A clustering technique to detect the abnormal voltage using sensors

Abhirup Ghosh,
Department of CSE,
Chandigarh University,
Mohali, India,
Abhirupghosh24@gmail.com

Simarpreet Kaur,
Department of CSE,
Chandigarh University,
Mohali, India,
spreet.kaur1203@gmail.com

Abhishek Mishra,
Department of CSE,
Chandigarh University,
Mohali, India,
abhishek706416@gmail.com

Arav Kumar Prasad,
Department of CSE,
Chandigarh University,
Mohali, India
arav3737@gmail.com

Rajendra Prasad Issar
Department of Computer Science
Indiana State University
Terre Haute IN, USA
rissar@sycamores.indstate.edu

***Abstract*— In order to resolve the issue of abnormal voltage of household appliances and electrocution. It is necessary to check the abnormal conditions in the households to avoid the losses. In current days, the abnormal voltage detection is important in terms of research. Voltage stability is an important aspect of power system and maintaining a stable voltage is also important that much. Abnormal voltage conditions, such as overvoltage and undervoltage can rise to power quality issues which leads to the appliances damage due to load fluctuations, faults etc. factors. It is too much important to have real-time monitoring and situational awareness to respond against the abnormal voltage conditions. This paper proposes an abnormal processing method based on K means algorithm with the help of Smart IoT sensors leading to system break and alerting through the help of alarm.**

Keywords— Voltage, Waveforms, K-Means Algorithm, Sensors, ATmega360, ACS723, Internet of Things (IOT)

I. INTRODUCTION

We cannot imagine the modern society without the electricity as it plays an indispensable role in our daily lives. With the pace of development of modern society and technology, the lifestyle of humans has been improved by more household appliances in the life. But due to this electrocution demand for electric current increased widely. The increased demand for electric energy, particularly from home appliances has been driven force behind this transformation. Because irregularities in electric load might unexpected consumption data has become a crucial task. Voltage anomalies can impair electronic equipment, machinery, and even human safety. Voltage anomalies include fluctuations, sags, spikes, and harmonics[1].

Electrocution is death caused by the passage of electric current. The main concern about the electricity is the condition of the abnormal current waveforms or abnormal voltage in the circuit. In India, there is an alternate current (AC) 220-230 V, 5A-10 A. As per the abnormal voltage cases occurrence data, 47.75 lakh cases in the households i.e., home, 29.5% other than workplace or house and 22.7 % cases at workplace. Usually, such deaths are accidental which can be prevented if we have taken the proper safety measures.

In spite India may have achieved 100% household electrification but 28% of consumers remain unhappy over the quality of voltage fluctuations or abnormal conditions which result in the damage of [2]appliances .For power distribution systems and the equipment attached to them to remain reliable, aberrant voltage circumstances must be quickly detected and addressed. The reliable and effective detection of abnormal voltage in household electrical circuits is overload importance in ensuring the safety and mobility of electrical appliances.

As the name suggests, anomaly detection is a technique for identifying data that deviates from the normal. Data anomalies are events that do not confirm to the expected pattern of behaviour. When referring to power consumption, the term abnormal voltage it refers to the voltage consumption having gap between the model's predicted and actual voltage consumption within the household. [3] The ability of identifying and responding to voltage fluctuations beyond acceptable range can avoid costly damage of appliances, safety hazards and decreasing the risk of electrical short circuits or fires. Traditional models often use anomaly detection and univariate prediction in situations where there are too many additional features is challenging. However, as technology and science improve, the power system is getting more intelligent and flexible, and people's daily use of electric energy is changing. As the combination of Machine Learning algorithm and Internet of Things gives an innovative approach to solve the complex problems or issues.[4] The low-cost, high-precision ACS723 current sensor can be used to measure current in both AC and DC circuits. ACS723 sensor is most accurate and secured device known to monitor current and voltage accuracy. The voltage at a load can be calculated using the current passing through the load as measured by the ACS723 sensor. The very sensitive ACS723 current sensor can measure the current flowing through a conductor with outstanding accuracy. Then the data or output fed into the microcontroller algorithm processes them to make real-time voltage assessment. If the abnormal voltage level is detected, the system takes immediate action, ringing the alarm and disconnecting the circuit to avoid the damage or fires.[1]In this situation,[5] the K-means algorithm can find patterns and trends in voltage fluctuations that may not be visible using conventional techniques when it is applied to data gathered by the ACS723 sensor.

K-means algorithm is an unsupervised learning algorithm which have no labelled data for the clustering. It performs the division of objects into the clusters having similarity and dissimilarity to the objects belonging to another cluster. Here, the term 'K' represents the number. It is useful for grouping data points into discrete groups based on similarities.

This research presents a novel method for identifying abnormal voltage conditions in electrical systems by applying this sensor in combination with the K-means clustering algorithm. A well-known machine learning method called the K- means algorithm. The ACS723 sensor is most accurate and secured device known to monitor current and voltage with accuracy. This paper integrates into the development and deployment of an Abnormal Voltage Detection using Machine learning Algorithm i.e., K-means algorithm. The main aim of this research is to leverage the capabilities of K-means with ACS723 sensor to detect and analyse the electrical voltage in real-time manner and present it to the user in real time for their own benefits. a.

The proposed system is designed to identify voltages which are exceedingly above 230 Volts to trigger the alarm and break the circuit to avoid potential electrical issues. Compared to conventional methods of abnormal voltage detection, the suggested method provides a number of benefits. First, the proposed solution is more robust to changes in the power system because it is based on machine learning. Second, the proposed system is simpler to maintain and modify because no specified rules are needed. Third, the suggested method makes advantage of the high-precision, low-cost ACS723 current sensor. This approach can help electricity systems become more dependable and effective which can aid in preventing equipment failure and blackouts by early detection of anomalous voltage patterns.

II. Literature Survey

The author [6] during his study in 2014 used LEACH- based energy-efficient and K-means-based quick clustering algorithms to develop a hybrid cluster scheme for Wireless Sensor Networks (WSNs). The optimal "k" value for data modeling was determined using the Elbow technique and kmeans algorithm. Building clusters with the Low Energy Adaptive Cluster Hierarchy (LEACH) protocol decreased energy consumption and increased network lifespan. Automatic cluster number selection is used in the proposed hybrid energy-efficient quick k- means and LEACH-based clustering algorithm.

The authors [7] and [8] noticed production issue with electricity surfaced in 2016, calling for effective management. The use of clustering techniques to identify patterns in electricity use is highlighted in this paper. To comprehend electricity usage profiles, data mining techniques such as power calculation and clustering were applied. Data was processed and grouped using the K- Means algorithm to examine how similar clients used electricity. Finding noise and outliers in numerical data is possible using this non-hierarchical clustering algorithm. Power load forecasting, seasonal and intermittent energy demand changes, and the design of energy-efficient solutions are all made possible by data mining on electricity consumption. The analysis showed that summer usage is more prevalent than winter use, which is less prevalent.

According to analysis of authors [9] and [10] in 2019, Because of aging infrastructure, lack of DER integration, and frequency of natural disasters, the current power distribution system faces difficulties. Future systems should concentrate on enhancing power reliability to critical loads during emergencies to address these issues. To maximize power availability, advanced Micro Grid systems can be visualized using machine learning algorithms. To increase grid reliability, the US Department of Energy suggests implementing microgrids, which are collections of interconnected loads and distributed energy resources. Advanced Micro Grid (AMG) Systems were created as a result of recent advancements in microgrid technology, which added features like increased dependability, supply demand balance and economical dispatching.

Another thesis by author [11], [12] suggests that in order to lower greenhouse gas emissions, this thesis investigates how much energy Controller Area Network (CAN) and Ethernet use in electric vehicles (EVs). It demonstrates that Ethernet uses 2.5–4 times as much energy as CAN and, based on actual measurements, 4.5–6 times as much energy. EEE provides more bandwidth but introduces brief delays as well. The study emphasizes the significance of using Ethernet and EEE effectively to lessen the negative effects on the environment and problems with battery charging in the EV sector. In order to determine whether higher bandwidth requirements can be met without significantly increasing power consumption, the thesis also compares the energy consumption of high-speed CAN and Ethernet.

In 2021, author [13] analyzed that Monitoring electricalenergy loads is crucial for power grid maintenance and power theft detection as the power system becomes more intelligent and flexible. In order to estimate power consumption over time and spot anomalies, this paper combines the Transformer deep learning model with the K-means clustering strategy. The K-means clustering technique optimizes prediction results while the Transformer model predicts the power consumption for the upcoming hour.

According to author [14],[15] Popular data mining algorithms like the k-means clustering algorithm have drawbacks like random centroids initialization and the requirement for predefined clusters. It also has trouble handling different data types. This paper provides a thorough review of the k-means algorithm research, discussing variations and recent advancements. To find out how effective they are, different datasets are subjected to experimental analysis. The focus on unsupervised learning algorithms sets this survey paper apart from others in the field. Clustering algorithms use the structure of the data distribution to their advantage and specify criteria for combining data that share similar traits. Each cluster in a perfect clustering scenario would be made up of similar data instances that are very different from the instances in the other clusters. The study of clustering spans many disciplines, including statistics, pattern recognition, computational geometry, bioinformatics, optimization, and image processing. Clustering is essential to many data-driven applications. In the past ten years, a plethora of clustering techniques have been developed and used in numerous application domains. The k-means algorithm, a well-liked and frequently used data analysis tool, is examined in this survey along with its drawbacks and potential solutions. It focuses on its capabilities to handle various types of data and its limitations in terms of assigning centroids and clusters. Despite the suggested alternatives, these are domain-specific and have poor generalizability. The paper compares other clustering surveys and identifies the main contributions of this one. The goal of the study is to offer remedies for resolving these problems.

According to Authors [16][17], in 2022 By tracking and managing loads at the appliance level, a methodology was developed to study and harvest flexibility in the residential building sector using a cluster of smart homes with energy management systems (SHEMS). The method makes use of three home hubs, five specially designed smart plugs, and a vendor-based smart plug. According to the findings, both types of smart plugs successfully measured loads and communicated with other layers, which led to lower electricity costs. Additionally, a hybrid cloud-fog- based architecture for synchronizing local and cloud databases was created. The author [18] says, To examine and take advantage of the adaptability of smart homes with energy management systems (SHEMS) in the residential building sector, a methodology has been developed. The methodology uses a cluster of SHEMS prototypes to monitor and manage loads at the appliance level. In order to reduce electricity costs and adhere to grid constraints, the devices are distributed among three residential hubs and used at a cluster scale. The findings demonstrate that both types of smart plugs are effective at communicating with other layers and measuring loads, which reduces the cost of electricity.

III. METHODOLOGY

The k means algorithm is a clustering algorithm which will provide clusters at the end of training of the dataset. The k means algorithm is best suited for high accuracy on a large and efficient dataset. The K means algorithm is first trained on a dataset and then the accuracy is tested, whilst dividing the clusters into three halves- first one for the normal flow of current and the other one for the abnormal flow of current and the third one for the underflow of current.

The ACS 723 sensor is a circuit breaking sensor which contains of 5 pins. The sensor will take AC current as input and then output the proportional voltage by measuring the magnetic field which is perpendicular to the electric flowing Electricity field. The Voltage which is calculated is then given as input to the K means algorithm which then predicts if the current is above the normal value or not.

After the ACS723 sensor senses the voltage, it is then fed into the algorithm which is pre-installed into the micro controller, and the micro controller predicts in which clustered region the voltage is suited and on the basis of that the circuit is broken if the prediction is greater than normal range of 220-230 volts.

3.1 Data collection and training

A curated unlabelled dataset which is pre available on the internet is used to train the valuable datapoints for their respective centroid. The dataset contains the value for 200 values with the features of current, frequency, impedance value as the independent variables and the Voltage as dependent variable. The first 5 rows of the dataset are as follows: -

Table 1. Dataset Sample Table

Time Stamp	Frequency	Impedance value	Current (in A)	Voltage (in V)
10:11	50	44	5	220
11:11	50	44	5	220
12:11	50	44	5	220
13:11	50	44	10	230
14:11	50	44	5.7	221

The dataset is applicable for unsupervised learning as it is not a labelled dataset, and the dataset contains the current and voltage value which is calculated at a 1hr time-stamp to find the graphical clustered relationship between the current and voltage.

3.1 Data sensing

The ACS 723 sensor is a circuit breaking sensor which has 8 pins which is used to measure voltage. The ACS 723 is mostly used because of its accuracy, easy to use, inexpensive nature. The ACS 723 is a Hall effect sensor which measures the proportional voltage given the alternating current. The ACS 723 has 8 pins. Following are the function and the working of each pin

a) The IP+ and IP- pin is used to measure the measure the positive and negative current input from the AC current carrying wire

b) VCC- This pin powers the sensors to function properly

c) GND- The grounding pin for the sensor

d) RSENSE- The pin is used to set the sensitivity of the sensor. The more the sensitivity the lesser the resistance of the RSENSE resistor.

e) BW_SEL- The BW_SEL is the pin used to select the bandwidth of the sensor that means the frequency range for the current that it can measure accurately.

f) OUT_EN- This pin is used to enable or disable the output of the sensor.

g) FAULT- The FAULT pin outputs the high voltage if there is a fault in the sensor letting the user knows that the sensor is damaged.

The voltage value which is obtained is given as input via the jumper wires to the micro controller which contains the K means algorithm which is used for the near to accurate prediction.

Table 2. Pin information

Pin name	Quantity and function
IP+ and IP-	2, measures positive and negative current
VCC	1, supply the power to sensor
GND	1, Ground the sensor
RSENSE	1, set the sensitivity of the sensor
OUT_EN	1, gives the proportional voltage
FAULT	1, Lets the user know if there is a fault in the sensor or not
BW_SEL	1, Selects the bandwidth of the frequency on which the voltage measuring is done

3.2 Data cleaning and conversion

The data which is obtained will contain the value of the voltage upon reading the value of current will then be cleaned and made it into a form which is understandable by the machine learning algorithm. To pre-process the data, cleaning need to be done by removing the noise and normalizing the data as per the scale of the trained data. The data is the converted by the help of NumPy array or a Pandas Data Frame.

Figure 1.Acs 723 Sensor

The figure above is a ACS 723 sensor used for sensing the data in real time and displaying it via the clustering methods. The ACS 723 sensor is a voltage measuring sensor by sensing the current in real time.

3.3 Data Prediction

The K-means algorithm follows the following steps: -

I. By using the elbow method find the appropriate number of clusters.

II. After assigning the clustering centroids, find the Euclidian appropriate distances between the data points and the centroid.

III. After finding the appropriate data points and assigning the clusters of the data points, recompute the data points again and again by calculating the mean of the data points .

IV. After assigning the clusters to the data points, check for the convergence.

V. If the convergence doesn't change then pick out the final centroids for the respective data point clusters and plot the output.

While doing this algorithm intelligent centroid allocation technique is used by assigning the in it value of the K-means function to "k-means ++" to automatically change it form the in it value of 10 to auto. After finding and training the model on the dataset the clustering points are displayed in the graphical format and on the basis of the input prediction is done. The following is the systematic diagram of the Methodology

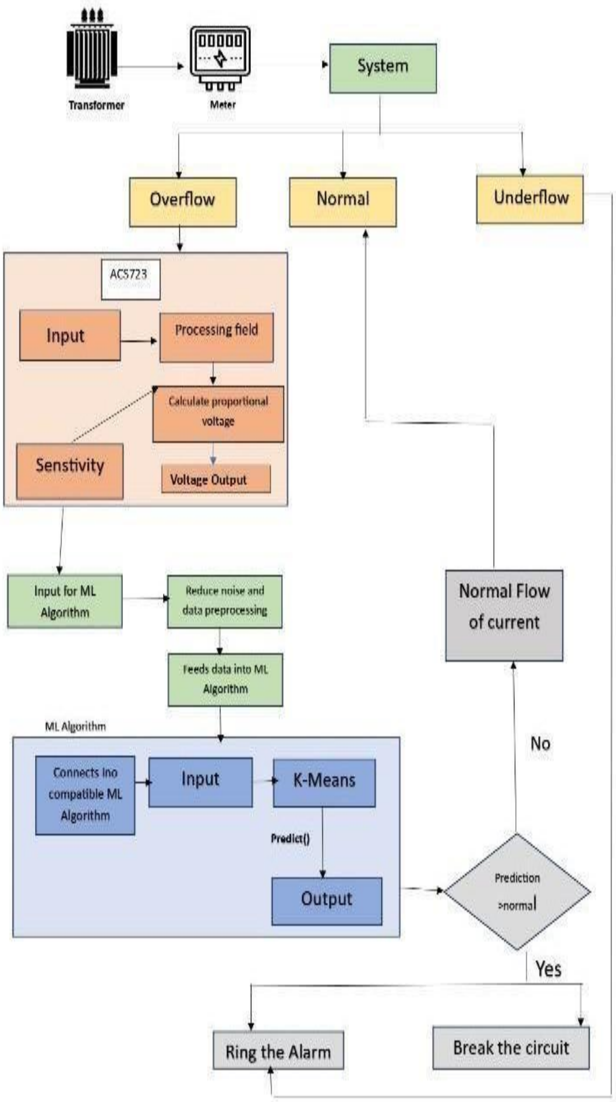

Figure2. Workflow

After doing the prediction the voltage is calculated on the basis of the intervals of the if-else and else if ladder structure and as a result the final output is produced.

```
[2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 2 2 2 2 2 2 2 2 2 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1
 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1]
```

Figure 3. Output analysis

Here each level of voltage is assigned a numeric value so that the model can train accurately without displaying the result in the form of string here in this scenario the numeric digit "2" represents abnormal voltage, numeric digit "1" represents normal voltage and numeric digit "0" represents abnormal voltage as a result the accuracy of the model is also increased abruptly. So the accuracy increases abruptly leading to a good prediction of the model.

Table 2. Significance of voltage clustering

Number	Representing value
1	Low voltage
2	Normal voltage
3	Abnormal voltage

IV. RESULT

At the beginning the number of clusters are to be calculated with the help of the elbow method. Elbow method is a very well-known method which is used to find the number of clusters for the particular dataset. While implementing the elbow method we will find the number of clusters to be needed to be implemented for the best training of our model. The model is trained on the basis of the number of clusters.

There are many ways to find the number of clusters such as the Silhouette method, Cross validation method in the k means algorithm but the most appropriate and the innovative way by using the elbow method which gets its name because of the elbow like structure that is formed in the graph giving the actual number of clusters in the k means algorithm to be used for appropriate results.

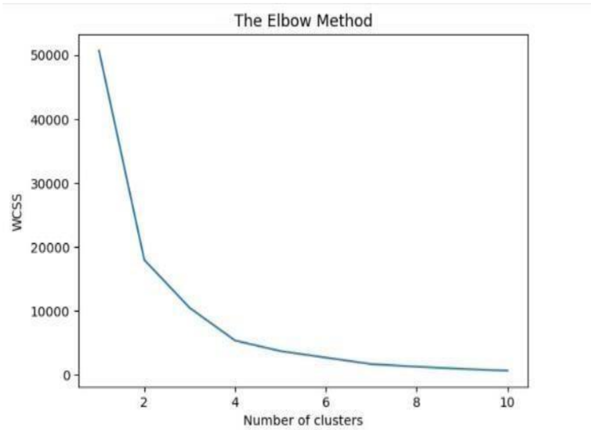

Figure 4. Elbow method

After applying the elbow method, we find out that the required number of clusters for our specific algorithm is 3. The elbow method also helps to keep as assured that the clustering number is going to be safe and will be used in a proper manner. After the number of clusters is decided the clustering is done on the basis of K-means algorithm. The k-means algorithm takes the number of clusters as 3 as input and also the value of in it as "k-means++" for intelligently assigning of the centroid for the clustered graph.

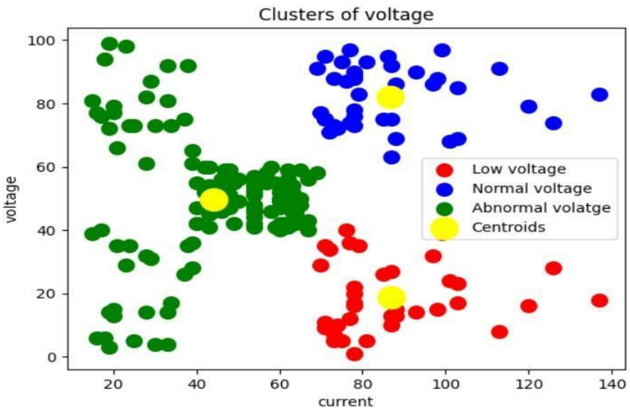

Figure 5. Cluster Graph showing the voltage points

The figure above shows the clustering of the points of the voltage of the based on the data points as input.

The K means algorithm is then used to predict the accurate value of the voltage which is taken as an input. After predicting the value, it then checks against the normal range of the value. After the prediction 3 cases might arise:

a) If the predicted value is greater than the normal range of the values then the circuit will be broken instantly and then alarm will be triggered.

b) If the predicted value is same as the normal value, then nothing will happen normal current flow will occur

c) If the predicted value is lesser than the normal range of value then the alarm will be triggered but no circuit breaking will be there as in most cases underflow of current in general does not hamper the appliances nor has high risk of catching fire After the clustering is done then comes the part for prediction as in the algorithm will be able to predict the value of the model accurately or not for the unknown data. To see the prediction for the unknown data we need to input two values one for the current and their respective voltage as the dataset is trained on a 2-D array. The values which are then input for the prediction will be placed on the clustering algorithm which will then plotted on the graph by calculating their Euclidean distance and checking their similarities and then plotting current vs voltage graph. To understand the working of the model for the unknown values.

Table 3. Output on Test values

Test Value	Output
[6,228]	Normal Voltage
[2,220]	Low Voltage
[14,240]	Abnormal Voltage

In the above table prediction is done on 6A current for which the voltage at a particular timestamp it is showing as 228V. Now as per Indian household the value if normal voltage is between 220-230V so it is showing correct values.

After that, prediction is done on 2A current for which the voltage at a particular timestamp it is showing as 200V. Now as per Indian household the value if normal voltage is between 220-230V so it is showing low voltage which is correct values.

Furthermore, prediction is done on 14A current for which the voltage at a particular timestamp it is showing as 240V. Now as per Indian household the value if normal voltage is between 220-230V so it is showing abnormally high voltage which is correct values.

V. CONCLUSION

Electrical systems may experience unexpected abnormal conditions, such power outages, leading to interruptions in quality of power. Relays and systems are designed for the isolation of a fault element as fast as possible should be used to protect these faults so that they are clearly distinguished by their direction and classification. In this paper, we have presented a digital signal processing techniques for Electrical systems to protection of all types of household electrical components.

Measured current signals are used to train the proposed algorithm to classify, identify and differentiate the path of various faults within the restricted protection zone. In this method, the protection function identifies the occurrence of a fault, and it will immediately break the circuit.

VI. FUTURE WORK

a. Supervised Machine learning can be used to train machine learning models by providing Labelled datasets, and it can also be used to develop new features that are indicative of abnormal voltage and current waveforms.
b. constructing strategies to lower costs and increase accessibility of the algorithm No matter who they are or where they live, everyone should have access to this technology.
c. Developing techniques for EVs to recognize abnormal voltage and current waveforms. The creation of techniques to identify anomalous voltage and current waveforms in EVs may aid in avoiding issues with both the infrastructure used for charging and the actual vehicles.
d. Constructing techniques for industrial waveform detection of abnormal voltage and current. Industrial facilities frequently consume electricity, making them vulnerable to electrical issues. It might be possible to reduce the likelihood of fires and other mishaps by developing techniques to identify anomalous voltage and current waveforms in industrial environments.
e. Creating techniques for spotting unusual voltage and current waveforms during cybersecurity threats. Atypical voltage and current waveforms may be used to identify cyberattacks on electrical systems, which are becoming more susceptible to them.

VII. REFERENCES

[1] Geon Yong Park, Heeseong Kim, Hwi Woon Jeong, and Hee Yong Youn, "A Novel Cluster Head Selection Method based on K-Means Algorithm for Energy Efficient Wireless Sensor Network", 2013 27th International Conference on Advanced Information Networking and Applications Workshops, 20130325.

[2] Murty MN and Devi VS 2011 Pattern recognition, An algorithmic approach (London: Springer)

[3] Hodge V H and Austin Jim 2004 A survey of outlier detection methodologies J. Artificial Intelligence Review 22 85-126

[4] B. Plumer, "Bad news: The U.S. power grid is getting pricier, less reliable," The Washington Post, March 8 2013. [Online],

[5] A. Smith, N. Lott, T. Houston, K. Shein, J. Crouch, J. Enloe, "U.S. Billion-Dollar Weather & Climate Disasters 1980-2019," NOAA National Centers for Environmental Information, [Online]

[6] S. Chakraborty, M. Lukasiewycz, C. Buckl, S. Fahmy, N. Chang, S. Park, Y. Kim, P. Leteinturier, and H. Adlkofer, "Embedded systems and software challenges in electric vehicles", in Proceedings of

the conference on design, automation and test in Europe, EDA Consortium, 2012.

[7] S. Shanker, “Enhancing automotive embedded systems with FPGAs”, PhD thesis, Nanyang Technological University, 2016.

[8] Ahn, B.-H., Choi, H.-R., and Lee, H.-C. (2015). Regional Long-Term/midTerm Load Forecasting Using Sarima in south korea. J. Korea AcademiaIndustrial cooperation Soc. 16, 8576–8584. doi:10.5762/ kais.2015.16.12.8576

[9] Mohri, M.; Rostamizadeh, A.; Talwalkar, A. Foundations of Machine Learning; MIT Press: Cambridge, MA, USA, 2012. [Google Scholar]

[10] Bishop, C.M. Neural Networks for Pattern Recognition; Oxford University Press: Oxford, UK, 1995. [Google Scholar]

[11] Afonso, J. A., Rodrigues, F., Pereira, P., Gonçalves, H., & Afonso, J. L. (2015). Wireless monitoring and management of energy consumption and power quality events. Lecture Notes in Engineering and Computer Science, 2217, 338–343. Ahmed, M. S., Mohamed, A., Homod, R. Z.,

[12] Shareef, H., Sabry, A. H., & Bin Khalid, K. (2015). Smart plug prototype for monitoring electrical appliances in Home Energy Management System. 2015 IEEE Student Conference on Research and Development, SCOReD 2015, 32–36. https://doi.org /10.1109/SCORED.2015.7449348

[13] Afonso, J. A., Rodrigues, F., Pereira, P., Gonçalves, H., & Afonso, J. L. (2015). Wireless monitoring and management of energy consumption and power quality events. Lecture Notes in Engineering and Computer Science, 2217, 338–343.

[14] de Sousa Jr, Rogerio R., et al. "Electrical Conductivity and In Situ SAXS Probing of Block Copolymer Nanocomposites Under Mechanical Stretching." *ACS Applied Materials & Interfaces* (2023).

[15] Obaid, Sofian N., et al. "Design and Fabrication of a Flexible Opto-Electric Biointerface for Multimodal Optical Fluorescence and Electrical Recording." *ACS Applied Electronic Materials* 5.3 (2023): 1688-1696.

[16] Li, Junyu, et al. "Direct-Laser-Writing Manipulating α/δ-FAPbI3 Array Enables Low-Crosstalk Terahertz Sensing and Imaging." *ACS Photonics* 10.8 (2023): 2755-2764.