Quarterly Journal of
Optimization In Soft Computing

Vol. 3, Issue 2, Summer 2025

- **Quadrants Dynamic Fuzzy Histogram Equalization for Color Images with Brightness Preservation**
  Nooshin Allahbakhshi, Yashar Salami, Mohammad Bagher Karimi, Kioumars Abdi, Yaser Pourshadlou
- **A Novel Algorithm for Enhancing Fault Tolerance and Reliability in Wireless Body Area Networks**
  Majid Ghazanfari, Babak Nikmard, Golnaz Aghaee Ghazvini
- **An Efficient Approach for Multi-Label Streaming Feature Selection**
  Azar Rafie, Parham Moradi
- **Optimizing Argumentative Text Comprehension via Inverted Classroom**
  Mohamadreza Rafizade tafti, Fariba Rahimi Esfahani, Azar Alisoltani
- **Implementation and comparison of active queue management algorithms in traditional and SDN networks**
  Khoshnam Salimi Beni, Mohammadreza Soltanaghaei, Rasool Sadeghi
- **Simultaneous Network Reconfiguration and Capacitor Placement in Distribution Systems Using the Proposed Discrete PSO Algorithm with Chaos Module**
  Fahimeh Sayadi Shahraki

Journal of Optimization of Soft Computing (JOSC)
Vol. 3, Issue 1, pp: (1-10), Summer-2025
Journal homepage: https://sanad.iau.ir/journal/josc

Paper Type (Research paper)

Quadrants Dynamic Fuzzy Histogram Equalization for Color Images with Brightness Preservation

Nooshin Allahbakhshi[1], Yashar Salami[2]*, Mohammad Bagher Karimi[3], Kioumars Abdi[4], Yaser Pourshadlou[5]

1. Department of Computer and Information Technology Engineering, Khoy Branch, Islamic Azad University, Khoy, Iran
2. Faculty of Computer and Information Technologies, Cappadocia University, Nevsehir, Turkey
3. Department of Computer Engineering, Tabriz branch, Islamic Azad University, Tabriz, Iran
4. Kapadokya Üniversitesi, Kapadokya Meslek Yüksekokulu, Bilgisayar Programcılığı, Nevşehir, Türkiye
5. Department of Electrical and Electronics Engineering, Tabriz branch, Islamic Azad University, Tabriz, Iran

Article Info

Article History:
Received: 2025/04/15
Revised: 2025/08/13
Accepted: 2025/09/17

DOI:

Keywords:
Color Images, Histogram Equalization, Fuzzy, HSV, Threshold Value, Dynamic Sub-Histograms

**Corresponding Author's Email Address:*
Yashar.salami@gmail.com

Abstract

Contrast enhancement is essential in image processing and contributes to image enhancement. Histogram equalization is perhaps the most common way operators enhance the contrast of digital images. Easy and handy, this method often has too much contrast enhancement, making the output images' visual quality look unnatural. Moreover, it usually cannot also preserve the mean of the image substantially. This paper presents a color image equalization technique that takes a better guess to conserve the brightness. In other words, it is a method based on some image histogram modification using fuzzy and a clipping process for equalization rate applied to the original image. Initially, the histogram is split into two parts hinged on the mean gray level. Then, it is divided into four sections by calculating an average of the two sub histograms. The dynamic equalization is defined for a new range, and the sub-histogram equalization is independent. The simulation results prove that this new method can significantly improve the spatial characteristics of color images and keep a high brightness level.

1. Introduction

In today's digital era, the rise of the Internet of Things [1], [2] and blockchain technology [3][4] has transformed data-driven applications [5][6], highlighting the need for secure and reliable information exchange[4], [7]. This reliance on trustworthy multimedia and image data makes digital image processing increasingly vital, a field that has historically advanced in response to real-world demands.

In the 1960s, NASA's Ranger 7 spacecraft transmitted unclear television images of the Moon's surface to Earth, marking the initial steps toward global human-space communication [8]. With the increasing demand for extracting image details and identifying suitable landing sites for the Apollo missions, the field of digital image processing emerged, leading to the rapid advancement and adoption of this technology [9]. However, the images obtained from these early transmissions were often affected by noise and distortions, including blurring and image fading, significantly diminishing their clarity and

quality[10]. As a result, removing noise and addressing visual defects such as improper lighting parameters and poor color composition have become critical concerns in image processing. All techniques and methods employed to enhance image quality and reduce visual imperfections fall under the domain of image processing [11].

Since its inception in 1964, image processing has witnessed remarkable growth and has extended beyond space research to various fields.[12]. It is now integral to various applications such as medicine, speech recognition, handwriting recognition, archaeology, astronomy, biology, nuclear medicine, and industries including aerospace, packaging, automotive, pharmaceuticals, medical diagnostics[13], [14] , and meteorology[15]. One standard method for improving image quality is image enhancement, particularly contrast enhancement. Image enhancement involves adjusting the intensity values of an input image so that the output image appears visually enhanced. The primary objective of image enhancement is to make the information in the image more interpretable for human viewers or to optimize it as an input for automated image processing systems.[16].

Histogram equalization (HE) is a widely recognized method for contrast enhancement that redistributes an image's intensity values.[17], [18]. The fundamental concept of HE is to map the input image intensity values to new intensities using a cumulative distribution function (CDF)[19]. This process effectively broadens and flattens the image's histogram, improving overall contrast. Initially, histogram equalization transforms the original image’s histogram into a uniform distribution based on the average grayscale levels. Consequently, the average brightness of the output image is centered around the mean brightness of the input image. This adjustment is particularly significant for images with low or high brightness, as it enhances contrast. Subsequently, the second phase of HE performs contrast enhancement based on the overall content of the image.[20].

Several HE methods have been proposed to preserve image brightness while improving contrast. These methods are typically classified into partitioned histogram equalization (PHE) and dynamic partitioned histogram equalization (DPHE) [21], [22]. Both approaches rely on statistical information to divide the original histogram into multiple sub-histograms. The primary distinction between them is that DPHE assigns a new dynamic range instead of utilizing the original range. PHE-based methods include brightness-preserving bi-histogram equalization (BBHE)[23], [24] and multi-peak histogram equalization with brightness preservation (MPHEBP)[25]. In contrast, DPHE methods are fewer in number, with examples such as dynamic histogram equalization (DHE) and brightness-preserving dynamic histogram equalization (BPDHE) [26]. Furthermore, to enhance images captured in low-light conditions, quarter dynamic histogram equalization (QDHE) has been introduced [27]. Another method was presented.

In recent years, additional methods have been developed, including color image enhancement based on gamma encoding and histogram equalization [28], and low-contrast enhancement for color images using intuitionistic fuzzy sets with adaptive histogram equalization [29]. In another paper, they present a method that uses a variational approach including an energy function to determine local transformations in the luminance (L) and chroma (C) channels of the CIE LCH color space[30] . Another paper introduces exposure-based recursive histogram equalization techniques along with an energy curve instead of the conventional histogram[31] . Another paper presented is fuzzy logic-based histogram segmentation based on maximum and minimum peaks, which combines it with an entropy-controlled coefficient correction system[32].

The proposed technique offers a fresh method for these problems, characterized by fuzzy histograms with an idealized membership function.

The strategy successfully counters issues of grey or color imprecisions; it not only eradicates the random variations but also prevents loss of intensity levels without extra smoothing. Fuzzy statistics indeed deliver a significant gain in performance, as is demonstrated through experimental results, too. Merging quarter histograms and fuzzy statistics improves the capability of the proposed method to solve incompletely developed narrow sub-histogram problems, as low-light images will be better reconstructed.

1.1. Contribution

Image processing and contrast enhancement are important and are number one in image enhancement. Histogram Equalization is perhaps the most popular way for operators to brighten contrast on digital images. Easy and handy Sadly, this method also has the problem of output images themselves looking artificial due to the notorious contrast enhancement. Also, it is hard to keep the mean of the image at a significant level. This paper introduces a color equalization method of images, which gives a better idea of maintaining lightness. It is a Fuzzy and Blending method of equalization rate applied to the original image that is reduced to level histogram modification tandem. In the beginning, the histogram is split into two parts according to the mean gray level, then divided into four parts, which uses an average of two sub-histograms. Dynamic equalization is described independently for a new range and sub-histogram equalization. Simulation results indicate that this novel method effectively affects much of spatial color images in terms of improvement, keeping high brightness.

1.2. Paper organization

The paper's organization is as follows: In the second section, the phases of the proposed method are described in detail, outlining each step systematically. The third section covers the simulation environment and discusses the obtained results, providing an in-depth analysis of the proposed approach's performance and effectiveness. Finally, the concluding section summarizes the key findings and reflects on the implications of the results.

2. Proposed Scheme

The proposed method begins by transforming the image from the RGB color space to the HSV color space, allowing for better image contrast and intensity manipulation. This transformation is crucial because the HSV color space separates the intensity (value) component from the color information, making enhancing the contrast of low-light and low-contrast images easier. Next, the input histogram is converted into a fuzzy histogram to smooth out the intensity values and prepare the data for more precise processing. The fuzzy histogram introduces a level of uncertainty, helping to preserve image details during the enhancement process. A clipping process is then applied to tackle saturation issues, which often arise in contrast enhancement. The mean intensity value of the image is used as a threshold, and intensity values exceeding this threshold are clipped, preventing over-bright areas and ensuring that the enhancement remains natural. The mean intensity of the image is used to divide the fuzzy histogram into four parts, and then it is in the mean range. This separated way allows for adjusting the contrast of each sub-histogram in its dynamic range, so you slowly increase saturation for all intensity levels but not so much the brightness. After equalization, the image was converted from HSV color space to RGB. This step aims to enhance contrast further but keep the image's original color. The proposed method increases contrast in such images, improving visual clarity and detail preservation. Next, a detailed description of each step in this process is given in the following parts.

2.1. Calculation of the Fuzzy Histogram

After converting the image to the HSV color space, the input histogram is transformed into a fuzzy histogram, as Equation (1) describes. Here, fuzzy Histogram(i) represents the frequency of gray levels around intensity i. The function I(x, y) denotes the gray level intensity at pixel coordinates (x, y), and its corresponding fuzzy value is computed based on the fuzzy histogram.

$$\text{fuzzyHistogram}(i), i \in \{0,1,2,\ldots,L-1\} \quad (1)$$

Where fuzzyHistogram(i) is actually the number of repetitions of gray levels around i. I(x,y) represents the gray value and represents its fuzzy value. a membership function must first be defined using Equation (2) to achieve this transformation. In the proposed method, a triangular membership function is employed. The fuzzy triangular membership function is defined a s follows:

$$\mu_{\tilde{I}(x,y)} = \max\left(0.1 - \frac{I(x,y) - i}{4}\right) \quad (2)$$

$\mu_{\tilde{I}(x,y)}$ is the fuzzy triangular membership function . This equation accumulates the fuzzy values corresponding to each gray level, thus creating the fuzzy histogram, which is subsequently used for further processing in the algorithm. The parameters a and c in the equation are adjustable elements of the membership function, allowing for control over the shape and sensitivity of the fuzzy set. The gray level intensity i is an essential component, as it is used to compute the fuzzy value for each gray level in the histogram.

The reasons for using the fuzzy triangular membership function include: gradual modeling of changes in brightness, increasing accuracy in processing brightness levels, reducing the effect of noise and improving image smoothness, and preserving details in dark or bright areas.

Equation (3) calculates the fuzzy histogram, incorporating these parameters to accurately reflect the distribution of gray levels in the image in a fuzzy framework. This approach enhances the histogram's ability to capture subtle differences in intensity, making it more effective in contrast enhancement and image segmentation.

$$\text{fuzzyHistogram}(i) = \text{fuzzyHistogram}(i) + \sum_x \sum_y \mu_{\tilde{I}(x,y)i} \quad (3)$$

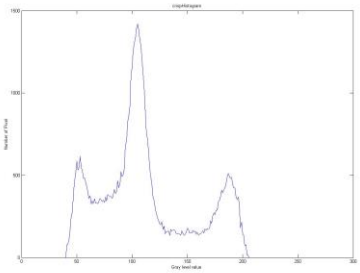


Figure 1. Crisp histogram of the input image

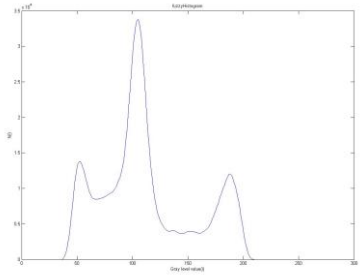


Figure 2. Fuzzy histogram obtained from the image

Figure1 shows the crisp histogram of the input image,which is transformed into Figure 2 by applying the fuzzy process that can be performed using equations (2) and (3). The fuzzy statistic is able to apply more precision to the gray value than the classic crisp histogram, thus producing a smooth histogram.

2.2. Application of Clipping Process Based on a Threshold Value

One key reason for utilizing the clipping process is to regulate the equalization rate, thereby avoiding unnatural image processing and excessive enhancement that could distort the original image. To simplify and reduce computational complexity, the mean intensity value of the image can be used to determine the clipping point, which is the approach adopted in this study. In the clipping process, the mean intensity of the input image is first calculated, and this value is taken as the clipping point or threshold (denoted as TCT_CTC). This threshold is then applied to the fuzzy histogram obtained from the image. Specifically, all values in the fuzzy histogram that exceed this threshold are replaced by the threshold value itself, while values below the threshold remain unchanged. This method ensures that the image's intensities are controlled, preventing overly bright or saturated regions, which ultimately enhances the image's overall quality and clarity.

2.3. Fuzzy Histogram Segmentation

The steps involved in the segmentation process based on the mean of the fuzzy histogram of the image are crucial for enhancing image analysis. Here is an expanded explanation of the process:

2.3.1. Segmentation of the Fuzzy Histogram: The process begins by dividing the fuzzy histogram into two sub-histograms based on the mean value of the entire histogram. This segmentation is achieved by separating the pixel intensity values into two categories: those above the mean and those below the mean. These two sub-histograms correspond to higher and lower intensity values relative to the mean, ensuring that the image's darker and brighter areas are distinctly addressed.

2.3.2. Calculation of Sub-Histogram Means: After the initial division, the mean value of each sub-histogram is calculated. These new mean values become the key separation points between the sub-histograms. By doing so, the algorithm identifies significant intensity boundaries within high and low-intensity regions, preparing the data for further segmentation.

2.3.2. Division into Four Sub-Histograms: The fuzzy histogram is divided into four distinct sub-histograms using the previously calculated mean values as separation points. Each sub-histogram represents a different image intensity range, with two sub-histograms dedicated to the lower intensities and two to the higher intensities. This finer segmentation allows for a more precise analysis of the intensity distribution across the image.

This segmentation process divides the image into distinct regions based on intensity levels and enhances the overall image analysis by providing more detailed information about how intensity values are distributed. As a result, it significantly improves the quality and clarity of the image, making it easier to analyze subtle details, contrasts, and textures. The additional information gained from this segmentation plays a key role in various image processing tasks, including contrast enhancement, edge detection, and texture analysis.

2.4. Assignment of New Gray Level Ranges

Dynamic equalization processes each sub-histogram to achieve a balanced and precise equalization. By dynamically assigning a gray level range based on the span of gray levels and the total number of pixels within each sub-histogram, this method ensures that each sub-histogram receives an appropriately tailored range for optimal equalization. This approach significantly improves the equalization of each sub-histogram, minimizing the risk of losing image details or encountering intensity saturation effects, which are common in conventional equalization techniques.

The dynamic range assignment, designed to enhance different image sections adaptively, is mathematically formulated using Equations (4) and (5). These equations dictate how the gray level ranges are distributed, ensuring that the intensity transitions between different parts of the image remain smooth while preventing over-compression or excessive expansion of the gray levels. The result is an image with better-preserved details, contrast, and overall quality.

$$\mathrm{span_i} = \mathrm{high_i} - \mathrm{low_i} \qquad (4)$$

$$\mathrm{range_i} = (\mathrm{L}-1) \times \mathrm{span_i} / \sum_{\mathrm{k}=1}^{4} \mathrm{span_k} \qquad (5)$$

The parameters $high_i$ and low_i are the maximum and minimum intensity values under histogram i are respectively. The dynamic range used by sub-histogram i in the input image is denoted by $\mathrm{span_i}$ While the dynamic range applied in the output image is denoted by $\mathrm{range_i}$ For the i-the sub-histogram, the new dynamic range is assigned to the interval $[\mathrm{i_{start}}\ \mathrm{i_{end}}]$ Which is determined by Equations (6) and (7).

$$\mathrm{i_{start}} = (\mathrm{i}-1)_{\mathrm{end}} + 1 \qquad (6)$$

$$i_{end} = i_{start} + range_i \quad (7)$$

The first value i_{start} is the smallest intensity value of the new dynamic range.

2.5. Dynamic Equalization of Each Sub-histogram

Dynamic equalization for each sub-histogram facilitates the individualized equalization of every sub-histogram, preventing the issues of under or over-equalization in different regions of the image. This is achieved by allocating unique, non-overlapping gray level ranges to each sub-histogram. This ensures that gray levels from distinct sub-histograms are not mapped to the same gray level in the final output image. The method guarantees that the entire gray level spectrum is utilized efficiently, maintaining the contrast and detail across all image regions.

For sub-histogram iii, which operates within the range [i_{start} i_{end}], the transfer function responsible for equalizing the output histogram is mathematically expressed by Equation (8). This equation governs how pixel intensities are redistributed within the designated range, ensuring optimal enhancement without introducing artifacts or excessive brightness changes. The adaptive nature of this approach contributes to more accurate image enhancement, particularly in cases with varied lighting conditions or uneven intensity distributions across different parts of the image.

$$y(x) = (i_{start} - i_{end}) \times cdf(X_k) + i_{start}\, cdf(X_k) \quad (8)$$

$cdf(X_k)$ is the cumulative distribution function in that sub histogram. This formula is the same as the HE formula, but instead of maximum and minimum intensity, i_{start} and i_{end} are used in the dynamic range of the output.

3 Simulation

3.1. Simulation Environment

A proposed method was implemented and checked on a Fujitsu laptop (4 GB RAM, 128 GB disk, Intel Core i5). The simulations were performed using MATLAB, a very well-known platform commonly used for image processing and algorithm development.

The critical aspect of this approach was accomplished using extensive libraries (built-in functions) in the MATLAB environment and specialized toolboxes to leverage fuzzy histogram equalization and intricate following calculations, ensuring that the simulations establish correct and efficient results.

For testing and evaluating the method, we used the test suite of the publicly available Kodak Lossless True Color Image, available from https://r0k.us/graphics/kodak/. This dataset is typically used in image processing research, as the images are highly quality and provide a reliable basis for evaluating contrast enhancement algorithms' performance. The dataset also includes several images that allow for an in-depth assessment of how well the method performs, particularly in low-contrast and low-light situations. This simulation environment, along with the Kodak dataset, assures a statistically sound testing ground to judge the efficacy of contrast enhancement methodology in enhancing contrast without compromising image quality for clarity and details.

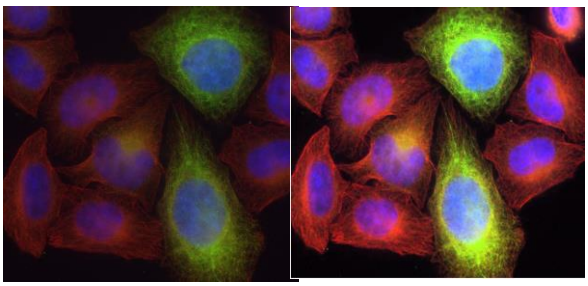
All of these components involve more than enough discernment to demonstrate that the method can deal with many imaging problems in different conditions.

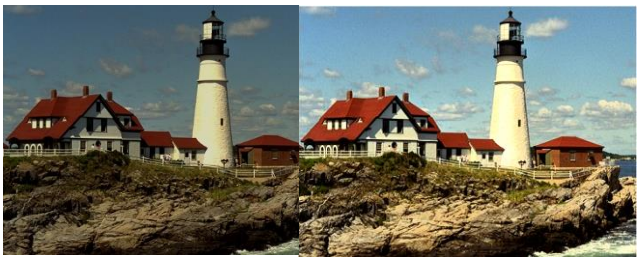
3.2. Simulation result

The simulation results prove that the proposed method successfully resolved the deficiencies of the earlier techniques, showing significant increases in different areas compared to the earlier techniques as presented. Also, one benefit offered by this technique is the most effective reduction in saturation level while avoiding noise amplification, which is crucial to preserving sharpness and nuance within an image.

One of the benefits this method offers is to reduce or minimize saturation effects without amplifying noise, which will be a big part of ensuring the model can maintain quality and detail fracture in an image. On the contrary, most traditional strategies are strongly or

improperly equalized, leading to losing some information in distinct image regions. Whilst this is not a problem solved with my method, the final image remains fine details and gains an enormous contrast throughout, and on the screenshot, it is a better visual. Of course, this previous approach is distinct because it takes this innovative attempt to create an even more flexible/dynamic equalization process. This guarantee is that the brightness and contrast of an image would be found in the better middle state without making it overexposed or underexposed. As the method performs well even in varying lighting environments, this added advantage makes it practical for many tasks. Figure 1 gives results of the original images and improved versions by using the proposed method to show that the proposed enhancement gives a greater detail fidelity and contrast improvement than an initially ordinary eye.





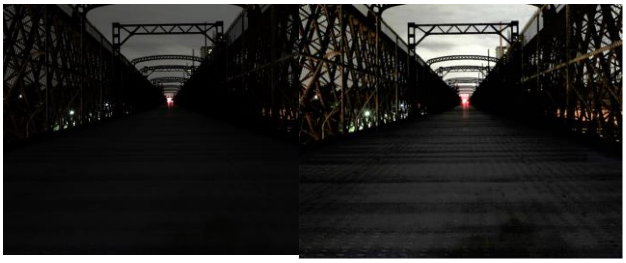


Fig 3. Original image and results of the proposed method.

4. Future Work

Future investigations may focus on merging advanced causal reasoning methods with biomarker detection techniques approaches that have been highlighted in recent studies on protein corona profiling and actual causality modeling as potential tools for application in image processing and contrast enhancement.

First, the methodology introduced in[33] which integrates mass spectrometry-based proteomics, machine learning, and causal analysis demonstrates the value of combining multi-modal data analysis with dynamic fuzzy histogram equalization. Within our framework, such integration could support adaptive tuning of histogram equalization parameters based on complex and heterogeneous datasets, thereby achieving improved contrast preservation in domain-specific imaging tasks, such as biomedical image interpretation.

Second, the causality-driven strategy described in[34] shows that pinpointing "actual causes" within variable interactions leads to system outputs that are more predictable and controllable. Applying this principle to our algorithm could help identify which image attributes or histogram regions act as the true causal drivers of perceived visual quality, thus enabling targeted enhancement methods that reduce the likelihood of unwanted artifacts.

Third, the formal reasoning approach for identifying root causes in engineered systems proposed in [35] could guide the design of automated diagnostic modules in our image enhancement pipeline. Such modules would be capable of detecting and isolating the primary sources of visual degradation such as uneven illumination or noise and activating selective enhancement processes tailored to each specific issue.

Overall, future research will aim to integrate our fuzzy dynamic histogram equalization method with causality-based analysis and cross-domain data fusion, in order to develop intelligent, context-aware enhancement systems. This path has the potential to extend the scope of our approach beyond general-purpose color image processing to specialized domains such as medical imaging, remote sensing, and industrial inspection, where maintaining high contrast alongside preservation of critical details is essential.

5. Conclusion

This paper proposes an original solution for enhancing color images using histogram equalization to preserve brightness. This solution overcomes some limitations of existing approaches by quenching the saturation, noise enhancement, and improper equalization, resulting in the preservation of image details. It is especially well-suited for low-light images. Fuzzy logic-based, dividing histograms into four sub-histograms results in better output and decreased computational overhead. Suggestions for future research are to use advanced fuzzy logic techniques, improve ways of maintaining brightness and preserving image details, and apply dynamic equalization methods via new boundary invariance.

References

[1] Y. Salami, "SO-ITS: a secure offloading scheme for intelligent transportation systems in federated fog-cloud," *Iran J. Comput. Sci.*, 2025, doi: 10.1007/s42044-025-00318-9.

[2] Y. Salami, "SOBT-UF: Secure Offloading in Blockchain Infrastructure for Intelligent Transportation Systems Using 5G-Enabled UAVs Within a Fog-Edge Computing Federation," in *2024 19th Iranian Conference on Intelligent Systems (ICIS)*, IEEE, 2024, pp. 217–222. doi: 10.1109/ICIS64839.2024.10887460.

[3] Y. Salami, F. Taherkhani, Y. Ebazadeh, M. Nemati, V. Khajehvand, and E. Zeinali, "Blockchain-Based Internet of Vehicles in Green Smart City: Applications and Challenges and Solutions," *Anthropog. Pollut.*, vol. 7, no. 1, pp. 87–96, 2023, doi: 10.22034/AP.2023.1978624.1144.

[4] Y. Salami, V. Khajehvand, and E. Zeinali, "LSMAK-IOV: Lightweight Secure Mutual AKE Scheme in Fog-Based IoV," in *2024 10th International Conference on Artificial Intelligence and Robotics (QICAR)*, IEEE, 2024, pp. 1–5. doi: 10.1109/QICAR61538.2024.10496659.

[5] Y. Salami, Y. Ebazadeh, M. Hamrang, and N. Allahbakhshi, "A Novel Approach for Intrusion Detection System in IoT Using Correlation-Based Hybrid Feature Selection and Harris Hawk Optimization Algorithm," *J. Optim. Soft Comput.*, vol. 2, no. 3, pp. 7–21, 2024.

[6] Y. Salami, V. Khajehvand, and E. Zeinali, "A new secure offloading approach for internet of vehicles in fog-cloud federation," *Sci. Rep.*, vol. 14, no. 1, p. 5576, 2024, doi: 10.1038/s41598-024-56141-y.

[7] Y. Salami, vahid khajehvand, and E. Zeinali, "Efficiency Simultaneous key Exchange-Cryptography Extraction from Public key in Fog-Cloud Federation-based Secure Offloading for Automatic Weather Stations Observing Systems," *Nivar*, vol. 47, no. 120–121, pp. 153–165, 2023, doi: 10.30467/nivar.2023.416270.1261.

[8] R. D. Launius *et al.*, "Spaceflight: the development of science, surveillance, and commerce in space," *Proc. IEEE*, vol. 100, no. Special Centennial Issue, pp. 1785–1818, 2012, doi: 10.1109/JPROC.2012.2187143.

[9] M. B. Quadrelli *et al.*, "Guidance, navigation, and control technology assessment for future planetary science missions," *J. Guid. Control. Dyn.*, vol. 38, no. 7, pp. 1165–1186, 2015, doi: 10.2514/6.2013-5411.

[10] D. M. Chandler, "Seven challenges in image quality assessment: past, present, and future research," *Int. Sch. Res. Not.*, vol. 2013, no. 1, p. 905685, 2013, doi: 10.1155/2013/905685.

[11] R. C. Gonzalez, *Digital image processing*. Pearson education india, 2009.

[12] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, 2010, doi: 10.1109/RBME.2010.2084567.

[13] Z. Hossein-Nejad and M. Nasri, "Social Spider Optimization Algorithm in Multimodal Medical Image Registration," *J. Optim. Soft Comput.*, no. 1, pp. 23–30, 2023, doi: https://doi.org/10.82553/josc.2023.14020714783333.

[14] A. Banitalebidehkordi, "Using the fuzzy methods to examine changes in brain lesions and atrophy from MRI images for rapid diagnosis of MS," *J. Optim. Soft Comput.*, no. 1, pp. 19–25, 2024, doi: https://doi.org/10.82553/josc.2024.140302141118861.

[15] N. Allahbakhshy, M. B. Karimi, and Y. Salami, "A Novel Histogram Equalization Method for Enhancing the Quality of Black-and-White Meteorological Images Maintaining Brightness," *Nivar*, vol. 48, no. 126–127, pp. 83–96, 2024, doi: 10.30467/nivar.2024.465389.1297.

[16] R. Ghabousian and N. Allahbakhshi, "Survey of Contrast Enhancement Techniques based on Histogram Equalization," *Int. J. Rev. Life Sci*, vol. 5, no. 8, pp. 901–908, 2015.

[17] K. G. Dhal, A. Das, S. Ray, J. Gálvez, and S. Das, "Histogram equalization variants as optimization problems: a review," *Arch. Comput. Methods Eng.*, vol. 28, pp. 1471–1496, 2021, doi: https://doi.org/10.1007/s11831-020-09425-1.

[18] K. Jha, A. Sakhare, N. Chavhan, and P. P. Lokulwar, "A Review on Image Enhancement Techniques using Histogram Equalization.," *Grenze Int. J. Eng. Technol.*, vol. 10, no. 1, 2024.

[19] S. A. Durai and E. A. Saro, "Image compression with back-propagation neural network using cumulative distribution function," *World Acad. Sci. Eng. Technol.*, vol. 17, pp. 60–64, 2006, doi: 10.1016/j.phpro.2012.03.132.

[20] Y. Zhu and C. Huang, "An adaptive histogram equalization algorithm on the image gray level mapping," *Phys. Procedia*, vol. 25, pp. 601–608, 2012.

[21] B. S. Rao, "Dynamic histogram equalization for contrast enhancement for digital images," *Appl. Soft Comput.*, vol. 89, p. 106114, 2020, doi: 10.1016/j.asoc.2020.106114.

[22] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 593–600, 2007.

[23] S.-D. Chen and A. R. Ramli, "Preserving brightness in histogram equalization based contrast enhancement techniques," *Digit. Signal Process.*, vol. 14, no. 5, pp. 413–428, 2004.

[24] Y.-T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 43, no. 1, pp. 1–8, 1997, doi: 10.1109/APCCAS.1998.743808.

[25] K. Wongsritong, K. Kittayaruasiriwat, F. Cheevasuvit, K. Dejhan, and A. Somboonkaew, "Contrast enhancement using multipeak histogram equalization with brightness preserving," in *IEEE. APCCAS 1998. 1998 IEEE Asia-Pacific Conference on Circuits and Systems. Microelectronics and Integrating Systems. Proceedings (Cat. No. 98EX242)*, IEEE, 1998, pp. 455–458. doi: 10.1109/APCCAS.1998.743808.

[26] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1752–1758, 2007.

[27] C. H. Ooi and N. A. M. Isa, "Quadrants dynamic histogram equalization for contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 56, no. 4, pp. 2552–2559, 2010, doi: 10.1109/TCE.2010.5681140.

[28] P. Kaur, B. S. Khehra, and A. P. S. Pharwaha, "Color image enhancement based on gamma encoding and histogram equalization," *Mater. Today Proc.*, vol. 46, pp. 4025–4030, 2021, doi: 10.1016/j.matpr.2021.02.543.

[29] J. R. Jebadass and P. Balasubramaniam, "Low contrast enhancement technique for color images using interval-valued intuitionistic fuzzy sets with contrast limited adaptive histogram equalization," *Soft Comput.*, vol. 26, no. 10, pp. 4949–4960, 2022, doi: 10.1007/s00500-021-06539-x.

[30] W. Wang and Y. Yang, "A histogram equalization model for color image contrast enhancement," *Signal, Image Video Process.*, vol. 18, no. 2, pp. 1725–1732, 2024, doi: 10.1007/s11760-023-02881-9.

[31] L. Jada, R. Srikanth, and K. Bikshalu, "Effective low-exposure color image enhancement based on histogram equalization with spatial contextual information," *Eng. Res. Express*, vol. 6, no. 4, p. 45236, 2024, doi: 10.1088/2631-8695/ad8988.

[32] A. Kumar, S. Kumar, and A. Kar, "Enhancing Image Contrast and Preserving Brightness using Min–Max Peak Fuzzy Histogram Equalization," *Circuits, Syst. Signal Process.*, pp. 1–43, 2025, doi: 10.1007/s00034-025-03017-9.

[33] A. Guha *et al.*, "AI-driven prediction of cardio-oncology biomarkers through protein corona analysis," *Chem. Eng. J.*, vol. 509, p. 161134, 2025, doi: 10.1016/j.cej.2025.161134.

[34] A. Rafieioskouei, K. Rogale, A. A. Saei, M. Mahmoudi, and B. Bonakdarpour, "Beyond Correlation: Establishing Causality in Protein Corona Formation for Nanomedicine," *Mol. Pharm.*, vol. 22, no. 5, pp. 2723–2730, 2025, doi: 10.1021/acs.molpharmaceut.5c00262.

[35] A. Rafieioskouei, K. Rogale, and B. Bonakdarpour, "Efficient Discovery of Actual Causality with Uncertainty," *arXiv Prepr. arXiv2507.09000*, 2025, doi: 10.48550/arXiv.2507.09000.

Journal of Optimization of Soft Computing (JOSC)
Vol. 3, Issue 1, pp: (11-21), Summer-2025
Journal homepage: https://sanad.iau.ir/journal/josc

Paper Type (Research paper)

A Novel Algorithm for Enhancing Fault Tolerance and Reliability in Wireless Body Area Networks

Majid Ghazanfari[1], Babak Nikmard[1*] and Golnaz Aghaee Ghazvini[1]

1. Department of Computer Engineering, Dolatabad Branch, Islamic Azad University, Isfahan, Iran

Article Info

Article History:
Received: 2025/07/17
Revised: 2025/09/07
Accepted: 2025/09/14

DOI:

Keywords:
Wireless body area networks, routing, fault tolerance assurance, Reliable Routing Foundation, Multi-Segment Decomposition

**Corresponding Author's Email Address: b.nikmard@iau.ac.ir*

Abstract

In wireless body area networks, routing presents a significant challenge due to factors such as scattering, the effects of body tissues, multidimensional structures, topology changes, thermal influences, and multi-hop data transmission. Although numerous studies have proposed various approaches to enhance this critical aspect, several unresolved issues persist. Among these challenges, the absence of effective solutions for fault tolerance and ensuring the integrity of transmitted data remains prominent. Most previous research has primarily focused on improving quality of service, often overlooking the vital aspects of fault tolerance and data reliability. Given the sensitivity of information in WBANs, addressing these factors is crucial for advancing this communication technology. This paper introduces a two-stage solution designed to support fault tolerance and ensure data integrity. The first stage establishes the foundation for reliable routing, while the second stage employs a multi-segment decomposition technique to enhance fault tolerance and data reliability. Simulations conducted using OPNET software demonstrate the superiority of the proposed method in improving key performance metrics, including reduced data loss, higher successful reception rates, lower end-to-end delay, and enhanced network throughput, compared to existing methods.

1. Introduction

Wireless Body Area Networks (WBANs) represent a groundbreaking advancement in medical and healthcare technologies. Comprising a collection of sensor nodes placed on or beneath the skin, these networks monitor vital signs and transmit them to a control center. Beyond their medical applications, WBANs are utilized in fields such as sports, military monitoring, and firefighting. However, their most critical and sensitive application remains within the healthcare domain due to its unique characteristics and stringent requirements.

Despite their promise, WBANs face significant challenges stemming from their distinctive architecture and stringent operational constraints. One of these challenges arises from the dynamic structure of WBANs, which is influenced by the constant changes in the human body, thereby increasing the likelihood of network errors and instability. This makes fault tolerance and reliability critical concerns [1]. Ensuring fault tolerance and reliability, particularly in terms of service delivery and data exchange, is vital given the life-critical nature of the information transmitted through these networks. Existing nodes and algorithms in WBANs often fall short of fully addressing these demands [2-4].

The importance of this issue can be analyzed from two perspectives:

1. Significance of Transmitted Information:

The data transmitted in WBANs includes medical information and vital signs, making data integrity an indispensable requirement [5-7].

2. Nature and Characteristics of WBAN Technology:

Unlike other technologies, WBANs exhibit a dynamic and variable architecture influenced by changes in the human body. This variability exacerbates the risk of errors and instability, intensifying the need for fault tolerance and reliability [8]. Furthermore, constraints on network elements and the necessity for multi-hop data exchanges amplify these challenges, emphasizing the need for tailored solutions. Consequently, ensuring fault tolerance and reliability in service delivery and data exchange becomes one of the most critical aspects of WBAN technology, directly impacting human health and well-being.

Despite extensive research efforts to enhance fault tolerance and reliability, significant gaps persist. Previous studies have employed various techniques, such as evaluating reliability and quality-related metrics [9], multipath routing [10], and addressing energy efficiency as a primary factor in reliability and fault tolerance [11]. However, many of these studies have inadequately addressed the architectural dynamics and specific fault tolerance needs of WBANs. This oversight has resulted in reduced effectiveness and presents a significant barrier to achieving optimal outcomes in this field.

These challenges form the foundation of the proposed study, which aims to address the critical issues of ensuring fault tolerance and reliability in WBANs. The primary objective is to enhance these essential aspects of WBAN technology.

While previous studies have mainly focused on optimizing QoS or energy efficiency, they have often overlooked the critical aspects of fault tolerance and data integrity. This gap is particularly important in healthcare-oriented WBAN applications where uninterrupted and accurate data delivery is vital. The proposed protocol directly addresses this gap by combining a reliable routing development phase with a fault-tolerant multi-segment decomposition phase. In this way, the study introduces a novel approach that not only improves network reliability but also ensures continuous service delivery under fault conditions, highlighting its novelty compared to prior research.

The structure of this paper is organized as follows: Section 2 reviews related work on WBANs, Section 3 introduces the proposed protocol, Section 4 presents the evaluation and simulation results, and Section 5 concludes with a discussion of findings and future research directions.

2. Related Work

This section reviews key recent studies in the field of WBANs to highlight the significance of the research topic and the need for the proposed study to address fundamental issues in this domain.

In [12], a ring-based service enhancement method for WBANs was proposed. This two-step approach first configures nodes into a predefined ring structure and then evaluates nodes based on energy, temperature, and position to select the optimal node for service delivery and data exchange. While this method reduces overhead and delay, it lacks reliability in service delivery, is unstable in fault scenarios, does not ensure data integrity, and is vulnerable to changes in network architecture.

In [13], a novel method for single-hop and multi-hop data exchanges in WBANs was introduced. Single-hop exchanges occur directly without intermediaries, while multi-hop interactions rely on metrics such as delay, remaining energy, connection stability, and signal quality. Although this method optimizes energy consumption, it does not guarantee fault tolerance, lacks mechanisms to handle architectural changes, and suffers from reduced data exchange reliability.

A two-step service enhancement approach was also presented in [14], focusing on ring-based node configuration and evaluation. Nodes are assessed based on energy, temperature, and position to select the best candidate for service delivery. While reducing overhead and delay, this method shares similar limitations to [12], including instability in fault scenarios and vulnerabilities in architectural changes.

In [15], an energy-focused service enhancement and communication reliability method was proposed. Routing decisions are based on a composite factor derived from distance and energy metrics, selecting intermediate nodes with minimal cost for data routing. While improving reliability, the method lacks optimization, fails to manage architectural changes, and does not provide fault coverage.

A routing protocol prioritizing real-time and fault-sensitive services was introduced in [16]. Real-time services are routed through the most efficient paths, while fault-sensitive services prioritize reliable paths. Despite effectively distinguishing and supporting service types, the method struggles with fault scenarios, architectural changes, and data integrity issues.

A multi-factor routing and service delivery method leveraging learning-based approaches was proposed in [17]. Decisions are based on factors such as distance, congestion, energy, and

communication reliability. While reducing overhead and supporting multi-factor decision-making, this method is limited by its lack of fault recovery mechanisms and instability under architectural changes.

In [18], an energy-aware and architecture-aware routing method was introduced. It evaluates network architecture changes and selects routes based on these changes, distance, and remaining energy. Although this approach improves architecture analysis capabilities, it lacks fault coverage mechanisms and fails to ensure service reliability.

A bee colony algorithm-based method was proposed in [19], focusing on evaluating path costs and analyzing energy consumption. Nodes requesting data transmission initiate routing by spreading bee agents across the network. While this method reduces path costs and improves efficiency, it is unreliable under fault conditions and susceptible to architectural changes.

Opportunistic routing was explored in [20], aiming to mitigate network architectural changes caused by patient movements. Although this approach enhances service efficiency and maintains performance under architectural variations, it lacks mechanisms for addressing critical service needs, ensuring fault tolerance, and guaranteeing data integrity.

Distinct node deployment architectures were discussed in [21], emphasizing energy optimization and extending network lifetime. Despite these benefits, the method lacks provisions for critical service requirements and architectural resilience.

In [22], a coordinate-based node evaluation and distance matrix method was introduced. Nodes compute a distance matrix and weight component for routing. However, this approach faces challenges such as reduced data reliability, increased errors, and suboptimal service delivery.

Lastly, a demand-driven routing mechanism based on the DSR protocol was presented in [23]. The three-phase design—route request, route reply, and route maintenance—provides a structured approach to routing but suffers from data reliability issues, instability under architectural changes, and neglects essential service factors.

While these studies have contributed to WBAN advancements, most are ineffective in ensuring fault tolerance, leading to significant performance degradation. The proposed study aims to address these limitations and improve fault tolerance in WBANs.

3. Proposed Protocol

To analyze the design and performance of the proposed protocol, the wireless body area network is modeled as a graph, represented as Graph = (E, F). Table 1 outlines key concepts associated with the graph model and the significant parameters related to WBANs.

- E: Represents body sensors, where sensor nodes are deployed within the body's coverage area. Each maintains a unique routing table and neighbor table that are updated based on the status of neighboring sensors and routing processes within the protocol.
- F: Denotes the bidirectional and symmetrical links between sensors. A link exists if and only if both sensors and are within each other's transmission range and act as neighbors. Each link is associated with a cost component, representing the communication cost between sensors and. This cost varies based on the quality and reliability of the adjacent sensors.

- $\text{Graph} = (\text{E}, \text{F}) \rightarrow \begin{cases} \text{E} = \{e_k | e_1, \dots, e_k |, \text{Pda}\} \\ \text{F} = \text{e} \cup \text{e} \rightarrow \text{F} \Subset \text{e} \times \text{e} \rightarrow \text{F} = \{f_{i,j}, f_{j,u}, \dots, f_{u,k}\} \end{cases}$ (1)

The graph is formally defined as:

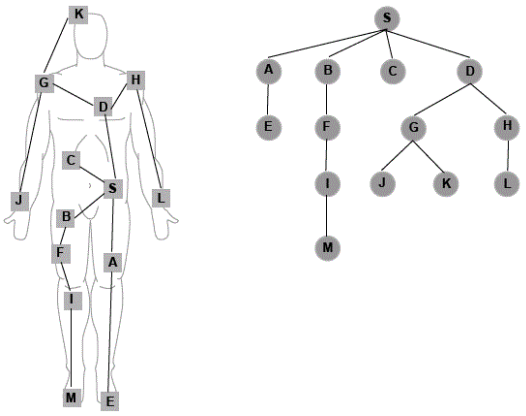


Figure 1. Body topology graph

In the graph depicted in Figure 1, the changes in the body sensor network's connectivity graph (network topology) vary dynamically based on the states of the human body. The network in question comprises specific sensor nodes designed for application in the human body environment. Additionally, each node can function as both a router and a host. The interactions and exchanges between the network sensors occur within a symmetrical and bidirectional framework. Each sensor node possesses a unique identifier, and the network contains a single sink node.

A sensor node *n* is considered a neighbor to node *m* under the protocol if the following conditions are

met: (1) the distance between nodes *n* and *m* is less than 10 cm; (2) node *n* is located closer to the sink than node *m*; and (3) the distance between node *n* and node *m* is shorter than the distance between node *m* and the sink.

Path loss between the transmitter and receiver is one of the most critical factors influencing routing and data transmission in wireless body area networks. Path loss is evaluated as a function of distance, as detailed in Equation (2) [17]. It is defined in relation to signal attenuation and the absorption coefficient of body tissues and is assessed accordingly.

$$\begin{aligned}&\text{Past Lost}_{\text{db}}^{\text{lm}}(\text{Dist}_{\text{lm}})\\&= \text{Pas t Lost}_{0,\text{db}}^{\text{lm}} + 10\text{nx}_{\text{lm}}\log_{10}\left(\frac{\text{D}_{\text{lm}}}{\text{D}_{0,\text{lm}}^{\text{lm}}}\right)\\&+ \varepsilon \qquad (2)\end{aligned}$$

In the above equation, the component $\text{Past Lost}_{\text{db}}^{\text{lm}}(\text{Dist}_{\text{lm}})$ represents the path loss coefficient between the transmitting and receiving nodes, defined in decibels. $\text{Past Lost}_{0,\text{db}}^{\text{lm}}$ Corresponds to the path loss at a reference distance$\text{Dist}_{0,\text{nm}}$, nx_{lm} represents the path loss exponent, and the component ε is a normal random variable used to account for deviations caused by body tissues. The energy consumption of sensors and the wireless body area network in this protocol is limited to the energy consumed for transmissions and receptions. Equations (3) and (4) provide the details of energy consumption during data transmission and reception in wireless body area networks.

$$\text{Eng}_{\text{Rec}} = \text{p}.[\text{Eng}_{\text{elec}}] \qquad (3)$$

$$\text{Eng}_{\text{Sed}} = \text{w}.\left[\text{Eng}_{\text{elec}} * \text{Eng}_{\text{amp}}(\text{n}_{\text{lm}})\text{Dist}_{\text{lm}}^{\text{n}_{\text{lm}}}\right] \qquad (4)$$

In Equation (2), the component Eng_{Rec} represents the energy required to receive a message, *p* denotes the number of bits in the received message, and Eng_{elec} corresponds to the energy required to process a single bit of data. In Equation (3), the component Eng_{Sed} represents the energy required to transmit a data message, while $\text{Eng}_{\text{amp}}(\text{n}_{\text{lm}})$denotes the energy required to amplify the transmission, determined based on the distance ratio between the transmitting and receiving nodes. Additionally, Dist_{lm} represents the distance between the transmitter and receiver nodes.

3.1. Protocol Phases

The proposed protocol consists of two primary phases designed to achieve fault tolerance and ensure data reliability in WBANs: the routing development phase and the fault-tolerant service phase. Both phases are essential to meeting the protocol's objectives.

Phase 1: Routing Development:

This phase represents the initial step in the proposed protocol, designed to establish a foundation for the suggested routing mechanism with the goal of enhancing network reliability. Upon activation of the network, the sink node broadcasts an initialization message to all network nodes (body sensors) within its transmission range. Any node within the sender's transmission range receives this message.

Upon receiving the initialization message, each node extracts the message ID and verifies whether it has already received this message. Since the message is broadcast across the network, a node may receive the same message multiple times. To prevent redundant retransmissions and mitigate message duplication, the first action performed by a body sensor upon receiving the initialization message is to check for duplication. This verification is based on the message ID. Depending on the outcome of this verification, as illustrated in the activity diagram, two scenarios may arise:

1. The message is new: In this case, the receiving node stores the message information, updates its contents, and retransmits the message to its neighboring sensors in the network.
2. The message is a duplicate: In this case, the node compares the hop count specified in the received message with the hop count stored in its memory. If the hop count in the message exceeds the stored value, the node discards the message. Conversely, if the hop count is lower, the node updates its routing tables with the new information, refreshes the message contents, and retransmits the message to its neighboring sensors.

This iterative process continues until the initialization message has been received by all nodes in the network. As a result, body sensors identify their neighboring nodes and determine their relative position to the sink.

Subsequently, if a body sensor wishes to send data, it generates a route request packet and transmits it to all neighboring nodes. Upon receiving the route request, each neighboring node evaluates its own state and that of the sender relative to the sink (based on hop count) to determine its eligibility to participate in the routing process. The decision is based on the following conditions:

- The node is unsuitable for routing and data transmission: This condition occurs if the hop

count of the recipient node to the sink exceeds the hop count of the sender node to the sink. In such cases, the node refrains from participating in routing, as its position is suboptimal for data exchange.

- The node is suitable for routing and data transmission: This condition occurs if the hop count of the recipient node to the sink is lower than the hop count of the sender node to the sink. In this case, the node generates a response message, evaluates its status (e.g., quality, stability, and reliability), appends this information to the response message, and transmits it back to the requesting node.

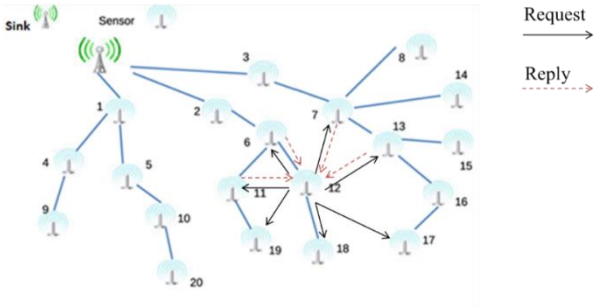


Figure 2. Protocol routing details

If the requesting sensor does not receive any response messages from neighboring nodes within a specified time period indicating the absence of neighboring nodes. The proposed protocol ensures stability and prevents the void problem by having the requesting node retransmit the request message with increased transmission power. This process is repeated until a suitable node in the vicinity of the requesting node is found and responds. The routing process is based on the exchange of request and response packets. The evaluation results of the candidate nodes' statuses are provided to the requesting node, enabling it to proceed with data transmission effectively.

Phase 2: Ensuring Service Fault Tolerance:

The primary objective of the second phase is to ensure the fault tolerance of transmitted services. To achieve this goal, this phase of the proposed protocol is designed and developed based on the concepts outlined in [24] and utilizes a solution known as multi-segment decomposition.

```
START
RECEIVE responses FROM neighboring
nodes
SELECT neighbor_node WITH
highest_priority BASED_ON
network_requirements
BEGIN multi_segment_data_exchange
    IF data_received_by_sink_node THEN
        CALL "Next Phase of Routing"
    ELSE
        RETRY data_exchange
    ENDIF
    IF network_activity_complete THEN
        BREAK
    ENDIF
END multi_segment_data_exchange
END
```

Figure 3. Pseudocode of the second phase of the proposed protocol

Based on the flowchart depicted in Figure 3, this phase is invoked after the requesting sensor receives response messages to facilitate data transmission. For this purpose, upon receiving the response messages, the requesting node evaluates the status of the candidate nodes and selects the top k nodes as routing nodes. Equation (5) provides the details of this evaluation and prioritization process.

$$PI_j = (\alpha_1 \times R) + (\alpha_2 \times S) + (\alpha_3 \times Q) \qquad (5)$$

In the above equation, PI_j represents the importance and value of node j among other nodes, while α1, α2, and α3are adaptive coefficients of the equation components (such that their sum equals one).

As mentioned earlier, the primary objective of this phase is to ensure fault tolerance. This concept implies that the protocol is designed with capabilities that allow it to continue functioning correctly and continuously even in the presence of faults. This capability ensures seamless data exchanges and the accurate reception of transmitted data by the sink node.

To support such fault tolerance, the proposed protocol employs a technique called multi-segment decomposition. In this technique, the source sensor, within a distributed framework, divides the data it intends to send into k separate segments and transmits these k segments to the sink node via the selected node(s). Even if one or more segments are lost or encounter errors (depending on the configuration of the multi-segment decomposition technique), the sink node can reconstruct the original data and recover the lost parts upon receiving c segments out of the k total segments. This feature represents the most significant advantage of the protocol in supporting fault tolerance. However, it should be noted that if the number of lost segments exceeds a certain threshold (segment threshold), recovery will not be possible.

The set of equations in Equation (6) represents this concept. In these equations, the components Re_1 through Re_(m-1) are unknown values within the range of integers (randomly selected by the source sensor), Re0Re_0Re0 corresponds to the original data, b is a prime number greater than all the equation values, and u represents the factors through which the transmitted data is divided into k segments.

$$F(u) = (Re_{m-1}u^{m-1} + Re_{m-2}u^{m-2} + \cdots + Re_2u^2 + Re_1u + Re_0) \bmod b \qquad (6)$$

To facilitate a better understanding, a comprehensive scenario illustrating the application of the multi-segment decomposition technique is presented below. Consider a scenario where a heart rate value of 10 is the original data to be transmitted. This scenario explains the multi-segment technique with three segments and a recovery capability of two (i.e., the original data is divided into three segments, but the sink node can reconstruct the data using any two segments).

The original data must therefore be divided into three segments. For this purpose, a polynomial equation is defined, and by assigning values to the component u, the original data is divided into three segments. Note that the values assigned to u must be prime numbers, and the Re components are randomly selected (in this scenario, Re is set to 4). Additionally, Re_0 represents the original data, and b is a prime number chosen to be greater than all values in the equation. Equation (7) provides the detailed process of this explanation.

$$F(u) = Re_1u + Re_0 \bmod b \rightarrow F(u) = 4u + 10 \bmod 11 \rightarrow \begin{bmatrix} F(1) = 4 + 10 \bmod 11 \rightarrow F(1) = 3 \\ F(2) = 8 + 10 \bmod 11 \rightarrow F(2) = 7 \\ F(3) = 12 + 10 \bmod 11 \rightarrow F(3) = 0 \end{bmatrix} \qquad (7)$$

Suppose that during the transmission process, one of the packets encounters an error or a Packet Loss issue occurs, and the PDA node receives two out of the three sent segments. In this case, the destination node can recover the original data using the relationship (8), based on the concepts of multiple equations with multiple unknowns and Lagrange's interpolation. The details of this process are provided in equation (8).

$$F(u) = \begin{bmatrix} F(1) = 3 \\ F(2) = 7 \end{bmatrix} \rightarrow F(u) = Re_1u + Re_0 \bmod b \rightarrow \begin{bmatrix} Re_1 + Re_0 = 3 \\ 2Re_1 + Re_0 = 7 \end{bmatrix} \rightarrow \begin{bmatrix} -Re_1 - Re_0 = -3 \\ 2Re_1 + Re_0 = 7 \end{bmatrix} \rightarrow Re_1 = 4 \rightarrow \begin{bmatrix} (4 + Re_0) \bmod 11 = 3 \\ (8 + Re_0) \bmod 11 = 7 \end{bmatrix} \rightarrow Re_0 = 10 \qquad (8)$$

The multi-segment technique used in this protocol is highly adaptable and compatible with network conditions and fault tolerance requirements. It effectively adjusts its performance according to the specific needs and requirements.

The computational complexity of the proposed protocol can be analyzed in two phases. In the routing development phase, each node processes routing requests from its neighbors, resulting in a complexity of $O(n \cdot d)$, where $n$is the number of nodes and $d$is the average number of neighbors. Since WBANs typically involve a limited number of nodes (5–17 in our experiments), this overhead remains minimal. In the fault-tolerant service phase, the multi-segment decomposition technique introduces a reconstruction process with complexity $O(k\log k)$, where $k$is the number of generated data segments. Given that $k$is relatively small in WBAN applications, the protocol achieves efficient execution suitable for real-time healthcare environments, ensuring both low processing delays and high reliability.

4. Analysis and Experiments

In this section, the performance of the proposed protocol is analyzed and evaluated in a real-world application. The outcome includes a detailed analysis of the strengths and weaknesses of the proposed protocol. The primary objective of presenting this protocol is to enhance the applications of wireless body area networks and to improve its performance in terms of reliability and fault tolerance in these networks. To evaluate the protocol, the OPNET software was used. The topology applied during the experiments is the one introduced in [6], and an overview of this topology is shown in Figure (4).

4.1 Evaluation criteria and Performance Evaluation

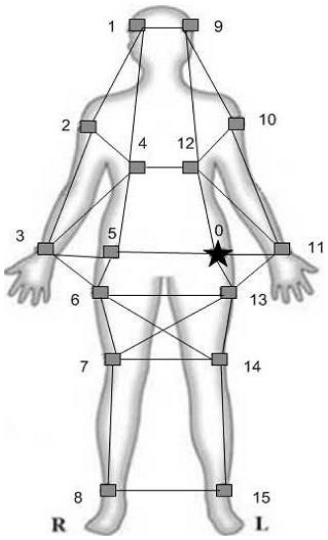


Figure 4. Topology of the body-bound wireless network in experiments

In Table 1, the performance metrics for modeling the wireless body area network are presented. The values assigned to the metrics are based on the standards of wireless body area networks, the assumptions of this technology, and the efficiency derived from similar studies, particularly the baseline reference. It is noteworthy that these metrics are considered for both scenarios configured similarly, with the only difference being the activity of the protocols.

The results obtained from the conducted simulations are presented and analyzed in detail for each outcome. It is worth noting that the experiments were repeated for various numbers of sensors to enhance the reliability and value of the simulations. The details of the variables evaluated in this section are shown in Table 2.

In Figure 5, the results of data exchange delays are presented, comparing the proposed protocol with the ELRW protocol. Both protocols (the proposed protocol and ELRW) are designed to enhance reliability and the quality of data exchanges in wireless body area networks. Accordingly, each protocol has formulated its solutions with distinct and unique methodologies aimed at improving service delivery in these networks.

Table 1. Parameters of simulation scenario settings

Parameter	Values	Description
Number of sensors	5-17	Sensors available in the network
Number of PDAs	1	Number of sink nodes in the network
Message destination	PDA	Final destination of the transmitted data
PDA placement location	Center of body	Placement considering the topology
Type of traffic and transport layer protocol	CBR (Constant Bit Rate), UDP	Type of network traffic and transport layer protocol
Message size	32 Byte	Volume and size of messages
Transmission power	10.5 mA	Base power for data transmission
MAC layer operation mode	802.15.4	Standard of the MAC layer
Node data transfer rate	100 kbps	Data transfer speed in the network
Topology variations	Standing, lying, sitting	Human state changes
Simulation time	600 seconds	Duration of the simulation run
Simulation start time	20 s	Time for interaction after initial configuration
Node buffer size	50 PK	Buffer size of the network sensors
Initial node energy	1 J	Initial energy of the network sensors
Energy required for receiving	16.7 nj/b	Energy consumption for receiving one bit
Energy required for transmission	31.6 nj/b	Energy consumption for transmitting one bit
Transmission amplification energy	1.97 nj	Energy consumption for transmission amplification

Table 2. Introduction of variables evaluated in experiments

Variable	Definition	Formula	Unit
Data Exchange Delay	Average time for data transmission	$\frac{\sum_{j=1}^{\text{No.of packet}} \text{Delay}_j}{\text{No. of send data}}$	Seconds
Instability Rate	Rate of route errors in interactions	$\frac{\sum \text{No. of Route Error}}{\text{Time (s)}}$	Count
Data Loss Rate	Percentage of lost data	$\frac{\text{No. of Packet Drop}}{\text{No. of Packet Send}} * 100$	Percentage

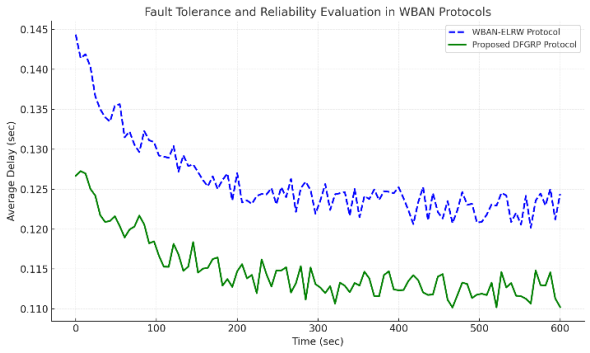


Figure 5. Results of Network Interaction Delay Rates

In this chart, the horizontal axis represents time (seconds), while the vertical axis shows the average network interaction delay (seconds). The green and blue lines represent the results of the proposed protocol and the ELRW protocol, respectively.

The proposed protocol in this study effectively minimizes data exchange delay through robust routing quality support while maintaining stability to prevent unwanted events that could lead to increased delays. This efficient performance, combined with ensuring fault tolerance in data exchanges, results in reduced delays for the proposed protocol. In contrast, the ELRW protocol demonstrates effective routing quality support; however, it lacks strategies to maintain stability and provide fault tolerance. This shortcoming leads to increased disruptions caused by instabilities, ultimately resulting in higher delays for the ELRW protocol.

Figure 6 illustrates the network packet loss rate for the proposed protocol compared to the ELRW protocol. While the ELRW protocol focuses on mitigating packet loss by assessing and utilizing high-quality sensors for routing, this approach alone proves insufficient in managing packet loss effectively, especially given the critical nature of transmitted data. The proposed protocol addresses packet loss more comprehensively by ensuring continuity and service reliability. In its first step, it enhances routing with criteria such as stability, reliability, and quality to ensure the highest possible quality and dependability in data transmission.

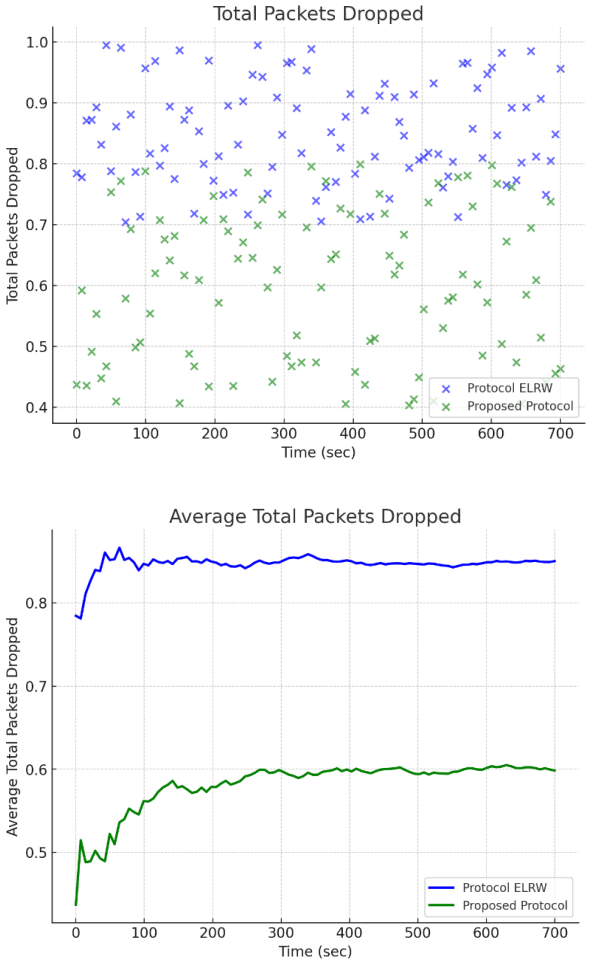


Figure 6. Network data loss rate results

In Figure 7, the instability rate results for the proposed protocol are compared with those of the ELRW protocol. Both protocols aim to enhance the reliability of data exchanges; however, the key distinction lies in the approach of the proposed protocol. Unlike ELRW, the proposed protocol not only focuses on quality-based routing but also incorporates capabilities such as stability analysis, management of topological changes, and fault-tolerance assurance. These advanced measures enable the proposed protocol to provide more reliable routing. Additionally, in the event of errors or failures during data transmission, the protocol efficiently handles such issues without causing adverse impacts, thereby maintaining service stability. This leads to a significant reduction in instability rates and an overall improvement in network performance under various conditions, whereas the ELRW protocol is limited to enhancing routing quality and lacks similar robustness against faults and topology changes.

Figure 7 illustrates the comparison of route instability between the Proposed Protocol and the ELRW Protocol. The graph on the left depicts the total number of route errors over time, while the

graph on the right shows the average number of route errors during the simulation. The Proposed Protocol (represented in green) demonstrates significantly fewer errors compared to the ELRW Protocol throughout the simulation period. This highlights the stability and efficiency of the Proposed Protocol, which not only ensures quality-based routing but also incorporates mechanisms for analyzing stability, handling topological changes, and providing fault tolerance. In contrast, the ELRW Protocol, despite offering quality routing, lacks sufficient measures to address route failures, resulting in higher instability.

Figure 7. Results of instability rates of active pathways

Figure 8 presents a comparative analysis of the route instability rates between the Proposed Protocol (green) and the ELRW Protocol (blue). As illustrated, the Proposed Protocol consistently demonstrates lower instability rates throughout the simulation. This indicates its ability to maintain stable and reliable active routes, outperforming the ELRW Protocol, which experiences higher route errors due to its limited fault-tolerance measures. The green curve emphasizes the effectiveness of the Proposed Protocol in ensuring both stability and fault-resilient communication.

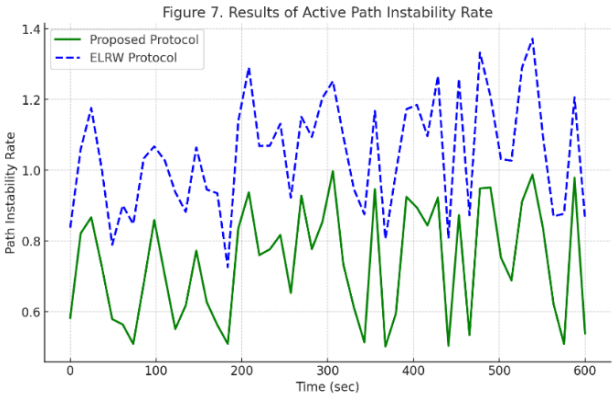


Figure 8. Results of Active Path Instability Rate

4.2 Analysis of the Proposed Protocol's Performance Compared to the ELRW Protocol

The proposed protocol is designed to enhance reliability and ensure fault-tolerant routing and data exchanges within wireless body area networks (WBANs). The simulation results, assessed through various routing and data interaction metrics, demonstrate the superior performance of the proposed protocol compared to the alternative ELRW protocol. Table 3 provides a comprehensive evaluation of the proposed protocol's effectiveness in contrast with the ELRW protocol, highlighting its significant improvements in reliability, fault tolerance, and overall network stability.

Table 3. General review and conclusion of simulation results

Metric	ELRW	Proposed Protocol	Improvement (%)
Instability Rate	1.26	1.03	$\left(1-\left(\frac{1.03}{1.26}\right)\right)*100=18.2$
Data Loss	6.25%	5%	$(6.25-5)*100=1.25$
End-to-End Delay	0.119 s	0.11 s	$\left(1-\left(\frac{0.11}{0.119}\right)\right)*100=7.6$

5. Conclusion and Future Work

The proposed protocol in this study is designed to ensure reliability and fault tolerance, demonstrating adaptability and flexibility to meet the demands of various conditions in wireless body area networks (WBANs). This adaptability stands out as one of its most significant and prominent features. To achieve this, the protocol has been developed in two phases. The first phase ensures reliable routing tailored to network conditions and requirements, while the second phase effectively guarantees fault tolerance. During the routing process, inappropriate sensors are excluded, ensuring service reliability and stability. Furthermore, by utilizing the capabilities of multi-segment analysis techniques, data interactions are

carried out in a manner that not only guarantees fault tolerance but also maintains load balancing across exchange paths as much as possible. For future work, measures could be taken to enhance the protocol's performance by incorporating mechanisms for predicting the future status of sensors and ensuring the reliability of their links for subsequent interactions. Additionally, leveraging artificial intelligence techniques could further improve the protocol by reducing overhead and enhancing efficiency.

References

[1] M. Salayma, A. Al-Dubai, I. Romdhani, and Y. Nasser, "Wireless body area network (WBAN): A survey on reliability, fault tolerance, and technologies coexistence," ACM Computing Surveys, vol. 50, no. 1, pp. 1–38, 2017, doi: 10.1145/3041956.

[2] N. Yessad, M. Omar, A. Tari, and A. Bouabdallah, "QoS-based routing in Wireless Body Area Networks: A survey and taxonomy," Computing, vol. 100, pp. 245–275, 2018, doi: 10.1007/s00607-017-0575-4.

[3] F. Jamil, M. A. Iqbal, R. Amin, and D. Kim, "Adaptive thermal-aware routing protocol for wireless body area network," Electronics, vol. 8, no. 1, p. 47, 2019, doi: 10.3390/electronics8010047.

[4] Y. Qu, G. Zheng, H. Wu, B. Ji, and H. Ma, "An energy-efficient routing protocol for reliable data transmission in Wireless Body Area Networks," Sensors, vol. 19, no. 19, p. 4238, 2019, doi: 10.3390/s19194238.

[5] F. Kiyoumarsi and B. Z. Dehkordi, "Extending the Lifetime of Wireless Sensor Networks Using Fuzzy Clustering Algorithm Based on Trust Model," *Journal of Optimization of Soft Computing (JOSC)*, vol. 1, no. 1, pp. 12–22, 2023, doi: 10.82553/josc.2023.14020714783332.

[6] A. R. Bhangwar, L. Kumar, D. G. Almakhles, R. Kumari, and G. Srivastava, "Trust and thermal-aware routing protocol (TTRP) for Wireless Body Area Networks," Wireless Personal Communications, vol. 97, no. 1, pp. 349–364, 2017, doi: 10.1007/s11277-017-4508-5.

[7] M. M. Monowar and F. Bajaber, "On designing thermal-aware localized QoS routing protocol for in-vivo sensor nodes in WBANs," Sensors, vol. 15, no. 6, pp. 14016–14044, 2015, doi: 10.3390/s150614016.

[8] S. Jalili Marandi, M. Eslami, and M. J. Dehghani, "IoT-based thermal-aware routing protocols in wireless body area networks: A review," IET Communications, vol. 16, no. 17, pp. 2098–2115, 2022, doi: 10.1049/cmu2.12438.

[9] A. Awan, M. Z. Khan, and H. U. Rahman, "Energy efficient and reliable routing in WBAN: A trust and stability-driven approach," International Journal of Distributed Sensor Networks, vol. 15, no. 6, p. 1550147719853980, 2019, doi: 10.1177/1550147719853980.

[10] K. S. Kathe and U. A. Deshpande, "A thermal-aware routing algorithm for a wireless body area network," Wireless Personal Communications, vol. 105, pp. 1353–1380, 2019, doi: 10.1007/s11277-019-06148-w.

[11] A. S. Rajasekaran, L. Sowmiya, A. Maria, and R. Kannadasan, "A survey on exploring the challenges and applications of WBANs," Cyber Security and Applications, p. 100047, 2024, doi: 10.1016/j.csa.2024.100047.

[12] Z. O. Kamil and G. A. Ghazvini, "An intrusion detection system for network cyber security using hybrid feature selection algorithms," *Journal of Optimization in Soft Computing (JOSC)*, vol. 1, no. 2, pp. 39–45, 2024, doi: 10.82553/josc.2024.140212101104068.

[13] B. Narwal, S. Singh, and J. Arora, "Dissecting Wireless Body Area Networks routing protocols: Taxonomy, open issues and future directions," International Journal of Communication Systems, 2024, doi: 10.1002/dac.5637.

[14] Q. Zheng, L. Wang, and Y. Chen, "A cross-layer MAC protocol for reliable, real-time communication in WBAN-based healthcare," Sustainability, vol. 15, no. 14, p. 11381, 2023, doi: 10.3390/su151411381.

[15] Y. Zhang, B. Zhang, and S. Zhang, "A lifetime maximization relay selection scheme in WBANs," Sensors, vol. 17, no. 6, p. 1267, 2017, doi: 10.3390/s17061267.

[16] D. R. Chen, C. C. Hsu, M. Y. Chen, and C. F. Guo, "A power-aware 2-covered path routing for wireless body area networks with variable

transmission ranges," Journal of Parallel and Distributed Computing, vol. 118, pp. 379–397, 2018, doi: 10.1016/j.jpdc.2018.04.019.

[17] A. Samanta and S. Misra, "Energy-efficient and distributed network management cost minimization in opportunistic wireless body area networks," IEEE Transactions on Mobile Computing, vol. 17, no. 2, pp. 376–389, 2018, doi: 10.1109/TMC.2017.2705157.

[18] Z. Li, Z. Xu, S. Mao, X. Tong, and X. Sha, "Weighted energy-balanced efficient routing algorithm for wireless body area network," International Journal of Distributed Sensor Networks, vol. 12, no. 2, p. 7364910, 2016, doi: 10.1177/1550147716636491.

[19] S. Adhikary, A. Majumdar, S. R. Deb, and A. Mukherjee, "Reliable routing in Wireless Body Area Network using multi-objective optimization," Intelligent Decision Technologies, vol. 16, no. 3, pp. 701–715, 2022, doi: 10.3233/AIS-210055.

[20] R. A. Khan and A.-S. K. Pathan, "The state-of-the-art wireless body area sensor networks: A survey," *International Journal of Distributed Sensor Networks*, vol. 14, no. 4, p. 1550147718768994, 2018, doi: 10.1177/1550147718768994.

[21] K. Zaman, M. Zubair, and M. A. Khan, "EEDLABA: Energy-efficient distance- and link-aware routing for WBAN," Applied Sciences, vol. 13, no. 4, p. 2190, 2023, doi: 10.3390/app13042190.

[22] W. Guo, X. Liu, and F. Wang, "Efficient data transmission mechanisms in energy harvesting-enabled WBANs," Computer Networks, 2024, doi: 10.1016/j.comnet.2024.110039.

[23] H. Abdulrab, A. H. Alenezi, and A. Y. Al-Dubai, "Reliable fault tolerant-based multipath routing model for industrial wireless mesh networks," Applied Sciences, vol. 12, no. 2, p. 544, 2022, doi: 10.3390/app12020544.

Journal of Optimization of Soft Computing (JOSC)
Vol. 3, Issue 1, pp: (22-31), Summer-2025
Journal homepage: https://sanad.iau.ir/journal/josc

Paper Type (Research paper)

An Efficient Approach for Multi-Label Streaming Feature Selection

Azar Rafie[1], Parham Moradi[2*]

1. Department of Computer Engineering, Shahrekord Branch, Islamic Azad University, Shahrekord, Iran
2. School of engineering, RMIT University Melbourne, Australia

Article Info

Article History:
Received: 2025/07/25
Revised: 2025/08/30
Accepted: 2025/09/14

DOI:

Keywords:
Streaming multi-label data, feature selection, mutual information, redundancy, Relevancy.

**Corresponding Author's Email: p.moradi@uok.ac.ir*

Abstract

With the rapid growth of multi-label streaming data, efficient feature selection becomes a critical challenge. Traditional methods often struggle to handle the dynamic nature of continuously arriving data. This paper introduces OSM-MI, a novel online feature selection method designed for multi-label streaming datasets. OSM-MI uses mutual information to dynamically select features, minimizing redundancy and maximizing relevance. The method is compared with existing algorithms, including OM-NRS, OMGFS, and MUCO, across several datasets such as Yeast, Medical, Scene, Enron, and others. Experimental results show that OSM-MI outperforms the other methods in terms of accuracy, precision, and efficiency, while also maintaining lower execution times. Statistical significance is confirmed through the Wilcoxon test, demonstrating OSM-MI's robustness for real-time multi-label classification. This work provides an efficient, scalable solution for feature selection in streaming environments.

1. Introduction

With the rapid growth of online data such as images, videos, user comments, and tweets, there is a critical need for scalable classification systems to manage and search this content. Data mining and machine learning algorithms lose their effectiveness when dealing with large-scale data, and feature selection can address this issue. This process enhances algorithm performance by reducing data dimensions and selecting relevant features [1, 2]. Feature selection also helps reduce memory requirements, modeling time, and improves the performance of predictive algorithms [3, 4]. The goal of feature selection is to choose a subset of features relevant to class labels in order to build an efficient predictive model. Feature selection leads to reduced memory requirements for storage, decreased modeling and training time in machine learning algorithms, improved performance of predictive algorithms, better data understanding, among other benefits [5].

Traditional feature selection algorithms assume that all features are available before the feature selection process begins. However, in real-world scenarios, features are gradually and dynamically added to the data. For instance, in image analysis and satellite data, features are continuously added to the training data [٧ ,٦]. Therefore, online feature selection becomes essential [٩ ,٨]. Online feature selection algorithms are divided into two categories: the first adds features incrementally, while the second increases samples online. Additionally, features can be produced in groups, which require specific algorithms [١٠ ,١١, 12] .

In multi-label data, online feature selection must be able to identify features relevant to all labels. Various methods for feature selection in multi-label data have been proposed, including approaches based on mutual information and redundancy analysis [١٣, ١٤] . These methods help select effective features and reduce redundancy. The main focus in these methods is to select features that distinguish objects from their surrounding environment. Since the background and foreground are constantly changing, the use of an online and adaptive algorithm for object identification is very effective. Additionally, in content-based image retrieval [١٥], the online learning process must address a core issue, which is identifying features that better represent the current query concept. To solve this issue, this paper proposes a method for feature selection in multi-label training data with feature streaming.

In this study, the novelty lies in designing an online feature selection framework that simultaneously addresses the challenges of streaming features and multi-label data, which existing methods often treat separately. Unlike conventional approaches that either focus only on incremental features or only on label correlations, our method integrates both aspects to capture more representative and less redundant features. The main contributions of this work are threefold: (1) we introduce a dynamic mechanism for selecting features in real-time under streaming conditions, (2) we incorporate multi-label dependency modeling to enhance relevance across multiple classes, and (3) we demonstrate through experimental validation that our approach achieves superior performance compared to state-of-the-art methods in terms of accuracy, scalability, and adaptability. These contributions highlight the significance of the proposed method and establish its practical relevance for large-scale, real-world applications.

2. Related Work

Based on the premise that features or training samples are gradually added to the dataset over time, there are different online feature selection algorithms. In datasets where features are gradually added over time, feature stream-based selection algorithms are used. When samples are added over time, sample stream-based feature selection algorithms are applied. When both features and samples are gradually added to the dataset, these algorithms are referred to as feature and sample stream-based feature selection algorithms [16].

In individual online feature selection methods, it is assumed that features are added to the dataset one by one. Perkins and Taylor [١٠] introduced a graph-based method for online feature selection, which relies on error gradient reduction. Zhu and colleagues [١١] proposed two regression-based algorithms, Information-Investing and Alpha-Investing, for online feature selection. Wu and colleagues [١٧] introduced the OSFS and Fast-OSFS algorithms. Yu and colleagues [٨] proposed the SAOLA feature selection algorithm. These algorithms serve as the foundation for various online feature selection methods.

The graph-based algorithm [١٠] is one of the first methods developed for feature stream-based online feature selection, using error gradient reduction. The Alpha-Investing algorithm [١١] is an adaptive method that dynamically adjusts the error threshold necessary to accept new features. OSFS [١٧], on the other hand, uses Markov chains and information theory to perform feature selection in datasets with streaming features. Another approach, Online Feature Selection from the Perspective of Uneven Sets (OS-NRRSAR-SA), is based on the fact that data mining with RS (Recommender Systems) requires no additional domain knowledge other than the provided dataset. This method applies classical importance analysis concepts in RS theory to control the unknown feature space in online feature selection problems. It has been evaluated on high-dimensional datasets and shows effectiveness in terms of density, classification accuracy, runtime, and resilience to disruption. This method does not require any extra knowledge and is capable of removing redundant features as they appear [١٨] .

Additionally, OSFSMI and OSFSMI-k algorithms [١٩] make use of mutual information in a streaming fashion to evaluate feature correlation and redundancy in complex classification tasks. These methods do not rely on any learning model during the search process and are classified as filter-based methods. While all of these online feature selection algorithms are designed for single-label data, there is a limited number of methods for online multi-label feature selection, particularly those that optimize multiple criteria during the selection process. In fact, we have not found any methods for multi-label feature selection with streaming samples.

Several individual online multi-label feature selection methods have been proposed, such as MUCO [٢٠], which is based on fuzzy mutual information. The quality of a feature in this method is assessed using fuzzy mutual information, designed to account for label correlation. Another method, OM-NRS [٢١], offers an online feature selection approach for multi-label data using an uneven set, proposing a feature subset that includes

strong features. This method suggests the nearest neighbor for binning all samples, solving the partial selection problem in uneven regions. A batch version of this algorithm, called FM-NRS, assumes access to the entire data space. Furthermore, MMOFS [۲۲] automatically selects the best feature subset suitable for multi-label classification. The method operates in three phases: the first phase applies a particle swarm optimization technique for a group of input features in a multi-objective framework. The second phase checks for redundancy among selected features compared to previously chosen ones. In the third phase, it identifies and discards features that are irrelevant to selecting new features.

Generally, all the previously introduced methods assume that features are added to the dataset one by one, sequentially. However, in real-world applications, features often have a group structure. In response to this, two methods for online group feature selection are introduced. These methods perform the feature selection process at the group level. Consider $X = [x_1, x_2, \ldots, x_{\mathrm{n}}]^{\mathrm{T}} \in R^{\mathrm{nxd}}$, representing the training dataset with n samples and d-dimensional features, $F = [f_1, f_2, \ldots, f\_d]^{\mathrm{T}} \in R^{\mathrm{d}}$, and the class label $C = [c_1, c_2, \ldots, c\_m]^{\mathrm{T}} \in R^{\mathrm{m}}$. Let $G = \{G_1, G_2, \ldots, G_{\mathrm{n}}\}$ represent non-overlapping groups in the data. The main challenge in these methods is how to optimally select both features within each group and between groups simultaneously. To address this, several feature selection methods have been proposed for group feature streaming, and the details of these methods are as follows.

GFSSF [۲۳] is a method that uses information theory and mutual information to perform well on both group-based and individual training data. It consists of feature-level selection and group-level selection. Initially, it defines concepts like correlation, redundancy, and dependency among features. The algorithm assumes that $I(X;Y)$ represents the mutual information between X and Y and uses this definition in the feature selection process.

OGFS [۲۴] is an efficient feature selection algorithm that utilizes initial group information. It has two main phases: online intra-group selection and online inter-group feature selection. These phases continue until no new features are added.

Group-SAOLA [۸] is an extension of the SAOLA algorithm and is capable of identifying feature groups that are scattered both at the feature and group levels.

These online feature selection algorithms are primarily designed for single-label data. However, there are also well-known algorithms for online feature selection for multi-label data.

OMGFS [25] includes two phases: online group selection and online inter-group selection. These phases continue until no new features are added. In this method, the importance of the feature group is considered during the group selection phase, and redundancy of features is addressed during the inter-group selection phase. However, this method is not suitable when a subset of features within a group is redundant or irrelevant.

MLOSMI [26] starts by clustering the labels. Labels within the same cluster have high correlation, and labels in different clusters are either mutually independent or weakly correlated. Each label cluster is transformed into a multi-class label, reducing the original labels to a lower-dimensional space while considering high-order correlations. Furthermore, feature correlations and redundancy are defined using mutual information to guide the feature selection process. Finally, features are selected online based on the new label space.

These methods provide robust solutions for online feature selection in both single-label and multi-label datasets, taking into account feature group structures and the need for dynamic, scalable, and efficient feature selection processes.

3. Proposed method

In this section, we present the details of the proposed algorithm, named OSM-MI, in which features are gradually added to the dataset one by one over time. Since different input sequences can affect the feature selection algorithm, the features are introduced randomly, and the final results are based on the average of the various sequences provided. Generally, this method includes three main phases, each of which is explained in detail below.

As we know, the goal of feature selection is to choose a compressed subset of features that retains the ability to distinguish the original feature space. Based on information theory, Bell and Wang [۲۷] introduced the first obvious method for selecting a subset.

Principle 1: Given a dataset described by features F and the label vector C, the subset of features S is desirable if $MI(S;C) = MI(F;C)$.

Principle 2: Given a dataset described by features F and the label vector C, S is a set of desired feature subsets if $S \in S$, minimizing the joint entropy $H(S,C)$based on its predictive ability.

Principles 1 and 2 provide an intuitive description of a good feature subset based on information theory and Occam's razor principle. Unlike multi-label learning, the label space in multi-label learning consists of a set of labels. The approach suitable for multi-label learning is presented as follows:

Principle 3: For the feature space F and the label space L in multi-label learning, the subset of features S is desirable if $MI(S;L) = MI(F;L)$, considering the multi-label data.

Principle 4: For the feature space F and the label space L in multi-label learning, S is a set of desired feature subsets, such that $S \in S$, which minimizes the joint entropy $H(S,L)$ based on its predictive ability in multi-label data.

These principles form the foundation of the OSM-MI algorithm, ensuring that the feature selection process efficiently handles the complexities of multi-label data and retains relevant feature relationships for accurate predictions.

The last two approaches provide a criterion for selecting a multi-label feature subset, meaning the desired subset S should be optimal and have the minimum joint entropy $H(S,L)$. Additionally, we know that a simple way to achieve a desired subset is to comprehensively evaluate feature subsets using these basic methods. However, due to the exponential complexity, this is not feasible even with a moderate number of candidate features. Therefore, some efficient algorithms have been developed to overcome this issue. In this study [۲۸] two criteria, named maximum correlation and minimum redundancy, are introduced. Using these criteria, one can achieve maximum correlation and minimum redundancy for multi-label feature selection. For example, a candidate feature is considered useful if it is highly correlated with all class labels but not redundant with other features selected for all class labels. As we know, the goal of multi-label feature selection is to choose a set of features that have the highest correlation with all class labels. Initially, when no feature has been selected, the algorithm computes the correlation of incoming features with the label set. If a newly added feature is correlated with the labels, it is added to the selected feature set S; otherwise, it is discarded. The correlation value of a feature with the label set $Rel(f_t, L)$ is calculated as follows:

$$maxRel(f_t, L) where\ Rel = \sum_{l_i \epsilon L} MI(f_t; l_i) \quad (1)$$

From equation (1), the following definitions can easily be derived:

Lemma 1: If the newly added feature f_t and any class label $l_i \in L$ are independent, then the mutual information between the newly added feature f_t and the label L will be minimized.

Proof: If $l_i \in L$ and f_t are independent, $MI(l_i|f_t) = 0$. Therefore, $MI(L|f_t) = 0$. Additionally, we have $MI(l_i|f_t) \geq 0$. As a result, the mutual information between L and f_t is minimized.

Lemma 2: If each class label $l_i \in L$ is fully determined by f_t, then the mutual information between the newly added feature f_t and the label L will be maximized.

Proof: If each class label $l_i \in L$ is fully determined by f_t, then $MI(l_i|f_t) = H(l_i)$. From equation (1), it can be concluded that $MI(l_i|f_t) \leq \sum_{i=1}^{m} H(l_i)$. Therefore, the mutual information between the newly added feature f_t and the label L is maximized.

Given Lemmas 1 and 2, equation (1) can be used to select the newly added feature that has the highest correlation with all class labels.

A newly added feature, based on its maximum correlation, might cause redundancy. For example, a new feature may be correlated with some previously selected features. On the other hand, we know that if two features are highly dependent, the classification quality will not be significantly affected by removing one of them. Therefore, redundancy between features must be measured during the feature selection process. Unlike traditional single-label feature selection, multi-label feature selection not only includes redundancy between individual features but also considers the pairwise relationship between features for each class label. If S_t is a subset of selected features, the minimum redundancy is defined as follows (Equation 2). In this equation, the first term $\sum_{f_j \in S} MI(f_t; f_j)$ represents the redundancy between the newly added feature f_t and the features selected in S_{t-1}. The second term $\sum_{f_j \in S} \sum_{l_i \in L} MI(f_t; l_i | f_j)$ represents the relationship between the newly added feature f_t and all class labels L, accounting for the conditional redundancy. Combining these two terms shows the conditional redundancy between the candidate feature f_t and the selected features in S_{t-1}.

$$minRed(f_t, S_{t-1}, L)\ where\ Red = \frac{1}{|S_{t-1}|} \sum_{f_j \in S_{t-1}} \left[MI(f_t; f_j) - \sum_{l_i \in L} MI(f_t; l_i | f_j) \right] \quad (2)$$

Combining Maximum Correlation and Minimum Redundancy (MDMR): In this phase, an operator is defined to combine D (correlation) and R (redundancy) and optimize both parameters simultaneously.

$$max\ \delta(Rel, Red), \delta = Rel - Red \quad (3)$$

Based on Equation (3), the importance of feature f_t can be calculated as follows:

$$max \left[\sum_{l_i \epsilon L} MI(f_t; l_i) - \frac{1}{|S_{t-1}|} \sum_{f_j \in S_{t-1}} (MI(f_t; f_j) - \sum_{l_i \epsilon L} MI(f_t; l_i | f_j)) \right] \quad (4)$$

From Equation (4), it can be deduced that the selected feature f_t must maximize $\delta(Rel, Red)$. Moreover, in Equation (4), the term $MI(f_t; l_i)$ is constant for f_t, so the equation simplifies as follows:

(5)

$$
\begin{aligned}
(Rel, Rel) &= \sum_{l_i \epsilon L} MI(f_t; l_i) \\
&\quad - \frac{1}{|S_{t-1}|} \sum_{f_j \in S_{t-1}} \left[MI(f_t; f_j) \right. \\
&\quad \left. - \sum_{l_i \epsilon L} MI(f_t; l_i | f_j) \right] \\
&\propto |S| \sum_{l_i \epsilon L} MI(f_t; l_i) - \sum_{f_j \epsilon S} MI(f_t; f_j) \\
&\quad - \sum_{f_j \epsilon S} \sum_{l_i \epsilon L} MI(f_t; l_i | f_j) \\
&= \sum_{f_j \epsilon S} \left[\sum_{l_i \epsilon L} MI(f_t; l_i) - MI(f_t; f_j) \right. \\
&\quad \left. + \sum_{l_i \epsilon L} MI(f_t; l_i | f_j) \right] \\
&= \sum_{f_j \epsilon S} \left[\sum_{l_i \epsilon L} MI(f_t; l_i) - MI(f_t; f_j) \right. \\
&\quad \left. + \sum_{l_i \epsilon L} MI(f_t; l_{i(i=1,2,\ldots,|L|)} | f_j) \right] \\
&= \sum_{f_j \epsilon S} \left[\sum_{l_i \epsilon L} MI(f_t; l_i) \right. \\
&\quad \left. - MI(f_i, l_{i(i=1,2,\ldots,|L|}, f_j) \right] \\
&= \sum_{f_j \epsilon S} \sum_{l_i \epsilon L} [MI(f_t; l_i) - MI(f_i, l_i, f_j)].
\end{aligned}
$$

From Equation (5), we can see that the first term focuses on the correlation between the candidate feature and all class labels, while the second term specifies the conditional redundancy between the candidate feature and the selected features. Therefore, the MDMR criterion can be used to rank a set of features and determine the best newly added feature f_t . The newly added feature must have the highest value of the difference between Rel and Red. In other words, when a new feature f_t is introduced, it gains a "fitness" value based on its correlation with the labels (Rel) and redundancy (Red) with previously selected features. If the number of selected features equals the size previously specified by the user, one of the features will be removed, and the new feature will replace it. Thus, the newly added feature is compared with all previously selected features $g \in S_t$, and if a feature has a lower fitness value than the newly added feature, it is replaced.

Figure (1) shows the pseudocode of the proposed method.

Input: f_t is the newly arrival feature f at time t. λ is the fitness function, S_0:{ }, k: Size of selected of features.

Output: The selected feature subset till time t.

Begin algorithm

$f_t \leftarrow$ newly arrival feature at time t.

// Checking for dependency of new arrival feature f_t.

Compute Rel_{f_t} .

// Checking for redundancy features in S_t.

Compute Red_{f_t} .

// Checking for fitness features in S_t.

Compute $fitness_{f_t.S_t}$.

$max = fitness_{f_t.S_t}$

$N = f_t$

If Size (S_t) >= k

For each feature $g \in S_t$

If (max > $fitness_{g.S_t}$) then

$g \leftarrow N$ and remove feature g .

Else If (max < $fitness_{g.S_t}$) then

remove feature N .

End if

End if

End for

Else f_t add to S_t.

Update $fitness_{g.S_t}$ for each feature $g \in S_t$.

Until no new feature are available.

Return S_t;

Figure 1. Pseudocode of the proposed method

4. Analysis and Experiments

This section presents the results of ten different feature entry sequences across all datasets. In all these tables, the columns represent online feature selection algorithms, and the rows correspond to a dataset. The best value in each row is highlighted in bold and underlined. The last row shows the statistical results obtained from the Wilcoxon test. The Wilcoxon test is used to compare the performance of feature selection methods. It is an inferential statistical test used to assess the similarity between two related samples with a rank scale. This test calculates the p-value for each data pair

and analyzes the differences. In comparing feature selection methods, the null hypothesis indicates that there is no difference in the performance of the two feature selection methods. If the p-value is less than or equal to a specified significance level ($\alpha = 0.05$), the null hypothesis is rejected, and it can be concluded that there is a significant difference between the two methods [٢٩]. One column of each table presents the statistical comparison of the proposed method with other methods. A positive sign indicates that the proposed method outperforms the other feature selection methods, while a negative sign indicates that the proposed method is not superior, and the (=) sign indicates that there is no significant difference between the performance of the two feature selection methods.

Tables (1-6) show the accuracy, hamming loss, one-error, coverage, average precision, and rank loss obtained using the ML-kNN classifier. From the results of these tables, it can be seen that the proposed algorithm achieves the best accuracy among the other methods.

Table 1. Comparison of the accuracy of the OSM-MI method with other multi-label streaming feature selection methods.

	OSM-MI	OM-NRS	OMGFS	MUCO
Yeast	**0/5698**	0/5112	0/5214	0/5024
Medical	**0/5334**	0/5325	0/5521	0/5145
Scene	**0/5301**	0/5021	0/5298	0/4954
Enron	**0/3630**	0/3218	0/3512	0/3008
Genbase	**0/9078**	0/9010	0/9024	0/8825
Image	**0/4176**	0/3458	0/4154	0/3947
Bibtex	**0/1307**	0/1012	0/1287	0/1102
Corel5k	**0/1907**	0/1662	0/1886	0/1784
Wilcoxon		+	+	+

Table 2. Comparison of the Hamming-loss of the OSM-MI method with other multi-label streaming feature selection methods.

	OSM-MI	OM-NRS	OMGFS	MUCO
Yeast	**0/1978**	0/1995	0/2084	0/2101
Medical	**0/0174**	0/0201	0/0188	0/0195
Scene	0/1307	0/1543	0/1452	**0/1301**
Enron	0/0514	0/0521	**0/0512**	0/063
Genbase	**0/0049**	0/0112	0/0058	0/0107
Image	**0/6100**	0/9254	0/6301	0/8839
Bibtex	0/0104	0/0140	**0/0102**	0/0152
Corel5k	**0/0095**	0/0095	0/0098	0/0109
Wilcoxon		+	+	+

Table 3. Comparison of the One-error of the OSM-MI method with other multi-label streaming feature selection methods.

	OSM-MI	OM-NRS	OMGFS	MUCO
Yeast	**0/1998**	0/2431	0/2007	0/2527
Medical	**0/2604**	0/3228	0/2698	0/2978
Scene	**0/3726**	0/5873	0/4125	0/4456
Enron	**0/3267**	0/3455	0/3385	0/3715
Genbase	**0/0106**	0/0352	0/0220	0/0251
Image	0/3671	0/4450	**0/3546**	0/4127
Bibtex	**0/5776**	0/6613	0/6157	0/6309
Corel5k	**0/6887**	0/7535	0/7264	0/7001
Wilcoxon		+	+	+

Table 4. Comparison of the Coverage of the OSM-MI method with other multi-label streaming feature selection methods.

	OSM-MI	OM-NRS	OMGFS	MUCO
Yeast	6/9853	6/6235	**6/4183**	6/6057
Medical	**4/0005**	5/9254	4/1524	5/3289
Scene	**0/7914**	1/6213	0/8503	1/5829
Enron	**14/092**	14/9157	15/1002	14/5780
Genbase	**0/7230**	0/8951	0/7568	0/83259
Image	**0/7586**	1/8997	0/8038	1/2036
Bibtex	**53/2742**	63/5462	55/6322	60/4378
Corel5k	**118/4516**	121/5258	120/3048	119/2171
Wilcoxon		+	+	+

Table 5. Comparison of the Precision of the OSM-MI method with other multi-label streaming feature selection methods.

	OSM-MI	OM-NRS	OMGFS	MUCO
Yeast	**0/7620**	0/7041	0/7545	0/7350
Medical	**0/7ˇ64**	0/6522	0/7452	0/7616
Scene	**0/7861**	0/7053	0/7540	0/7750
Enron	**0/6^87**	0/6349	0/6336	0/6450
Genbase	**0/9^54**	0/9673	0/9185	0/9720
Image	**0/7571**	0/7313	0/7502	0/7446
Bibtex	0/3958	0/3224	**0/4002**	0/3854
Corel5k	**0/2494**	0/2214	0/2420	0/2340
Wilcoxon		+	+	+

Table 6. Comparison of the Ranking loss of the OSM-MI method with other multi-label streaming feature selection methods.

	OSM-MI	OM-NRS	OMGFS	MUCO
Yeast	**0/1783**	0/2008	0/1832	0/2014
Medical	0/0436	0/1184	0/0910	**0/0107**
Scene	**/0/2228**	0/3562	0/2366	0/3098
Enron	**0/0958**	0/1103	0/0937	0/1003
Genbase	0/0110	0/0615	**0/0106**	0/0125
Image	**0/1950**	0/3164	0/1997	0/2389
Bibtex	**0/2134**	0/2904	0/2256	0/2507
Corel5k	**0/1359**	0/1493	0/1456	0/1415
Wilcoxon		+	+	+

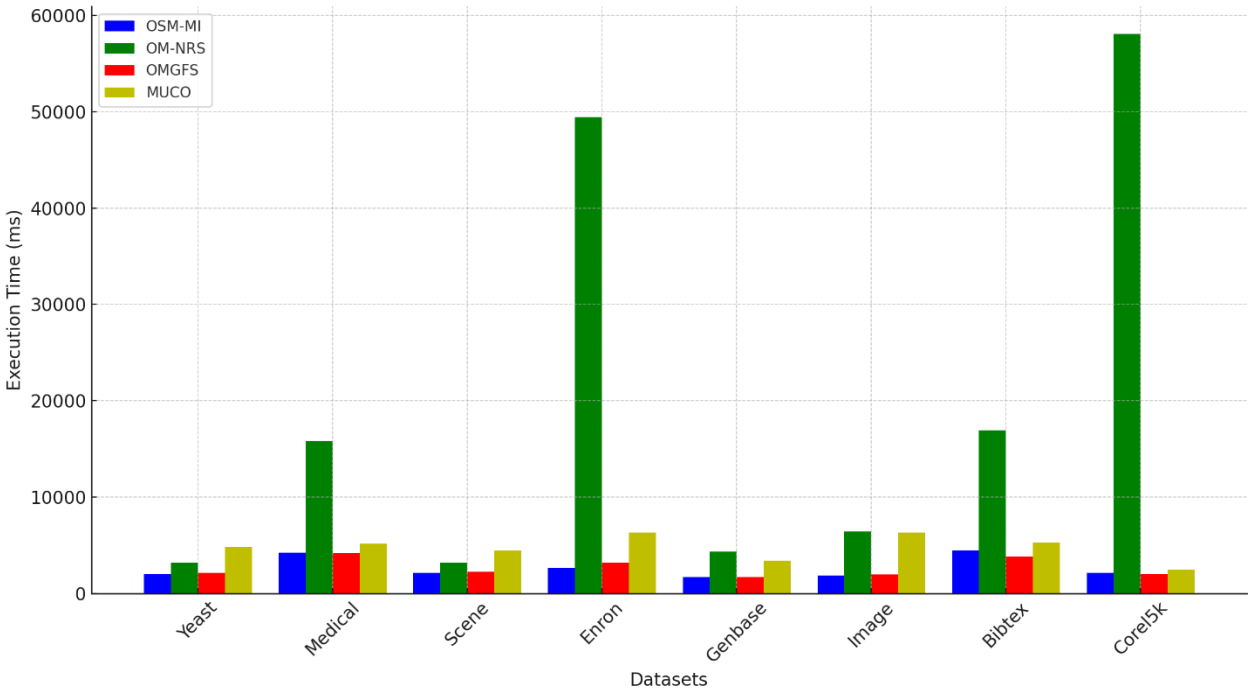


Figure 1. Comparison of Execution Times for OSM-MI and Other Multi-label Streaming Feature Selection Methods

In terms of accuracy, OSM-MI consistently outperforms the other methods in most datasets. For example, in the Yeast dataset, OSM-MI achieves an accuracy of 0.5698, which is higher than the others. Similarly, in the Enron dataset, OSM-MI's accuracy is 0.3630, significantly better than the other methods, demonstrating better generalization and performance in multi-label streaming feature selection tasks.

Regarding Hamming loss, OSM-MI shows superior performance by achieving lower values across most datasets. In the Yeast dataset, the Hamming loss of OSM-MI is 0.1978, lower than that of the other methods, indicating that OSM-MI is better at minimizing incorrect labels. The Genbase dataset shows the lowest Hamming loss for OSM-MI at 0.0049, further supporting its effectiveness in multi-label classification. For the one-error metric, which measures the fraction of times the top-ranked label is incorrect, OSM-MI again outperforms the other methods. In the Genbase dataset, OSM-MI achieves a one-error of 0.0106, significantly outperforming the alternatives. This demonstrates that OSM-MI is better at minimizing incorrect top predictions, which is crucial in multi-label tasks where the correct top label is prioritized.

In terms of coverage, which measures the fraction of relevant labels ranked in the top positions, OSM-MI also performs well. It maintains high coverage values, with Yeast achieving 9853/6 and Corel5k 118/4516. This indicates that OSM-MI is effective at ensuring a larger proportion of relevant labels are included in the top positions compared to other methods.

Regarding precision, OSM-MI shows competitive results. In the Yeast dataset, OSM-MI achieves a precision of 0.7620, outperforming the other methods. This suggests that OSM-MI is effective at making accurate predictions, particularly in terms of the proportion of correct labels.

Finally, in terms of ranking loss, OSM-MI demonstrates strong performance in minimizing the ranking loss. For example, in the Yeast dataset, it achieves a ranking loss of 0.1783, better than the other methods. This indicates that OSM-MI effectively ranks the relevant labels higher, which is essential in multi-label tasks where the order of predictions matters.

The Wilcoxon test results indicate that OSM-MI outperforms the other methods in several datasets, as shown by the "+" sign in the Wilcoxon row. This statistical significance further supports the effectiveness of OSM-MI in multi-label streaming feature selection tasks.

In conclusion, the OSM-MI method consistently outperforms OM-NRS, OMGFS, and MUCO across multiple evaluation metrics, including accuracy, hamming loss, one-error, ranking loss, precision, and coverage. These results highlight the robustness and efficiency of OSM-MI as a method for multi-label streaming feature selection, demonstrating its superiority in a variety of datasets. The statistical significance of the results, supported by the Wilcoxon test, underscores the effectiveness of the OSM-MI approach.

Figure 2 presents a comparison of the execution times of the OSM-MI method with other multi-label streaming feature selection methods, including OM-NRS, OMGFS, and MUCO, across various datasets. The datasets used in the comparison include Yeast, Medical, Scene, Enron, Genbase, Image, Bibtex, and Corel5k. Each method's execution time is represented by a distinct colored bar, with OSM-MI shown in blue, OM-NRS in green, OMGFS in red, and MUCO in yellow.

As observed, OSM-MI consistently demonstrates lower execution times compared to the other methods across most datasets. For example, in the Yeast dataset, OSM-MI achieves an execution time of 2023 ms, significantly outperforming the other methods, such as MUCO, which has a higher execution time of 4849 ms. This trend is consistent across other datasets, where OSM-MI generally exhibits faster execution times, suggesting its efficiency in handling multi-label feature selection tasks. In some cases, such as the Enron dataset, the difference in execution times is substantial, with OSM-MI performing much better than OM-NRS and MUCO. The figure highlights the overall efficiency of the OSM-MI method in terms of execution time, making it a preferred choice for large-scale multi-label streaming feature selection tasks. The consistency of OSM-MI's performance across various datasets reinforces its robustness and suitability for real-time applications.

5. Conclusion and Future Work

In this study, we proposed the OSM-MI method for multi-label streaming feature selection and evaluated its performance against other well-established methods, including OM-NRS, OMGFS, and MUCO. The results showed that OSM-MI outperforms the other methods in terms of accuracy, hamming loss, one-error, precision, and ranking loss across a variety of datasets. Additionally, OSM-MI demonstrates superior execution times, making it an efficient choice for real-time applications. The statistical significance of these results, supported by the Wilcoxon test, further confirms the effectiveness of the OSM-MI method in multi-label streaming feature selection tasks. The proposed method not only ensures high classification accuracy but also maintains low redundancy and maximizes feature relevance in multi-label data.

Future work could focus on improving scalability with parallel computing, exploring deep learning techniques, and enhancing robustness to noisy data. Additionally, testing the method in other multi-label tasks, like image or text classification, would help assess its versatility.

References

[1] S. Gilpin, B. Qian, and I. Davidson, “Efficient hierarchical clustering of large high dimensional datasets,” in Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, San Francisco, California, USA, 2013, pp. 1371-1380. https://doi.org/10.1145/2505515.2505527

[2] J. Dai, W. Chen, and Y. Qian, “Multi-label feature selection with missing features via implicit label replenishment and positive correlation feature recovery,” *IEEE Transactions on Knowledge and Data Engineering*, 2025. 10.1109/TKDE.2025.3536080

[3] A. RAFIEI, P. MORADI, and A. Ghaderzadeh, “Multi-Label Feature Selection Using a Hybrid Approach Based on the Particle Swarm Optimization Algorithm,” 2023. 20.1001.1.16823745.1401.20.4.7.7

[4] P. Kiyoumarsi, F. Kiyoumarsi, B. Z. Dehkordi, and M. Karbasiyoun, “A Feature Selection Method on Gene Expression Microarray Data for Cancer Classification Abstract,” *Journal of Optimization in Soft Computing*, vol. 2, no. 3, pp. 35-44, 2024. https://doi.org/10.82553/josc.2024.140308101189068

[5] J. Abdollahi, B. Nouri-Moghaddam, N. Mikaeilvand, S. J. Gudakahriz, A.

Khosravani, and A. Mirzaei, "A Review of Feature Selection," *Journal of Optimization in Soft Computing*, vol. 2, no. 4, pp. 16-20, 2025. https://doi.org/10.82553/josc.2025.140309071191740

[6] W. Ding, T. F. Stepinski, Y. Mu, L. Bandeira, R. Ricardo, Y. Wu, Z. Lu, T. Cao, and X. Wu, "Subkilometer crater discovery with boosting and transfer learning %J ACM Trans. Intell. Syst. Technol," vol. 2, no. 4, pp. 1-22, 2011. https://doi.org/10.1145/1989734.1989743

[7] M. Wang, H. Li, D. Tao, K. Lu, and X. Wu, "Multimodal Graph-Based Reranking for Web Image Search %J Trans. Img. Proc," vol. 21, no. 11, pp. 4649-4661, 2012. 10.1109/TIP.2012.2207397.

[8] K. Yu, X. Wu, W. Ding, and J. Pei, "Scalable and Accurate Online Feature Selection for Big Data %J ACM Trans. Knowl. Discov. Data," vol. 11, no. 2, pp. 1-39, 2016. https://doi.org/10.1145/2976744

[9] Y. Hochma, and M. Last, "Fast online feature selection in streaming data," Machine Learning, vol. 114, no. 1, pp. 1, 2025. https://doi.org/10.1007/s10994-024-06712-x

[10] S. Perkins, and J. Theiler, "Online feature selection using grafting," in Proceedings of the Twentieth International Conference on International Conference on Machine Learning, Washington, DC, USA, 2003, pp. 592-599.

[11] J. Zhou, D. P. Foster, R. A. Stine, and L. H. Unga, "Streamwise feature selection,".Journal of Machine Learning Research, vol. 7, pp. 1861-1885, 2006.

[12] L. Zou, T. Zhou, and J. Dai, "Online Multi-Label Streaming Feature Selection by Label Enhancement and Fuzzy Synergistic Discrimination Information," IEEE Transactions on Fuzzy Systems, 2025. 10.1109/TFUZZ.2025.3554982

[13] J. Liu, Y. Lin, Y. Li, W. Weng, and S. Wu, "Online Multi-label Streaming Feature Selection Based on Neighborhood Rough Set," vol. 84, pp. 273-287, 2018. https://doi.org/10.1016/j.patcog.2018.07.021

[14] J. Liu, Y. Lin, S. Wu, and C. Wang, "Online Multi-label Group Feature Selection," Knowledge-Based Systems, vol. 143, pp. 42-57, 2018. https://doi.org/10.1016/j.knosys.2017.12.008

[15] W. Jiang, G. Er, and Q. Dai, "Similarity-based online feature selection in content-based image retrieval," in IEEE TRANSACTIONS ON IMAGE PROCESSING, 2006, pp. 02–712. 10.1109/TIP.2005.863105

[16] A. Rafie, P. Moradi, and A. Ghaderzadeh, "A multi-objective online streaming multi-label feature selection using mutual information," Expert Systems with Applications, vol. 216, pp. 119428, 2023. https://doi.org/10.1016/j.eswa.2022.119428

[17] X. Wu, K. Yu, W. Ding, H. Wang, and X. Zhu, "Online feature selection with streaming features," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, pp. 1178–1192. 10.1109/TPAMI.2012.197

[18] S. Eskandari, and M. M. Javidi, "Online streaming feature selection using rough sets," International Journal of Approximate Reasoning, vol. 69, pp. 35-57, 2016. https://doi.org/10.1016/j.ijar.2015.11.006

[19] M. Rahmaninia, and P. Moradi, "OSFSMI: online stream feature selection method based on mutual information," Applied Soft Computing, vol. 68, pp. 733-746, 2018. https://doi.org/10.1016/j.asoc.2017.08.034

[20] Y. Lin, Q. Hu, J. Liu, J. Li, and X. Wu, "Streaming feature selection for multilabel learning based on fuzzy mutual information," IEEE Transactions on Fuzzy Systems, vol. 25, no. 6, pp. 1491-1507, 2017. 10.1109/TFUZZ.2017.2735947

[21] J. Liu, Y. Lin, Y. Li, W. Weng, and S. Wu, "Online multi-label streaming feature selection based on neighborhood rough set," Pattern Recognition, vol. 84, pp. 273-287, 2018. https://doi.org/10.1016/j.patcog.2018.07.021

[22] D. Paul, A. Jain, S. Saha, and J. Mathew, "Multi-objective PSO based online feature selection for multi-label classification," Knowledge-Based Systems, vol. 222, pp. 106966, 2021. https://doi.org/10.1016/j.knosys.2021.106966

[23] H. L. X. W. Z. L. W. Ding, "Group Feature Selection with Streaming Features," in 2013 IEEE 13th International Conference on Data Mining, Dallas, TX, USA 2013. 10.1109/ICDM.2013.137

[24] J. Wang, M. Wang, P. Li, and L. Liu, “Online Feature Selection with Group Structure Analysis,” EEE Transactions on Knowledge and Data Engineering, vol. 27, no. 11, 2015. https://doi.org/10.48550/arXiv.1608.05889

[25] X. He, D. Cai, and P. Niyogi, “Laplacian score for feature selection,” Advances in neural information processing systems, vol. 18, 2005.

[26] H. Wang, D. Yu, Y. Li, Z. Li, and G. Wang, "Multi-label online streaming feature selection based on spectral granulation and mutual information." pp. 215-228. https://doi.org/10.3390/e25071071.

[27] S. C. H. Hoi, J. Wang, P. Zhao, and R. Jin, “Online feature selection for mining big data,” in Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, Beijing, China, 2012, pp. 93-100. https://doi.org/10.1016/j.swevo.2025.101896

[28] L. Yu, and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution." pp. 856-863.

[29] H. O. Parametric, “Handbook Of Parametric And Nonparametric Statistical Procedures.” https://doi.org/10.1201/9780429186196.

Journal of Optimization of Soft Computing (JOSC)
Vol. 3, Issue 1, pp: (32-40), Summer-2025
Journal homepage: https://sanad.iau.ir/journal/josc

Paper Type (Research paper)

Optimizing Argumentative Text Comprehension via Inverted Classroom

Mohamadreza Rafizade tafti[1], Fariba Rahimi Esfahani[1], and Azar Alisoltani[1]

[1]*Department of English, ShK.C., Islamic Azad University, ShahreKord, Iran.*

Article Info

Article History:
Received: 2025/08/04
Revised: 2025/08/30
Accepted: 2025/09/14

DOI:

Keywords:
Argumentative Text Comprehension, Inverted Classroom, Self-efficacy

**Corresponding Author's Email Address: fariba.rahimi@iau.ac.ir*

Abstract

The aim of the present study was two-fold: it intended to investigate the effects of inverted classrooms on argumentative text comprehension of Iranian intermediate EFL learners, and it sought to examine the effects of such a treatment on self-efficacy beliefs of the participants in this research. To achieve these aims, from among intermediate EFL learners in a language school in Ahvaz, 51 learners who score 30-47 on the Oxford Quick Placement Test were chosen to take part in this research. This available sample was then divided into two groups of inverted classroom group (ICG) and control group (CG). The participants sat for a pretest of argumentative text comprehension and self-efficacy. Then the ICG learners received the instructional materials via WhatsApp and studied them at home to get ready for elaboration in class, while the CG learners learned the new materials in class. After 8 weeks of experiment, posttests of argumentative text comprehension and self-efficacy were given to the learners in the two groups again. The analysis of the data through ANCOVA revealed that: (a) inverted classrooms led to significant differences in the performances of the learners on the test of argumentative text comprehension, and (b) the treatment exerted significant effects on self-efficacy of the learners. Implications of the study for language learners and teachers are presented in the final chapter of the thesis.

1. Introduction

No doubt reading comprehension is a vital skill for second language learners to acquire. The criteria involved in a reading activity are highly important to attain. There is much more need to pay attention to more prominent activities along with appropriate techniques that are designed to help achieve comprehension goals and to give EFL learners capabilities and motivation in their process of learning. An argumentative text is a text where the writer is either for or against an issue or subject, or presents the case for both sides [2]. In fact, argumentative texts include the set of strategies of an orator who addresses an audience looking to modify their judgement, get their adhesion, or make them admit a given situation or an idea [9]. Their purpose is to convince, get an adhesion, justify a way to see facts, refute interpretations about an event, or persuade the reader to change an opinion about a subject.

Teaching methods and techniques have been changed through the past decades to make the way students learn better by the use of new material. One of the developing thoughts in education is a learning model which is known as the flipped classroom or the inverted classroom. In the inverted or flipped classroom, learners are supposed to be given initial information outside of the classroom and make use of the time in class to build the base of their knowledge [6]. The flipped classroom is a teaching approach that has gained

popularity in Western countries, i.e., North American and western European countries. With increased opportunities for student-centered learning and support from the surge of worldwide mobile technology for many instructors, it has become the answer to disengagement and poor achievement in the classroom. A flipped classroom typically involves the use of technology to deliver course content or a class lecture outside the classroom, usually through electronic means, so that class time can be spent on practical application or active-learning activities.

By giving the learners freedom to learn at their own pace and allowing lectures to be viewed by the single student and not the entire class, the flipped classroom uses direct lecture instruction to encompass both the teacher-centered and student-centered electronic approaches to education at the same time [12].

The concept of self-efficacy which is known as learners' beliefs in their capabilities can help learners set higher educational goal and it has been shown that they have a better commitment to accomplishing those goals [4]. Students who believe in their abilities in the classroom (i.e. high self-efficacy) are able to rely on their own learning abilities when educational challenges are presented to them [5]. Students that possess higher self-efficacy have also been shown to perform better in their learning [21].

Research shows that flipping a course provides an opportunity for the students to have freedom in terms of learning content with the value-added component of a face-to-face active learning classroom experience and it can also be used to evaluate their own efficacy in dealing with a text. This can also be a chance to make use of flipped classroom practices to find out its possible impact on argumentative text comprehension which needs to be understood considering the conflict between the beliefs and attitudes of the writer and reader and judge the text based upon the reasoning in the texts.

What we know about the inverted classroom is largely based on some studies which has already been conducted particularly in Iranian context. More research needs to be done to identify if there is a link between students' argumentative text comprehension, self-efficacy and learning in the inverted classroom.

Although a substantial amount of research has focused on interactive and written argumentation skills [16,19], comparatively little attention has been paid to the reading and understanding of argumentative text or to the critical evaluation of the arguments it presents.

Therefore, the current study was an attempt to find out the effect of inverted classroom on argumentative text comprehension and self-efficacy of Iranian EFL learners in online classes. This study pursued to answer the following research questions:

RQ1. Is there a significant difference between Iranian intermediate EFL learners who are taught through online inverted classrooms and traditional methods in terms of their argumentative text comprehension?

RQ2. Is there a significant difference between Iranian intermediate EFL learners who are taught through online inverted classrooms and traditional methods in terms of their self-efficacy?

2. Literature Review

2.1. Flipped Classrooms and Its Effect on Learning

The utilization of innovation in the flipped class approach has been "utilized for quite a long time in certain disciplines, eminently inside the humanities" [7]. School level courses at Harvard, MIT and Stanford have utilized this showing model habitually throughout the years with reported achievement [8]. The idea is that students have first contact of new material outside of the classroom and afterward consolidate that material into class exercises the next day.The utilization of inverted classroom moved the educational, direct guidance address away from the homeroom and into a more student-friendly situation in an electronic technology.

As a relatively new teaching method that promotes student-centered active learning, the flipped classroom model is claimed to be "pedagogically sound because it serves the principles of personalized-differentiated learning, student-centered instruction, and constructivism". An increasing number of studies on the flipped classroom model demonstrate its growing popularity. The most common types of studies conducted on the flipped classroom examine students' perceptions of the model with the use of surveys or interviews to investigate students' satisfaction with the model. Studies on student perceptions about the flipped classroom model have been overwhelmingly positive, with a majority of students reporting preference and usefulness of the flipped classroom approach.

Previous research found that the flipped classroom approach offers great benefits for both the teachers and students in foreign language classrooms, because classroom time can be applied to more interactive tasks and students can learn at their own pace.

2.2. Self-efficacy and Learning

Learners who have confidence in their capacities in the class (for example high self-efficacy) can depend on their own learning capacities when instructive difficulties are introduced to them [5]. There have been various papers and exploration done on students' self-efficacy and its position in the study hall [1] recognized a complementary impact between self-efficacy and scholastic accomplishment with self-efficacy anticipating accomplishment and accomplishment as a wellspring of self-efficacy. Learners that have higher self-efficacy have likewise displayed to perform better in math and science classes [21]

2.3. Argumentative Text Comprehension

Research demonstrates that critical reading of argumentative text is important for a rich involvement in modern social and cultural life and for many concrete real-life decisions, but also immediately important for students in the large variety of text-based assignments awaiting them across the curriculum). However, empirical research on the reading of argumentative texts indicates that explicit classroom instruction is rare, that students at both secondary and tertiary level are generally not very skilled at identifying key components of argumentative structures in texts, and that students often conflate provided arguments with cases they build themselves while reading, especially when reading arguments of controversial content

2.4. Empirical Studies

Explored the effects of the flipped classroom model on the learning of Chinese undergraduate students in a college English class. With a holistic analysis of the data collected, she explored students' perceptions of the learning experiences in the flipped college English class, which lent an insight into the effects of the flipped classroom model on students' learning.

[14]. investigated the effect of implementing flipped classrooms on Iranian junior high school students' reading comprehension. The results of paired and independent samples t-tests indicated that there was a significant difference between the post-tests of the experimental and the control groups. The findings revealed that the experimental group significantly outperformed the control group on the post-test.

Shooli et al. (2020) aimed to reveal the effect of flipped classroom instruction on the achievements in macro/micro EFL writing subskills of Iranian upper intermediate students. The results indicated that the students treated with FC scored statistically higher on the macro-subskills and micro-subskills than the students who experienced conventional instruction. The statistical analysis of the quantitative data revealed that FC was an efficient means of developing writing subskills for the Iranian EFL learners. Moreover, results indicated a certain amount of pedagogical implications for teachers, learners, curriculum designers, and administrators.

[15] Examined the effects of using flipped instruction on Iranian EFL learners' speaking complexity, accuracy, and fluency (CAF). The findings of Independent Samples T-test and Paired Samples T-test revealed that there was a significant difference between the post-tests of the experimental and the control groups in the favor of the experimental group. Moreover, the findings of One-Sample T-Test showed that Iranian EFL learners had positive attitudes towards using flipped instruction for speaking classes.

The review of the studies revealed that there are a number of research studies which are concerned specifically with the inverted classroom, but it can be argued that there is a need to move beyond the pervious views and find out about its effect on argumentative text comprehension and learners' self-efficacy. The present study was thus set up to fill this gap in the body of studies on the topics of inverted classrooms, argumentative text comprehension, and self-efficacy in the literature.

3. Methodology

3.1. Research Design

This study had a quasi-experimental design [10] as random sampling was not possible in the current study, but other components of experimental research (such as pretest, posttest, treatment, placebo, experimental group, and control group) were all included in the design of the study. The independent variable in this study was the type of instruction (i.e., via inverted classroom or traditional instruction), and the dependent variables included argumentative text comprehension and self-efficacy.

3.2. Participants

The participants of the present study ($N = 51$) were intermediate EFL learners studying English in one of the language institutes in Ahvaz. They were selected from a pool of intermediate learners (118 learners) based on an Oxford Quick Placement Test (OQPT).. Their proficiency level at the institute was intermediate, but for good measure, their proficiency level was also measured by the OQPT. They ranged in age from 18 to 27, and their mother language was Persian. The participants of the study

were selected based on non-random convenience sampling as random selection of the learners in a language school was not possible for the researcher.

3.3. Instruments and Materials

The instruments which were used in the study included an OQPT, pretest and posttest of argumentative text comprehension, and a self-efficacy questionnaire. The texts and worksheets which were used to teach argumentative text comprehension were accessed from a website on argumentative texts (https://en.islcollective.com/english-esl-worksheets/search/argumentative).

3.4. Procedure

As the first step in the collection of the data for the present study, the manager of the language institute was informed so that the researcher could have permission to conduct the study in the institute. Then four of the language classes at the intermediate level were selected and assigned to the two groups of ICG and CG. The OQPT was then administered to the learners in the two groups and those whose scores were not in the 30-47 range were excluded from the study (though they were present in class and they were exposed to the testing instruments and instructional materials). The pretest of argumentative texts and the self-efficacy questionnaire were given to the learners at the beginning of the study. Then, during the 8-week instruction, the learners in the ICG received argumentative texts and instructions on how to comprehend them via WhatsApp and they had to read the materials at home and be ready to do the exercises and supplementary materials in their online classes. During the first and last two sessions, the learners received the pretest, self-efficacy questionnaire, and the posttest. The learners in the CG were exposed to similar materials, though in a different way: they were taught the reading texts in class and they were supposed to do the exercises at home, just like any other traditional class. As it was mentioned above, after the instructional period ended, the posttest of argumentative texts and the self-efficacy questionnaire were given to the learners in the two groups. The data obtained from the pretest, posttest, and questionnaire were coded (in the form of scores) and made ready for analysis by SPSS (version 26).

4. Results

A Sig. value larger than .05 for the Kolmogorov-Smirnov test shows that the assumption of normality has not been violated. In Table 4.1, it could be observed that all the Sig. values for the OQPT and all the pretests and posttests in the two groups were larger than .05, which indicates that all the distributions were normal, and the researcher could proceed with conducting the parametric statistical test required in this study. Moreover, for the ANCOVA analysis presented in the following sections, other assumptions such as the assumptions of linearity and homogeneity of regression slopes were checked and no violation of these assumptions was ensured.

Table 4.1 *Results for the Tests of Normality*

Groups	Tests	Kolmogorov-Smirnov		
		Statistic	*df*	Sig.
ICG	OQPT	.137	27	.200*
	Pretest	.141	27	.176
	Posttest	.144	27	.158
	SE Pretest	.139	27	.193
	SE Posttest	.137	27	.200*
CG	OQPT	.165	24	.091
	Pretest	.132	24	.200*
	Posttest	.154	24	.107
	SE Pretest	.089	24	.200*
	SE Posttest	.143	24	.163

Note: SE stands for self-efficacy

In order to find out whether the ICG and CG learners were at the same level of proficiency at the outset of the study, an independent-samples *t* test was conducted:

Table 4.2
Descriptive Statistics for the OQPT Scores of ICG and CG Learners

Groups	Mean	*N*	Std. Deviation	Std. Error Mean
ICG	38.92	27	4.22	.812
CG	38.41	24	4.60	.940

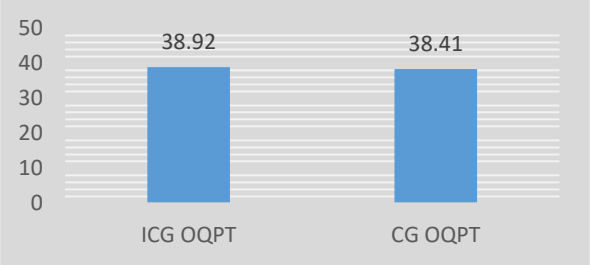
It can be seen in Table 4.2 that there was a slight difference between the OQPT scores of the learners in the ICG (M = 38.92) and CG (M = 38.41). To make sure this difference between the OQPT scores of the two groups was not statistically significant, the researcher had to consult the following *t* test table (Table 4.3):

Table 4.3
Independent-Samples t Test for the OQPT Scores of ICG and CG Learners

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.20	.64	.41	49	.68	.50	1.23	-1.97	2.99
Equal variances not assumed			.41	47.00	.68	.50	1.24	-1.99	3.00

Table 4.3 shows the fact that there was no statistically significant difference between the OQPT scores of the ICG and CG learners because the *p* value under the Sig. (2-tailed) column was found to be larger than the significance level (i.e., $p = .68 > .05$). This means that the learners in the two groups were at a similar level of proficiency before the experiment began. This result is also graphically represented in

Figure 4.1. OQPT mean scores of the ICG and CG learners



The bar chart in Figure 4.1 shows that the OQPT scores of the ICG and CG learners did not differ considerably, and that the two groups of learners were at the same proficiency level when the instruction commenced.

4.2. Testing Research Question Two

To examine whether inverted classroom had significant effects on the self-efficacy beliefs of Iranian intermediate EFL learners, one-way ANCOVA was conducted again to compare the post-instruction self-efficacy scores of the ICG and CG learners:

Table 4.6
Descriptive Statistics for the Self-efficacy Posttest Scores of ICG and CG Learners

Groups	Mean	Std. Deviation	*N*
ICG	24.85	4.34	27
CG	23.25	2.80	24
Total	23.03	3.67	51

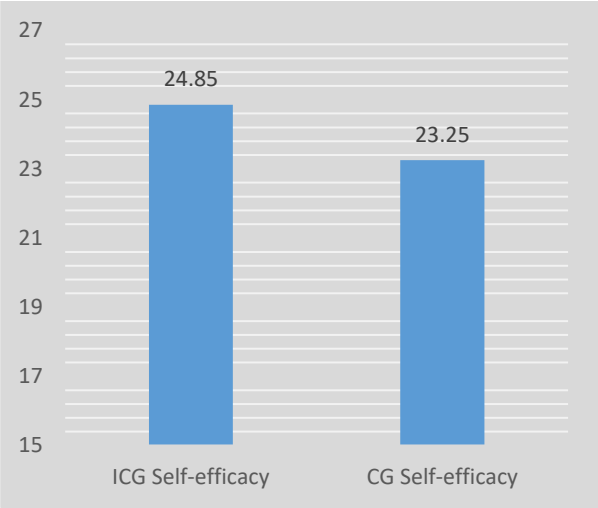
Table 4.6 demonstrates that the self-efficacy posttest mean score of the ICG learners ($M = 24.85$) was larger than the self-efficacy posttest mean score of the CG learners ($M = 23.25$). To figure out whether this difference could reach statistical significance or not, the researcher needed to examine the *p* value under the Sig. (2-tailed) column in Table 4.7:

Table 4.7
One-way ANCOVA for the Self-efficacy Posttest Scores of the ICG and CG Learners

Source	Type III Sum of Squares	*df*	Mean Square	*F*	*Sig.*	Partial Eta Squared
Corrected Model	676.54	2	338.27	580.68	.000	.96
Intercept	10.46	1	10.46	17.96	.000	.27
Pretest	643.94	1	643.94	1105.39	.000	.95
Groups	43.44	1	43.44	74.57	.488	.60
Error	27.96	48	.58			
Total	30321.00	51				
Corrected Total	704.51	50				

In Table 4.5, it could be seen that there is a significant difference between the self-efficacy posttest scores of the learners in the two groups because the *p* value was found to be lower than the .05 level of significance ($p = .000 < .05$). The bar graph below shows the self-efficacy posttest mean scores of the two groups of learners:

Figure 4.3. Self-efficacy posttest mean scores of the ICG and CG learners



As it was mentioned above, the difference between the self-efficacy posttest scores of the ICG and CG learners reached statistical significance.

5. Discussion and Conclusion

5.1. Discussion

The results of the data analysis indicated that there was a statistically significant difference between argumentative text comprehension scores of the learners in the inverted classroom group and the control group (rejecting the first null hypothesis of the study). Moreover, it was revealed that the treatment applied in this study exerted significant impacts on the self-efficacy beliefs of Iranian intermediate EFL learners (rejecting the second null hypothesis of the study).

Regarding the answer to the first research question of the study, the results are in line with previous research findings on flipped instruction and its effects and L2 skills and components. As a case in point, Wei (2019) showed in her study that inverted/flipped classrooms had significant impacts on learning of English by Chinese speakers. She found that the flipped learning tasks prepared students for the active learning in class. She found that the students perceived improved learning in the active learning environment in class. Besides, they perceived enhanced autonomy in learning, improvement in their English listening and speaking proficiency, and opportunities for cultivating higher order thinking skills.

Moreover, the present findings lend further support to those of [15] who examined the effect of flipped classrooms on high school students' reading comprehension, and found that there was a significant difference between the posttests of the experimental and the control groups, leading to the conclusion that flipped instruction had significant effects on high school students' reading comprehension outperformed the control group on the post-test.

The findings of this study are also compatible with those of a study conducted by [22]. They investigated the effect of flipped instruction on the improvements of L2 macro/micro writing subskills of Iranian upper-intermediate students, and found that flipped classroom was an efficient means of developing writing subskills for the Iranian EFL learners

In a more recent study, [15] explored the effect of flipping a classroom on Iranian EFL learners' speaking complexity, accuracy, and fluency. The results of their study revealed that there was a significant difference between the posttests of the experimental and the control groups in the favor of the experimental group, insinuating the effectiveness of inverted classrooms. They also found that their participants had positive attitudes towards using flipped instruction for speaking classes.

The findings of the above-mentioned studies all point to the fact that the current study corroborated much of previous research findings, showing the effectiveness of inverted instruction. Regarding the studies on argumentative text comprehension, [12] used a reading intervention project and examined its effects on argumentative reading and writing. This researcher found the employed treatment significantly effective for writing, but not for reading of argumentative texts. Nonetheless, the results of the present study showed that inverted classrooms can have significant effects on argumentative text comprehension.

Regarding the second research question, self-efficacy is correlated, is influenced by, and influences a large number of variables related to teaching and learning. In the same vein, the results of the current study revealed that the method of instruction (i.e., inverted classroom vs. traditional instruction) exerted significant effects on EFL learners' self-efficacy.

These results could resemble the ones by study, in which 100 participants joined the experiment and the Foreign Language Learning Anxiety Scale and the Self-Efficacy Scale were administered to them. The results showed that both aspects are correlated but gender plays no important role in terms of the anxiety level and self-perception ratings of these junior teacher trainees.

5.2. Conclusions and Implications

The present study was designed to examine the effects of inverted classroom and traditional classes on the comprehension of argumentative texts and on self-efficacy of Iranian intermediate EFL learners. The results of data analysis resulted in two

major conclusions drawn from the study. Firstly, it was revealed that flipped instruction led to significantly better comprehension of argumentative texts. There have been many studies in the literature that confirm the positive effects of flipped instruction on different language skills and subskills.

Secondly, the results of data analysis comparing the effectiveness inverted and traditional classes for self-efficacy revealed that there was a significant difference between these two methods of instruction. In other words, learners in both of the conditions had similar levels of self-efficacy.

In sum, flipped instruction, as a popular method of instruction, was shown to be effective for argumentative text comprehension, as it has been shown to be effective for many other language skills and components. Also, this study shed light on the reality that self-efficacy could also be affected by the methods of instruction employed in this experiment just as it has relationships with a lot of other traits pertinent to language learning and teaching.

The findings of the present study have implications for EFL learners, teachers, and materials developers in the realms of EFL and ESL teaching in particular and education in general. It helps teachers in accomplishing their challenging task of teaching reading comprehension in general, and argumentative text comprehension in particular, more effectively in various EFL contexts such as Iranian language schools where learners have less exposure to language compared to ESL contexts.

EFL learners must know that technology-enhanced methods such as inverted classes are versatile ways for teaching and learning different language skills and areas of language such reading and writing. The contribution of technology should be fully realized by EFL learners to alleviate their L2 learning difficulties. Students will be able to not only develop communication skills but also exchange ideas and gain other benefits from their interactions in their collaborative learning environment of such platforms. Integration of new technologies in education will improve students' classroom engagement and increase their academic achievements.

Moreover, EFL teachers and materials developers are highly recommended to integrate technology-enhanced language learning tools such as inverted classes into their EFL classrooms and materials. It is high time technology received more serious attention from all practitioners and policy-makers in educational and language teaching circles. The Covid-19 pandemic changed the normal teaching environment conditions allowing teaching activities to move from offline to online, from teaching face to face to online and from students' learning from classroom to autonomous learning where students can take control of their learning by learning independently. The use of online opportunities will help generate excitement and enthusiasm towards learning, mainly when it caters to difficult learning situations.

References

[1]. Arslan, A. (2013). Investigation of relationship between sources of self-efficacy beliefs of secondary school students and some variables. Educational Sciences: Theory and Practice, 13(4) 1983-1993. DOI:10.12738/estp.2013.4.1753

[2].Assadi Aidinlou, N., & Taghinezhad Vaskehmahalleh, M. (2017). The Relationship between Iranian EFL Learners' Reading Comprehension, Vocabulary Size and Lexical Coverage of the Text: The Case of Narrative and Argumentative Genres. *The Journal of Language Pedagogy and Practice, 10*(21), 49-71.

[3].Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review, 84*(2), 191. https://doi.org/10.1037/0033-295X.84.2.191

[4].Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. Developmental Psychology, 25(5) 729-735. https://doi.org/10.1037/0012-1649.25.5.729

[5].Bandura, A. (1997). Self-Efficacy: The Exercise of Control. New York, New York: W.H. Freeman and Company.

[6].Bergmann, J., & Sams, A. (2012). Flip Your Classroom: Reach every student in every class every day. Eugene, OR: International Society for Technology in Education.

[7].Brame, C.J. (2012). Flipping the Classroom. Retrieved from http://cft.vanderbilt.edu/teaching-guides/teaching-activities/flipping-the-classroom.

[8].Bull, G., Ferster, B., & Kjellstrom, W. (2012, August). Inventing the Flipped Classroom. Learning & Leading with Technology 40(1).

Retrieved from: http://www.learningandleading-digital.com/learningandleading/ DOI:10.1016/B978-0-12-814702-3.00009-3

[9].Chala, Pedro & Chapetón, Claudia. (2012). EFL argumentative essay writing as a situated-social practice: review of concepts. *Folios,1*, 23-36. DOI:10.17227/01234870.36folios23.36

[10].Farhady, H. (1995). *Research methods in applied linguistics*. Tehran: Payame-Noor University Press. ISBN, 9644552695, 9789644552694

[11].Garay, L.W.P. & Soto, D.M. (2021). From traditional learning to Flipped learning as a continuity of the educational process in the context of COVID-19. *Rev. Mendive*, 19, 214–226. http://mendive.upr.edu.cu/index.php/MendiveUPR/article/view/2191

[12].Grogan, M.S. (2014). *Reading, argumentation, and writing: Collaboration and development of reading comprehension intervention for struggling adolescents*. Unpublished doctoral dissertation. University of Arkansas, US. DOI:10.4324/9781003156857-23

[13].Haria, P.D. (2010). *The effects of teaching genre-specific reading cconmprehension strategy on struggling fifth grade students' ability to summarize and analyze argumentative texts.* Unpublished doctoral dissertation. University of Delware.

[14].Hashemifardnia, A., Namaziandost, E., Shafiee, S. (2018). The Effect of Implementing Flipped Classrooms on Iranian Junior High School Students' Reading Comprehension, *Theory and Practice in Language Studies, 8* (6), 665-673. DOI:10.17507/tpls.0806.17

[15].Hashemifardnia, A., Shafiee, S., Rahimi Esfahani, F., & Sepehri, M. (2021). Effects of flipped instruction on Iranian intermediate EFL learners' speaking complexity, accuracy, and fluency. *Cogent Education, 8*(1), 1-19. DOI:10.1080/2331186X.2021.1987375

[16].Iordanou, K., & Constantinou, C. P. (2014). Developing pre-service teachers' evidence-based argumentation skills on socio-scientific issues. Learning & Instruction, 34, 42–57. doi:10.1016/j.learninstruc.2014.07.004.

[17].Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' reasoning. Psychological Science, 22, 545–552. doi:10.1177/0956797611402512.

[18].Louis, R. A., & Mistele, J. M. (2012). The differences in scores and self-efficacy by student gender in mathematics and science. *International Journal of Science and Mathematics Education, 10*, 1163-1190. DOI:10.1007/s10763-011-9325-9

[19].Midgette, E., Haria, P., & MacArthur, C. (2008). The effects of content and audience awareness goals for revision on the persuasive essays of fifth- and eighth-grade students. Reading & Writing: An Interdisciplinary Journal, 21, 131–151. doi:10.10007/s11145.007-9067-9.

[20].Nussbaum, E. M. (2008). Using argumentation Vee diagrams (AVDs) for promoting argument-counterargument integration in reflective writing. Journal of Educational Psychology, 100, 549–565. doi:10.1037/0022-0663.100.3.549.

[21].Peters, M. L., (2013). Examining the relationships among classroom climate, self-efficacy, and achievement in undergraduate mathematics: a multi-level analysis. *International Journal of Science and Mathematics Education, 11*(2), 459-480. DOI:10.1007/s10763-012-9347-y

[22].Shooli, E., Rahimi Esfahani, F., & Sepehri, M. (2021). Impacts of Flipped Classroom on Micro/Macro Writing Subskills in Iranian EFL Context. *Journal of Modern Research in English Language Studies*, *8*(4), 85-109. DOI:10.30479/jmrels.2020.13367.1649

[23].Velayutham, S., Aldridge, J. M., & Fraser, B. (2012). Gender differences in student motivation and self-regulation in science learning: a multi-group structural equation modeling analysis. *International Journal of Science and Mathematics Education, 10*, 1347-1368. DOI:10.1007/s10763-012-9339-y

[24].ZhengWei, X. (2019). *The effects of the flipped classroom model on students' learning*

in a college English class in Shanghai, China. Unpublished doctoral dissertation, University of the Pacific, Stockton, California, US. DOI:10.12973/iji.2018.11226a

Journal of Optimization of Soft Computing (JOSC)
Vol. 3, Issue 1, pp: (41-52), Summer-2025
Journal homepage: https://sanad.iau.ir/journal/josc

Paper Type: (Research paper)

Implementation and comparison of active queue management algorithms in traditional and SDN networks

Khoshnam Salimi Beni[1], Mohammadreza Soltanaghaei [1*] and Rasool Sadeghi[2]

[1]*Institute of Artificial Intelligence and Social and Advanced Technologies, Isf.C., Islamic Azad University, Isfahan, Iran.*

[2]*Department of Electrical Engineering, Dolatabad Branch, Islamic Azad University, Isfahan, Iran*

Article Info

Article History:
Received: 2025/07/28
Revised: 2025/08/30
Accepted: 2025/09/13

DOI:

Keywords:
Network, SDN, resource allocation, queuing mechanisms

**Corresponding Author's Email Address: soltan@iau.ac.ir*

Abstract

In the past decade, networks have experienced significant improvements in scale and data transfer rates, and network traffic rates will soon increase dramatically. Network management and traffic control play key roles in real-time data transmission (such as video conferencing, high-bandwidth streams, video calls, etc.) and data transmission in the Internet of Things (IoT). Although technologies such as SSD storage and virtualization are very effective in meeting network traffic needs. Future networks will require centralized management, easy upgradeability, application optimization, efficient resource allocation, and dynamic routing. To meet these requirements, the benefits of software-defined networking (SDN) must be used. By separating the control part from the data part, SDN will lead to scalability, flexibility and centralized management of the network. With excessive demands on limited network resources, it is inevitable to create long queues of information packets in intermediate routers, and the use of active queue management (AQM) algorithms of TCP/IP network in order to make more use of available bandwidth and reduce Transmission delay is necessary. In this article, we examine some of the most important active queue management algorithms including PFIFO_fast, ARED, CoDel, FQ-CoDel and PIE in traditional and SDN networks. The results of the simulation show that the use of AQM algorithms in the SDN network reduces the average delay and packet loss rate and increases the network efficiency.

1. Introduction

Configuring and implementing different scenarios in traditional networks not only has problems in the field of managing these networks and wasting time, but also has its own errors and problems in the field of extensibility. Therefore, it is necessary to change the network architecture and use other structures such as software-defined networks (SDN)[1]. SDN is one of the new network architectures that separates the data part from the control part to improve the use of network resources, reduce operational costs and provide network innovation and evolution. However, the main challenge in SDN is to provide high quality services and resource allocation in these programmable networks. Proper allocation of resources improves network performance and reduces overall network costs. In this regard, various techniques are used to allocate resources in SDN in order to increase network efficiency, one of which is the optimal allocation of resources to each task in the network[2]. Various parameters are defined in these techniques for resource

allocation in SDN. Overview the controller in SDN can easily collect data from available network resources and basically allocate resources to different services through the OpenFlow protocol[3]. Various techniques and methods have been used to improve network resource allocation in SDN. SDN can be used in various technologies such as virtual networks (VN), data centers (DC), cloud environment, 5G and wireless networks, and can also be used in combination to improve network performance[4]. In SDN networks, two important resources, the bandwidth capacity of the switch to the controller and the capacity of the flow table, which mutually influence each other, must be carefully analyzed[5]. For example, the acceleration of the incoming flow to a switch can greatly intensify the message exchange between the switch and the controller, which causes more bandwidth consumption[6]. In order to optimally transmit data in the network, some challenges such as congestion, delay and packet loss must be considered. Some papers focus on the channel congestion problem and suggest the use of queue management algorithms. These algorithms are categorized into Active Queue Management (AQM) or Passive Queue Management (PQM) such as Drop-Tail[7] depending on the congestion control mechanism. Some of the famous AQM algorithms are RED[8], ARED[8], PIE[9], CoDel[10], FQ-CoDel[11] and PFIFO_fast[12], which are used in various papers to address these challenges. The queue management system controls the size of the communication channel queue by enabling or disabling queue management. One passive queue management algorithm is Drop-Tail, which drops packets when the queue is full, but in active systems, such as Random Early Detection (RED), network packets are dropped before the queue becomes saturated. In addition, other queue management algorithms such as CoDel have proposed Adaptive Random Early Detection (ARED), Packet-First-In-First-Out (PFIFO), to solve the congestion problem.

SDN is one of the new network architectures in which the information control part is separated from the data part. In traditional networks, routers and network switches, data transfer and information control operations are performed together. In SDN architecture, the control part is separated from the switch and router hardware and is performed by software at a higher layer. Therefore, the speed, flexibility, scalability, availability and reliability of the network are improved. Researchers can centralize and integrate network management by creating programming interfaces. Another advantage of using SDN is network and hardware reconfiguration without the need for the involvement of hardware manufacturers. In traditional networks, they must use the technology and architecture provided by hardware manufacturers, and network development is not possible, but in SDN networks, according to the needs, the network can be localized[13].

OpenFlow is a key protocol in SDN that enables centralized control and programmability of network devices. In an SDN architecture, OpenFlow decouples the control plane from the data plane, allowing a centralized controller to manage and direct network traffic dynamically. This separation facilitates the implementation of network policies and configurations through a logically centralized controller, leading to more efficient network management and flexibility. OpenFlow operates by defining a set of communication messages between the SDN controller and network devices, such as switches and routers, allowing the controller to instruct these devices on how to forward, modify, or block packets based on the network's current state and requirements. The protocol enhances network agility, scalability, and programmability, making it a fundamental component in the evolution of modern networking architectures[14].

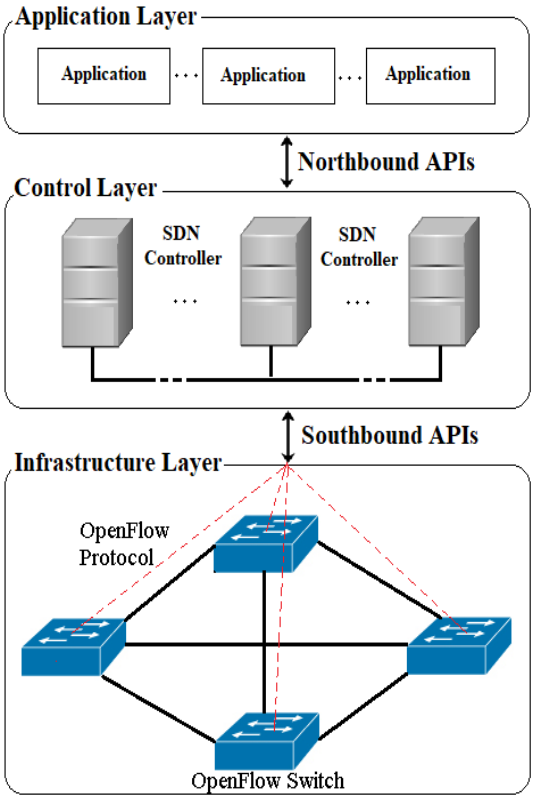


Figure (1). SDN Architecture

In this article, we examine the use of queue management algorithms in congestion control. For this purpose, the most important queue management algorithms including PFIFO_fast, ARED, CoDel, FQ-CoDel and PIE in traditional networks and Software-Defined Networks have been investigated separately. One of the most important factors affecting network efficiency is the average delay and packet loss rate that have been evaluated. In the second part, software-defined networks and basic concepts will be examined, and the studies and works done on queue management algorithms will be examined. In the third part, some of the most important queue management algorithms have been implemented in the SDN network using the NS3 simulator. In the fourth part, the results of the implementation of queue management algorithms in traditional and SDN networks will be analyzed and evaluated, and finally, the fifth part will include the conclusion.

2. Related works

In this section, we examine the research that has been done on queue management algorithms. Optimal allocation of resources in SDN plays a key role in improving performance and is very challenging. To achieve this goal, many solutions have been proposed in the existing research. Some resources, such as the capacity of the communication link between the switch and the controller, the rate of messages sent and received from the switch to the controller, and vice versa, have limitations. Therefore, bandwidth management is very important for resource sharing in SDN [15]. Various techniques and methods have been proposed for optimal bandwidth allocation in SDN[16], which we have discussed in this section.

Some research investigates resource consumption in SDN controller and switches using queuing mechanisms. Packet loss may occur simultaneously while transmitting traffic with the same queue priority. J. Hao and his colleagues in [17] presented a flow-level bandwidth provisioning algorithm (FBP) to deal with the switch scheduling problem using a fair queuing algorithm. This algorithm schedules multiple fair queues on OpenFlow switches to separate flows and allocate bandwidth between flows on a shared link. However, this paper has disadvantages such as time complexity, large hardware, unreliability of service guarantee, and inefficient processing of variable-length packets. Some of these problems were solved by traffic classification and queue prioritization, which was studied by H. Cui and his colleagues in the article [18]. In this article, the queuing mechanism is used to classify traffic, collect information about the network status, and determine the optimal route for allocating network resources to different services. The purpose of the proposed plan is to guarantee QoS for different services and balance the load on the communication link and prevent time wastage and congestion in the network. This mechanism divides the core network's proposed queue into smaller virtual subnets and allocates resources to each service. Multiple queues with different priorities can be configured on a switch port. However, when a new flow arrives on a port, other lower-priority flows in the queues may experience delays and jitter. However, dynamic queue mapping that can improve resource allocation in the network is not studied in this paper.

Connections that have a common destination in the network share their communication links to use these links. However, without providing a solution to protect and isolate services, connections with high data transfer rates will send more traffic to the core network than others. Therefore, packet loss occurs more often in low-rate connections. To address this problem J. Guo. and colleagues in [19] proposed an application-layer fair bandwidth allocation (FBA) protocol called Falloc to distribute network resources at the virtual machine (VM) level in IaaS data centers. Falloc assigns a base bandwidth and a weight to each virtual machine. Virtual machines with less required bandwidth than the original bandwidth share the remaining bandwidth among all machines in proportion to the weight. Therefore, fairness can be ensured by balancing the bandwidth allocated to VMs and the shared bandwidth between virtual machines. However, due to competition between other VMs, Falloc cannot guarantee bandwidth for all VMs.

Another fair bandwidth allocation scheme in the application layer named UFalloc is proposed in the article [20] to achieve fairness in bandwidth allocation between virtual machines using the max-min algorithm [21]. UFalloc not only limits the bandwidth of each flow on OpenFlow switches to guarantee performance, but also shares bandwidth resources across dense switches and links. UFalloc uses a factor called relaxation-fairness to maintain a certain degree of fairness for bandwidth allocation. As shown in the results of this paper, UFalloc can reduce the application utilization switching degree by 5.9% to 10.9% compared to the traditional TCP rate control mechanism and max-min fair allocation

algorithm. However, the computational overhead is one of the drawbacks of this method.

Heuristic methods can be used to solve problems that are not guaranteed to be optimal, logical, or complete, but can be sufficient to arrive at an approximate solution. In cases where it is impractical or impossible to identify an optimal solution, heuristic strategies can be used to speed up the process of finding a satisfactory solution. Using heuristic algorithms, KT. Bagci and his colleagues in [22] proposed a heuristic model based on a packet-based shortest path (GCSP) for fair allocation of resources among a group of requests with the same service level. This method is close to the optimal solution and uses a divide-and-conquer strategy in networks that are divided into smaller subnets. So many service requests can be handled by processing groups of service requests in minimum time. Another algorithm based on near-optimal heuristic methods was presented by W. Aljoby and his colleagues in the paper [23] to share bandwidth between several active applications in SDN. This algorithm is obtained from stream processing in programs and formulation of bandwidth allocation between streams belonging to these streams. Although this method can be used in a variety of platforms, including parallel and pipelined network flows, however, balancing bandwidth between multiple broadcast applications with different performance and bandwidth-optimizing communication overheads is a challenge. This article is important.

A. Marin and colleagues in [24] analyzed AQM techniques for bandwidth sharing in TCP and UDP traffic and analyzed the performance of these techniques in different scenarios using mean field methods. Because TCP and UDP streams exist on the same channel, and because TCP's congestion control mechanism uses more resources, packets for UDP-based applications must be queued in buffers. Therefore, UDP-based applications will experience unfavorable latency. To deal with this challenge, AQM mechanisms are used with the help of congestion control mechanism to avoid congestion in bottleneck links and make optimal use of available bandwidth. In this study, their proposed method is compared with RED. Another optimal model of Internet congestion control using AQM is proposed by C. Han and his colleagues in [25]. In this paper, a state monitoring system is used to collect system state information. This observer can estimate the window size in the real network and estimate the queue length obtained by measuring the output of the system. Another efficient AQM algorithm is proposed by L. Chrost and A. J. T. S. Chydzinski in the paper [26]. This algorithm keeps the queue size short and stable to reduce packet loss and high throughput. The proposed algorithm reduces the energy consumption in routers by reducing the complexity of calculations. Some AQM schemes such as RED calculate the probability of dropping packets using the average queue size. In heavy traffic, increasing the frequency of crossing the maximum threshold value will lead to frequent dropping of packets. To address this problem, S. Patel and S. J. T. S. Bhatnagar in [27] proposed an adaptive queue management mechanism using information obtained from the average queue size and the rate of change of the queue size. Therefore, using the rate of change in queue size as an additional parameter leads to an increase in system efficiency in terms of average queue size, throughput and queue delay compared to other AQM algorithms such as RED. The problem of bandwidth consumption in wireless networks of TCP streams is investigated by K. O. Okokpujie and his colleagues in the article [28]. Using AQM algorithms in this paper, two adaptive TCP strategies are proposed for queue management by implementing feedback control techniques in AQM. The comparison result states that these models have better performance than PI and RED proportional integral controllers.

In large networks, simultaneous communication between network equipment leads to increased channel congestion. One of the important challenges in these networks is to maintain fairness between TCP-based packets that are sent on a congested channel. Some papers suggested using Drop-Tail for fair channel allocation to each stream. But dropping the packet in the Drop-Tail method reduces the flow throughput. To overcome this challenge, M. M. Hamdi and his colleagues in [29] proposed the use of AQM to solve the problem of packet loss and delay. In this article, Drop-Tail is compared with different active queue management methods such as PFIFO, RED, ARED and CoDel. The comparison results show that CoDel and RED have higher throughput with minimum delay. Also, ARED performs better than RED. In addition, the use of RED algorithms reduces the congestion problem, but it does not guarantee network efficiency and QoS due to the predefined and fixed parameters in RED algorithms. To address this problem, A. F. AL-Allaf and A. A Jabbar in [30] proposed an improved adaptive RED algorithm using reconfigurable policy for multimedia traffic. In this approach, the maximum drop probability (maxp) in RED is replaced by another parameter obtained from the network traffic load. According

to this policy, the average queue size and queue delay time are reduced without increasing the packet drop rate or reducing the link utilization.
In order to control congestion and improve network performance, D. Kumhar et al in [31] proposed an AQM method based on the RED algorithm named QRED random early detection (QRED). Compared with RED, the simulation results showed that QRED performs better in terms of end-to-end delay, packet loss, packet delivery conditions, and jitter. Although AQM is considered as a solution for congestion control, the selection of accurate parameters for AQM methods is an important problem in inter-domain structures due to the dynamics of IP networks. To address this challenge, C. A. Gomez and his colleagues in [32] proposed an architecture called (FIAQM) to adjust AQM parameters dynamically in a multi-domain network. In this method, artificial neural network is used to check and predict congestion. Therefore, the performance of inter-domain communication is improved by reducing link congestion.

3. Appling of AQM algorithms in SDN architecture

In this section, we apply some of the most important queue management algorithms including PFIFO_fast, ARED, CoDel, FQ-CoDel and PIE first in the traditional network and then in the SDN network. In the traditional network, according to “figure 2”, node1 is connected to node2 through a switch, and also in “figure 3”, node1 is connected to node2 through a switch equipped with SDN technology. The required parameters in the simulation are determined according to “table 1”.

Table 1. Parameters required in the simulation

Parameters	Value
Delay	0.1 ms
Datarate	20-100 mb/s
Queue capacity	1000 packet
Similation time	60 s
Packet size	1024 bit

Different metrics exist in the network to assess and monitor the performance and reliability of computer networks. These metrics provide a significant understanding of various dimensions of network performance, encompassing elements such as data transfer speed, bandwidth utilization, latency, jitter, packet loss, and other fundamental indicators of network performance. Through meticulous monitoring and analysis of these network metrics, network administrators can identify existing constraints, diagnose issues, and improve network configurations to enhance network performance, reduce downtime, and ensure a better end-user experience. In this article, three metrics, namely average delay, average jitter, and packet loss ratio, have been used for network evaluation[33].

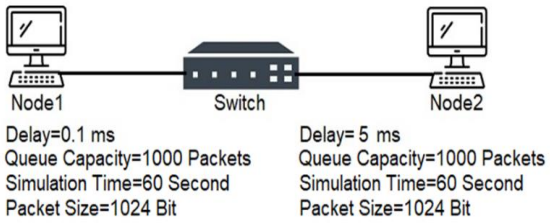


Figure 2. Scenario in Traditional network

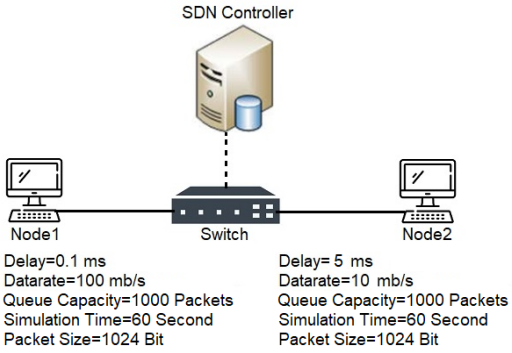


Figure 3. Scenario in SDN Architecture

The structured methodology for evaluating and comparing Active Queue Management (AQM) algorithms in both traditional and Software-Defined Networking (SDN) architectures is shown in Figure 4. The process begins by defining key simulation parameters such as delay, data rate, queue capacity, and packet size, which establishes a consistent baseline for testing. The critical branching point involves selecting either a traditional network or an SDN-based environment, allowing for a direct comparison of how the centralized control and programmability of SDN impact network performance. Each network type then undergoes systematic testing with five prominent AQM algorithms (PFIFO_fast, ARED, CoDel, FQ-CoDel, and PIE) during a 60-second simulation, after which three crucial performance metrics—average delay, average jitter, and packet loss ratio—are collected and analyzed. This comprehensive approach enables researchers to quantitatively assess the effectiveness of each algorithm in both conventional and modern SDN environments, ultimately providing insights into how SDN's centralized control can optimize traffic management and improve overall network quality of service.

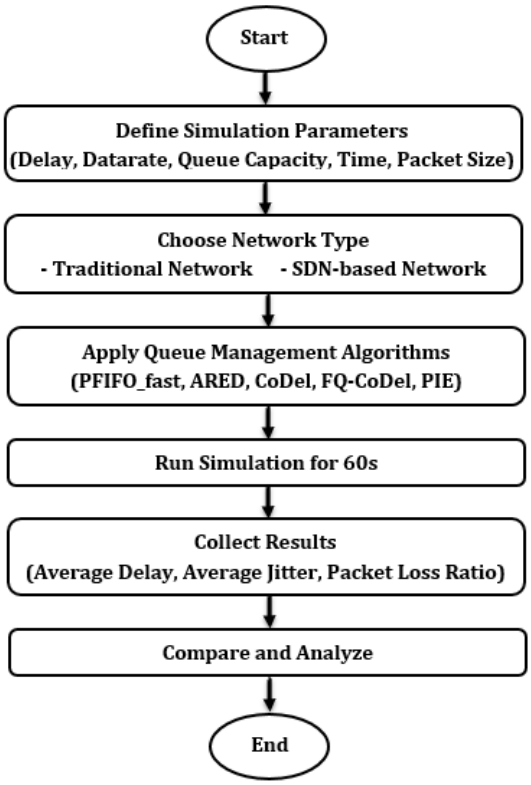


Figure 4. The structured methodology of the proposed algorithm

4. Evaluation and comparison of results

In this section, we implement some of the most important queue management algorithms including PFIFO_fast, ARED, CoDel, FQ-CoDel and PIE first in the traditional network and then in the SDN network. For simulation, the NS3 simulator[34], which is one of the most powerful network simulators, has been used.

A. Evaluation Based on Mean Delay

The metric known as mean delay, or commonly referred to as "average delay," serves as a vital indicator of network performance. It measures the average time it takes for data packets or information to traverse from a source to a destination within the network, typically expressed in milliseconds. Formula (2) is employed to calculate the mean delay, providing a quantitative insight into the efficiency and responsiveness of the network's data transmission.

$$\overline{delay} = \frac{delaySum}{rxPackets} \quad (2)$$

In this formula, "delaySum" represents the cumulative sum of all end-to-end delays incurred by every packet received within a specific flow. Meanwhile, "rxPackets" denotes the total count of packets received for that particular flow. This information is crucial for assessing the overall performance and efficiency of the network, offering insights into the latency experienced by the transmitted data and the volume of packets successfully received[35].

In Figure (4), the diagram depicts the average delay across varying data rates within the range of 20 to 100 Mbps in a traditional network. The visualization provides a comparative analysis of how different AQM algorithms, including PFIFO_fast, ARED, CoDel, FQ-CoDel, and PIE, influence the average delay for data transmissions. This representation is instrumental in understanding the performance of these AQM algorithms across a spectrum of data rates in a traditional network environment. This visualization provides a comparative analysis of the average delay experienced by data packets as they traverse the network under different AQM strategies. The graph serves as a valuable tool for understanding how these AQM algorithms impact the latency of data transmission in a traditional network setting.

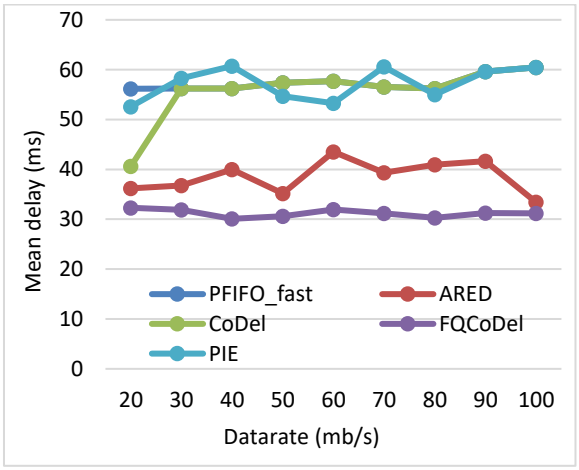


Figure 5. Average delay diagram in traditional network

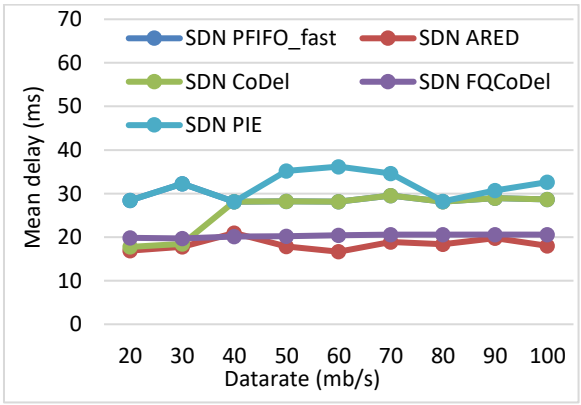
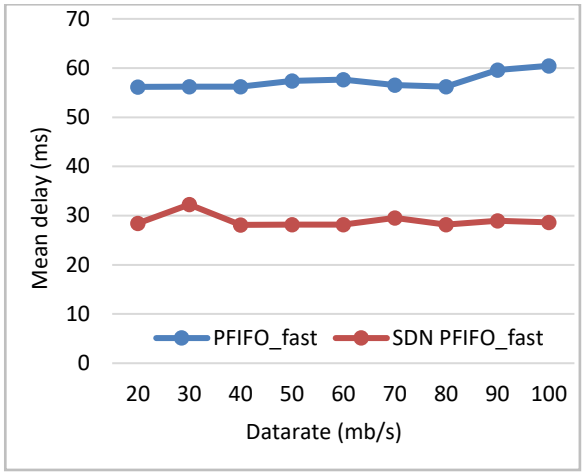


Figure 6. Average delay diagram in SDN network

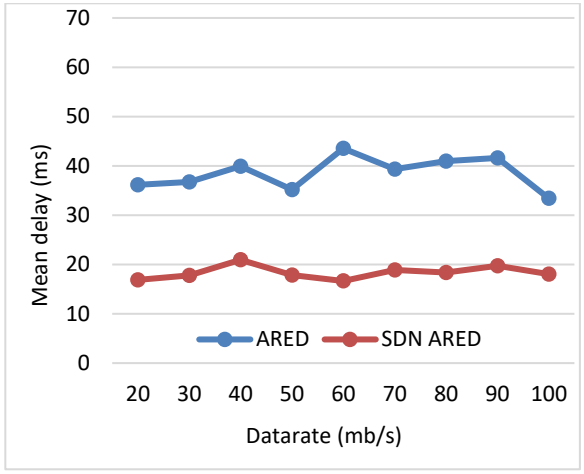
In Figure (5), the diagram illustrates the average delay across a range of data rates from 20 to 100 Mbps within an SDN architecture. This

visualization offers a comparative analysis of the impact of various AQM algorithms, including PFIFO_fast, ARED, CoDel, FQ-CoDel, and PIE, on the average delay experienced by data packets during their traversal in the network. Providing insights into how these AQM algorithms influence latency, this graph serves as a valuable tool for understanding the performance of data transmissions within the context of an SDN environment.

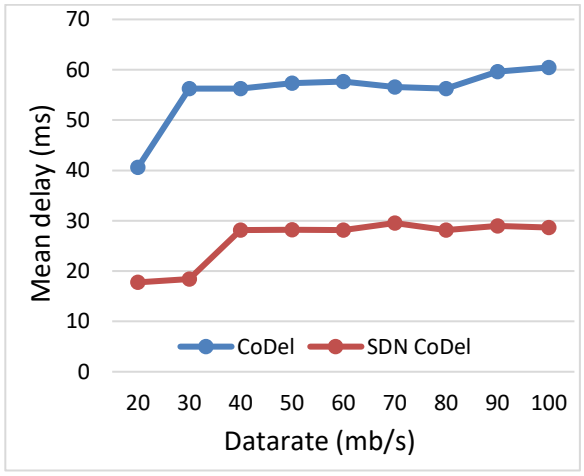
In SDN architecture, implemented AQM algorithms excel at early congestion detection through the monitoring of queue lengths or patterns of packet drops. Upon congestion detection, proactive measures are taken, selectively dropping or marking packets before the queue reaches excessive congestion levels. This preemptive action prevents the network from reaching a critical state, effectively reducing the mean delay for packets. Furthermore, these algorithms support traffic prioritization, offering preferential treatment to specific traffic types (e.g., voice-over IP) or ensuring superior service for time-sensitive applications. This prioritization significantly contributes to reducing the mean delay for critical data, enhancing overall network responsiveness. SDN-based AQM algorithms prove to be a versatile solution, extending beyond mere maintenance of low and consistent delay to actively minimizing overall delay. This dual focus on delay management positions SDN-based AQM algorithms as valuable tools for optimizing network performance. The selection of an appropriate AQM policy within the SDN framework, tailored to the unique network requirements, holds the potential for significant improvements in mean delay. The adaptability to network specifics, coupled with the overarching goal of delay reduction, underscores the potential of SDN-based AQM algorithms to enhance the efficiency of network traffic management. The dynamic adjustment of parameters based on network conditions, facilitated by their implementation in an SDN environment and the centralized aggregation of required information in the SDN controller, allows SDN-based AQM algorithms to fine-tune their settings in response to changing network conditions. The comparative results indicate a significant reduction in mean delay for each of these algorithms. Figure (6) illustrates the comparison chart of each of these algorithms in both traditional and SDN networks.



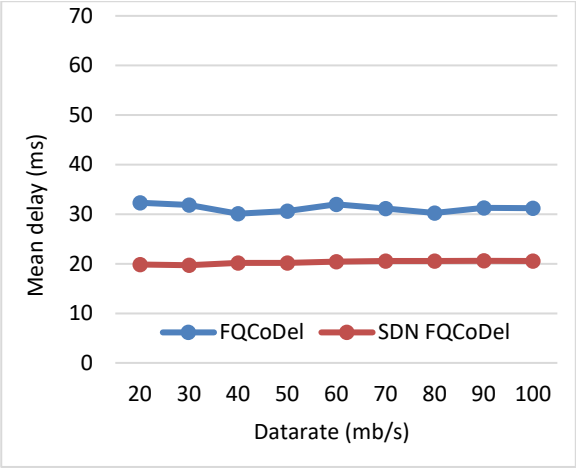
a) Average Mean delay in PFIFO-fast and SDN PFIFO-fast



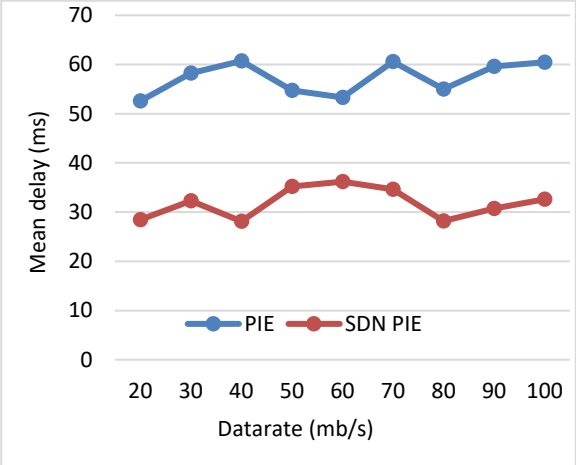
b) Average Mean delay in ARED and SDN ARED



c) Average Mean delay in CoDel and SDN CoDel



d) Average Mean delay in FQCoDel and SDN FQCoDel



e) Average Mean delay in PIE and SDN PIE

Figure 7. Average delay diagram in traditional network and SDN architecture for a) PFIFO-fast b) ARED c) CoDel d) FQ-CoDel e) PIE

B. Evaluation Based on Packet loss Ratio

Packet loss ratio, commonly known as "packet loss," serves as a networking metric gauging the proportion of data packets that do not successfully reach their intended destination or are discarded during transmission across a network. There are various factors contributing to packet loss, such as network congestion, hardware malfunctions, software glitches, or the deliberate discarding of packets by network devices. Mitigating packet loss stands as a primary objective in the realm of network management and optimization, necessitating actions like the enhancement of network infrastructure, implementation of Quality of Service (QoS) mechanisms, or the deployment of error correction techniques to bolster the dependability of data transmission. Formula (4) is provided to articulate the calculation of the Packet loss ratio.

$$q = \frac{lostPackets}{rxPackets + lostPackets} \quad (4)$$

Where the lostPackets variable represents the total count of packets assumed to be lost, meaning those that were transmitted but have not been reported as received or forwarded within an extended timeframe. By default, packets not acknowledged within a duration exceeding 10 seconds are considered lost, though this threshold is adjustable during runtime. On the other hand, the rxPackets parameter denotes the overall number of received packets for the specific flow[35].

In Figure (7), the graphic illustrates the packet loss ratio across diverse data rates ranging from 20 to 100 Mbps within a traditional network. The diagram offers a comparative examination of the impact of various Active Queue Management (AQM) algorithms, namely PFIFO_fast, ARED, CoDel, FQ-CoDel, and PIE, on the Packet Loss Ratio during data transmissions. This visualization proves crucial in comprehending the efficacy of these AQM algorithms across a spectrum of data rates in a conventional network setting. It facilitates a side-by-side comparison of the Packet Loss Ratio encountered by data packets as they navigate the network under distinct AQM strategies. The graph acts as a valuable instrument for gaining insights into how these AQM algorithms influence the latency of data transmission within a traditional network environment.

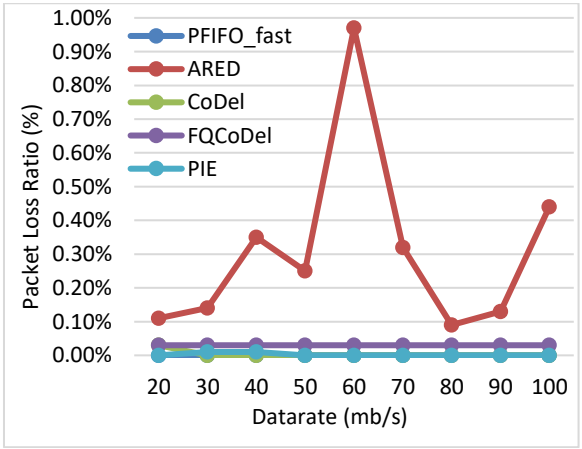


Figure 8. Packet loss ratio diagram in traditional network

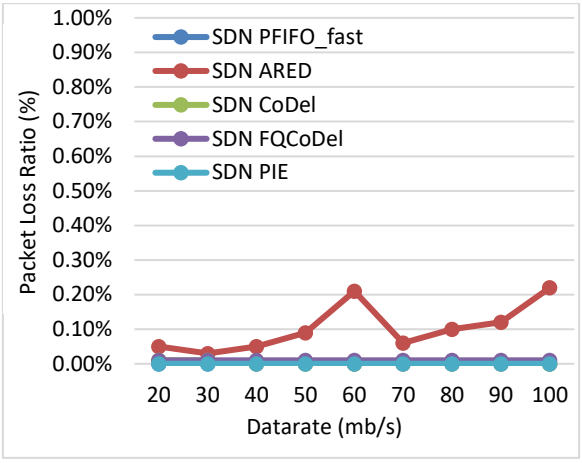
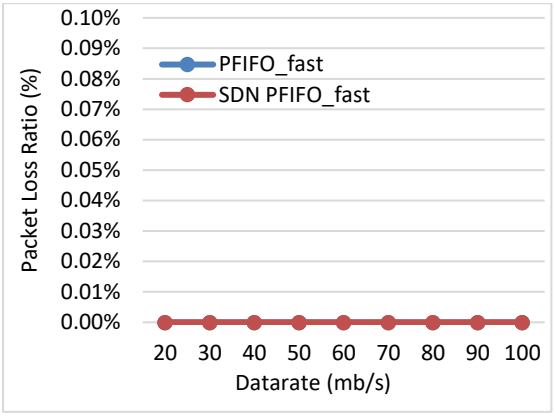


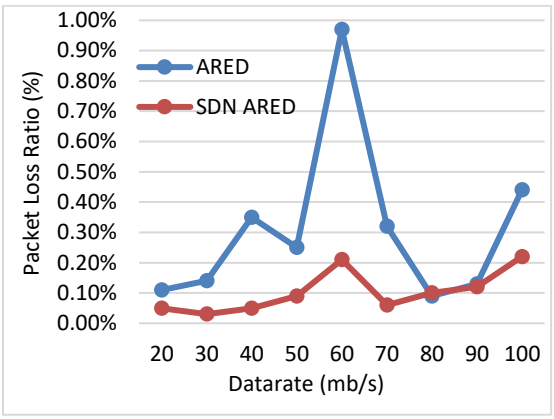
Figure 9. Packet loss ratio diagram in SDN network

In Figure (8), the diagram showcases the Packet Loss Ratio across a spectrum of data rates, spanning from 20 to 100 Mbps within an SDN architecture. This visual representation facilitates a comparative assessment of the influence of different Active Queue Management (AQM) algorithms, such as PFIFO_fast, ARED, CoDel, FQ-CoDel, and PIE, on the Packet Loss Ratio encountered by data packets as they traverse the network. Offering insights into the impact of these AQM algorithms on latency, this graph acts as a valuable resource for comprehending the performance of data transmissions within the framework of an SDN environment.

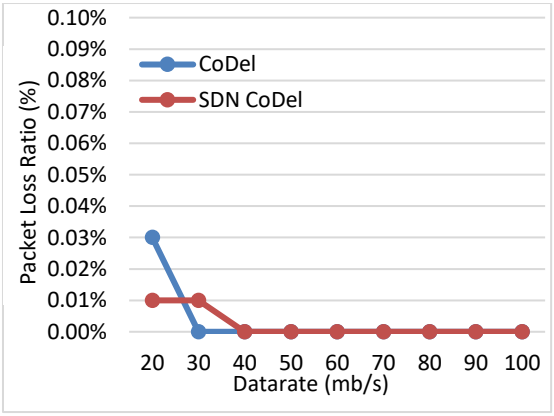
As outlined in Section A, SDN-based algorithms enable traffic differentiation and prioritization, ensuring that critical or time-sensitive data, such as voice or video packets, receives preferential treatment. This prioritization minimizes the risk of losing crucial packets, leading to an overall reduction in packet loss. These algorithms adeptly manage the network queue, exerting a significant influence on mitigating packet loss. By achieving low and consistent queue delays while proactively avoiding congestion-related losses, this algorithm emerges as a pivotal factor in the reduction of packet loss. The comparative findings reveal a notable decrease in the packet loss ratio across all of these algorithms. Figure (9) visually represents the comparative analysis of each algorithm in both traditional and SDN networks.



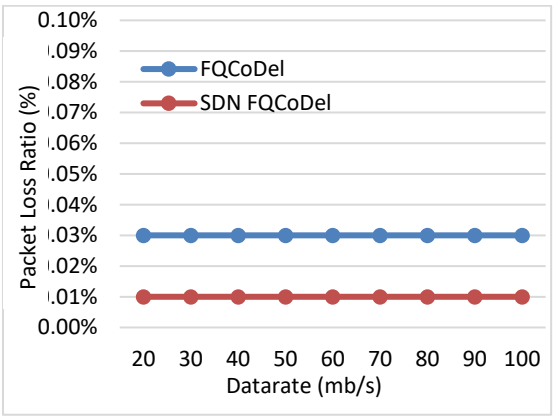
a) Packet loss ratio in PFIFO-fast and SDN PFIFO-fast



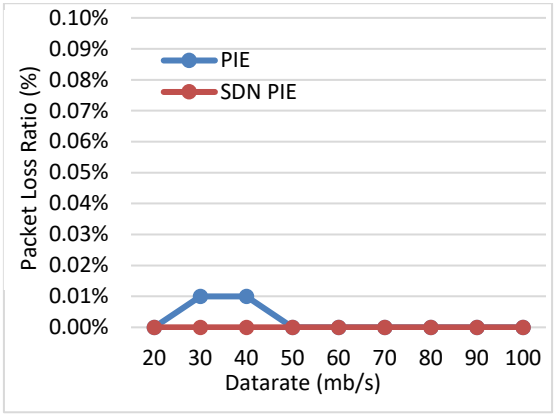
b) Packet loss ratio in ARED and SDN ARED



c) Packet loss ratio in CoDel and SDN CoDel



d) Packet loss ratio in FQ- CoDel and SDN FQ- CoDel



e) Packet loss ratio in PIE and SDN PIE

Figure 10. Packet loss ratio diagram in traditional network and SDN architecture for a) PFIFO-fast b) ARED c) CoDel d) FQ-CoDel e) PIE

According to the scenario in "figure 2" and "figure 3", the bandwidth and transmission delay between Node1 and the switch are considered to be 100 mb/s and 0.1 ms. Also, the bandwidth and transmission delay between Node2 and the switch are 10 mb/s and 5 ms for traditional and SDN networks. To generate congestion and get better simulation results, the bandwidth between Node2 and the switch varies within the range of 20 to 100 Mbps.The simulation time is 60 seconds and the size of each packet is 1024 bits. Based on the simulation results in the graphs of "figure 4" to "figure 9", in traditional networks, the FQ-CoDel algorithm has a lower average delay than other algorithms, and in SDN networks, the ARED algorithm has a lower average delay than other algorithms. Also, in traditional networks, the packet loss rate in PIE, CoDel and PFIFO_fast algorithms has the lowest value compared to other algorithms, and for SDN networks, the packet loss rate in PIE and PFIFO_fast algorithms is zero and it has the lowest value compared to other algorithms. The comparison results of each of the aforementioned algorithms in traditional and SDN networks indicate that the average delay and packet loss rate in SDN networks have been significantly reduced.

5. Conclusion

In this article, using SDN technology, we implemented active queue management (AQM) algorithms of TCP/IP network in SDN environment. This caused more utilization of the available bandwidth and reduced transmission delay. In this article, we examined some of the most important active queue management algorithms including PFIFO_fast, ARED, CoDel, FQ-CoDel and PIE in traditional and SDN networks. The results of the simulation show that the use of AQM algorithms in the SDN network reduces the average delay and packet loss ratio and increases the network efficiency.

References

[1] Al-Darrab, A., I. Al-Darrab, and A. Rushdi, Software-Defined Networking load distribution technique for an internet service provider. Journal of Network and Computer Applications, 2020. 155: p. 102547, https://doi.org/10.1016/j.jnca.2020.102547.

[2] Al-Tarawneh, L. and O.A. Saraereh, An optimal method for resource allocation in SDN optical networks. Optical Fiber Technology, 2022. 74: p. 103120, https://doi.org/10.1016/j.yofte.2022.103120.

[3] Prabha, C., A. Goel, and J. Singh. A survey on sdn controller evolution: A brief review. in 2022 7th International Conference on Communication and Electronics Systems (ICCES). 2022. IEEE, https://doi.org/10.1109/ICCES54183.2022.9835810.

[4] Salami, Y., et al., A Novel Approach for Intrusion Detection System in IoT Using Correlation-Based Hybrid Feature Selection and Harris Hawk Optimization Algorithm. Journal of Optimization in Soft Computing, 2024: p. 54 - 63, https://doi.org/10.82553/josc.2024.140308141189548.

[5] Banitalebi Dehkordi , A., An optimal approach to detect anomalies in intrusion detection systems. Journal of Optimization in Soft Computing, 2024: p. 54 - 63, https://doi.org/10.82553/josc.2024.140212131104279.

[6] Aldabbas, H., Efficient bandwidth allocation in SDN-based peer-to-peer data streaming using machine learning algorithm. The Journal of Supercomputing, 2023. 79(6): p. 6802-6824, http://dx.doi.org/ 10.1007/s11227-022-04929-y.

[7] Hamdi, M.M., et al. A review on queue management algorithms in large networks. in IOP Conference Series: Materials Science and Engineering. 2021. IOP Publishing, DOI 10.1088/1757-899X/ 1076/1/012034.

[8] Hassan, S., AD-RED: A new variant of random early detection AQM algorithm. Journal of High Speed Networks, 2023(Preprint): p. 1-15, https://doi.org/ 10.3233/JHS-222.

[9] Menacer, O., A. Messai, and L. Kassa-Baghdouche, Improved variable structure Proportional–Integral Controller for TCP/AQM network systems. Journal of Electrical Engineering & Technology, 2021. 16: p. 2235-2243, doi:10.1007/s42835-021-00737-1.

[10] Kundel, R., et al. P4-codel: Experiences on programmable data plane hardware. in ICC 2021-IEEE International Conference on Communications. 2021. IEEE, https://doi.org/10.1109/ICC42927.2021.9500943

[11] Kua, J., P. Branch, and G. Armitage. Detecting bottleneck use of pie or fq-codel active queue management during dash-like content streaming. in 2020 IEEE 45th conference on local computer networks (LCN). 2020. IEEE, http://dx.doi.org /10.1109/LCN48667.2020.9314804.

[12] Park, J., Application Study of FQ-CoDel Algorithm based on QoS-guaranteed Class in Tactical Network. The Journal of The Institute of Internet, Broadcasting and Communication, 2019. 19(3): p. 53-58, http://dx.doi.org/10.7236/JIIBC.2019.19.3.53.

[13] Bhuiyan, Z.A., et al., On the (in) Security of the Control Plane of SDN Architecture: A Survey. IEEE Access, 2023, https://doi.org/10.1109/ACCESS. 2023.3307467.

[14] Alsaeedi, M., M.M. Mohamad, and A.A. Al-Roubaiey, Toward adaptive and scalable OpenFlow-SDN flow control: A survey. IEEE Access, 2019. 7: p. 107346-107379, https://doi.org/10.1109/ACCESS.2019.2932422.

[15] Cao, J., et al., Towards tenant demand-aware bandwidth allocation strategy in cloud datacenter. Future Generation Computer Systems, 2020. 105: p. 904-915, https://doi.org/10.1016/j.future.2017.06.005.

[16] Mosa Al-Mha, A. and P. Khosravian Dehkordi, Using Machine Learning to Discover Traffic Patterns in Software Defined Networks. Journal of Optimization in Soft Computing, 2025: p. 24 - 29, https://doi.org/10.82553/josc.2025.140305271129363.

[17] Jin, H., et al., OpenFlow-based flow-level bandwidth provisioning for CICQ switches. IEEE Transactions on Computers, 2012. 62(9): p. 1799-1812, https://doi.org/10.1109/TC.2012.167.

[18] Cui, H., et al. Accurate network resource allocation in SDN according to traffic demand. in 2015 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering. 2015. Atlantis Press, https://doi.org/10.2991/ icmmcce-15.2015.227.

[19] Guo, J., et al. Falloc: Fair network bandwidth allocation in iaas datacenters via a bargaining game approach. in 2013 21st IEEE international conference on network protocols (ICNP). 2013. IEEE, https://doi.org/10.1109/ICNP.2013.6733583.

[20] Xu, F., et al., Ufalloc: Towards utility max-min fairness of bandwidth allocation for applications in datacenter networks. Mobile Networks and Applications, 2017. 22(2): p. 161-173, https://doi.org/10.1007/s11036-016-0739-z.

[21] Obaid, H.S.B. and T.B. Trafalis, An approximation to max min fairness in multi commodity networks. Computational Management Science, 2020. 17(1): p. 65-77, https://doi.org/10.1007/s10287-018-0336-7.

[22] Bagci, K.T. and A.M. Tekalp, Dynamic resource allocation by batch optimization for value-added video services over sdn. IEEE Transactions on Multimedia, 2018. 20(11): p. 3084-3096, https://doi.org/10.1109/ TMM.2018.2834699.

[23] Aljoby, W., et al., On SDN-enabled online and dynamic bandwidth allocation for stream analytics. IEEE Journal on Selected Areas in Communications, 2019. 37(8): p. 1688-1702, https://doi.org/ 10.1109/JSAC.2019.2917541.

[24] Marin, A., et al. Performance evaluation of AQM techniques with heterogeneous traffic. in 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC). 2016. IEEE, https://doi.org/10.1109/CCNC.2016.7444755.

[25] Han, C., et al. Optimal active queue management for internet with reduced-order state-observer. in 2016 Chinese Control and Decision Conference (CCDC). 2016. IEEE, https://doi.org/10.1109/CCDC.2016. 7492857.

[26] Chrost, L. and A.J.T.S. Chydzinski, On the deterministic approach to active queue management. 2016. 63(1): p. 27-44, https://doi.org/10.1007/s11235-015-9969-9.

[27] Patel, S. and S.J.T.S. Bhatnagar, Adaptive mean queue size and its rate of change: queue management with random dropping. 2017. 65(2): p. 281-295, https://doi.org/10.1007/s11235-016-0229-4.

[28] Okokpujie, K.O., et al., Novel active queue management scheme for routers in wireless networks. International Journal on Communications Antenna and Propagation (I. Re. CAP), 2018. 8(1): p. 53-61, https://doi.org/10.15866/irecap.v8i1.14589.

[29] Hamdi, M.M., et al. Performance evaluation of active queue management algorithms in large network. in 2018 IEEE 4th International Symposium on Telecommunication Technologies (ISTT). 2018. IEEE, https://doi.org/10.1109/ISTT.2018.8638590.

[30] AL-Allaf, A.F. and A. A Jabbar, RED with Reconfigurable Maximum Dropping Probability. International Journal of Computing and Digital Systems, 2019. 8(01): p. 61-72, https://doi.org/10.12785/ijcds/080107.

[31] Kumhar, D. and A.J.I.J.o.I.T. Kewat, QRED: an enhancement approach for congestion control in network communications. 2021. 13(1): p. 221-227, https://doi.org/10.1007/s41870-020-00538-1.

[32] Gomez, C.A., X. Wang, and A.J.I.A. Shami, Federated Intelligence for Active Queue Management in Inter-Domain Congestion. 2021. 9: p. 10674-10685, https://doi.org/10.1109/ACCESS.2021.3050174.

[33] Chen, Q., et al., Research on assessment method of network quality performance for a private data network. Microsystem Technologies, 2021. 27: p. 1475-1482, https://doi.org/10.1007/s00542-020-05669-7.

[34] Campanile, L., et al., Computer network simulation with ns-3: A systematic literature review. Electronics, 2020. 9(2): p. 272, https://doi.org/10.3390/electronics9020272.

[35] Eyinagho, M.O., Minimum and maximum packets' delays determination for communication flows' delays jitters computation. Australian Journal of Electrical and Electronics Engineering, 2023. 20(1): p. 27-34, https://www.tandfonline.com/doi/full/10.1080/1448837X.2023.2000000.

Journal of Optimization of Soft Computing (JOSC)
Vol. 3, Issue 1, pp: (53-60), Summer-2025
Journal homepage: https://sanad.iau.ir/journal/josc

Paper Type (Research paper)

Simultaneous Network Reconfiguration and Capacitor Placement in Distribution Systems Using the Proposed Discrete PSO Algorithm with Chaos Module

Fahimeh Sayadi Shahraki[1*]

1. Department of Electrical Engineering, ShQ.C.,Islamic Azad University, Shahre-e- Qods, Iran.

Article Info

Article History:
Received: 2025/08/04
Revised: 2025/08/30
Accepted: 2025/09/14

DOI:

Keywords:
distribution system reconfiguration, sensitivity analysis, capacitor placement, discrete PSO, Chaos module.

**Corresponding Author's Email Address: sayadi.class@gmail.com*

Abstract

In this study, a simultaneous optimization method is proposed for distribution network reconfiguration in the presence of harmonic disturbances, along with determining the optimal size and location of switchable capacitors. The main objectives are to reduce active power losses and improve voltage profiles while considering operational constraints and power quality. The optimization objective function includes active power loss costs, capacitor installation costs, and penalty terms for constraint violations. To enhance convergence speed and optimization accuracy, candidate buses for capacitor placement are selected using sensitivity analysis, and the search space is efficiently reduced. The proposed algorithm is a novel Discrete Particle Swarm Optimization (PSO) with chaos module (PSOCM), which delivers fast and superior results compared to conventional methods. The applied constraints include the maximum allowable reactive power of installed capacitors and bus voltage limits according to the IEEE-519 standard. The algorithm is implemented on the Sirjan distribution network, and the results demonstrate significant performance improvements.

1. Introduction

The installation of shunt capacitors in distribution networks is generally one of the most effective methods for reducing power losses in distribution systems. Capacitor placement is also used for reactive power compensation, voltage regulation, and power factor correction. The effectiveness of compensation largely depends on the capacitor's location within the distribution system. Therefore, determining the optimal placement, sizing, and type of capacitors in the distribution network is essential [1-2].

Distribution network reconfiguration is another effective method for loss reduction. Medium-voltage distribution networks are typically designed with a loop structure but operated radially. These networks contain normally closed switches and several normally open switches that can be reconfigured to achieve an optimal radial configuration, thereby reducing losses while improving bus voltage profiles. Numerous techniques with various approaches have been proposed for optimal capacitor placement [3-4] and distribution network reconfiguration [5-9]. Some studies have addressed simultaneous reconfiguration and capacitor placement [10-15]. In [10], the status of capacitors and network branches is modified using an ant colony algorithm, ultimately determining which branches should remain open upon convergence. In [12] a P-PSO algorithm is employed to capacitor placement and reconfiguration in the presence of non-linear loads. In [13], an improved adaptive genetic algorithm is employed for optimal capacitor placement as the primary objective. Although the evaluation of the results indicates that the condition of preventing loop formation in the reconfiguration process has not been met. In [14], simultaneous reconfiguration and capacitor

placement are performed using a binary genetic algorithm, considering different load patterns to reduce losses; however, a closer examination reveals that the radiality constraint was not strictly enforced.

While most of these techniques are computationally fast, their main weakness lies in their susceptibility to local optima. To date, the simultaneous optimization of network configuration and capacitor placement considering harmonic loads has not been adequately addressed. This paper presents a comprehensive approach that incorporates harmonic conditions under varying load levels in the network. The methodology first performs optimal capacitor placement with the objective of loss reduction while adhering to voltage magnitude constraints. Subsequently, the same optimization is conducted simultaneously with network reconfiguration to determine the optimal system configuration.

Given the inherent complexity of this optimization problem, a novel two-layer Particle Swarm Optimization (PSO) algorithm is proposed. This innovative approach enhances particle diversity and significantly improves the algorithm's ability to avoid premature convergence to local optima, thereby ensuring more robust and globally optimal solutions. The proposed optimization method is similar to the one suggested in [15], with the difference that the chaos generation mechanism in particles has been enhanced, further reducing the likelihood of getting trapped in local optima. The rest of the paper is organized as follows. In section 2 problem formulation consist of objective function and constraints formulations are presented. Section 3 Describes how to implement PSOCM method. The implementation method of the reconfiguration process is described in Section 4. Simulation scenarios and results are provided in section 5 and section 6 discusses the results and concludes the paper.

2. Problem Formulation

This section presents the mathematical formulation of the problem, including the objective function, problem constraints, and the power flow calculation framework.

2.1. Objective Function and Problem Constraints

The primary objectives of optimal network reconfiguration and capacitor placement are to determine:

1.The optimal network configuration

2.The optimal locations and sizes of capacitor units in the distribution system

3.The minimization of energy losses and active power losses at both fundamental and harmonic frequencies

The capacitor placement objective function to be minimized is formulated as follows:

$$F = K_e T P_{loss} + K_p P_{loss} + \sum_{j=1}^{nc} C_j B_j + \alpha \sum_{i=1}^{n} |(1 - V_i)| \quad (1)$$

In equation (1), ke and kp represent the constant cost coefficients of energy losses and power losses respectively and Bj and Cj denote the available capacitor kVAr and the cost coefficient per kVAr of capacitors installed at each candidate bus, respectively. This function consists of the sum of four components: power loss cost, energy loss cost, capacitor installation cost, and a voltage violation penalty to maintain maximum voltage regulation.

Candidate buses are determined through sensitivity analysis [16].

Capacitor Unit Constraints: The total installed capacitor units are also constrained by the following equation:

$$\sum_{j=1}^{nc} B_j < B \quad (2)$$

Typically, B is selected as the sum of reactive power loads connected to the studied network.

For the network reconfiguration problem, the following constraints are applied:

1. Fixed number of system branches (the total number of lines remains constant)

2. Radial topology preservation (the network must maintain a radial structure after reconfiguration).

2.2. Power Flow Calculations

template, prepare your technical work in single-For distribution system analysis, a specialized power flow algorithm is required to determine power losses and bus voltages at the fundamental frequency. This study employs the backward-forward sweep method, specifically designed for radial distribution systems.

Implementation Steps:

1.Formation of BIBC Matrix: A three-phase Bus Injection to Branch Current (BIBC) matrix is constructed based on the system's topological structure. This matrix establishes the relationship between bus current injections and branch currents

2.Iterative Voltage Calculation: All bus voltages are computed through an iterative process using the BIBC matrix. The voltage at each three-phase bus during the k-th iteration is calculated using Equation (3) [17]:

$$[V]^{k} = [V_0] - [BIBC]^{T}[Z][BIBC][I]^{(k)} \quad (۳)$$

Where:
- $[V]^{(k)}$: Vector of bus voltages at iteration k
- V_0: Vector of slack bus voltages (reference node)
- [BIBC]: Bus Injection to Branch Current matrix (topology-based current distribution mapping)
- [Z]: Primitive impedance matrix (includes line impedances and mutual couplings)
- $[I]^{(k)}$: Load current vector at iteration k (calculated from power demands and voltages)

3. PSO Algorithm

PSO is an optimization technique designed to solve complex optimization problems. The fundamental concept relies on generating a random population where each individual, called a "particle," represents a potential solution. Each particle dynamically adjusts its position and velocity in the search space based on: Its own flight experience (cognitive component) and the collective knowledge of neighboring particles (social component)

3.1. Basic PSO Algorithm

current solution coordinates of particle i, Particle Velocity is defined as: V_i=($V_{i1}, V_{i2},\ldots, V_{iD}$) Determines the direction and magnitude of movement. P_i=($P_{i1}, P_{i2},\ldots, P_{iD}$) is best position found by particle i so far. P_g=($P_{g1}, P_{g2},\ldots, P_{gD}$) is best position found by the entire swarm.

At each iteration k+1, velocity and position update as [18]:

$$(t+1) = V_{id}(t) + c_1 r_1 \left(P_{id}(t) - X_{id}(t)\right) + c_2 r_2 \left(P_{gd}(t) - X_{id}(t)\right) \quad (4)$$

$$X_{id}(t+1) = X_{id}(t) + V_{id}(t+1) \quad (5)$$

c_1 and c_2 are acceleration constants controlling cognitive and social influence and r_1 and r_2 are uniformly distributed random numbers $\in$ [0,1]. Particle velocities are constrained by V_{max} to prevent overshooting. For capacitor placement and switch status optimization: Binary representation for capacitor units (0=absent, 1=present) Switch states (0=open, 1=closed). When minimizing objective function f in D-dimensional space, particle ii updates its position at iteration t+1 as:

$$P_i(t+1) = \begin{cases} X_i(t+1) & if\ f\left(X_i(t+1)\right) < f(P_i(t)) \\ P_i(t) & other\ wise \end{cases} \quad (6)$$

3.2. Proposed PSO Algorithm

Here, to better control the exploration and exploitation capabilities, the parameter ω depends on the fitness of the particles (rather than time). Therefore, particles with lower fitness are assigned lower velocities to aid exploitation, while particles with higher fitness values are assigned higher velocities, guiding them towards greater exploration. The velocity of the i-th particle is calculated as follows:

$$\omega_i = (\omega_{max} - \omega_{min}) * G_i + \omega_{min} \quad (7)$$

ω_{max} and ω_{min} are the maximum and minimum velocity values, respectively, equal to 0.9 and 0.4. The fitness G_i is normalized as follows:

$$G_i = \frac{f(P_i) - f_{min}}{f_{max} - f_{min}} \quad (8)$$

$f_{max}$and f_{min} represent the maximum and minimum fitness values of the personal experience of each particle in the population. According to Equation (8), G_i decreases for a particle with lower fitness and vice versa. Finally, the velocity V_i of the ith particle is updated as follows:

$$V_{id}(t+1) = \omega_i V_{id}(t) + c_1 r_1 \left(P_{id}(t) - X_{id}(t)\right) + c_2 r_2 \left(P_{gd}(t) - X_{id}(t)\right) \quad (9)$$

P_k represents the best personal experience among the neighboring particles in the vicinity of the ih particle. N_i is the number of particles in the neighborhood, c_k is the acceleration coefficient, which is uniformly distributed among the neighboring particles as ($c_k = c/|N_i|$), where c = 4.1, and r_k is a random number in the range [0, 1].

3.2.1. Proposed Chaos module

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

The last line of a paragraph should not be printed by itself at the beginning of a column nor should the first line of a paragraph be printed by itself at the end of a column.

This module prevents premature convergence by introducing controlled chaos when optimization stagnates (no improvement for 5 or more iterations).

Counter f_c increments by 1 each iteration if no fitness improvement occurs and triggers disturbance when f_c>m (threshold m = 5).

Perturbation Process:

- Randomly select 10%–50% of dimensions (D) from the global best solution (P_g).
- Modify selected dimensions using:

$$P_{g-disturbed} = P_g + \sigma \cdot N(0'1) \cdot I_{selected} \quad (10)$$

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

Where:

σ: Scaling factor (0.4).

(0,1): Standard Gaussian noise.

I_selected: Binary mask (1 for selected

dimensions, 0 otherwise).

Here, since the presence or absence of each capacitor unit and the open/closed status of each line switch are determined using binary values (0 and 1) respectively, Binary PSO is employed. In this approach, whenever the particle's position is updated, it is converted to binary form using the following sigmoid function:

$$\text{Sigmoid}\left(P_{id}{}^{k}\right) = \frac{1}{1+e^{-P_{id}{}^{k}}} \quad (11)$$

$$P_{id}{}^{k} = \begin{cases} 1 & \text{if rand} < S\left(P_{id}{}^{k}\right) \\ 0 & \text{other wise} \end{cases} \quad (12)$$

4. Reconfiguration Process

To increase the speed and quality of optimization, appropriate tools have been used wherever possible. In the case of capacitor placement, candidate buses are determined using sensitivity analysis. For the reconfiguration process, if the status of all network switches were to be determined solely by optimization under the constraint of maintaining a radial network, the search space would become excessively large, reducing the likelihood of reaching an optimal solution. Therefore, to reduce the search space, first, switches whose status must clearly remain closed are eliminated from the problem's solution space. That is, switches feeding end buses must definitely remain closed and are not considered as unknowns in the problem.

However, in most conventional methods, a potential solution is not considered valid until it passes the radiality test, which itself hinders the speed and accuracy of optimization. In this paper, a method of dividing the network into different loops is used, the details of which are presented in [19]. Each loop consists of one normally open switch and several normally closed switches. Then, following, the loop spreading matrix and the T-node degree method are employed to ensure that the resulting configuration from optimization is radial and supplies all buses in the system. The flowchart of the proposed method for solving the problem is shown in Figure 1.

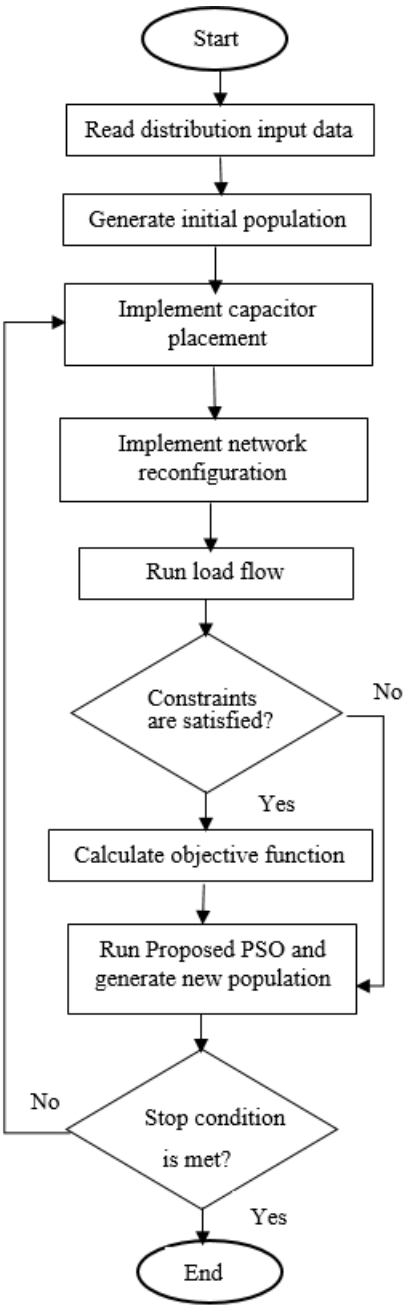


Figure 1: Flowchart of the proposed method

5. Numerical Results

The effectiveness of the proposed method for simultaneous reconfiguration and capacitor placement in the distribution network is demonstrated using the 77-bus Sirjan network (figure 1), whose load and line specifications are provided in [14].

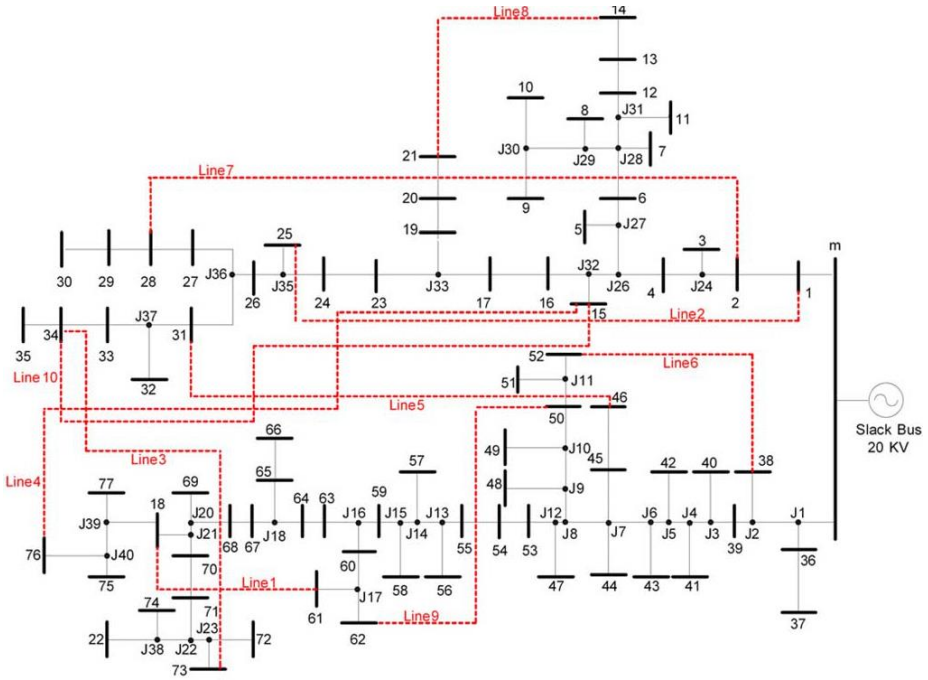


Figure 2: Single-line diagram of the 77-bus Sirjan distribution network

Table 1: Capacitor Units (in kVAR) and Their Cost Coefficients (in $/kVAR)

Capacitor (kVAR)	**150**	**300**	**450**	**600**	**750**	**900**	**1200**	**1350**
Cost Coefficient ($/kVAR)	0.5	0.35	0.253	0.22	0.276	0.183	0.17	**0.207**

The network consists of 114 normally closed switches (numbered) and 10 normally open switches, with their corresponding lines labeled as $Line_1$, $Line_2$, …, $Line_{10}$.

Using sensitivity analysis, the most sensitive buses in the system are identified as Buses 76, 75, 42, 59, 39, 38, 37, 24, 11, and 9.

To demonstrate the effectiveness of the proposed method for simultaneous network reconfiguration and capacitor placement, as well as the capability of the proposed optimization technique, six different cases are implemented on the Sirjan network. The coefficient K_p in Equation (1) is set to 150 according to reference [12], while the value of K_e is considered as a variable based on reference [20]. Optimization is performed using the proposed method in all cases except Case 4:

Case 1: Network reconfiguration only
Case 2: Capacitor placement only
Case 3: Reconfiguration followed by capacitor placement
Case 4: Simultaneous optimal capacitor placement and network reconfiguration using standard PSO
Case 5: Simultaneous optimal capacitor placement and network reconfiguration using the proposed method

Table 2: Network Reconfiguration Results

Case	**Switches to be Opened**
Cases 1 & 3	12-13, 15-32j, 16-17, 27-28, 31-36j, 7j-8j, 9j-10j, 63-16j, 70-71, Line1

Case	Switches to be Opened
Case 4	12-31j, 15-32j, 24-35j, 26-35j, 31-36j, 7j-8j, 9j-10j, 17j-61, 64-18j, 18-21j
Case 5	12-13, 15-32j, 18-21j, 26-35j, 27-28, 50-10j, 67-18j, 7j-8j, 4-26j, 18-39j

Table 3: Program Execution Results

Parameter	Before Compensation	Case 1	Case 2	Case 3	Case 4	Case 5
Minimum Bus Voltage (pu)	0.944	0.998	0.987	0.998	0.998	0.998
Maximum Bus Voltage (pu)	0.998	1.000	1.000	1.000	1.000	1.000
Active Power Loss (kW)	229.443	201.85	208.00	196.00	192.00	

The available capacitor units and their cost coefficients considered in the objective function are listed in Table 1.
The coefficient α in the term included in the objective function to enhance voltage regulation is set to 10,000. For the optimization process, the population size is set to 50 and the maximum iterations to 100. The results obtained from running the program are presented in Tables 2 and 3. The parameters for conventional PSO are set as follows: c_1 and c_2 equal to 1.99 and 2.05 respectively, r_1 and r_2 are random numbers between 0 and 1, the maximum velocity is set to 1.05, and the population size is 50. Table 2 shows the results of the network reconfiguration program execution. According to Table 3, standalone capacitor placement (Case 2) improves voltage but remains below optimal (0.987 pu min). Combined approaches show better voltage profile improvement. proposed method achieves optimal voltage profile (0.998-1.033 pu), Lowest power losses (177 kW) and 23% reduction compared to base case.

Table 4: Capacitor kVAR Allocation at Candidate Buses

Bus Number	76	75	42	59	39	38	37	24	11	9	Total
Case 2	1050	-	-	900	1650	-	-	1650	-	-	5250
Case 3	300	-	600	-	900	600	300	600	450	900	4650
Case 4	300	-	750	1050	-	600					

The results demonstrate that implementing simultaneous network reconfiguration and capacitor placement reduces the total required kVAr capacity compared to standalone capacitor placement. This reduction is particularly significant when using co-optimization, which harmonizes both network topology and capacitor values for optimal performance. The proposed method achieves faster convergence speed

(reduced iteration count) and higher solution quality (improved objective function values). The convergence characteristics are visualized in Figure 3.

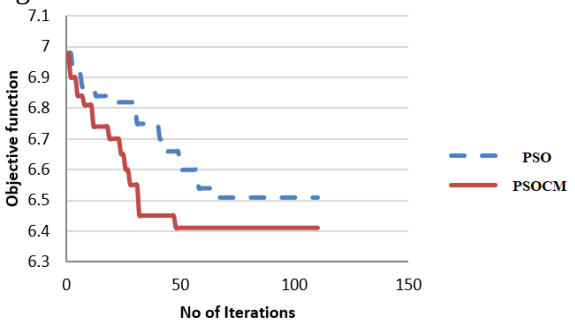


Figure 3: Convergence Comparison Between the PSO and PSOCM

6. Conculusion

Implementation on the 77-bus Sirjan test system proves the algorithm's practical effectiveness, delivering simultaneous improvements across three critical aspects: 23% reduction in power losses, voltage profile enhancement within 0.998-1.00 pu range, and 11.4% decrease in required capacitor investment (4650 kVAr vs 5250 kVAr). The coordinated optimization of capacitor placement and network reconfiguration yields solutions that properly balance technical and economic objectives.

The results demonstrate that the proposed chaotic-enhanced PSO algorithm successfully overcomes the limitations of conventional optimization approaches in solving complex distribution network problems. By intelligently integrating chaotic search mechanisms with sensitivity analysis and loop-based network partitioning, the method achieves superior performance in both solution quality and computational efficiency. These improvements stem from the algorithm's adaptive search strategy that dynamically adjusts exploration/exploitation balance while maintaining feasible radial configurations through innovative constraint-handling techniques.

References

[1] Al-ammar, E. A. et al., "Comprehensive impact analysis of ambient temperature on multi-objective capacitor placements in a radial distribution system", Ain Shams Eng. J. 12(1), 717–727 (2021), 10.1016/j.asej.2020.05.003.

[2] Asabere, P., Sekyere, F., Ayambire, P. & Ofosu, W. K., "Optimal capacitor bank placement and sizing using particle swarm optimization for power loss minimization in distribution network", J. Eng. Res,10.1016/j.jer.2024.03.007.

[3] Mouwafi, M. T., El-Sehiemy, R. A. & El-Ela, A. A. A., "A two-stage method for optimal placement of distributed generation units and capacitors in distribution systems", Appl. Energy 307, 118188 (2022), 10.1016/ j.apenergy.2021.118188.

[4] Elseify, M. A., Hashim, F. A., Hussien, A. G. & Kamel, "S. Single and multi-objectives based on an improved golden jackal optimization algorithm for simultaneous integration of multiple capacitors and multi-type DGs in distribution systems", Appl. Energy 353, 122054 (2024), 10.1016/ j.apenergy.2023.122054.

[5] Hoseini, S. E., Simab, M., & Bahmani-Firouzi, B.,"AI-Based Multi-Objective Distribution Network Reconfiguration Considering Optimal Allocation of Distributed Energy Storages and Renewable Resources", International Journal of Smart Electrical Engineering, 14(2), 67-82, 2025, https://doi.org/10.82234/ijsee.2025.1197454.

[6] G. Vulasala, S. Sirirgiri, and S. Thiruveedula, "Feeder reconfiguration for loss reduction in unbalanced distribution system using genetic algorithm," Int. J. Elect. Power Energy Syst. Eng., vol. 2, no. 4, pp. 240–248, Feb. 2009, https://doi.org/10.1007/s00500-023-09472-3.

[7] E. López, H. Opazo, L. García, and P. Bastard, "Online reconfiguration considering variability demand: Applications to real networks," IEEE Trans. Power Syst., vol. 19, no. 1, pp. 549–553, Feb. 2004,10.1109/TPWRS.2003.821447.

[8] J. Z. Zhu, "Optimal reconfiguration of electrical distribution network using the refined genetic algorithm", Electric Power Systems Research, vol. 62, no. 1, pp. 37-42, 2002, https://doi.org/10.1016/S0378-7796(02)00041-X.

[9] A.Y. Abdelaziz, F.M. Mohammed, S.F. Mekhamer and M.A.L. Badr, "Distribution Systems Reconfiguration using a modified particle swarm optimization algorithm", Electric Power Systems Research, vol. 79, no. 11, pp. 1521-1530, 2009,https://doi.org/10.1016/j.epsr.2009.05.004.

[10] C. F. Chang, "Reconfiguration and capacitor placement for loss reduction of distribution systems by ant colony search algorithm," IEEE Trans. Power Syst., vol. 23, no. 4, pp. 1747–1755, Nov. 2008, 10.1109/TPWRS.2008.2002169.

[11] Z. Rong, P. Xiyuan, H. Jinliang, and S. Xinfu, "Reconfiguration and capacitor placement for loss reduction of distribution systems," in Proc. IEEE TENCON'02, 2002, pp. 1945–1949, 10.1109/TENCON.2002.1182719.

[12] Sayadi F., Esmaeili S., and Keynia F., "Feeder reconfiguration and capacitor allocation in the presence of non-linear loads using new PPSO algorithm", IET. Gener. Transm. Distrib., 2016, 10, (10), pp. 2316–2326, https://doi.org/10.1049/iet-gtd.2015.0936.

[13] D.Zhang, Z. Fu, and L. Zhang, "Joint optimization for power loss reduction in distribution systems," IEEE Trans. Power Syst., vol. 23, no. 1, pp. 161–169, Feb. 2008, 10.1109/TPWRS.2007.913300.

[14] V. Farahani, B. Vahidi, "Reconfiguration and Capacitor Placement Simultaneously for Energy Loss Reduction Based on an Improved Reconfiguration Method", IEEE Trans. Power Syst., vol. 27, no. 2, pp. 587-595, 2012, 10.1109/TPWRS.2011.2167688.

[15] Sayadi Shahraki. F, Bakhtiari. Sh, Zamani Nouri, "A, Optimal use of photovoltaic systems in the distribution network considering the variable load and production profile of Kerman city", Optimization in Soft Computing, pp. 56-65, 2025, doi.org/10.82553/josc.2025.140310121195201.

[16] Sayadi, F., Esmaeili S., Keynia F., "Two-layer volt/var/total harmonic distortion control in distribution network based on PVs output and load forecast errors", IET Gener. Transm. Distrib. 11(8), 2130–2137 (2017), 10.3390/ electricity6020028.

[17] Jen-HaoTeng, Chuo-Yean Chan "Backward/ ForwardSweep- Based Harmonic Analysis Method for Distribution Systems", IEEE Transactions on Power Delivery, VOL. 22, NO. 3, JULY 2007, 10.1109/TPWRD.2007.899523.

[18] N. Mohsenifar, "Investigating the Impact of Distributed Generation (DG) in Radial Distribution Networks and Optimizing Protective Devices Using the PSO Optimization Algorithm", Journal of Optimization in Soft Computing, Vol.2, 2024, doi.org/10.82553/josc.2024.140302091118461.

[19] Yu J, Zhang F, Ni F, Ma Y., "Improved genetic algorithm with infeasible solution disposing of distribution network reconfiguration", IEEE Proc 2009 WRI Global Congr Intell Syst 2009;2:48–52, 10.1109/GCIS.2009.194.

[20] S.P. Singh , A.R. Rao, "Optimal allocation of capacitors in distribution systems using particle swarm optimization", Electrical Power and Energy Systems, Volume 43, Pages 1267-1275, 2012, https://doi.org/10.1016/j.ijepes.2012.06.059.