Application of Ensemble Learning with Mean Shift Clustering for Output Profile Classification and Anomaly Detection in Energy Production of Grid-Tied Photovoltaic System

***Abstract*— Fault detection and monitoring system in photovoltaic (PV) energy management system is important in achieving its optimal performance. An effective diagnostic system involves correct analysis of electrical parameters of a PV array on a given weather condition. In the study, mean-shift clustering was applied for pre-classification and anomaly detection of time-series data of electrical parameters from grid-tied inverter, and solar-irradiance. Classification and anomaly detection applied is based in ensemble learning, where its base learners are based from multilayer perceptron. A stacking ensemble is used in classification of energy production profile while bagging ensemble is used detecting anomalous trend in time-series data. A stacking ensemble got a highest accuracy value of 94% compared to single classifiers which have accuracy value of 85.25%, 84.14%, and 63.4%, respectively. The bagging ensemble autoencoders have the lowest mean squared error during model reconstruction compared to single autoencoder. It has a fair performance in classifying anomaly points from normal datapoints, having an AUC value of 0.795 and F1-score of 0.71, given that the hyperparameter is 0.5. Overall, ensemble learners improve the performance in classification and detection tasks.**

***Keywords*—anomaly detection, energy production classification, energy management, ensemble learning, multi-layer perceptron**

I. Introduction

Fault detection and monitoring system in photovoltaic (PV) system is important because it ensure its optimal performance in terms of efficiency, reliability, and safety. The ideal scenario in an energy management system is to maximize energy production, minimize energy losses, and ensure safety in the facility. However, subjection of PV system to faulty conditions is unavoidable. Fault conditions usually arise from factors such as string faults, hot-spotting, array degradation, low solar irradiance, and partial shading [1][2]. An effective fault diagnosis in PV system involves correct analysis of electric parameters such as current, voltage, and energy given different weather patterns[3]. Electric parameters of a PV system can be fluctuating due to erratic nature of weather conditions. There are scenarios where a normal condition can closely resemble a faulty condition on a certain weather season and can result to false anomaly detection. Other challenges include the presence of noise in electrical parameters time-series data which lead to difficulty in distinguishing anomalies from normal points.

Researches involving data-driven approach in monitoring PV system status and detecting anomalous trend in electrical parameter time-series data is widespread. Machine learning algorithms and other unsupervised learning models such as neural networks are applied for PV monitoring and fault diagnostics. However, issues in single model such as underfitting and overfitting problems, affects the overall performance in anomaly detection and classification tasks [4][5]. An algorithm that utilizes the combination of predictions made from models is needed for improvement in performance. Linear combination of methods for time series data are likely to produce unbiased and more accurate results compared to single learners.

In this paper, mean shift clustering algorithm was applied to detect unique profiles or characteristics of energy production of a grid-tied PV system [6][7]. The said clustering algorithm is used for ensemble-based pre-classification and anomaly detection. A stacking-based ensemble algorithm is applied for classification of clusters predicted. A bagging-based ensemble autoencoder based on multilayer perceptron is applied for detecting anomalous trend in the dataset. A metric for accuracy is used to evaluate the performance of ensemble models in classification and anomaly detection.

II. Energy Production Classification

A. Multilayer Perceptron as Classifiers

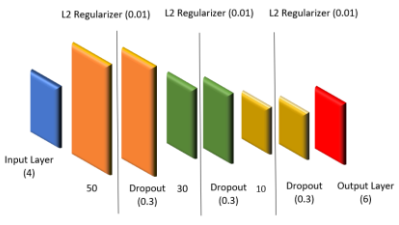

Fig. 1. Plot of Validation Loss and Accuracy for Base Classifiers

Figure 1 shows the architecture of neural network model used as base learner in stacking-based ensemble classifiers. The base learner is based from multilayer perceptron, which is a neural network with multiple hidden layers. The input layer has 4 nodes, which corresponds to the number of columns of input sequence. The model has 3 layers, where each have 50, 30, and 10 nodes, respectively. Each layer has dropout layer with dropout rate of 0.3 and L2 regularizers with weight decay of 0.01 to avoid overfitting. The output layer has 6 nodes which corresponds to the number of predicted clusters made by mean-shift clustering.

B. *Stacking Based Ensemble Learning*

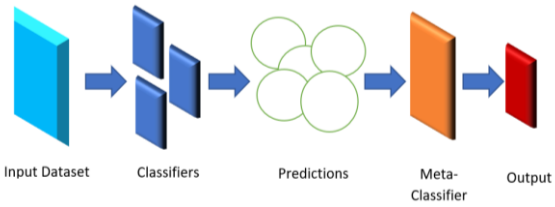

Fig. 2. Stacking-based Ensemble Learning

Figure 2 shows the principle behind stacking-based ensemble learning. Stacking is an ensemble learning algorithm where a meta-classifier combines the predictions made by well-performing base classifiers [8][9]. There are 20 base classifiers trained for stacking ensemble and is saved as hierarchical data format (.h5 files). These files are fed to a meta-classifier defined by a neural network with 10 nodes on its hidden layer, and 6 nodes on its output layer which pertains to the number of predicted clusters. The meta-classifier takes the predictions made by base classifiers as features.

III. Anomaly Detection Model

A. *Multilayer Perceptron as Autoencoders*

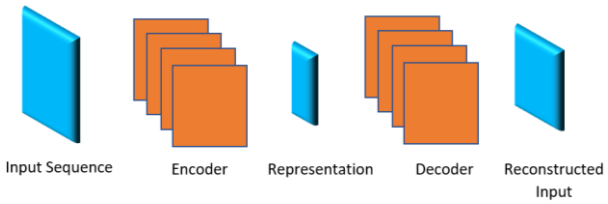

Fig. 3. Autoencoders

Figure 3 shows the principle behind autoencoders. Autoencoders are neural networks which learns from compressed version of the input sequence. These unsupervised learning algorithms is done for dimensional reduction by training the neural network in noise-less input signals and attempt to reconstruct it. Autoencoders are commonly used for dimension reduction, denoising, and outlier detection.

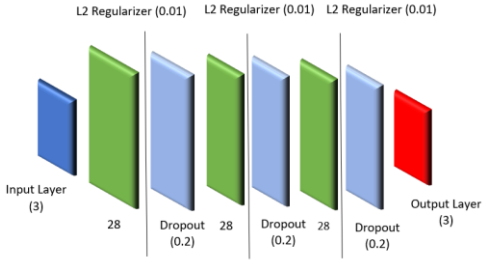

Fig. 4. Plot of Validation Loss and Accuracy for Base Classifiers

Figure 4 shows the architecture of a single autoencoder. The input layer has 3 nodes, which corresponds to the electrical parameter gathered from the grid-tied inverter. The model has 3 layers, where each have 28 nodes. Each layer has dropout layer with dropout rate of 0.3 and L2 regularizers with weight decay of 0.01 to avoid overfitting. The output layer is the same as input layer since it is meant for reconstructing the input sequence.

B. *Bagging Based Ensemble Learning*

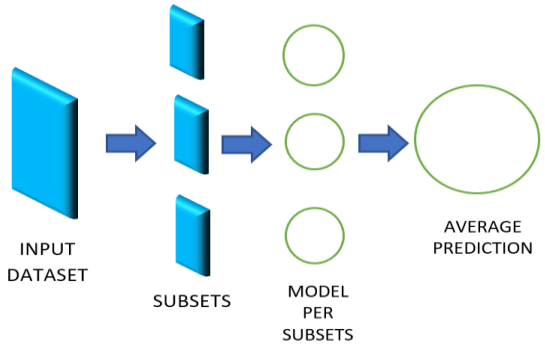

Fig. 5. Bagging-based Ensemble Learning

"Bagging" or bootstrap aggregation is an ensemble learning algorithm where it combines all predictions made by its base learners [10]. It reduces the variance error of the prediction by taking the average of the estimates from the base learners. The subsets of the data are used to train neural networks and compute the ensemble expressed mathematically as

$$f(x) = \frac{1}{M}\sum_{1}^{M} f_m(x) \qquad (1)$$

It is noted that the learners for each random subset are independent to one another and works in parallel. Figure 5 shows the illustration of how bagging-based ensemble learning works. In the study, there are 20 base autoencoders for bagging-based ensemble learning.

IV. Methodology

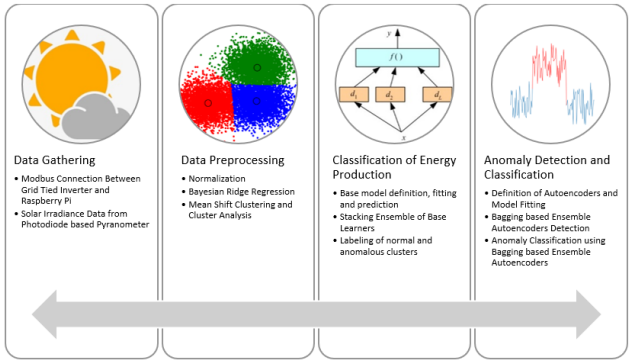

Fig. 6. Methodology of the Study

Figure 6 shows the process involved in conducting the methodology of the study. First process involves data gathering of electrical parameters from the grid-tied inverter and solar irradiance measured by photodiode based pyranometer. Data gathered are saved in comma separated values (.csv) format. Missing values found in .csv file is handled by applying imputation based on regression values on rows, specifically applying Bayesian Ridge Regression in the dataset. Dataset is then normalized in order to

improve the clustering performance, as well as to achieve faster convergence during model training.

To check for unique profile of the energy production dataset, mean shift clustering is applied. The predicted cluster is saved to a column in a dataset later to be used for classification of energy problem and detecting of anomalous trend from normal datapoints. A stacking ensemble was developed for classifying the clusters.

In anomaly detection, dataset is split into training and testing set. Each set consists of combination of normal and anomalous condition. A multilayer perceptron is then defined for classifying PV energy production and checking for anomalous trend in dataset. A bagging-based ensemble learner is used for model reconstruction during anomaly detection. To assess the performance of ensemble learners in detecting anomalous, classification metrics is applied.

V. Data Gathering And Dataset Exploration

A. Grid-Tied Photovoltaic System Setup

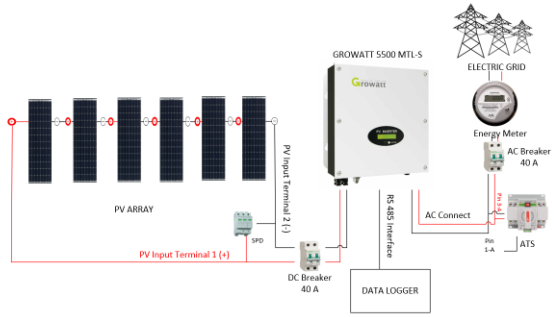

Fig. 7. Grid Tied Photovoltaic Setup with Modbus Connection

Figure 7 shows the experimental setup of the study. The figure shows the detailed connection of grid-tied photovoltaic system. It consists of 6 monocrystalline solar panels, each rated of 300 W. The solar array is connected to a grid-tied inverter, which is responsible for converting the direct current (DC) electricity generated by solar arrays into alternating current (AC) electricity. AC electricity is then fed to electrical grid for the purpose of selling it to local electricity utility provider. The data from the inverter is accessed by establishing a Modbus connection and is connected to a Raspberry Pi. RS-232 cable was used as a medium for transferring the data [11]. Data for solar irradiance was gathered by using a photodiode based pyranometer [12][13].

B. Dataset Description

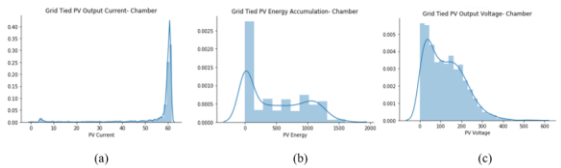

Fig. 8. Histogram of Columns in Dataset (a) Output Current (b) Output Voltage (c) Output Energy

Figure 8 shows the data distribution of electrical parameters gathered from the grid-tied inverter and Table 1 shows its statistical summary. The data is gathered by accessing the specific registers of the inverter. The list of accessible registers of the inverter is found at the datasheet.

From the plot of the data distribution, the data for output current shows a longer tail on the left side of distribution, as indicated by its negative skewness value. It also has the smallest value of standard deviation, implying that it has smallest range in the dataset. The data for energy production shows a negative kurtosis, implying that the outliers found in its dataset is less extreme compared to the expected outliers derived from normal distribution. Lastly, the data of output voltage shows positive skewness and kurtosis, meaning that the distribution is right tailed and has few extremes of outliers.

TABLE I. Statistical Description of The Parameters

Parameter	***PV Voltage***	***PV Current***	***PV Energy***
Mean	122.5143 V	56.1073 A	476.182 W
Standard Deviation	90.4922	12.2955	467. 1023
Skewness	0.7276	-3.3208	0.4702
Kurtosis	0.2986	10.2858	-1.2077

VI. Cluster Analysis

Figure 9a shows the plot of squared error of K-means clustering model against the varying number of *k* number of clusters. The plot can be utilized to find the optimal number of clusters to partition the given input sequence, which is indicated by "elbow". In the plot, it is recommended to use 2 clusters for K-means.

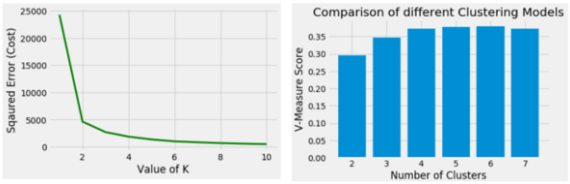

Fig. 9. (a) Plot of Squared Error for K-means to Determine the Optimal Number of Clusters. (b) Plot of V- Measure Score for Performance Comparison of K-means Clustering Model

Figure 9b shows the histogram of V-measure score with the *k* number of clusters on the x-axis. V-measure evaluates the performance of the algorithm in partitioning the given input sequence. In the plot, having 2 clusters to partition the input sequence records the lowest V-measure score. To achieve high V-measure score, it requires 6 clusters. One of the downsides of K-means algorithm is its sensitivity in outliers. The given dataset is not normally distributed and classifying it requires another clustering algorithm.

Mean shift clustering is a non-parametric, unsupervised clustering algorithm based on kernel density estimation (KDE) [14][15]. The algorithm works by assigning a weighting function on each data points. The summation of all assigned weighting function generates a density function. The resultant density function varies depending on bandwidth or the size scale of weighting function. One of the advantages of mean shift clustering is robust to outlier and does not require the user to specify a hyperparameter to partition the given input sequence[16][17]. The number of clusters is determined by the algorithm itself, considering the given input sequence.

The kernel density estimator is expressed as

$$f_k(\mathrm{u}) = \frac{1}{nh^d}\sum_{i=1}^{n} K\frac{\mathrm{u}-\mathrm{u}_i}{h} \qquad (2)$$

The kernel density estimator needs to satisfy the following conditions; the estimates should be normalized, and there should be a symmetry on a given space. These conditions are expressed mathematically as

$$\int \mathrm{K(u)}du = 1 \quad (3)$$

$$K(\mathrm{u}) = K(|\mathrm{u}|) \quad (4)$$

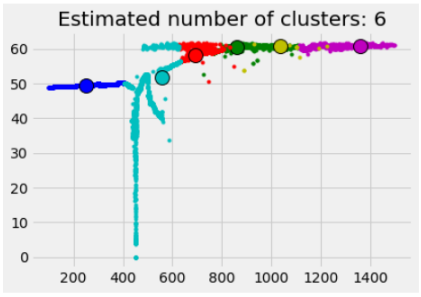

Fig. 10. Clustering of Dataset using Mean Shift Clustering

Figure 10 shows the clusters predicted by mean shift clustering on a given dataset. The algorithm predicted that there are 6 unique clusters in the dataset.

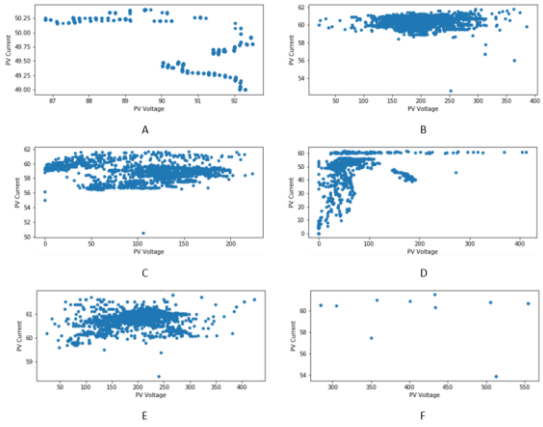

Fig. 11. Scatter plot of I-V Curves derived from Mean Shift Clustering

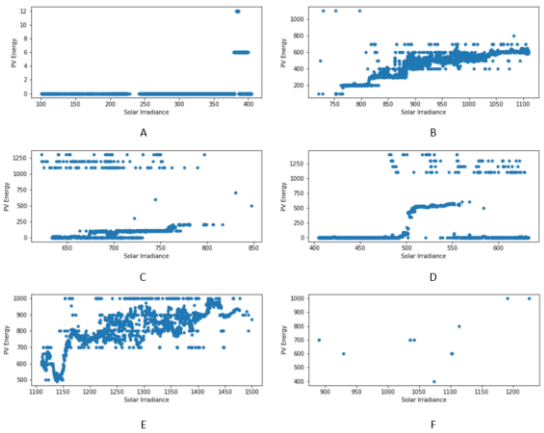

Fig. 12. Scatter plot of Energy and Solar Irradiance derived from Mean Shift Clustering

Table 2 shows the explanation of the clusters derived from mean shift clustering. The clustering algorithm predicted that there are 6 unique profile in the energy production of grid-tied PV system. Figure 11 shows the scatter plot of the I-V curve and figure 12 shows the scatter plot of energy and solar irradiance. Status E is the optimal profile of the PV array during energy production since it is during peak sunlight hours and has highest number of energy accumulated. The status recorded are later labeled as 0 and 1 for anomaly detection. Status B, C, and E are labeled as normal conditions and is indicated as '0'. Status A, D, and F are labeled as anomalous and is labeled as '1'.

TABLE II. STATUS OF THE PV-ARRAY

Status	***Explanation***
A	Voltage increases, current decreases. No energy harvested yet. Dawn time
B	Voltage ranging from 100 to 300 V. Current ranging from 60 to 62 A. Early morning time
C	Voltage is lesser than Status B. Afternoon time
D	Smallest range of voltage and current values. Dusk time
E	Peak sunlight hours
F	Constant current and maximum voltage. Special case since it has few points

VII. PERFORMANCE METRICS

A. *Accuracy*

Accuracy is the total number of correctly classified instances made by a classification algorithm. It is expressed mathematically as the ratio of correctly classified validation (sum of true positive and true negative) to the total number of observations.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Where TP is true positives, TN is true negative, FP is false positive, and FN is false negative.

B. *Precision*

Precision is mathematically expressed as the percentage of the ratio between correctly predicted positive instances and the total number of predicted positive instances.

$$Precision = \frac{TP}{TP+FP} x100\% \quad (6)$$

C. *Recall or Sensitivity*

Recall, also known as Sensitivity, is mathematically expressed as the percentage of the ratio between correctly predicted positive instances and the total number of positive instances.

$$Recall = \frac{TP}{TP+FN}\ x\ 100\% \quad (7)$$

D. *F1 Scores*

F1 scores refers to the weighted average of precision and recall. The metric considers the balance between precision and recall since there is a tradeoff. It is mathematically expressed as

$$F1\ Score = \frac{2(precision)(recall)}{precision+recall}\ x\ 100\% \quad (8)$$

E. *AUC-ROC Curve*

AUC-ROC stands for Area Under the Curve (AUC) of Receiver Characteristic Operator (ROC). It gives information about the model's performance in classification. ROC is a probability curve where it plots the true positive rate against false positive threshold given range of threshold values. AUC quantifies the ability of the model to classify various classes. A higher AUC indicates better performance of model in classification problems.

VIII. RESULTS

A. *Stacking-based Ensemble Classifiers*

TABLE III. CLASSIFIERS ACCURACY

Classifier	*Accuracy*
Single Classifier 1	85.25%
Single Classifier 2	84.14%
Single Classifier 3	63.44%
Stacked Classifiers	94.00%

Table 3 shows the performance comparison of single classifier and stacking-based ensemble classifier using the accuracy metric. From the table, it is observed that the accuracy of stacked classifiers is higher compared to single classifier, having an accuracy value of 0.9400. It implies that it is 94% accurate in performing classification tasks.

B. *Bagging-based Ensemble Autoencoders*

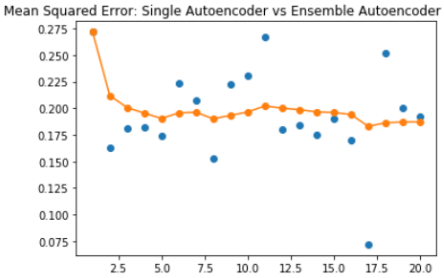

Fig. 13. Plot of Mean Squared Error of Model Reconstruction using Single (blue dot) and Bagging-based Ensemble Autoencoder (orange line with markers)

Figure 13 shows the comparison of performance of single and ensemble autoencoders in reconstructing the training dataset by plotting its mean squared error. The plot shows that at some iteration, bagging-based ensemble autoencoders produce less error in reconstructing the model compared to single autoencoder. It is noted that the performance of ensemble learner is getting stable at some iteration compared to single autoencoder by observing the plot's downward trend.

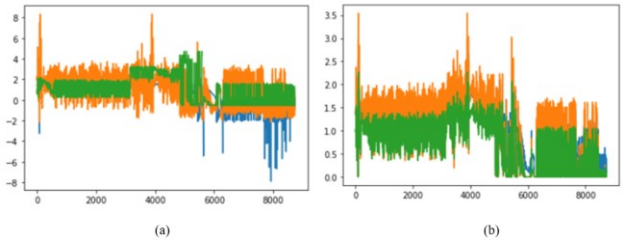

Fig. 14. (a) Comparison of Plot of the Training Dataset (b) Reconstructed Dataset

Figure 14a shows the comparison of the plot of training dataset and figure 14b the reconstructed dataset using bagging-based ensemble reconstruction autoencoders. From the figure, it shows that the original and reconstructed model has almost identical shape.

The maximum value of mean squared error (MSE) of the training dataset is assigned as the threshold for anomaly detection. The threshold indicates the worst-case scenario in reconstructing the training dataset. The testing dataset is used in validating the performance of model reconstruction. If the MSE of testing dataset is greater than or equal to MSE of training dataset, it is labeled as anomaly. In the study, the reconstruction error threshold is 0.221 Using the threshold, the model labeled 4775 datapoints as anomaly.

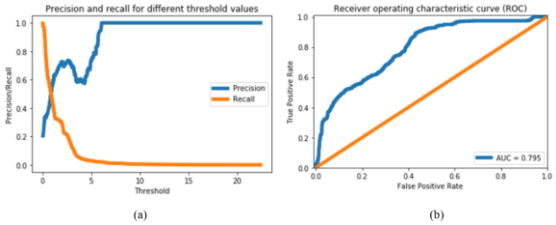

Fig. 15. (a) Recall-Precision Curve with varying Threshold of Bagging-based Ensemble Autoencoders. (b) Receiver Operating Characteristic Curve of Bagging-based Ensemble Autoencoders.

Figure 15b shows the ROC curve of the bagging-based ensemble autoencoders. The curve tells the information about the model's performance in distinguishing classes. The plot shows that the model has a fair performance in classification task. The recorded AUC is 0.795, implying that there is a 79.5% chance that the model is capable of distinguishing normal and anomalous datapoints. The model will obtain a true positive rate of 70% at false positive rate of 30%.

Figure 15a shows the recall-precision curve of the ensemble autoencoder with varying thresholds. The intersection of precision and recall can be used as a new hyperparameter for optimizing the filtering of false positives and negatives. In the study, the new hyperparameter is set to 0.5.

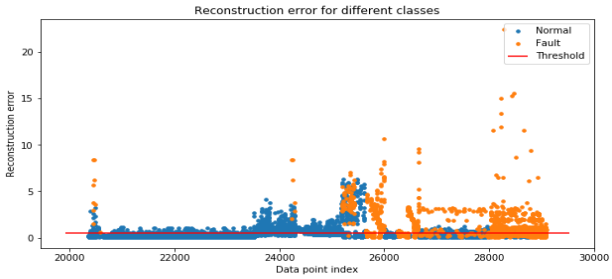

Fig. 16. Plot of Reconstruction Error for Different Classes

Figure 16 shows the visualization of classification between anomaly and non-anomalous data points. Blue points represent normal conditions and orange points as anomalous points.

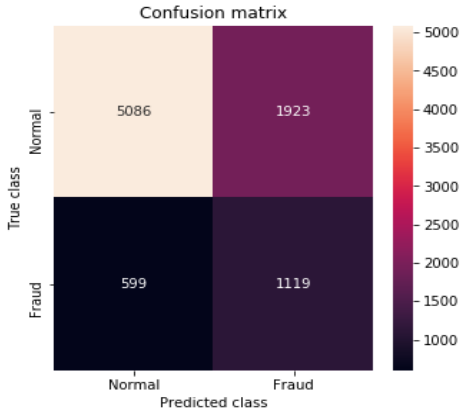

Fig. 17. Confusion Matrix of Anomaly Classification

TABLE IV. CLASSIFICATION REPORT

	Precision	***Recall***	***F1-score***
0	0.89	0.73	0.80
1	0.37	0.65	0.47
Accuracy	-	-	0.71
Macro average	0.63	0.69	0.64
Weighted Average	0.79	0.71	0.74

Figure 17 shows the confusion matrix and Table 4 shows the classification report based on the provided confusion matrix. It is shown that there are 5086 points as true positive, 1923 as false negative, 599 as false positive and 1129 as true negatives. The ensemble autoencoder has an F1-score of 0.71. Considering the unweighted (macro-average) and support-weighted mean (weighted average) per label, it has an F1 score of 0.69 and 0.74, respectively.

IX. CONCLUSION

In this paper, mean-shift clustering was used to determine the unique profile in dataset containing the electrical parameters of PV array and solar irradiance. A stacking-based ensemble classifier was used to classify the predicted clusters. It shows that it has higher accuracy value compared to single classifiers. A bagging-based ensemble autoencoders was used to detect anomalous trend in the dataset. Using bagging-based ensemble learner shows lesser mean squared error in reconstructing the input sequence compared to single autoencoder. The ensemble autoencoder has fair performance in classifying anomalous datapoints from normal based on AUC and F1 scores. It is recommended to use ensemble LSTM autoencoders and classifiers for better performance.

ACKNOWLEDGMENT

The researchers would like to express its gratitude to the

and

for their continuous encouragement and support.

REFERENCES

[1] E. Pedersen *et al.*, "PV Array Fault Detection using Radial Basis Networks," *10th Int. Conf. Information, Intell. Syst. Appl. IISA 2019*, pp. 12–15, 2019, doi: 10.1109/IISA.2019.8900710.

[2] S. V Spataru, A. Gavriluta, D. Sera, L. Maaløe, and O. Winther, "Development and Implementation of a PV Performance Monitoring System Based on Inverter Measurements," 2016, doi: 10.1109/ECCE.2016.7855039.

[3] M. Patil and T. Hinge, "Improved Fault Detection and Location Scheme for Photovoltaic System," *2019 Innov. Power Adv. Comput. Technol. i-PACT 2019*, pp. 12–15, 2019, doi: 10.1109/i-PACT44901.2019.8960246.

[4] Z. M. Omer and H. Shareef, "Adaptive boosting and bootstrapped aggregation based ensemble machine learning methods for photovoltaic systems output current prediction," *2019 29th Australas. Univ. Power Eng. Conf. AUPEC 2019*, 2019, doi: 10.1109/AUPEC48547.2019.211856.

[5] M. Q. Raza, M. Nadarajah, and C. Ekanayake, "An Improved Neural Ensemble Framework for Accurate PV Output Power Forecast," *2016 Australas. Univ. Power Eng. Conf.*, pp. 1–6, 2016, doi: 10.1109/AUPEC.2016.7749296.

[6] T. Niimura, K. Ozawa, D. Yamashita, K. Yoshimi, and M. Osawa, "Profiling residential PV output based on weekly weather forecast for home energy management system," *IEEE Power Energy Soc. Gen. Meet.*, pp. 1–5, 2012, doi: 10.1109/PESGM.2012.6345020.

[7] C. S. Lai, Y. Jia, M. D. McCulloch, and Z. Xu, "Daily Clearness Index Profiles Cluster Analysis for Photovoltaic System," *IEEE Trans. Ind. Informatics*, vol. 13, no. 5, pp. 2322–2332, 2017, doi: 10.1109/TII.2017.2683519.

[8] S. Zirpe and B. Joglekar, "Negation Handling using Stacking Ensemble Method," *2017 Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2017*, pp. 1–5, 2018, doi: 10.1109/ICCUBEA.2017.8463946.

[9] K. Leartpantulak and Y. Kitjaidure, "Music genre classification of audio signals using particle swarm optimization and stacking ensemble," *iEECON 2019 - 7th Int. Electr. Eng. Congr. Proc.*, 2019, doi: 10.1109/iEECON45304.2019.8938995.

[10] X. D. Zeng, S. Chao, and F. Wong, "Optimization of bagging classifiers based on SBCB algorithm," *2010 Int. Conf. Mach. Learn. Cybern. ICMLC 2010*, vol. 1, no. July, pp. 262–267, 2010, doi: 10.1109/ICMLC.2010.5581054.

[11] S. Kumar, "Instrumentation for Solar Photovoltaic System Efficiency Monitoring through Modbus Protocol," *2018 Second Int. Conf. Electron. Commun. Aerosp. Technol.*, no. Iceca, pp. 232–240, 2018, doi: 10.1109/ICECA.2018.8474601.

[12] F. Vignola, J. Peterson, R. Kessler, M. Dooraghi, M. Sengupta, and F. Mavromatakis, "Evaluation of Photodiode-based Pyranometers and Reference Solar Cells on a Two-Axis Tracking System," *2018 IEEE 7th World Conf. Photovolt. Energy Conversion, WCPEC 2018 - A Jt. Conf. 45th IEEE PVSC, 28th PVSEC 34th EU PVSEC*, vol. 22, pp. 2376–2381, 2018, doi: 10.1109/PVSC.2018.8547299.

[13] A. Driesse, W. Zaaiman, D. Riley, N. Taylor, and J. S. Stein, "Investigation of pyranometer and photodiode calibrations under different conditions," *2017 IEEE 44th Photovolt. Spec. Conf. PVSC 2017*, pp. 1–6, 2017, doi: 10.1109/PVSC.2017.8366307.

[14] K. Virupakshappa and E. Oruklu, "Unsupervised Machine Learning for Ultrasonic Flaw Detection using Gaussian Mixture Modeling, K-Means Clustering and Mean Shift Clustering," *IEEE Int. Ultrason. Symp. IUS*, vol. 2019-Octob, pp. 647–649, 2019, doi: 10.1109/ULTSYM.2019.8926078.

[15] H. C. Lee and G. Yang, "Integrating dimension reduction with mean-shift clustering for biological shape classification," *2014 IEEE 11th Int. Symp. Biomed. Imaging, ISBI 2014*, pp. 254–257, 2014, doi: 10.1109/isbi.2014.6867857.

[16] Q. Y. Xie and Y. Cheng, "K-Centers Mean-shift Reverse Mean-shift clustering algorithm over heterogeneous wireless sensor networks," *Wirel. Telecommun. Symp.*, 2014, doi: 10.1109/WTS.2014.6835019.

[17] J. M. Beaulieu and R. Touzi, "Mean-shift and hierarchical clustering for textured polarimetric SAR image segmentation/classification," *Int. Geosci. Remote Sens. Symp.*, pp. 2519–2522, 2010, doi: 10.1109/IGARSS.2010.5653919.