JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 1, Maret 2025, Pp. 589-598

ANALYZING TEMPEARTURE ANOMALIES IN MONITORING DATA USING CONVOLUTIONAL NEURAL NETWORK

Nurhidayah Puteri*1), Widi Astuti2), Aditya Firman Ihsan3)

1. Informatics, Informatics, Telkom University, Indonesia
2. Informatics, Informatics, Telkom University, Indonesia
3. Informatics, Informatics, Telkom University, Indonesia

Article Info
Keywords: Convolutional Neural Network, Anomaly, Temperature, Classification, Machine Learning.

Article history:
Received 13 October 2024
Revised 10 November 2024
Accepted 11 December 2024
Available online 1 March 2025

DOI :
https://doi.org/10.29100/jipi.v10i1.5616

* Corresponding author.
Corresponding Author
E-mail address:
dayah.putri07@gmail.com

ABSTRACT
Temperature is a tool that shows the degree or measure of how hot or cold an object is. Incorrect temperature measurement can be fatal and cause various problems. Abnormal temperatures can prevent the temperature detection system from running optimally. Therefore, it is necessary to classify temperatures into normal and anomalous. Machine learning can be used as an alternative for temperature classification. By utilizing machine learning methods, one of which is Convolutional Neural Network. 3688569 temperature data were tested, dividing the results into 80% training data and 20% testing data. Accuracy, Precision, Recall, and F1 Score get a score of 100% and the CNN model graph is very good.

.

I. INTRODUCTION

Temperature is a measure of the average kinetic energy of the particles in a substance or system. In simpler terms, it tells us how hot or cold something is [1]. When the temperature of a substance increases, its particles move faster on average, while a decrease in temperature results in slower particle movement [2]. Temperature is typically measured in degrees Celsius (°C), Kelvin (K), or Fahrenheit (°F). It’s a fundamental concept in physics, chemistry, engineering, meteorology, and many other fields, influencing various aspects of our daily lives [3].

Temperature monitoring is crucial because temperature affects various processes and systems in daily life, as well as in many industries and technical applications [4]. Temperature influences human health and safety, the efficiency and quality of production processes, the security of electronic devices, plant growth and ecosystem balance, as well as transportation and storage product safety. By monitoring temperature, we can prevent dangers related to extreme temperature, maintain product quality, manage agriculture more effectively, secure electronic devices, and ensure environmental sustainability and product safety during transportation and storage [5].

In the oil and gas industry, monitoring temperature fluctuation is crucial for ensuring operational efficiency, safety, and the longevity of equipment [6]. Temperature anomalies, if left undetected, can lead to equipment malfunction, production disruption, and safety hazard [7].

Data anomaly in temperature refers to values that are unusual or unexpected within the observed temperature dataset [8]. This could include temperature significantly higher or lower than surrounding values within a specific time frame or location [9].

Detecting temperature anomalies is important because they can indicate significant weather changes, natural events such as storms or wildfires, or even errors in data measurement or recording [10]. Information about temperature anomalies can be helpful in climate monitoring, weather forecasting, natural disaster risk management, and climate change policy [11].

Machine learning techniques offer a promising avenue for temperature classification, with Convolutional Neural Network (CNN) emerging as a particularly effective method. CNN excels at capturing spatial patterns in multidimensional data, making them well-suited for anomaly detection tasks [12].

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 1, Maret 2025, Pp. 589-598

The use of deep learning techniques, especially Convolutional Neural Network (CNN) and transfer learning, for anomaly detection in time series data has been successfully used with satisfactory results, and further research could involve comparing performance with benchmark algorithms as well as exploring factors that influence transfer learning performance [13].

CNNs excel at extracting complex and deep spatial features from multidimensional data. This capability enables CNNs to detect unusual patterns or anomalies that might not be evident with traditional anomaly detection methods [14]. Pooling layers help reduce the dimensionality of data while preserving important features, enhancing the efficiency of anomaly detection by reducing noise and emphasizing anomalous signals [15]. Common data augmentation techniques used with CNNs help the model learn to detect anomalies under diverse conditions. This includes rotating, scaling, and translating images, making the model more robust to variations in data [16].

Developing an air temperature prediction model in Padang City, West Sumatera, using Convolutional Neural Network (CNN), Multilayer Perceptron (MLP), and a hybrid of the two models based on monthly air temperature data from January 2015 to December 2017. The research results show that the CNN model provides the most satisfactory results with an R2 value of 0.9965 which shows that the CNN model is best used in predicting air temperature [17].

Soil temperature predictions get accurate results using the Convolutional Neural Network (CNN) model. The data used is hourly climate data to train and test the CNN model, with data normalization to reduce the model's sensitivity to the scale of climate features. The research results show that the one-dimensional CNN model is better at predicting soil temperature than the multilayer perceptron model, especially in normal and hot weather conditions, and is able to predict daily maximum soil temperature [18].

Convolutional Neural Network (CNN) is used to detect mask use and body temperature. The system will close the portal automatically if someone is not wearing a mask or has a body temperature above 37.5°C. Test results show mask detection accuracy of 94% with a computing time of 9.09 seconds. The system also achieved 100% accuracy in integrating mask detection and infrared temperature sensors, marking an important step in preventing the spread of COVID-19 [19].

In diabetics, foot ulcers are a serious complication, however, they can be detected early through thermogram images that measure temperature differences in the foot area. Using four pre-trained CNN models accelerates the process of adapting this technology in clinical settings, providing a solid foundation for further implementation in medical practice [20].

II. Research Methods

The research stages regarding the analysis of temperature anomaly data using the CNN method are as follows:

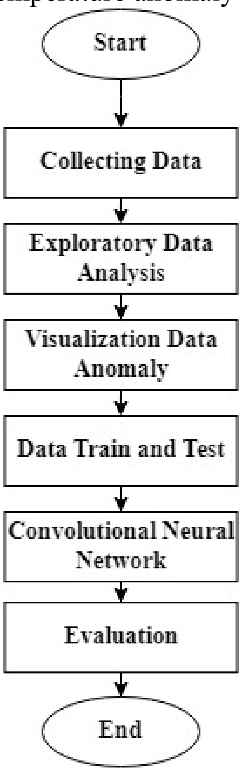

Fig. 1. Flow Diagram

At the data collecting stage, the data collected is date stamp, asset id, and temperature, totaling 3688569 data. Data was taken every minute from January 2020 to December 2020 the data source was taken from Kaggle with the data criteria taken being oil and gas temperature. The data was then processed using Google Collaboration using the Python programming language. The detailed data for each month and the total data can be seen in the following table.

TABLE I
NUMBER OF DATASETS

Month	Total
January	312473
February	292301
March	312463
April	302393
May	312466
June	302393
July	312011
August	312466
September	302393
Oktober	312432
November	302305
December	312373

Exploratory data analysis is an initial exploratory process that aims to identify patterns, find anomalies data, test hypotheses and validate assumptions. By running Exploratory Data Analysis (EDA), users can be greatly helped in detection errors, identifying outliers, finding relationships between data, and exploring important elements from scratch. The EDA process performed is:

1. View the amount of data per month
2. Combine all months into one data frame
3. Converting inappropriate data types
4. Identifying for missing values
5. Identifying for anomalous data in the temperature column

At this stage, data visualization is carried out to see the distribution of anomalous data in the dataset. The data visualization is divided into 3 categories, namely: every month, based on asset ID, and based on temperature.

The aim of dividing training data and testing data is to divide the dataset so that later the divided dataset can be processed into the CNN classification model. The distribution of the dataset is 80% training data totaling 5874979 data and 20% testing data totaling 1468745 data. Data sharing is based on research conducted by previous research [21] which obtained an accuracy of up to 97.3%, therefore a sampling size of 80% train and 20% test was chosen in this study. The output of the convolutional will be passed through an activation function using ReLU activation function.

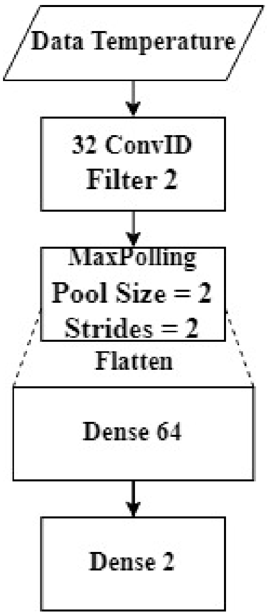

Fig. 2. Flow CNN Methods

The type of pooling used by CNN is max pooling. The pooling takes the maximum value in each window/block. The details of the max pooling used are pool_size=2, strides=2, and padding='same'.
When using CNN, there are 4 important hyperparameters that we need to determine, namely:

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 1, Maret 2025, Pp. 589-598

1. Kernel/filter size, which is the length and width of the filter to be used.
2. Number of filters, namely how many filters we will use.
3. Stride, namely how far the filter is when it is shifted
4. Padding

This architecture is used for classification tasks on one-dimensional sequence data. Conv1D layers are used to extract local features from sequences, MaxPooling1D is used to reduce dimensions and focus on important features, and Dense layers are used to make a final decision about the input class based on the extracted features.

Evaluation uses a confusion matrix, in the evaluation you will see accuracy, f1 score, precision, and recall which are used as material for analysis and conclusion. The confusion matrix helps understand how well the model can distinguish between normal and anomaly classes. Accuracy provides a general overview of the overall model performance. Precision is crucial to ensure that when the model detects anomalies, its detections are indeed accurate. Recall is important to ensure that the model does not miss too many cases of anomalies that should be detected. F1 score is useful for measuring the overall performance of the model in detecting temperature anomalies, considering both precision and recall. By understanding and using these metrics, we can measure and analyze how effective the model is in detecting and classifying temperature anomalies, as well as identify areas that need improvement to enhance model performance.

III. Results And Discussion

At the data collection stage, the data collected is date stamp, assed id, and temperature, total 3688569 data. Data was taken every minute from January 2020 to December 2020

Exploratory Data Analysis in an initial exploratory process that aims to identify patterns, find anomalies, test hypotheses, and validate assumptions. By running EDA, users can greatly help in detecting errors, identifying outliers, finding relationships between data, and exploring important elements from the start. The amount of data for each month can be seen in the following image:

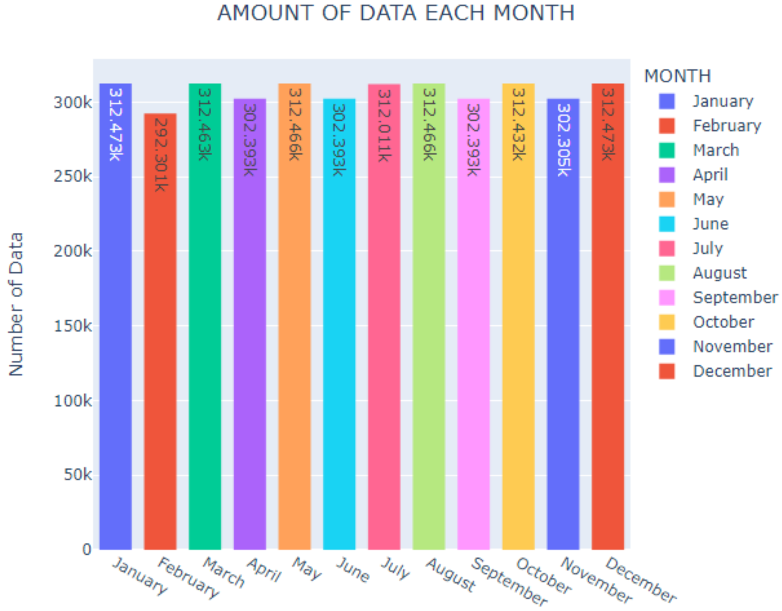

Fig. 3. Amount of Data Each Month

To make it easier to preprocess the data, all tables are combined into one data frame. After that, delete all columns that are not needed so that the data is easier to read and process. At this stage the columns taken are 3 columns, namely DATE_STAMP, ASSET_ID, and TEMPERATURE. Next, check the data type for each column. Inappropriate data types will be changed to match the column value.

In the DATE_STAMP column, the data type is still an object or string, while the data type should be datetime, therefore it must be changed to datetime data type. In the TEMPERATURE column, the data type is still a float where the value is a decimal. To make data processing easier, we change the temperature data type to integer so that the value is a whole number. Next step identification for missing values. At this stage, missing values are checked. The result was that no missing values were detected in each column. Next step identification for anomalous data in the temperature column. At this stage, anomalous data is checked on the available dataset. The results are that 99.5% of the data is normal and 0.5% of data is anomalous.

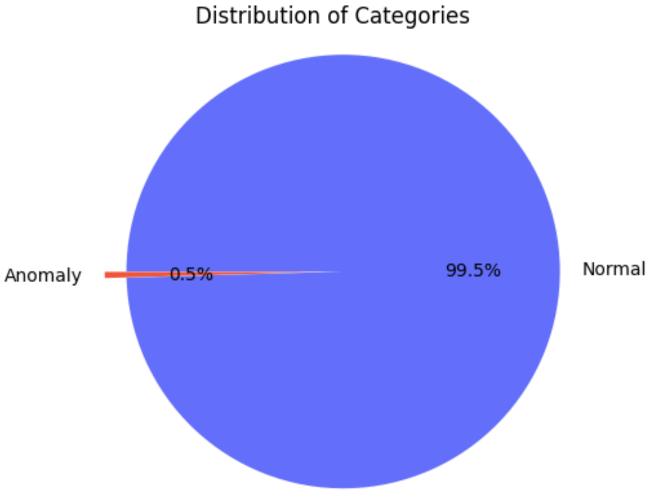

Fig. 4. Distribution of Categories

At this stage, data visualization is carried out to see the distribution of anomalous data in the dataset. Data visualization is divided into 3 categories, namely:

1. Visualization of status data every month
   Based on the data visualization results, only January and September had no anomalous data, while the other months had anomalous data, this is because there are various types of gas stored in each room which are represented by asset_id. Under certain circumstances, substances stored in each asset_id cause temperature changes and significant temperature changes occur in months other than January and September. The amount of anomalous and normal data each month can be seen as follows.

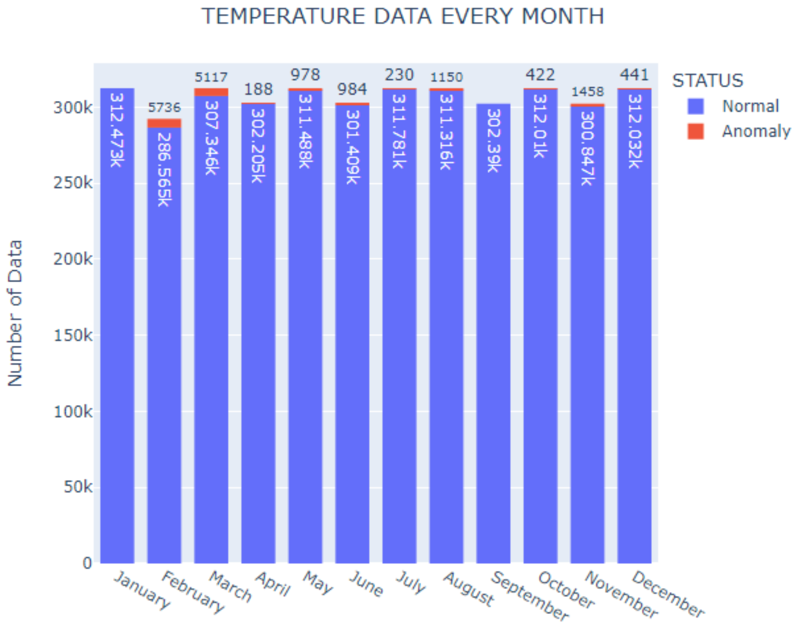

Fig. 5. Temperature Data Every Month

2. Visualization of label data based on asset ID
   Based on the data visualization results, only asset ID 133002 does not have anomaly data, while the other asset IDs have anomaly data. The asset ID is the ID of the oil and gas storage room. In other asset IDs, temperature changes tend to be significant, this is caused by the type of gas stored in the asset ID causing extreme temperature changes. The size of the anomalous and normal data base on the asset ID can be seen as follows.

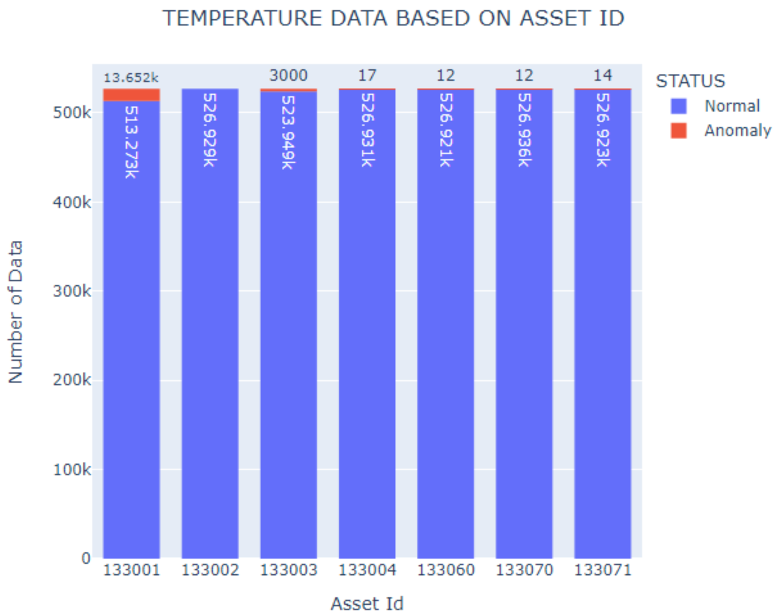

Fig. 6. Temperature Data Based on Asset ID

3. Visualization of label data based on temperature
   Based on the visualization results, there are 414 temperatures that fall into the normal category, while 91 temperatures fall into the anomaly category. The temperature collection here is an interval of temperatures that are categorized as normal and anomalous, not the total temperature data in the dataset. The temperature interval that is categorized as normal is above 31 degrees Fahrenheit and below 212 degrees Fahrenheit. Apart from the interval above, it is said that this temperature is included in the anomalous temperature interval.

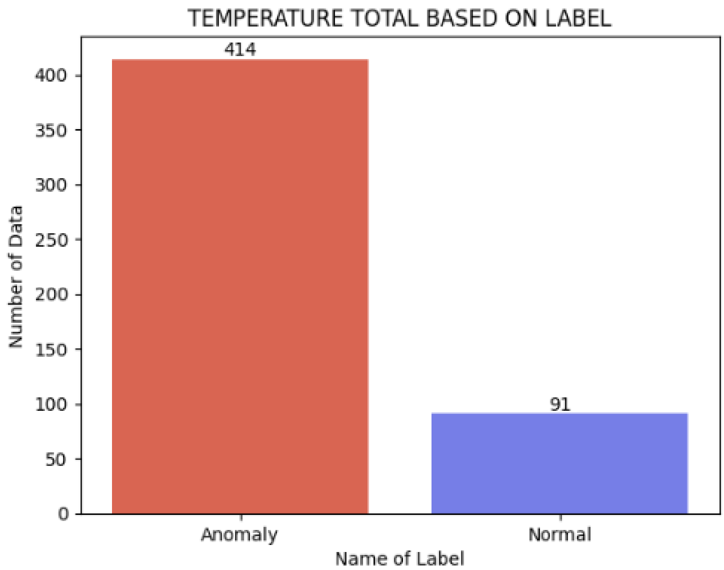

Fig. 7. Temperature Total Based on Label

The top 5 temperatures that fall into the normal category based on the largest amount of the data are as follows.

TABLE II
TOP 5 TEMPERATURE NORMAL CATEGORIES

Temperature	Status	Normal
84	Normal	240130
83	Normal	189295
82	Normal	157720
81	Normal	157613
78	Normal	155040

The top 5 temperatures included in the anomaly category based on the largest amounts of data are as follows.

TABLE III
TOP 5 TEMPERATURE ANOMALY CATEGORIES

Temperature	Status	Total
0	Anomaly	14887
1	Anomaly	35
248	Anomaly	14
63424	Anomaly	10
63456	Anomaly	9

At this stage, the number of labels is balanced first so that the data will be trained and tested later has no data imbalances, unbalanced data affects the model accuracy results. The data imbalance technique used is random oversampling, the initial number of anomaly label data was 16707 while the normal label was 3671862. After balancing the data using random over sampling the result was that the number of normal and anomaly label data each amounted to 3671862. Random oversampling is a technique in data processing used to address class imbalance in datasets, especially when dealing with anomaly data. In the case of temperature anomaly data, the minority class represents the anomaly temperature data, while the majority class represents the normal temperature data. The benefits of random oversampling in anomaly data detection are that it addresses class imbalance, allows the model to learn better from the anomaly data, and improves the model's performance in detecting anomalies. However, the drawbacks of random oversampling in anomaly data include the risk of overfitting and longer processing times.

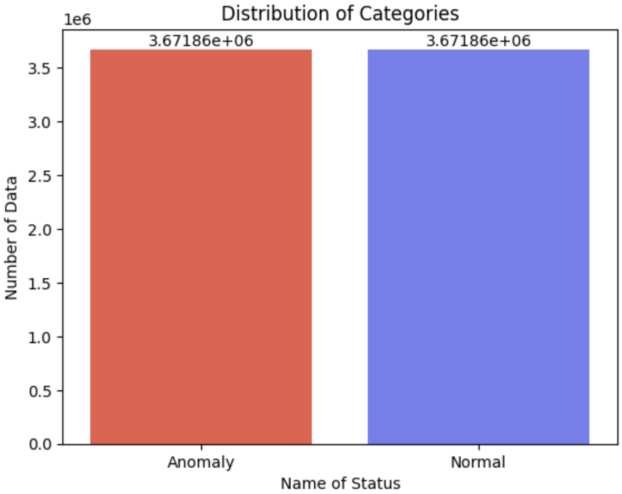

Fig. 8. Random Oversample Label

Before entering into the distribution of training and testing data, the temperature column data must first be reshaped. The reshape function is used to create or change column rows. After that, change the multilabel column label to multiclass with the command pd.get_dummies. Divides training and testing data. The data division is 80% training data totaling 5874979 data and 20% testing data totaling 1468745 data.

Creating of CNN models using Keras. The model created is a 1-dimensional CNN. Using ReLU and sigmoid activation. Because the multiclass optimizer uses adam and loss uses binary_crossentropy.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv1d (Conv1D)             (None, 1, 32)             96

 max_pooling1d (MaxPooling1  (None, 1, 32)             0
 D)

 flatten (Flatten)           (None, 32)                0

 dense (Dense)               (None, 64)                2112

 dense_1 (Dense)             (None, 2)                 130

=================================================================
Total params: 2338 (9.13 KB)
Trainable params: 2338 (9.13 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Fig. 9. Model CNN

Train loss and val loss graphs are very useful tools for understanding the performance and training dynamics of a CNN model. They help identify overfitting and underfitting, as well as determine when training should be stopped

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 1, Maret 2025, Pp. 589-598

to achieve the best performance on unseen data. The model was trained using 3 epochs with a batch_size of 255. The results were that the accuracy in the last epoch reached 99% and the loss was 9.0392e-04. Based on the results of the accuracy and loss graphs, the model created has a good graph, meaning the model does not experience under fitting or over fitting.

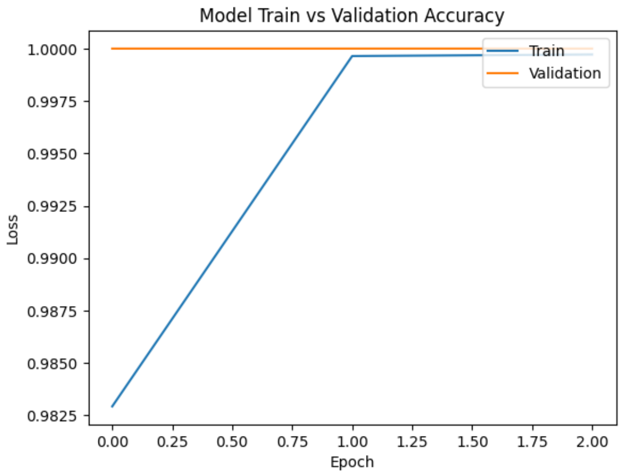

Fig. 10. Accuracy

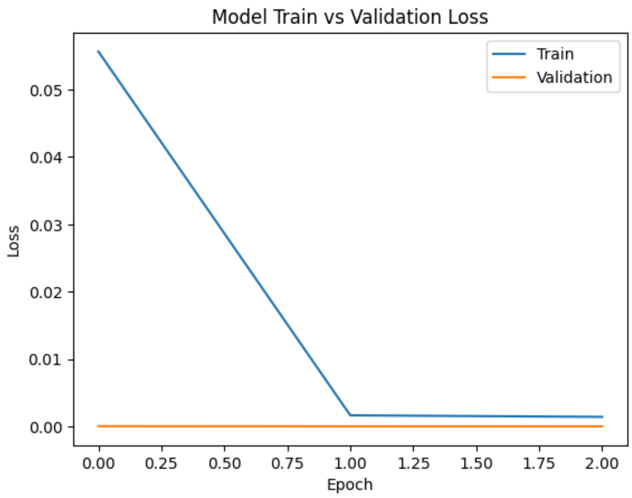

Fig. 11. Loss

Final stage is evaluated using a confusing matrix. The results for the normal label get 100% accuracy, 100% precision, 100% f1-score, and 100% recall. The results of the classification report can be seen in the following image.

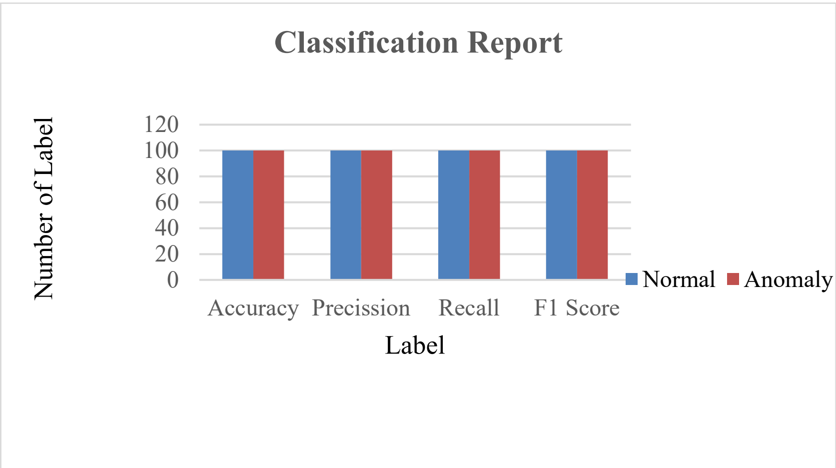

Fig. 12. Classification Report

In the prediction and actual matrix, the data results are categorized into the original data according to the CNNN classification results. The following in the confusion matrix in this research.

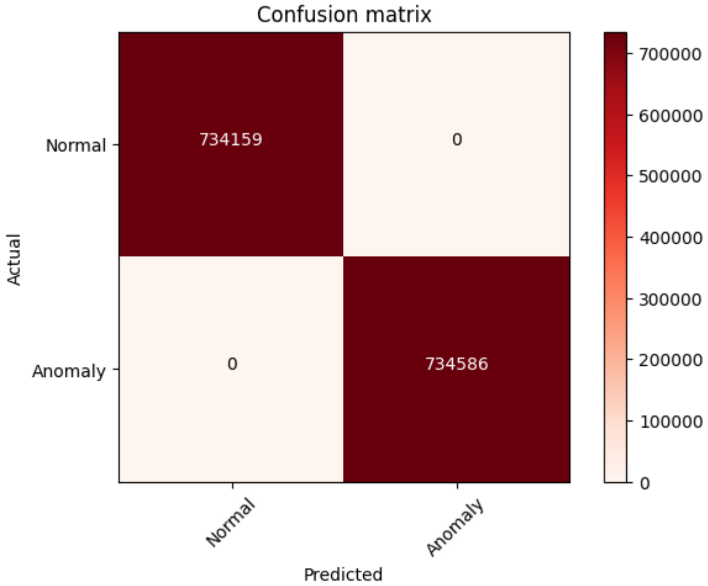

Fig. 13. Confusion Matrix

The use of CNN in data anomaly detection achieves high accuracy. This is proven by previous research which discusses human face recognition in video conference applications using the CNN pipeline method, saying that CNN succeeded in classifying and obtaining high accuracy [22]. However, this is not in line with research on real-time handwritten character recognition using a Convolutional Neural Network architecture. In this study, the results stated that the accuracy obtained was not high enough [23].

The results obtained by this research are in the form of a CNN model architecture implemented using the Keras library. The architecture obtained uses 4 layers (32, 32, 32, 64), the optimizer uses Adam, 3 epochs, and batch size 225. Results in 100% accuracy and 0% loss. The accuracy and loss graphs can be seen in figure 9 and 10. The accuracy is very high and the accuracy model graph is classified as good, showing neither underfitting nor overfitting. The results of this research are slightly different from research conducted by Gunawan who used 80:20 train and test data with an accuracy of 97.3, in this study the accuracy was higher, namely 100%, this is due to differences in the CNN architecture and the data used. So, the accuracy results are different.

IV. Conclusion

Based on the results and discussion of data anomaly detection in temperature using the Convolutional Neural Network method, the results were 100% accuracy with precision, recall, and f1 score getting a value of 100%. This means that Convolutional Neural Networks can be used as a method for detecting data anomalies. If seen from results of the CNN model graph that has been created, it shows that the CNN model is very good.

However, the model that has been created has not been tested with samples other than the available dataset samples, therefore new, relevant data is needed to test the accuracy of the model that has been created.

References

[1] G. Gethin, J. D. Ivory, D. Sezgin, H. Muller, G. M. O'Connor and A. Vellinga, "What is the 'Normal' Wound Bed Temperature? A Scoping Review and New Hypothesis," *Wound Repair and Regeneration,* vol. 29, no. 5, pp. 843-847, September 2021.

[2] W. Kurniasih, "Pengertian Suhu: Rumus, Faktor, Alat Ukur dan Skala," 2021. [Online]. Available: https://www.gramedia.com/literasi/pengertian-suhu/.

[3] R. P. Rasyid, M. Todingan and M. Sirman, "Analisis Hubungan Massa Jenis Dan Suhu Terhadap Kalkukasi," *Andromeda,* vol. 6, no. 1, 2022.

[4] D. Y. Pimenov, A. Bustillo, S. Wojciechowski, V. S. Sharma, M. K. Gupta and M. Kuntoğlu, "Artificial Intelligence Systems for Tool Condition Monitoring in Machining: Analysis and Critical Review," *Journal of Intelligent Manufacturing,* vol. 34, no. 5, 2023.

[5] Y. Bai, Y. Duan, M. Qian, S. Zhu, X. Ma, Y. Yang and H. Gao, "Temperature Dependence of Electrical Properties and Conduction Mechanism of SiNx-Based Resistive Random Access Memory," *Physica Status Solidi A: Applications and Materials Science,* vol. 220, no. 21, 2023.

[6] B. M. Erland, . A. K. Thorpe and J. A. Gamon, "Recent Advances Toward Transparent Methane Emissions Monitoring: A Review," *Environmental Science & Technology,* vol. 56, no. 23, 2022.

[7] J. Jhony, S. Amos and R. Prabhu, "Optical Fibre-Based Densors for Oil and Gas Applications," *Sensors,* vol. 21, no. 18, 2021.

[8] H. Chen, "Data Anomaly Diagnosis Method of Temperature Sensor Based on Deep Neural Network," *Mobile Information Systems,* 2022.

[9] C. Liu, X. Su and C. Li, "Edge Computing for Data Anomaly Detection of Multi-Sensors in Underground Mining," *Electronics,* vol. 10, no. 3, 2021.

[10] A. T. M. Fisch, I. A. Eckley and P. Fearnhead, "A Linear Time Method For The Detection of Collective and Point Anomalies," *Statistical Analysis and Data Mining,* vol. 15, no. 4, 2022.

[11] V. Mahato, M. Obeidi, D. Brabazon and P. Cunningham, "Detecting Voids in 3D Printing Using Melt Pool Time Series Data," *Journal of Intelligent Manufacturing,* vol. 33, no. 3, 2022.

[12] S. Maiti and R. K. Chiluvuru, "A Deep CNN-LSTM Model for Predicting Interface Depth From Gravity Data Over Thrust and Fold Belts of North East India," *Journal of Asian Earth Sciences,* vol. 259, 2024.

[13] K. Tzoumpas, A. Estrada, P. Miraglio and P. Zambelli, "A Data Filling Methodology for Time Series Based on CNN and (Bi)LSTM Neural Networks," *IEEE Access,* vol. 12, 2024.

[14] J. Yopento, E. Ernawati and F. F. Coastera, "Identifikasi Pneumonia Pada Citra X-Ray Paru-Paru Menggunakan Metode Convolutional Neural Network (CNN) Berdasarkan Ekstraksi Fitur Sobel," vol. 10, no. 1, 2022.

[15] E. A. Mohamed, T. Gaber, O. Karam and E. A. Rashed, "A Novel CNN pooling layer for breast cancer segmentation and classification from thermograms," vol. 17, no. 10, 2022.

[16] T. B. Sasongko, H. Haryoko and A. Amrullah, "Analisis Efek Augmentasi Dataset dan Fine Tune pada Algoritma Pre-Trained Convolutional Neural Network (CNN)," vol. 10, no. 4, 2023.

[17] I. Kurniawan, L. S. Silaban and D. Munandar, "Implementation of Convolutional Neural Network and Multilayer Perceptron in Predicting Air Temperature in Padang," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi),* vol. 4, no. 6, pp. 1165-1170, December 2020.

[18] V. Farhangmehr, J. H. Cobo, A. Mohammadian, P. Payeur, H. Shirkhani and H. Imanian, "A Convolutional Neural Network Model for Soil Temperature Prediction under Ordinary and Hot Weather Conditions: Comparison with a Multilayer Perceptron Model," *Sustainability,* vol. 15, no. 10, 2023.

[19] I. A. Rachimi and F. Utaminingrum, "Deteksi Masker dan Suhu Tubuh untuk Kendali Portal Otomatis Menggunakan CNN sebagai Pencegahan Penularan SARS-CoV-2," vol. 5, no. 8, pp. 3472-3477, 2021.

[20] W. M. A. A. Pradhana, G. A. Pradipta and R. R. Huizen, "Analisis Performa Pre-Trained Model CNN pada Citra Termogram dalam Deteksi Dini Pasien Diabetes," 2024.

[21] B. Gunawan and M. E. Al Rivan, "Klasifikasi Jenis Beras Putih menggunakan CNN Residual Network Optimizer SGD," vol. 2, no. 1, 2023.

[22] E. Tanuwijaya, R. L. Lordianto and R. A. Jasin, "Recognition Of Human Faces In Video Conference Applications Using The CNN Pipeline," *Jurnal Teknik Informatika,* vol. 3, no. 2, pp. 421-427, April 2022.

[23] M. S. Gumilang and D. Avianto, "Recognition of Real-Time Handwritten Characters Using Convolutional Neural Network Architecture," *Jurnal Teknik Informatika,* vol. 4, no. 5, pp. 1143-1150, October 2023.