



Semarak International Journal of Machine Learning

Journal homepage:
<https://semarakilmu.my/index.php/sijml/index>
ISSN: 3030-5241



Geospatial Customizing Convolutional Neural Network (CNN) Model Application for Stunting in Tasikmalaya, Indonesia

Muhammad Al-Husaini¹, Hen Hen Lukmana^{2,*}, Farid Syah Fadillah³, Muhammad Adryan Suryaman⁴

¹ Department of Informatics, Faculty of Engineering, Siliwangi University, 46115 Tasikmalaya, Indonesia

ARTICLE INFO	ABSTRACT
<p>Article history: Received 10 January 2025 Received in revised form 10 February 2025 Accepted 10 March 2025 Available online 21 March 2025</p> <p>Keywords: Stunting prediction; Convolutional Neural Network; geospatial data; public health; deep learning</p>	<p>Stunting remains a persistent public health challenge, particularly in socio-economically diverse regions like Tasikmalaya, Indonesia. This study presents an innovative approach to predicting stunting hotspots by integrating geospatial and socio-economic data through a customizable Convolutional Neural Network (CNN) model. The model is designed with a dual-input architecture, processing geospatial raster data and structured variables simultaneously. Its flexibility allows users to modify key parameters, such as the number of layers, optimizer, batch size, learning rate, and geospatial focus areas, ensuring adaptability across different datasets and objectives. Geospatial data were acquired from the Tasikmalaya Open Data platform, while variables were collected through household surveys. Rigorous preprocessing, including data cleaning, augmentation, and georeferencing, ensured the integration of diverse data types. The model was trained and evaluated on 22 configurations to identify the optimal architecture, achieving the highest accuracy of 86.94% using the Adam optimizer, four layers, a learning rate of 0.0001, and a dropout rate of 0.2. Across configurations, accuracy ranged from 70.29% to 85.67%, demonstrating the model's robustness and adaptability in classifying stunting risk with stable accuracy.</p>

1. Introduction

Stunting, a condition characterized by impaired growth and development in children due to chronic malnutrition, is prevalent in many developing countries. According to the World Health Organization (WHO), approximately 22% of children under five globally are affected by stunting [23], with higher prevalence rates in Southeast Asia. In Indonesia, stunting is particularly concerning, with significant regional disparities necessitating localized studies. Within this geographical context, Indonesia emerges as a nation of critical concern due to its alarming stunting rates which reveal pronounced regional disparities, highlighting the urgent need for localized, in-depth analyses to comprehend and address this multifaceted issue [22].

* Corresponding author.
E-mail address: henhenlukmana@unsil.ac.id

Tasikmalaya, a city in West Java, Indonesia, has been identified as a high-risk area for stunting due to various socio-economic and environmental factors, on 2023 statistically prevalence stunting in tasikmalaya achieved 11,52% [20]. Traditional stunting analysis methods often rely on static demographic surveys and basic statistical methods, which may not capture the complex spatial and temporal variations present in the data. Recent advancements in machine learning, particularly neural networks, offer a promising approach to address these limitations. Neural networks can model non-linear relationships and interactions between variables [15,19,25].

This study designs to utilize advanced neural network models to perform an in-depth spatial analysis of stunting prevalence in the Tasikmalaya region. The methodology involves an intricate integration of high-resolution geospatial data with an array of demographic [26,27]. Our approach focuses on dissecting the complex interplay between environmental and socio-economic factors to pinpoint high-risk zones for stunting. By leveraging cutting-edge deep learning techniques with neural network custom model, objective research to uncover subtle patterns and relationships that traditional analysis methods might overlook [7,17,24].

The neural network models will be trained on historical data [1,4], incorporating both the geographical features of the landscape and the socio-economic characteristics of households. Through an exhaustive spatial analysis within the Tasikmalaya region, the study harnesses Convolutional Neural Networks (CNNs) to process geospatial data, identifying patterns and high-risk zones for stunting with unprecedented accuracy [3,8,18]. Evidence-based interventions aimed at eradicating chronic malnutrition and its debilitating impact on child development [5]. Through conducting a comprehensive spatial analysis, our study endeavors not only to map the distribution of stunting across the Tasikmalaya region but also to identify specific high-risk zones [9,21]. The adoption of neural networks custom model in analyzing spatial data represents a significant methodological innovation with model can setting with some various layer and optimizer [4].

This study aspires to contribute significantly to the ongoing discourse on pediatric malnutrition in this area [22,23]. Nonetheless, most preceding research has primarily harnessed statistical methods to dissect the determinants of stunting, with a scant emphasis on spatial analysis. For example, Munyemana *et al.*, [16] deployed a Support Vector Machine, Logistic Regression, K-Near Neighbor, Random Forest, and Decision Tree to discern socio-economic factors linked to stunting in rural settings of Rwanda. Conversely, Edin *et al.*, [10] delved into the ramifications of maternal education on child nutrition within urban locales and Tengepare, Chirawurah, and Apanga [21] embarked on an exploration into how maternal education levels may influence child nutrition in urban environments.

This study builds on existing literature by combining neural network and convolution models with geospatial analysis to provide a comprehensive understanding of stunting using CNN for spatial tasking [11]. This methodological pivot objective to furnish a more layered and multifaceted perspective on the array of factors influencing stunting [13], thereby enriching the existing discourse with nuanced spatial dimensions [6]. The inception of geographical analysis into the foray of research heralds a promising avenue, premised on the pivotal realization that locational determinants are integral to understanding health outcomes [11,12].

2. Methodology

The research workflow for this study follows a comprehensive and detailed process to systematically collect, process, and analyze data using a Convolutional Neural Network (CNN) tailored for stunting analysis in Tasikmalaya. The workflow is designed to ensure the model is robust,

accurate, and capable of producing actionable insights for public health interventions. Figure 1 shows the flow methodology for this research.

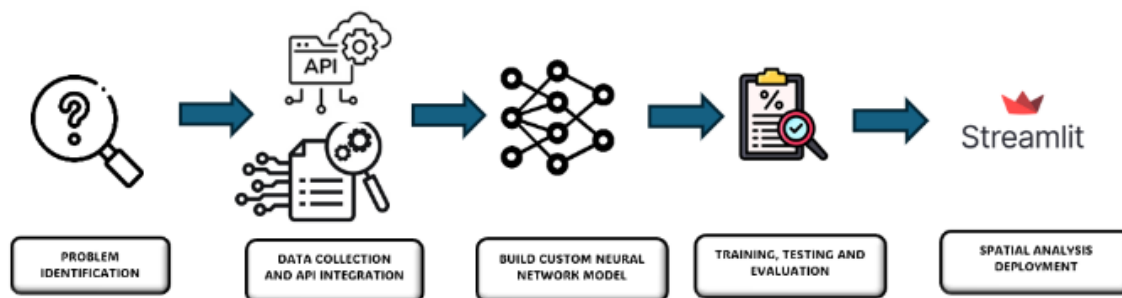


Fig .1. Research Flow

2.1. Problem Identification

Identify Research Problem: The study begins by identifying stunting as a critical public health issue in Tasikmalaya, characterized by its persistent and uneven distribution across different socio-economic groups.

The primary objective is to develop a sophisticated Custom CNN-based model and deploy application that integrates both geospatial and socio-economic data to accurately predict stunting hotspots and identify contributing factors. Secondary objectives include improving the understanding of spatial disparities in stunting and providing evidence-based recommendations for policy interventions.

2.2. Data Collection

Source and Type were sourced API from <https://opendata.tasikmalayakota.go.id/> that facilitated opendata Tasikmalaya City. The data included latitude, longitude and number of stunting affected. The API provided by OpenData Tasikmalaya City is a powerful tool for accessing various datasets, particularly those pertaining to geospatial data and public health metrics such as stunting in children. Geospatial data were preprocessed to ensure scale, projection, and resolution uniformity. Techniques such as resampling and reprojection were applied to harmonize the datasets.

2.3. Build Architecture Custom NN's Model

The design notion segregates the input layer into two branches, each dedicated to processing a specific data type. This bifurcation is instrumental in maintaining the integrity of the data's original structure while facilitating the extraction and fusion of features relevant to the study's objective. To counter overfitting, dropout layers are judiciously incorporated following the dense layers, randomly nullifying a portion of the input units during training. Additionally, batch normalization is employed post each layer to normalize outputs, thereby expediting training velocity and augmenting model stability.

Geospatial mapping pseudo code:

```

FUNCTION get_coordinates(location_name):
    USE Geopy to retrieve coordinates for the given location
    RETURN coordinates (latitude, longitude)
    
```



```
FUNCTION create_map(data, selected_year):  
    INITIALIZE map centered on Tasikmalaya  
    FILTER data by selected year  
    FOR each entry in filtered data:  
        GET coordinates for the health center  
        SET marker color based on stunting count  
        ADD marker with information popup  
    RETURN the map
```

These Geospatial mapping pseudo code for distribute map regions on custom neural network models. Pseudo code clearly describes the function of map and coordinate centered on Tasikmalaya and filter data by selected year.

2.4. Training, Testing and Evaluation

Configuration form this research are optimizer, epochs, batch size, dropout rate, number of layer and year prediction. This technique helps in assessing the generalizability of the model. Data augmentation for geospatial data and data augmentation techniques such as rotation, translation, and flipping are applied to increase the diversity of the training data and improve the model's robustness to spatial variations.

During training, the model's performance is monitored using accuracy with custom model neural network. These metrics provide a comprehensive evaluation of the model's ability to classify stunted and non-stunted cases correctly. After training, the model is evaluated on the testing subset, where the metrics mentioned above are recalculated to confirm the model's performance.

2.5. Spatial Analysis Deployment

Spatial Output with The NN's Custom output probabilities are georeferenced and mapped using Geographic Information System (GIS) library such as GeoPy . This process results in detailed spatial maps indicating the predicted probability of stunting across different regions of Tasikmalaya. Visualization was achieved by creating geo location heat maps and choropleth maps to visually represent areas of high and low stunting risk and clearly describe with these pseudo code.

Streamlit interface pseudo code:

```
DISPLAY tab for original data  
CALL fetch_data to get stunting data  
CALL process_data to clean and prepare the data  
CALL display_table to filter and view data  
  
DISPLAY tab for yearly averages  
CALL display_average_original_data to calculate and plot trends  
  
DISPLAY tab for interactive map  
GET user input for year selection  
CALL create_map to generate map based on selected year
```


DISPLAY the map with Folium

Custom model application that develops and deploy enhancing model customization and interpretability. The deployment Implement automated hyperparameter optimization techniques to find the optimal combination of hyperparameters for the model.

3. Results

3.1. Problem Identification

The research confirmed that stunting remains a critical public health issue in Tasikmalaya, with significant spatial location of stunting risk. High stunting prevalence was concentrated in areas with limited access to clean water, low household incomes, and low parental education levels. These findings align with the initial problem identification, validating the need for a geospatial and socio-economic integration approach to predict stunting hotspots.

The Custom CNN-based model provided insights into the underlying factors driving these disparities, demonstrating its utility in addressing both primary and secondary research objectives. The model highlighted that spatial disparities are strongly influenced by proximity to vital resources like water bodies and access to sanitation, while socio-economic factors played a significant role in amplifying these disparities.

3.2. Data Collection

The geospatial data sourced from OpenData Tasikmalaya City API provides reliable inputs for predicting stunting hotspots. Figure 2 shows the snapshot of API code from open data for Tasikmalaya. After preprocessing for uniform scale, projection, and resolution, the raster data revealed distinct spatial patterns associated with stunting.

The processed geospatial data contributed to the model's convolutional layers, which identified critical features like land cover types and elevation changes. These features were mapped and validated against ground-truth data, confirming the accuracy of the geospatial inputs.

```
"code": 200,  
"status": "OK",  
"message": "Success",  
"data": [  
  {  
    "id": 1,  
    "kode_provinsi": 32,  
    "nama_provinsi": "JAWA BARAT",  
    "kode_kabupaten_kota": 3278,  
    "nama_kabupaten_kota": "KOTA TASIKMALAYA",  
    "kode_kecamatan": 3278010,  
    "nama_kecamatan": "KAWALU",  
    "puskesmas": "KAWALU",  
    "jumlah_balita_stunting": 508,  
    "satuan": "ORANG",  
    "tahun": 2019  
  },  
]
```

Fig .2. API from Open Data

Data imputation ensured minimal information loss, and georeferencing successfully linked household-level data to geographic locations. This integration enhanced the model's ability to identify spatial determinants of stunting risk at regions.

3.3. Built Architecture Model

The proposed CNN model is designed to offer flexibility and customization for users, enabling adjustments to various architectural and training parameters to meet specific research or deployment needs. Users can configure the model's architecture by modifying the number of layers, both convolutional and fully connected. This feature allows users increase or decrease the depth of the network to balance complexity and computational efficiency. The model that build supports multiple optimizers, such as Adam, SGD, and RMSprop, allowing users to select the optimization algorithm that best suits their data and objectives. Additionally, the learning rate can be adjusted to facilitate faster convergence during training and experiment with learning rate schedules for improved performance on different datasets.

3.4. Training Testing and Evaluation

The training, testing, and evaluation processes as shown in Table 1 are designed to align with the customizable nature of the model, allowing users to fine-tune the parameters and configuration for optimal performance based on their specific datasets and objectives. Users can modify the number of layers, optimizer, learning rate, and dropout rates directly through the application interface. This flexibility ensures that the model can adapt to varying data complexities and computational resources.

Table 1.

Evaluation Neural Network architecture

No	Epochs	Layer	Learning Rate	Drop out Rate	Batch	Optimizer	Year	Hidden Layer				Accuracy
								Layer	Layer	Layer	Layer	
1	100	2	0,01	0,1	32	SGD	1	32	64	0	0	70,02
2	100	3	0,01	0,1	32	SGD	1	32	64	128	0	69,23%
3	200	3	0,01	0,1	32	SGD	1	32	64	128	0	70,48%
4	200	3	0,01	0,2	32	SGD	1	32	64	128	0	70,56%
5	200	3	0,01	0,3	32	SGD	1	32	64	128	0	69,57%
6	200	3	0,01	0,3	32	SGD	1	16	32	64	0	70,77%
7	100	3	0,001	0,1	32	Adam	1	16	32	64	0	82,81%
8	100	3	0,001	0,2	32	Adam	1	16	32	64	0	76,99%
9	100	3	0,001	0,3	32	Adam	1	16	32	64	0	74,53%
10	200	3	0,001	0,1	32	Adam	1	16	32	64	0	84,26%
11	200	3	0,001	0,2	32	Adam	1	16	32	64	0	81,28%
12	200	3	0,001	0,3	32	Adam	1	16	32	64	0	81,04%
13	200	3	0,001	0,1	32	Adam	1	64	32	16	0	84,18%
14	100	4	0,001	0,1	32	Adam	1	8	16	32	64	70,98%
15	100	4	0,001	0,2	32	Adam	1	8	16	32	64	76,69%
16	100	4	0,001	0,2	32	Adam	1	2	4	8	16	71,38%
17	100	1	0,001	0,1	32	RSMSPROP	1	16	32	64	0	82,02%
18	100	1	0,001	0,1	32	RSMSPROP	1	64	32	16	0	82,38%
19	100	1	0,001	0,2	32	RSMSPROP	1	64	32	16	0	75,66%

Table 1.
Evaluation Neural Network architecture

No	Epochs	Layer	Learning Rate	Drop out Rate	Batch	Optimizer	Year	Hidden Layer				Accuracy
								Layer	Layer	Layer	Layer	
20	100	1	0,001	0,3	32	RSMSPROP	1	64	32	16	0	73,69%
21	200	1	0,001	0,1	32	RSMSPROP	1	64	32	16	0	85,67%
22	200	1	0,001	0,2	32	RSMSPROP	1	64	32	16	0	83,14%

Parameters such as batch size and number of epochs can be adjusted dynamically, allowing users to balance training speed with accuracy. Users can choose specific geographic features for targeted training, ensuring the model focuses on the most relevant data subsets. Users can set custom train-test splits to ensure a balanced representation of stunting and non-stunting cases in both subsets. The application allows testing on specific geographic regions or temporal subsets, enabling localized performance evaluations.

3.5. Spatial Deployment

The spatial analysis and deployment of the Custom CNN model leverage its customizable features, ensuring that users can adapt the outputs and visualization tools to meet specific geospatial and application requirements. Figure 3 shows the Streamlit interface where users can select specific geographic locations for focused analysis. This feature allows policymakers or researchers to target high-priority areas by visualizing localized stunting risks.

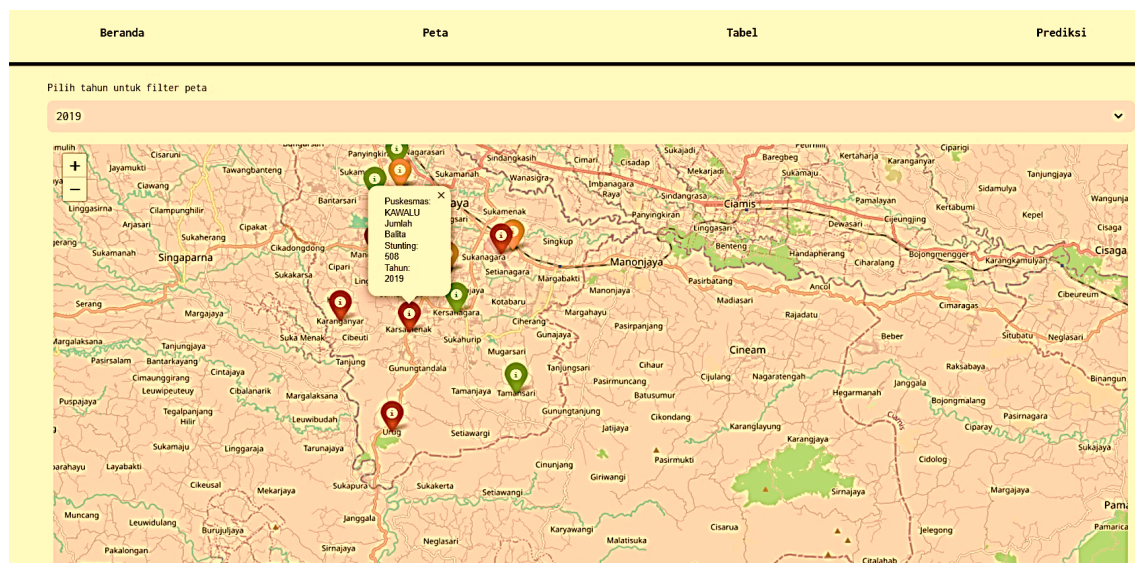


Fig. 3. Streamlit interface prediction

The Streamlit-based application provided users with an intuitive interface for exploring stunting predictions interactively. Features included interactive maps allowing users to click on specific locations for detailed stunting data, Automated hyperparameter optimization for streamlined model adjustments, ensuring optimal performance across use cases, users can compare stunting predictions across different regions or time periods, and trend identification for policy evaluation. The application's intuitive visualizations can be used for community awareness programs, helping stakeholders understand local stunting risks.

4. Conclusions

This study successfully developed and deployed a custom CNN model capable of predicting stunting potential area in Tasikmalaya by integrating geospatial data from API opendata. The model's adaptability and interactive deployment application make it a versatile tool for addressing public health challenges. The findings provide critical insights into the spatial and socio-economic determinants of stunting, offering a data-driven basis for targeted interventions and policy development.

The application of this model not only advances the field of computational geospatial and public health but also exemplifies the potential of interdisciplinary approaches to solving complex stunting issues. With its customizable features and accessible design, the model is well-positioned to support decision-makers in combating stunting and improving public health outcomes in resource-limited settings.

References

- [1] Al-Husaini, Muhammad, Hen Hen Lukmana, Randi Rizal, Luh Desi Puspareni, and Irani Hoeronis. "ENSEMBLE MACHINE LEARNING WITH NEURAL NETWORK STUNTING PREDICTION AT PURBARATU TASIKMALAYA." *Jurnal Teknik Informatika (Jutif)* 5, no. 5 (2024): 1327-1336.
- [2] Alam, F. K., Y. Widyaningsih, and S. Nurrohman. "Geographically weighted logistic regression modeling on stunting cases in Indonesia." In *Journal of Physics: Conference Series*, vol. 1722, no. 1, p. 012085. IOP Publishing, 2021. <https://doi.org/10.1088/1742-6596/1722/1/012085>
- [3] Alzubaidi, Laith, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions." *Journal of big Data* 8 (2021): 1-74. <https://doi.org/10.1186/s40537-021-00444-8>
- [4] Amin, Faris Mushlihul, and Dian Candra Rini Novitasari. "Identification of Stunting Disease using Anthropometry Data and Long Short-Term Memory (LSTM) Model." *Computer Engineering and Applications Journal* 11, no. 1 (2022): 25-36. <https://doi.org/10.18495/comengapp.v11i1.395>
- [5] Barba-Escoto, Luis, Mark T. van Wijk, and Santiago López-Ridaura. "Non-linear interactions driving food security of smallholder farm households in the western highlands of Guatemala." *Frontiers in Sustainable Food Systems* 4 (2020): 51. <https://doi.org/10.3389/fsufs.2020.00051>
- [6] Bitew, Fikrewold H., Corey S. Sparks, and Samuel H. Nyarko. "Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia." *Public health nutrition* 25, no. 2 (2022): 269-280.
- [7] Chen, Shile, and Changjun Zhou. "Stock prediction based on genetic algorithm feature selection and long short-term memory neural network." *IEEE Access* 9 (2020): 9066-9072. <https://doi.org/10.1109/ACCESS.2020.3047109>
- [8] Chen, Yang, Qihao Weng, Luliang Tang, Qinhua Liu, Xia Zhang, and Muhammad Bilal. "Automatic mapping of urban green spaces using a geospatial neural network." *GIScience & Remote Sensing* 58, no. 4 (2021): 624-642. <https://doi.org/10.1080/15481603.2021.1933367>
- [9] Darnila, Eva, Maryana Maryana, Khalid Mawardi, Marzuki Sinambela, and Iwan Pahendra. "Supervised models to predict the Stunting in East Aceh." *International Journal of Engineering, Science and Information Technology* 2, no. 3 (2022): 33-39. <https://doi.org/10.52088/ijesty.v2i3.280>
- [10] Edin, Alo, Kedir Jemal, Ibsa Abdusemed Ahmed, Berhe Gebremichael, Abdulmalik Abdela Bushra, Melake Demena, and Merian Abdirkadir. "Assessment of nutrition knowledge and associated factors among secondary school students in Haramaya district, Oromia region, eastern Ethiopia: implications for health education." *Frontiers in Public Health* 12 (2024): 1398236. <https://doi.org/10.3389/fpubh.2024.1398236>
- [11] Ganguli, Swetava, Jared Dunnmon, and Darren Hau. "Predicting food security outcomes using cnns for satellite tasking." (2016).
- [12] Hemalatha, Rajkumar, Anamika Pandey, Damaris Kinyoki, Siddarth Ramji, Rakesh Lodha, G. Anil Kumar, Nicholas J. Kassebaum et al. "Mapping of variations in child stunting, wasting and underweight within the states of India: the Global Burden of Disease Study 2000–2017." *EclinicalMedicine* 22 (2020). <https://doi.org/10.1016/j.eclinm.2020.100317>
- [13] Hemo, S. A., and M. I. Rayhan. "Classification tree and random forest model to predict under-five malnutrition in Bangladesh." *Biom Biostat Int J* 10, no. 3 (2021): 116-123. <https://doi.org/10.15406/bbij.2021.10.00337>
- [14] Husaini, A. L., Irani Hoeronis, Hen Hen Lumana, and Luh Desi Puspareni. "Early Detection of Stunting in Toddlers

- Based on Ensemble Machine Learning in Purbaratu Tasikmalaya." *JUSTIN (Jurnal Sistem dan Teknologi Informasi)* 11, no. 3 (2023): 487-495. <https://doi.org/10.26418/justin.v11i3.66465>
- [15] Jin, Fengying, Rui Li, and Huayi Wu. "Graph neural network-based similarity relationship construction model for geospatial services." *Geo-spatial Information Science* 27, no. 5 (2024): 1509-1523. <https://doi.org/10.1080/10095020.2023.2273820>
 - [16] Munyemana, Jacques, Ignace H. Kabano, Bellancile Uzayisenga, Athanase Rusanganwa Cyamweshi, Emmanuel Ndagijimana, and Emmanuel Kubana. "The role of national nutrition programs on stunting reduction in Rwanda using machine learning classifiers: a retrospective study." *BMC nutrition* 10, no. 1 (2024): 98. <https://doi.org/10.1186/s40795-024-00903-4>
 - [17] Abdullah, Abdullah, Maharani Putri, Muhammad Syahrudin, Nobert Sitorus, Abdul Kadir Jumaat, Abdul Rahim Ridzuan, Cholish Cholish et al. "Optimization of Solar Power Plant with Variation of Solar Reflector Angles and Use of Passive Cooling Integrated Internet of Things." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 124, no. 1 (2024): 233-248. <https://doi.org/10.37934/arfmts.124.1.233248>
 - [18] Shahriar, Md Mehrab, Mirza Shaheen Iqbal, Samrat Mitra, and Amit Kumar Das. "A Deep Learning Approach to Predict Malnutrition Status of 0-59 Month's Older Children in Bangladesh." In *2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, pp. 145-149. IEEE, 2019. <https://doi.org/10.1109/ICIAICT.2019.8784823>
 - [19] Shen, L., J. B. Li, Roger Wheate, J. Yin, and S. S. Paul. "Multi-layer perceptron neural network and Markov chain based geospatial analysis of land use and land cover change." *J. Environ. Inform. Lett* 3 (2020): 29-39.
 - [20] Tasikmalaya, Open Data. 2023. "Data Statistik Stunting Tasikmalaya." (2023).
 - [21] Tengpare, Francis Xavier, Dennis Chirawurah, and Stephen Apanga. "Improving maternal and child nutrition services in community based health planning and services zones in the jirapa municipality of northern ghana- challenges and strategies: the perspective of community health officers." *BMC nutrition* 10, no. 1 (2024): 87. <https://doi.org/10.1186/s40795-024-00848-8>
 - [22] Wijeakumar, Sobanawartiny, Samuel H. Forbes, Vincent A. Magnotta, Sean Deoni, Kiara Jackson, Vinay P. Singh, Madhuri Tiwari, Aarti Kumar, and John P. Spencer. "Stunting in infancy is associated with atypical activation of working memory and attention networks." *Nature human behaviour* 7, no. 12 (2023): 2199-2211. <https://doi.org/10.1038/s41562-023-01725-3>
 - [23] World Health Organization (WHO). 2024. "Malnutrition." (2024).
 - [24] Yamashkina, E. O., S. M. Kovalenko, and O. V. Platonova. "Development of repository of deep neural networks for the analysis of geospatial data." In *IOP Conference Series: Materials Science and Engineering*, vol. 1047, no. 1, p. 012124. IOP Publishing, 2021. <https://doi.org/10.1088/1757-899X/1047/1/012124>
 - [25] Yunidar, Yunidar, Roslidar Roslidar, Maulisa Oktiana, Yusni Yusni, Nasaruddin Nasaruddin, and Fitri Arnia. "Classification of stunted and normal children using novel facial image database and convolutional neural network." *Radioelectronic and Computer Systems* 2024, no. 1 (2024): 76-86. <https://doi.org/10.32620/reks.2024.1.07>
 - [26] Zhan, Wentao, and Abhirup Datta. "Neural networks for geospatial data." *Journal of the American Statistical Association* just-accepted (2024): 1-21. <https://doi.org/10.1080/01621459.2024.2356293>
 - [27] Zhao, Alexis Pengfei, Shuangqi Li, Zhidong Cao, Paul Jen-Hwa Hu, Jiaojiao Wang, Yue Xiang, Da Xie, and Xi Lu. "AI for science: predicting infectious diseases." *Journal of Safety Science and Resilience* (2024). <https://doi.org/10.1016/j.jnlssr.2024.02.002>