Indonesian Journal of Electrical Engineering and Computer Science
Vol. 37, No. 1, January 2025, pp. 540~550
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v37.i1.pp540-550

Classification of weather conditions based on automatic weather station data using a multi-layer perceptron neural network

Muhammad Aristo Indrajaya, Tan Suryani Sollu, Mery Subito, Yuli Asmi Rahman, Erwin Ardias Saputra
Department of Electrical Engineering, Faculty of Engineering, University of Tadulako, Palu, Indonesia

Article Info

Article history:

Received Jun 30, 2024
Revised Sep 11, 2024
Accepted Sep 29, 2024

Keywords:

Accuracy
Altair AI studio
Classification
Multi-layer perceptron
Weather

ABSTRACT

Weather is one of the important elements that greatly determines human activities, especially those related to economic factors. Therefore, understanding weather conditions using weather parameters as a reference is important for human life, so a method is needed to classify weather according to its category so that the information produced can be used for various needs. Determining weather conditions in an area will not run well without a reliable method that can analyze existing weather parameters. Therefore, in this study, the weather condition classification process was carried out using the multilayer perceptron algorithm, a type of neural network (NN) algorithm. All data analyzed were weather parameter data collected by mini weather stations placed on land. The weather parameters used were temperature, humidity, air pressure, wind speed, dew point, wind chill, daily rainfall, solar radiation, and UV index. This study was conducted in Palu city, Central Sulawesi Province, Indonesia. The classification process carried out by the multilayer perceptron algorithm was carried out on the Altair AI Studio application and produced an accuracy value of 93.87%, recall of 92.33%, and precision of 91.29%.

Corresponding Author:

Mery Subito
Department of Electrical Engineering, Faculty of Engineering, University of Tadulako
Soekarno-Hatta Road Kilometer 9, Palu 94112, Central Sulawesi, Indonesia
Email: mery.subito@untad.ac.id

1. INTRODUCTION

Weather is a dynamic aspect of the Earth's atmosphere that changes frequently. Temperature, air pressure, humidity, light intensity, and rainfall all have an impact on the weather in a given location. Weather conditions greatly determine whether or not community activities run smoothly [1]. In the field of transportation, especially aviation and sea transportation, weather conditions are critical in terms of safety [2]. Many accidents, especially in the aviation sector, are indirectly caused by the influence of weather, so weather prediction is significant in supporting the transportation sector, especially in sea and air transportation. In the agricultural and plantation sectors, the role of weather in determining agricultural production results is huge [3], [4]. Whether that is in good condition will encourage maximum agricultural production and provide large economic benefits for farmers [5], [6]. On the other hand, extreme weather such as long droughts and extreme rains will cause crop failures which have implications for a significant reduction in profits for farmers and under certain conditions will cause scarcity of several agricultural commodities that are needed by the community. Therefore, the ability to predict weather conditions is essential in anticipating the influence of weather on the economic activities of a region [7], [8].

Data mining is a process that comprises data collecting and the application of past data to discover an order, pattern, or relationship in massive data [9]. Data mining involves four disciplines: statistics, visualization, databases, and machine learning. Machine learning is a subset of artificial intelligence that deals with building programming approaches based on previous data acquisition and intersects with statistics and optimization [10]. Currently, numerous programs use data mining techniques to classify weather, and research related to weather classification is not something very new. There have been several studies that have been carried out previously which have also been published as scientific articles. Zhao and Wu [11], in his research, conducted research using the convolutional neural network (CNN) method in the classification process which he carried out based on the visual conditions of each weather condition (sunny, foggy, rainy, snowy). Goel *et al.* [12] also carried out a weather condition classification process based on captured images of each weather condition and processed using the VGG19 model which produced an accuracy value of 98.5%. Dritsas *et al.* [13] classified weather conditions using five data attributes, namely day, precipitation, temperature, wind, and weather, and involved several techniques in the classification process, namely support vector machine (SVM), random forest (RF), and neural network (NN).

In this research, we carried out a process of classifying weather conditions based on the weather parameters that were collected. The research location itself is precisely in Palu city, Central Sulawesi Province, Indonesia. The city of Palu itself is a city crossed by the equator, which in other words is a tropical area with high levels of rainfall throughout the year. The parameter data used in this research are temperature, humidity, air pressure, wind speed, dew point, wind chill, daily rainfall, solar radiation, and UV index. This data was collected using the MISOL WS2320 automatic weather station device. To carry out this classification process ourselves, we use the Altair AI studio application, which was previously called RapidMiner studio. This application is handy in the data mining process, data prediction, and machine learning [14]. This classification method employs a multi-layer perceptron (MLP) algorithm, a form of NN technique. The parameters used to assess the performance of the algorithm are accuracy, precision, and recall.

2. METHOD

This section explains the basic theory of the algorithm used in this research along with the profile of the location where the data was collected along with the automatic weather station equipment used in this research. The algorithm used in this research is MLP and the research location itself is in Palu city, Central Sulawesi Province, Indonesia. In this research, the MISOL WS2320 device was also used as an automatic weather station which functions to collect weather parameter data.

2.1. Multi-layer perceptron algorithm

A MLP is a form of artificial neural network (ANN) comprising several layers of neurons [15]. The neurons of the MLP typically use nonlinear activation functions, which allows the network to learn complex data patterns. MLPs are important in machine learning because they can learn nonlinear correlations in data, resulting in effective models for classification, regression, and pattern recognition tasks. A multilayer perceptron is a feedforward NN of fully linked neurons that use a nonlinear activation function. It is commonly used to differentiate data that is not linearly separable. MLP itself consists of three layers, namely input, output, and more than one hidden layer [16], [17]. The layer arrangement of the MLP can be seen in Figure 1. Each layer in the MLP has its functions. The functions are as follows:

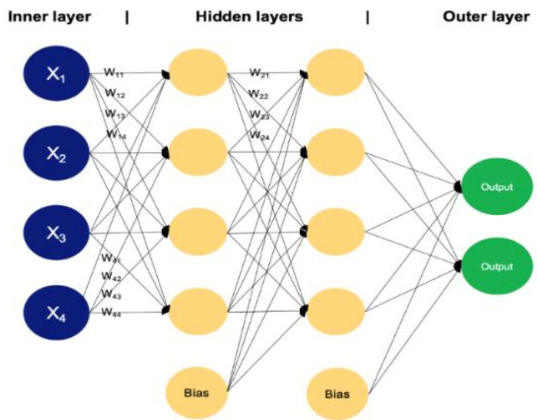


Figure 1. An example of a MLP with two hidden layers

2.1.1. Input layer

The input layer is made up of nodes or neurons that receive initial input data [18]. Each neuron represents a certain aspect or dimension of the input data. The number of neurons in the input layer is dependent on the input data's dimensionality [19]. The input layer of a MLP accepts input data, which may be features taken from input samples in a dataset. Each neuron in the input layer represents a single feature, and they do not do any computations; instead, they pass the input values to the neurons in the first hidden layer.

2.1.2. Hidden layer

Each neuron in a hidden layer receives inputs from all neurons in the preceding layer (either the input layer or another hidden layer) and creates an output that is passed on to the following layer [20]. The number of hidden layers and neurons in each hidden layer are hyperparameters that must be determined during the model creation stage [21]. The hidden layers of a MLP are made up of interconnected neurons that execute computations on the incoming data. Each neuron in a hidden layer receives input from every neuron in the previous layer [22]. The inputs are multiplied by the appropriate weights, represented as w. The weights govern how much influence one neuron's input has on another's output. In addition to weights, each neuron in the hidden layer is assigned a bias, denoted as b. The bias adds more input to the neuron, allowing it to change its output threshold. Biases, like weights, are learned during the training process. The weighted total of each neuron's inputs in a hidden or output layer is calculated. This entails multiplying each input by its matching weight, totaling the products, and adding the bias expressed in (1).

$$Weighted\ Sum = \sum_{i=1}^{n}(W_i \times X_i) + b \tag{1}$$

Where n is the total number of input connections, w_i is the weight of the i input, and x_i is the i input value. The weighted sum is subsequently processed by an activation function, indicated as f. The activation function adds nonlinearity to the network, letting it learn and express complicated correlations in data. The activation function defines the neuron's output range and behavior in response to various input values. The activation function used is determined by the nature of the task and the network's desired qualities.

2.1.3. Output layer

The neurons in this layer produce the network's final output. The number of neurons in the output layer varies according to the task. In binary classification, depending on the activation function, there may be one or two neurons expressing the likelihood of belonging to one class; however, in multi-class classification tasks, the output layer may have several neurons. The output layer of a MLP generates the network's final predictions or outputs. The number of neurons in the output layer depends on the task at hand (e.g., binary classification, multi-class classification, regression) [16]. Each neuron in the output layer receives input from the neurons in the last hidden layer and performs an activation function. This activation function is typically distinct from those employed in the hidden layers, and it generates the final output value or prediction.

2.2. Research location and dataset collection

As explained in the introduction, this research is located in Palu city, Central Sulawesi Province, Indonesia. Palu, the capital of Central Sulawesi Province, Indonesia, is located on the Palu Valley and Palu Bay plains. The region has five dimensions: mountainous terrain, valleys, rivers, bays, and oceans. Palu city has an area of 395.06 square kilometers divided into eight sub-districts. Palu city is located between 0°.36"-0°.56" South Latitude and 119°.45" - 121°.1" East Longitude, making it directly on the Equator with a 0-700 meters high above sea level. The map of Palu city itself, which is located on Sulawesi Island, Indonesia, can be seen in Figure 2.

Palu city, like other regions in Indonesia, has two seasons: summer and rainy season. Summer lasts from April to September, and the rainy season runs from October to March. The maximum rainfall recorded at the Mutiara Palu Meteorological Station in 2010 was 123.0 mm in June, with the lowest being 11.7 mm in March. Meanwhile, in 2010, the average wind speed was 3.7 knots. Palu city's location on the equator also contributes significantly to the present microclimate. Of course, geographical characteristics like this will cause Palu City to receive a lot of sunlight for an extended period, making the location hotter than other cities [23].

The weather parameter data used in this research are temperature, humidity, air pressure, wind speed, solar radiation, UV index, and rainfall levels. All of these parameters can be used using an automatic weather station device, namely MISOL WS2320. This product itself is made by Jiaxing Misol Electronics Co., Ltd. This device is often used by other researchers to detect weather parameters in an area and display them on a website page. This device consists of two main parts, namely an external device consisting of weather parameter sensors placed in the external environment which can be seen in Figure 3. In this research, we placed this device on a building that is the highest compared to its surroundings so that it can measure wind speed well.

Apart from the weather parameter sensors placed outside the building which are received via radio communication lines, this weather station has a device that functions as a receiver of the collected data which can be seen in Figure 4. This receiving device in addition to collecting the data collected obtained by the sensor can also display factual data regarding the condition of the measurement results of weather data via its LCD screen. Apart from that, this device is also capable of sending the collected data to the computer in real-time.

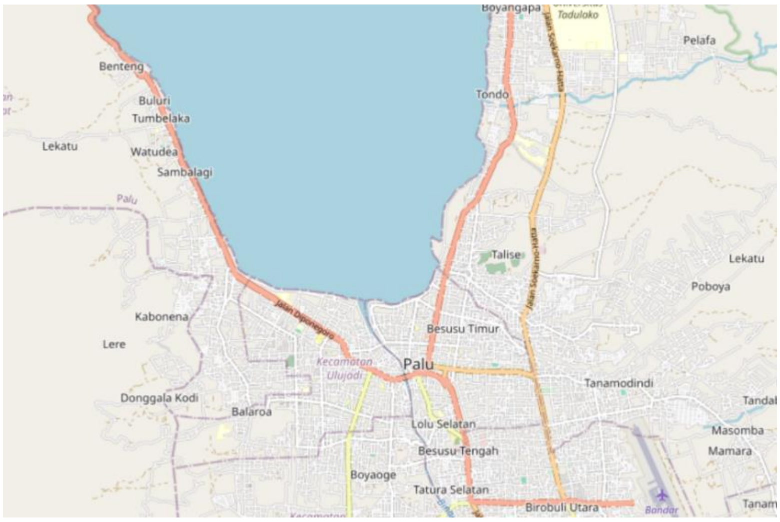


Figure 2. The maps of Palu city

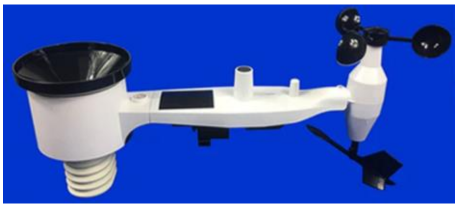


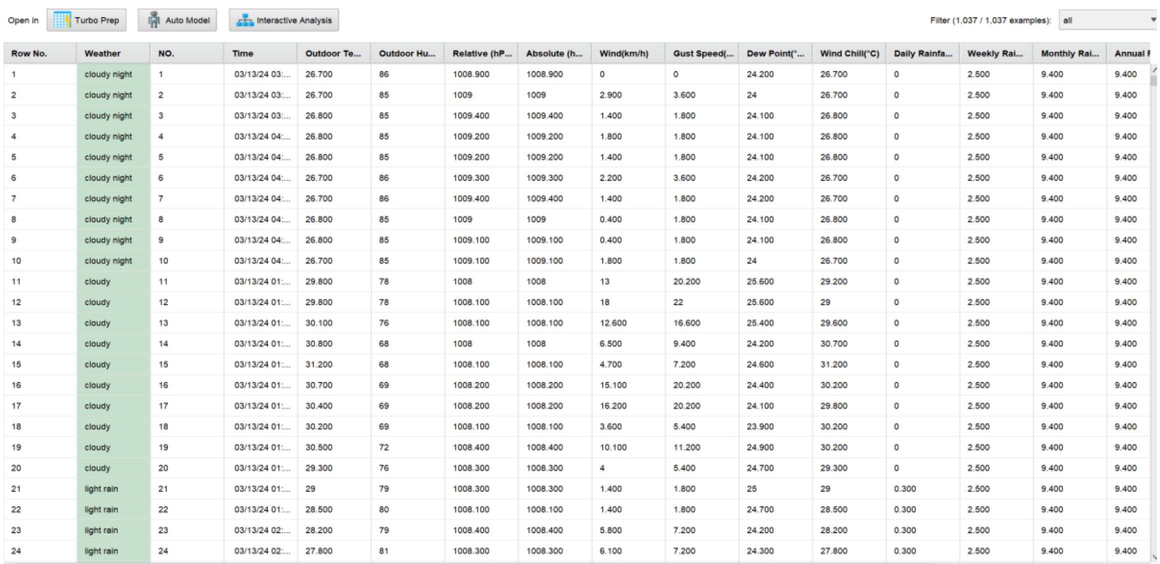
Figure 3. MISOL WS2320 outdoor sensor

Figure 4. MISOL WS2320 receiver

Data collection was carried out for six months starting from January to June, where the data excerpt can be seen in Figure 5. Data collection is carried out in weather conditions where the weather can change all the time. This is because, in May, Palu city experiences a transition season so that in one month the weather can change quite quickly. For example, during the day the weather is sunny, but during the day the weather conditions can immediately change to cloudy and not long after that it can rain quite heavily. Through data collection, 1037 weather parameter data were obtained which were divided into several classes, namely:

1. Sunny
   Sunny weather is a weather condition that describes the sky in bright conditions and sunlight can be felt clearly and has a high level of UV radiation.
2. Sunny cloudy
   Sunny, cloudy weather is characterized by the large number of clouds visible in the sky. In this condition, the intensity of sunlight is not as high as in sunny weather which is characterized by low levels of solar radiation and UV index.
3. Cloudy
   Cloudy weather is characterized by the sky being covered with thick clouds. These clouds block sunlight, causing very low solar radiation and UV index values.
4. Light rain
   Light rain is a weather condition with low intensity with a rain rate of less than 2.5 mm/h (<0.1”/hr) [24].
5. Moderate rain
   Moderate rain occurs when the rainfall rate ranges from 2.6 to 7.5 mm/hour (0.1 to 0.3"/hour) or 0.04 to 0.125 mm/minute (0.0017 to 0.005"/hour), which equates to 26 to 75 full tipping buckets per hour (26 to 75 pulses/hour) for a rain gauge with a resolution of 0.1 mm [24].
6. Heavy rain
   Heavy rain is a rainy weather condition in which the rain rate exceeds 7.6 to 50 mm/hour (0.3 to 2"/hour) or 0.125 to 0.83 mm/minute (0.005" to 0.033"/minute), which amounts to 76 or more full tipping buckets per hour (76+ pulses/hour) on a rain gauge with a resolution of 0.1 mm [24].
7. Clear night
   A clear night is a weather condition where the sky appears clear of clouds and the stars can be seen clearly.
8. Cloudy night
   This is a weather condition at night where clouds cover the sky and the eye cannot see the stars.
9. Light rainy night
   This is a rainy condition that occurs at night when the rain rate is less than 2.5 mm/hour (<0.1"/hour).
10. Moderate rainy night
   This is a rainy condition at night where rainfall ranges from 2.6 to 7.5 mm/hour (0.1 to 0.3"/hour) or 0.04 to 0.125 mm/minute (0.0017 to 0.005"/ hour), which is equivalent to 26 to 75 full tipping buckets per hour (26 to 75 pulses/hour) for a rain gauge with a resolution of 0.1 mm.
11. Heavy rainy night

It is a rainy weather condition that occurs at night when the rain rate exceeds 7.6 to 50 mm/hour (0.3 to 2"/hour) or 0.125 to 0.83 mm/minute (0.005" to 0.033"/minute), which amounts to 76 or more full tipping buckets per hour (76+ pulses/hour) on a rain gauge with a resolution of 0.1 mm.



Open in Turbo Prep Auto Model Interactive Analysis Filter (1,037 / 1,037 examples): all

Row No.	Weather	NO.	Time	Outdoor Te...	Outdoor Hu...	Relative (hP...	Absolute (h...	Wind(km/h)	Gust Speed(...	Dew Point(°...	Wind Chill(°C)	Daily Rainfa...	Weekly Rai...	Monthly Rai...	Annual R
1	cloudy night	1	03/13/24 03:...	26.700	86	1008.900	1008.900	0	0	24.200	26.700	0	2.500	9.400	9.400
2	cloudy night	2	03/13/24 03:...	26.700	85	1009	1009	2.900	3.600	24	26.700	0	2.500	9.400	9.400
3	cloudy night	3	03/13/24 03:...	26.800	85	1009.400	1009.400	1.400	1.800	24.100	26.800	0	2.500	9.400	9.400
4	cloudy night	4	03/13/24 04:...	26.800	85	1009.200	1009.200	1.800	1.800	24.100	26.800	0	2.500	9.400	9.400
5	cloudy night	5	03/13/24 04:...	26.800	85	1009.200	1009.200	1.400	1.800	24.100	26.800	0	2.500	9.400	9.400
6	cloudy night	6	03/13/24 04:...	26.700	86	1009.300	1009.300	2.200	3.600	24.200	26.700	0	2.500	9.400	9.400
7	cloudy night	7	03/13/24 04:...	26.700	86	1009.400	1009.400	1.400	1.800	24.200	26.700	0	2.500	9.400	9.400
8	cloudy night	8	03/13/24 04:...	26.800	85	1009	1009	0.400	1.800	24.100	26.800	0	2.500	9.400	9.400
9	cloudy night	9	03/13/24 04:...	26.800	85	1009.100	1009.100	0.400	1.800	24.100	26.800	0	2.500	9.400	9.400
10	cloudy night	10	03/13/24 04:...	26.700	85	1009.100	1009.100	1.800	1.800	24	26.700	0	2.500	9.400	9.400
11	cloudy	11	03/13/24 01:...	29.800	78	1008	1008	13	20.200	25.600	29.200	0	2.500	9.400	9.400
12	cloudy	12	03/13/24 01:...	29.800	78	1008.100	1008.100	18	22	25.600	29	0	2.500	9.400	9.400
13	cloudy	13	03/13/24 01:...	30.100	76	1008.100	1008.100	12.600	16.600	25.400	29.600	0	2.500	9.400	9.400
14	cloudy	14	03/13/24 01:...	30.800	68	1008	1008	6.500	9.400	24.200	30.700	0	2.500	9.400	9.400
15	cloudy	15	03/13/24 01:...	31.200	68	1008.100	1008.100	4.700	7.200	24.600	31.200	0	2.500	9.400	9.400
16	cloudy	16	03/13/24 01:...	30.700	69	1008.200	1008.200	15.100	20.200	24.400	30.200	0	2.500	9.400	9.400
17	cloudy	17	03/13/24 01:...	30.400	69	1008.200	1008.200	16.200	20.200	24.100	29.800	0	2.500	9.400	9.400
18	cloudy	18	03/13/24 01:...	30.200	69	1008.100	1008.100	3.600	5.400	23.900	30.200	0	2.500	9.400	9.400
19	cloudy	19	03/13/24 01:...	30.500	72	1008.400	1008.400	10.100	11.200	24.900	30.200	0	2.500	9.400	9.400
20	cloudy	20	03/13/24 01:...	29.300	76	1008.300	1008.300	4	5.400	24.700	29.300	0	2.500	9.400	9.400
21	light rain	21	03/13/24 01:...	29	79	1008.300	1008.300	1.400	1.800	25	29	0.300	2.500	9.400	9.400
22	light rain	22	03/13/24 01:...	28.500	80	1008.100	1008.100	1.400	1.800	24.700	28.500	0.300	2.500	9.400	9.400
23	light rain	23	03/13/24 02:...	28.200	79	1008.400	1008.400	5.800	7.200	24.200	28.200	0.300	2.500	9.400	9.400
24	light rain	24	03/13/24 02:...	27.800	81	1008.300	1008.300	6.100	7.200	24.300	27.800	0.300	2.500	9.400	9.400

Figure 5. A snapshot of the data that has been entered into the Altair AI studio application

3. RESULTS AND DISCUSSION

This section explains the data processing process - parameter data that has been obtained using the MISOL WS2320 automatic weather station. The data obtained in real-time is then processed using the MLP algorithm in the Altair AI studio application. Altair AI studio formerly known as RapidMiner is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analysis. It is used for business and commercial purposes, as well as for research, education, training, rapid prototyping, and application development, and supports all steps in the machine learning process including data preparation, results visualization, model validation, and optimization.

3.1. Data preprocessing

Processing of the data obtained is carried out through several stages. The process steps carried out using Altair AI Studio can be seen in Figure 6. Previously, the data sent by the automatic weather station device was in a file with the CSV extension. To make processing easier, the file is then converted to the xlsx extension.

Before the data is processed by the MLP algorithm, the data first goes through a stage called preprocessing. This process begins with the select attributes process. In this process, the year, month, and time of data collection are filtered to be processed into the algorithm. This aims to avoid errors in the three training processes in the algorithm. Furthermore, in the remove duplicates section, duplicate data will be removed. Furthermore, in the split data section, the weather parameter data that has been collected is divided based on the needs for the training and testing processes. The division itself is 70% for the training process and 30% for the testing process. The performance of the training process is measured using three parameters, namely :

1. Accuracy
   Accuracy is an evaluation metric that measures how well a model makes correct predictions out of the total predictions it makes. In the context of classification, accuracy provides an idea of how often a model predicts the correct class, whether it is positive or negative [25].
2. Precision
   Precision is an evaluation metric that measures how well a model makes correct predictions for the positive class out of the total positive predictions. In the context of classification, precision provides an idea of how often a model correctly predicts the positive class, among all the positive predictions made by the model [25].
3. Recall
   Recall is a metric that measures how many positive cases the classifier accurately predicted out of all the positive cases in the dataset. It is sometimes referred to as sensitivity [25].

In the data processing stage, in addition to the data analysis process involving the MLP algorithm, one of the important stages in data processing is cross-validation. Cross-validation is a statistical technique used in machine learning and other predictive modeling to assess the performance and generalization ability of a model. This technique provides a more accurate estimate of model performance on unseen data. Another important benefit of cross-validation is that it helps data analysts overcome the problem of overfitting or conditions when the model is too specific to the training data so that it is less good at analyzing new data [26].

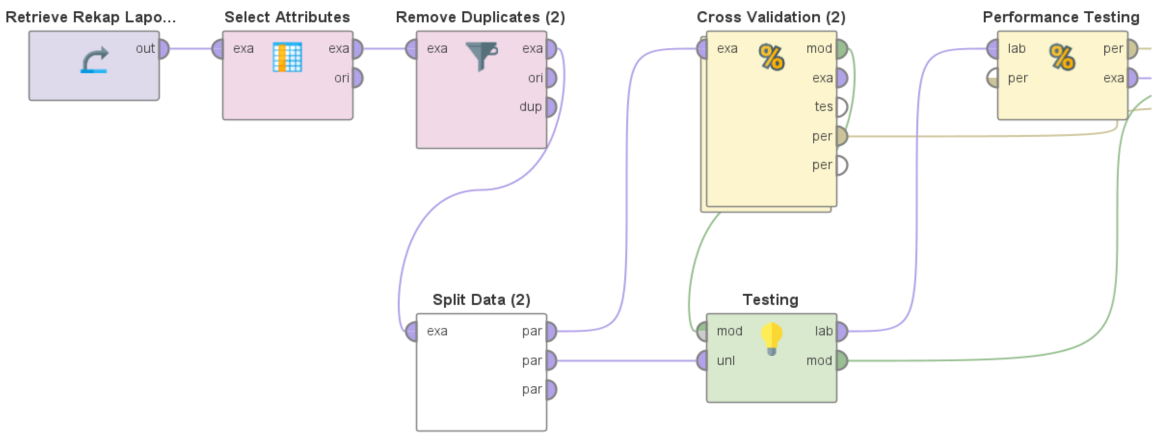


Figure 6. Data processing diagrams

3.2. Performance analysis

As previously explained, the data analysis used in this study uses a MLP algorithm commonly abbreviated as MLP. The analysis process carried out in this study was carried out through several stages. In the first stage, weather parameter data analysis was carried out using the standard MLP configuration in the Altair AI Studio application. In Altair AI Studio, the performance of the MLP is configured using parameters, namely training cycle, number of generations, and number of ensembles. These parameters will affect the ability of the algorithm to learn the input data. By using the default configuration of Altair AI Studio, namely the training cycle is set to 10, the number of generations is set to 10, and the number of ensembles is set to 4, the results of the analysis performance carried out by the learning and training process can be seen in Figures 7 and 8.

accuracy: 87.79% +/- 3.34% (micro average: 87.79%)

	true cloudy night	true cloudy	true light rain	true sunny clou...	true clear night	true moderate r...	true heavy rain	true sunny	true light rainy ...	true moderate r...	true heavy rain...	class precision
pred. cloudy ni...	37	0	1	0	0	0	0	0	2	0	0	92.50%
pred. cloudy	0	117	2	3	3	0	0	4	0	0	0	90.70%
pred. light rain	1	1	82	2	2	0	0	2	3	0	0	88.17%
pred. sunny clo...	0	1	4	138	0	0	0	1	0	0	0	95.83%
pred. clear night	5	0	0	0	50	0	0	3	0	0	0	86.21%
pred. moderate...	0	0	3	0	0	32	3	0	0	1	0	82.05%
pred. heavy rain	0	0	0	0	0	2	7	0	0	0	0	77.78%
pred. sunny	0	3	2	2	1	0	0	36	0	0	0	81.82%
pred. light rain...	0	0	13	0	0	0	0	4	29	1	0	61.70%
pred. moderate...	0	0	1	0	1	0	0	0	2	50	0	92.59%
pred. heavy rai...	0	0	1	0	0	0	2	0	0	6	55	85.94%
class recall	86.05%	95.90%	75.23%	95.17%	87.72%	94.12%	58.33%	72.00%	80.56%	86.21%	100.00%	

Figure 7. Accuracy value of testing in the training process

accuracy: 85.16%

	true cloudy night	true cloudy	true light rain	true sunny clou...	true clear night	true moderate r...	true heavy rain	true sunny	true light rainy ...	true moderate r...	true heavy rain...	class precision
pred. cloudy ni...	14	1	0	0	0	0	0	0	0	0	0	93.33%
pred. cloudy	0	48	0	0	1	0	0	0	0	0	0	97.96%
pred. light rain	0	3	29	1	0	1	0	0	0	0	0	85.29%
pred. sunny clo...	0	0	2	57	0	0	0	1	0	0	0	95.00%
pred. clear night	4	0	1	0	22	0	0	2	0	0	0	75.86%
pred. moderate...	0	0	0	0	0	13	1	0	0	0	0	92.86%
pred. heavy rain	0	0	0	0	0	0	4	0	0	0	0	100.00%
pred. sunny	0	1	3	4	2	0	0	19	0	0	0	65.52%
pred. light rain...	0	0	12	0	0	0	0	0	16	4	0	50.00%
pred. moderate...	0	0	0	0	0	0	0	0	0	19	0	100.00%
pred. heavy rai...	0	0	0	0	0	0	0	0	0	2	23	92.00%
class recall	77.78%	90.57%	61.70%	91.94%	88.00%	92.86%	80.00%	86.36%	100.00%	76.00%	100.00%	

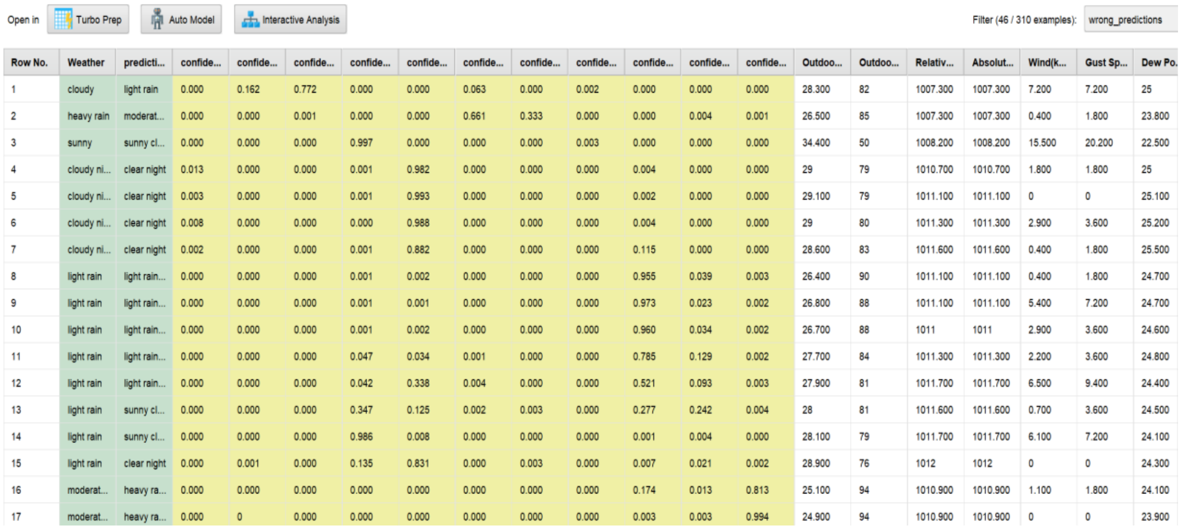
Figure 8. Accuracy value of testing in the testing process

Based on the analysis process carried out using standard configuration values, the accuracy value in the training process was 87.79% and 85.16% in the testing process. The detailed performance comparison values in the training and testing processes can be seen in Table 1.

Although based on the performance generated from the analysis carried out, the accuracy of the classified data is still quite high, which is above 85% in the training and testing process. Also, the accuracy results in the learning process and testing process are not striking so they can be categorized as not overfitting. However, the number of incorrect data in the classification process is still quite a lot. In the testing process, out of 310 data used in the testing process, 46 data were predicted incorrectly. A snippet of incorrectly predicted data can be seen in Figure 9.

Table 1. Performance results of the analysis performed by the MLP algorithm using the standard configuration

Parameters	Training	Testing
Accuracy	87.79% +/- 3.34%	85.16%
Precision	84.58% +/- 6.52%	86.17%
Recall	84.28%	85.93%



Open in Turbo Prep Auto Model Interactive Analysis

Filter (46 / 310 examples): wrong_predictions

Row No.	Weather	predicti...	confide...	confide...	confide...	confide...	confide...	confide...	confide...	confide...	confide...	confide...	confide...	Outdoo...	Outdoo...	Relativ...	Absolut...	Wind(k...	Gust Sp...	Dew Po.
1	cloudy	light rain	0.000	0.162	0.772	0.000	0.000	0.063	0.000	0.002	0.000	0.000	0.000	28.300	82	1007.300	1007.300	7.200	7.200	25
2	heavy rain	moderat...	0.000	0.000	0.001	0.000	0.000	0.661	0.333	0.000	0.000	0.004	0.001	26.500	85	1007.300	1007.300	0.400	1.800	23.800
3	sunny	sunny cl...	0.000	0.000	0.000	0.997	0.000	0.000	0.000	0.003	0.000	0.000	0.000	34.400	50	1008.200	1008.200	15.500	20.200	22.500
4	cloudy ni...	clear night	0.013	0.000	0.000	0.001	0.982	0.000	0.000	0.000	0.004	0.000	0.000	29	79	1010.700	1010.700	1.800	1.800	25
5	cloudy ni...	clear night	0.003	0.000	0.000	0.001	0.993	0.000	0.000	0.000	0.002	0.000	0.000	29.100	79	1011.100	1011.100	0	0	25.100
6	cloudy ni...	clear night	0.008	0.000	0.000	0.000	0.988	0.000	0.000	0.000	0.004	0.000	0.000	29	80	1011.300	1011.300	2.900	3.600	25.200
7	cloudy ni...	clear night	0.002	0.000	0.000	0.001	0.882	0.000	0.000	0.000	0.115	0.000	0.000	28.600	83	1011.600	1011.600	0.400	1.800	25.500
8	light rain	light rain...	0.000	0.000	0.000	0.001	0.002	0.000	0.000	0.000	0.955	0.039	0.003	26.400	90	1011.100	1011.100	0.400	1.800	24.700
9	light rain	light rain...	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.973	0.023	0.002	26.800	88	1011.100	1011.100	5.400	7.200	24.700
10	light rain	light rain...	0.000	0.000	0.000	0.001	0.002	0.000	0.000	0.000	0.960	0.034	0.002	26.700	88	1011	1011	2.900	3.600	24.600
11	light rain	light rain...	0.000	0.000	0.000	0.047	0.034	0.001	0.000	0.000	0.785	0.129	0.002	27.700	84	1011.300	1011.300	2.200	3.600	24.800
12	light rain	light rain...	0.000	0.000	0.000	0.042	0.338	0.004	0.000	0.000	0.521	0.093	0.003	27.900	81	1011.700	1011.700	6.500	9.400	24.400
13	light rain	sunny cl...	0.000	0.000	0.000	0.347	0.125	0.002	0.003	0.000	0.277	0.242	0.004	28	81	1011.600	1011.600	0.700	3.600	24.500
14	light rain	sunny cl...	0.000	0.000	0.000	0.986	0.008	0.000	0.000	0.000	0.001	0.004	0.000	28.100	79	1011.700	1011.700	6.100	7.200	24.100
15	light rain	clear night	0.000	0.001	0.000	0.135	0.831	0.000	0.003	0.000	0.007	0.021	0.002	28.900	76	1012	1012	0	0	24.300
16	moderat...	heavy ra...	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.174	0.013	0.813	25.100	94	1010.900	1010.900	1.100	1.800	24.100
17	moderat...	heavy ra...	0.000	0	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.003	0.994	24.900	94	1010.900	1010.900	0	0	23.900

Figure 9. Incorrect data in the analysis process

To overcome the many errors, in the next stage we took the initiative to increase the training cycle, the number of ensembles, and the number of generations. The addition of values set for the three parameters will affect the algorithm's ability to learn input data. The process of adding parameter values is carried out gradually and results in an increase in the accuracy, precision, and recall values. This can happen because when the algorithm's learning cycle increases, it will also increase the number of neurons used by the algorithm. As a comparison, in the process using the default value, the number of neurons used in the hidden layer is 43 units. The relationship between the addition of training cycles, the number of ensembles, and the number of generations for the performance of the MLP algorithm in the training and testing processes can be seen in Table 2 and Table 3.

Table 2. Performance of the training process with modified configuration

Training cycles	Number of ensemble	Number of generation	Accuracy	Precision	Recall
15	15	15	89.87% +/- 3.48%	87.94% +/- 5.57%	88.36% +/- 4.68%
20	20	20	88.91% +/- 3.58%	86.12% +/- 5.27%	86.07% +/- 3.66%
25	25	25	89.60% +/- 4.72%	87.16% +/- 5.94%	86.98% +/- 6.34%
30	30	30	89.32% +/- 4.04%	89.01% +/- 5.09%	88.32% +/- 5.10%
35	35	35	90.01% +/- 4.24%	88.65% +/- 5.26%	87.83% +/- 4.26%
40	40	40	90.29% +/- 1.95%	86.82% +/- 5.85%	86.47% +/- 5.07%
42	42	42	92.10% +/- 3.06%	88.79% +/- 5.46%	88.87% +/- 5.09%
44	44	44	92.09% +/- 4.09%	90.26% +/- 4.76%	90.24% +/- 3.80%
45	45	45	90.43% +/- 3.35%	87.81% +/- 6.83%	87.34% +/- 5.68%
46	46	46	90.85% +/- 3.97%	88.25% +/- 5.99%	87.76% +/- 6.15%
47	47	47	91.82% +/- 2.89%	90.06% +/- 4.55%	89.22% +/- 4.21%
48	48	48	91.82% +/- 3.29%	88.50% +/- 5.84%	87.74% +/- 5.86%

Table 3. Performance of the testing process with modified configuration

Training cycles	Number of ensemble	Number of generation	Accuracy	Precision	Recall
15	15	15	87.74%	88.92%	90.36%
20	20	20	88.39%	86.12%	87.15%
25	25	25	94.52%	94.22%	93.50%
30	30	30	92.90%	94.41%	94.20%
35	35	35	90.97%	89.71%	92.93%
40	40	40	91.29%	92.10%	92.23%
42	42	42	88.39%	89.46%,	91.39%
44	44	44	93.87%	91.29%	92.39%
45	45	45	93.87%	92.75%	92.72%
46	46	46	91.94%	91.13%	91.44%
47	47	47	89.03%	87.06%	91.92%
48	48	48	89.68%	87.59%	90.48%

Based on the calculations, the best accuracy value for the MLP algorithm was 90.43% +/- 3.35% in the training process and 93.87% in the testing process. This value was achieved by setting the training cycle configuration, number of ensembles, and number of generations to 45. At this figure, the precision was 87.81% +/- 6.83% for training and 92.75% for testing, while the recall was 87.34% +/- 5.68% for the training process and 92.72% for training. Although testing was conducted until the three parameters reached 48, there was a risk of overfitting. This is evident from the increasingly significant discrepancy in accuracy scores between the training and testing processes, indicating that the testing method is overly reliant on a single model.

By increasing the training cycle value, the number of ensembles, and the number of generations that affect the level of learning carried out by the algorithm, it shows that its performance is getting better, where previously the accuracy value obtained was 87.79% +/- 3.34% for the learning process and 85.16% for the testing process. In addition to increasing performance, increasing the value of the three parameters also affects the number of neurons in the hidden layer used in data processing, if previously in the default parameters, the number of neurons used was 43 units, then after the value of the three parameters was increased to reach the best value of 45, the number of neurons used became 145 units. The increase in performance, especially on the accuracy side, finally had an impact on reducing the number of incorrect data in the classification process. If in the initial test with the default parameters the number of incorrect data was 46 data, then at the best accuracy value, the number of incorrect data dropped to 19 data. The wrong prediction data snippet can be seen in Figure 10.

Open in Turbo Prep Auto Model Interactive Analysis

Filter (19 / 310 examples): wrong_predictions

Row No.	Weather	predicti...	confide...	confide...	confide...	confide...	confide...	confide...	confide...	confide...	confide...	confide...	confide...	Outdoo...	Outdoo...	Relativ...	Absolut...	Wind(k...	Gust Sp...	Dew Po...
1	cloudy	light rain	0.000	0.317	0.605	0.000	0.000	0.000	0.000	0.077	0	0.000	0.000	30.800	68	1008	1008	6.500	9.400	24.200
2	cloudy	light rain	0.000	0.242	0.756	0.000	0.000	0.002	0.000	0.000	0	0.000	0.000	28.300	82	1007.300	1007.300	7.200	7.200	25
3	sunny	sunny cl...	0.000	0.000	0	1.000	0.000	0.000	0.000	0.000	0	0	0.000	34.400	50	1008.200	1008.200	15.500	20.200	22.500
4	moderat...	moderat...	0	0.000	0.000	0.000	0.000	0.464	0.001	0.000	0.000	0.535	0.000	28.100	88	1008.700	1008.700	11.900	20.200	26
5	moderat...	heavy ra...	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.995	25.100	94	1010.900	1010.900	1.100	1.800	24.100
6	moderat...	heavy ra...	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	24.900	94	1010.900	1010.900	0	0	23.900
7	light rain...	clear night	0.000	0.000	0.000	0.000	0.985	0.001	0.002	0.000	0.012	0.000	0.000	26	91	1009.600	1009.600	0	0	24.400
8	light rain...	clear night	0.000	0.000	0.000	0.000	0.967	0.000	0.017	0.000	0.009	0.007	0.000	26	91	1009.700	1009.700	0	0	24.400
9	moderat...	light rain...	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.982	0.011	0.002	25.800	93	1010.100	1010.100	5.400	7.200	24.600
10	clear night	moderat...	0.000	0.000	0.000	0.000	0.000	0.998	0.000	0.000	0.000	0.002	0.000	26.900	88	1006	1006	5	7.200	24.800
11	clear night	moderat...	0.001	0.000	0.000	0.000	0.000	0.995	0.000	0.000	0.000	0.003	0.000	26.800	89	1006	1006	3.600	3.600	24.900
12	clear night	moderat...	0.000	0.000	0.000	0.000	0.000	0.998	0.000	0.000	0.000	0.001	0.000	26.700	90	1005.900	1005.900	1.400	1.800	25
13	moderat...	light rain...	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.923	0.072	0.005	25	95	1009.500	1009.500	5	5.400	24.200
14	cloudy ni...	cloudy	0.040	0.466	0.174	0.000	0.065	0.165	0.080	0.000	0.011	0.000	0.001	25.500	93	1008.700	1008.700	2.900	3.600	24.300
15	sunny	cloudy	0.000	0.751	0.002	0.000	0.206	0.005	0.034	0.000	0.002	0.000	0.000	29.400	76	1009.200	1009.200	0	0	24.800
16	sunny	clear night	0.000	0.033	0.001	0.000	0.929	0.006	0.028	0.000	0.003	0.000	0.000	28.900	78	1009.300	1009.300	0	0	24.700
17	sunny cl...	sunny	0.000	0.000	0.000	0.398	0	0.000	0.000	0.602	0.000	0.000	0.000	30.900	79	1006.100	1006.100	13	18.400	26.900
18	sunny cl...	sunny	0.000	0.000	0.000	0.354	0.000	0.000	0.000	0.646	0.000	0.000	0.000	30.800	79	1006.100	1006.100	14.800	23.800	26.800
19	light rain...	moderat...	0.000	0.000	0.000	0.000	0.000	0.006	0.002	0.000	0.000	0.992	0.000	25.700	95	1007.500	1007.500	1.800	1.800	24.900

Figure 10. Incorrect data snippets in the data processing process using MLP with modified parameters

4. CONCLUSION

The purpose of this study was to test whether the MLP algorithm is capable of classifying weather conditions with a high level of accuracy and a low error rate. Through the tests that have been carried out and based on the performance values obtained, it can be concluded that the MLP algorithm is capable of classifying weather with a fairly high level of accuracy, which is 93.87% in the testing process. This test also shows that changing the training cycles, number of generations, and number of ensemble parameters that affect the learning ability of the MLP algorithm will significantly increase the accuracy of the MLP algorithm. The increase in performance has an impact on reducing the number of incorrect data in the classification process, where if in the initial test, the number of incorrect data was 46 data in the testing process, then in the process with the best level of accuracy, the number of incorrect data dropped to 19 data. However, this increase also has limitations, where when the three parameters are set to values that are to high, the resulting performance values tend to experience overfitting which is marked by a striking difference in accuracy values in the training and testing processes. Therefore, to obtain better accuracy values from this study, it is highly recommended to use superior algorithms such as deep learning in further research.

REFERENCES

[1] C. Fang, C. Lv, F. Cai, H. Liu, J. Wang, and M. Shuai, "Weather classification for outdoor power monitoring based on improved SqueezeNet," in *Proceedings - 2020 5th International Conference on Information Science, Computer Technology and Transportation, ISCTT 2020*, Nov. 2020, pp. 11–15, doi: 10.1109/ISCTT51595.2020.00009.

[2] R. Kumar, "Fuzzy intensive decision function-based weather prediction using multi-perception neural classification for successive crop recommendation in big data analysis," in *2nd IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics, ICDCECE 2023*, Apr. 2023, pp. 1–5, doi: 10.1109/ICDCECE57866.2023.10150625.

[3] M. Tian, X. Chen, H. Zhang, P. Zhang, K. Cao, and R. Wang, "Weather classification method based on spiking neural network," in *2021 International Conference on Digital Society and Intelligent Systems, DSInS 2021*, Dec. 2021, pp. 134–137, doi: 10.1109/DSInS54396.2021.9670557.

[4] R. Meenal, K. Kailash, P. A. Michael, J. J. Joseph, F. T. Josh, and E. Rajasekaran, "Machine learning based smart weather prediction," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 28, no. 1, pp. 508–515, Oct. 2022, doi: 10.11591/ijeecs.v28.i1.pp508-515.

[5] R. Meenal, P. A. Michael, D. Pamela, and E. Rajasekaran, "Weather prediction using random forest machine learning model," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 2, pp. 1208–1215, May 2021, doi: 10.11591/ijeecs.v22.i2.pp1208-1215.

[6] N. Bogdanovs, V. Bistrovs, E. Petersons, A. Ipatovs, and R. Belinskis, "Weather prediction algorithm based on historical data using Kalman filter," in *Proceedings - 2018 Advances in Wireless and Optical Communications, RTUWO 2018*, Nov. 2018, pp. 94–99, doi: 10.1109/RTUWO.2018.8587795.

[7] Q. Fu, D. Niu, Z. Zang, J. Huang, and L. Diao, "Multi-stations' weather prediction based on hybrid model using 1D CNN and Bi-LSTM," in *Chinese Control Conference, CCC*, Jul. 2019, vol. 2019-July, pp. 3771–3775, doi: 10.23919/ChiCC.2019.8866496.

[8] G. Meti, R. K. G. Krishnegowda, and G. S. Swamy, "Rainfall analysis and prediction using ensemble learning for Karnataka State," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 32, no. 2, pp. 1187–1198, Nov. 2023, doi: 10.11591/ijeecs.v32.i2.pp1187-1198.

[9] D. K. Singh and N. Goel, "Analysing data mining techniques on bank customers for credit score," in *ICRITO 2020 - IEEE 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*, Jun. 2020, pp. 1293–1295, doi: 10.1109/ICRITO48877.2020.9197909.

[10] H. Erol, B. M. Tyoden, and R. Erol, "Classification performances of data mining clustering algorithms for remotely sensed multispectral image data," in *2018 IEEE (SMC) International Conference on Innovations in Intelligent Systems and Applications, INISTA 2018*, Jul. 2018, pp. 1–4, doi: 10.1109/INISTA.2018.8466320.

[11] X. Zhao and C. Wu, "Weather classification based on convolutional neural networks," in *Proceedings - 2021 International Conference on Wireless Communications and Smart Grid, ICWCSG 2021*, Aug. 2021, pp. 293–296, doi: 10.1109/ICWCSG53609.2021.00064.

[12] S. Goel, S. Markanday, and S. Mohanty, "Analysis of multi-class weather classification using deep learning models and machine learning classifiers," in *Proceedings - 2022 OITS International Conference on Information Technology, OCIT 2022*, Dec. 2022, pp. 223–227, doi: 10.1109/OCIT56763.2022.00050.

[13] E. Dritsas, M. Trigka, and P. Mylonas, "A multi-class classification approach for weather forecasting with machine learning techniques," in *2022 17th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2022*, Nov. 2022, pp. 1–5, doi: 10.1109/SMAP56125.2022.9942121.

[14] T. Pawar, P. Kalra, and D. Mehrotra, "Analysis of sentiments for sports data using RapidMiner," in *Proceedings of the 2nd International Conference on Green Computing and Internet of Things, ICGCIoT 2018*, Aug. 2018, pp. 625–628, doi: 10.1109/ICGCIoT.2018.8752989.

[15] A. D. Novika and A. S. Girsang, "Multi-layer perceptron hyperparameter optimization using Jaya algorithm for disease classification," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 35, no. 1, pp. 620–630, Jul. 2024, doi: 10.11591/ijeecs.v35.i1.pp620-630.

[16] M. Tamulionis and A. Serackis, "Comparison of multi-layer perceptron and cascade feed-forward neural network for head-related transfer function interpolation," in *2019 Open Conference of Electrical, Electronic and Information Sciences, eStream 2019 - Proceedings*, Apr. 2019, pp. 1–4, doi: 10.1109/eStream.2019.8732158.

[17] H. Cai and X. Wang, "Steady-state torque prediction of inverter-fed induction machine based on multi-layer perceptron neural network," in *2023 IEEE Energy Conversion Congress and Exposition, ECCE 2023*, Oct. 2023, pp. 3802–3809, doi: 10.1109/ECCE53617.2023.10362191.

[18] S. Jena, D. P. Mishra, and S. Mishra, "Detection and classification of permanent fault using multi-layer perceptron model in a distribution network," in *2023 IEEE 3rd International Conference on Smart Technologies for Power, Energy and Control, STPEC 2023*, Dec. 2023, pp. 1–6, doi: 10.1109/STPEC59253.2023.10431048.

[19] X. Xu, S. L. Keoh, C. K. Seow, Q. Cao, and S. K. Bin Abdul Rahim, "Towards prediction of bus arrival time using multi-layer perceptron (MLP) and MLP regressor," in *2023 8th International Conference on Business and Industrial Research, ICBIR 2023 - Proceedings*, May 2023, pp. 669–674, doi: 10.1109/ICBIR57571.2023.10147614.

[20] A. Sadiq and N. Yahya, "Fractional stochastic gradient descent based learning algorithm for multi-layer perceptron neural networks," in *International Conference on Intelligent and Advanced Systems: Enhance the Present for a Sustainable Future, ICIAS 2021*, Jul. 2021, pp. 1–4, doi: 10.1109/ICIAS49414.2021.9642687.

[21] H. Yang, "Transmission line fault detection based on multi-layer perceptron," in *Proceedings - 2022 International Conference on Big Data, Information and Computer Network, BDICN 2022*, Jan. 2022, pp. 778–781, doi: 10.1109/BDICN55575.2022.00151.

[22] J. Sathya and F. Mary Harin Fernandez, "Crime detection using multi-layer perceptron in social media platforms," in *12th IEEE International Conference on Advanced Computing, ICoAC 2023*, Aug. 2023, pp. 1–8, doi: 10.1109/ICoAC59537.2023.10250014.

[23] M. L. Allokendek and E. Ellisa, "Reconditioning the resilience of Palu city within the natural disaster stories," *IOP Conference Series: Earth and Environmental Science*, vol. 1082, no. 1, p. 012020, Sep. 2022, doi: 10.1088/1755-1315/1082/1/012020.

[24] Meteorological Service of Canada, "MANOBS - manual of surface weather observations," p. 475, 2015.

[25] L. Andrade-Arenas, I. Rubio-Paucar, and C. Yactayo-Arias, "Predictive models in Alzheimer's disease: an evaluation based on data mining techniques," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 3, pp. 2988–3002, Jun. 2024, doi: 10.11591/ijece.v14i3.pp2988-3002.

[26] K. Mnich, A. Polewko-Klim, A. K. Golinska, W. Lesinski, and W. R. Rudnicki, "Super learning with repeated cross validation," in *IEEE International Conference on Data Mining Workshops, ICDMW*, Nov. 2020, vol. 2020-November, pp. 629–635, doi: 10.1109/ICDMW51313.2020.00089.

BIOGRAPHIES OF AUTHORS

Muhammad Aristo Indrajaya received a bachelor's degree in electrical engineering from Tadulako University in Palu, Central Sulawesi, Indonesia, in 2013 and later earned a master's degree from the Sepuluh Nopember Institute of Technology in Surabaya, Indonesia, in 2015 at the field of Telematics. In addition to having an academic degree, he has a certificate of expertise, namely Cisco Certified Networking Associate (CCNA) Routing and Switching (RS), and can operate Cisco devices such as routers and switches. Before becoming a lecturer, he worked as an Engineer on-site at one of the telecommunications companies, namely PT. Telkom Indonesia Tbk. He is currently working as a lecturer at Tadulako University and researching telecommunications and computer networks. He can be contacted at email: aristo@untad.ac.id.

Tan Suryani Sollu is an assistant professor at the Electrical Engineering Department, at Tadulako University, Palu, Indonesia. She received his bachelor's and master's degrees from the University of Hasanuddin, Makassar, Indonesia. She leads the Electronic and MicroprocessOR Laboratorytor of the Engineering Department, at Tadulako University. She has authored papers/technical reports and holds patents and copyrights in electronic implementation and Telecommunications. Her current research is in Electronics, Microcontroller Systems, and Telecommunication. She joins in professional societies of FORTEI Indonesia. She is currently working as a lecturer at Tadulako University. She can be contacted at email: tansuryani171265@gmail.com.

Mery Subito is a senior lecturer in the Department of Electrical Engineering at Tadulako University, known for her contributions to electronics. Her higher education began at Hasanuddin University, Makassar, where she earned a Bachelor of Engineering (S1) degree in Electrical Engineering. As an academic, Mery Subito has extensive research experience in the field of electronics. She has been involved in various research projects focused on innovation and the development of electronic technologies, which have contributed to advancements in scientific knowledge and the application of technology in daily life. Her research findings have been widely published in numerous scientific journals, both nationally and internationally, showcasing her deep expertise and commitment to her field she can be contacted at email: mery.subito@untad.ac.id.

Yuli Asmi Rahman was an assistant professor of the electrical engineering department, at Tadulako University, Palu, Indonesia. She received her B.Tech (ST) and Ph.D. degrees from the University of Hasanuddin, Makassar, Indonesia. She leads the electrical engineering department, at Tadulako University. She has authored papers/technical reports and holds a patent and copyright in renewable energy implementation. Her current research is in power systems, and renewable energy integration, including energy management and optimization. Dr. Yuli Asmi Rahman joins in professional societies of FORTEI Indonesia. She is involved in the organization of conferences. She has served as editor/guest editor of several journals. She can be contacted at email: asmi_yuli@yahoo.co.id.

Erwin Ardias Saputra is a young lecturer at the Faculty of Engineering, Department of Electrical Engineering, Tadulako University. He earned his Bachelor of Engineering degree from Tadulako University and his Master of Engineering degree from the Sepuluh Nopember Institute of Technology in Surabaya. Mr. Saputra has over 4 years of professional experience as a software engineer and data scientist, followed by 3 years as an academic. He actively engages in several research projects published in national and international journals and proceedings. His research focuses on computer vision, artificial intelligence, and intelligent systems. He can be contacted at email: erwin.ardias@untad.ac.id.