International Journal of Computer Applications Technology and Research
Volume 13–Issue 11, 49 – 53, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1311.1008

Efficient Data Recognition and Classification in IoT Ecosystem using Optimized K-Means Algorithm and Hybrid Deep Learning Model

*Ihediuche Evangeline Ndidi
*Dennis Memorial Grammar school Onitsha
Anambra State

#Ike Mgbeafulike
#Department of Computer Science, Chukwuemeka
Odumegwu Ojukwu University, Uli AN, NG

Abstract: With the exponential increase in IoT devices, there is a growing demand for efficient threat recognition and classification to handle massive data generated in Cloud-IoT environments. This research introduces a hybrid machine learning framework that combines an optimized K-Means clustering algorithm with a machine learning model to enhance data processing in IoT systems. The optimized K-Means algorithm facilitates initial data grouping, significantly reducing computational complexity and improving clustering accuracy. Subsequently, a hybrid model integrates Convolutional Neural Network (CNN) for feature extraction and Support Vector Machine (SVM) for precise classification, effectively handling the diverse and high-dimensional data in Cloud-IoT systems. Experimental results show that the proposed method achieves superior accuracy and processing efficiency compared to conventional approaches, making it a robust solution for scalable IoT data management.

Keywords: Internet of Things (IoT), Cloud Computing, K-Means Clustering, Machine Learning, Data Recognition, Data Classification.

INTRODUCTION

The Internet of Things (IoT) is transforming a wide range of industries by connecting devices to the internet, enabling seamless interaction between physical objects and digital systems. With IoT's expansion, the number of connected devices is projected to reach 29.3 billion by 2023, generating vast amounts of data that pose significant challenges for data management, storage, and processing within IoT environments. This exponential data growth demands efficient solutions for data recognition and classification, which are essential for timely and accurate IoT operations. The IoT market is projected to grow significantly, with estimates suggesting that by 2030, there could be over 30 billion connected devices worldwide (Statista, 2021).

To address these issues, cloud computing has become integral to IoT, offering scalable data storage, high processing power, and on-demand computational resources. Major cloud platforms, such as AWS, Microsoft Azure, and Google Cloud, provide Infrastructure as a Service (IaaS), Software as a Service (SaaS), and Platform as a Service (PaaS) solutions, making cloud resources accessible and economical for IoT applications. Figure 1 illustrates the diverse applications of cloud-integrated IoT in fields like healthcare, smart cities, agriculture, and transportation. Given these vulnerabilities, it is crucial to develop security frameworks specifically designed for IoT environments. Existing cybersecurity approaches are often ill-suited for the unique challenges of IoT. A hybrid security framework combines various strategies to provide a more comprehensive defence. Key components of this approach include layered security, real-time analysis, and adaptability. Layered security involves implementing multiple layers of protection to defend against a variety of threats, ensuring a more comprehensive security framework (NIST, 2018). Real-time analysis leverages machine learning algorithms to continuously monitor and analyze data, enabling the system to detect anomalies and respond to potential threats as they arise (Scully et al., 2018). Adaptability refers to the system's ability to evolve in response to new and emerging threats, adjusting to the diverse landscape of IoT devices. By integrating machine learning techniques, such as K-means clustering, this approach enhances the framework's ability to detect and respond to threats dynamically, addressing current vulnerabilities while also preparing for future challenges.

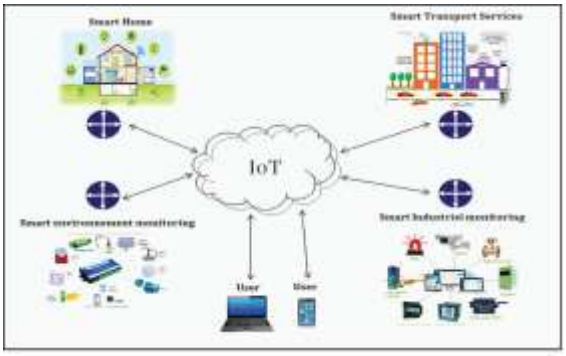

Figure 1: A generic IoT network architecture

Despite these advancements, data recognition in IoT systems remains challenging, particularly in cloud environments where high data transfer rates can complicate real-time processing and analysis. This study leverages Convolutional Neural Networks (CNN) and machine learning techniques to enhance IoT data recognition and classification, creating a robust and scalable solution for cloud-based IoT systems. Common security challenges in the realm of IoT include data breaches, malware attacks, privacy concerns, and denial of service (DoS) attacks. Data breaches occur when unauthorized access is gained to sensitive information, often due to weak authentication protocols

(Weber, 2010). Another significant threat is malware attacks, where IoT devices can be hijacked and used to form botnets, which can then launch Distributed Denial of Service (DDoS) attacks (Symantec, 2019). Additionally, the privacy concerns associated with IoT are substantial, as these devices often collect vast amounts of personal data, which can lead to regulatory scrutiny and privacy violations (Zhou et al., 2019). Finally, Denial of Service (DoS) attacks can overwhelm IoT devices by flooding them with excessive traffic, causing the devices to become inoperable and disrupting the services they provide (Miorandi et al., 2012).

Security in IoT environments is inherently complex due to the heterogeneity of connected devices and their resource constraints (Rukmony & Gnanamony 2023). Traditional security mechanisms fall short in addressing the dynamic nature of IoT ecosystems, prompting the need for a theoretical foundation that integrates multiple security paradigms.

REVIEW OF RELATED WORKS

The Internet of Things (IoT) has become a transformative force, facilitating data sharing and enhancing various applications across fields such as smart grids, healthcare, agriculture, and autonomous vehicles (Cai et al., 2017; Celesti et al., 2018; Yang et al., 2018). IoT enables real-time interaction between physical and virtual entities, which can occur directly or through intermediaries like cloud computing (Chaudhry et al., 2020). However, the rapid growth in the number of IoT devices brings challenges related to storage, processing capacity, and security, necessitating integration with cloud computing for efficient data management (Haghi et al., 2020; Jiang et al., 2019).

To address these challenges, the concept of the "Cloud of Things" (CoT) has emerged, merging IoT with cloud services to enhance scalability, flexibility, and security (Manoharan et al., 2023). CoT benefits from advanced encryption methods, such as ciphertext policy attribute-based encryption, which secures data transmission and helps maintain data integrity across IoT systems (Samanta et al., 2021). The use of deep learning models, particularly Convolutional Neural Networks (CNNs), has been explored to improve data recognition and anomaly detection in IoT applications, providing robust solutions for cybersecurity and monitoring (Peng et al., 2019; Ullah & Mahmoud, 2021).

CNNs, which are widely used in image processing and object recognition, offer substantial benefits in applications requiring high accuracy, such as autonomous driving and obstacle detection (Li et al., 2020). However, traditional CNN models are computationally intensive, making them challenging to deploy on IoT devices with limited resources. To address this issue, hybrid models combining CNN with Support Vector Machine (SVM) classifiers have been developed. These models use CNN as a feature extractor while relying on SVM for classification, achieving a balance between accuracy and computational efficiency (Ahmed et al., 2020). Such hybrid CNN-SVM models are particularly suited for resource-constrained IoT environments, as they reduce computational demands without compromising on classification accuracy (Ferrag et al., 2021; Ray et al., 2019).

In addition to hybrid models, optimized clustering algorithms like K-Means have shown promise in organizing high-dimensional IoT data into manageable clusters. Enhanced K-Means algorithms improve cluster center selection, thus reducing computational complexity and achieving higher clustering precision, which is crucial for applications involving large volumes of data, such as smart city infrastructure and healthcare monitoring (Manoharan et al., 2021).

This study builds upon these foundational works by proposing a novel framework that integrates optimized K-Means clustering with a CNN-SVM hybrid model. The objective is to enhance data recognition and classification within Cloud-IoT systems, addressing both accuracy and efficiency challenges associated with traditional data processing techniques in large-scale IoT applications.

PROPOSED METHOD

This study presents a robust two-phase framework to address the challenges of data recognition and classification in Cloud-IoT systems, where both accuracy and computational efficiency are critical. The proposed framework combines an optimized K-Means clustering algorithm for data organization and a hybrid deep learning model integrating Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for classification. This method capitalizes on the strengths of each algorithm to deliver a scalable and effective solution for IoT data processing.

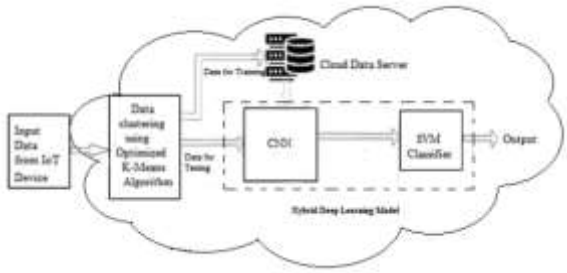

Figure 2: the workflow of the proposed model

A. Data Clustering Phase

The initial phase of this framework employs an optimized K-Means clustering algorithm, which plays a crucial role in reducing data dimensionality and computational load before classification. Traditional K-Means clustering relies on randomly selecting initial cluster centers, which can lead to inconsistencies in cluster quality and higher computational costs due to the increased number of iterations required to reach convergence. In response, this study enhances the K-Means clustering algorithm by using an optimization technique that strategically selects initial centroids based on data distribution patterns.

By employing this optimized approach, the clustering algorithm achieves faster convergence with fewer iterations, effectively reducing processing time and resource usage. This optimized K-Means clustering not only organizes data into more accurate clusters but also significantly improves the relevance of features fed into the CNN-SVM model in the next phase. Furthermore, this clustering phase groups data points based on their similarity, making it easier for the deep learning model to distinguish

between different data classes, ultimately enhancing classification accuracy.

B. Data Training and Classification Phase

Following clustering, the data undergoes classification through a hybrid CNN-SVM model, which leverages the advantages of CNN's feature extraction capabilities and SVM's strong classification boundaries. CNNs are well-regarded for their ability to automatically learn hierarchical features from high-dimensional data, making them ideal for IoT applications where data variability is high. The CNN component processes the clustered data to identify essential features, such as textures, shapes, and patterns, which are crucial for distinguishing between classes. This study's CNN architecture includes several convolutional layers, pooling layers, and activation functions, designed to capture complex features while maintaining computational efficiency.

To enhance classification performance, the SoftMax layer typically used in CNNs is replaced by an SVM classifier. Unlike SoftMax, which tends to perform best on linearly separable classes, SVM provides more robust decision boundaries and is well-suited for handling complex and overlapping data distributions. The SVM classifier in this model uses the features extracted by CNN to categorize data into distinct classes, achieving higher precision, especially in binary and multi-class classification tasks. This CNN-SVM hybrid model is optimized for IoT data, balancing the high accuracy of deep learning with the efficiency needed for IoT environments.

C. Implementation and Deployment Considerations

For practical deployment in Cloud-IoT systems, this framework is designed to be scalable and compatible with cloud infrastructures. The data clustering and classification phases are implemented as independent modules, enabling parallel processing to reduce latency and accommodate large volumes of IoT data. This modular design also allows the framework to be easily integrated into existing cloud-based IoT architectures, where data preprocessing, feature extraction, and classification can be distributed across cloud resources. Additionally, the framework's reliance on optimized algorithms ensures that it can operate within the constraints of typical IoT devices and networks, minimizing the computational burden on individual devices.

By combining optimized K-Means clustering with CNN-SVM hybrid classification, this framework addresses the core challenges of accuracy, computational efficiency, and scalability in Cloud-IoT data recognition. This approach not only improves classification accuracy but also reduces processing time, making it a practical solution for real-time IoT applications such as smart city infrastructure, healthcare monitoring, and autonomous vehicle navigation. The proposed method demonstrates significant advantages over traditional CNN and SVM models, providing an effective balance between precision and efficiency, essential for large-scale IoT deployments.

RESULT AND DISCUSSION

The proposed model's results highlight its improved efficiency in data recognition and classification in a Cloud-IoT system using an optimized K-Means clustering algorithm and a hybrid CNN-SVM model. The following key metrics were evaluated:

Clustering Convergence Rate: The optimized K-Means clustering algorithm demonstrated superior convergence compared to traditional K-Means and fuzzy c-means algorithms. Convergence was achieved in fewer iterations, as shown in Figure 4, due to the resemblance-based clustering approach. While traditional methods required additional iterations, the optimized approach converged rapidly, indicating reduced computational time and resources.

Clustering Accuracy: As seen in Figure 5, the optimized K-Means algorithm achieved higher clustering accuracy than traditional clustering algorithms. By using zero-mean normalization and optimized cluster center calculations, the proposed algorithm minimized cluster overlap and improved accuracy in grouping IoT data. This accuracy enhancement is critical for IoT applications, where precise clustering can lead to more reliable data processing and better decision-making.

Computation Time: The optimized K-Means clustering algorithm reduced computation time significantly compared to conventional methods (Figure 6). This efficiency stems from the reduced iteration count and optimized calculations, making the model suitable for real-time IoT data processing, where rapid clustering is essential to support timely data recognition and action.

Error Rate: The proposed clustering approach achieved a lower error rate, as shown in Figure 7, compared to traditional K-Means and fuzzy c-means algorithms. Lower error rates indicate that the clustering algorithm could accurately group data with minimal misclassification, thereby enhancing the reliability of subsequent data analysis and classification stages.

Hybrid CNN-SVM Model Performance: The hybrid CNN-SVM model showed superior performance in comparison to other hybrid models, including CNN-KNN and CNN-GA, in terms of running time, data recognition rate, and classification accuracy.

Running Time: Figure 8 shows that the CNN-SVM model required less running time than other hybrid models. This efficiency is attributed to CNN's effective feature extraction, which reduced the computational load on the SVM classifier. As a result, the model is well-suited for environments where processing speed is critical.

Recognition Rate: The proposed CNN-SVM model achieved a high recognition rate (Figure 9). By integrating CNN's feature extraction capabilities with the SVM classifier's robust separation ability, the model accurately recognized IoT data patterns. This recognition efficiency is essential for identifying and categorizing data in complex IoT systems, particularly in real-time applications.

Classification Accuracy: As demonstrated in Figure 10, the hybrid CNN-SVM model achieved higher classification accuracy than other hybrid models. This improvement indicates that the model can effectively classify IoT data into distinct categories,

supporting more accurate decision-making processes in Cloud-IoT systems.

Effectiveness of SVM Verdict Functions: The model was evaluated under two SVM verdict functions: one-versus-one and one-versus-rest. As shown in Table 1, the one-versus-one function achieved slightly higher accuracy than one-versus-rest in both training and testing phases, suggesting that the choice of SVM configuration can further optimize classification performance.

Comparative Analysis and Overall Efficiency

The comparative analysis presented in Figure 11 confirms that the proposed model's efficiency surpasses that of other hybrid approaches. This is due to the synergy between optimized K-Means clustering and the hybrid CNN-SVM architecture. The optimized K-Means clustering reduces the computational load on the CNN-SVM model by pre-processing and structuring data clusters, while CNN-SVM improves recognition and classification accuracy.

The results demonstrate that integrating an optimized K-Means clustering algorithm with a CNN-SVM hybrid model leads to substantial improvements in computational efficiency, accuracy, and error reduction. This combination provides a robust solution for data recognition and classification in large-scale Cloud-IoT systems. The proposed model not only accelerates data processing but also enhances the accuracy of IoT data interpretation, supporting various real-time applications that rely on rapid and accurate data analysis.

REFERENCES

Chaudhry, S. A., Yahya, K., Al-Turjman, F., & Yang, M.-H. (2020). A secure and reliable device access control scheme for IoT-based sensor cloud systems. IEEE Access, 8, 139244–139254. https://doi.org/10.1109/ACCESS.2020.3012755

Venkannaudutalapally, S. P., Mohanty, S. P., Pallagani, V., & Khandelwal, V. (2020). sCrop: A novel device for sustainable automatic disease prediction, crop selection, and irrigation in Internet-of-Agro-Things for smart agriculture. IEEE Sensors Journal, 21(16), 17525–17538. https://doi.org/10.1109/JSEN.2020.2985545

Cai, H., Xu, B., Jiang, L., & Vasilakos, A. V. (2017). IoT-based big data storage systems in cloud computing: Perspectives and challenges. IEEE Internet of Things Journal, 4(1), 75–87. https://doi.org/10.1109/JIOT.2016.2619367

Celesti, A., Galletta, A., Carnevale, L., Fazio, M., Lay-Ekuakille, A., & Villari, M. (2018). An IoT cloud system for traffic monitoring and vehicular accidents prevention based on mobile sensor data processing. IEEE Sensors Journal, 18(12), 4795–4802. https://doi.org/10.1109/JSEN.2018.2829865

Chen, H., Zhang, Y., Cao, Y., & Xie, J. (2021). Security issues and defensive approaches in deep learning frameworks. Tsinghua Science and Technology, 26(6), 894–905. https://doi.org/10.26599/TST.2020.9010125

Ferrag, M. A., Friha, O., Maglaras, L., Janicke, H., & Shu, L. (2021). Federated deep learning for cybersecurity in the Internet of Things: Concepts, applications and experimental analysis. IEEE Access, 9, 138509–138542. https://doi.org/10.1109/ACCESS.2021.3117035

Peng, Y., Liao, M., Song, Y., Liu, Z., He, H., Deng, H., & Wang, Y. (2019). FB-CNN: Feature fusion-based bilinear CNN for classification of fruit fly image. IEEE Access, 8, 3987–3995. https://doi.org/10.1109/ACCESS.2019.2961099

Manoharan, J. S., et al. (2021). A hybrid approach to accelerate the classification accuracy of cervical cancer data with class imbalance problems. International Journal of Data Mining, 25(3/4), 234–259. https://doi.org/10.1504/IJDMB.2021.10040853

Ullah, I., & Mahmoud, Q. H. (2021). Design and development of a deep learning-based model for anomaly detection in IoT networks. IEEE Access, 9, 103906–103926. https://doi.org/10.1109/ACCESS.2021.3099885

Singh, J., Pasquier, T., Bacon, J., Ko, H., & Eyers, D. (2016). Twenty security considerations for cloud-supported Internet of Things. IEEE Internet of Things Journal, 3(3), 269–284. https://doi.org/10.1109/JIOT.2015.2505902

Chi, F., Wang, X., Cai, W., & Leung, V. C. M. (2015). Ad-hoc cloudlet based cooperative cloud gaming. IEEE Transactions on Cloud Computing, 6(3), 625–639. https://doi.org/10.1109/TCC.2015.2415796

Jiang, W., Wang, Y., Jiang, Y., Chen, J., Xu, Y., & Tan, L. (2019). Research on mobile Internet mobile agent system dynamic trust model for cloud computing. China Communications, 16(7), 174–194. https://doi.org/10.23919/JCC.2019.07.015

Yang, G., Jiang, M., Ouyang, W., Ji, G., Xie, H., Rahmani, A. M., Liljeberg, P., & Tenhunen, H. (2018). IoT-based remote pain monitoring system: From device to cloud platform. IEEE Journal of Biomedical and Health Informatics, 22(6), 1711–1719. https://doi.org/10.1109/JBHI.2018.2807212

Haghi, M., Neubert, S., Geissler, A., Fleischer, H., Stoll, N., Stoll, R., & Thurow, K. (2020). A flexible and pervasive IoT-based healthcare platform for physiological and environmental parameters monitoring. IEEE Internet of Things Journal, 7(6), 5628–5647. https://doi.org/10.1109/JIOT.2019.2953635

Li, G., Xie, H., Yan, W., Chang, Y., & Qu, X. (2020). Detection of road objects with small appearance in images for autonomous driving in various traffic situations using a

deep learning-based approach. IEEE Access, 8, 211164–211172. https://doi.org/10.1109/ACCESS.2020.3037802

Guan, Z., Li, J., Wu, L., Zhang, Y., Wu, J., & Du, X. (2017). Achieving efficient and secure data acquisition for cloud-supported Internet of Things in smart grid. IEEE Internet of Things Journal, 4(6), 1934–1944. https://doi.org/10.1109/JIOT.2017.2754912

Ray, P. P., Thapa, N., & Dash, D. (2019). Implementation and performance analysis of interoperable and heterogeneous IoT-edge gateway for pervasive wellness care. IEEE Transactions on Consumer Electronics, 65(4), 464–473. https://doi.org/10.1109/TCE.2019.2953756

Manoharan, J. S., et al. (2023). A comprehensive study and review of tuning the performance on database scalability in big data analytics. Journal of Intelligent and Fuzzy Systems, 44, 5231–5255. https://doi.org/10.3233/JIFS-189999

Guansheng, W., Liu, X., Zheng, X., & Li, J. (2018). IF-CNN: Image-aware inference framework for CNN with the collaboration of mobile devices and cloud. IEEE Access, 6, 68621–68633. https://doi.org/10.1109/ACCESS.2018.2879784

Ahmed, I., Din, S., Jeon, G., Piccialli, F., & Fortino, G. (2020). Towards collaborative robotics in top view surveillance: A framework for multiple object tracking by detection using deep learning. IEEE/CAA Journal of Automatica Sinica, 8(7), 1253–1270. https://doi.org/10.1109/JAS.2020.1003354

Esposito, C., Castiglione, A., Frattini, F., Cinque, M., Yang, Y., & Choo, K.-K. R. (2018). On data sovereignty in cloud-based computation offloading for smart cities applications. IEEE Internet of Things Journal, 6(3), 4521–4535. https://doi.org/10.1109/JIOT.2018.2882341

Samanta, D., Alahmadi, A. H., P, K. M., Khan, M. Z., Banerjee, A., Dalapati, G. K., & Ramakrishna, S. (2021). Cipher block chaining support vector machine for secured decentralized cloud-enabled intelligent IoT architecture. IEEE Access, 9, 98013–98025. https://doi.org/10.1109/ACCESS.2021.3108619