1st Proceeding of International Conference on Science and Technology UISU (ICST)
"Contribution of Engineering Disciplines to Achieving Sustainable Development Goals Research Realizing the Vision of Golden Indonesia 2045"
ISSN: 3063-7929 (Online)
DOI: https://doi.org/10.30743/icst

Comparison of Accuracy between Random Forest Method and K-Nearest Neighbors Method for Recognizing Solar Panel Energy Conversion Temperature

Habib Satria[1*], Rika Rosnelly[2], Wanayumini[3]
[1]Master of Computer Science, Faculty of Engineering and Computer Science, Universitas Potensi Utama, Medan. Indonesia
[2]Department of Computer Science, Faculty of Engineering and Computer Science, Universitas Potensi Utama, Medan. Indonesia
[3]Department of Informatic Engineering, Faculty of Engineering, Universitas Asahan, Kisaran. Indonesia
[*]Corresponding author: habibsatria222@gmail.com

Abstract - This study analyzes the classification of temperature image pattern recognition on solar panels to improve the accuracy of energy conversion performance affected by weather changes. The process begins by capturing the surface temperature image of the panel as primary data, which is then processed through the pattern recognition stage. The pattern recognition method is chosen to detect and understand patterns in temperature images that will be used as datasets. This study also compares the results of pattern recognition using the Random Forest classification method with the K-Nearest Neighbors (KNN) method, in order to build an effective model in analyzing images based on weather temperature in Medan City. The results of the study obtained that the solar panel produced a maximum output of 15.74 Wp at 12:00 pm, when the temperature tends to be higher and sunlight is optimal. Then the results of the random forest method showed good performance with an accuracy of 84% and the K-Nearest Neighbors method had an accuracy of 78%..

Keywords : Pattern Recognition, Random Forest Method, KNN Method, Solar Panel Temperature, Energy Conversion.

1. INTRODUCTION

North Sumatra as a tropical climate region has high solar radiation potential. This condition provides a great opportunity for the development and installation of renewable energy systems, especially solar panels or photovoltaics (PV). However, despite its enormous potential, the energy conversion performance of solar panels is often disrupted by temperature fluctuations due to weather changes [1]–[3]. The use of solar panels is also getting more attention along with the increasing demand for energy worldwide. The availability of unlimited energy resources, ease of implementation, and environmental friendliness are the main advantages of this system [4]. However, the high initial cost and low conversion efficiency of solar panels, as well as intensive land use, are the main obstacles faced. Over the years, various improvements have been made by manufacturers, research centers, and researchers around the world to overcome the most significant shortcomings of the Solar Panel system and improve its performance [5]. These improvements are made both at the material level, such as increasing the conversion efficiency of Solar Panel panels while minimizing production costs, and at the overall system level, such as maximizing or optimizing the power taken from the solar panels. MPPT controllers, cooling systems, cleaning systems, solar tracking systems, and floating Solar Panel systems are some of the most popular techniques introduced to improve the performance of Solar Panel systems and maximize the use of available solar energy [6]. Then the optimization of solar panel power using solar tracking systems and floating Solar Panel systems becomes a hot topic in solar power generation and attracts the attention of researchers around the world, especially the developing floating Solar Panel system. This technology can be implemented and optimized in the context of efficient and sustainable solar power generation. However, the problem is that the ambient temperature is also one of the key factors that determine the energy output of solar panels. When the temperature increases or decreases drastically, or when the position of the sun is not optimal to the surface of the panel, the energy conversion efficiency can be significantly reduced. Therefore, the use of methods to implement pattern recognition becomes the main solution to help identify and understand patterns that occur in weather temperature fluctuations.

The use of the method used to help recognize the temperature of this solar panel, uses the Random Forest classification method with the K-Nearest Neighbors (KNN) method. By applying this method, it is expected that the use of energy conversion generated from solar panels can be managed more efficiently and effectively. The main focus of this study is to analyze the performance of solar panel energy conversion by considering weather fluctuations that occur in Medan City. In this analysis, an algorithm is used to calculate important parameters such as Accuracy, Precision, Recall, and F1-Score [7]–[10]. Data collection is carried out in real-time using thermovisi technology, which makes it possible to record temperature images accurately. The recorded data is then processed using a classification technique using Random Forest with the K-Nearest Neighbors (KNN) method. In this method, identification is carried out to detect and understand weather temperature image patterns that will be used as datasets for the recognition or classification stage. The Random Forest algorithm was chosen because of its ability to model non-linear relationships between features and targets, as well as its ability to handle overfitting, so that it can build a reliable model in understanding image data based on weather temperature in Medan City and the comparison of the KNN method was chosen [11]. K-Nearest Neighbors has several advantages in pattern recognition yaitu first, this

ISSN: 3063-7929 (Online)
DOI: https://doi.org/10.30743/icst

algorithm is simple and easy to understand, making it a good choice for beginners [12], [13]. In addition, KNN does not assume a particular data distribution, so it can work effectively on datasets that are not normally distributed [14], [15]. This algorithm is also adaptive and can be used for various types of data, both numeric and categorical, as well as for classification and regression [16], [17]. KNN often gives good results on small datasets without many features and its performance can be improved through the selection of the right features[18]–[20]. By using this algorithm, it is expected to gain a deeper understanding of image classification or images related to weather conditions in Medan. The resulting model will make a significant contribution to renewable energy management, especially in maximizing energy output from solar panels by considering the dynamics of the weather that occurs. Thus, this study is expected to provide practical solutions in optimizing the use of energy from renewable sources, as well as providing new insights into the effect of temperature fluctuations on solar panel performance. Through a data-based approach and appropriate algorithms, the potential of solar energy in North Sumatra can be explored optimally for the sake of sustainability and improving the quality of life of the community.

.

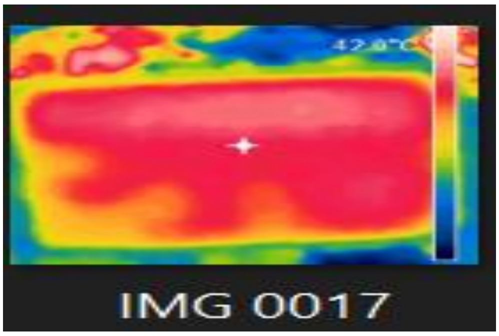
2. METHOD

The classification technique for recognizing image patterns and temperature values on solar panels uses the Random Forest Method and the K-Nearest Neighbors (KNN) method to determine the accuracy value. With optimal accuracy to obtain energy conversion performance results based on weather changes that occur, it will later be considered in installing solar panels. For performance analysis on this solar panel energy conversion to capture images or images of surface temperatures, thermovision technology is used which will later be used as primary data to be processed to the pattern recognition stage. The pattern recognition method is an identification process that is chosen to detect and understand image patterns or weather temperature images that are used as datasets to be processed to the recognition or classification stage. And this study also uses pattern recognition which is used as a comparison using the Random Forest Method and the K-Nearest Neighbors (KNN) method to build a model that is in accordance with the aim of understanding image data or images based on weather temperatures in Medan City. This modeling is expected to help in the recognition and analysis of solar panel surface temperature patterns against weather changes that occur in the area. The projection map of the position of the solar panel installation in Medan City, North Sumatra is shown in Figure 1.

Figure 1. Projected map of solar panel installation positions in Medan City, North Sumatra.

Then the data that will be used later for the training process (training set) as many as 210 images and the testing process (testing set) as many as 90 images. The data used in this study consists of the results of energy conversion from solar panels in the form of 2-dimensional images. The image data used in this study can be shown in Figure 2 where Figure 2 (a) is a hot temperature image and Figure 2 (b) a cold temperature image.

1st Proceeding of International Conference on Science and Technology UISU (ICST)
"Contribution of Engineering Disciplines to Achieving Sustainable Development Goals Research Realizing the Vision of Golden Indonesia 2045"
ISSN: 3063-7929 (Online)
DOI: https://doi.org/10.30743/icst



(a). Hot Temperature Image

(b). Cold Temperature Image

Figure 2. Surface temperature image of hot red and cool blue color panel

The use of thermovision technology in capturing hot and cold temperature images on solar panels will make it easier to detect the temperature output produced by the panel, as well as identify potential problems such as overheating or weaknesses in the cooling system, which can affect the productivity and life of the panel device. In Figure 2. is an image of the panel surface temperature, namely hot temperature in red and cold blue.

3. RESULTS AND DISCUSSION

The data taken is primary data where the data is very important in the process of creating a machine learning model by collecting it directly through a thermovision tool based on weather fluctuations in the city of Medan. The thermovision shooting distance to collect image data is set at one meter from the solar panel. This measurement is carried out to obtain accurate information about the temperature produced by the solar panel in its operational conditions, as well as to understand the effect of temperature on the efficiency of the panel in producing energy. Then the collection of DC power data functions to analyze energy conversion to determine the output value of the solar panel. Then, accurate observation of the panel surface temperature is very important to produce the efficiency and performance of the solar panel. The main stage in data collection is to check the watt meter to be calibrated according to standards. After that, choose a test location that gets direct sunlight and is free from shadows, such as buildings or trees around it, so that the measurement results are more accurate. Next, set the thermovision tool so that it points directly at the surface of the solar panel and position the watt meter correctly so that it is connected to the panel to measure the power output. Data collection at 1 hour intervals, starting from 10.00 am in the morning to 15.00 pm in the afternoon, which is the peak load period for solar panels. During each interval, the panel surface temperature is shown on the thermovision screen and the output power value is shown by the watt meter. After all the data is collected, a thorough analysis is performed to identify the panel surface temperature pattern and its relationship to the power output generated.

3.1 Results of Solar Panel Energy Conversion Testing in Medan City

The results of the solar panel energy conversion test in Medan City show varying efficiency depending on the intensity of sunlight and weather conditions. The significant potential for developing renewable energy in Medan City is very promising if it can be exploited properly. The solar panel performance measurement data shows a clear relationship between power output throughout the measurement hours from 10:00 am to 3:00 pm. The energy conversion performance of the solar panel power output based on weather fluctuations in Medan City is shown in Figure 3.

ISSN: 3063-7929 (Online)
DOI: https://doi.org/10.30743/icst

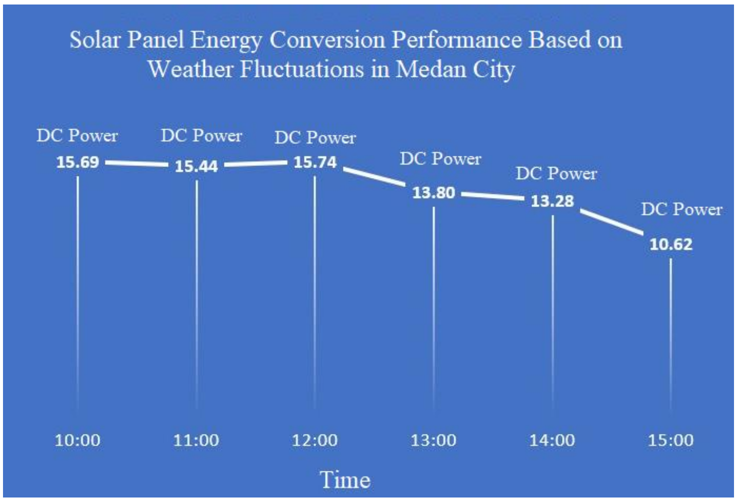


Figure 3. Solar panel power output energy conversion performance based on weather fluctuations in Medan City

Based on Figure 3 at 10:00 am, the solar panel produces a voltage of 17.63 V and a current of 0.89 Amp, resulting in an output power of 15.69 W with a panel surface temperature of 32.3 °C. As time progresses to 11:00 am, there is a slight decrease in power output to 15.44 W, with the current decreasing slightly to 0.88 Amp and the voltage stable at 17.55 V. The temperature also decreases to 30.1 °C. Entering 12:00 pm, the power output reaches its highest point of 15.74 W with a temperature that increases sharply to 43.5 °C. The voltage increases slightly to 17.68 V, while the current remains relatively stable at 0.89 Amp. The optimal performance of the panel during peak sunlight has reached its peak starting at 1:00 pm, the output power decreases to 13.80 W with a temperature reaching 45.1 °C. The current also started to decrease to 0.81 Amp, although the voltage remained at 17.04 V. The power decrease became more significant at 2:00 pm, where the power output was 13.28 W with a temperature of 41.5 °C and the current slightly decreased to 0.79 Amp, while the voltage was at 16.81 V. Finally, at 3:00 pm, the solar panel recorded the lowest power output of 10.62 W. At this time, the panel surface temperature was 42.8 °C, with the current dropping to 0.68 Amp and the voltage dropping to 15.62 V.

3.2 Accuracy of Random Forest Method

After the image results are obtained using thermovision technology, the image results are then processed using a method to obtain Accuracy, Precision, Recall, and F1-Score. The method that will be used is the Random Forest method, where the accuracy in this method is capable of processing data often higher than other algorithms because of its ability to handle data that has diverse and complex features. In addition, this technique is also effective in reducing the risk of overfitting and providing more stable results through the ensemble process. The results of machine learning processing using the Random Forest method with visualization of prediction accuracy, namely the actual value and predicted value on the test data and training data are shown in Figure 4 below.

ISSN: 3063-7929 (Online)
DOI: https://doi.org/10.30743/icst

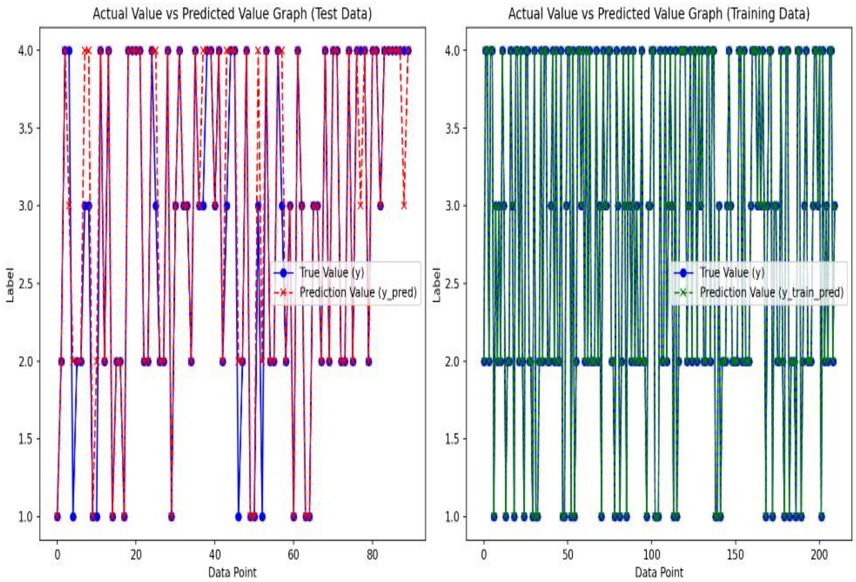


Figure 4. Random Forest Method Value visualization of prediction accuracy with actual values and predicted values on test data and training data

The test results of Figure 4 on the actual value graph with the predicted value for the Random Forest method show that in the test data, there is a significant variation between the actual value (blue dots) and the predicted value (red dots), indicating that the model may not be able to capture the pattern well. In contrast, in the training data, the predicted value looks more in line with the actual value, indicating that the model can learn well from the data. However, the better performance on the training data compared to the test data suggests the potential for overfitting. This analysis emphasizes the need for further development so that the model can show more consistent results. Therefore, the results of the analysis obtained that have been extracted values can be seen in Table 1 below.

Table 1. Random Forest Method Test Results

Class	precision	recall	f1-score	support
1	1.00	0.71	0.83	14
2	0.85	1.00	0.92	22
3	0.77	0.59	0.67	17
4	0.83	0.92	0.87	37
accuracy			0.84	90
macro avg	0.86	0.81	0.82	90
weighted avg	0.85	0.84	0.84	90
Accuracy	0.8444			

The classification model using the random forest method in Table 1 shows good performance with an accuracy of 84%, meaning that most of the predictions are correct. Class 1 has perfect precision (1.00), but its recall is lower (0.71), indicating that there are some undetected examples. Class 2 performs very well with a precision of 0.85 and a recall of 1.00, indicating that the model is able to detect all examples of this class. However, class 3 shows challenges with a precision of 0.77 and a recall of 0.59, indicating difficulties in identification, possibly due to data imbalance or less informative features. Class 4 also shows good performance with a precision of 0.83 and a recall of 0.92. The macro averages for precision (0.86), recall (0.81), and F1-score (0.82) show good stability across classes, while the weighted averages also reflect comparable values. Overall, this model shows good effectiveness in classification, although more attention is needed to improve the detection of class 3.

3.3 Accuracy of the K-Nearest Neighbors (KNN) Method

After the image is obtained using thermovision technology, the image will be processed with a method that aims to obtain Accuracy, Precision, Recall, and F1-Score values. In this analysis, the method used is K-Nearest Neighbors (KNN), which has a good reputation in data classification, especially when the dataset size is small. This method works by identifying the nearest neighbors of the data to be predicted, so that it can produce better and more

ISSN: 3063-7929 (Online)
DOI: https://doi.org/10.30743/icst

consistent accuracy. The results of machine learning processing using KNN will be visualized through a comparison of accuracy between actual values and predicted values on test data and training data, as shown in Figure 5 below.

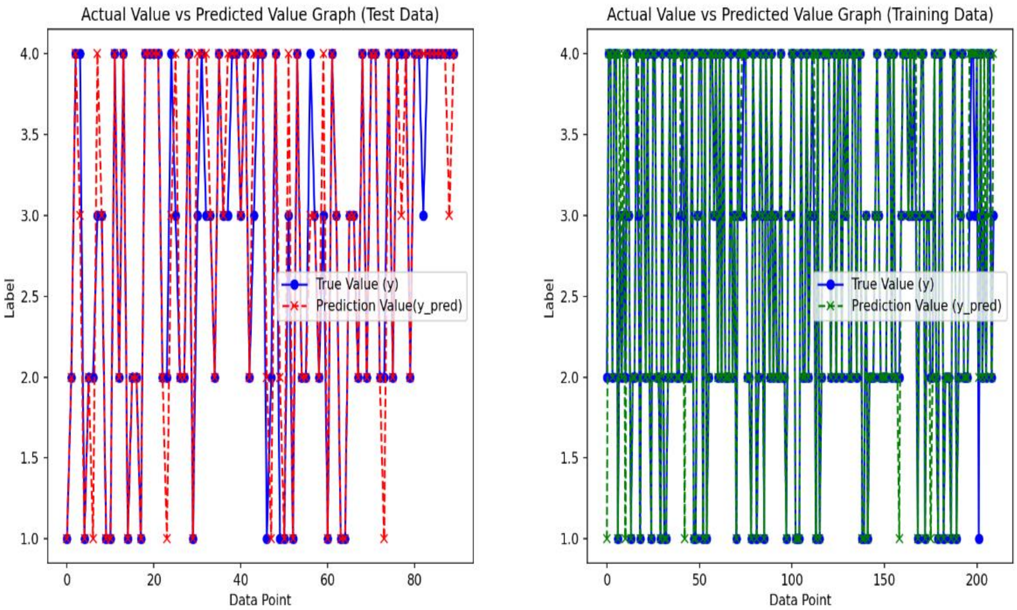


Figure 5. The value of the K-Nearest Neighbors (KNN) method visualizes the prediction accuracy with actual values and predicted values on test data and training data.

The value of the K-Nearest Neighbors (KNN) method in Figure 5 shows the comparison between the actual value (True Value) and the predicted value (Prediction Value) on the test data and training data. In the test data graph, there is a significant variation between the actual value and the predicted value, with some predictions far from the actual value, indicating the difficulty of the model in generalizing on data that has never been seen before. Therefore, the results of the analysis obtained that have been extracted values can be seen in Table 2 below.

Table 2. Performance Analysis of the KNN Method Classification Model

Class	precision	recall	f1-score	support
1	0.75	0.86	0.8	14
2	0.9	0.82	0.86	22
3	0.62	0.47	0.53	17
4	0.78	0.86	0.82	37
accuracy			0.78	90
macro avg	0.76	0.75	0.75	90
weighted avg	0.77	0.78	0.77	90
Accuracy	0.7778			

Classification Model Performance The K-Nearest Neighbors (KNN) method based on Table 2, has an accuracy of 78%, indicating that 78% of predictions are correct. The highest precision is found in class 2 (0.90), while class 3 has the lowest precision (0.62). The best recall is found in classes 1 and 4 (0.86), but class 3 records the lowest recall (0.47), indicating the challenge in detecting this class. F1-score shows good performance in most classes, but class 3 needs improvement. Overall, the model shows strength in some classes, but there is room for improvement especially in less detected classes.

4. CONLUSION

Based on the results of the research that has been done, it can be concluded that the application of the Random Forest algorithm for the classification of solar panel surface temperatures based on weather fluctuations in Medan City shows that there is a clear relationship between power output, temperature, current, and voltage throughout the measurement hours from 10:00 am to 3:00 pm. The highest energy conversion from the Solar Panel occurred at

12.00 pm with a voltage of 17.68 V, a current of 0.89 A, a power of 15.74 W and a temperature of 43.5 ° C, but there was a decrease in the performance of the solar panel at high temperatures such as 41 ° C, which caused the energy conversion effect of the solar panel to also decrease. Therefore, although solar panels use sunlight to generate electricity, temperatures that are too high can cause a decrease in efficiency. Then the implementation process of the Pattern Recognition method using the Random Forest Method and the K-Nearest Neighbors (KNN) Method includes several important stages, starting with collecting data through digital images of solar panels that represent various weather conditions, which are very crucial. Furthermore, feature extraction techniques such as Local Binary Pattern (LBP) and RGB mean values are used in pattern recognition to describe image features, contributing significantly to temperature classification. Model evaluation using metrics such as accuracy, precision, recall, and F1-score shows good performance in temperature classification, proving the effectiveness of the techniques applied in analyzing solar panel energy performance. The modeling testing process using Random Forest in data classification achieved an accuracy of 84% and the K-Nearest Neighbors (KNN) method had an accuracy of 78%.

ACKNOWLEDGMENT

The authors would like to thank the Master of Computer Science for its support and sponsorship. And we also thank the research collaboration of Universitas Potensi Utama and Universitas Asahan for supporting this research.

REFERENCES

[1] I. Laabab, S. Ziani, and A. Benami, "Solar panels overheating protection: a review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 1. 2023.

[2] R. Dallaev, T. Pisarenko, N. Papež, and V. Holcman, "Overview of the Current State of Flexible Solar Panels and Photovoltaic Materials," *Materials*, vol. 16, no. 17. 2023.

[3] D. Maizana, A. Rezky, S. M. Putri, H. Satria, and M. Mungkin, "Stability analysis of smart grid management system on campus building," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 3, 2023.

[4] M. Mungkin, H. Satria, D. Maizana, M. Isa, Syafii, and M. Y. Puriza, "Analysis of the feasibility of adding a grid-connected hybrid photovoltaic system to reduce electrical load," *Int. J. Power Electron. Drive Syst.*, vol. 14, no. 2, 2023.

[5] H. Satria, R. Syah, N. A. Silviana, and Syafii, "Sensitivity of solar panel energy conversion at sunrise and sunset on three weather fluctuations in equatorial climate," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 3, 2023.

[6] J. Huang, K. Zeng, Z. Zhang, and W. Zhong, "Solar panel defect detection design based on YOLO v5 algorithm," *Heliyon*, vol. 9, no. 8, 2023.

[7] A. Sekulić, M. Kilibarda, G. B. M. Heuvelink, M. Nikolić, and B. Bajat, "Random forest spatial interpolation," *Remote Sens.*, vol. 12, no. 10, 2020.

[8] S. Georganos *et al.*, "Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling," *Geocarto Int.*, vol. 36, no. 2, 2021.

[9] D. R. Anamisa, A. Jauhari, and F. Ayu Mufarroha, "K-Nearest Neighbors Method for Recommendation System in Bangkalan's Tourism," *ComTech Comput. Math. Eng. Appl.*, vol. 14, no. 1, 2023.

[10] M. Kim, Y. Kim, H. Kim, W. Piao, and C. Kim, "Evaluation of the k-nearest neighbor method for forecasting the influent characteristics of wastewater treatment plant," *Front. Environ. Sci. Eng.*, vol. 10, no. 2, 2016.

[11] M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," *Stata J.*, vol. 20, no. 1, 2020, doi: 10.1177/1536867X20909688.

[12] E. S. Almutairi and M. F. Abbod, "Machine Learning Methods for Diabetes Prevalence Classification in Saudi Arabia," *Modelling*, vol. 4, no. 1, 2023.

[13] J. J. S. De Figueiredo *et al.*, "Automatic detection and imaging of diffraction points using pattern recognition," *Geophys. Prospect.*, vol. 61, no. SUPPL.1, 2013.

[14] I. H. Sourov *et al.*, "EEG-Based Preference Classification for Neuromarketing Application," *Comput. Intell. Neurosci.*, vol. 2023, no. 1, 2023.

[15] N. Ukey, Z. Yang, B. Li, G. Zhang, Y. Hu, and W. Zhang, "Survey on Exact kNN Queries over High-Dimensional Data Space," *Sensors*, vol. 23, no. 2. 2023.

[16] D. Rahmawati, M. Puspa Putri I, M. Ulum, and K. Joni, "Identification and Classification of Pathogenic Bacteria Using the K-Nearest Neighbor Method," *JEEE-U (Journal Electr. Electron. Eng.*, vol. 5, no. 1, 2021.

[17] P. Y. Chen, C. P. Tung, J. H. Tsao, and C. J. Chen, "Assessing future rainfall intensity–duration–frequency characteristics across taiwan using the k-nearest neighbor method," *Water (Switzerland)*, vol. 13, no. 11,

ISSN: 3063-7929 (Online)
DOI: https://doi.org/10.30743/icst

2021.
[18] R. Andrian, D. Maharani, M. A. Muhammad, and A. Junaidi, "Butterfly identification using gray level co-occurrence matrix (Glcm) extraction feature and k-nearest neighbor (knn) classification," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 6, no. 1, 2020.
[19] R. Bakare, V. V. Shete, I. Kompatsiaris, and M. Tsolaki, "Revolutionizing the Alzheimer's Disease Stage Diagnosis through AI-Powered approach," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 13s, 2024.
[20] M. Wang *et al.*, "Functional localization in the brain of a cynomolgus monkey based on spike pattern recognition with machine learning," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 11, 2023.