The Process of Identifying Outliers within Continuous Human Activity Data Streams

Prabhu Patel

Institution of Electronics and Telecommunication Engineers (IETE), New Delhi

ABSTRACT

Technological advancement progresses rapidly while the quantity of data generated from multiple sources especially those linked to human activities expands creating a necessity for strong frameworks to conduct stream data analysis and identify outliers. This article initiates a detailed exploration into the role these analytical frameworks play in identifying statistically significant outliers that deviate from expected patterns within extensive datasets. The numerous traditional methods for detecting outliers encounter major difficulties when processing streaming data because of its vast volume and rapid speed which leads to computational inefficiencies that make these methods unsustainable for real-time applications [1]. The realm of stream analytics addresses these challenges through its capability to process data points as they arrive which leads to enhanced decision-making power by obtaining immediate insights [2]. The pursuit of detecting anomalies within human activity datasets represents a critical endeavor because these anomalies possess potential implications that extend across multiple applications such as health monitoring systems and security networks. The present study examines contemporary challenges such as excessive memory usage alongside dimensionality issues which complicate outlier detection within large data streams [3]. This article embarks on an exploration into how contemporary methodologies and frameworks crafted for real-time analysis empower stream data processing advancements to reach heightened precision in detecting human behavioral anomalies. The progression of these technological advancements serves to deepen our understanding of normal activity patterns while at the same time laying down essential initial steps for developing proactive responses to detected anomalies [4]. Our detailed investigation into specialized frameworks and techniques for outlier detection reveals the necessity for ongoing research to address emerging challenges while improving stream data analytics performance in the rapidly evolving digital landscape.

Keywords: stream data big data human activity recognition outlier detection.

INTRODUCTION

The rapidly increasing volume of data produced by IoT devices and digital platforms necessitates the development of novel stream data analysis techniques to accurately identify human activities. Tugrul's 2023 study indicates that traditional outlier detection methods encounter major challenges when handling the vast size and quick velocity of modern datasets which renders them ineffective for real-time applications. The exceptional power of stream analytics emerges through its real-time data point processing capability which supports timely decision-making while enhancing the detection of statistically significant anomalies that diverge from expected patterns. This article undertakes a detailed examination of specialized frameworks designed to detect outliers in human activities by analyzing their methods and effectiveness in identifying discrepancies that warrant further investigation [2]. The document places additional focus on the vast array of challenges faced during real-time data processing which encompasses both extreme memory consumption and complex issues that arise from dealing with high-dimensional datasets [6]. The examination of human activities via wearable technology combined with smart environmental systems has become more intense

making it crucial to understand these frameworks to improve user experience while ensuring safety and security across different applications. This article embarks on a thorough investigation of thematic elements to elucidate current stream data analysis methodologies that identify human activity outliers while addressing the complex challenges professionals face in dynamic environments [7].

A COMPREHENSIVE EXAMINATION OF TECHNIQUES FOR STREAM DATA ANALYSIS

Techniques for stream data analysis serve as fundamental tools to manage and interpret the vast amounts of information generated in real-time settings which become particularly crucial for the recognition of human activities. According to Tugrul's 2023 study traditional data processing methods exhibit major deficiencies since they demand vast computational resources while proving ineffective at handling high-velocity data streams. Stream analytics assigns its operational focus to managing data as it arrives which permits the creation of immediate insights while supporting timely decision-making. The methods in question typically utilize specially designed algorithms for incremental learning which allow systems to adapt to new data inputs without requiring a complete reassessment of the entire dataset. A vast array of analytical methodologies for stream data examination exists which encompasses statistical approaches designed to identify outliers through the analysis of deviations from established dataset patterns. Manolova's 2025 study demonstrates how the combined use of clustering methods with density estimation techniques functions as fundamental tools to identify normal behavior patterns while distinguishing them from anomalous activities. An increasing quantity of machine learning models—especially those utilizing deep learning methods—are being implemented to enhance outlier detection systems by leveraging both historical data patterns and real-time input information together [11]. The models make frequent use of sliding window mechanisms to maintain relevance while simultaneously addressing resource limitations related to both memory usage and computational demand. Rezk's 2024 documentation indicates that ongoing challenges with dimensionality reduction techniques persist together with the curse of dimensionality which stands as a basic problem in high-dimensional datasets typical of human activity monitoring systems. Stream data analysis demands the intricate preservation of a fragile equilibrium between accuracy and processing velocity to maintain robust and reliable outlier detection systems as new data entries flow into the system [10]. The continuous development of specialized research frameworks for specific applications necessitates a deep understanding of foundational techniques to advance human activity recognition through real-time analytics. Fig -1
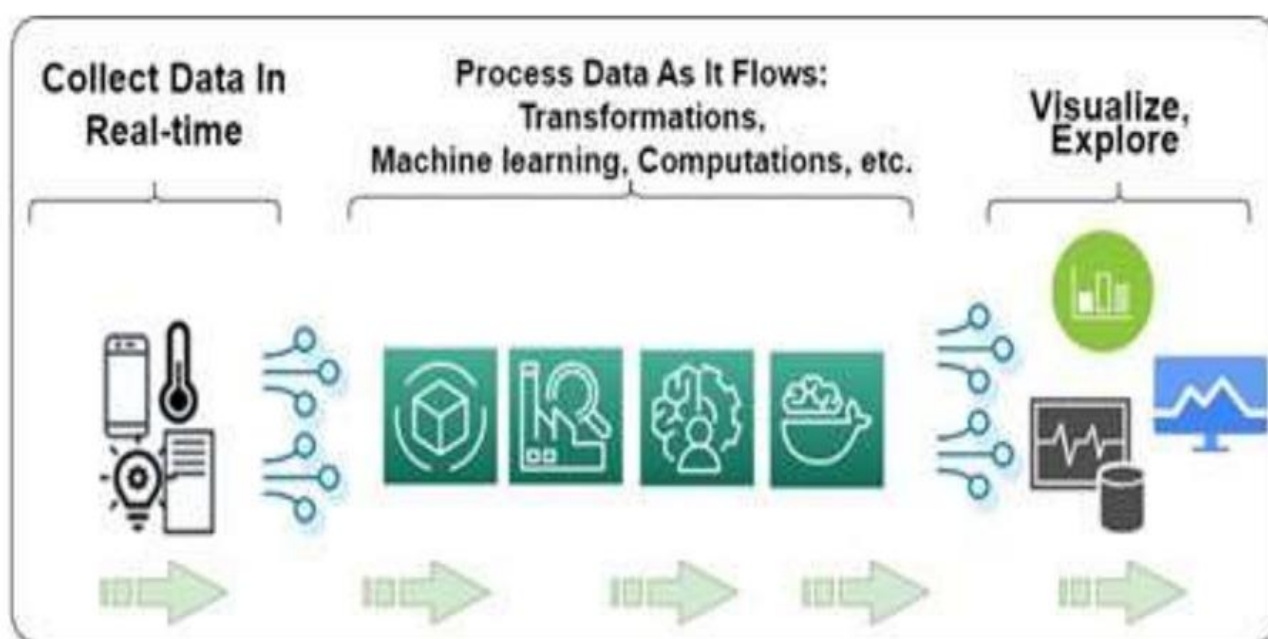

Fig -1

Algorithmic Explanations

The detection of outliers in data streams necessitates the use of algorithms which possess the capability to process information in real-time while simultaneously adjusting to emerging patterns. The following sections present detailed analyses of two fundamental techniques Incremental LOF and Adaptive Random Cut Forest through explicit explanations and accompanying pseudocode which demonstrate their practical applications.

The Incremental Local Outlier Factor (I-LOF)

The I-LOF method detects outliers through a process that involves measuring the density of a new data point in relation to its neighboring points. A point whose density measurement is markedly lower than that of its surrounding area gets identified as an anomaly.

The operational mechanism of this system functions in the following manner:

The sliding window technique considers exclusively the latest data points within a predetermined range which helps to decrease memory consumption.

Upon the arrival of a new point, the algorithm initiates a complex process to refresh the list of nearest neighbors situated within its defined window.

The process of density calculation involves determining the degree to which a point stands apart from nearby points by evaluating its local reachability density.

The LOF score when it reaches high values indicates that the data point is considered an outlier.

The reason for its significance can be understood through examination of various factors that contribute to its importance.

The system possesses an efficient mechanism where it recalculates scores solely for the affected points instead of processing the entire dataset.

Adaptability: The system functions effectively when faced with unexpected anomalies such as a fall detected by wearable devices without prior warning.

The Adaptive Random Cut Forest (ARCF) technique

ARCF represents an ensemble technique that constructs multiple compact decision trees to identify outlier data points. The system demonstrates adaptability to slow modifications in user activity patterns such as the progressive reduction in walking speed that occurs due to aging processes.

The operational mechanism:

1. The process of forest construction involves the training of numerous tree models on various randomly selected data subsets.
2. The concept of Path Length Scoring identifies anomalies as data points which manifest in shorter tree paths because they are more readily isolated.
3. Drift Detection: Whenever shifts in data distribution occur due to emerging behavior patterns, the model responds by resetting certain trees.

The significance of this issue emerges from the fact that it plays a crucial role in many aspects

- Concept Drift Handling: This mechanism adapts to persistent alterations such as the gradual deterioration of a patient's mobility capabilities.

- Scalability: The ability of the system to handle increasing workloads is demonstrated by its fast performance on large data streams through parallel tree updates.

An Exhaustive Comparative Examination of Various Outlier Detection Techniques Applied to Human Activity Datasets

Selecting an appropriate outlier detection technique for human activity recognition (HAR) requires an examination of various methods through criteria such as accuracy, processing speed, adaptability to different conditions, and practical applicability in real-world scenarios. In the following examination we delve into four fundamental techniques by applying them to authentic HAR datasets such as UCI HAR and MobiAct while simultaneously emphasizing their respective strengths and weaknesses.

Method	Type	Best For
Incremental LOF (I-LOF)	Density-based	Sudden anomalies (e.g., falls)
Adaptive Random Cut Forest (ARCF)	Tree-based	Evolving behavior patterns
Deep Autoencoders	Neural Networks	Complex, high-dimensional data
Sliding Window Z-Score	Statistical	Fast, low-resource environments

DIVERSE METHODOLOGICAL STRUCTURES FOR DETECTING ANOMALOUS ACTIVITY PATTERNS

Frameworks engineered to identify outlier activities become essential tools when addressing the multifaceted complexities present in streaming data generated from human behaviors. Tugrul's 2023 study identifies numerous deficiencies in traditional methods due to their excessive computational resource demands and their ineffectiveness in processing the fast data streams required by real-time applications. Manolova's 2025 study indicates that stream analytics frameworks dedicate their efforts to achieving operational efficiency by processing data points immediately which allows for the quick identification of statistically significant anomalies that deviate from established behavioral patterns. This field has seen the emergence of numerous methodologies which include statistical techniques for time-series analysis to detect interval-based changes as well as machine learning models that evolve by learning from incoming data[8]. The domain of deep learning methodologies has gained widespread acknowledgment as an effective technique for outlier detection. According to Silvia 2022, deep neural networks (DNNs) exhibit remarkable skill in detecting complex patterns across large datasets while maintaining high levels of accuracy and recall. The examined frameworks employ sliding window techniques to scrutinize continuous activity data streams which enable them to adjust their operations dynamically upon the availability of new data. The integration of explainable AI technology into these systems has emerged as a critical element because the ability to comprehend the rationale behind outlier detection enhances user trust while enabling more detailed explorations into identified anomalies (Custom)[9]. The development of robust systems for identifying anomalies in human activities requires a multifaceted approach that combines real-time processing capabilities with advanced analytical techniques. The intricate exploration of efficiency and interpretability challenges within streaming data environments enables these frameworks to deliver significant progress in decision-making techniques across various human activity recognition applications.

THE NUMEROUS DIFFICULTIES ASSOCIATED WITH REAL-TIME DATA PROCESSING SYSTEMS:

The realm of real-time data processing faces numerous challenges that deeply impact the effectiveness of stream data analysis frameworks as they work to detect anomalies in human activity patterns. The vast number of incoming data streams combined with their swift arrival presents a core difficulty because these elements consistently exceed the capabilities of traditional analytical methods [1]. The relentless production of data from numerous sources such as IoT devices and sensors creates escalating challenges in achieving

low latency performance together with maintaining detection accuracy. The curse of dimensionality introduces complications into outlier detection processes because when datasets expand across more dimensions it becomes increasingly difficult to discern meaningful patterns and anomalies without incurring significant computational expenses [2]. The management of memory resources encounters substantial challenges during the execution of real-time processing operations. The successful management of vast streaming data volumes for storage and retrieval requires advanced memory techniques to avoid bottlenecks which might slow down analytical processes according to Rezk 2024. The concept of scalability stands as an essential requirement since frameworks demand design principles which allow them to adapt dynamically to fluctuating data loads while sustaining both performance levels and accuracy standards. The endeavor to control inherent noise together with variability found within human activity data emerges as a major difficulty to overcome [5]. The identification and control of outliers in real-time analytics systems constitute an essential requirement because these anomalies can skew analytical results if they remain unmanaged. The current situation necessitates the creation of intricate algorithms capable of distinguishing genuine anomalies from non-threatening behavioral variations according to Silvia's 2022 study. Researchers who create effective outlier detection frameworks for dynamic environments still face the essential challenge of achieving a balance between processing speed and detection accuracy. Real-time stream data analysis systems encounter numerous challenges which necessitate the creation of innovative techniques to enhance both operational performance and system reliability.

HUMAN ACTIVITY RECOGNITION:

The fields of healthcare technology and smart home systems stand to gain tremendous advantages from the implementation of human activity recognition systems. The accomplishment of this goal has involved the deployment of numerous sensors including wearable LED lights, cameras, and smartphones. Fig-2 , A significant number of Human Activity Recognition systems depend on accelerometers to identify and distinguish common daily movements such as standing walking sitting jogging and lying down. An investigation utilized accelerometer data collected from thirty individuals engaged in everyday activities to identify patterns that could aid in preventing falls among older adults within smart environments. Previous scientific investigations have demonstrated that wearable sensor technology enhances activity recognition accuracy to a significant degree. A number of initiatives have directed their attention toward improving current HAR systems through the application of more sophisticated data collection methods alongside advanced modeling techniques. A particular research investigation developed a bespoke human activity recognition system using a multi-sensor device which collected data from 28 volunteers representing various ages genders weights and heights. Subsequent to data collection processes, researchers employed the gathered information to construct complex hybrid models aimed at distinguishing various activities. A different research investigation gathered acceleration data through wearable devices and extracted twenty efficient features to classify human activities which resulted in a recognition accuracy rate of 94%. A distinct research initiative employed three wearable accelerometers to record lower body movements in a group of 10 patients. In a separate investigation, scientists successfully identified eight distinct activities—running, walking, standing, sit-ups, vacuuming, brushing teeth, and stair movements—through data collected by a single triaxial accelerometer positioned at the pelvis. The multifaceted capabilities of smartphones including their accelerometers gyroscopes and wireless functions enable them to serve as practical substitutes for dedicated wearable sensors. The ubiquitous nature of smartphone usage creates an exceptionally convenient yet robust platform for implementing HAR applications. A particular research investigation suggested a digital approach to human activity classification that utilizes statistical features extracted from time-series data while eliminating the need for individualized training sessions. Within the scope of another research project scientists integrated smartwatch and smartphone data to successfully distinguish thirteen separate daily activities. A multitude of research studies demonstrate that an individual accelerometer device achieves precise human activity recognition when positioned on the right waist. A technique involving a sliding window approach was employed to partition signals into segments which enabled the detection of physical movements. The absence of a standardized technique for sensor signal interpretation in HAR leads to the utilization of various algorithmic methods. The standard procedure requires examining sensor signals to

identify handcrafted features which represent statistical descriptors of the raw data and these features are subsequently input into machine learning algorithms for classification purposes. The handcrafted features commonly encompass statistical measures such as standard deviation mean duration between peaks and binned distribution. A diverse array of classifiers can be developed through training processes to associate distinct features with their corresponding behaviors. A number of scientific investigations have undertaken experiments with ensemble classifiers which involve the combination of several modeling approaches to enhance predictive accuracy. A body of research work has concentrated on deriving features directly from time-domain signals which include mean values, energy metrics, and root mean square calculations. The employment of classification algorithms such as Random Forests has been a common practice among these approaches to attain exceptional accuracy levels. Researchers have examined frequency-domain features alongside time-domain analysis techniques. Throughout their investigations researchers have utilized mathematical techniques such as the Fast Fourier Transform (FFT) and Discrete Fourier Transform (DCT) to extract meaningful signal characteristics. The application of Principal Component Analysis PCA together with Haar filters and Ensemble Empirical Mode Decomposition EEMD has achieved successful results in HAR systems. The process of planning a HAR project necessitates consideration of numerous essential elements which include the specific activities intended for recognition, the chosen sensing methods, the algorithms to be deployed, the data sources to be utilized, and the targeted application domain.
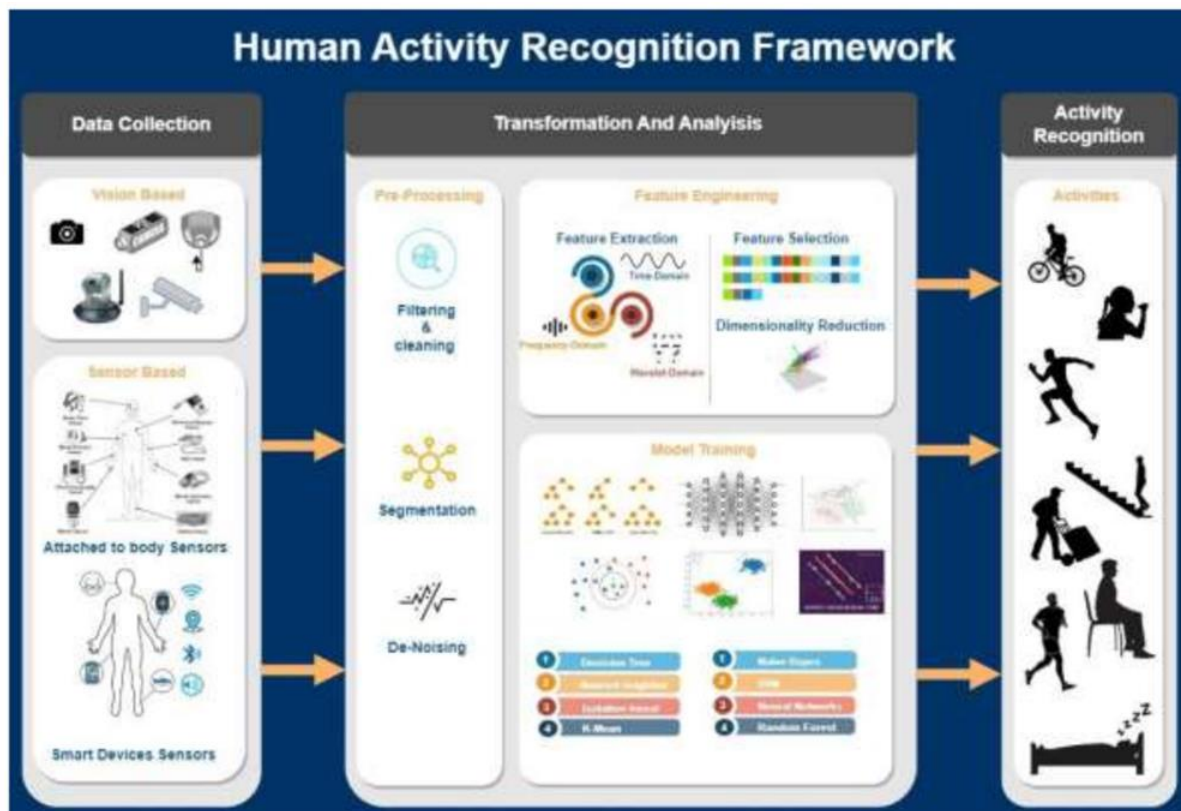

Fig- 2

CONCLUSION

The study of stream data analysis and processing frameworks for detecting anomalies in human activities reveals complex interactions between technological advancement and practical challenges. The exploration of stream data analysis techniques unfolds the creation of methodologies that enable real-time processing while simultaneously emphasizing the essential requirement for efficient algorithms capable of handling vast quantities of continuous data. Detailed examination of numerous frameworks uncovers their distinct outlier detection techniques which illustrate the adaptability of these systems to meet particular requirements across diverse applications such as healthcare monitoring and smart environments. Real-time data processing encounters numerous complex challenges which demand focused attention because elements like noise

management and dynamic behavior patterns together with computational limitations form substantial obstacles that necessitate resolution to attain improved accuracy and reliability. The ongoing research initiatives within this scientific domain necessitate the development of advanced models that boost detection capabilities while maintaining both scalability and adaptability in rapidly changing environments. Future research work must aim to combine machine learning methods with current systems to improve outlier detection while also tackling ethical issues about privacy and data protection. By meticulously examining and resolving these varied challenges we create opportunities to devise enhanced solutions that expand our understanding of human activities through the use of advanced stream data analysis techniques. The foundational progress of research initiatives aimed at improving these frameworks for practical application in our data-driven society will rely on the critical joint efforts between academic institutions and industrial entities.

Future Work:

- **Investigate federated learning**for privacy-preserving human activity recognition (HAR).
- **Improve explainable AI (XAI) techniques**to enable real-time anomaly interpretation.

REFERENCES

1. B. Tugrul, "Stream data analysis and processing frameworks for detecting ...," WSEAS Transactions on Information Science & Applications, 2025. [Online]. Available: https://wseas.com/journals/isa/2025/a225109-007(2025).pdf.
2. A. Manolova, "Stream data analysis and processing frameworks for detecting ...," WSEAS Journal Articles, 2025. [Online]. Available: https://wseas.com/journals/articles.php?id=10028.
3. A. Rezk, "Outlier detection in streaming data for telecommunications and ...," MDPI Electronics, vol. 13, no. 16, 2024. [Online]. Available: https://www.mdpi.com/2079-9292/13/16/3339.
4. S. Silvia, "Towards a deep learning-based outlier detection approach in the ...," Journal of Big Data, 2022. [Online]. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00670-8.
5. "A survey on outlier explanations," The VLDB Journal, 2022. [Online]. Available: https://link.springer.com/article/10.1007/s00778-021-00721-1.
6. MOHAMMED SABHA, BULENT TUGRUL, Stream Data Analysis and Processing Frameworks for Detecting Outliers in Human Activities: A Review [Online]. Available: https://www.wseas.com/journals/isa/2025/a225109-007(2025).pdf
7. Gazder, Uneb, "Studying Patterns of Rainfall and Topographical Clustering for Kingdom of Bahrain: An Application of Big Data", Engineering World, vol. 6, pp. 29–34, 2024, doi: 10.37394/232025.2024.6.5.
8. Ramaswamy, Sridhar and Rastogi, Rajeev and Shim, Kyuseok, "Efficient algorithms for mining outliers from large data sets", In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 427–438, 2000, doi: 10.1145/342009.335437.
9. Angiulli, Fabrizio and Basta, Stefano and Pizzuti, Clara, "Distance-based detection and prediction of outliers", IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 2, pp. 145–160, 2005, doi: 10.1109/TKDE.2006.29.
10. Gyllensten, Illapha Cuba and Bonomi, Alberto G, "Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life", IEEE Transactions on Biomedical Engineering, vol. 58, no. 9, pp. 2656–26
11. Krishnan, Narayanan C and Colbry, Dirk and Juillard, Colin and Panchanathan, Sethuraman, "Real time human activity recognition using tri-axial accelerometers", In Sensors, Signals and Information Processsig Workshop, vol. 2008, pp. 3337– 3340, 2008. [60] Ravi, Nishkam and Dandekar, Nikhil and Mysore, Preetham and Littman, Michael L, "Activity recognition from accelerometer data", In Proceedings of the 17th Conference on Innovative Applications of Artificial Intelligence, pp. 1541– 1546, 2005.