Journal of Computer Science

Research Article

Detecting Anomalies in IoT Using Intuitionistic Fuzzy Clustering Algorithms

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

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

Article history
Received: 19-07-2025
Revised: 08-09-2025
Accepted: 23-09-2025

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

Abstract: Intuitionistic Fuzzy Sets (IFS) are convenient ways to represent the vagueness and uncertainty inherent in any dataset. There are many uses of these. One such use is the Fuzzy C-Means (FCM) algorithm, which clusters the data objects into a pre-assigned (c) number of fuzzy clusters, which has been adopted in many fields, namely, pattern classification, anomaly detection, fraud identification, etc. One of the important applications is finding anomalies in Internet of Things (IoT) data. IoT is made up of a vast network of digital devices that are constantly producing enormous amounts of data and performing live calculations. Owing to their high susceptibility to the Internet, IoT nodes frequently face issues from illegitimate access, such as intrusions, anomalies, and fraud. Detecting such anomalies in the IoT domain might be a fascinating research challenge. In this article, we propose to develop and evaluate Intuitionistic Fuzzy Clustering (IFC) and Interval-Valued Intuitionistic Fuzzy Clustering (IVIFC) methods for the detection of IoT anomalies by extending the widely acknowledged FCM algorithm and establishing their efficacies through complexity analysis, experimental studies, and comparative analysis.

Keywords: Anomaly Detection, Distance Function, FCM, Fuzzy Clustering Method, Correlation Coefficient, Interval Valued Intuitionistic Fuzzy Sets, Intuitionistic Fuzzy Sets

Introduction

Anomaly detection in the IoT domain has turned out to be an increasingly vital aspect of cybersecurity due to the increase in frequent illegitimate accesses or attacks (Alsaedi et al., 2020). As more people are using IoT devices, there is a huge growth in data generation, which makes attackers more interested in using these devices. Consequently, information security, and in particular anomaly detection, is becoming more and more important as the attraction of IoT devices grows. Finding anomalies in IoT data has a broad spectrum of real-world uses, including fault finding, fraud protection, pre-emptive maintenance, and monitoring. Therefore, in situations where reliable responses are lacking, anomaly detection can offer useful information. In order to address the IoT anomaly, trustworthy solutions to the issues are presented.

IoT refers to interlinked smart gadgets with computing and networking capabilities implanted within the object to perform various tasks (Sethi and Sarangi, 2017). Improving and personalizing the user's interactions and experiences with a range of applications is the main objective of IoT. Significant technological breakthroughs are made possible by IoT in a variety of industries, such as retail, health and fitness, smart cities, logistics, traffic, and agriculture. IoT is often regarded as a global infrastructure that, building upon existing frameworks, architectures, and earlier generations of IoT devices, enables seamless connectivity between the cyber and physical worlds (Sethi and Sarangi, 2017). IoT devices have become an integral part of our everyday lives nowadays. They have been extensively used in various domains such as agriculture (Kopawar Atul and Gajanan Wankhede, 2024) including precision farming, livestock management, smart irrigation systems, and smart farming; healthcare (Atadoga et al., 2024) such as virtual health monitoring, heart rate monitoring, depression monitoring, mood monitoring, ingestible sensors, and robotic surgery; and education (Dake Kwasi et al., 2023) including distance learning, attendance automation and monitoring, smart boards, smart classes, augmented reality, automated student tracking, and

personalized and adaptive learning. IoT devices are also used in other areas (Masmali et al., 2023), such as IoT-enabled cities, home management, logistics, production, and supply network management. Due to their reliance on communications and Internet technology, IoT devices are vulnerable to malicious activities and therefore require a comprehensive approach to safeguard against malicious users and intruders.

There are a couple of approaches proposed for the aforesaid problem, and one such method is clustering-based IoT anomaly detection (Teh et al., 2021; Ren et al., 2009). Clustering has been widely accepted as a process to identify the patterns and distribution of data (Mazarbhuiya and Abulaish, 2012; Shenify and Mazarbhuiya, 2023), and anomaly detection has made extensive use of it. Mazarbhuiya et al. (2019) proposed a hierarchical agglomerative approach for network data anomaly detection. A partitioning and hierarchical-based hybrid approach was presented by Mazarbhuiya et al. (2020) for the mixed data anomaly detection. Similarly, another hybrid approach using a density-based rough set technique was presented in Mazarbhuiya and Shenify (2023a) for the multi-dimensional IoT anomaly detection. A two-phased method consisting of both hierarchical partitioning strategies was proposed in Mazarbhuiya and Shenify (2023), which took into account the temporal features of the datasets for the detection of real-time anomalies. Similar works were presented by Mazarbhuiya and Shenify (2023c-d); 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. (2019); Chenaghlou et al. (2018); Firoozjaei et al. (2022). Mazarbhuiya (2023) discussed the internal threat, which creates major issues for the cybersecurity of process control systems. An online technique for detecting anomalies based on random forests was presented by Chen et al. (2022). Zhao et al. (2018) proposed a detailed review of IoT anomaly detection techniques.

Most of the previously suggested algorithms have certain drawbacks. For example, very few are good at detecting anomalies in IoT data. However, most of the drawbacks can be addressed by incorporating fuzziness in the clustering methods for the following fundamental reasons. First of all, fuzzy clustering facilitates overlapping clusters that are helpful in handling datasets with overlapping class borders, complex structure, or ambiguity. Secondly, owing to the gradual nature of the transition between the clusters, they are more resilient to noise and anomalies. Thirdly, fuzzy clustering provides a more complex perspective on the structure of the data by allowing a comprehensive representation of the link between the data objects and clusters. Samara et al. (2022) presented a novel approach, which increases intrusion detection accuracy by utilizing Mahalanobis distance. Wang et al. (2021) introduced an FCM approach for intrusion detection in network data, employing principal component analysis to detect the most significant discriminative characteristics. Similar studies were conducted by Harish and Kumar (2017); Gustafson and Kessel (1978); Haldar et al. (2017); Zhao et al. (2015); Ghorbani (2019).

Shenify et al. (2024) presented the idea of Intuitionistic Fuzzy Sets (IFS) by associating each element with its membership and non-membership. Occasionally, it is impossible to define exactly the membership grade and non-membership grade of an element; rather, the value ranges or intervals can be given. Such IFSs are referred to as Interval-Valued Intuitionistic Fuzzy Sets (IVIFS). Atanassov (1983), whose elements' membership and non-membership are intervals instead of numbers. The IFS and IVIFS are found to be more useful than conventional fuzzy sets to handle uncertainty, imprecision, and vagueness. Using the concept of hierarchical clustering and the IFS averaging operator in Atanassov and Gargov (1989), a new IFS clustering algorithm was introduced. Xu (2009) introduced IFCM for IFS clustering, which is based on FCM and the basic distance measures on IFSs (Szmidt and Kacprzyk, 2000; Xu and Wu, 2010). Xu (2007) presented Intuitionistic fuzzy hierarchical clustering algorithms an approach using rough set theory and IFSs for network data anomaly detection. They introduced an α-relation derived from the correlation coefficient of IFS to construct intuitionistic fuzzy rules. Mazarbhuiya and Shenify (2023b) proposed a distance-based clustering approach derived from the application of IFS to the possibilistic FCM algorithm.

With the quick propagation of IoT devices across various spheres such as healthcare, agriculture, and smart cities, the complexity and volume of data being produced have increased substantially, making the IoT systems more vulnerable to cyberattacks and anomalies. Traditional anomaly detection methods discussed here often fall short in addressing the uncertainty, vagueness, imprecision, ambiguity, etc., inherent in IoT data. To address this, the paper explores intuitionistic fuzzy and interval-valued intuitionistic fuzzy clustering approaches, which provide a more nuanced and robust mechanism for anomalies by integrating membership, non-membership, and hesitancy. The proposed methods aim to enhance the reliability and accuracy of anomaly detection in IoT environments. The proposed algorithms used the correlation coefficient (Xuan Thao, 2018) as a metric to determine clusters.

Then the paper's objective is described as follows:

- First, the correlation coefficient is defined in terms of the covariance and the mean expressed using the membership and non-membership

- Secondly, using the aforesaid metric, two fuzzy approaches, namely the IFCM and the IVIFCM clustering algorithms, are introduced for the generation of clusters
- Finally, a comparative assessment is carried out with the conventional FCM to determine the efficacy of the suggested approaches

The time complexities of the algorithms are also computed. Then, using MATLAB and the KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018) datasets, the suggested methods are evaluated, and comparisons are performed. The results convincingly show that IFCM and IVIFCM are more effective than FCM, and the IVIFCM algorithm is found to be the best-performing.

Problem Statement

The following section presents key concepts and definitions employed in this article.

Definition 2.1 Fuzzy Set (Zadeh, 1978)

As defined in Zadeh (1978), a fuzzy set A on a universe of discourse $X = \{x_1, x_2, \ldots x_n\}$ is given by:

$$A=\{(x_i, \mu_A(x_i));\ \mu_A(x_i) \in [0,1], x_i \in X\} \quad (1)$$

Here $\mu_A: X \rightarrow [0,1]$ gives the membership value of each element of X in A. Also, $0 < \mu_A(x_i) < 1$ is the partial belongingness of x_i in A. In this article, an attempt is made to establish the role of μ in determining how the IoT data instances belong to a cluster. It helps us to assign IoT data instances to the cluster with a higher membership grade, indicating the level of association.

Definition 2.2 Intuitionistic Fuzzy Set (Atanassov and Gargov, 1989)

Atanassov and Gargov (1989) defined an Intuitionistic Fuzzy Set (IFS) A on X as:

$$A = \{(x_i; \mu_A(x_i), \nu_A(x_i));\ x_i \in X\} \quad (2)$$

Where $\mu_A: X \rightarrow [0,1]$ and $\nu_A: X \rightarrow [0,1]$ are the membership and non-membership functions of A, respectively, subject to the constraint $0 \leq \mu_A(x_i) + \nu_A(x_i) \leq 1\ \forall\ x_i \in X$. Here, ν_A denotes non-membership degree quantifying the extent to which an element does not belong to the set. The idea of an independent non-membership parameter was introduced by Atanassov and Gargov (1989) by extending Zadeh (1978) classical notion: $\nu_A(x_i) = 1 - \mu_A(x_i) \Rightarrow \mu_A(x_i) + \nu_A(x_i) = 1 \forall\ x_i \in X$. Atanassov and Gargov (1989) extended the aforesaid equality with the inequality, $\mu_A(x_i) + \nu_A(x_i) \leq 1$, and the resulting set is referred to as IFS. By taking into account the data instances that are minimally either included in a cluster or not included, ν plays a significant part in the clustering method of IoT data instances. It offers versatility in simulating a scenario in which we are unsure about membership or non-membership. It complements the membership value by capturing, either strongly or weakly, how a data instance might be removed from a cluster. It helps in anomaly detection by improving the clustering technique's capacity to manage the noise. The fundamental requirement of IFS (Atanassov and Gargov, 1989) is the inequality $0 \leq \mu(x) + \nu(x) \leq 1$ that allows an additional degree of uncertainty known as hesitation. Thus, Atanassov's IFS (Atanassov and Gargov, 1989) model is found to be more expressive than Zadeh (1978) model.

In IFS (Atanassov and Gargov, 1989), besides membership and non-membership values, another component is allowed, which is known as hesitation, thus providing a more thorough depiction of uncertainty associated with many real-life problems. Zadeh (1978) fuzzy sets and crisp sets could not conveniently characterize IoT models, as they produced inconsistent, incomplete, ambiguous, vague, and uncertain data. The IFS (Zadeh, 1978) model can accomplish that commendably.

In order to effortlessly depict IoT uncertainty and vagueness caused by malfunction, imprecise data, errors caused by communication technology, interference due to intermediary media, anomalies, and noises, the IFS is evolved.

Definition 2.3

For an IFS A defined over X, if $\xi_A(x_i) = 1 - \mu_A(x_i) - \nu_A(x_i)\ \forall x_i \in X$, then $\xi_A(x_i)$ is known as a hesitation function representing the level of hesitation of x_i in A such that $0 \leq \xi_A(x_i) \leq 1$. It represents the amount of hesitation in classifying x_i.

If $\xi_A(x_i) = 0\ \forall x_i \in X$, then $\nu_A(x_i) = 1 - \mu_A(x_i)$, and the IFS A takes the form of a fuzzy set (Zadeh., 1978). Furthermore, if $\xi_A(x_i) = 0 = \nu_A(x_i)\ \forall x_i \in X$, then A becomes a crisp set.

Definition 2.4 Correlation Coefficient Between Two IFSs

Let us take two IFSs, $A=\{(x_i; (\mu_A(x_i), \nu_A(x_i)), x_i \in X\}$ and $B=\{(x_i; (\mu_B(x_i), \nu_B(x_i)), x_i \in X\}$, over $X=\{x_1, x_2, \ldots x_n\}$, where each A and B are represented by respective membership and non-membership functions $\mu_A, \mu_A: X \rightarrow [0,1]$, $\nu_B, \nu_B: X \rightarrow [0,1]$. Also, let $\lambda_A(x_i) =$

$\mu_A(x_i)$-$\nu_A(x_i)$ and $\lambda_B(x_i) = \mu_B(x_i)$-$\nu_B(x_i)$; $x_i \in X$. According to Xuan Thao (2018), the correlation coefficient between *A* and *B* is expressed mathematically as:

$$\rho(A,B) = \frac{cov(A,B)}{\sqrt{cov(A,A).cov(B,B)}} \quad (3)$$

where $\frac{\sum_{i=1}^{n}(\frac{(\lambda_A(x_i)-\overline{\lambda_A(x)}}{2}).(\frac{(\lambda_B(x_i)-\overline{\lambda_B(x)}}{2})}{n-1}$ = covariance of *A* and *B*, (4)

$$cov(A,A) = \frac{\sum_{i=1}^{n}(\frac{(\lambda_A(x_i)-\overline{\lambda_A(x_i)}}{2})^2}{n-1} = \text{covariance of } A. \quad (5)$$

And:

$$cov(B,B) = \frac{\sum_{i=1}^{n}\left(\frac{(\lambda_B(x_i)-\overline{\lambda_B(x_i)}}{2}\right)^2}{n-1} = .\text{covariance of } B \quad (6)$$

$$\text{Mean } (A) = \overline{\lambda_A(x)} = \frac{\sum_{i=1}^{n}\lambda_A(x_i)}{n} = \frac{\sum_{i=1}^{n}(\mu_A(x_i)-\nu_A(x_i))}{n} \quad (7)$$

$$\text{Mean } (B) = \overline{\lambda_B(x)} = \frac{\sum_{i=1}^{n}\lambda_B(x_i)}{n} = \frac{\sum_{i=1}^{n}(\mu_B(x_i)-\nu_B(x_i))}{n} \quad (8)$$

Using (4), (5), and (6), (3) can be expressed as:

$$\rho(A,B) = \frac{\frac{\sum_{i=1}^{n}(\frac{(\lambda_A(x_i)-\overline{\lambda_A(x)}}{2}).(\frac{(\lambda_B(x_i)-\overline{\lambda_B(x)}}{2})}{n-1}}{\sqrt{\frac{\sum_{i=1}^{n}\left(\frac{(\lambda_A(x_i)-\overline{\lambda_A(x_i)}}{2}\right)^2}{n-1}.\frac{\sum_{i=1}^{n}\left(\frac{(\lambda_B(x_i)-\overline{\lambda_B(x_i)}}{2}\right)^2}{n-1}}}$$

$$= \frac{\sum_{i=1}^{n}(\lambda_A(x_i)-\overline{\lambda_A(x_i)})(\lambda_B(x_i)-\overline{\lambda_B(x)})}{\sqrt{\sum_{i=1}^{n}(\lambda_A(x_i)-\overline{\lambda_A(x_i)})^2.\sum_{i=1}^{n}(\lambda_B(x_i)-\overline{\lambda_B(x_i)})^2}} \quad (9)$$

Theorem 1. If $A = \{(x_i, \mu_A(x_i), \nu_A(x_i)); x_i \in X\}$ and $B = \{(x_i; \mu_B(x_i), \nu_B(x_i)); x_i \in X\}$ are two IFSs, then $-1 \leq \rho(A,B) \leq 1$.

Definition 2.5 Inter-Valued Intuitionistic Fuzzy Set (Atanassov and Gargov, 1989)

Atanassov and Gargov (1989) defined an interval-valued intuitionistic fuzzy set (IVIFS) *A* over $X=\{x_1, x_2,\ldots x_n\}$ as:

$$A = \{(x_i; \bar{\mu}_A(x_i), \bar{\nu}_A(x_i)); x_i \in X\} \quad (10)$$

Here the membership $\bar{\mu}_A(x_i) = [\bar{\mu}_A^L(x_i), \bar{\mu}_A^U(x_i)] \subset [0,1]$ and the non-membership $\bar{\nu}_A(x_i) = [\bar{\nu}_A^L(x_i), \bar{\nu}_A^U(x_i)] \subset [0,1]$ are interval-valued functions, and $\bar{\mu}_A^L(x_i) = inf\bar{\mu}_A(x_i)$, $\bar{\mu}_A^U(x) = sup\bar{\mu}_A(x_i)$, $\bar{\nu}_A^L(x_i) = inf\bar{\nu}_A(x_i)$, $\bar{\nu}_A^U(x) = sup\bar{\nu}_A(x_i)$, satisfying the conditions $\bar{\mu}_A^L(x) + \bar{\nu}_A^L(x) \geq 0$ and $\bar{\mu}_A^U(x) + \bar{\nu}_A^U(x) \leq 1$ for $x_i \in X$. If $\bar{\mu}_A^L(x_i) = \bar{\mu}_A^U(x_i)$ and $\bar{\nu}_A^L(x_i) = \bar{\nu}_A^U(x_i)$, then IVIFS becomes IFS.

Definition 2.6 Correlation Coefficient Between Two IVIFSs

Let us take two IVIFSs, $A = \{(x_i; \bar{\mu}_A(x_i), \bar{\nu}_A(x_i)); x_i \in X\}$, and $B = \{(x_i; \bar{\mu}_B(x_i), \bar{\nu}_B(x_i)); x_i \in X\}$ over $X = \{x_1, x_2,\ldots x_n\}$ as defined in the *Definition 2.5*. Also, let $\bar{\lambda}_A(x_i) = \bar{\mu}_A(x_i) - \bar{\nu}_A(x_i) = [\bar{\mu}_A^L(x_i), \bar{\mu}_A^U(x_i)] - [\bar{\nu}_A^L(x_i), \bar{\nu}_A^U(x_i)] = [\bar{\mu}_A^L(x_i) - \bar{\nu}_A^U(x_i), \bar{\mu}_A^U(x_i) - \bar{\nu}_A^L(x_i)] = [\bar{\lambda}_A^L(x_i), \bar{\lambda}_A^U(x_i)], \bar{\lambda}_B(x_i) = \bar{\mu}_B(x_i) - \bar{\nu}_B(x_i) = [\bar{\mu}_B^L(x_i), \bar{\mu}_B^U(x_i)] - [\bar{\nu}_B^L(x_i), \bar{\nu}_B^U(x_i)] = [\bar{\mu}_B^L(x_i) - \bar{\nu}_B^U(x_i), \bar{\mu}_B^U(x_i) - \bar{\nu}_B^L(x_i)] = [\bar{\lambda}_B^L(x_i), \bar{\lambda}_B^U(x_i)]$; $x_i \in X$. Then, the correlation coefficient (Xuan Thao, 2018) of *A* and *B* is given by:

$$\rho(A,B) = \frac{cov(A,B)}{\sqrt{cov(A,A).cov(B,B)}} \quad (11)$$

Where:

$$cov(A,B) = \frac{\sum_{i=1}^{n}\left(\frac{\bar{\lambda}_A^L(x_i)-\overline{\bar{\lambda}_A^L(x)}}{2}\right).\left(\frac{\bar{\lambda}_A^U(x_i)-\overline{\bar{\lambda}_A^U(x)}}{2}\right).\left(\frac{\bar{\lambda}_B^L(x_i)-\overline{\bar{\lambda}_B^L(x)}}{2}\right).\left(\frac{\bar{\lambda}_B^U(x_i)-\overline{\bar{\lambda}_B^U(x)}}{2}\right)}{(n-1)} \quad (12)$$

Is the covariance of IVIFSs *A* and *B*:

$$cov(A,A) = \frac{\sum_{i=1}^{n}\left(\frac{\bar{\lambda}_A^L(x_i)-\overline{\bar{\lambda}_A^L(x)}}{2}\right).\left(\frac{\bar{\lambda}_A^U(x_i)-\overline{\bar{\lambda}_A^U(x)}}{2}\right).\left(\frac{\bar{\lambda}_A^L(x_i)-\overline{\bar{\lambda}_A^L(x)}}{2}\right).\left(\frac{\bar{\lambda}_A^U(x_i)-\overline{\bar{\lambda}_A^U(x)}}{2}\right)}{n-1} =$$

$$\frac{\sum_{i=1}^{n}\left(\frac{\bar{\lambda}_A^L(x_i)-\overline{\bar{\lambda}_A^L(x)}}{2}\right)^2.\left(\frac{\bar{\lambda}_A^U(x_i)-\overline{\bar{\lambda}_A^U(x)}}{2}\right)^2}{n-1}$$ is the covariance of the IVIFS *A* (13)

And:

$$cov(B,B) = \frac{\sum_{i=1}^{n}\left(\frac{\bar{\lambda}_B^L(x_i)-\overline{\bar{\lambda}_B^L(x)}}{2}\right)^2.\left(\frac{\bar{\lambda}_B^U(x_i)-\overline{\bar{\lambda}_B^U(x)}}{2}\right)^2}{n-1}$$ is the covariance of the IVIFS *B* (14)

$$\text{Mean } (A) = \left[\overline{\bar{\lambda}_A^L(x)}, \overline{\bar{\lambda}_A^U(x)}\right] = \left[\frac{\sum_{i=1}^{n}\bar{\lambda}_A^L(x_i)}{n}, \frac{\sum_{i=1}^{n}\bar{\lambda}_A^U(x_i)}{n}\right]$$

$$= \left[\frac{\sum_{i=1}^{n}(\bar{\mu}_A^L(x_i)-\bar{\nu}_A^U(x_i))}{n}, \frac{\sum_{i=1}^{n}(\bar{\mu}_A^U(x_i)-\bar{\nu}_A^L(x_i))}{n}\right] \quad (15)$$

$$\text{Mean } (B) = \left[\overline{\bar{\lambda}_B^L(x)}, \overline{\bar{\lambda}_B^U(x)}\right] = \left[\frac{\sum_{i=1}^{n}\bar{\lambda}_B^L(x_i)}{n}, \frac{\sum_{i=1}^{n}\bar{\lambda}_B^U(x_i)}{n}\right]$$

$$= \left[\frac{\sum_{i=1}^{n}(\bar{\mu}_B^L(x_i)-\bar{\nu}_B^U(x_i))}{n}, \frac{\sum_{i=1}^{n}(\bar{\mu}_B^U(x_i)-\bar{\nu}_B^L(x_i))}{n}\right] \quad (16)$$

Using (13), (14), and (15), (12) becomes:

$$\rho(A,B)=\frac{\frac{\sum_{i=1}^{n}\left(\frac{\bar{\lambda}_A^L(x_i)-\overline{\bar{\lambda}_A^L(x)}}{2}\right).\left(\frac{\bar{\lambda}_A^U(x_i)-\overline{\bar{\lambda}_A^U(x)}}{2}\right).\left(\frac{\bar{\lambda}_B^L(x_i)-\overline{\bar{\lambda}_B^L(x)}}{2}\right).\left(\frac{\bar{\lambda}_B^U(x_i)-\overline{\bar{\lambda}_B^U(x)}}{2}\right)}{n-1}}{\sqrt{\frac{\sum_{i=1}^{n}\left(\frac{\bar{\lambda}_A^L(x_i)-\overline{\bar{\lambda}_A^L(x)}}{2}\right)^2.\left(\frac{\bar{\lambda}_A^U(x_i)-\overline{\bar{\lambda}_A^U(x)}}{2}\right)^2}{n-1}.\frac{\sum_{i=1}^{n}\left(\frac{\bar{\lambda}_B^L(x_i)-\overline{\bar{\lambda}_B^L(x)}}{2}\right)^2.\left(\frac{\bar{\lambda}_B^U(x_i)-\overline{\bar{\lambda}_B^U(x)}}{2}\right)^2}{n-1}}}$$
$$=\frac{\sum_{i=1}^{n}(\bar{\lambda}_A^L(x_i)-\overline{\bar{\lambda}_A^L(x)})(\bar{\lambda}_A^U(x_i)-\overline{\bar{\lambda}_A^U(x)})(\bar{\lambda}_B^L(x_i)-\overline{\bar{\lambda}_B^L(x)})(\bar{\lambda}_B^U(x_i)-\overline{\bar{\lambda}_B^U(x)})}{\sqrt{\sum_{i=1}^{n}\left(\bar{\lambda}_A^L(x_i)-\overline{\bar{\lambda}_A^L(x)}\right)^2\left(\bar{\lambda}_A^U(x_i)-\overline{\bar{\lambda}_A^U(x)}\right)^2.\sum_{i=1}^{n}\left(\bar{\lambda}_B^L(x_i)-\overline{\bar{\lambda}_B^L(x)}\right)^2\left(\bar{\lambda}_B^U(x_i)-\overline{\bar{\lambda}_B^U(x)}\right)^2}} \quad (17)$$

Example 1. Let us take a universe of discourse $X=\{x_1, x_2, x_3\}$. Let us take two IVIFSs, *A* and *B*, over *X* as:

$$A=\{(x_1,[0.2,0.4],[0.3,0.5]),(x_2,[0.3,0.4],[0.2,0.5]),(x_3,[0.1,0.5],[0.2,0.4])\} \quad (18)$$

And:

$$B=\{\{(x_1,[0.1,0.4],[0.2,0.4]),(x_2,[0.2,0.3],[0.3,0.4]),(x_3,[0.2,0.6],[0.1,0.3])\} \quad (19)$$

Then, the correlation coefficient between *A* and *B* is expressed as:

$$(A,B)=\frac{\begin{matrix}(-0.3-(-0.2667))(0.1-0.2)(-0.3-(-0.2))(0.2-0.1667)+\\(-0.2-(-0.2667))(0.2-0.2)(-0.2-(-0.2))(0-0.1667)\\+(-0.3+(-0.2667))(0.3-0.2)(0.1-(-0.2))(0.5-0.1667)\end{matrix}}{\sqrt{\begin{matrix}\left[\begin{matrix}(-0.3-(-0.2667))^2(0.1-0.2)^2+(-0.2+0.2667)^2(0.2-0.2)^2\\+(-0.3+0.2667)^2(0.3-0.2)^2\end{matrix}\right]\\\left[\begin{matrix}(-0.3+0.2)^2(0.2-0.1667)^2+(-0.2+0.2)^2(0-0.1667)^2\\+(0.1+0.2)^2(0.5-0.1667)^2\end{matrix}\right]\end{matrix}}}$$
$$=\frac{-0.0001220778}{\sqrt{(0.0000110889+0+0.0000110889)\times(0.0000110889+0+0.001110889)}}=-0.0036090604 \quad (20)$$

Theorem 2. If $A=\{(x_i;\ \bar{\mu}_A(x_i),\ \bar{\nu}_A(x_i));\ x_i\in X\}$ and $B=\{(x_i;\ \bar{\mu}_B(x_i),\ \bar{\nu}_B(x_i));\ x_i\in X\}$ are two IVIFSs over *X*, then $-1\le\rho(A,B)\le 1$.

Definition 2.7

An *n*-dimensional vector $x_i=[x_{i1}, x_{i2},\ldots x_{in}]$, $x_i\in R^n$ is made up of *n* measured variables for each IoT data instance. $X=\{x_i;\ i=1, 2,\ldots N\}$ represents a set of *N* data instances, which can be expressed by $N\times n$ matrix as follows:

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

The following matrix represents the intuitionistic fuzzy clustering, which is the partitioning space of IFS for *X*:

$$Fi_{fc}=\{[\boldsymbol{\mu_{ij}},\ \boldsymbol{\nu_{ij}}]_{c\times n};\ \boldsymbol{\mu_{ij}},\ \boldsymbol{\nu_{ij}}\in[0,\ 1],\ \forall\ i,\ j;\ \sum_{i=1}^{c}\boldsymbol{\mu_{ij}}=\boldsymbol{1},\ \sum_{i=1}^{c}\boldsymbol{\nu_{ij}}=\boldsymbol{1},\ \boldsymbol{0}<\sum_{j=1}^{n}\boldsymbol{\mu_{ij}},\sum_{j=1}^{n}\boldsymbol{\nu_{ij}}<\boldsymbol{N}\ \forall\ i\} \quad (22)$$

Where, $\boldsymbol{\mu_{ij}}$, $\boldsymbol{\nu_{ij}}$, the $j^{\text{-th}}$ column of Fi_{fc}, provides the membership and non-membership values of the $i^{\text{-th}}$ cluster.

Methods

In this section, the two algorithms, namely, the IFCM clustering algorithm and IVIFCM clustering algorithm, are proposed.

IFCM Clustering Algorithm

We formulate our IFCM for IFSs using the correlation coefficient instead of the distance norm (the definition of the correlation coefficient is given in Section 2). One can formulate IFCM's objective function in the following manner:

$$\boldsymbol{J(X;F,V)=\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^m\rho(x_j,v_i)^2} \quad (23)$$

Where $F\in Fi_{fc}$ and $V=[v_1, v_2,\ldots v_c]$, $v_i\in R^n$ cluster's mean vector must be evaluated.

$\boldsymbol{D_{ij}^2=\rho(x_j,v_i)^2}$ represents the squared correlation coefficient, and $m\in[1,\infty]$ denotes the fuzziness parameter of the resulting cluster.

By solving the optimization problem (24) using Picard's iterative approach, the first-order stationary points conditions are obtained that lead to the IFCM Algorithm Xu and Wu (2010).

By introducing constraints to *J* with Lagrange's multipliers, the stationary points of (24) can be found:

$$J(X;F,V,\lambda)=\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^m D_{ij}^2+\sum_{j=1}^{N}\lambda_j\left[\sum_{i=1}^{c}\mu_{ij}-1\right] \quad (24)$$

By equating to 0, the partial derivatives of *J* with respect to *F*, *V*, and λ, and using the conditions if $\boldsymbol{D_{ij}^2>0}$ $\forall$ *i*, *j*, and *m*>1, then $\boldsymbol{(F,V)\in F_{ifc}\times R^{c\times n}}$, then we get the following:

For $\boldsymbol{1\le i\le c}$, $\boldsymbol{1\le j\le N}$, we have $\frac{\boldsymbol{\delta J}}{\boldsymbol{\delta\mu_{ij}}}=\boldsymbol{m\mu_{ij}^{m-1}D_{ij}^2+\lambda_j}$

Therefore, $\frac{\boldsymbol{\delta J}}{\boldsymbol{\delta\mu_{ij}}}=\boldsymbol{0\Longrightarrow m\mu_{ij}^{m-1}D_{ij}^2+\lambda_j=0\Longrightarrow\mu_{ij}^{m-1}=-\frac{\lambda_j}{mD_{ij}^2}}$

$\Rightarrow \mu_{ij}^{m-1} = \frac{m\alpha_j}{mD_{ij}^2}$, where $\alpha_j = -m\lambda_j$ (since $\lambda_j < 0$ and $\mu_{ij} > 0 \; \forall \; i = 1, 2, .., c, \text{and } j = 1, 2, \ldots, N$.

$$\Rightarrow \mu_{ij} = \left(\frac{\alpha_j}{D_{ij}^2}\right)^{\frac{1}{m-1}}$$

Using the above result in $\sum_{i=1}^{c} \mu_{ij} = 1$, we get $\sum_{i=1}^{c} \left(\frac{\alpha_j}{D_{ij}^2}\right)^{1/(m-1)} = 1$

$$\Rightarrow \alpha_j^{1/(m-1)} \sum_{i=1}^{c} \left(\frac{1}{D_{ij}^2}\right)^{1/(m-1)} = 1$$

$$\Rightarrow \alpha_j^{\frac{1}{m-1}} = \frac{1}{\sum_{i=1}^{c} \left(\frac{1}{D_{ij}}\right)^{\frac{2}{m-1}}}$$

Therefore, $\mu_{ij} = \left(\frac{\alpha_j}{D_{ij}^2}\right)^{1/(m-1)} = \frac{(\alpha_j)^{1/(m-1)}}{(D_{ij}^2)^{1/(m-1)}} = \frac{\frac{1}{\sum_{i=1}^{c}\left(\frac{1}{D_{ij}}\right)^{2/(m-1)}}}{(D_{ij}^2)^{1/(m-1)}} = \frac{1}{\sum_{r=1}^{c}\left(\frac{D_{ij}}{D_{rj}}\right)^{2/(m-1)}}$ is one of the conditions of optimization of (24). Similarly, other conditions can be obtained by differentiating J partially with respect to V and equating it to 0.

Hence, J will be minimized only if:

$$\mu_{ij} = \frac{1}{\sum_{r=1}^{c}\left(D_{ij/D_{rj}}\right)^{2/(m-1)}}, 1 \le i \le c, \; 1 \le j \le N \qquad (26)$$

And:

$$v_i = \frac{\sum_{j=1}^{N}(\mu_{ij})^m x_j}{\sum_{j=1}^{N}(\mu_{ij})^m}, \; 1 \le i \le c \qquad (27)$$

The first-order necessary conditions for the existence of stationary points of the objective function (24) are represented by Equations (26) and (27) subject to the constraint (23).

Algorithm 1: IFCM

For dataset X, select c clusters ($1 < c < N$), a fuzziness parameter $m > 1$, a convergence threshold $\varepsilon > 0$, $D_{ij} = \rho(x_j, v_i)$ is the correlation coefficient between data instance x_j and cluster–mean v_i, where $v_i \in R^n$, $(1 \le i \le c)$, μ_{ij} is the membership value of j-th data instance to the i-th cluster-mean v_i, $F=(\mu_{ij})_{c\times N}$ and $V= [v_1, v_2, \ldots v_c]$ is a matrix of $c \times n$.

Initialize cluster-means $V=V^{(0)}$ and $F^{(0)}$

for each k =1, 2, ...

Step 1 compute cluster mean $v_i^{(k)} = \frac{\sum_{j=1}^{N}\left(\mu_{ij}^{(k-1)}\right)^m x_j}{\sum_{j=1}^{N}\left(\mu_{ij}^{(k-1)}\right)^m}$, i=1, 2,..c

Step 2 compute $D_{ij}^{(k)} = \rho(x_j, v_i^{(k)})$, **i=1, 2,..c, j=1, 2,...N**

if $\forall j, p, D_{ij}^{(k)} > 0$

$$\mu_{ij}^{(k)} = \frac{1}{\sum_{r=1}^{c}\left(\frac{D_{ij}^{(k)}}{D_{pj}^{(k)}}\right)^{2/(m-1)}}, \; 1 < i < c, \; 1 < j < N$$

else if $\exists$ *j, p such that* $D_{pj}^{(k)} = 0$ *then let* $\mu_{pj}^{(k)} = 1$ *and* $\mu_{ij}^{(k)} = 0$ *for* $i \neq p$

Step 3 compute $F^{(k)} = (\mu_{ij}^{(k)})_{c\times N}$

Step 4 compute $V^{(k+1)} = [v_1^{(k+1)}, v_2^{(k+1)}, \ldots v_c^{(k+1)}]$ *where* $v_i^{(k+1)} = \frac{\sum_{j=1}^{N}\left(\mu_{ij}^{(k)}\right)^m x_j}{\sum_{j=1}^{N}\left(\mu_{ij}^{(k)}\right)^m}$, i=1, 2,..c

Step 5 if $\|V^{(k)} - V^{(k+1)}\| < \varepsilon$, *then go to step6*

else let k:=k+1, go to step1.

Step 6 End.

In this context, each cluster belonging to the output set is represented by an IFS comprising IoT data. Any IoT data instance belonging to a cluster can have any of the following options: Both membership and non-membership values can be high or low simultaneously, or membership value high and non-membership value low or vice-versa. A data instance belonging to all the clusters with minimum membership value and maximum non-membership value, or minimum membership and non-membership value, can be viewed as an anomaly. The IFCM's flowchart is shown in Figure 1.

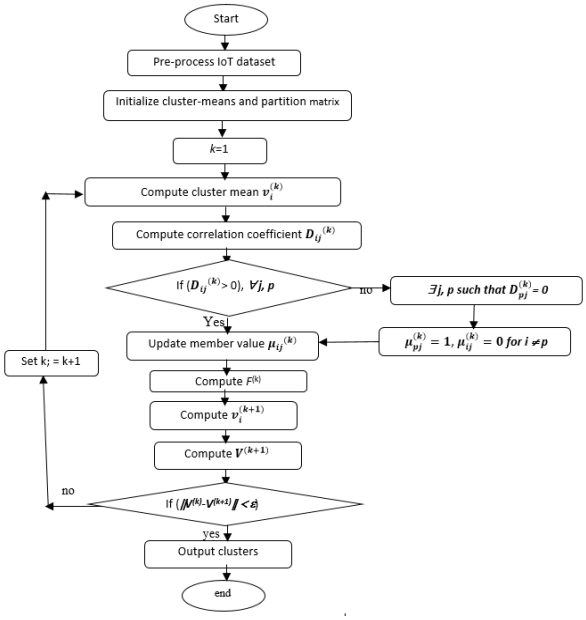

Fig. 1: Flowchart of the IFCM algorithm

IVIFCM Clustering Algorithm

We formulate our IVIFCM clustering algorithm for the IVIFSs in a similar fashion. The distance function is represented by Equation (18). The objective function for IVIFCM can be expressed as:

$$\text{Minimize } \boldsymbol{J(X;F,V)} = \sum_{i=1}^{c}\sum_{j=1}^{N}(\boldsymbol{\mu_{ij}})^{m}\boldsymbol{\rho}(\overline{\boldsymbol{X}_j},\boldsymbol{v}_i)^2 \quad (28)$$

Where $\overline{\boldsymbol{X}}$= $\{\overline{\boldsymbol{X_1}}, \overline{\boldsymbol{X_2}},\ldots.\overline{\boldsymbol{X_N}}\}$ are N IVIFSs, each of which has N elements $\{x_1, x_2,\ldots x_N\}$, and the number of clusters is c ($1 < c < N$), and $V = [v_1, v_2,\ldots v_c]$, $v_i \in R^n$ is the cluster prototype-vector of IVIFSs. The $m > 1$ is the fuzziness parameter, $\boldsymbol{\mu_{ij}}$ is the membership value of *the j*-th data instance $\overline{\boldsymbol{X}_j}$ to the *i*-th cluster, and $F=\{\boldsymbol{\mu_{ij}}\}_{c\times N}$. For finding the stationary points of (28), the Lagrange multiplier is employed in (28). Let:

$$\boldsymbol{J'} = \sum_{i=1}^{c}\sum_{j=1}^{N}(\boldsymbol{\mu_{ij}})^{m}\boldsymbol{D_{ij}^2} - \sum_{j=1}^{N}\boldsymbol{\lambda_j}\left[\sum_{i=1}^{c}\boldsymbol{\mu_{ij}} - \boldsymbol{1}\right] \quad (29)$$

Where $\boldsymbol{D_{ij}^2}$= $\boldsymbol{\rho}(\overline{\boldsymbol{X}_j}, \boldsymbol{v}_i)^2$, which is expressed in (18).

By equating to 0, the partial derivatives of J' with respect to F and λ_j, we obtain the following minimization conditions:

$$\boldsymbol{\mu_{ij}} = \frac{\boldsymbol{1}}{\sum_{r=1}^{c}\left(\boldsymbol{D_{ij/D_{rj}}}\right)^{2/(m-1)}}, 1\le i \le c,\ 1 \le j \le N \quad (30)$$

$$\boldsymbol{v_i} = \boldsymbol{f}(\overline{\boldsymbol{Z}}) = \left\{\left(\boldsymbol{x_l}, \left[\sum_{j=1}^{N}\boldsymbol{\mu}_{\overline{\boldsymbol{X}}_j}^{\boldsymbol{L}}(\boldsymbol{x_l}), \sum_{j=1}^{N}\boldsymbol{\mu}_{\overline{\boldsymbol{X}}_j}^{\boldsymbol{U}}(\boldsymbol{x_l})\right], \left[\sum_{j=1}^{N}\boldsymbol{v}_{\overline{\boldsymbol{X}}_j}^{\boldsymbol{L}}(\boldsymbol{x_l}), \sum_{j=1}^{N8}\boldsymbol{v}_{\overline{\boldsymbol{X}}_j}^{\boldsymbol{U}}(\boldsymbol{x_l})\right]\right), \boldsymbol{1} < \boldsymbol{l} < \boldsymbol{n}\right\},\ 1\le i \le c. \quad (31)$$

Algorithm 2: IVIFCM

Initialize cluster-means $V=V^{(0)}$ let $k = 0$ and set $\varepsilon > 0$

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

Step 1 compute $\boldsymbol{F^{(k)}} = (\boldsymbol{\mu_{ij}^{(k)}})_{\boldsymbol{c\times N}}$, *where*

(a) If $\forall j, r, \boldsymbol{\rho}\left(\overline{\boldsymbol{X}_j}, \boldsymbol{v_i^{(k)}}\right) > \boldsymbol{0}$, *then*

$$\boldsymbol{\mu_{ij}^{(k)}} = \frac{\boldsymbol{1}}{\sum_{r=1}^{c}\left(\frac{\boldsymbol{\rho}\left(\overline{\boldsymbol{X}_j}, \boldsymbol{v_i^{(k)}}\right)}{\boldsymbol{\rho}\left(\overline{\boldsymbol{X}_j}, \boldsymbol{v_r^{(k)}}\right)}\right)^{2/(m-1)}},\ 1< i < c,\ 1 < j < N$$

(b) If $\exists j, r$ *such that* $\boldsymbol{\rho}\left(\overline{\boldsymbol{X}_j}, \boldsymbol{v_r^{(k)}}\right) = 0$ *then let* $\boldsymbol{\mu_{rj}^{(k)}} = \boldsymbol{1}$ *and* $\boldsymbol{\mu_{ij}^{(k)}} = \boldsymbol{0}$ *for* $i \neq r$

Step 2 compute $\boldsymbol{V^{(k+1)}} = [\boldsymbol{v_1^{(k+1)}}, \boldsymbol{v_2^{(k+1)}}, \ldots \boldsymbol{v_c^{(k+1)}}]$ *where* $\boldsymbol{v_i^{(k+1)}}$ i=1, 2,..c, *computed using (31)*

Step 3 if $\|V^{(k)}-V^{(k+1)}\| < \varepsilon$, *then go to step4*

else let k:=k+1, go to step1.

Step 4 End.

Here, each cluster belonging to the output set is represented by an IVIFS comprising IoT data instances. So, the algorithm presents the output clusters in the most generalized manner, i.e., IVIFSs. The IVIFCM's flowchart is depicted in Fig. 2.

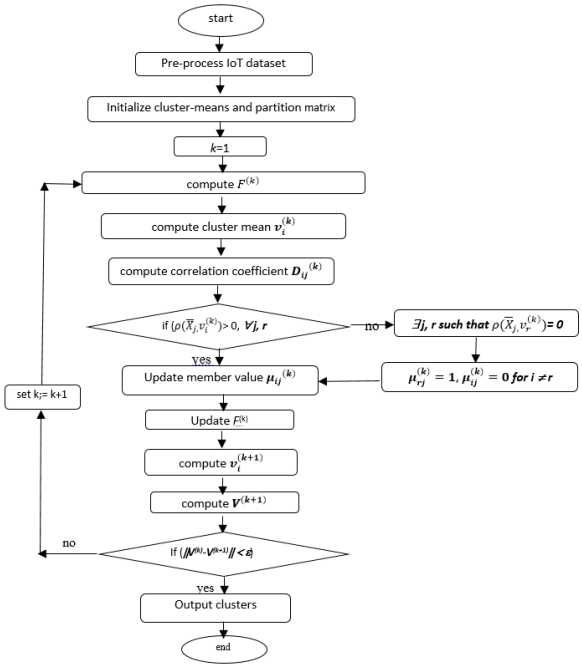

Fig. 2: Flowchart of the IVIFCM algorithm

Complexity Analysis

Since the IFCM clustering algorithm is not dependent on any of the given parameters, it takes O (1) time to compute both membership and non-membership values. Also, the computation of correlation coefficients takes constant time. Such computations do not contribute any extra costs. Therefore, the algorithms IFCM and FCM are computationally compatible. Thus, the worst-case computational cost of IFCM is found to be O ($i.c^2.N.n$), where i, c, N, n are the number of iterations, number of clusters, number of IoT data instances, and dimension of the dataset, respectively. Obviously, I = O (N), c being small, can be neglected, and $n <= N$. Similarly, the IVIFCM clustering algorithm is not dependent on any of the given parameters and takes O (1) time to compute both membership and non-membership intervals. It also takes constant time to compute correlation coefficients. So these factors do not include any extra costs to the algorithm. Therefore, the computational cost in the worst-case of IVIFCM is also O ($i.c^2.N.n$). Using the similar conditions of IFCM, it can be found that the IVIFCM's computational cost in the worst-case is O ($N^2.n$), which in turn demonstrates that both the proposed algorithms exhibit quadratic time and linear time complexity with respect to the size and dimensionality of the dataset, respectively.

Experimental Findings and Discussion

Experimental Findings

To evaluate the performance of the two approaches, the following benchmark datasets are considered.

KDDCup'99 datasets (UCI, 1999): This is a multivariate dataset representing intrusion scenarios in the military network, collected over 9 weeks, having 4,898,431 instances and 41 attributes.

Smartphone dataset (Irfan et al., 2018): It is a time-series dataset collected from smartphone sensors. It consists of 14,221 data instances.

Both the aforementioned datasets were acquired from the UCI machine learning repository, and the features are listed in detail in Table 1.

The proposed methods, together with the FCM algorithm, were executed using MATLAB on the aforementioned datasets using a standard computing environment. The obtained results are illustrated graphically in Figures 3-11.

Figure 3 depicts the rates of anomaly detections of the methods, namely, *k*-means, FCM, IFCM, and IVIFCM, using a bar diagram from which comparative analysis can be made easily.

Figure 4 describes the accuracy of the anomaly detections of the aforesaid algorithms using the datasets KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018). This can be effectively used for comparing the performances of the algorithms based on accuracy.

Table 1: Dataset characteristics

Dataset	Dataset char[c]	Attribute char[c]	No. of instances	No. of attributes
KDDCup'99 (UCI, 1999)	Multivariate	Numeric, categorical, temporal	4,898,432	41
Smartphone (Irfan et al., 2018)	Multivariatr	Numeric, time-series	14,221	15

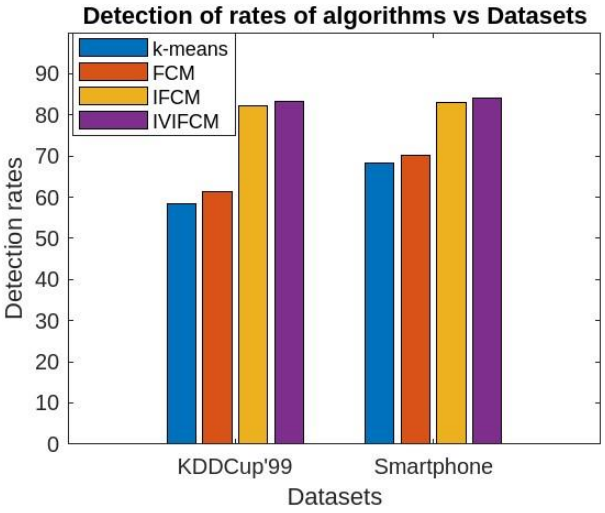

Fig. 3: Comparison of the algorithms based on detection rates

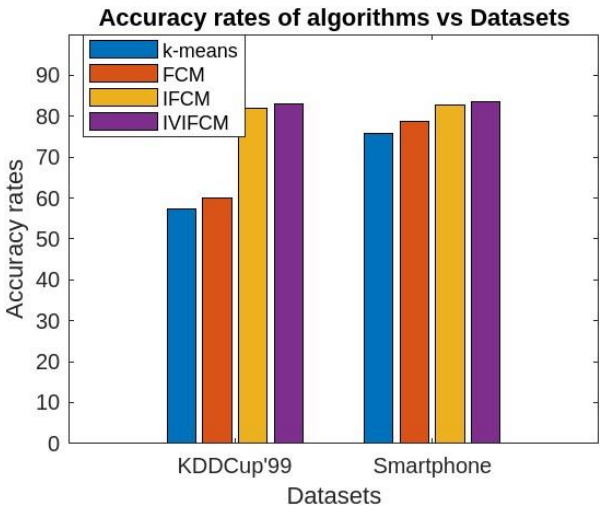

Fig. 4: Comparison of the algorithms based on the accuracy of detection

Figure 5 illustrates the obtained False alarm rates on KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018). This can easily be used for analysis of comparative performances of the approaches based on False alarm rates.

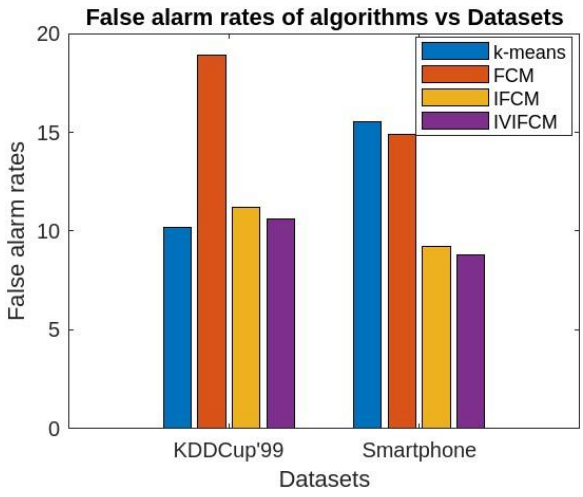

Fig. 5: Comparison of the algorithms based on false alarm rates

Figure 6 presents the algorithm's Denial of service rates on KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018). This facilitates the analysis of comparative performances of the algorithms based on Denial of service rates.

Figure 7 gives the User to root percentage of the algorithms on KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018) this provides an easy analysis of comparative performances of the algorithms based on the User to root.

Figure 8 illustrates the obtained Probe percentage on KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018). This can easily be utilized for the analysis of comparative performances of the algorithms based on the Probe percentage.

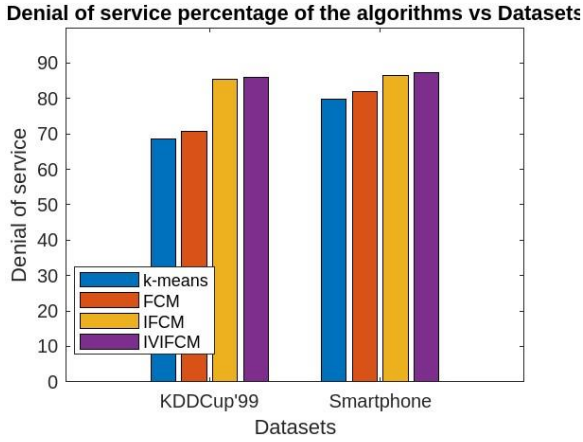

Fig. 6: Comparison of algorithms based on Denial of service

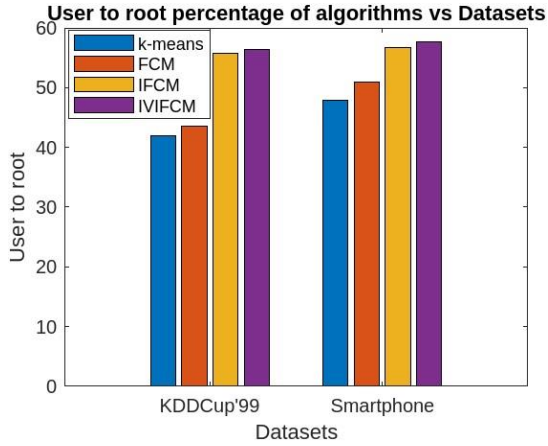

Fig. 7: Comparison of algorithms based on User to root

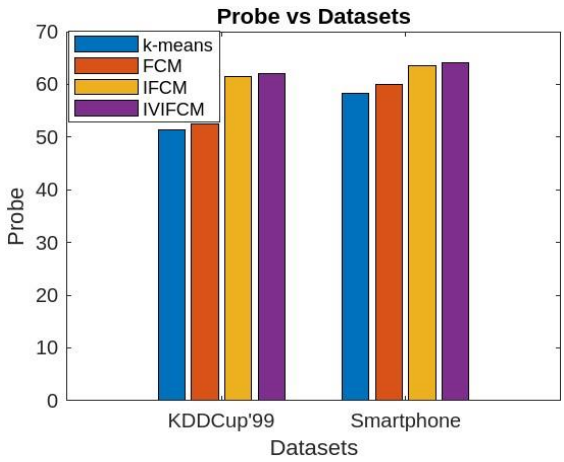

Fig. 8: Comparisons of the Probe with the two datasets

Figure 9 gives the precision percentages of the algorithms on KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018). This can easily be used for comparative analysis of performances based on precision.

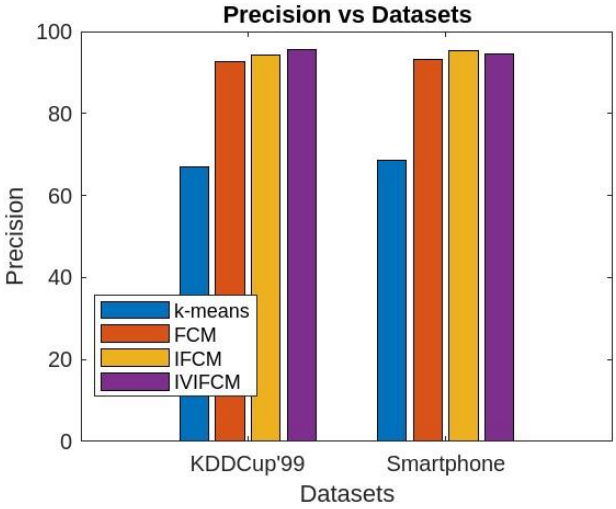

Fig. 9: Comparison based on precision

Figure 10 gives the Recall percentages of the algorithms on KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018). This can easily be used for the analysis of the comparative performances of the algorithms based on Recall.

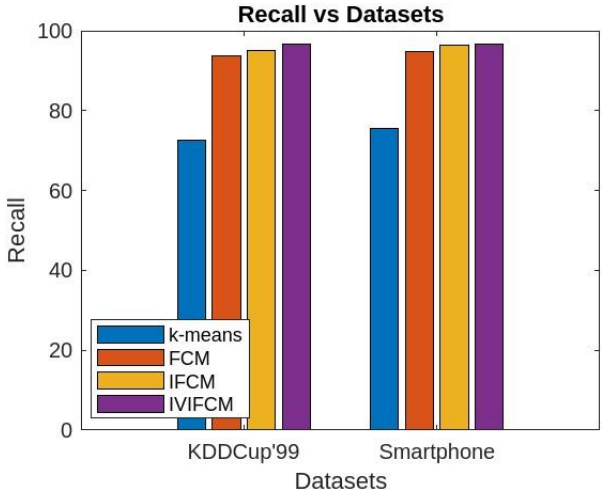

Fig. 10: Comparison based on Recalls

Figure 11 gives the F-scores of the approaches on KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018). This can easily be used for the assessment of the comparative performances based on the F-score.

Likewise, the execution times of IFCM and IVIFCM in terms of size and dimensionality of the datasets are illustrated in Figures 12 and 13.

Discussion

From the findings of the proposed algorithms, the following observations can be made.

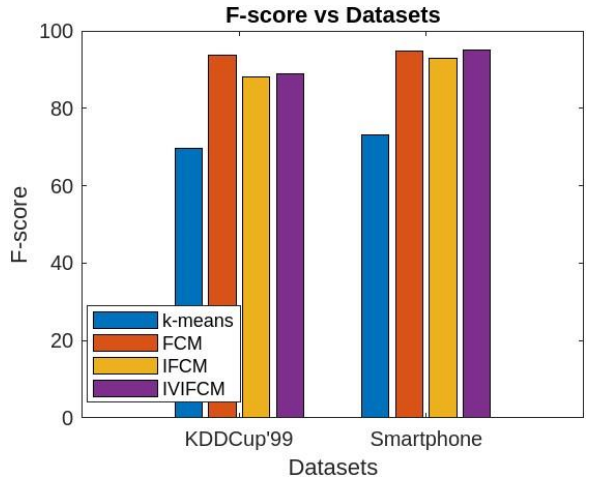

Fig. 11: Comparison based on F-scores

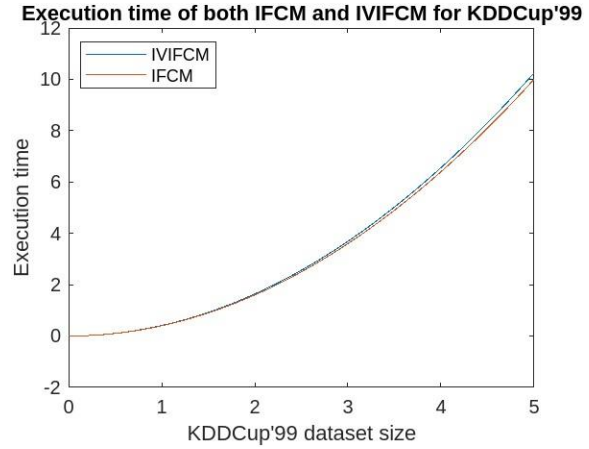

Fig. 12: Execution time of the proposed algorithms with KDDCup'99

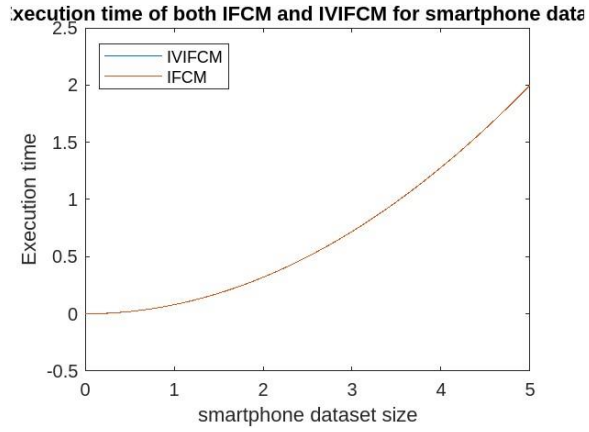

Fig. 13: Execution time of the proposed algorithms with the Smartphone dataset

For both the datasets KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018), the detection rates of the proposed algorithms are much higher than those of *k*-means and FCM. However, the rate of detection of IVIFCM is slightly higher than that of IFCM. So, the detection rates of the proposed algorithms (82.01, 83, 83.9, and 84.09%) do not vary so much. In fact, the IVIFC is more efficient than the others as far as the detection rate is concerned.

Similarly, the rates of accuracy of the proposed algorithms are much higher for both the datasets KDDCup'99 (82.01 and 82.9%) and Smartphone (82.7 and 83.6%) than those of k-means and FCM; however, both algorithms' accuracy rates are almost the same, and the IVIFCM's rate is slightly better than IFCM.

Again, the false alarm rates of both the proposed approaches are significantly lower than that of k-means and FCM for both the datasets, KDDCup'99 (UCI, 1999), and Smartphone (Irfan et al., 2018), and the IVIFCM's false alarm rate is the lowest.

Similarly, in the case of the attack parameters, Denial of service, User to root, and Probe, the performances of the introduced approaches exhibit superior performance than the others. Also, in the case of the other evaluation metrics, the methods outperform the conventional methods.

Anomaly detection rates and accuracy improve gradually from *k*-means → FCM → IFCM →IVIFCM across the datasets KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018). IVIFCM consistently has higher performance, indicating that it is more effective than others. Smartphone (Irfan et al., 2018) yields better performance. IVIFCM has the lowest False alarm rates, suggesting its better reliability than others. For both datasets, IVIFCM and IFCM significantly outperform *k*-means and FCM in all the performance metrics across both datasets. The comparative analysis of the above four algorithms can be presented in Table 2.

Table 2: Comparative analysis of performance metrics

Performance metric	Best algorithm	Worse algorithm
Detection rate	IVIFCM and IFCM	*k*-means algorithm
Accuracy rate	IVIFCM	*k*-means algorithm
False alarm rate	IVIFCM	FCM algorithm
Denial of services rate	IVIFCM and IFCM	*k*-means algorithm
User to root rate	IVIFCM and IFCM	*k*-means and FCM algorithms
Probe rate	IVIFCM and IFCM	*k*-means algorithm
Precision	IVIFCM and IFCM	*k*-means algorithm
Recall	IVIFCM and IFCM	*k*-means algorithm
F-score	IVIFCM and IFCM	*k*-means algorithm

Both IFCM and IVIFCM show a non-linear increase in execution with respect to the dataset sizes, and the former is slightly slower than the latter. IVIFCM offers a superior performance across all the performance metrics and datasets, KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018), making it practically an upgrade over IFCM and others.

Conclusion

This work introduces a fuzzy clustering algorithm for anomaly detection within the IoT environment. The proposed algorithms are the IFCM and IVIFCM algorithms. Both algorithms have used the correlation coefficient of IFSs instead of a distance measure to determine clusters. The proposed algorithms generate a specified number of clusters, where each IoT instance belongs to every cluster and is associated with a membership value and a non-membership value, whose sum lies in [0, 1]. An IoT instance that does not belong to any cluster, or belongs to all clusters with minimal membership values, or exhibits both minimal membership and non-membership values across all clusters, can be labelled as an anomaly. The algorithm's efficacies were demonstrated by the experimental studies with the datasets KDDCup'99 (UCI, 1999) and Smartphone (Irfan et al., 2018), and comparative analysis with *k*-means and FCM. The findings evidently demonstrate that the algorithms IFCM and IVIFCM outperform the *k*-means and FCM algorithms in all the parameters. Furthermore, IVIFCM is the most efficient one.

The run-time complexity of both IFCM and IVIFCM depends on the dimensionality and the size of the datasets, and their computational complexity is quadratic in the dataset size and linear in the dimensionality. As the dimension of any dataset is significantly smaller than its size, the computational complexities of both the IFCM and IVIFCM are assumed to be quadratic. Therefore, both the IFCM and IVIFCM demonstrate efficacy in IoT anomaly detection.

Limitations and Lines for Future Work

Even though the IFCM and IVIFCM are quite efficient, they still have several shortcomings. Firstly, like many other partitioning-based approaches, IFCM and IVIFCM are sensitive to the initialization of the cluster centroid. Secondly, they struggle to address the curse of high dimensionality, which makes them less efficient in high-dimensional data. Finally, the optimal solution may not be reached, as the methods may be stuck at local minima.

The subsequent points outline the possible future direction of work:

- In the subsequent research, algorithms may be proposed to deal with high dimensionality in the IoT environment
- In future works, supervised and hybrid approaches can be explored for similar problems
- In future studies, a more generalized approach, like the picture fuzzy or bipolar fuzzy clustering approach, can be proposed for IoT anomaly detection

Acknowledgment

Thank you to the publisher for their support in the publication of this research article. We are grateful for the resources and platform provided by the publisher, which have enabled us to share our findings with a wider audience. We appreciate the efforts of the editorial team in reviewing and editing our work, and we are thankful for the opportunity to contribute to the field of research through this publication.

Funding Information

The authors have not received any financial support or funding to report.

Author's Contributions

Vijo Arul Selvi M. and Fehmin Nadira Laskar: Conceptualization, data curation, formal analysis, methodology, software, validation, visualization, writing original draft, writing review and edited.

Fokrul Alom Mazarbhuiya, Mohamed Shenify and M. Alliheedi: Conceptualization, formal analysis, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing original draft, writing review and edited.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

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

Atanassov, K., & Gargov, G. (1989). Interval valued intuitionistic fuzzy sets. In *Fuzzy Sets and Systems* (Vol. 31, Issue 3, pp. 343–349). https://doi.org/10.1016/0165-0114(89)90205-4

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

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

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

Chaudhury, A. (2015). Hindawi Publishing Corporation, Advances in Fuzzy Systems. *Fuzzy Methods for Data Analysis*, 1–17. https://doi.org/10.1155/2015/238237

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

Dake Kwasi, D., 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

Firoozjaei, M. D., Mahmoudyar, N., Baseri, Y., & Ghorbani, A. A. (2022). An evaluation framework for industrial control system cyber fincidents. *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.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

Irfan, M., Tokarchuk, L., Marcenaro, L., & Regazzoni, C. (2018). Anomaly Detection in Crowds Using Multi Sensory Information. *Proceeding of the IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 1–6. https://doi.org/10.1109/avss.2018.8639151

Kopawar Atul, N., & 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

Masmali, F. H.., 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. In *ICIC Express Letters* (Vol. 17, Issue 1, pp. 73–80). https://doi.org/10.24507/icicel.17.01.73

Mazarbhuiya, F. A., & Abulaish, M. (2012). Clustering Periodic Patterns using Fuzzy Statistical Parameters. *International Journal of Innovative Computing Information and Control (IJICIC*, *8*(3), 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. Y., & Mahanta, A. K. (2020). Detecting Anomaly Using Partitioning Clustering with Merging. In *ICIC Express Letters* (Vol. 14, Issue 10, pp. 951–960).

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

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. Alom. (2023). Documents Clustering using Subspace Clustering Algorithm. *ICIC Express Letters*, *17*(12), 1405–1415. https://doi.org/10.24507/icicel.17.12.1405

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

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

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

UCI. (1999). KDD Cup 1999 Data. *Https://Kdd.Ics.Uci.Edu/Databases/Kddcup99/Kddcup99.Html*.

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. (2009). Intuitionistic fuzzy hierarchical clustering algorithms. In *Journal of Systems Engineering and Electronics* (Vol. 20, Issue 1, pp. 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

Xuan Thao, N. (2018). A new correlation coefficient of the intuitionistic fuzzy sets and its application. *Journal of Intelligent & Fuzzy Systems*, *35*(2), 1959–1968. https://doi.org/10.3233/jifs-171589

Younas, M. Z. (2020). Anomaly Detection using Data Mining Techniques: A Review. In *International Journal for Research in Applied Science and Engineering Technology* (Vol. 8, Issue 11, pp. 568–574).

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, Rui, Robu, B., Bouchenak, S., Ben Mokhta, S., & Chen, L. Y. (2019). RAD: On-line Anomaly Detection for Highly Unreliable Data. *Machine Learning*. https://doi.org/https://doi.org/10.48550/arXiv.1911.04383

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