoch; an argument of perigee; inclination; mean anomaly; mean motion and right ascension of the ascending node) and the response output, y_t (ro_x, ro_y, ro_z) position. Then, the network is created and trained in open-loop form as it is more efficient and can supply the network with correct past output.

Meanwhile, Figure 7 shows the NARX Neural Network orbit propagation model architecture using two (2) input data predictors, x_t (most affected features: the argument of perigee and mean anomaly). The training input data consist of two (2) predictors (an argument of perigee and mean anomaly) and the response output, y_t (ro_x, ro_y, ro_z) position.

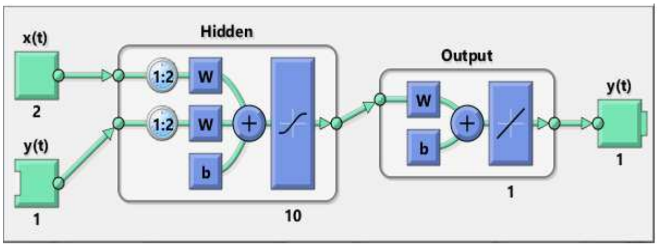

Fig. 7 NARX Neural Network Orbit Propagation Model using Two (2) Features

The data size used is 30 days that, is 43200 samples. Then it is divided into 30240 samples for training data; validation data is 6480 and testing data about 6480. Both models used ten (10) hidden layers, and the network is trained using the Levenberg-Marquardt algorithm. In addition, network training functions update weight and bias values are also according to this algorithm. After each iteration, the network updates the weights and bias values to get predicted values closer to target values. It minimizes the combination squared error and then determines the right combination to produce an optimized network. However, this algorithm requires more memory and less time to process. The training automatically stops when the generalization stops improving. This is indicated by the increase in the mean square error of the validation samples. Matlab is used as a tool for the modeling process.

E. Evaluation

Comparing the trained model and the actual result evaluates the trained model's performance for the evaluation process. The methods used to assess and validate the improved model's performance are the root mean squared error (RMSE). These methods can be calculated using the following equations.

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2} \quad (3)$$

Whereby y_t is the actual output, $\hat{y}_t$ is the predicted output, and N is the number of samples used. These functions also can be used as an optimization criterion of the improved model. Meanwhile, equation (4) and equation (5) are used to evaluate the model's performance.

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \quad (4)$$

$$P_{ML} = 100 - MAPE \quad (5)$$

The smaller the RMSE and MAPE score, the better the model performance. In contrast, the P_{ML} result is higher, the better the model's performance. At the end of this phase, the result will decide if the trained model can assist the orbit propagation process or vice versa.

III. RESULTS AND DISCUSSION

In this section, the result of this study is shown and discussed. Table 2 shows the trained model performance result using eight (8) and two (2) features.

TABLE II
TRAINED MODEL PERFORMANCE RESULTS

Position	8 Features		2 Features	
	MSE (km)	Speed Time (hh:mm:ss)	MSE (km)	Speed Time (hh:mm:ss)
ro_x	1.60e-2	0:02:00	1.16e-4	0:01:01
ro_y	0.029	0:02:07	5.64e-4	0:00:57
ro_z	5.59e-3	0:02:13	3.68e-4	0:00:56

The result shows that using two (2) features has the lowest MSE value and speed time result for each position performance value. Besides that, a one-day prediction ahead is also made using the trained model. The results show that using only two (2) features is sufficient to develop an accurate model compared to eight features with performance, P_{ML} up to 99.49%, and a distance error of 18.73km per minute for X Position (ro_x). While using eight (8) features, the achieved P_{ML} is 90.43% and distance error of 151.14km per minute. The same goes for Y Position (ro_y) and Z Position (ro_z) values, where the result using two (2) features is better than eight (8) features. However, the network configuration setting for Y Position (ro_y) and Z Position (ro_z) will be further investigated and retuning for a more accurate result. The summary of the evaluation results is listed in Table 3.

TABLE III
SUMMARY OF EVALUATION RESULTS

Number of Features	Position	RMSE (km)	MAPE (%)	P_{ML} (%)
2	ro_x	18.73	0.51	99.49
	ro_y	202.45	40.95	59.05
	ro_z	192.30	18.09	81.91
8	ro_x	151.14	9.57	90.43
	ro_y	218.04	46.48	53.53
	ro_z	1103.93	51.60	48.40

The results also indicate that the proper learning techniques must be considered to develop an accurate orbit propagation model. The ANN technique is performed well, and it is recommended due to its flexibility to deal with time-series data and proven to provide accurate and reliable results. Figure 8 shows the X position comparison value for one (1) pass of satellite 25994 orbiting the Earth, which takes about 125 minutes. It shows that the model trained used two (2) features is nearer to the actual result than the model trained used the eight (8) features.

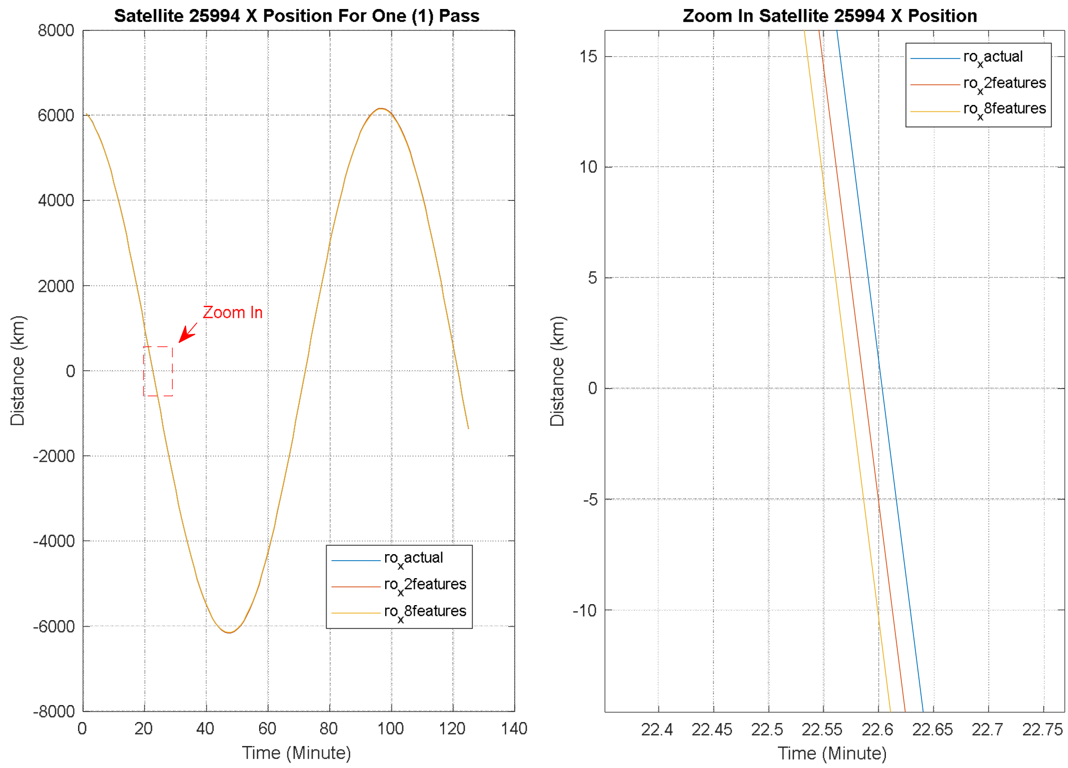

Fig. 8 Comparison between Two (2) Features and Eight (8) Features Trained Model for X Position

Meanwhile, Figure 9 shows the 3D illustration of satellite 25994 positions (ro_x, ro_y, ro_z) orbiting the Earth for one (1) pass. Through this figure, the track of space objects orbiting the Earth can be seen and analyzed. Also, the predicted position used two (2) features close to the space object's actual position compared to the predicted position used eight (8) features.

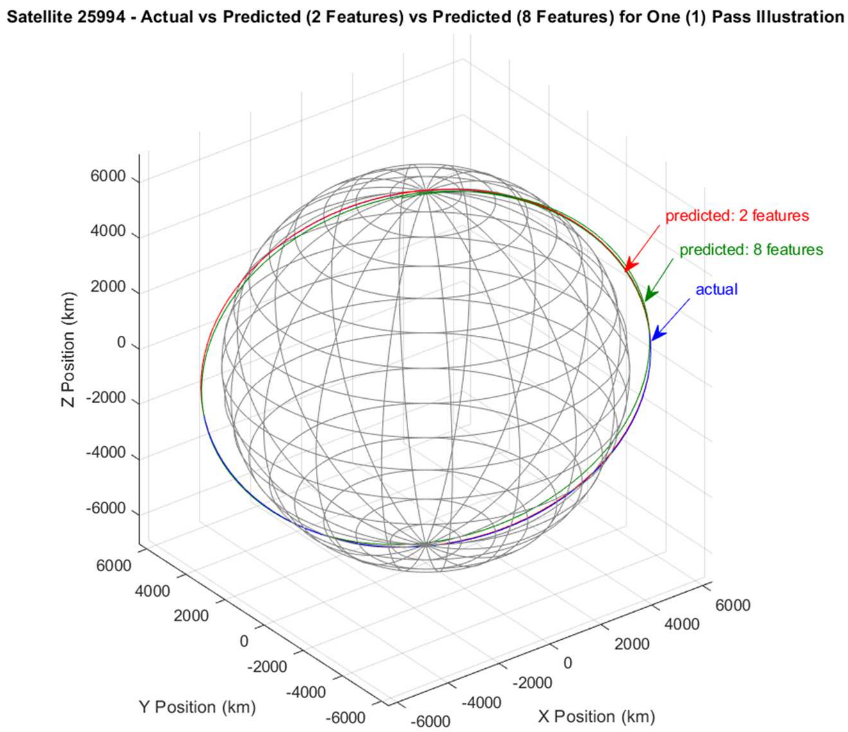

Fig. 9 3D Illustration of Satellite 25994 Orbital Propagation Actual vs. Predicted

In general, this study shows that the ANN can provide more flexible solutions. Compared to the previous studies, most of them used more features can produce a more accurate orbit propagation model. However, it is different from this study as it shows the opposite way. For example, a study by [9] shows that using more input features gave higher accuracy. However, they are using the Holt Winter, a tricky statistical technique that requires more input data. A similar problem can be solved more easily through ANN as it does not require much input to produce an accurate model. However, this should not make the primary basis as there is a possibility that the data configuration used in their study might be different from this study.

Other studies also select features by training all features, so it takes time. Meanwhile, this study chooses the features through the linear regression technique. The results can then identify the features that most affect the final development of the model by comparing all features as in Table I. Then, the researcher can explore further with various approaches such as network structure diversity for ANN, data configuration such as sizing training data, time steps, and others.

Nevertheless, to ensure the accuracy of the trained model, the TLE data has to be updated. It should not become an issue because NORAD is released the TLE data at regular periodic intervals [26]. The TLE data is known as the most comprehensive space object cataloging system in which the information is updated every 1-2 days for the expected target, and for the critical target, it will be updated 2-3 times every day [23][27][28]. However, this issue still needs to be considered because we do not want to rely too heavily on TLE data to make orbital propagation. Hence, further studies shall be done, especially in countries with no facilities and expertise in this field.

IV. Conclusion

In conclusion, the regression technique and ANN, the learning method, can do an orbital propagation model using the observed and historical data with minimal input data. Therefore, it is simpler and easier to be accessed by any organization or anyone interested. Using two (2) most affected features is enough to develop to train the model instead of using all features. It eventually has expedited the process and save processing time. Also, it reduces computational and cost-saving. The trained orbital propagation model's performance can also be up to 99.49%, with the 18.73 km distance error per minute. The main contributions of this work include the following fold. First, the NARX model using ANN is developed to learn the TLE dataset's complex data distribution. Second, the model can learn from the historical data by using minimal features. Finally, the evaluation of the trained model is done and proved to have an accurate orbit propagation model.

In general, this study can also conclude that many input features may reduce accuracy. Some of the input features do not have clear values and fluctuate due to surrounding factors such as perturbation and solar drag, resulting in unstable developed models.

Despite that, the trained model is only used for one class of satellites in this study. Other classes of space objects shall be studied in the future to ensure this kind of research can contribute more to space awareness. Besides that, some of the future work that can be done is studying other learning techniques that can create updated Keplerian elements instead of downloading the TLE data from NORAD. Such as randomly creating the TLE and optimizing using the genetic algorithm (GA) methods, particle swarm optimization (PSO), etc.

ACKNOWLEDGMENT

The authors are grateful to the Malaysian Space Agency and special gratitude to the Universiti Teknologi Malaysia for realizing this study.

REFERENCES

[1] J. N. Pelton, "A path forward to better space security: Finding new solutions to space debris, space situational awareness and space traffic management," *J. Sp. Saf. Eng.*, vol. 6, no. 2, pp. 92–100, 2019.

[2] E.-J. Choi *et al.*, "Performance analysis of sensor systems for space situational awareness," *J. Astron. Sp. Sci.*, vol. 34, no. 4, pp. 303–314, 2017.

[3] H. Peng and X. Bai, "Improving orbit prediction accuracy through supervised machine learning," *Adv. Sp. Res.*, vol. 61, no. 10, pp. 2628–2646, 2018.

[4] H. Peng and X. Bai, "Exploring capability of support vector machine for improving satellite orbit prediction accuracy," *J. Aerosp. Inf. Syst.*, vol. 15, no. 6, pp. 366–381, 2018.

[5] S. A. Kaiser, "Legal and policy aspects of space situational awareness," *Space Policy*, vol. 31, pp. 5–12, 2015.

[6] W. Lu *et al.*, "Microservice-based platform for space situational awareness data analytics," *Int. J. Aerosp. Eng.*, vol. 2020, 2020.

[7] "ARES | Orbital Debris Program Office | LEGEND." .

[8] N. Salleh, S. S. Yuhaniz, N. F. M. Azmi, and S. F. Sabri, "Enhancing simplified general perturbations-4 model for orbit propagation using deep learning: A review," in *ACM International Conference Proceeding Series*, 2019, vol. Part F1479, doi: 10.1145/3316615.3316675.

[9] J. F. San-Juan, I. Pérez, M. San-Martín, and E. P. Vergara, "Hybrid SGP4 orbit propagator," *Acta Astronaut.*, vol. 137, pp. 254–260, 2017.

[10] A. Celletti, C. Efthymiopoulos, F. Gachet, C. Galeş, and G. Pucacco, "Dynamical models and the onset of chaos in space debris," *Int. J. Non. Linear. Mech.*, vol. 90, pp. 147–163, 2017.

[11] T. Carrico, J. Carrico, L. Policastri, and M. Loucks, "Investigating orbital debris events using numerical methods with full force model orbit propagation," *Adv. Astronaut. Sci*, vol. 130, no. PART 1, pp. 407–426, 2008.

[12] S. Valk, A. Lemaître, and L. Anselmo, "Analytical and semi-analytical investigations of geosynchronous space debris with high area-to-mass ratios," *Adv. Sp. Res.*, vol. 41, no. 7, pp. 1077–1090, 2008.

[13] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of machine learning in wireless networks: Key techniques and open issues," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 4, pp. 3072–3108, 2019.

[14] I. E. Dawoodjee and M. Rajeswari, "Establishing a regression baseline for predicting satellite motion," *J. Appl. Technol. Innov. (e-ISSN 2600-7304)*, vol. 5, no. 1, p. 47, 2021.

[15] H. Peng and X. Bai, "Artificial Neural Network–Based Machine Learning Approach to Improve Orbit Prediction Accuracy," *J. Spacecr. Rockets*, vol. 55, no. 5, pp. 1248–1260, 2018.

[16] J. M. Valente and S. Maldonado, "SVR-FFS: A novel forward feature selection approach for high-frequency time series forecasting using support vector regression," *Expert Syst. Appl.*, vol. 160, p. 113729, 2020.

[17] A. Khosravi, L. Machado, and R. O. Nunes, "Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil," *Appl. Energy*, vol. 224, pp. 550–566, 2018.

[18] S. Bhardwaj, E. Chandrasekhar, P. Padiyar, and V. M. Gadre, "A comparative study of wavelet-based ANN and classical techniques for geophysical time-series forecasting," *Comput. Geosci.*, vol. 138, p. 104461, 2020.

[19] A. H. Nury, K. Hasan, and M. J. Bin Alam, "Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in northeastern Bangladesh," *J. King Saud Univ.*, vol. 29, no. 1, pp. 47–61, 2017.

[20] F. Di Martino and F. Delmastro, "High-Resolution Physiological Stress Prediction Models based on Ensemble Learning and Recurrent Neural Networks," in *2020 IEEE Symposium on Computers and Communications (ISCC)*, 2020, pp. 1–6.

[21] L. Chen *et al.*, *Orbital Data Applications for Space Objects*. Springer, 2017.

[22] H. Peng and X. Bai, "Machine learning approach to improve satellite orbit prediction accuracy using publicly available data," *J. Astronaut. Sci.*, vol. 67, no. 2, pp. 762–793, 2020.

[23] D. Vallado and P. Crawford, "SGP4 orbit determination," in *AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, 2008, p. 6770.

[24] M. Lin, M. Xu, and X. Fu, "A parallel algorithm for the initial screening of space debris collisions prediction using the SGP4/SDP4 models and GPU acceleration," *Adv. Sp. Res.*, vol. 59, no. 9, pp. 2398–2406, 2017.

[25] B. Alizadeh Savareh, A. Bashiri, A. Behmanesh, G. H. Meftahi, and B. Hatef, "Performance comparison of machine learning techniques in sleep scoring based on wavelet features and neighboring component analysis," *PeerJ Prepr.*, vol. 6, p. e27020v1, 2018.

[26] F. R. Hoots and R. L. Roehrich, "Models for Propagation of Norad Element Sets," Aerospace Defense Command Peterson AFB Co Office of Astrodynamics, 1980.

[27] B. Li, J. Sang, and J. Chen, "Achievable orbit determination and prediction accuracy using short-arc space-based observations of space debris," *Adv. Sp. Res.*, vol. 62, no. 11, pp. 3065–3077, 2018.

[28] J. Choi, J. H. Jo, H.-S. Yim, E.-J. Choi, S. Cho, and J.-H. Park, "Optical tracking data validation and orbit estimation for sparse observations of satellites by the OWL-Net," *Sensors*, vol. 18, no. 6, p. 1868, 2018.