Poverty Prediction System on Indonesian Population Using Deep Learning

Suaib Halim, Titik Khawa Abdul Rahman, Hoga Saragih, and Basuki Rahmat

***Abstract*—The poverty rate is an important metric to measure the welfare of the population of a country. Furthermore, the yearly data of this rate is also useful for predicting the number of poor people in subsequent years. Accurate poverty prediction and measurement is an important instrument for policymakers such as the government in order to pay attention to the living conditions of the poor. In this research, deep learning was used to predict the number of poor people living in urban areas, rural areas, and both combined. The model resulting from the training process to predict the number of poor people living in urban areas, rural areas, and both combined showed successful prediction percentages of 99.9542%, 95.8261%, and 99.9419%, respectively, while the testing results for these areas were 95.0139%, 91.1321%, and 96.5633%, respectively.**

***Index Terms*— Deep learning, poverty, prediction system, web server.**

I. Introduction

THE poor often live in deprivation. For example, they either live in a small house or have no shelter, have insufficient food, experience bad health conditions, as well as being illiterate and unemployed. Consequently, their prospects for achieving a better life are usually limited or uncertain [1]. In March 2020, the population of the poor in Indonesia was 26.42 million, and compared to March 2019, the number had increased by 1.28 million people. Similarly, the percentage of poor people in March 2020 was recorded as 9.78%, showing an increase of 0.37 points against March 2019, when it amounted to 9.41%. Meanwhile, one of the factors contributing to this increase in the poverty rate was the Covid-19 pandemic, which had an impact on the economic behavior and activities of the people in Indonesia [2].

One of the characteristics of poverty in Indonesia is the high disparity between urban and rural areas. For example, in March 2020, the poverty rate in urban areas was 7.38% (11.16 million people), while in rural areas it was almost twice as high, at 12.82% (15.26 million people) [2].

To be able to develop policies to reduce the poverty rate, accurate data needs to be obtained in order to predict the number of poor people in Indonesia, both in urban and rural areas, in the following years. Reliable prediction and measurement of poverty can serve as a valuable instrument for policymakers to focus attention on the living conditions of the poor. For example, the results of good data predictions are useful for evaluating government policies on poverty. In addition, comparing poverty across time and regions helps in determining targets for the poor in order to improve their conditions.

II. Related Works

Research on poverty in Indonesia is related to the government's poverty data collection system and the distribution of poverty aid funds to the poor. One of the systems used for data collection on the poor and the distribution of aid funds to the poor in Indonesia is e-Govind (Electronic Government for Aid Funds Distribution). E-Govind is an important software used by the government of Indonesia for data collection on poor families and for distributing financial assistance to poor families in Indonesia. This system uses fourteen criteria that describe poor people's conditions according to the Central Bureau of Statistics (BPS). Among the fourteen criteria for poverty, there are criteria in the form of photos of the conditions of the walls and roofs of houses that are identified as houses of poor people who are eligible to assistance. Photos of houses identified as poor people's houses are low-quality bamboo/thatch/wall houses with thatch roofs and without plaster walls.

The e-Govind system, apart from being used for channeling government assistance, should also be able to assist the government in predicting how many poor people there are in Indonesia. Therefore, this study aimed to predict the poverty rate in urban and rural areas using a deep learning method.

Deep learning is a method that has been proven to be able to solve time series prediction problems. Some examples can be found in [3]-[7]. Other relevant studies include Earthquake precursors by means of deep learning [8], time series category using Deep Learning [9], Deep learning technique aimed at source and amount of work prediction [10], Stability of time series prediction using deep learning [11], and Deep learning

Manuscript received February 4, 2022.

S. Halim is with Asia e University, Wisma Subang Jaya, No. 106, Jalan SS 15/4, 47500 Subang Jaya, Selangor, Malaysia. (email: suaibhalim@gmail.com).

T.K.A. Rahman is with Asia e University, Wisma Subang Jaya, No. 106, Jalan SS 15/4, 47500 Subang Jaya, Selangor, Malaysia. (email: titik.khawa@aeu.edu.my).

H. Saragih is with Asia e University, Wisma Subang Jaya, No. 106, Jalan SS 15/4, 47500 Subang Jaya, Selangor, Malaysia. (email: suaibhalim@gmail.com).

B. Rahmat is with the Informatics Department Universitas Pembangunan Nasional 'Veteran' Jawa Timur, Indonesia. (email: basukirahmat.if@upnjatim.ac.id).

to capture mutation information within multivariate time series estimation [12].

Furthermore, a study in [13] dealts more specifically with poverty prediction. Four deep learning approaches, namely, VGG-Net, Inception-Net, ResNet, and DenseNet, were used in [14] to extract inner features from daytime satellite imagery and then absolute least shrinkage and selection (LASSO) regression was applied to make poverty predictions. Performance improvements were made by integrating the squeeze and excitation (SE) module and focus loss into ResNet and DenseNet. The selection of wrapper features coupled with a decision tree algorithm to predict poverty levels in order to complete surveys and censuses has been conducted by the Central Statistics Agency (CSA) [15].

III. Proposed Methodology

A. Research Data

Data collection through e-Govind, as shown in Fig. 1, is primarily needed by local governments in Indonesia to find out the condition of the poor population in their area. The private sector may also contribute in terms of assisting in channeling funds, together with other partners. Meanwhile, some data can also be open to the public.

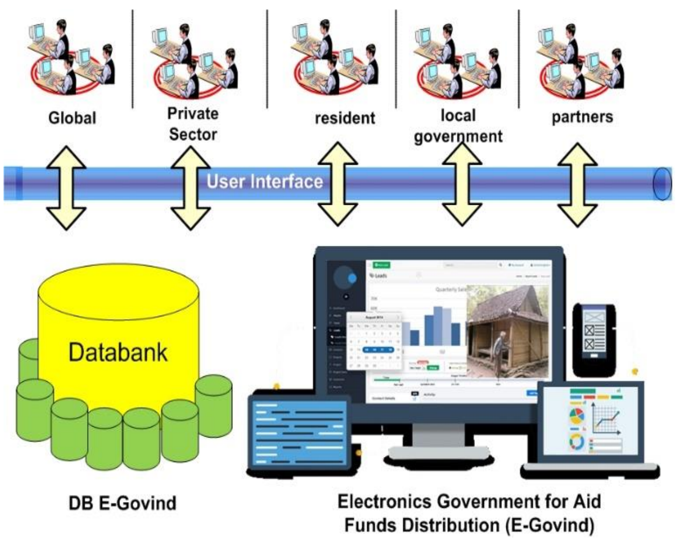

Fig. 1. E-Govind system.

The network architecture shown in Fig. 2 is commonly used for the implementation of e-Govind in Indonesia. A virtual private server (VPS) is needed, on which a web server and e-Govind web are installed. Admins and operators at the Data Center Office and the village office can access the online e-Govind system according to their respective authorities.

The authors are of the opinion that E-Govind should also be equipped with a poverty prediction system for the Indonesian population. This research provides an alternative to the existing poverty prediction system, using deep learning to improve the features of e-Govind.

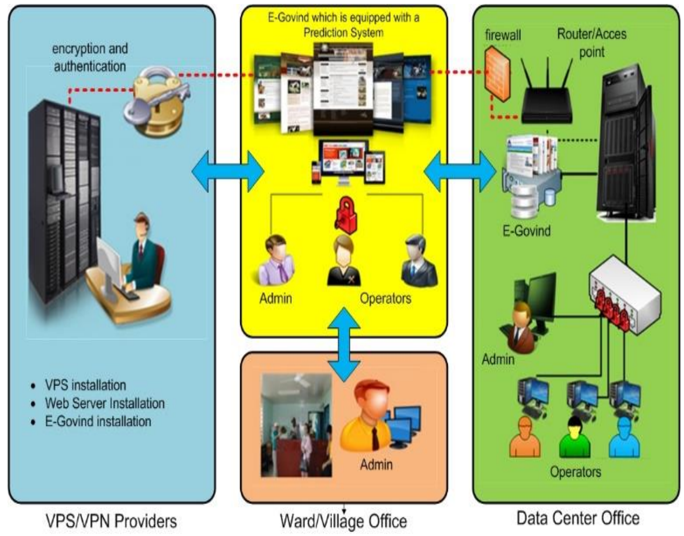

Fig. 2. E-Govind network architecture.

B. Model Structure

The data used in this study were collected from the Central Statistics Agency (CSA) and included data on the poor in urban and rural areas, and both combined, from 1999 to 2020. Some of these data were used for training and validation, while the rest were used for predictive testing. For example, the data for training was from 1999 to 2016 and was used to predict data in 2017 and 2018, while the poverty data for 2017 and 2018 was used for validation. Furthermore, by grouping the poverty data per two years from 1999 to 2016, nine group of years were obtained, which were used as input. Similarly, the predicted two-year data for the poor were used as a prediction group. Hence, a system model with nine inputs and one output was used, according to the design shown in Fig. 1, to ensure the structure of the training and validation data.

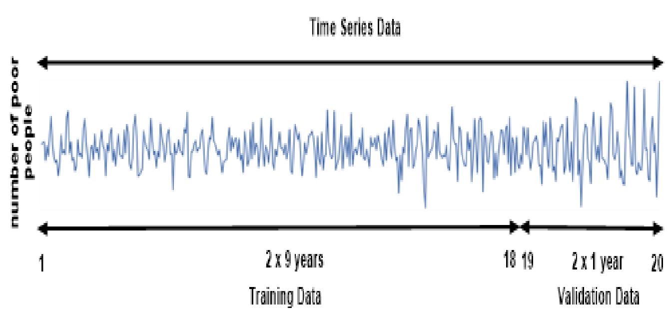

Fig. 3. The structure of the population poverty training data model and data validation.

C. Deep Learning Architecture

In accordance with the system model structure with nine inputs and one output for the number of poor people in urban and rural areas, the deep learning system architectural design is shown in Fig. 4.

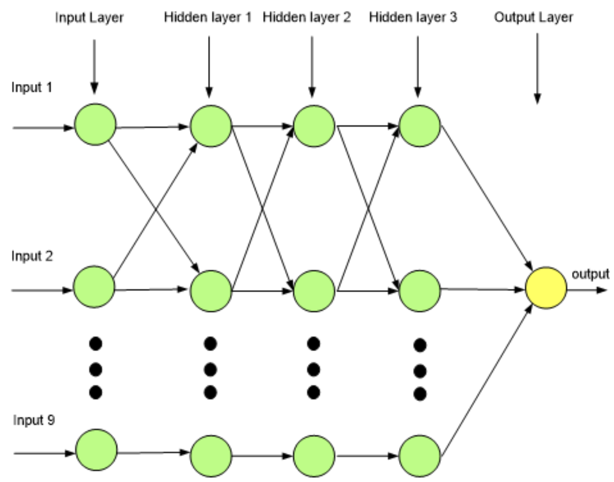

Fig. 4. Deep learning architecture.

The green nodes in the input layer and the hidden layers were designed to use the ReLU activation function, while the yellow node in the output layer was designed to use the Sigmoid activation function.

D. *Training and Testing Flowchart*

Furthermore, by using the deep learning network architecture design in Fig. 4, the training and testing processes for deep learning were designed as shown in Fig. 5.

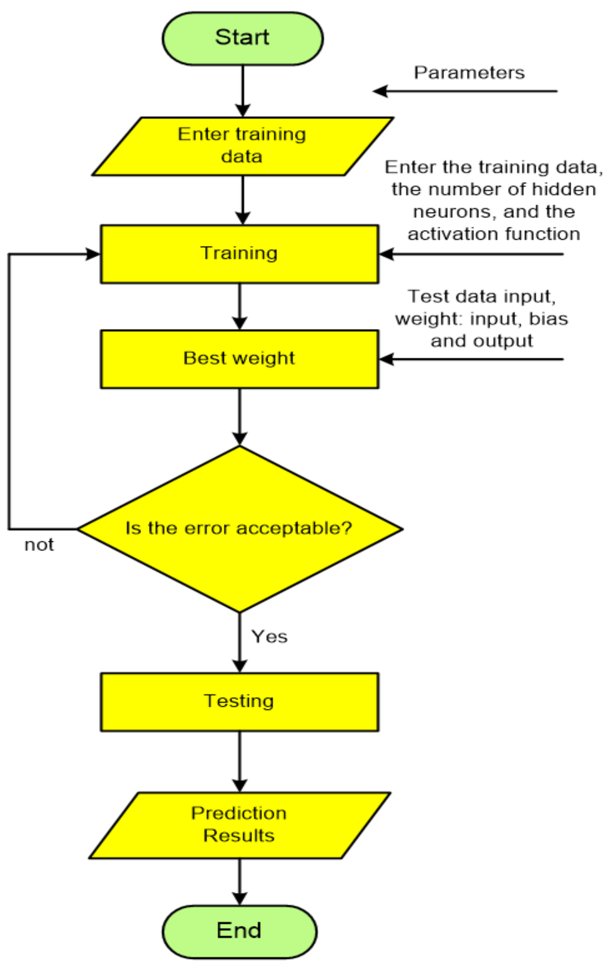

Fig. 5. Flowchart of the training and testing processes for deep learning.

Following the steps depicted in Fig. 5, the deep learning networks were trained to use data in the form of input and output pairs, which were eventually processed with weight and bias by using the appropriate activation function. Afterward, the deep learning network output was compared with the desired output until the error was accepted. The best weights from the training were used in the testing process, while the new data obtained were processed at each layer until the network output (a prediction) was obtained. The results of this prediction were used later in this study to predict the number of poor people in urban and rural areas. The results of the performance assessment in the prediction process were expressed in the form of mean square error (MSE), root mean square error (RMSE), absolute average error percentage (MAPE), and percentage of success (%).

IV. Results and Discussion

Based on the data used, namely the urban poor population, the rural poor population and both combined, the experimental results were as described below. The first, second, and third training processes were the predictions of the poor living in urban areas, rural areas, and both combined, respectively. The training data used referred to the poor from 1999 to 2016 and were used to predict the poor population in 2017 and 2018. The results of the training process of the deep learning system using data on the poor population in urban areas, rural areas, and both combined until the 1000th iteration are shown in Fig. 6, Fig. 7, and Fig. 8, respectively.

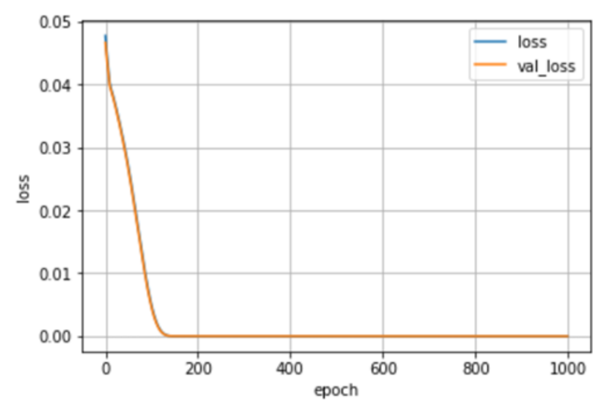

Fig. 6. Results of the process of reducing the mean squared error of the deep learning training process for data on the urban poor.

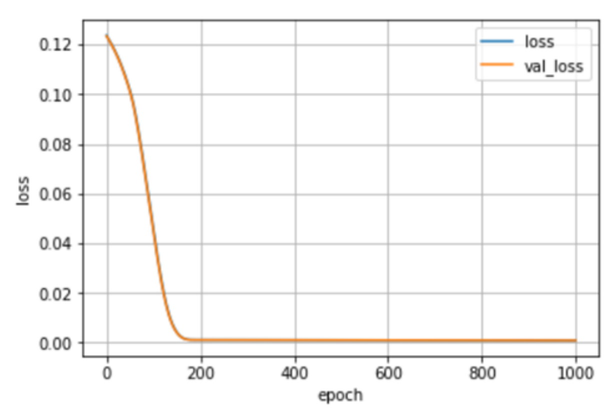

Fig. 7. Results of the process of reducing the mean squared error of the deep learning training process for data on the rural poor.

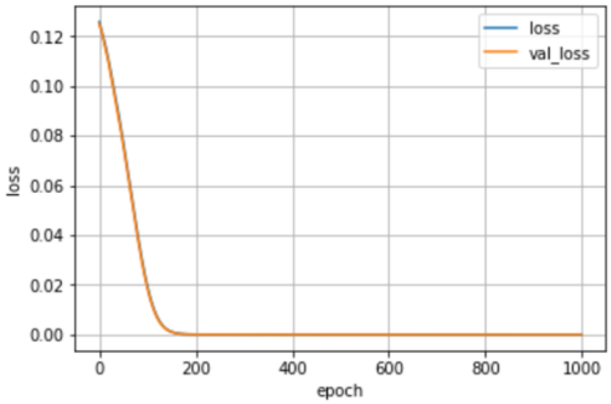

Fig. 8. Result of the process of decreasing the mean squared error of the deep learning training process for data on the urban and rural poor.

Similarly, the results of the deep learning training process for the poverty prediction system for urban, rural, and urban and rural residents are presented briefly in tabular form in Table I.

TABLE I.
RESULTS OF THE TRAINING PROCESS

Population	Iterations	MSE	RMSE	MAPE	Success percentage (%)
Urban residents	1000	21804232.5	4669.5	0.0458	99.9542
Rural residents	1000	441723364952.5	664622.7	4.1739	95.8261
Urban and rural residents	1000	233228690.0	15271.8	0.0581	99.9419

Furthermore, to test the poverty prediction system for these areas, the data from 2001 to 2018 were used to predict the poor populations in 2019 and 2020. The results of the prediction testing process using the proposed deep learning method are shown in Fig. 9, 10, and 11 respectively.

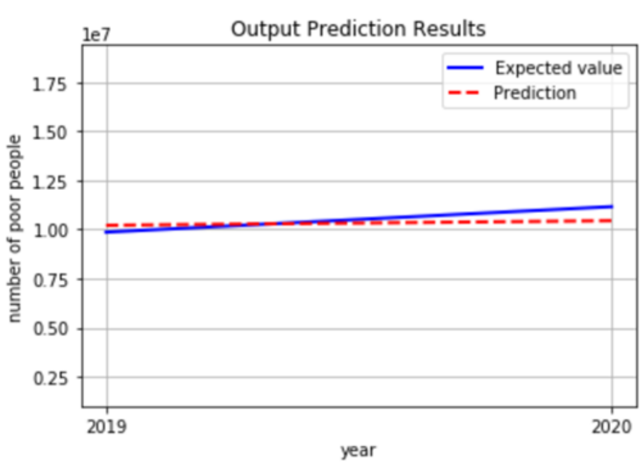

Fig. 9. Output of the first Extreme Learning Machine testing process.

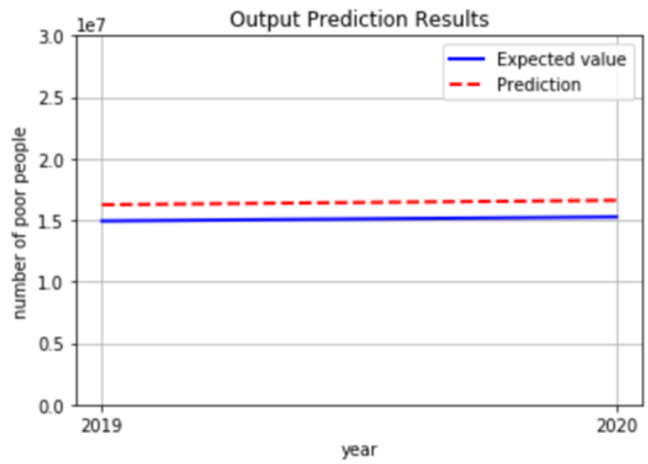

Fig. 10. Output of the deep learning testing process for the rural poor.

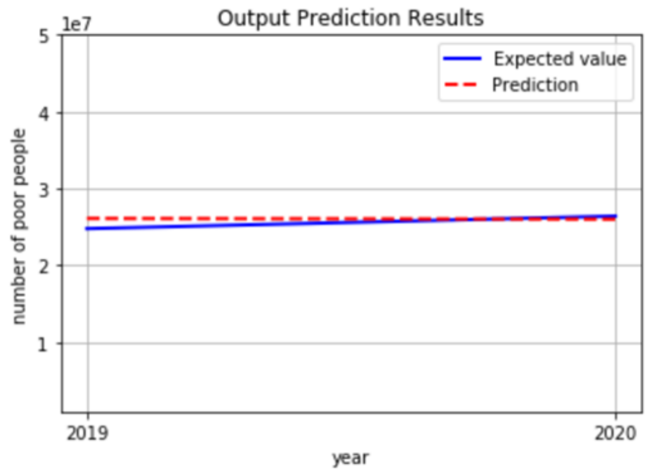

Fig. 11. Output of the deep learning testing process for the urban and rural poor.

According to Table II, the results of the deep learning testing of the prediction of the poor populations in 2019 and 2020 are presented in full in tabular form.

TABLE II
RESULTS OF POVERTY PREDICTION TESTING OF INDONESIAN POPULATIONS

Population	MSE	RMSE	MAPE	Success Percentage (%)
Urban residents	318072817354.0	563979.4	4.9861	95.0139
Rural residents	1792048067300.0	1338673.9	8.8679	91.1321
Urban and rural residents	961048828514.0	980330.9	3.4367	96.5633

From the results of the prediction testing process for the poor living in cities, villages, and both combined, as shown in Table II, it can be seen that changes in data patterns for poor people living in cities are relatively easier to recognize than for those in living villages. This is indicated by the percentage of success for predicting the poor living in cities being higher than for those living in villages. For example, the prediction success percentage of those living in cities was 95.0139%, while for those living in villages it was 91.1321%. Likewise, the changes in data patterns for the poor living in cities and villages with a larger number of data on the poor, are more easily recognized than for residents of cities or villages only. This is indicated by the percentage of success in predicting the number of poor people living in cities and villages combined, which was 96.5633%. Therefore, the deep learning method provided excellent predictive results for the number of poor people, with a prediction success percentage higher than 90%.

V. Conclusion

This study investigated the implementation of deep learning to predict the number of poor people in Indonesia. The results showed percentages of successful prediction of the poor population in urban areas, rural areas, and both combined of 95.0139%, 91.1321%, and 96.5633% respectively. The method proposed in this study can be valuable as an instrument for policymakers focusing on poverty reduction in Indonesia.

References

[1] M. P. Todaro and S. C. Smith, "Economic Development, 12th Edition," Pearson, 2015.

[2] S. D. S. K. Sosial, "Penghitungan Dan Analisis Kemiskinan Makro Indonesia Tahun 2020," Badan Pusat Statistik, 2020.

[3] W. Liu, G. Zeng, and K. Hu, "Growth Scale Prediction of Big Data for Information Systems Based on a Deep Learning SAEP Method," *IEEE Access*, vol. 8, pp. 62883–62894, 2020.

[4] S. Cui and I. Joe, "Collision prediction for a low power wide area network using deep learning methods," *J. Commun. Networks*, vol. 22, no. 3, pp. 205–214, Jun. 2020.

[5] J. Qu, F. Liu, Y. Ma, and J. Fan, "Temporal-Spatial Collaborative Prediction for LTE-R Communication Quality Based on Deep Learning," *IEEE Access*, vol. 8, pp. 94817–94832, 2020.

[6] B. Chen, Y. Liu, C. Zhang, and Z. Wang, "Time Series Data for Equipment Reliability Analysis With Deep Learning," *IEEE Access*, vol. 8, pp. 105484–105493, 2020.

[7] W. Jiang and H. D. Schotten, "Deep Learning for Fading Channel Prediction," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 320–332, 2020.

[8] B. Rahmat, E. Joelianto, F. Afiadi, A.D.L. Frandeza, R.A. Kurniawan, E.Y. Puspaningrum, B. Nugroho and D.S.Y. Kartika, "Comparison of B-Value Predictions as Earthquake Precursors using Extreme Learning Machine and Deep Learning," *Internetworking Indones. J.*, vol. 12, no. 1, pp. 47–52, 2020.

[9] J. Azar, A. Makhoul, R. Couturier, and J. Demerjian, "Robust IoT time series classification with data compression and deep learning," *Neurocomputing*, vol. 398, pp. 222–234, 2020.

[10] J. Bi, S. Li, H. Yuan, and M. C. Zhou, "Integrated deep learning method for workload and resource prediction in cloud systems," *Neurocomputing*, vol. 424, pp. 35–48, Feb. 2021.

[11] P. Liu, J. Liu, and K. Wu, "CNN-FCM: System modeling promotes stability of deep learning in time series prediction," *Knowledge-Based Syst.*, vol. 203, p. 106081, 2020.

[12] J. Hu and W. Zheng, "A deep learning model to effectively capture mutation information in multivariate time series prediction," *Knowledge-Based Syst.*, vol. 203, p. 106139, 2020.

[13] K. D. Hartomo, S. Yulianto, and A. Valentina, "A New Model of Poverty Index Prediction Using Triple Exponential Smoothing Method," in *2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 2020, pp. 76–79.

[14] Y. Ni, X. Li, Y. Ye, Y. Li, C. Li, and D. Chu, "An Investigation on Deep Learning Approaches to Combining Nighttime and Daytime Satellite Imagery for Poverty Prediction," *IEEE Geosci. Remote Sens. Lett.*, pp. 1–5, 2020.

[15] A. Pangestu, D. R. Wijaya, E. Hernawati, and W. Hidayat, "Wrapper Feature Selection for Poverty Level Prediction Based on E-Commerce Dataset," in *2020 International Conference on Data Science and Its Applications (ICoDSA)*, 2020, pp. 1–7.

Suaib Halim earned his informatics degree from STMIK Budi Luhur in 1997. He obtained a master's degree in informatics engineering from the Surabaya Institute of Technology in 2005. He is currently pursuing an ICT doctoral degree at Asia e University and is a staff member of the Department of Informatics, Jayapura University of Science and Technology, Papua. His research interests are deep learning, intelligent systems, and software.

Titik Khawa Abdul Rahman received her bachelor's degree in electronic and electrical engineering from Loughborough University, UK in 1988. She received her doctor's degree in power system engineering from the University of Malaya in 1998. She is a senior member of IEEE. Her research interest is primarily in the area of intelligence systems for power system engineering and many other application areas of intelligent systems.

Hoga Saragih received his bachelor's degree in electrical engineering from Krida Wacana Christian University, Indonesia in 1998. He completed his master's degree in electrical telecommunication engineering at the Department of Electrical Engineering, University of Indonesia in 2001. He received his PhD in electrical telecommunication engineering from the Department of Electrical and Engineering, University of Indonesia in 2008. His research interest is primarily in the area of mobile communications, with special emphasis on ALOHA and CDMA.

Basuki Rahmat received his bachelor's degree in instrumentation physics from Institut Teknologi Sepuluh Nopember Surabaya in 1995. He received his master's degree in instrumentation and control from Institut Teknologi Bandung in 2000. He received his doctoral degree in electrical engineering from Institut Teknologi Sepuluh Nopember Surabaya in 2018. Currently, he is a staff member of the Informatics Department, Universitas Pembangunan Nasional 'Veteran' Jawa Timur, Indonesia. His research interests are machine learning, deep learning, intelligent systems, soft computing, image and video processing, intelligent control, drone and robotics systems, PHP, Arduino, and Python Programming. He is a member of ASIOTI, IEEE and IAENG.