An Analytical Study of Areas with Intelectuall Disability in East Java and Their Prediction Using Ensemble Filtering

R. B. Utami [1,4*], M. Fannani [2], S. Sumardiyono [2] and T. Herlambang [3]

[1] *Doctoral Program on Public Health, Universitas Sebelas Maret, Surakarta, Indonesia.*
[2] *Department of Occupational Health and Safety, Vocational School, Universitas Sebelas Maret, Surakarta 57126, Indonesia.*
[3] *Department of Information System, Faculty of Economic Business and Digital Technology, Universitas Nahdlatul Ulama, Surabaya, Indonesia.*
[4] *Satria Bhakti College of Health Sciences, Nganjuk, Indonesia.*

Abstract: According to the WHO, intellectual disability refers to individuals with limited intellectual abilities, who face difficulties in meeting basic needs. This condition affects both children and adults, with children requiring full support from parents and their environment, as they are prone to discrimination. It is common in several Indonesian provinces, notably West Java, Central Java, and East Java, particularly in rural or remote areas where access to health facilities and knowledge is limited. In this study, the prevalence of intellectual disability in East Java is analyzed and predicted using machine learning methods based on the number of affected regions per district/city. The methods applied are the Ensemble Kalman Filter (EnKF) and Ensemble Kalman Filter Square Root (EnKF-SR), both of which are developments of the Kalman Filter. The EnKF method achieved the best RMSE of 50.68 with 600 ensembles and R = 0.01, while the EnKF-SR achieved the best RMSE of 69.16 with 800 ensembles and R values of 0.01 and 0.03.

Keywords: *intellectual disability; East Java; prediction; machine learning; ensemble filtering.*

Mathematics Subject Classification (2020): 62M20, 68T07, 92C50.

---

* Corresponding author: `mailto:budiutami2201@gmail.com`

1   Introduction

The World Health Organization's International Classification of Disease (ICD-10) defines intellectual disability as a mental disorder that simultaneously refers to the people of below average intellectual capability marked by the limitation of two or more skills, namely communication, self-care, social skills, safety, and personal health. In Indonesia, intellectual disability is experienced not only by adults, but also by children. According to Pendataan Program Perlindungan Sosial (PPLS) in 2011, the number of children with intellectual disability was 30,460 children out of a total of 13,572. This data is spread throughout Indonesia, with the most provinces in Central Java, East Java, and West Java [1].

According to the Basic Health Research Indonesia, 2013 (Riset Kesehatan Dasar known as Riskesdas), 14.3% of Indonesian household have a patient with a mental disorder and a majority are in rural area [2]. Knowledge, economic, and information factors and access to health facilities in rural areas are also limited. This lack of knowledge, for example, can be evidenced by the many cases of shackling that occur. Until now, most cases of shackling are in rural areas with the main factors being low economic problems and the inability to go to a mental hospital [3].

Almost all of Indonesia is familiar with the term "Pasung". Pasung is a long-standing custom in West Java and beyond. Pasung has traditionally involved a person who is agitated or considered at high risk of doing harm, either to themselves or others, being shackled using a wooden log [4]. Although it sounds less humane, it is still widely found in rural areas in Indonesia.

Especially for people with intellectual disability who are still children, this requires special attention and treatment. One of the most significant factors that can help improve quality of life for children with developmental disabilities is parental resilience [5]. On a broader scale, the role of the local government is needed as a health facilitator so that all people with intellectual disability have the same rights and access to health services in terms of legal guarantees in the form of Regional Regulations (Perda) [6].

In this research, the authors try to analyze and predict the number of regions in East Java province with intellectual disability survivors. Predictions are performed using Machine Learning (ML). ML is a subset of Artificial Intelligence [7]. Machine Learning (ML) has potential to enable a machine to learn from experience and apply this learning in future uncertainties [8]. ML methods can be divided into supervised learning, unsupervised learning, and reinforce learning [9].

The prediction methods used in this research are Ensemble Kalman Filter (EnKF) and Ensemble Kalman Filter-Square Root (EnKF-SR). The Ensemble Kalman Filter (EnKF) is an algorithm which is widely employed in the engineering community for the estimation of the state of partially-observed dynamical systems whose dynamics are governed by a nonlinear agent [10]. In the EnKF, the system's mean and covariance are approximated by an ensemble of states randomly sampled from the probability distribution. In the forecast, each state is advanced by the dynamical model; in the analysis step, each ensemble member is updated with the new observation to minimize the posterior covariance [11]. Meanwhile, EnKF-SR is a development of EnKF.

Previously, in 2023, S. Yoo et al. [12] conducted research about predicting wildfire spread using Ensemble Kalman Filter and polyline simplification. The result is algorithm successfully reduced computational time required for data assimilation [12]. In 2022, M. Y. Anshori et al. [13] conducted research about estimation XYZ company profitability

using H-Infinity and Ensemble Kalman Filter (EnKF). The result is that the performance of EnKF is more promising than H-Infinity with error percentage around 5-8% [13]. Still in 2022, A. Muhith et al. [14] conducted research about the stock of TC blood in PMI Gresik. The result is that EnKF-SR is better than UKF with an error less than 3% [14].

R. Ghostine et al. [15] in 2021, employed EnKF to predict COVID-19 pandemic evolution in Saudi Arabia complementing SEIR with vaccination and achieved promising short-term predictions with a relatively small ensemble size [15]. In 2021, H. Cevallos et al. [16] conducted research about electrical estimation using EnKF and Particle Filter. The result is that the Particle Filter achieves greater stability, higher estimation accuracy, and better performance [16]. In 2020, E. K. Korir et al. [17] conducted research on forecasting electricity prices using EnKF. The result is that EnKF gives better prediction with an RMSE value of 0.0089 [17].

2 Research Method

In this research, the data source refers to the official website of BPS Jatim (`www.jatim.bps.go.id`). The data contains 3 columns and 30 rows which represent the results recorded in 2014 and 2018. After that, the data is processed statistically to find insights and identify patterns. Table 1 shows the dataset used in this research, while Figure 1 illustrates the research methodology.

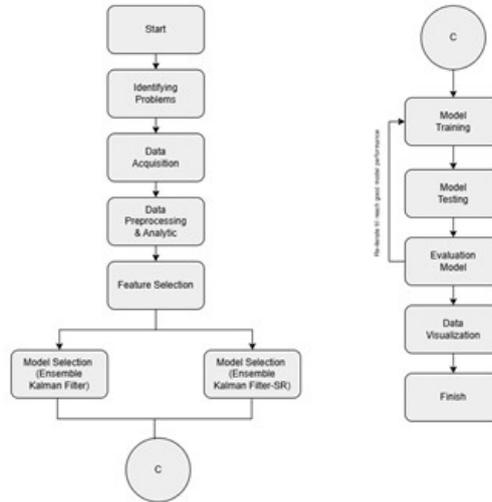| City/Regency	Number of Villages with Intellectual Disability in 2014	Number of Villages with Intellectual Disability in 2018
Pacitan	150	142
Ponorogo	236	238
Trenggalek	137	139
Tulungagung	239	243
Blitar	227	226
Kediri	277	298
Malang	338	303
Lumajang	147	140
Jember	192	169
Banyuwangi	166	172
Bondowoso	169	151
Situbondo	99	86
Probolinggo	254	209
Pasuruan	274	248
Sidoarjo	273	265
Mojokerto	234	229
Jombang	266	279
Nganjuk	231	214
Madiun	182	159
Magetan	202	199
Ngawi	175	159
Bojonegoro	345	327
Tuban	235	247
Lamongan	360	325
Gresik	276	235
Bangkalan	173	150
Sampang	145	148
Pamekasan	136	122
Sumenep	231	209
Kota Kediri	35	37
Kota Blitar	18	18
Kota Malang	24	38
Kota Probolinggo	27	26
Kota Pasuruan	27	26
Kota Mojokerto	28	28
Kota Madiun	9	17
Kota Surabaya	21	20
Kota Batu	22	20

Table 1: Dataset.

Figure 1: Research Methodology.

2.1   Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a statistical approach that examines the presence of multiple hidden features and patterns in a dataset [18]. In other words, Exploratory Data Analysis (EDA) is a way to explore the data with the aim of extracting useful and actionable information from it [19]. The data was collected in 2014 and 2018 and covers the entire province of East Java. Table 2 presents a summary of the central tendency of the data and Figure 2 plots the number of villages with intellectual disability in East Java.

Statistic	Number of Villages with Intellectual Disability in 2014	Number of Villages with Intellectual Disability in 2018
mean	174.97	38.0
min	9.0	17.0
max	360.0	327.0
std	100.24	93.73

Table 2: Measurement of the central tendency of villages with intellectual disability in East Java.
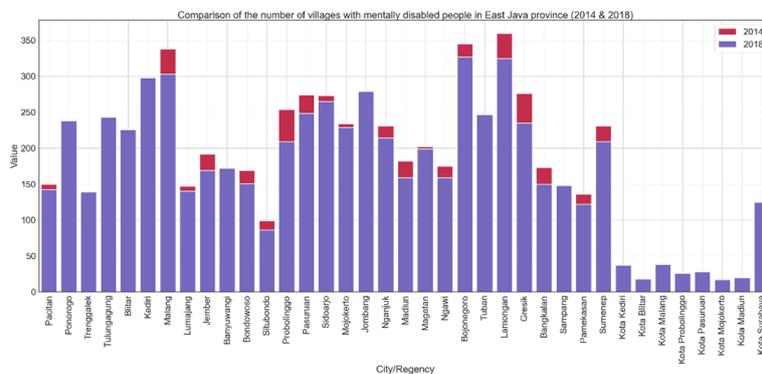

Figure 2: Plot comparison of the number of areas with people with intellectual disability in East Java.

From Figure 2, it can be seen that in some areas of East Java province, there has been an increase in areas with people with intellectual disability. Lamongan district has the highest level of mentally retarded people in East Java province and Blitar City along with Mojokerto City became the regions with the lowest levels. Then, the previous dataset can be tested using the Wilcoxon test to determine the relationship between condition variables in 2014 and 2018. Below, Figures 3 and 4 plot the distribution of data and Table 3 presents the result of the Wilcoxon test.
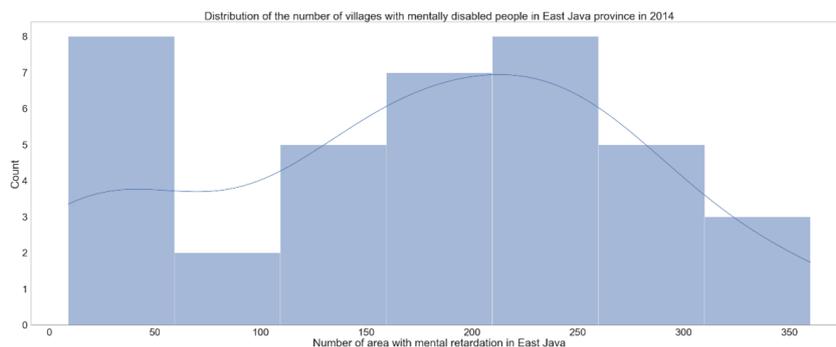

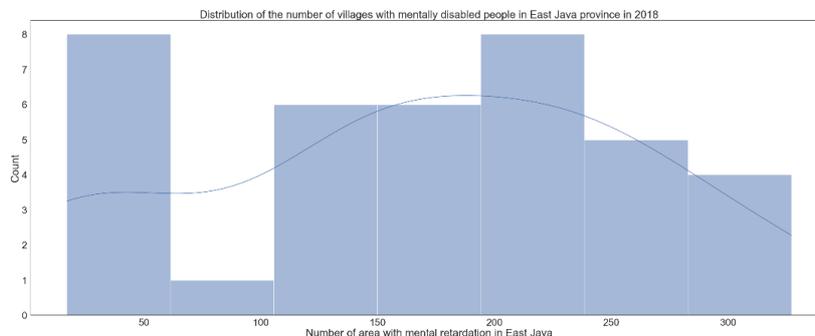**Figure 3**: Histogram of distribution of data in 2014.

Figure 4: Histogram of distribution of data in 2018.

Wilcoxon statistic	176.0
P-Value	0.013

Table 3: Result of the Wilcoxon test.

From Figures 3 and 4 and also Table 3 above, it can already be determined which distribution of data is not normal and based on the Wilcoxon test, it is clear that there is a relationship between the values in the columns in 2014 and 2018. Indeed, many papers apply non-parametric tests such as the Mann-Whitney test or the Wilcoxon test to compare groups when the data do not seem completely normally distributed [20].

2.2 Preprocessing

After that, the next step is normalization. It refers to scaling the dataset so that the data values lie between 0 and 1 [21]. One advantage of MinMax Scaler is that it allows mapping features with very different ranges into the same interval while keeping all information since distance ratios are preserved [22]. Below in Equation (1) is the MinMax Scaler function

$$x^1 = \frac{x_i - \min(x)}{\max(x) - \min(x)}. \tag{1}$$

2.3 Feature selection

Next, the process moves on to feature selection. Based on the case study and dataset variables, the variable *Number of Villages with Impaired Patients in 2018* is selected as the dependent/target variable.

2.4 Ensemble Kalman Filter (EnKF) prediction model

The EnKF was developed as a version of the Kalman Filter for large-scale problems. It is a Monte Carlo based implementation of the Kalman Filter for non-linear state estimation problems [23]. Similar to the KF, the EnKF consists of recursive application of a forecast step and an analysis step [24]. Below in Figure 5, the EnKF algorithms are presented.
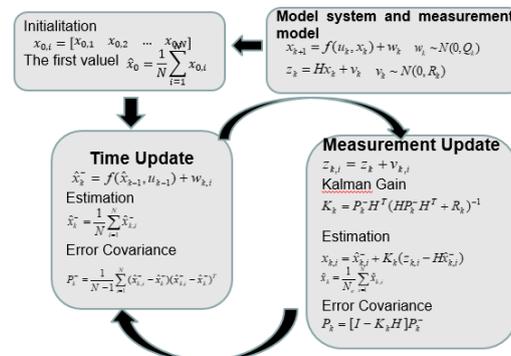

Figure 5: Ensemble Kalman Filter (EnKF) algorithm.

2.5 Ensemble Kalman Filter Square Root (EnKF-SR) prediction model

Below Figure 6 present the function of Ensemble Kalman Filter Square Root.

2.6 Model evaluation

At the evaluation stage, the model trained and tested is calculated for accuracy based on the resulting error value. The root mean squared error (RMSE) and the mean absolute error (MAE) are two standard metrics used in model evaluation. This study uses the Root Mean Square Error (RMSE) method to calculate the error value generated by the
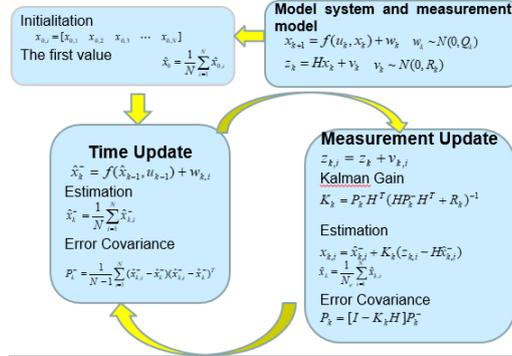
Figure 6: Ensemble Kalman Filter Square Root (EnKF-SR) algorithm.

model. The function of the Root Mean Square Error (RMSE) is presented in Equation (2) as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}.$$

(2)

3   Results and Discussion

In this research, the authors tried to implement Ensemble Kalman Filter (EnKF) and Ensemble Kalman Filter Square Root (EnKF-SR) to predict the number of areas in East Java with people with intellectual disability. The ensemble numbers used are 600 and 800. The overall simulation results are shown in Figures 7 to 10.

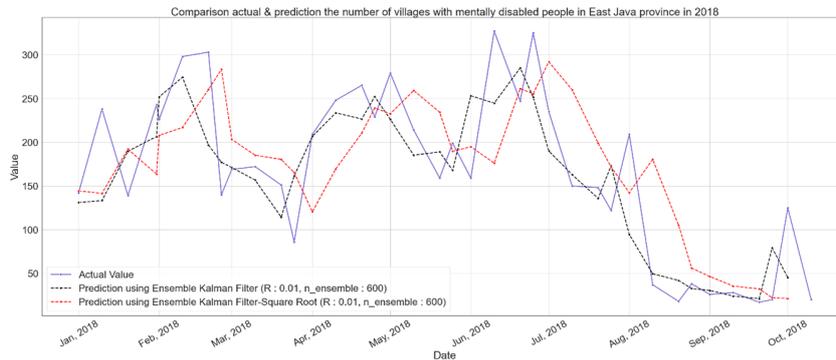### 3.1   Simulation result of EnKF and EnKF-SR with 600 ensembles

Figure 7: 1st Simulation Plot with EnKF and EnKF-SR with 600 ensembles and R value 0,01.

The first simulation can be seen in Figure 7, it shows the result of comparison between Ensemble Kalman Filter (EnKF) and Ensemble Kalman Filter Square Root (EnKF-SR). EnKF prediction results are shown by the black line, while those of EnKF-SR are shown

by the red line. All methods use 600 ensembles and R value is 0,01. The result of EnKF with 600 ensembles and R value of 0,01 has a better accuracy value with an RMSE value of 50,68. Meanwhile, EnKF-SR gives an RMSE value of 69.19.
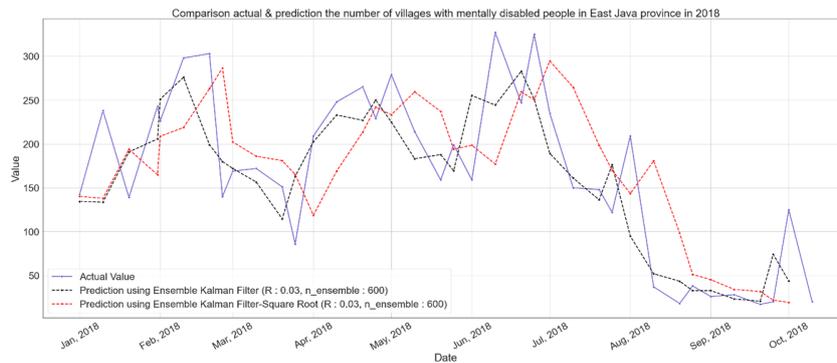

Figure 8: 2nd Simulation Plot with EnKF and EnKF-SR with 600 ensembles and R value 0,03.

The second simulation can be seen in Figure 8, it shows the result of comparison between Ensemble Kalman Filter (EnKF) and Ensemble Kalman Filter Square Root (EnKF-SR). EnKF prediction results are shown by the black line, while those of EnKF-SR are shown by the red line. All methods use 600 ensembles and R value is 0,03. The result of EnKF with 600 ensembles and R value of 0,03 has a better accuracy value with an RMSE value of 50,80. Meanwhile, EnKF-SR gives an RMSE value of 69.64.

3.2   Simulation result of EnKF and EnKF-SR with 800 ensembles
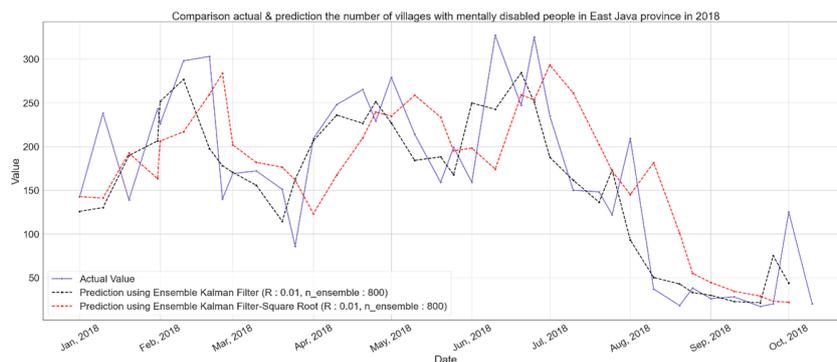

Figure 9: 3rd Simulation Plot with EnKF and EnKF-SR with 800 ensembles and R value 0,01.

The third simulation can be seen in Figure 9, it shows the result of comparison between Ensemble Kalman Filter (EnKF) and Ensemble Kalman Filter Square Root (EnKF-SR). EnKF prediction results are shown by the black line, while those of EnKF-SR are shown by the red line. All methods use 800 ensembles and R value is 0,01. The result of EnKF with 800 ensembles and R value of 0,01 has a better accuracy value with an RMSE value of 50,89. Meanwhile, EnKF-SR gives an RMSE value of 69.16.
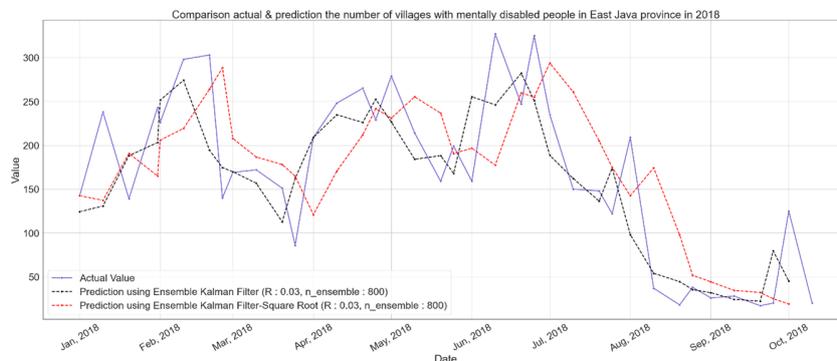
Figure 10: 4th Simulation Plot with EnKF and EnKF-SR with 800 ensembles and R value 0,03.

The fourth simulation can be seen in Figure 10, it shows the result of comparison between Ensemble Kalman Filter (EnKF) and Ensemble Kalman Filter Square Root (EnKF-SR). EnKF prediction results are shown by the black line, while those of EnKF-SR are shown by the red line. All methods use 800 ensembles and R value is 0,03. The result of EnKF with 800 ensembles and R value of 0,03 has a better accuracy value with an RMSE value of 50,99. Meanwhile, EnKF-SR gives an RMSE value of 69.16. For a clearer interpretation, below is a recapitulation table of all simulations shown in Tables 4 and 5

3.3   Comparison of RMSE Values

Method	Number of Ensembles	R Value	RMSE Value
EnKF	600	0.01	50.68
		0.03	50.80
EnKF-SR	600	0.01	69.19
		0.03	69.64

Table 4: Comparison of RMSE values of EnKF and EnKF-SR using 600 ensembles.

Method	Number of Ensembles	R Value	RMSE Value
EnKF	800	0.01	50.89
		0.03	50.99
EnKF-SR	800	0.01	69.16
		0.03	69.16

Table 5: Comparison of RMSE values of EnKF and EnKF-SR using 800 ensembles.

From Tables 4 and 5, it can be seen that both methods perform well. The EnKF method achieves the best RMSE value when the number of ensembles is 600 with an R value of 0.01. Meanwhile, the EnKF-SR method achieved the best RMSE value when the number of ensembles was 800 with R values of 0.01 and 0.03. Overall, the accuracy

rate produced by the EnKF method is about 80 percent. Meanwhile, the accuracy rate produced by the EnKF-SR method is around 77 percent.

4   Conclussion

From the overall simulation results that have been obtained using the EnKF and EnKF-SR methods, the prediction performance is quite good. Both methods maximize the accuracy value based on the number of ensembles and the R value. The EnKF method achieves the best RMSE value of 50,68 when the number of ensembles is 600 with an R value of 0.01. Meanwhile, the EnKF-SR method achieved the best RMSE value of 69,16 when the number of ensembles was 800 with R values of 0.01 and 0.03. Both methods can be used and optimized with other variations.

References

[1] M. H. Imaduddin, N. Febriyana, Y. Setiawati and I. Irwanto. Risk Factor Mild Mental Retardation in Extraordinary School at Surabaya. *Surabaya Journal of Psychiatry* **9** (2) (2020) 34—45.

[2] M. A. Wahyu and T. M. A. Elven. Protecting the Rights of Mental Health Patients: Comparative Study between Indonesia and Taiwan. *Indonesian Comparative Law Review* **2** (2) (2020) 110—125.

[3] N. V. E. S. Hidayanti, G. Alfianto, N. M. A. Kurniyanti and N. Soebagijono. Case Study: People with Severe Mental Disorders with Physical Mobility Barriers After in Pasung. *Journal of Health Research and Technology* **1** (1) (2023) 36—46.

[4] M. T. Hidayat, C. Oster, E. Muir-Cochrane and S. Lawn. Indonesia Free from Pasung: A Policy Analysis. *International Journal of Mental Health Systems* **17** (1) (2023) 1—22.

[5] Y. Widyawati, R. H. J. Scholte, T. Kleemans and R. Otten. Positive parenting and its mediating role in the relationship between parental resilience and quality of life in children with developmental disabilities in Java Island, Indonesia. *Research in Developmental Disabilities*. **112** (2021) 103911.

[6] T. I. Pramesti and S. Wibawani. Implementation of the Protection Policy for Children with Disabilities in the Rumah Anak Prestasi Surabaya City. *Journal La Sociale* **5** (4) (2024) 1139—1151.

[7] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist and H. Oza. Application of machine learning and artificial intelligence in oil and gas industry. *Petroleum Research* **6** (4) (2021) 379—391.

[8] P. Kumar and M. Sharma. Data, Machine Learning, and Human Domain Experts: None Is Better than Their Collaboration. *International Journal of Human-Computer Interaction* **38** (14) (2021) 1307—1320.

[9] N. Peiffer-Smadja et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clinical Microbiology and Infection* **26** (5) (2019) 584—595.

[10] A. Abdulle, G. Garegnani and A. Zanoni. Ensemble Kalman filter for multiscale inverse problems. *Multiscale Modeling and Simulation* **18** (4) (2020) 1565—1594.

[11] M. L. Provost and J. D. Eldredge. Ensemble Kalman filter for vortex models of disturbed aerodynamic flows. *Physical Review Fluids* **6** (5) (2021).

[12] S. Yoo and J. Song. Rapid prediction of wildfire spread using ensemble Kalman filter and polyline simplification. *Environmental Modelling & Software* **160** (2022) 105610.

[13] M. Y. Anshori, T. Herlambang, P. Katias, F. A. Susanto and R. R. Rasyid. Profitability estimation of XYZ company using H-infinity and Ensemble Kalman Filter. *Journal of Physics Conference Series* **2157** (2022) 012027.

[14] A. Muhith, T. Herlambang, D. Rahmalia, N. Irhamah and D. F. Karya. Estimation of Thrombocyte Concentrate (TC) in PMI Gresik using unscented and square root Ensemble Kalman Filter. *Journal of Physics Conference Series* **2157** (1) (2022) 012029.

[15] R. Ghostine, M. Gharamti, S. Hassrouny and I. Hoteit. An Extended SEIR Model with Vaccination for Forecasting the COVID-19 Pandemic in Saudi Arabia Using an Ensemble Kalman Filter. *Mathematics* **9** (6) (2021) 636.

[16] H. Cevallos, G. Intriago, and D. Plaza. Ensemble kalman filter and particle filter-based state estimation on electrical power systems. *Mathematics* **2090** (1) (2021) 012016.

[17] E. K. Korir, J. Aduda and T. Mageto. Forecasting Electricity Prices Using Ensemble Kalman Filter Forecasting Using Ensemble kalman Filter. *Journal of Statistical and Econometric Methods* **9** (1) (2020) 27–45.

[18] U. Javed et al. Exploratory Data Analysis Based Short-Term Electrical Load Forecasting: A Comprehensive analysis. *Energies* **14** (17) (2021) 1-22.

[19] N. S. Mittal et al. An exploratory data analysis of COVID-19 in India. *International Journal of Engineering Research* **9** (4) (2020) 580.

[20] S. L. Cessie, J. J. Goeman and O. M. Dekkers. Who is afraid of non-normal data? Choosing between parametric and non-parametric tests. *European Journal of Endocrinology* **182** (2) (2020) E1—E3.

[21] W. M. Ridwan et al. Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. *Ain Shams Engineering Journal* **12** (2) (2021) 1651—1663.

[22] D. Chanal, N. Y. Steiner, D. Chamagne and M. -C. Pera. Impact of standardization applied to the diagnosis of LT-PEMFC by Fuzzy C-Means clustering. *2021 IEEE Vehicle Power and Propulsion Conference (VPPC)* (2021).

[23] B. Nilam and S. T. Ram. Forecasting Geomagnetic activity (Dst Index) using the ensemble kalman filter. *Monthly Notices of the Royal Astronomical Society* **511** (1) (2022) 723—731.

[24] M. Fan, Y. Bai, L. Wang and L. Ding. Combining a fully connected neural network with an ensemble Kalman filter to emulate a dynamic model in data assimilation. *IEEE Access* **9** (2021) 144952–144964.