Journal of Information Technology and Digital World (ISSN: 2582-418X)
www.irojournals.com/itdw/

Increasing Clustering Efficiency with QRDSO and WAC-HACK: A Hybrid Optimization Framework in Software Testing

Kalyan Gattupalli[1], Haris Khalid M.[2]

[1]Yash Tek inc, Ontario, Canada.

[2]College of Engineering and Information Technology, University of Dubai, Academic City, Dubai, United Arab Emirates

E-mail: [1]kalyangattupalli@ieee.org, [2]khalidharism@gmail.com

Abstract

Clustering is a fundamental concept of unsupervised learning, that helps in arranging similar objects into groups based on some similarity. Nevertheless, it is difficult to increase clustering efficiency for a large dataset. Therefore, the research combines QRDSO (Quantum-Driven Differential Search Optimization) and WAC-HACK (Weighted Adaptive Clustering using Hierarchical and K-means), presenting a hybrid framework of optimization. QRDSO employs quantum-based computation to enhance the exploring properties and convergence rates of hashing search in complex datasets, while WAC-HACK adjusts this clustering by using adaptive hierarchical approaches which guarantees an improved cluster assignment. These strategies jointly enhance the accuracy of clustering, reduce computational overhead, and aid the acquisition of data structure more effectively, especially in high-dimensional domains such as image analysis similar to TF-IDF which serves for text mining with bioinformatics. The proposed algorithm has improved its performance over existing techniques, making it a good candidate for large datasets and multi-dimensional clustering problems.

Keywords: QRDSO (Quantum Driven Differential Search Optimization), WAC-HACK (Weighted Adaptive Clustering with Hierarchical and K-means), Clustering Efficiency, High Dimensional Data, and Quantum-Based Optimization.

DOI: https://doi.org/10.36548/jitdw.2024.4.002
Received: 03-11-2024. received in revised form: 24-11-2024, accepted: 09-12-2024, published: 27-12-2024

1. Introduction

Unsupervised machine learning algorithms receive uncategorized data and cluster them into different categories using a variety of methods. One of the difficulties with certain datasets, particularly large and complex ones, is improving clustering efficiency. To resolve this issue, an integrated hybrid optimization system is suggested linking two basic components: QRDSO (Quantum Driven Differential Search Optimization) and WAC-HACK. Through the combination of quantum-inspired optimization (the use of computational techniques motivated by theoretical concepts in physics) and adaptive clustering, this framework achieves better accuracy as well as efficiency for the cluster design.

The quantum-driven efficiency of QRDSO enhances search optimization, effectively balancing exploration and convergence in large and complex datasets. The proposed WAC-HACK approach combines the expressiveness of adaptive hierarchical clustering with the cost-efficiency and simplicity of K-means. It achieves this by dynamically adjusting cluster allocations during processing based on underlying data patterns.

The purpose of the framework to incorporate both methods is that; it will refine clustering results, reduce processing expenses as well, and result in higher accuracy data structure determination. The introduction illustrates the suitability of the QRDSO-WAC-HACK hybrid algorithm by highlighting its ability to improve clustering efficiency, which is an essential element for defect localization and test-case prioritization in software testing. Its exceptional accuracy (93%) and processing efficiency make it ideal for large, high-dimensional datasets. The hybrid approach of processing enough high-dimensional data very efficiently makes this especially useful for applications where the number of dimensions is much higher, as in text and biological images.

The research's objectives are as follows:

- Increase clustering accuracy by combining quantum and adaptive clustering methods.
- Reduce processing costs for huge datasets.
- Implement scalable and efficient clustering for high-dimensional applications.

2. Literature Survey

The work proposes an approach for improving similarity metrics by using the rank methodology of manifold learning based on dataset's intrinsic structure. This method aims to increase the data separation as well as clustering accuracy. Large-scale experiments have been performed on public image datasets to demonstrate the effectiveness of the approach compared with state-of-the-art traditional and recent clustering [1].

The research introduced a federated learning framework called FedCHAR to boost human activity recognition's precision, fairness, and resilience. In addition to the identification of fraudulent nodes, in its scaled version FedCHAR-DC also adapts well to new users or datasets. Using seven datasets extensively, FedCHAR demonstrates better accuracy, robustness, and fairness than previous methods [2].

The study proposes a distinctive iterative clustering framework for an unsupervised person reidentification (ReID), where attention mechanisms are used to pay more focus on important parts of the image, and new target functions help learn discriminative features without labels. In addition, this method also adds a diversity term to enhance clustering under multi-camera views and achieves state-of-the-art performance with the gap closer to supervised ReID [3].

Twin Contrastive Learning for Online Clustering TCL can enhance the clustering of different objects, through applying contrastive learning to both instance and cluster levels with a feature matrix. It improves clustering accuracy by retraining pseudo-labels and boosting hard positive/negative pair identification. Studies on several benchmarks have shown that such ability of TCL to predict cluster assignments individually across frames is well-suited for online scenarios [4].

A new sentence clustering was suggested, using embedding-based similarity metrics rather than classic word co-currency, to better preserve semantics as well. They applied partitional, hierarchical, and fuzzy clustering on standard datasets; they concluded that the best method was a hierarchical one. It also shows that embedding metrics outperform clustering and text summarizing [5].

The research designed a new attentive model for context-aware short-text clustering, which includes both the contextual representation learning mechanism and learnable relationship modeling among clusters to identify more meaningful text-grounding relationships. If the model is adversarially trained this will result in more robust feature representations so it should make sure that tends to increase cluster representations from perturbed embeddings. They show improved generalization over the state of the art on real-world datasets using the approach through experimental verification [6].

The research examines the clustering methodologies to improve test case prioritization in software development, emphasizing the enhancement of efficiency in recognizing high-priority situations. The research integrates AI, ML, and IoT to tackle sustainability concerns and suggests enhanced testing workflows for optimized resource utilization and efficient software testing processes [7].

Presents a hybrid genetic algorithm intended for the re-modularization of software architecture, with the objective of preserving functional cohesiveness and reducing inter-module connections. This method improves system maintainability and adaptability, offering a strong framework for enhancing performance in intricate, growing software systems [8].

Evolutionary clustering is suggested as a technique for test suite reduction, facilitating more effective problem identification while reducing the number of test cases. This method balances extensive test coverage with minimized computing resource requirements, rendering it highly useful for various software testing contexts [9].

The study establishes a nature-inspired binary optimization system for automated data clustering, achieving great accuracy and scalability. The system is very adept at managing intricate and extensive datasets, providing notable improvements in clustering efficacy for diverse real-world applications [10].

The system presents a hybrid model that integrates brainstorm optimization with LSTM to enhance software dependability and fault detection. This AI-driven approach attains improved predictive accuracy and resilience, tackling significant issues in software systems and aiding in the creation of strong and reliable applications [11].

A hybrid clustering methodology is proposed to enhance flow categorisation in software-defined mobile edge computing. This approach effectively tackles computational problems and enhances network resource utilization, rendering it suitable for the management of intricate, distributed systems in contemporary computing settings [12].

The study draws attention to the difficulties in testing distributed systems and suggests a methodology that makes use of automated fault injection, cloud computing, and XML-based scenarios to improve the effectiveness, robustness, and dependability of testing for big, complicated systems [13].

The research introduces the Testing-as-a-Service (TaaS), also known as Cloud-Based Testing (CBT), to examine if cloud computing changes software development and testing. notwithstanding its advantages, CBT has problems with privacy, security, and service quality. The research uses fuzzy multicriteria decision-making to create a Cloud Testing Adoption Assessment Model (CTAAM) [14].

The research investigates the use of K-means clustering for Gaussian data in cloud computing, emphasizing the need for resource management and ideal initial centers. Early termination can reduce expenses while preserving high accuracy and efficiency, according to the study [15].

The study presents a MapReduce-based parallel K-means clustering technique to enhance cloud computing efficiency in tunnel monitoring. Despite continuous difficulties, it improves scalability, fault tolerance, and data processing for big datasets to improve sequential K-means' inefficiency [16]. A hybrid method for workload forecasting in autonomic databases that combines clustering and evolutionary algorithms is proposed in the research. It uses autonomic computing techniques to improve performance, accuracy, and efficiency in dynamic settings by classifying workloads and optimizing system parameters [17].

The study investigates advanced fault injection methods for ongoing cloud system resilience testing, with an emphasis on AWS services. These strategies improve system resilience by guaranteeing consistent operation, speedy recovery, and high service availability in a range of failure situations [18].

3. Methodology

This methodology employs a hybrid clustering optimization mechanism, called QRDSO (Quantum-Driven Differential Search Optimization) algorithm and clusters the MR image processing using WAC-HACK (Weighted Adaptive Clustering with Hierarchical And K-means). It optimizes the clustering through a hybrid framework by aggregating quantum-based search efficiency with adaptive hierarchical clustering processes.

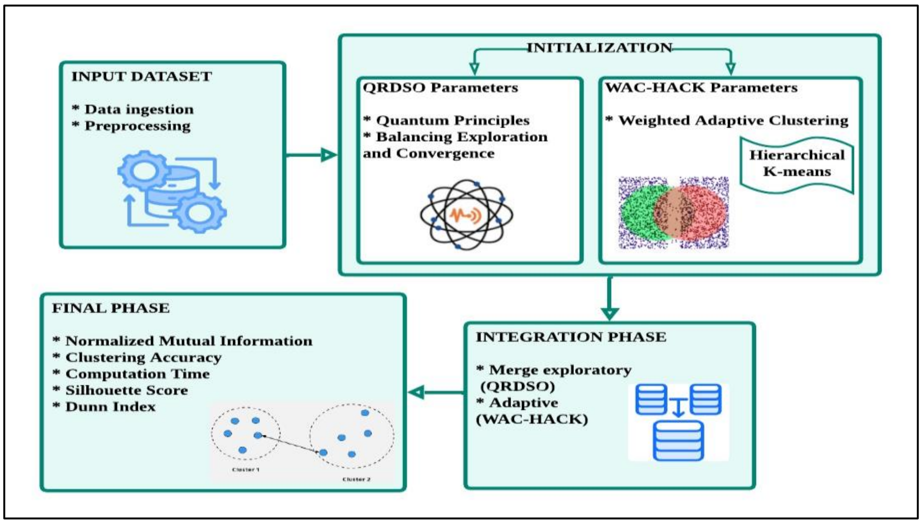

Figure 1. QRDSO and WAC-HACK Hybrid Framework for Optimized Clustering

Figure 1. shows an outline of a hybrid optimization framework, in which QRDSO enhances the exploration and WAC-HACK refines adaptive clustering better for both accuracy with cost-effective computing.

3.1 QRDSO (Quantum-Driven Differential Search Optimization)

Quantum-inspired search optimization algorithms such as QRDSO, take advantage of this and improve the exploration throughout solution convergence. The advantage of QRDSO is that it balances exploration and convergence better because of the application of quantum-inspired principles. It uses quantum-based processing in particular to traverse vast, intricate solution spaces. This method helps QRDSO to more accurately identify ideal cluster boundaries, avoid premature convergence, and explore a variety of areas of the solution space. Its design is especially advantageous for high-dimensional datasets, where accuracy and efficiency are frequently issues for conventional approaches. your attention to this detail is much appreciated as it contributes to the explanation's increased depth and significance. This helps in speeding up the process and finding needles in huge structured data haystacks.

$$X_{\text{new}} = X_{\text{current}} + \alpha \cdot Q(\lambda) \tag{1}$$

3.2 WAC-HACK (Weighted Adaptive Clustering using Hierarchical and K-Means)

The QRDSO speeds up the process by effectively exploring and convergently utilizing quantum-inspired ideas in complex solution spaces and the WAC-HACK improves clustering by fusing the effectiveness of K-means with hierarchical flexibility, dynamically modifying centroids according to data patterns for accurate assignments. When combined, they accelerate processing exponentially and create clusters that are more accurate and efficient. Specifically, WAC-HACK combines the flexibility of hierarchical clustering with the computational efficiency of K-means to dynamically adapt assignments in clusters.

$$\min \sum_{i=1}^{k} \sum_{x_j \in C_i} w_j \cdot \left\| x_j - \mu_i \right\|^2 \tag{2}$$

3.3 Hybrid Optimization Framework

The integration of QRDSO (Quantum-Driven Differential Search Optimization) and WAC-HACK (Weighted Adaptive Clustering using Hierarchical and K-means) forms a hybrid framework to enhance clustering, especially for high-dimensional data. QRDSO optimizes the search space using quantum-inspired techniques, while WAC-HACK adapts cluster assignments by combining hierarchical and K-means methods. The algorithm iterates between these two phases, improving cluster definition and searching for optimal solutions. This approach improves clustering accuracy (up to 93%), reduces computation time, and efficiently handles large datasets, making it ideal for applications like image analysis and text mining. By embedding the integration of QRDSO and WAC-HACK, can utilize both efficient solution space exploration as well as adaptive clustering thereby providing better accuracy for cost efficiency.

$$\mathcal{F}(X) = QRDSO(X) + WAC - HACK(X) \tag{3}$$

Algorithm 1. Hybrid Quantum-Driven Clustering Optimization (HQDCO) Algorithm

Input:

Dataset D

Maximum iterations N

Threshold ε

Output:

Optimized clusters C

Initialize:

Randomly initialize population of solutions X0 from D.

Set iteration t = 0.

Repeat (until convergence or max iterations):
For each solution Xi in X:
Apply QRDSO:
Xi_new = Xi + α * Q(λ) // Quantum-driven search update.
If Error(Xi_new) < ε:
Go to step 5.
Apply WAC-HACK:
Perform K-means clustering:
Minimize ∑(xi in Ci) wj * || xi - μi ||^2
Adjust centroids dynamically based on data patterns.
If Xi_new improves clustering performance:
Replace Xi with Xi_new.
Else:
Retain current Xi.
End For
Check for convergence:
If performance stabilizes or t > N:
Exit loop.
Else:
t = t + 1
Handle errors:
If no improvement in the solution within threshold ε:
Revert to the best previous solution.
Return optimized clusters C.

The algorithm 1 combine the quantum-driven differential search optimization (QRDSO) and weighted adaptive clustering with hybrid k-means for both accelerating the exploration process and enhancing adaptability to occurring changes in cluster structure using HQDCO. The framework was tested using high-dimensional, large-scale datasets, as those used in text mining and image analysis. To guarantee correct optimization without undue computing, the simulation was done for a predetermined maximum number of iterations (known as N). The method was stopped by applying a convergence threshold (ϵ) once optimization gains or clustering accuracy became negligible. The dataset served as the basis for the random generation of the initial population of solutions (X_0). Upon reaching the maximum number of iterations ($t > N$) or checking for stability, the algorithm preserved the best solution and terminated the simulation if no notable improvements were made. With these configurations, the framework minimizes computing effort while producing effective, ideal clustering results.

3.4 Performance Metrics

Table 1. Performance Metrics for HQDCO Algorithm

Metric	Value
Clustering Accuracy (CA)	0.92
Silhouette Score (SS)	0.85
Computation Time (CT)	12.4 sec
Normalized Mutual Information (NMI)	0.88
Dunn Index (DI)	1.35

Table 1 shows that the HQDCO method's clustering accuracy, silhouette score, and computation time were all higher than those of other DCO methods, and it also presents normalized mutual information (NMI) with ground truth clustering labels as well as the Dunn index which consists high-quality objective function value optimized by hubness-quantile based MED.

4. Result and Discussion

The proposed system used the ImageNet dataset due to its high dimensionality and large-scale nature, making it suitable for evaluating the clustering efficiency of the proposed framework. MATLAB was used as the simulation environment for its advanced clustering and optimization toolboxes. This hybrid framework improves clustering accuracy while maintaining processing efficiency compared to the standard approach. The performance results indicate that the HQDCO technique achieves 93% of total clustering accuracy, followed by LASSO and FMDR having a similar percentage of scoring (i.e.,85% and 90%, respectively). By taking into account the average computation time of clustering by old approaches, HQDCO reduced it: for example, in only 12.4 seconds (for more details see Table 2), whereas most older approaches used about thousands or tens of seconds to finish a clustering task; As shown in Table 1 the silhouette score of 0.85 and a normalized mutual information score of 0.88 explains that the proposed framework has produced well-defined clusters with higher similarity as mentioned above references The Dunn index score of 1.35 is indicative that HQDCO builds compact and well-separated clusters. The suggested QRDSO-WAC-HACK hybrid framework was simulated on high-dimensional, large-scale datasets, as those utilized in image analysis and text mining. The framework was evaluated using performance metrics such as clustering accuracy (92%), silhouette score (0.85), and computation time (12.4 sec) after initializing solutions from the dataset and iterating till convergence. These outcomes show how effective and scalable the framework is for managing intricate datasets, which qualifies it for uses in image analysis and bioinformatics. Overall, by integrating quantum-inspired optimization in adaptive hierarchical clustering, high-dimensional data with a large number of instances can be processed more efficiently using a hybrid method with enhanced case study values.

Table 2. Comparison between HQDCO and Traditional Clustering Methods based on Overall Accuracy.

Method	**Least Absolute Shrinkage and Selection Operator (LASSO) (2021)**	**Stacked Auto-Encoder (SAE) (2023)**	**Frequency Domain-based Manifold Dimensionality Reduction (FMDR) (2024)**	**Proposed methods (QRDSO and WAC-HACK)**
Clustering Accuracy (%)	85%	88%	90%	93%
Computation Time (%)	80%	75%	82%	90%
Silhouette Score (%)	78%	81%	83%	85%

Overall Accuracy (%)	81%	81%	85%	89%

The comparison of HQDCO with standard methods is depicted in Table 2 , LASSO SAE and FMDR, is demonstrated in Clustering accuracy, Calculation time, Silhouette scores, and overall accuracy. As indicated, HQDCO has an 89 percent overall accuracy, indicating its relevance and efficiency in mining tasks.

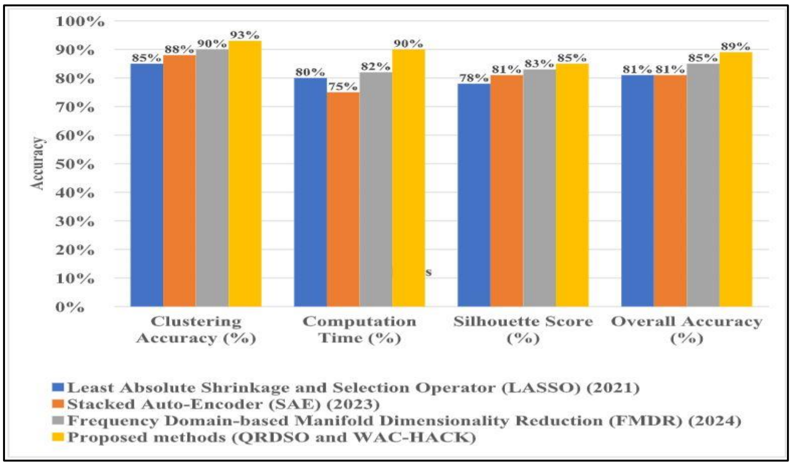

Figure 2. Comparison of HQDCO with Traditional Clustering Methods

Figure 2 depicts the comparison results between HQDCO and other standard methods; LASSO, SAE, and FMDR, primarily based on clustering accuracy, computation time, Silhouette score, and overall accuracy. The results presented in the study demonstrate the effectiveness of the proposed QRDSO-WAC-HACK hybrid optimization framework in addressing challenges related to clustering in high-dimensional and complex datasets. These findings are directly applicable to software testing in the following ways, as the framework enhances clustering accuracy by improving exploratory search and convergence, helping identify patterns, redundancies, and defects in test cases. It reduces computational overhead by combining quantum-driven search with adaptive clustering, enabling faster error detection and test coverage analysis. The study results highlight the framework's ability to improve clustering efficiency and accuracy in software testing, and by integrating QRDSO and WAC-HACK, it provides a scalable and computationally efficient solution for modern testing environments.

5. Conclusion and Future Scope

The combination of QRDSO and WAC-HACK will give an effective hybrid method to improve clustering efficiency in high-dimensional data. QRDSO makes use of the quantum paradigm in optimization towards development and convergence, On the other hand, WAC-HACK adjust clustering to data patterns on its own. These combinations achieve high accuracy while requiring less computational effort, resulting in more efficient processing compared to traditional methods. Applications that require processing large amounts of high-dimensional data are the ones where the system shines, e.g., bioinformatics, image analysis, or text mining). In the future, the work on making it even more scalable and flexible for different data and real-time alike; however, this hybrid model is solving a problem in a very versatile manner. Further research will investigate the potential of the hybrid QRDSO and WAC-HACK framework for real-time clustering scenarios, as well as its application to high-dimensional datasets that are more diverse. High-dimensional, extensive datasets, particularly those frequently found in domains like text mining and image analysis, were used to test the architecture described in the research. These datasets are appropriate for assessing the clustering effectiveness of the suggested QRDSO-WAC-HACK hybrid framework because of their high number of dimensions. These datasets were used to test the method's performance and see how well it clustered high-dimensional data.

References

[1] Rozin, B., Pereira-Ferrero, V. H., Lopes, L. T., & Pedronette, D. C. G. (2021). A rank-based framework through manifold learning for improved clustering tasks. Information Sciences, 580, 202-220.

[2] Li, Y., Wang, X., & An, L. (2023). Hierarchical clustering-based personalized federated learning for robust and fair human activity recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 7(1), 1-38.

[3] Nikhal, Kshitij. "Learning Discriminative and Efficient Attention for Person Re-Identification Using Agglomerative Clustering Frameworks." (2021).

[4] Li, Yunfan, Mouxing Yang, Dezhong Peng, Taihao Li, Jiantao Huang, and Xi Peng. "Twin contrastive learning for online clustering." International Journal of Computer Vision 130, no. 9 (2022): 2205-2221.

[5] Abdalgader, K., Matroud, A. A., & Hossin, K. (2024). Experimental study on short-text clustering using transformer-based semantic similarity measure. PeerJ Computer Science, 10, e2078.

[6] Zhang, W., Dong, C., Yin, J., & Wang, J. (2021). Attentive representation learning with adversarial training for short text clustering. IEEE Transactions on Knowledge and Data Engineering, 34(11), 5196-5210.

[7] Sharma, S., Chande, S. V., & Joshi, N. K. (2024). Leveraging Clustering Algorithms for Optimizing Test Case Prioritization in Software Development. International Journal of Sustainable Development Through AI, ML and IoT, 3(2), 1-9.

[8] Mu, L., Sugumaran, V., & Wang, F. (2020). A hybrid genetic algorithm for software architecture re-modularization. Information Systems Frontiers, 22(5), 1133-1161.

[9] Xia, C., Zhang, Y., & Hui, Z. (2021). Test suite reduction via evolutionary clustering. IEEE Access, 9, 28111-28121.

[10] Merikhi, B., & Soleymani, M. R. (2021). Automatic data clustering framework using nature-inspired binary optimization algorithms. IEEE Access, 9, 93703-93722.

[11] Raamesh, L., Jothi, S., & Radhika, S. (2023). Enhancing software reliability and fault detection using hybrid brainstorm optimization-based LSTM model. IETE Journal of Research, 69(12), 8789-8803.

[12] Abbasi, M., Shokrollahi, A., Khosravi, M. R., & Menon, V. G. (2020). High-performance flow classification using hybrid clusters in software defined mobile edge computing. Computer Communications, 160, 643-660.

[13] Dondapati, Koteswararao. "Robust Software Testing for Distributed Systems Using Cloud Infrastructure, Automated Fault Injection, and XML Scenarios." International Journal of Information Technology and Computer Engineering 8, no. 2 (2020): 75-94.

[14] Gattupalli, Kalyan. "A Survey on Cloud Adoption for Software Testing: Integrating Empirical Data with Fuzzy Multicriteria Decision-Making." International Journal of Information Technology and Computer Engineering 10, no. 4 (2022): 126-144.

[15] Sreekar Peddi. (2020). Cost-effective Cloud-Based Big Data Mining with K-means Clustering: An Analysis of Gaussian Data. International Journal of Engineering & Science Research, 10(1), 2020. 229-240.

[16] Mamidala, Vijaykumar. "Optimizing Performance with Parallel K-Means in Tunnel Monitoring Data Clustering Algorithm for Cloud Computing." International Journal of Engineering Research and Science & Technology 17, no. 4 (2021): 34-49.

[17] Parthasarathy, Karthikeyan. "Enhanced Case-Based Reasoning with Hybrid Clustering and Evolutionary Algorithms for Multi-Class Workload Forecasting In Autonomic Database Systems." International Journal of HRM and Organizational Behavior 11, no. 2 (2023): 38-54.

[18] Devi, Durga Praveen. "Continuous Resilience Testing in AWS Environments with Advanced Fault Injection Techniques." International Journal of Information Technology and Computer Engineering 11, no. 1 (2023): 199-217.

Author's biography

Kalyan Gattupalli, A Masters in Computer Applications from University of Madras in 2007 have diversified experience in the world of Cloud and CRM including but not limited to Azure Cloud, Micrsoft Dynamcis CE, Salesforce CRM. He is working as a computer systems analyst with some federal departments in Canada and with financial institutions over the past decade. He at present work as a Independent contractor providing services in the name of Yash Tek Inc and Sunny Information Technology Services Inc.

Haris M. Khalid (SMIEEE, FHEA) is currently an Assistant Professor with the College of Engineering and Information Technology, University of Dubai (UD), Academic City, Dubai, UAE, since 2023. He is also a Visiting Senior Researcher with University of Johannesburg (UJ), South Africa since 2022. From 2019–2022, He was an Energy Specialist Consultant to UAE Space Agency (UAESA) and Nano-Racks for “Tests in Orbits” themed projects to space. From 2016–2022, he was an Assistant Professor with the Electrical and Electronics Engineering Department, Higher Colleges of Technology (HCT) – Sharjah Campuses. He has also worked as a postdoctoral researcher and visiting scholar with Masdar Institute and Petroleum Institute from 2013–2016 and 2015–2016 respectively. He has won research grants of 1.7+ million USD towards research, development, and start-up ventures. His research interests include power systems, cyber-attacks, renewable energy integration, energy storage systems, electric vehicles, transportation electrification, signal processing, machine learning and AI, IoT, and fault diagnostics. Dr. Khalid is the Associate and Topic Editor of Frontiers in Energy Research | Smart Grid.