Journal of Computer Science

Research Article

IoT Anomaly Detection Using Picture Fuzzy Clustering Approach

Fehmin Nadira Laskar[1], Vijo Arul Selvi M.[2], Fokrul Alom Mazarbhuiya[2], Mohamed Shenify[3] and Vijay Prasad[1]

[1]*Department of Computer Applications, Assam Don Bosco University, Guwahati, India*
[2]*Department of Mathematics, Assam Don Bosco University, Guwahati, India*
[3]*College of Computer Science and IT, Albaha University, Albaha, KSA, Saudi Arabia*

Article history
Received: 20-07-2025
Revised: 01-12-2025
Accepted: 04-12-2025

Corresponding Author:
Fokrul Alom Mazarbhuiya
Department of Mathematics,
Assam Don Bosco University,
Guwahati, India
Email:
fokrul.mazarbhuiya@dbuniversity.ac.in

Abstract: Improving the effectiveness of security systems without slowing them down is a major challenge in cybersecurity. Many methods have been explored for detecting anomalous behaviour in network data, with fuzzy set-based approaches standing out for their potential. The Internet of Things (IoT) consists of many connected devices that constantly generate large amounts of data and perform tasks in real time. Because they are always online, these devices are especially vulnerable to cyberattacks. Detecting such malicious activity considered anomalies in the data is a key research issue. Picture Fuzzy Sets (PFSs) offer a robust means to deal with the uncertainty, vagueness, and imprecision found in IoT data. PFSs are constructed on Intuitionistic Fuzzy Sets (IFSs) by introducing a neutrality component, in addition to membership and non-membership values. In this paper, we propose a method based on Picture Fuzzy C-Means (PFCM) clustering to detect anomalies in IoT data. This method is an improved version of the traditional Fuzzy C-Means (FCM) algorithm and is better suited to handling the multifaceted uncertainty in IoT environments. We also evaluate the computational efficacy through complexity analysis. Experimental results using real-world datasets, NSL-KDD and SAB, show that the detection rate and accuracy improve by up to 10% compared to the k-means algorithm, 5% compared to FCM, and 3% compared to the Intuitionistic Fuzzy C-Means (IFCM) algorithm, demonstrating the superior efficacy of our approach.

Keywords: IoT System, IoT Anomalies, Detection of Anomalies, Canberra Distance, Picture Fuzzy Sets (PFS), Positive Membership Degree, Negative Membership Degree, Neutral Membership Degree, Picture Fuzzy C-Means (PFCM) Clustering Algorithm

Introduction

Anomaly detection within the Internet of Things (IoT) ecosystem has become an increasingly critical component of cybersecurity, primarily due to the rising frequency of unauthorized access and cyberattacks (Alsaedi et al., 2020). With the widespread adoption of IoT devices, there has been an exponential surge in data generation, making these devices attractive targets for malicious actors. As a result, ensuring information security particularly through effective anomaly detection has gained significant importance. Identifying anomalies in IoT data has numerous practical applications, such as fault diagnosis, fraud prevention, predictive maintenance, and system monitoring. In scenarios where consistent and dependable responses are unavailable, anomaly detection can provide valuable insights. To tackle these challenges, this work introduces reliable approaches for detecting anomalies in IoT environments.

The IoT refers to a network of interconnected devices embedded with computing and communication capabilities, enabling them to perform a variety of tasks autonomously (Sethi and Sarangi, 2017). The primary objective of IoT is to enrich and personalize user experiences by facilitating seamless interaction with physical objects. IoT has driven major technological advancements across diverse domains such as agriculture, smart cities, healthcare, transportation, retail, and

logistics. Sethi and Sarangi (2017) described it as a global infrastructure, integrating the cyber and physical worlds based on existing systems and early-generation IoT technologies.

Nowadays, IoT devices have become deeply integrated into everyday life. In agriculture, they support applications such as precision farming, livestock monitoring, and smart irrigation systems (Kopawar and Wankhede, 2024). In healthcare, IoT enables solutions like remote patient monitoring, heart rate and mood tracking, ingestible sensors, and robotic surgery (Atadoga et al., 2024). In education, it supports innovations such as distance learning, smart classrooms, attendance automation, augmented reality, and adaptive learning platforms (Dake et al., 2023). Additionally, IoT is widely applied in areas including smart cities, home automation, transportation, supply chain management, and manufacturing (Masmali et al., 2023). Given their constant connectivity to the Internet and to each other, IoT devices are particularly vulnerable to malicious actors. Therefore, implementing robust security mechanisms is essential to prevent and detect intrusions within these interconnected environments.

Several approaches have been proposed to address the aforementioned problem, among which clustering-based anomaly detection of IoT data is a prominent method (Teh et al., 2021; Ren et al., 2009). Clustering is widely recognized as an effective technique for uncovering patterns and understanding data distribution within datasets (Mazarbhuiya and Abulaish, 2012; Shenify and Mazarbhuiya, 2023), and it has been extensively applied in the context of anomaly detection. For instance, Ren et al. (2009) introduced a fuzzy *c*-means clustering approach for detecting anomalies in mixed-type data. Mazarbhuiya et al. (2019) developed an agglomerative hierarchical clustering algorithm for anomaly detection in network traffic, using the Canberra metric measure as the distance formula (Lance and Williams, 1966; 1967; Clifford and Stephenson, 1975; Emran and Ye, 2001). A hybrid model combining both partitioning and hierarchical techniques was proposed by Mazarbhuiya et al. (2020) to handle anomalies in mixed datasets. Additionally, Mazarbhuiya (2023) presented a hybrid method that integrates rough set theory with a density-based clustering technique for detecting anomalies in high-dimensional IoT data. A two-phase approach that incorporates both partitioning and hierarchical clustering, while also considering the temporal characteristics of real-time data, was introduced by Mazarbhuiya and Shenify (2023a). Further related works can be found in (Mazarbhuiya et al., 2023; Alguliyev et al., 2017; Hahsler et al., 2019; Song et al., 2017; Alghawli, 2022; Younas, 2020; Thudumu et al., 2020; Habeeb et al., 2019; Wang et al., 2022; Halstead et al., 2023; Zhao et al., 2021; Chenaghlou et al., 2018; Firoozjaei et al., 2022; Mazarbhuiya, 2023), highlighting the ongoing interest and development in this area. Notably, Chen et al. (2022) explored insider threats, which pose significant cybersecurity challenges for industrial control systems. An online anomaly detection method using random forests was presented by Zhao et al. (2018), offering a real-time solution. Finally, Samara et al. (2022) provided a comprehensive review of various anomaly detection techniques applicable to IoT systems.

Many of the existing algorithms proposed in the literature exhibit certain limitations, particularly in effectively detecting anomalies within IoT data. However, incorporating fuzziness into clustering techniques can help overcome several of these challenges for the following key reasons. First, fuzzy clustering enables data points to belong to multiple clusters simultaneously, which is advantageous when dealing with complex data structures, ambiguity, or overlapping class boundaries. Second, it demonstrates greater resilience to noise and anomalies, as the transition between clusters occurs gradually rather than abruptly. Third, fuzzy clustering provides a more detailed representation of the association between data points and clusters, offering a richer and more nuanced understanding of the data's inherent structure. Wang et al. (2021) introduced a novel algorithm that incorporates Mahalanobis distance to enhance the accuracy of intrusion detection. Harish and Kumar (2017) proposed a fuzzy c-means clustering method for network intrusion detection, utilizing principal component analysis to select the most discriminative features. Related research efforts can also be found in (Gustafson and Kessel, 1978; Haldar et al., 2017; Zhao et al., 2015; Ghorbani, 2019; Shenify et al., 2024; Zadeh, 1978) introduced the concept of fuzziness into mathematics by defining Fuzzy Sets (FS) based on membership degrees. This foundational idea led to the development of the mathematics of fuzziness, which has since been applied across nearly all domains of human knowledge. In response to real-world challenges, numerous extensions, generalizations, and variations of fuzzy sets have been proposed. One such extension is the Intuitionistic Fuzzy Set (IFS), introduced by Atanassov (1983), which incorporates both membership and non-membership degrees. Building on this, Cuong (2014) proposed the Picture Fuzzy Set (PFS), which further includes a degree of neutrality alongside membership and non-membership degrees.

Fuzzy sets (Zadeh, 1978) and their extensions (Atanassov, 1983) have been effectively employed in various applications, particularly in clustering and anomaly detection (Ren et al., 2009; Mazarbhuiya and Abulaish, 2012; Wang et al., 2021; Harish and Kumar, 2017; Gustafson and Kessel, 1978; Zhao et al., 2015; Ghorbani, 2019; Shenify et al., 2024; Bezdek et al., 1984; Butkiewicz, 2012; Chaira, 2011; Chaira and Panwar 2014; Thong and Son, 2016). For example, an IFS-based

hierarchical clustering algorithm was proposed by Xu (2009), leveraging traditional hierarchical clustering and intuitionistic fuzzy aggregation operators to cluster IFSs. (Xu and Wu, 2010) introduced the Intuitionistic Fuzzy C-Means (IFCM) algorithm, which extends the well-known fuzzy *c*-means method by incorporating distance measures specific to IFSs (Szmidt and Kacprzyk, 2000; Xu, 2007). Additionally, a hybrid method combining rough set theory and IFSs was proposed by Mazarbhuiya and Shenify (2023b) to detect anomalies in network data. This approach used the α-relation, based on the correlation coefficient of IFSs, to generate intuitionistic fuzzy rules.

As the demand for intelligent and autonomous systems grows particularly within the IoT domain, which continuously generates massive amounts of data marked by high volume, velocity, variety, variability, veracity, value, time-sensitivity, location-awareness, and a highly unstructured, semi-structured, and heterogeneous nature applying clustering algorithms like FCM and IFCM becomes increasingly challenging (Shenify et al., 2024). The clustering performance of FCM is often limited due to its reliance on classical fuzzy sets, which struggle with accurately modeling membership, hesitancy, and the vagueness of prototype parameters. Recent studies explored numerous cutting-edge models for anomaly detection. A Variable Temporal Transformer model Kang and Kang (2024) leveraged self-attention mechanisms to capture temporal dependencies among variables. An inclusive study by Carletti et al. (2025) examined two Graph Neural Networks for node anomaly detection in large-scale IoT traffic datasets. Hendaoui et al. (2025) proposed a machine learning model for efficient and privacy-preserving intrusion detection in IoT networks. A clustering-based outlier detection algorithm Huang et al. (2023) utilized mutual information matrices, spectral clustering, and Local Outlier Factor to identify anomalies. Iglesias Vázquez et al. (2023) evaluated eight unsupervised outlier detection methods for streaming data. Additionally, an approach proposed by Retiti Diop Emane et al. (2024) combined Graph Convolutional Networks with the DBSCAN algorithm to detect anomalies in graph-structured data.

Although existing approaches tried to address some of the aforementioned issues and improve clustering quality to a certain extent, their effectiveness remains limited. So, the challenges still exist in this area. For example, IoT data generated from multiple heterogeneous sources is often noisy, imprecise, and overlapping, with high levels of uncertainty, ambiguity, and vagueness. Existing approaches such as *k*-means, FCM, IFCM, etc. (Zhao et al., 2021; Mazarbhuiya, 2023; Wang et al., 2021; Harish and Kumar, 2017; Zhao et al., 2015; Atanassov, 1983) can only handle crisp or mildly fuzzy boundaries, making them inadequate for modelling such high uncertainty. Secondly, the existing fuzzy approaches primarily rely on the membership or at most non-membership aspects; however, the IoT data may also contain neutral states, which were not taken into consideration by any of the existing approaches. Finally, the IoT data are often high-dimensional, complex, and non-uniform. Mazarbhuiya et al. (2020); Mazarbhuiya (2023) tried to address these issues to a limited extent.

In spite of several efforts, the literature does not have a unified solution that can efficiently address the highly uncertain nature of IoT datasets, such as positive membership, negative membership, and neutral membership. Although FCM is widely used, it lacks the ability to handle negative and neutral memberships. IFCM tried to improve this by adding a degree of hesitation, but it still lacks a comprehensive mechanism for neutral responses that often exist in any IoT environment. Moreover, none of the existing studies systematically explored the application of PFSs to IoT anomaly detection by incorporating all three aspects of uncertainty, despite their robust theoretical potential. The PFS framework Cuong (2014) extends fuzzy sets and IFSs by incorporating a degree of neutral membership along with a positive membership and a negative membership. Therefore, it can be appropriate for IoT data, as the data might either be partially perceived or be ambiguous, and some data instances' behaviour is precisely neither normal nor anomalous.

Motivated by these challenges, this paper aims to develop a novel fuzzy clustering approach for IoT anomaly detection using Picture Fuzzy Sets (PFSs), with the goal of achieving higher clustering quality compared to both FCM- and IFCM-based approaches.

The objectives of this paper are outlined as follows:

- First, a distance formula based on the Canberra metric (Lance and Williams, 1966; Lance and Williams, 1967; Clifford and Stephenson, 1975; Emran and Ye, 2001) is defined for Picture Fuzzy Sets (PFSs)
- Second, building on this distance measure, we propose a novel fuzzy clustering method termed the PFCM algorithm to generate soft picture fuzzy clusters from IoT data
- Third, a comparative analysis is performed against existing clustering techniques, specifically FCM and IFCM, to evaluate the effectiveness of the proposed approach

Additionally, the time complexity of the PFCM-based approach is computed. The proposed PFCM algorithm is implemented and tested using MATLAB, with experiments conducted on the NSL-KDD and Skoltech Anomaly Benchmark (SAB) datasets. The results demonstrate that the PFCM-based method significantly outperforms both FCM and IFCM approaches.

Related Work

Anomaly detection in the IoT ecosystem is found to be an important aspect of cybersecurity, mainly because of the huge increase in unauthorized activities and cyberattacks (Alsaedi et al., 2020). The IoT talks about a network of interconnected devices armed with computing and communication capabilities, allowing them to autonomously perform a wide range of tasks (Sethi and Sarangi, 2017). The primary objective of the IoT is to enrich and personalize user proficiencies by enabling seamless interaction with physical objects. IoT has driven major technological advancements across a wide range of domains, including agriculture, smart cities, healthcare, transportation, retail, and logistics. Often considered a worldwide infrastructure, IoT links the physical and digital worlds by incorporating the existing systems and earlier generations of IoT technologies (Sethi and Sarangi, 2017). Nowadays, IoT devices are integrated into almost every aspect of our daily life, such as agriculture (Kopawar and Wankhede, 2024), healthcare (Atadoga et al., 2024), education (Dake et al., 2023), manufacturing (Masmali et al., 2023), and many more. Detecting anomalies in such applications is the most challenging research area of cybersecurity.

Several anomaly detection techniques have been successfully implemented in many IoT environments, and clustering-based approaches are among them. (Teh et al., 2021) proposed an unsupervised feature selection framework for anomaly detection in time-series data. (Ren et al., 2009) proposed an FCM-based approach for the intrusion detection of network data. Mazarbhuiya and Abulaish (2012) proposed a novel agglomerative hierarchical technique for the clustering of period patterns in transaction data. Mazarbhuiya and Shenify (2023c) proposed a similar clustering approach for the clustering of documents. Mazarbhuiya et al. (2019) proposed an agglomerative hierarchical clustering approach for the detection of anomalies in network data. Mazarbhuiya et al. (2020) proposed a hierarchical classification approach using the Canberra metric, which is superior to the clustering approach for quantitative data. Similar works using the Canberra metric were also reported in (Lance and Williams, 1966; 1967; Clifford and Stephenson, 1975; Emran and Ye, 2001). A hybrid approach consisting of rough set theory and density-based clustering for anomaly detection in high-dimensional data was reported by Mazarbhuiya (2023). A two-phase method using both partitioning and hierarchical clustering for real-time anomaly detection was proposed by Mazarbhuiya and Shenify (2023d).

Mazarbhuiya and Shenify (2023a) employed a real-time anomaly detection approach using rough set theory, a dynamic *k*-means clustering algorithm, and an interval superimposition-based subspace clustering approach. Alguliyev et al. (2017) proposed algorithms for clustering and anomaly detection, taking into account the compactness and separation of clusters. Hahsler et al. (2019) proposed an implementation of the DBSCAN clustering algorithm using R. An anomaly detector based on a hybrid semi-supervised approach consisting of a deep autoencoder and an ensemble *k*-nearest neighbor graphs was proposed by Song et al. (2017). Alghawli (2022) proposed a method consisting of components based on entropy analysis, signature analysis, and machine learning for anomaly detection in telecommunication traffic. An anomaly detection using data mining techniques was discussed by Younas (2020). A couple of methods addressing real-time and online anomaly detection in high-dimensional big data using supervised or semi-supervised approaches were discussed in detail in many of the recent research works.

Firoozjaei et al. (2022) conducted a detailed study about the foundation for cyber risk assessment for Operational Technology (OT) systems. An anomaly detection technique using a neighbourhood rough set-based classification approach was discussed by Mazarbhuiya (2023). Chen et al. (2022) made a comprehensive review of insider threat detection in cyber-physical systems. A random forest-based online anomaly detection approach was proposed by Zhao et al. (2018). Samara et al. (2022) provided a detailed review of various anomaly detection approaches for IoT systems. Wang et al. (2021) proposed a Mahalanobis distance-based clustering approach for intrusion detection. A principal component analysis-based FCM approach for network intrusion detection by selecting the most discriminative features for the intrusion detection was discussed by Harish and Kumar (2017). Similar research was reported in many recent articles. An IFS-based hierarchical clustering algorithm was proposed by Xu (2009); Xu and Wu (2010) proposed an Intuitionistic Fuzzy C-Means (IFCM) algorithm incorporating the distance measures specific to IFSs (Szmidt and Kacprzyk, 2000; Xu, 2007). A hybrid method combining rough set theory and IFSs was proposed by Mazarbhuiya and Shenify (2023b) to detect anomalies in network data.

A Variable Temporal Transformer model was proposed by Kang and Kang (2024), which uses the self-attention mechanism of transformers to understand the temporal dependencies and relationships among variables effectively. Carletti et al. (2025) proposed an inclusive study in a realistic setup of two Graph Neural Networks designed for node anomaly detection and applied to large-scale IoT network traffic datasets. A machine learning-based framework for intrusion detection on IoT networks was proposed by Hendaoui et al. (2025) that analyzed the data more efficiently and privately. Huang et al. (2023) proposed an innovative clustering-based outlier detection algorithm that computes a mutual information matrix between features,

partitions the attribute set using reduced spectral clustering, applies the Local Outlier Factor within each subset, and aggregates the scores to detect the anomalies. Iglesias Vázquez et al. (2023) conducted a comparative study of eight unsupervised outlier detection methods for streaming data. Retiti Diop Emane et al. (2024) proposed a new method by incorporating Graph Convolutional Networks with the DBSCAN algorithm to detect anomalies in graph-structured data.

Preliminaries

Below, we present some important terms and definitions used in this paper.

Definition 2.1 (Lance and Williams, 1966; 1967)

Let $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$ be two vectors, then the Canberra distance between X and Y is given by:

$$d(X,Y) = \sum_{i=1}^{n} \frac{|x_i - y_i|}{|x_i| + |y_i|} \quad (1)$$

Definition 2.2 Fuzzy Set (Zadeh, 1978)

Let $X=\{x_1, x_2, \ldots, x_n\}$ be the universe of discourse. A fuzzy set, A on X, is characterized by:

$$A=\{(x_i, \mu_A(x_i));\ x_i \in X, i = 1, 2, \ldots, n\} \quad (2)$$

Where $\mu_A: X \rightarrow [0,1]$, the membership function, gives the grade of membership of each element $x_i \in X$ in A.

Definition 2.3 Intuitionistic Fuzzy Set (Atanassov, 1983)

Atanassov (1983) proposed the definition of an Intuitionistic Fuzzy Set (IFS) A on X as:

$$A=\{(x_i; \mu_A(x_i), \nu_A(x_i));\ x_i \in X, i = 1, 2, \ldots, n\} \quad (3)$$

Where $\mu_A: X \rightarrow [0,1]$ and $\nu_A: X \rightarrow [0,1]$ are the membership function and non-membership function of the fuzzy set A, respectively, satisfying the condition $0 \leq \mu_A(x_i) + \nu_A(x_i) \leq 1$ for every $x_i \in X$. Obviously, $\pi_{A_1}(x_i) = 1 - \mu_A(x_i) + \nu_A(x_i)$ is the degree of hesitation of $x_i \in X$.

Definition 2.4 Distance Measure on IFSs

Let IFS(X) be the collection of all IFSs on $X = \{x_1, x_2, x_n\}$. A distance measure is a real-valued function $d_C: \mathrm{IFS}(X) \times \mathrm{IFS}(X) \longrightarrow \mathbb{R}$ defined by:

$$d^c(A_1,A_2) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{|\mu_{A_1}(x_i)-\mu_{A_2}(x_i)|}{\mu_{A_1}(x_i)+\mu_{A_2}(x_i)} + \frac{|\nu_{A_1}(x_i)-\nu_{A_2}(x_i)|}{\nu_{A_1}(x_i)+\nu_{A_2}(x_i)} + \frac{|\pi_{A_1}(x_i)-\pi_{A_2}(x_i)|}{\pi_{A_1}(x_i)+\pi_{A_2}(x_i)}\right] \forall\, A_1, A_2 \in IFS(X) \quad (4)$$

Definition 2.5 Picture Fuzzy Set (Cuong, 2014)

A PFS (Cuong, 2014) A over $X = \{x_1, x_2, \ldots, x_n\}$ is defined as:

$$A = \{(x_i; \alpha_A(x_i), \beta_A(x_i), \gamma_A(x_i));\ x_i \in X\} \quad (5)$$

With $\alpha_A(x_i) \in [0,1]$ is the degree of positive membership, and $\beta_A(x_i) \in [0,1]$ is the degree of neutral membership and $\gamma_A(x_i) \in [0,1]$ is the degree of negative membership satisfying the condition $\alpha_A(x_i) + \beta_A(x_i) + \gamma_A(x_i) \leq 1$ for every $x_i \in X$. Also, $\rho_A(x_i) = 1 - (\alpha_A(x_i) + \beta_A(x_i) + \gamma_A(x_i))$, is the degree of refusal membership of $x_i \in X$.

Definition 2.6 Distance Measure on PFSs

Let PFS(X) be the collection of all PFSs on X (discrete or continuous), then we define a metric measure on PFS(X) as follows:

For the discrete case:

$$d^c(A_1,A_2) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{|\alpha_{A_1}(x_i)-\alpha_{A_2}(x_i)|}{|\alpha_{A_1}(x_i)|+|\alpha_{A_2}(x_i)|} + \frac{|\beta_{A_1}(x_i)-\beta_{A_2}(x_i)|}{|\beta_{A_1}(x_i)|+|\beta_{A_2}(x_i)|} + \frac{|\gamma_{A_1}(x_i)-\gamma_{A_2}(x_i)|}{|\gamma_{A_1}(x_i)|+|\gamma_{A_2}(x_i)|} + \frac{|\rho_{A_1}(x_i)-\rho_{A_2}(x_i)|}{|\rho_{A_1}(x_i)|+|\rho_{A_2}(x_i)|}\right] \forall\, A_1, A_2 \in PFS(X) \quad (6)$$

For the continuous case (taking $X = [a,b]$):

$$d^c(A_1,A_2) = \frac{1}{(b-a)}\int_a^b \left[\frac{|\alpha_{A_1}(x)-\alpha_{A_2}(x)|}{|\alpha_{A_1}(x)|+|\alpha_{A_2}(x)|} + \frac{|\beta_{A_1}(x)-\beta_{A_2}(x)|}{|\beta_{A_1}(x)|+|\beta_{A_2}(x)|} + \frac{|\gamma_{A_1}(x)-\gamma_{A_2}(x)|}{|\gamma_{A_1}(x)|+|\gamma_{A_2}(x)|} + \frac{|\rho_{A_1}(x)-\rho_{A_2}(x)|}{|\rho_{A_1}(x)|+|\rho_{A_2}(x)|}\right] dx,\ \forall\, A_1, A_2 \in PFS([a,b]) \quad (7)$$

Obviously, in both cases, $0 \leq d^c(A_1,A_2) \leq 1$. (8)

Definition 2.7

Each IoT data instance consists of n measured variables grouped into an n-dimensional vector $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]$, $x_i \in R^n$. A set of N data instances is given by $X = \{X_i;\ i=1, 2, \ldots, N\}$ and is expressed as $N \times n$ matrix as follows:

$$X = \begin{bmatrix} X_{11} & X_{12} & \ldots & X_{1n} \\ X_{21} & X_{22} & \ldots & X_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ X_{N1} & X_{N2} & \ldots & X_{Nn} \end{bmatrix} \quad (9)$$

Proposed Algorithm

In this section, a picture fuzzy c-means clustering algorithm for the detection of IoT anomalies is presented. Suppose that there is an IoT dataset X consisting of N data instances of dimension n. Our aim is to devise X into c clusters, each of which is PFS, satisfying the following objective function:

$$\boldsymbol{Minimize}\ (\boldsymbol{J}) = \sum_{i=1}^{N}\sum_{j=1}^{c}(\boldsymbol{\alpha_{ij}}(\boldsymbol{2}-\boldsymbol{\gamma_{ij}}))^{m}\boldsymbol{d^{c}}(\boldsymbol{X_i},\boldsymbol{V_j})^{2} + \sum_{i=1}^{N}\sum_{j=1}^{c}\boldsymbol{\beta_{ij}}(\boldsymbol{log\beta_{ij}}+\boldsymbol{\gamma_{ij}}) \quad (10)$$

Subject to the constraints:

$$\boldsymbol{\alpha_{ij}}+\boldsymbol{\beta_{ij}}+\boldsymbol{\gamma_{ij}} \leq \boldsymbol{1} \quad \textbf{(11)}$$

$$\sum_{j=1}^{c}(\boldsymbol{\alpha_{ij}}(\boldsymbol{2}-\boldsymbol{\gamma_{ij}})) = \boldsymbol{1} \quad \textbf{(12)}$$

$$\sum_{j=1}^{c}\left(\boldsymbol{\beta_{ij}}+\frac{\boldsymbol{\gamma_{ij}}}{\boldsymbol{c}}\right) = \boldsymbol{1} \quad (13)$$

For $i = 1, \ldots, N$ and $j = 1, \ldots, c$.

The first term of Equation (10) tries to minimize the distance between each IoT data instance $\boldsymbol{X_i}$ with $\boldsymbol{V_j}$ $\forall$ i=1, 2,…, N and j=1,2,…, c. The distance function $\boldsymbol{d^{c}}(\boldsymbol{X_i},\boldsymbol{V_j})$ is weighted by $\boldsymbol{\alpha_{ij}}$ and another factor $\boldsymbol{\gamma_{ij}}$ $\forall$, i=1, 2,…, N and j=1,2,…, c. The exponent m and $\boldsymbol{d^{c}}(\boldsymbol{X_i},\boldsymbol{V_j})^{2}$ control the influence of distance on the objective function (10). The second term of (10) involves the weights $\boldsymbol{\beta_{ij}}$ and $\boldsymbol{\gamma_{ij}}$ $\forall$, i=1, 2,…, N, and j=1,2,…, c, which control the membership of each IoT data instance to the picture fuzzy clusters. This term includes linear and logarithmic components to adjust the membership.

This inequality (11) ensures that the total membership of an IoT data instance $\boldsymbol{X_i}$ across the three components $\boldsymbol{\alpha_{ij}}, \boldsymbol{\beta_{ij}}, \textbf{and } \boldsymbol{\gamma_{ij}}$ for a given cluster with cluster-mean $\boldsymbol{V_j}$ cannot exceed 1 $\forall$ i=1, 2,…, N and j=1,2,…, c, which stops over-assignment of an IoT data instance to multiple components.

The inequality (12) ensures that the sum of the weighted membership of an IoT data instance $\boldsymbol{X_i}$ across all the c clusters equals 1, which is a customary restriction in a fuzzy clustering, meaning each IoT data instance must belong to multiple clusters with partial membership.

The inequality (13) guarantees that the weighted sum of the parameters $\boldsymbol{\beta_{ij}}$ **and** $\boldsymbol{\gamma_{ij}}$ is normalized across all clusters, which ensures the proper distribution of all the IoT data instances across the clusters.

Using the Lagrangian method, the solution of the optimization problem (10) subject to (11-13) is obtained as follows:

$$\boldsymbol{\gamma_{ij}} = \boldsymbol{1} - (\boldsymbol{\alpha_{ij}}+\boldsymbol{\beta_{ij}}) - \left(\boldsymbol{1}-(\boldsymbol{\alpha_{ij}}+\boldsymbol{\beta_{ij}})^{\boldsymbol{\alpha}}\right)^{\frac{1}{\alpha}} \quad (i = 1, ,N, j = 1, c) \quad (14)$$

The above Equation (14) provides the value of $\boldsymbol{\gamma_{ij}}\forall$, i=1, 2,…, N, and j=1,2,…, c, which represents the degree of neutral membership of an IoT data instance $\boldsymbol{X_i}$,i=1, 2,…, N , with respect to cluster $\boldsymbol{V_j}$, j=1,2,…, c . The parameter $\boldsymbol{\alpha}$ controls the adjustment of the degree of neutral membership $\boldsymbol{\gamma_{ij}}$, which is adjusted based on $\boldsymbol{\alpha_{ij}}$ **and** $\boldsymbol{\beta_{ij}}$:

$$\boldsymbol{\alpha_{ij}} = \frac{\boldsymbol{1}}{\sum_{k=1}^{c}\left(\frac{\boldsymbol{d^{c}}(\boldsymbol{X_i},\boldsymbol{V_j})}{\boldsymbol{d^{c}}(\boldsymbol{X_i},\boldsymbol{V_k})}\right)^{\frac{2}{m-1}}}, \boldsymbol{i = 1, \ldots, N, j = 1,}\text{c} \quad (15)$$

Equation (15) computes the degree of membership $\boldsymbol{\alpha_{ij}}$ of IoT data instance $\boldsymbol{X_i}$, i=1, 2,…, N in cluster with mean $\boldsymbol{V_j}$, j=1,2,…, c. The distance $\boldsymbol{d^{c}}(\boldsymbol{X_i},\boldsymbol{V_j})$ between the $\boldsymbol{X_i}$ and $\boldsymbol{V_j}$ is raised to an exponent controlled by m. So, the equation (15) influences the degree of membership based on the relative distances between clusters:

$$\boldsymbol{\beta_{ij}} = \frac{e^{-\gamma_{ij}}}{\sum_{k=1}^{c}e^{-\gamma_{ik}}}\left(\boldsymbol{1}-\frac{\boldsymbol{1}}{\boldsymbol{c}}\sum_{k=1}^{c}\boldsymbol{\gamma_{ik}}\right)\boldsymbol{i = 1, .., N, j = 1, \ldots, c} \quad (16)$$

Equation (16) calculates the non-membership degrees $\boldsymbol{\beta_{ij}}$, which controls the degree of membership of $\boldsymbol{X_i}$, i=1, 2,…, N to cluster $\boldsymbol{V_j}$, j=1,2,…, c based on the degree of neutral membership $\boldsymbol{\gamma_{ij}}$. The exponential decay function involved in the term ensures that the degree of neutral membership has a decaying effect on the non-membership, and the term involving $\boldsymbol{\gamma_{ik}}$, i=1, 2,…, N, k=1,2,…, c adjusts for the whole neutral membership across all clusters:

$$\boldsymbol{V_J} = \frac{\sum_{i=1}^{N}(\alpha_{ij}(2-\gamma_{ij}))^{m}X_i}{\sum_{i=1}^{N}(\alpha_{ij}(2-\gamma_{ij}))^{m}}, \boldsymbol{j = 1, \ldots, c} \quad (17)$$

Equation (17) computes the new cluster mean $\boldsymbol{V_j}$ for the j th cluster as a weighted average of the IoT data instances. The weights are determined by the values of $\boldsymbol{\alpha_{ij}}, \boldsymbol{\beta_{ij}}, \textbf{and } \boldsymbol{\gamma_{ij}}$, which indicate how strongly each IoT data instance belongs to or does not belong to the cluster and how neutral it is. With the help of the equations (14-17), the steps of the picture fuzzy c-means (PFCM) clustering algorithm (Thong and Son, 2016) for the detection of IoT anomaly are described as follows:

Picture Fuzzy c-Means (PFCM) Clustering Algorithm

Given dataset X as expressed using (9).

***Initialize**: c (number of clusters), m > 1 (weighting exponent), and $\varepsilon > 0$, terminating threshold.*

***Randomly initialize:** $\boldsymbol{\alpha_{ij}}, \boldsymbol{\beta_{ij}}, \boldsymbol{and}\ \boldsymbol{\gamma_{ij}}$ subject to the given constraints*

for each iteration k =1, 2,.....

***Step1:** Compute cluster mean $\boldsymbol{V_j^{(k)}}, \boldsymbol{j}$=1,2,...,$c$ using equation (17).*

Step2**: Compute $\boldsymbol{d_{ij}^{c\,(k)} = d^{c}(X_i, V_j^{(k)})}$, **i=1,..N, j=1,…,c.

***Step3:** Update $\boldsymbol{\alpha_{ij}}, \boldsymbol{\beta_{ij}}, \boldsymbol{and}\ \boldsymbol{\gamma_{ij}}$ using the equations (14-16) subject to the conditions (11-13).*

***Step4:** Compute cluster mean $\boldsymbol{V_j^{(k+1)}}, \boldsymbol{j}$=1, 2,.., c*

***Step5:** Update $\boldsymbol{V_j^{(k+1)} = [v_1^{(k+1)}, v_2^{(k+1)}, \ldots v_c^{(k+1)}]}$ using equation (17).*

Step6**: if $\|V^{(k)}-V^{(k+1)}\| < \varepsilon$, then go to **Step6
*else let k:=k+1, go to **Step1**.*

***Step7**: End.*

Here, each cluster in the final output cluster set is a PFS consisting of IoT data instances along with a positive membership degree, a neutral membership degree, and a negative membership degree. A data instance that does not belong to any of the clusters or belongs to all the clusters with low positive membership value, high neutral membership value, and high non-membership value can be considered an anomaly. The flowchart of the PFCM clustering algorithm is shown in Fig. 1.

Complexity Analysis

The PFCM clustering algorithm computes the positive, neutral, and negative membership values in constant time, so the complexities of such computations are O (1), O (1), and O (1).

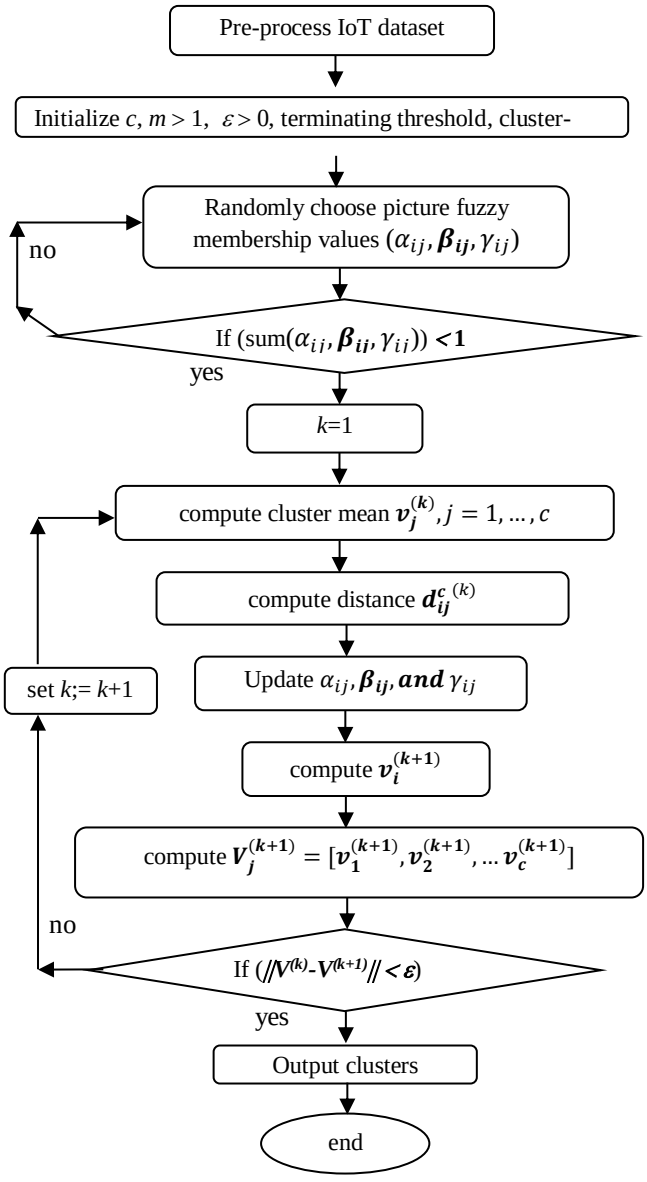

Fig. 1: Flowchart of the PFCM Clustering algorithm

The initialization step takes O($c \cdot N \cdot n$), where c is the number of clusters, N is the number of data instances, and n is the dimension of the IoT dataset. The distance computation also takes constant time. As a result, in each iteration, updating the membership values and cluster centroids requires O ($c \cdot N \cdot n + c \cdot N \cdot n + c \cdot N \cdot n + c \cdot N \cdot n$), while the convergence check requires O(c). Therefore, the total time complexity per iteration is O($c \cdot N \cdot n + c \cdot N \cdot n + c \cdot N \cdot n + c \cdot N \cdot n + c \cdot N \cdot n$ + c) = O($c \cdot N \cdot n$). If t represents the number of iterations, the overall computational complexity of PFCM is O($t \cdot c \cdot N \cdot n$). Assuming c is small and negligible, t = O(N), and $n \leq N$, the worst-case time complexity of the PFCM algorithm is O($N^2 \cdot n$). This shows that the proposed algorithm operates in quadratic time with respect to the dataset size and linear time with respect to the dataset's dimension.

Results and Discussion

Experimental Analysis and Results

To find the efficacy of the proposed approach, the following two well-recognized datasets are employed.

NSL-KDD dataset: This is a refined version of the synthetic dataset KDDCup'99 (Stolfo et al., 1999), constructed by removing the duplicates and redundant instances and extensively used for benchmarking and evaluating intrusion detection systems. The dataset is divided into normal and attack traffic. The attack traffic is further categorized into several types, such as:

- DoS (Denial of Service)
- Probe (Scans or Reconnaissance)
- R2L (Remote to Local)
- U2R (User to Root

It has around 147,000 data instances (normal or anomalous) with 41 features.

Skoltech Anomaly Benchmark (SAB) (Katser and Kozitsin, 2020)**:** The dataset was developed by the Skolkovo Institute of Science and Technology (Skoltech) to support research in anomaly detection, specifically within the areas of network security and system monitoring. SAB offers well-defined evaluation metrics to assess anomaly detection models, which usually consist of the following:

- Precision**:** The proportion of correctly identified anomalies
- Recall**:** The proportion of actual anomalies that are successfully detected
- F1-Score**:** A measure that balances precision and recall

The datasets NSL-KDD and SAB cannot be used directly as input to the proposed method. So, we apply the following preprocessing steps to make them well-suited for the proposed applications. The steps are as follows:

- Data cleaning: This step is intended to remove duplicate or inconsistent entries from the dataset. Since the NSL-KDD dataset is a refined version of the KDD Cup'99 dataset (Stolfo et al., 1999), it does not contain duplicate or inconsistent entries, and therefore, data cleaning is not required. However, for the SAB dataset, this step is necessary to address missing sensor values, which are handled using interpolation or statistical measures such as the mean, median, or mode
- Normalization: This step is necessary to minimize the dominance of any particular feature or attribute. It has been performed using the well-known min-max normalization technique
- Feature selection: Since the entire dataset is being used, feature selection is not required
- Fuzzification: This step transforms the data into the picture fuzzy domain. For each attribute, we define positive, negative, and neutral membership functions and assign corresponding values to each Picture Fuzzy Set (PFS)

Finally, the values of each attribute in the datasets are converted into picture fuzzy vectors for the NSL-KDD dataset and picture fuzzy sequences for the SAB dataset. As a result, both datasets are transformed into a format suitable for input to the proposed model.

The proposed algorithm, along with the classical *k*-means, FCM, and IFCM clustering methods, was implemented in MATLAB using the mentioned datasets on a standard computing machine. A partial graphical representation of the results is shown in Figs. 2-7.

Fig. 2 illustrates the anomaly detection and accuracy of the k-means, FCM, IFCM, and PCM clustering algorithms, evaluated using the NSL-KDD and SAB datasets. The results are presented through bar diagrams, facilitating a comparative performance analysis of the algorithms based on their accuracy.

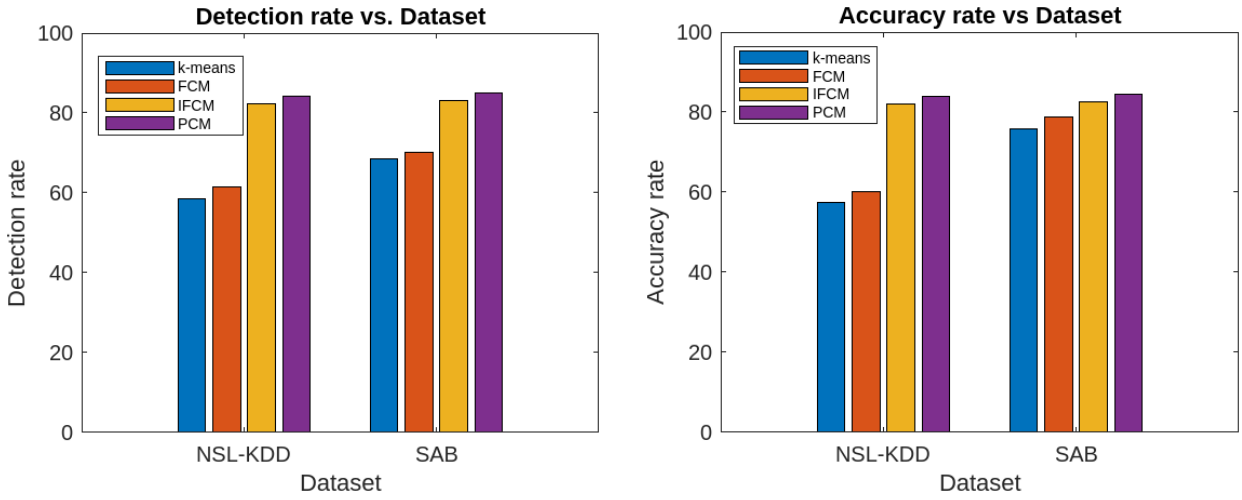

Fig. 2: Comparative analysis of the algorithms on Detection and accuracy rates across the datasets

Fig. 3 presents the false alarm rates of the aforementioned algorithms, evaluated using the NSL-KDD and SAB datasets. This enables a straightforward comparative analysis of the algorithms' performance in terms of false alarm rates.

Fig. 4 presents the Denial of Service (DoS) rates and Remote-to-Local (R2L) rates for the aforementioned algorithms, evaluated using the NSL-KDD and SAB datasets. This allows for an easy comparative performance analysis of the algorithms in terms of DoS rates.

Fig. 5 shows the User-to-Root (U2R) and Probe percentages for the aforementioned algorithms, evaluated using the NSL-KDD and SAB datasets. This facilitates a straightforward comparative analysis of the algorithms based on these parameters.

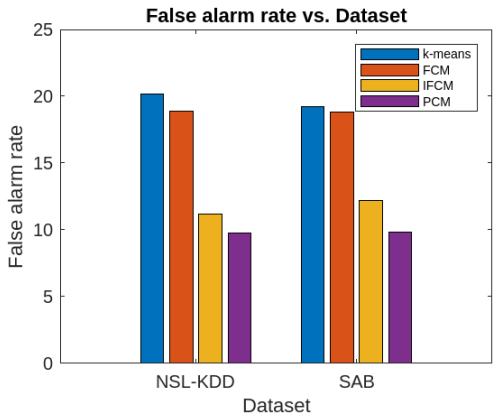

Fig. 3: Comparative analysis of the algorithms on False alarm rates across the datasets

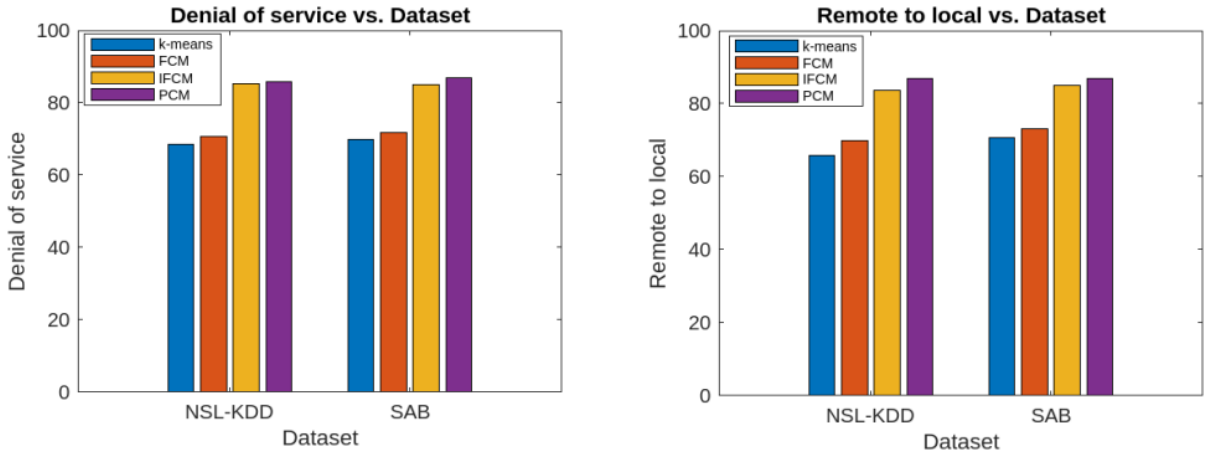

Fig. 4: Comparative analysis of the algorithms on Denial of service and Remote to local across the two datasets

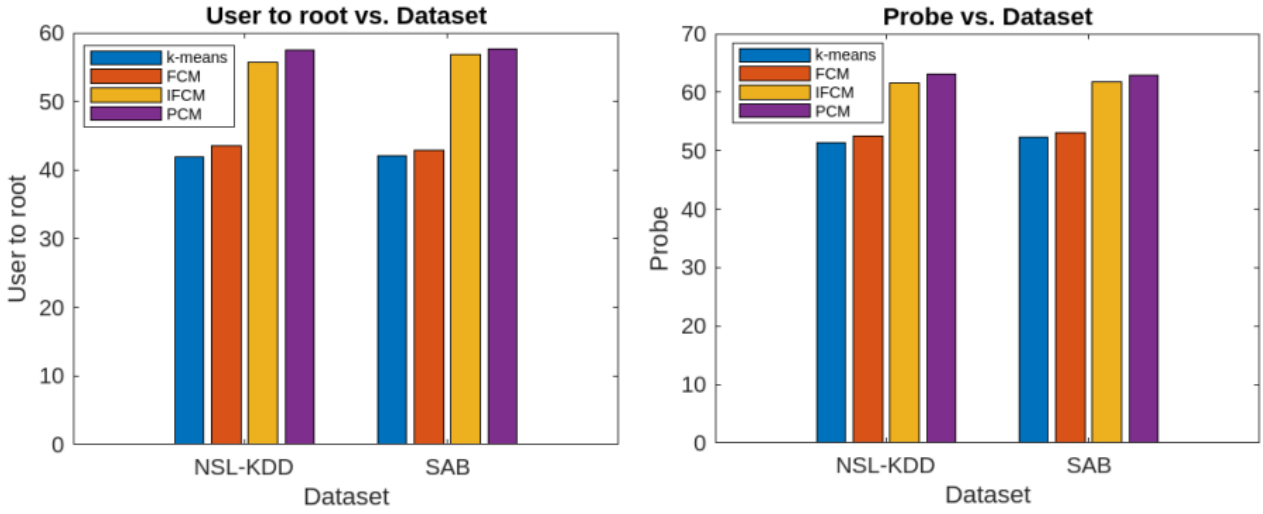

Fig. 5: Comparative analysis of algorithms on User to root and Probe across the two datasets

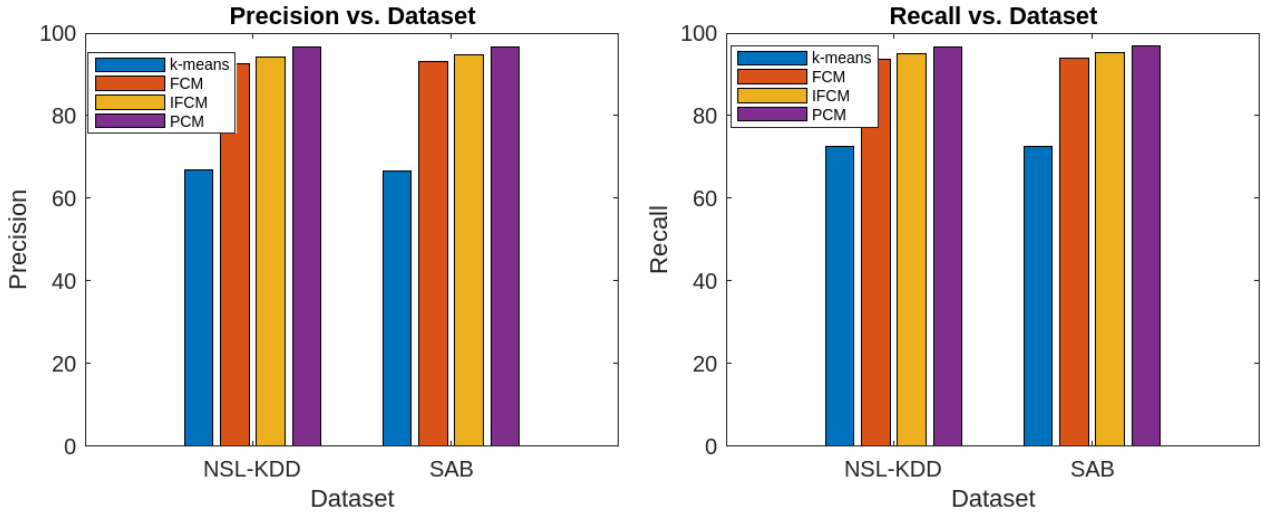

Fig. 6: Comparative analysis of algorithms on Precision and Recall across the two given datasets

Fig. 6 presents the precision and recall values for the aforementioned algorithms, evaluated using the NSL-KDD and SAB datasets. This enables a clear comparative performance analysis of the algorithms based on these parameters.

Fig. 7 displays the F-score of the aforementioned algorithms, evaluated using the NSL-KDD and SAB datasets. This allows for an effective comparative performance analysis of the algorithms based on this parameter.

Similarly, the execution times of the proposed algorithm with respect to the dimensions and sizes of the datasets are presented in Figs. 8 and 9.

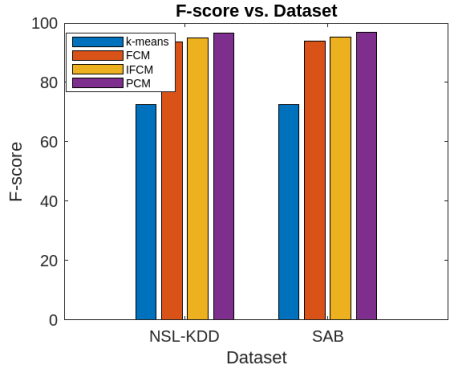

Fig. 7: Comparative analysis of the algorithms on F-score across the datasets

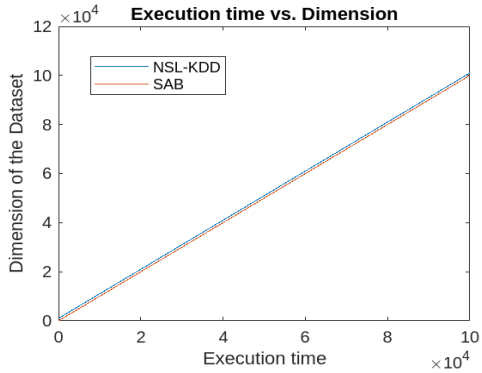

Fig. 8: Execution time with respect to dimension

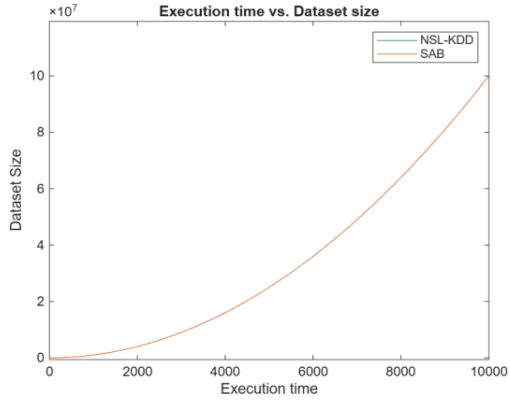

Fig. 9: Execution time with respect to dataset size

Discussion

Based on the results obtained from the proposed algorithm, the following conclusions can be drawn.

For both the NSL-KDD and SAB datasets, the detection rate of the proposed algorithm significantly surpasses that of k-means, FCM, and IFCM. Additionally, the detection rate remains almost identical for both datasets, indicating that the proposed algorithm is the most efficient in terms of detection rate.

Similarly, the accuracy rate of the proposed algorithm is considerably higher for both datasets compared to k-means, FCM, and IFCM. Moreover, the false alarm rate of the proposed algorithm is much lower than that of the other algorithms for both datasets.

Regarding the attack parameters (Denial of Service, Remote-to-Local, User-to-Root, and Probe), the proposed algorithm consistently outperforms the others. In terms of other performance metrics, the proposed method also surpasses *k*-means, FCM, and IFCM clustering algorithms. However, in terms of execution time, the proposed algorithm performs comparably to the FCM and IFCM algorithms.

Conclusion

This article proposes a picture fuzzy clustering-based approach for anomaly detection in the IoT domain. The proposed algorithm, the Picture Fuzzy C-Means (PFCM) clustering algorithm, utilizes a distance measure based on the Canberra metric to form clusters. It generates a predefined number of clusters, where each IoT data instance is associated with a positive, neutral, and negative membership value, all lying between 0 and 1, with their sum also between 0 and 1. An IoT data instance that either does not belong to any cluster, belongs to all clusters with minimal positive membership values, or belongs to all clusters with maximum neutral and negative membership values is considered an anomaly.

The efficacy of the proposed algorithm is demonstrated through experimental studies using the NSL-KDD and SAB datasets, along with a comparative analysis against traditional k-means, FCM, and IFCM algorithms. The proposed PFCM algorithm accomplished an accuracy of 94.6%**,** in comparison with 84.1% for *k*-means**,** 89.2% for FCM, and 91.4% for IFCM on the NSL-KDD dataset. Likewise, on the SAB dataset, PFCM reached an F1-score of 0.91, outperforming *k*-means**,** FCM, and IFCM. Also, the improvements of the proposed PFCM over FCM and IFCM were statistically significant on both datasets. Thus, the results clearly show that the proposed approach outperforms the other methods across all evaluated parameters on both datasets.

The runtime complexity of the proposed algorithm is dependent on the size and dimensions of the datasets. It operates in quadratic time with respect to the dataset size

and linear time with respect to the dataset's dimensions. Since the dataset's dimension is typically much smaller than its size, the overall time complexity of the algorithm is considered quadratic.

The convergence of the proposed PFCM algorithm is achieved after 15-20 iterations on both NSL-KDD and SAB datasets, which is almost the same as that of FCM and IFCM. Also, the curves of convergence show a smooth monotonic decline, demonstrating the stability of the proposed PFCM algorithm. Thus, the proposed PFCM algorithm consistently reached stable membership distributions, even if the datasets contain noisy or ambiguous data instances. It has been found that for $m \in [1.8, 2.2]$, detection accuracy remains stable. However, if there are greater deviations, then the proposed PFCM performance slightly reduces along with slower convergence. The proposed PFCM algorithm is found to be feasible for medium-to-large datasets; its application to very large IoT data can be accomplished through parallel/distributed implementations. Hence, the proposed PFCM clustering-based approach is efficient for IoT anomaly detection.

Limitations and Lines for Future Works

Although the proposed algorithm demonstrates significant efficiency compared to other methods, it still has some limitations. First, like many partitioning-based clustering algorithms, the proposed approach is sensitive to the initial selection of cluster centroids. Second, it struggles with the curse of high dimensionality, which reduces its efficiency when handling high-dimensional data. Lastly, the algorithm may not always converge to the optimal solution, as it can get trapped in local minima.

Future work can focus on the following areas:

- Comparing the proposed method with the state-of-the-art methods
- Developing algorithms to address high dimensionality in IoT datasets
- Exploring alternative approaches beyond unsupervised methods for IoT anomaly detection
- Investigating techniques like bipolar fuzzy or complex fuzzy clustering for IoT anomaly detection

Acknowledgment

The authors express their sincere gratitude to all individuals and institutions who contributed to the successful completion of this research work. We thank our colleagues and reviewers for their valuable comments and suggestions, which have greatly improved the quality of this article. The authors also acknowledge the support of our department and institution for providing the necessary facilities and academic environment for carrying out this study.

Funding Information

The authors declare that this research received no external funding.

Authors Contributions

Fehmin Nadira Laskar: Conceptualization, literature review, and initial development of the PFC framework. Contributed to methodology design and drafted the preliminary version of the manuscript.

Vijo Arul Selvi M.: Formulation of the core algorithmic structure, refinement of the theoretical model, and critical revision of the manuscript, implementation, and validation.

Fokrul Alom Mazarbhuiya: Supervision, Project management, dataset preprocessing, simulation experiments, comparative performance analysis, visualizations, and preparation of the final manuscript draft. Contributed significantly to the result interpretation and discussion.

Mohamed Shenify: Provided insights into IoT system requirements, contributed to application relevance, model evaluation, and validation of anomaly detection scenarios. Assisted in revising the manuscript for clarity and technical accuracy.

Vijay Prasad: Analysis, mathematical formulation, and proof verification. Contributed to editing, technical review, and overall improvement of the manuscript's scientific presentation.

Ethics

The research does not involve any human participation, animal experimentation, or any sensitive personal information.

References

Alghawli, A. S. (2022). Complex methods detect anomalies in real time based on time series analysis. *Alexandria Engineering Journal*, *61*(1), 549–561.https://doi.org/10.1016/j.aej.2021.06.033

Alguliyev, R., Aliguliyev, R., & Sukhostat, L. (2017). Anomaly Detection in Big Data based on Clustering. *Statistics, Optimization & Information Computing*, *5*(4), 325–340. https://doi.org/10.19139/soic.v5i4.365

Alsaedi, A., Moustafa, N., Tari, Z., Mahmood, A., & Anwar, A. (2020). TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems. *IEEE Access, 8*, 165130–165150. https://doi.org/10.1109/access.2020.3022862

Atanassov, K. (1983). Intuitionistic fuzzy sets. In *VII ITKR Session, Sofia* (Vol. 20, Issue S1, pp. S1–S6).

Atadoga, A., Tosanbami Omaghomi, T., Adijat Elufioye, O., Pamela Odilibe, I., Ifesinachi Daraojimba, A., & Rita Owolabi, O. (2024). Internet of Things (IoT) in healthcare: A systematic review of use cases and benefits. *International Journal of Science and Research Archive*, *11*(1), 1511–1517. https://doi.org/10.30574/ijsra.2024.11.1.0243

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2), 191–203. https://doi.org/10.1016/0098-3004(84)90020-7

Butkiewicz, B. S. (2012). *Fuzzy Clustering of Intuitionistic Fuzzy Data*. *7267*. https://doi.org/10.1007/978-3-642-29347-4_25

Carletti, V., Foggia, P., Rosa, F., & Vento, M. (2025). *Enhancing IoT Network Security with Graph Neural Networks for Node Anomaly Detection*. *15444*, 41–51. https://doi.org/10.1007/978-3-031-80507-3_5

Chaira, T. (2011). A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images. *Applied Soft Computing*, *11*(2), 1711–1717. https://doi.org/10.1016/j.asoc.2010.05.005

Chaira, T., & Panwar, A. (2014). An Atanassov's intuitionistic Fuzzy Kernel Clustering for Medical Image segmentation. *International Journal of Computational Intelligence Systems*, *7*(2), 360–370. https://doi.org/10.1080/18756891.2013.865830

Chen, Q., Zhou, M., Cai, Z., & Su, S. (2022). *Compliance Checking Based Detection of Insider Threat in Industrial Control System of Power Utilities*. 1142–1147. https://doi.org/10.1109/acpee53904.2022.9784085

Chenaghlou, M., Moshtaghi, M., Leckie, C., & Salehi, M. (2018). Online Clustering for Evolving Data Streams with Online Anomaly Detection. *Proceedings of the 22nd Pacific-Asia Conference*, *3–6*, 508–521. https://doi.org/10.1007/978-3-319-93037-4_40

Clifford, T. H., & Stephenson, W. (1975). *An Introduction to Numerical Classification*.

Cuong, B. C. (2014). Picture fuzzy sets. *J Comput Sci Cybern*, *30*(4), 409-420. https://doi.org/10.15625/1813-9663/30/4/5032.

Dake, D., K., Kudjo Bada, G., & Ekow Dadzie, A. (2023). Internet of Things (IoT) Applications in Education: Benefits and Implementation Challenges in Ghanaian Tertiary Institutions. *Journal of Information Technology Education: Research*, *22*, 311–338. https://doi.org/10.28945/5183

Emran, S. M., & Ye, N. (2001). Robustness of Canberra Metric in Computer Intrusion Detection. *Proceedings of 2001 IEEE Workshop on Information Assurance and Security, US Military Academy, NY*, 80–84.

Firoozjaei, M. D., Mahmoudyar, N., Baseri, Y., & Ghorbani, A. A. (2022). An evaluation framework for industrial control system cyber incidents. *International Journal of Critical Infrastructure Protection*, *36*, 100487. https://doi.org/10.1016/j.ijcip.2021.100487

Ghorbani, H. (2019). Mahalanobis distance and its application for detecting multivariate outliers. In *Facta Universitatis, Series: Mathematics and Informatics* (Vol. 34, Issue 3, pp. 583–892). https://doi.org/10.22190/fumi1903583g

Gustafson, D., & Kessel, W. (1978). *Fuzzy clustering with a fuzzy covariance matrix* (pp. 761–766). https://doi.org/10.1109/cdc.1978.268028

Habeeb, R. A., Nasaruddin, F., Gani, A., Targio Hashem, I. A., Ahmed, E., & Imran, M. (2019). Real-time big data processing for anomaly detection: A Survey. *International Journal of Information Management*, *45*, 289–307. https://doi.org/10.1016/j.ijinfomgt.2018.08.006

Hahsler, M., Piekenbrock, M., & Doran, D. (2019). dbscan Fast Density-Based Clustering with R. *Journal of Statistical Software*, *91*(1), 1–30. https://doi.org/10.18637/jss.v091.i01

Haldar, N. A. H., Khan, F. A., Ali, A., & Abbas, H. (2017). Arrhythmia classification using Mahalanobis distance based improved Fuzzy C-Means clustering for mobile health monitoring systems. In *Neurocomputing* (Vol. 220, Issue 12, pp. 221–235). https://doi.org/10.1016/j.neucom.2016.08.042

Halstead, B., Koh, Y. S., Riddle, P., Pechenizkiy, M., & Bifet, A. (2023). Combining Diverse Meta-Features to Accurately Identify Recurring Concept Drift in Data Streams. *ACM Transactions on Knowledge Discovery from Data*, *17*(8), 1–36. https://doi.org/10.1145/3587098

Harish, B. S., & Kumar, S. V. A. (2017). Anomaly based Intrusion Detection using Modified Fuzzy Clustering. *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*(6), 54–59. https://doi.org/http://doi.org/10.9781/ijimai.2017.05.002

Hendaoui, F., Meddeb, R., Trabelsi, L., Ferchichi, A., & Ahmed, R. (2025). FLADEN: Federated Learning for Anomaly DEtection in IoT Networks. *Computers & Security*, *155*, 104446. https://doi.org/10.1016/j.cose.2025.104446

Huang, Y., Liu, W., Li, S., Guo, Y., & Chen, W. (2023). A Novel Unsupervised Outlier Detection Algorithm Based on Mutual Information and Reduced Spectral Clustering. *Electronics*, *12*(23), 4864. https://doi.org/10.3390/electronics12234864

Iglesias Vázquez, F., Hartl, A., Zseby, T., & Zimek, A. (2023). Anomaly detection in streaming data: A comparison and evaluation study. *Expert Systems with Applications*, *233*, 120994. https://doi.org/10.1016/j.eswa.2023.120994

Kang, H., & Kang, P. (2024). Transformer-based multivariate time series anomaly detection using inter-variable attention mechanism. *Knowledge-Based Systems*, *290*, 111507. https://doi.org/10.1016/j.knosys.2024.111507

Katser, I., & Kozitsin, I. (2020). *SKAB: Skoltech Anomaly Benchmark*. https://www.kaggle.com/dsv/1693952

Kopawar, N., A & Gajanan Wankhede, K. (2024). Internet of Things in Agriculture : A Review. *International Journal of Scientific Research in Science, Engineering and Technology*, *11*(2), 161–165. https://doi.org/10.32628/ijsrset2411215

Lance, G. N., & Williams, W. T. (1966). Computer Programs for Hierarchical Polythetic Classification ("Similarity Analyses"). *The Computer Journal*, *9*(1), 60–64. https://doi.org/10.1093/comjnl/9.1.60

Lance, G. N., & Williams, W. T. (1967). Mixed-data classificatory programs I. *Agglomerative Systems. Australian Computer Journal*, *1*(1), 15–20.

Masmali, H., F., Miah, S. J., & Noman, N. (2023). Different Applications and Technologies of Internet of Things (IoT). *Proceedings of Seventh International Congress on Information and Communication Technology.*, *464*, 41–54. https://doi.org/10.1007/978-981-19-2394-4_5

Mazarbhuiya, F. A. (2023). Detecting Anomaly using Neighborhood Rough Set based Classification Approach. *ICIC Express Letters*, *17*(1), 73–80.

Mazarbhuiya, F. A., & Abulaish, M. (2012). Clustering Periodic Patterns using Fuzzy Statistical Parameters. *International Journal of Innovative Computing Information and Control (IJICIC*, *8*(3(b)), 2113–2124.

Mazarbhuiya, F. A., & Shenify, M. (2023a). Detecting IoT Anomaly Using Rough Set and Density Based Subspace Clustering. *ICIC Express Letters*, *17*(12), 1395–1403. https://doi.org/10.24507/icicel.17.12.1395.

Mazarbhuiya, F. A., & Shenify, M. (2023b). An Intuitionistic Fuzzy-Rough Set-Based Classification for Anomaly Detection. *Applied Science, MDPI*, *13*(9), 1–21.

Mazarbhuiya, F. A., & Shenify, M. (2023c). Mixed Clustering Approach for Real-Time Anomaly Detection. *Applied Sciences*, *13*, 4151. https://doi.org/10.3390/app13074151

Mazarbhuiya, F. A., & Shenify, M. (2023d). Real-Time Anomaly Detection with Subspace Periodic Clustering Approach. *Applied Sciences*, *13*(13), 1–21. https://doi.org/10.3390/app13137382

Mazarbhuiya, F. A., AlZahrani, M. Y., & Georgieva, L. (2019). Anomaly Detection Using Agglomerative Hierarchical Clustering Algorithm. *Information Science and Applications*, *514*, 475–484. https://doi.org/10.1007/978-981-13-1056-0_48

Mazarbhuiya, F. A., AlZahrani, M., & Mahanta, A. K. (2020). Detecting Anomaly Using Partitioning Clustering with Merging. *ICIC Express Letters*, *14*(10), 951-960. https://doi.org/https://doi.org/10.24507/icicel.14.10.951.

Ren, W., Cao, J., & Wu, X. (2009). *Application of Network Intrusion Detection Based on Fuzzy C-Means Clustering Algorithm*. 19–22. https://doi.org/10.1109/iita.2009.269

Retiti Diop Emane, C., Song, S., Lee, H., Choi, D., Lim, J., Bok, K., & Yoo, J. (2024). Anomaly Detection Based on GCNs and DBSCAN in a Large-Scale Graph. *Electronics*, *13*, 2625. https://doi.org/10.3390/electronics13132625

Samara, M. A., Bennis, I., Abouaissa, A., & Lorenz, P. (2022). A Survey of Outlier Detection Techniques in IoT: Review and Classification. *Journal of Sensor and Actuator Networks*, *11*(1), 1–13. https://doi.org/10.3390/jsan11010004

Sethi, P., & Sarangi, S. R. (2017). Internet of Things: Architectures, Protocols, and Applications. *Journal of Electrical and Computer Engineering*, *2017*, 25. https://doi.org/10.1155/2017/9324035

Shenify, M., & Mazarbhuiya, F. A. (2023). Documents Clustering using Subspace Clustering Algorithm. *ICIC Express Letters*, *17*(12), 1405–1415.

Shenify, M., Mazarbhuiya, F. A., & Wungreiphi, A. S. (2024). Detecting IoT Anomalies Using Fuzzy Subspace Clustering Algorithms. In *Applied Sciences* (Vol. 14, Issue 3, p. 1264). https://doi.org/10.3390/app14031264

Song, H., Jiang, Z., Men, A., & Yang, B. (2017). A Hybrid Semi-Supervised Anomaly Detection Model for High-Dimensional Data. *Computational Intelligence and Neuroscience*, *2017*, 1–9. https://doi.org/10.1155/2017/8501683

Stolfo, S., Fan, W., Lee, W., Prodromidis, A., & Chan, P. (1999). KDD Cup 1999 Data [Dataset. *UCI Machine Learning Repository*. https://archive.ics.uci.edu/dataset/130/kdd+cup+1999+data

Szmidt, E., & Kacprzyk, J. (2000). Distances between intuitionistic fuzzy sets. *Fuzzy Sets and Systems*,*114*(3), 505–518. https://doi.org/10.1016/s0165-0114(98)00244-9

Teh, H. Y., Wang, K. I.-K., & Kempa-Liehr, A. W. (2021). Expect the Unexpected: Unsupervised Feature Selection for Automated Sensor Anomaly Detection. *IEEE Sensors Journal*, *21*(16), 18033–18046. https://doi.org/10.1109/jsen.2021.3084970

Thong, P. H., & Son, L. H. (2016). Picture fuzzy clustering: a new computational intelligence method. *Soft Computing*, *20*(9), 3549–3562. https://doi.org/10.1007/s00500-015-1712-7

Thudumu, S., Branch, P., Jin, J., & Singh, J. (2020). A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of Big Data*, *7*(1), 42. https://doi.org/10.1186/s40537-020-00320-x

Wang, B., Hua, Q., Zhang, H., Tan, X., Nan, Y., Chen, R., & Shu, X. (2022). Research on Anomaly Detection and Real-Time Reliability Evaluation with the Log of Cloud Platform. *Alexandria Engineering Journal*, *61*(9), 7183–7193. https://doi.org/10.1016/j.aej.2021.12.061

Wang, L., Wang, J., Ren, Y., Xing, Z., Li, T., & Xia, J. (2021). A Shadowed Rough-fuzzy Clustering Algorithm Based on Mahalanobis Distance for Intrusion Detection. *Intelligent Automation & Soft Computing*, *29*(3), 31–47. https://doi.org/10.32604/iasc.2021.018577

Xu, Z. (2007). Some similarity measures of intuitionistic fuzzy sets and their applications to multiple attribute decision making. *Fuzzy Optimization and Decision Making*, *6*(2), 109–121. https://doi.org/10.1007/s10700-007-9004-z

Xu, Z. S., (2009). Intuitionistic fuzzy hierarchical clustering algorithms. *Journal of Systems Engineering and Electronics*, *20*(1), 90–97.

Xu, Z., & Wu, J. (2010). Intuitionistic fuzzy C-means clustering algorithms. *Journal of Systems Engineering and Electronics*, *21*(4), 580–590. https://doi.org/10.3969/j.issn.1004-4132.2010.04.009

Younas, M. Z. (2020). Anomaly Detection using Data Mining Techniques: A Review. *International Journal for Research in Applied Science and Engineering Technology*, *8*(11), 568–574. https://doi.org/10.22214/ijraset.2020.32188

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, *1*(1), 3–28. https://doi.org/10.1016/0165-0114(78)90029-5

Zhao, X., Li, Y., & Zhao, Q. (2015). Mahalanobis distance based on fuzzy clustering algorithm for image segmentation. In *Digital Signal Processing* (Vol. 43, Issue 12, pp. 8–16). https://doi.org/10.1016/j.dsp.2015.04.009

Zhao, Z., Birke, R., Han, R., Robu, B., Bouchenak, S., Ben Mokhtar, S., & Chen, L. Y. (2021). Enhancing Robustness of On-line Learning Models on Highly Noisy Data. *IEEE Transactions on Dependable and Secure Computing*, *18*(5), 2177–2192. https://doi.org/10.1109/tdsc.2021.3063947

Zhao, Z., Mehrotra, K. G., & Mohan, C. K. (2018). Online Anomaly Detection Using Random Forest. *Recent Trends and Future Technology in Applied Intelligence*, *10868*, E1–E1. https://doi.org/10.1007/978-3-319-92058-0_87