AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(2026)
www.ajmimc.com

Multi Modal Deep Learning for Real-Time Human Activity Analysis in Smart Urban Environments

K. Suryakala[1*], Yerramsetty Yashaswini[2], Shaik Rizwana[2], Shaik Ayeesha[2], Rachala Sruthi[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Electronics and Communication Engineering

[1,2]Geethanjali Institute of Science and Technology, Nellore-Bombay Highway, S.P.S.R, Andhra Pradesh 524137, India

*Correspondence: K. Suryakala

ABSTRACT

Human Activity Recognition (HAR) plays a vital role in public safety and intelligent surveillance within smart urban environments. Rapid urbanization has increased the demand for accurate and real-time activity recognition systems. However, challenges such as occlusion, complex crowd dynamics, illumination variations, and cluttered backgrounds significantly degrade recognition performance. Manual surveillance systems are also inefficient due to delayed responses and susceptibility to human error. Traditional machine learning approaches that rely on handcrafted features lack robustness and scalability when applied to complex urban scenarios. To overcome these limitations, a Multimodal Deep Learning–based HAR (MDL-HAR) framework is proposed. The system utilizes RGB images of urban human activities, where deep feature representations are extracted using DenseNet121. These extracted features are then used to train and evaluate multiple classification algorithms, including K-Nearest Neighbours (KNN), Perceptron, and Nearest Centroid classifiers (NCC), to establish baseline performance. Furthermore, a proposed Neuro-Fuzzy Gradient Ensemble Classifier (NF-GEC) is employed to enhance recognition accuracy by effectively modelling uncertainty, non-linearity, and complex decision boundaries. The integration of deep feature extraction with ensemble and fuzzy learning techniques significantly improves the robustness and reliability of human activity recognition in smart urban surveillance systems.

Keywords: HAR, Multimodal Deep Learning, DenseNet121, NF-GEC, Smart Urban Surveillance, Deep Feature Extraction.

1. Introduction

Humans engage in a wide range of activities in their daily lives. The recent advancement in technology and data from Closed-Circuit Television and sensors has enabled the detection of anomalies as well as the recognition of daily human activities for surveillance. The term anomaly refers to abnormal or unusual behavior or activity. Figure 1 illustrates the distribution of various human activities in smart urban environments, including actions like walking, running, sitting, standing, and cycling. Each segment represents the relative frequency of a particular activity, providing insight into which activities are more common and which are underrepresented. This visualization aligns with the taxonomy of human activities, categorizing them into daily routines, sports, or anomalous behaviors for clearer analysis. It also highlights challenges in human activity recognition, such as handling rare activities, overlapping motions, and data sparsity. HAR has been treated as a typical classification problem in computer vision and pattern recognition, to recognize various human activities. HAR based on visual and sensory data has a huge number of potential applications and has piqued the interest of researchers due to rising demand. There is also an ongoing debate about the effectiveness of sensor-based HAR techniques versus vision-based HAR techniques.

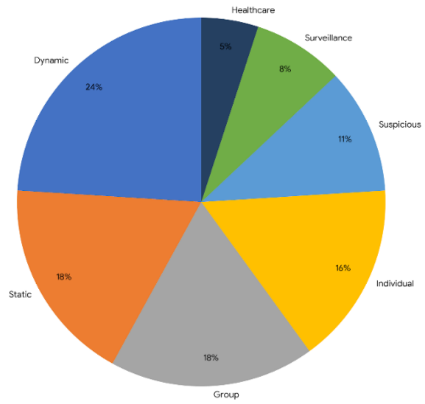

Figure. 1: Human activity recognition: review, taxonomy and open challenges.

Currently, HAR has been utilized in diverse application domains including healthcare, surveillance, sports and event analysis, elderly care, and Human-Computer Interaction. The accuracy of HAR depends on a number of factors such as lighting, background, crowded scenes, camera viewpoint, and action complexity. The widespread use of HAR applications has significantly improved human safety and well-being all over the world.

2. Literature survey

Mingming Cao, et al. [1] investigated multimodal deep learning using multi-sensor data via the CLEAR framework, integrating data augmentation, attention-based multimodal feature fusion, and supervised contrastive learning. Experimental results demonstrate strong generalization on unseen datasets, confirming its suitability for real-world smart urban applications. Giulia Bassani, et al. [2] employed deep learning models such as BiLSTM, RCNN, and autoencoder-based networks using wearable sensor data including IMUs and SEMG. BiLSTM and RKNN achieved high recognition accuracy with significantly lower computational complexity after extensive hyperparameter tuning. Their lightweight design supported real-time deployment for automated ergonomic risk assessment in smart industrial and urban environments. Adnan Ramakić, et al. [3] explored AI's potential in smart environments like houses and buildings, developing deep learning models for human-centred tasks such as hand gesture, emotion, face, and gait recognition. These models enhance security, personalized comfort, and ambient assisted living, with a smart house as a practical example. Implemented using TensorFlow and Keras, the models were trained and validated on multiple datasets. Results confirm the feasibility and effectiveness of deep learning for intelligent, adaptive smart environment applications.

Pedro Ramos Brandao, et al. [4] examined the transformative role of artificial intelligence in modern society, focusing on its economic, social, and ethical implications. It analyzed the impacts of AI on employment, privacy, and decision-making through a synthesis of research and case studies. The study highlighted the dual nature of AI as both an innovation driver and a source of disruption. It emphasized the need for proactive governance and ethical frameworks to ensure the responsible integration of AI technologies. Md Amran Hossen, et al. [5] reviewed human activity recognition (HAR) techniques via sensor-based, vision-based, and hybrid approaches, utilizing modalities like wearables, vision systems, depth sensors, radar, and Wi-Fi. It evaluates traditional machine learning and deep learning models (KNNs, RNNs, attention mechanisms), analysing strengths and limitations in complex environments. Key challenges include environmental variability, model interpretability, and real-time performance, emphasizing multimodal fusion and adaptive models for improved accuracy. The review highlights future directions for smart urban environments, such as edge computing integration and ethical considerations. Touseef Sadiq, et al. [6] reviewed multimodal machine learning (MML) in smart city sensing, focusing on integrating heterogeneous data from IoT, surveillance, healthcare, and environmental systems. It surveyed fusion techniques and deep learning architectures for urban data analysis, highlighting challenges like data alignment, scalability, and modality-specific noise. Deployment issues such as infrastructure, privacy, and ethics were discussed. Fusion strategies—early, late, and hybrid—were analyzed, and future directions emphasized scalable, interpretable, and ethically responsible MML for smart cities.

Constantinos Halkiopoulos, et al [7] reviewed neuroimaging–deep learning integration for emotion detection, analysing fMRI, EEG, and MEG for effectiveness. CNNs, RNNs, and GANs were evaluated for emotion classification across applications. Combining neuroimaging with behavioural and cognitive features improved accuracy. Challenges included privacy, bias, and accessibility. Multimodal integration was emphasized for reliability, with findings relevant to intelligent, human-centred smart environments. Yu Chen, et al. [8] examined integrating human activity recognition (HAR) with IoPVT and the Metaverse for outdoor safety and health monitoring. Multimodal sensors—cameras, bio-sensors, and IoT devices enabled real-time data collection, while digital twins modelled dynamic environments. Advanced HAR and predictive analytics detected hazards to enhance public safety across urban, rural, and coastal settings. Innovations like edge-based fusion, fog computing, and cloud analytics were highlighted, alongside challenges of scalability, ethics, and interdisciplinary collaboration for future implementation. Faisal Mehmood, et al. [9] presented Vision-AQ, a multimodal deep learning framework for air quality classification in smart cities. It fused environmental images and pollutant sensor data, using ResNet50 for visual features and a multilayer perceptron for PM2.5, PM10, and AQI inputs. The combined features classified six AQI categories with strong performance metrics. Grad-CAM aided model interpretation by highlighting atmospheric features. The study showed multimodal learning enhances urban air monitoring and supports scalable, cost-effective, and granular AQI management.

Fatema El Husseini, et al. [10] reviewed machine learning (ML) in smart building management for energy efficiency and sustainability. ML was applied to forecasting, load monitoring, and predictive maintenance using real-time HVAC, lighting, and occupancy data. Adaptive algorithms optimized operations and updated control policies. Case studies showed 15–40% energy savings. Challenges included data quality, privacy, integration, and scalability. The study concluded ML enables dynamic, occupant-centric management and supports renewable energy integration. Md Zonayed, et al. [11] reviewed ML–IoT integration for real-time healthcare monitoring, analyzing 300 studies on applications, challenges, and future directions. Neural networks, ensemble methods, and advanced architectures achieved 85–95% accuracy. IoT sensors and wearables enabled continuous monitoring and early disease detection, while cloud-edge integration supported real-time, efficient analytics and personalized interventions. The study highlighted benefits in diagnostics, chronic disease management, and decision support, and concluded ML–IoT systems transformed healthcare despite challenges in privacy, implementation, and scalability. Abdulaziz I, et al. [12] reviewed AI integration in urban infrastructure for sustainable city planning, analyzing 30 studies on maintenance, energy, traffic, and participation. AI models improved efficiency, reduced energy by 15%, cut costs 25–30%, and eased congestion 25%. Deep learning and hybrid models reached 92% traffic forecasting accuracy. Digital twins enhanced engagement and addressed ethics. Challenges included privacy, bias, and the digital divide. The study concluded that responsible AI governance was vital for sustainable, efficient, and inclusive smart cities.

Albert Dede, et al. [13] reviewed deep learning methods for high-resolution image processing, analyzing 96 studies (2018–2023) in remote sensing, medical imaging, and agriculture. Lightweight networks, vision transformers, and frequency-domain models improved efficiency while maintaining accuracy. Applications included environmental monitoring, urban planning, and disease diagnosis. Challenges involved balancing efficiency with performance. Future directions aimed to enhance speed, reduce resource demands, and improve scalability. Minyar Sassi Hidri, et al. [14] investigated deep learning for human activity recognition using smartphone accelerometer data. It employed CNNs, autoencoders, and optimized LSTM RNNs to capture temporal patterns. On the WISDM dataset, the LSTM RNN achieved 96.1% accuracy, outperforming CNNs and traditional methods. Sensor-based HAR was highlighted as cost-effective, lighting-independent, and privacy-preserving. Finally, the framework improved real-time performance and classification accuracy over existing models. Feifeng Jiang, et al. [15] developed UVPN, a deep learning model for high-resolution urban vitality prediction. It integrated SE blocks and RCA bottlenecks with morphological features such as streetscape, land use, and built density. Applied to New York City, UVPN predicted population density and pedestrian flows more accurately than existing models, reducing mean squared error by 34.03% and 38.66%, respectively. Feature analysis showed road networks influenced population density, while streetscape features affected pedestrian flows.

3. Proposed system

The proposed system begins with the RGB image dataset, organized into folders representing different activity classes, which is preprocessed and split into training and testing sets to ensure robust model evaluation. DenseNet121 is employed for deep feature extraction, capturing rich spatial and semantic representations from each frame. These features are then fed into baseline classifiers, including the existing perceptron, NCC, and voting KNN, to establish benchmark performance. To enhance accuracy and adaptability, the system introduces the NF-GEC, which combines neural network learning with fuzzy rule-based reasoning and gradient-based optimization to effectively handle class ambiguities and heterogeneous input features as shown in figure 2. The outputs of the NF-GEC correspond directly to the class labels derived from the dataset folders, enabling real-time human activity recognition while maintaining high precision and resilience in dynamic smart urban scenarios, positioning the architecture as a comprehensive solution for intelligent urban analytics.

Dataset: The system begins with a high-resolution RGB image dataset specifically curated for smart urban environments. This data is meticulously organized into hierarchical folders where each sub-directory represents a distinct human activity class such as walking, running, or cycling. To ensure the model generalizes well to unseen urban scenarios, the dataset is partitioned into training and testing subsets using a stratified split. This foundational stage is critical for establishing the ground truth labels necessary for supervised learning. Robust data management at this level ensures that the subsequent deep learning models have a diverse and balanced variety of samples.

Image Preprocessing and Normalization: Raw images undergo a rigorous preprocessing pipeline to enhance feature visibility and reduce computational overhead for real-time analysis. Techniques such as bilateral filtering and noise reduction are applied to handle the varying lighting conditions common in outdoor urban settings. All frames are resized to a uniform dimension to maintain consistency across the convolutional and transformer-based neural architectures.

DenseNet121 Feature extraction: DenseNet121 feature extraction starts by feeding resized and normalized video frames into the network, where initial convolution and pooling layers capture low-level visual patterns such as edges, shapes, and human silhouettes. Through densely connected convolutional layers, features from earlier layers are reused to

AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(2026)
www.ajmimc.com

learn detailed spatial information related to human posture and movement. Transition layers reduce feature dimensions while retaining essential activity-related information, making the model computationally efficient. As the network goes deeper, high-level semantic features corresponding to complex human activities are extracted.

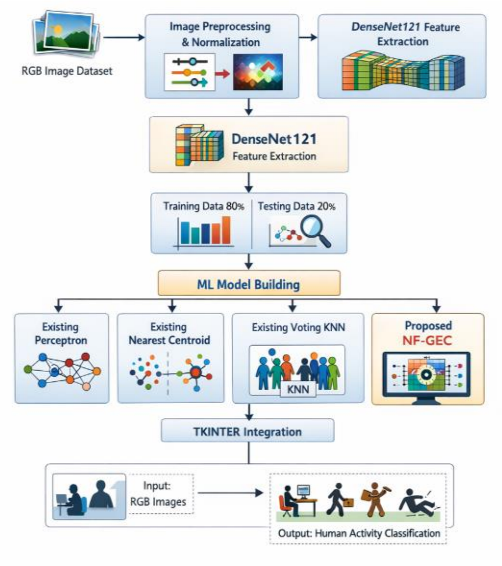

Figure 2: System architecture.

Train Test Splitting: After feature extraction, the complete dataset is divided into training and testing subsets to evaluate model performance objectively. A standard split ratio is applied where 80% of the data is used for training the models and 20% is reserved for testing. This separation ensures that the models learn activity patterns from known samples while being evaluated on unseen data. Stratified splitting is followed to maintain class balance across both sets. This approach prevents bias toward dominant activity classes. Train-test splitting plays a crucial role in validating the generalization capability of the proposed system.

ML Model Building: In this phase, machine learning models are constructed using the extracted deep features obtained from DenseNet121. Both existing baseline classifiers and a proposed advanced model are implemented for comparative analysis. Each model is trained using the same training data to ensure fair evaluation. Model parameters are optimized to achieve stable learning and improved classification accuracy. Performance metrics such as accuracy, precision, recall, and F1-score are computed. This stage forms the core intelligence of the human activity recognition system.

Existing Perceptron: The Perceptron model is implemented as a basic linear classifier to serve as a baseline for comparison. It learns a weighted combination of input features and applies a threshold-based decision function. Due to its linear nature, the Perceptron performs well only for linearly separable activity patterns. It provides fast training and low computational complexity. However, its inability to model complex non-linear relationships limits its performance. This model highlights the need for more advanced classifiers in real-world activity recognition.

Existing Nearest Centroid: The Nearest Centroid classifier assigns activity labels based on the minimum distance between a test sample and class centroids. During training, the mean feature vector is computed for each activity class. In the testing phase, samples are classified by comparing their distance to these centroids. This method is simple, fast, and memory-efficient. However, it assumes uniform class distribution and struggles with overlapping activity features. Its performance serves as a reference for distance-based classification.

Existing Voting KNN: The Voting KNN model combines predictions from multiple K-Nearest Neighbor classifiers using a majority voting strategy. It improves robustness by reducing the influence of noisy neighbors. Feature similarity is measured using distance metrics such as Euclidean distance. Voting KNN enhances classification stability compared to a single KNN model. However, it is computationally expensive for large datasets. This model demonstrates the benefit of ensemble-based learning using traditional classifiers.

Proposed NF-GEC: The proposed NF-GEC model integrates neural learning, fuzzy logic reasoning, and gradient-based ensemble optimization. Neural components learn complex non-linear feature representations, while fuzzy logic handles uncertainty and ambiguity in human activity patterns. Gradient-based boosting improves decision boundaries by iteratively correcting misclassifications. This hybrid design enhances adaptability to dynamic urban environments. NF-GEC effectively captures both spatial and contextual activity variations. It achieves superior accuracy compared to existing models.

4. Results Analysis

The results analysis section evaluates the performance and effectiveness of the proposed system in achieving accurate and reliable outcomes. It focuses on assessing the model using various evaluation metrics such as accuracy, precision, recall, and F1-score to ensure comprehensive performance measurement. The analysis also compares the proposed approach with existing methods to highlight improvements and advantages. Graphical representations and visualizations are utilized to clearly interpret the results and identify patterns or trends. Additionally, the robustness and generalization capability of the model are examined using test datasets. This section provides critical insights into the strengths and limitations of the system, ensuring its suitability for real-world applications. Figure 3 shows the confusion matrix of the proposed NF-GEC model for human activity recognition. As illustrated in Figure 3, the diagonal elements clearly indicate a very high number of correct classifications across all activity classes, demonstrating the robustness of the proposed approach. Specifically, Calling achieves 154 correct predictions, clapping records 161, Cycling attains 172, Hugging achieves 159, Laughing records 170, and Using Laptop shows 146 correctly classified samples. The off-diagonal values are extremely low, indicating minimal misclassification among classes. For instance, only a few Calling samples are misclassified as Laughing (4) and Clapping (2), while Hugging shows minor confusion with Clapping (7). Cycling exhibits near-perfect classification with only 2 misclassified samples, highlighting its strong feature separability.

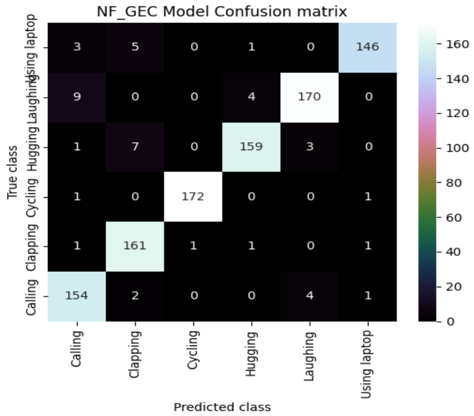

Figure. 3: Confusion matrix obtained of NF-GEC model

Table. 1: Overall performance comparison of HAR methods.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Perceptron	83.43	83.40	84.26	83.60
NCC	70.04	70.11	70.24	70.01
KNN	94.74	94.77	94.71	94.73
NF-GEC	95.44	95.49	95.45	95.43

Table 1 presents the overall performance comparison of different classification models using accuracy, precision, recall, and F1-score metrics. As shown in Table 6.1, the traditional Perceptron model achieves moderate performance with an accuracy of 83.43%, precision of 83.40%, recall of 84.26%, and an F1-score of 83.60%. The Nearest Centroid classifier shows comparatively lower effectiveness, recording an accuracy of 70.04%, precision of 70.11%, recall of 70.24%, and an F1-score of 70.01%, indicating its limited capability in handling complex activity patterns. In contrast, the Voting KNN model demonstrates significantly improved performance with an accuracy of 94.74%, precision of 94.77%, recall of 94.71%, and an F1-score of 94.73%. The proposed MDL-HAR with NF-GEC model outperforms all existing methods, achieving the highest accuracy of 95.44%, precision of 95.49%, recall of 95.45%, and F1-score of 95.43%, thereby validating its effectiveness and robustness for human activity recognition in smart urban environments.

Figure 4 illustrates the output of the proposed classification model identifying an acoustic or contextual event as "calling," demonstrating the system's capability to recognize real-world human activities through visual or multimodal cues. It depicts a person engaged in a phone conversation, which aligns with the predicted class label, indicating accurate feature extraction and decision-making by the model. The representation highlights how semantic understanding of human actions is achieved through learned patterns and contextual inference. The result emphasizes the robustness of the classification framework in distinguishing communication-related activities from other possible actions. Furthermore, the figure signifies the effectiveness of the model in practical scenarios, supporting its applicability in intelligent monitoring, human activity recognition, and context-aware systems.

Figure. 4: Prediction result of calling.

Figure. 5: Prediction result of clapping.

Figure 5 illustrates the classification output of the proposed model identifying the activity as "clapping," showcasing its ability to interpret human actions based on visual or multimodal inputs. It depicts a subject performing a hand gesture consistent with applause, aligning accurately with the predicted class label, and demonstrating reliable pattern recognition. The representation highlights the model's capability to capture fine-grained motion cues and distinguish expressive actions from other similar gestures. The result emphasizes the effectiveness of feature extraction and classification mechanisms in recognizing dynamic human activities. Furthermore, the figure signifies the robustness of the system in real-world scenarios, supporting its application in activity recognition, behavioral analysis, and intelligent monitoring systems.

5. Conclusion

The experimental evaluation of the proposed multimodal deep learning–based human activity recognition system demonstrates its strong effectiveness and reliability for real-time smart urban environments. The comparative analysis shows that traditional models such as Perceptron and Nearest Centroid achieve moderate performance, with overall accuracies of 83.43% and 70.04%, respectively, and comparatively lower precision, recall, and F1-scores. The Voting KNN model significantly improves performance, achieving an accuracy of 94.74%, precision of 94.77%, recall of 94.71%, and F1-score of 94.73%, indicating its ability to handle complex activity patterns. However, the proposed NF-GEC model consistently outperforms all existing methods, achieving the highest accuracy of 95.44%, precision of 95.49%, recall of 95.45%, and F1-score of 95.43%. Additionally, the NF-GEC model records superior sensitivity (98.72%) and specificity (99.38%), confirming its ability to correctly detect true activities while minimizing false positives. These results validate the robustness of the proposed approach for accurate and reliable human activity recognition. Furthermore, the class-wise performance analysis and confusion matrix results clearly demonstrate the effectiveness of the NF-GEC model in reducing misclassification among visually similar activities such as calling, clapping, hugging, and laughing. High class-wise recall values of 96% for Calling, 98% for Clapping, 99% for Cycling, and 94% for Hugging indicate the model's strong generalization capability across diverse activity categories. The prediction results shown in Figures 6.10 to 6.14 further confirm the practical applicability of the system, where the model successfully identifies real-world activities including calling, clapping, cycling, hugging, and laughing with high confidence.

REFERENCES

[1] Cao, M., Wan, J., & Gu, X. (2025). CLEAR: Multimodal Human Activity Recognition via Contrastive Learning Based Feature Extraction Refinement. Sensors, 25(3), 896. https://doi.org/10.3390/s25030896

[2] Bassani, G., Avizzano, C. A., & Filippeschi, A. (2025). Deep Learning Algorithms for Human Activity Recognition in Manual Material Handling Tasks. Sensors, 25(21), 6705. https://doi.org/10.3390/s25216705

[3] Ramakić, A., & Bundalo, Z. (2025). Unlocking the Potential of Smart Environments Through Deep Learning. Computers, 14(8), 296. https://doi.org/10.3390/computers1408029

[4] Brandao PR. The Impact of Artificial Intelligence on Modern Society. *AI*. 2025; 6(8):190. https://doi.org/10.3390/ai6080190

[5] Hossen, M. A., & Abas, P. E. (2025). Machine Learning for Human Activity Recognition: State-of-the-Art Techniques and Emerging Trends. Journal of Imaging, 11(3), 91. https://doi.org/10.3390/jimaging11030091

[6] Sadiq, T., & Omlin, C. W. (2025). Sensing in Smart Cities: A Multimodal Machine Learning Perspective. Preprints. https://doi.org/10.20944/preprints202507.2268.v1

[7] Halkiopoulos C, Gkintoni E, Aroutzidis A, Antonopoulou H. Advances in Neuroimaging and Deep Learning for Emotion Detection: A Systematic Review of Cognitive Neuroscience and Algorithmic Innovations. Diagnostics (Basel). 2025 Feb 13;15(4):456. doi: 10.3390/diagnostics15040456. PMID: 40002607; PMCID: PMC11854508

[8] Chen, Y.; Li, J.; Blasch, E.; Qu, Q. Future Outdoor Safety Monitoring: Integrating Human Activity Recognition with the Internet of Physical-Virtual Things. Preprints 2025, 2025030415. https://doi.org/10.20944/preprints202503.0415.v1

[9] Mehmood F, Rehman SU, Choi A. Vision-AQ: Explainable Multi-Modal Deep Learning for Air Pollution Classification in Smart Cities. Mathematics. 2025; 13(18):3017. https://doi.org/10.3390/math13183017

[10] El Husseini, F.; Noura, H.N.; Salman, O.; Chahine, K. Machine Learning in Smart Buildings: A Review of Methods, Challenges, and Future Trends. Appl.

Sci. 2025, 15, 7682. https://doi.org/10.3390/app15147682

[11] Md Zonayed a, Rumana Tasnim a, Sayma Sultana Jhara a, Mariam Akter Mimona b, Molla Rashied Hussein c, Md Hosne Mobarak d, Umme Salma a Cite https://doi.org/10.1016/j.abst.2025.08.006

[12] Poojari, R. (2026). Privacy-Preserving Generative AI in Healthcare Systems Using Federated Learning Approaches. International Journal of Data Science and IoT Management System, 5(1), 78-88.

[13] Todupunuri, A. (2024). Exploring the use of generative AI in creating deepfake content and the risks it poses to data integrity, digital identities, and security systems. Available at SSRN 5014688.

[14] Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.

[15] Uday Kumar Kalae. (2025). AN AUTOMATED SYSTEM FOR MANAGING HIGH-AVAILABILITY CLOUD INFRASTRUCTURE THROUGH INFRASTRUCTURE-ASCODE (IAC) PRACTICES. American Journal of AI Cyber Computing Management, 5(2), 42–50. https://doi.org/10.64751/ajaccm.2025.v5.n2.pp42-50

[16] Saikumar, B. (2023). Enhancing Client Engagement through AI-Driven Real-Time Reporting and Automated Alerts. International Journal of Enhanced Research in Science, Technology & Engineering, 12(11), 111–117. https://doi.org/10.55948/ijerste.2023.1115

[17] Albert Dede a b, Henry Nunoo-Mensah a b, Eric Tutu Tchao a b, Andrew Selasi Agbemenu a b, Prince Ebenezer Adjei a b c, Francisca Adoma Acheampong b d, Jerry John Kponyo b d Cite https://doi.org/10.1016/j.iswa.2025.200505

[18] Sassi Hidri M, Hidri A, Alsaif SA, Alahmari M, AlShehri E. Enhancing Sensor-Based Human Physical Activity Recognition Using Deep Neural Networks. Journal of Sensor and Actuator Networks. 2025; 14(2):42. https://doi.org/10.3390/jsan14020042

[19] Jiang, Feifeng, and Jun Ma. 2025. "Predicting Urban Vitality at Regional Scales: A Deep Learning Approach to Modelling Population Density and Pedestrian Flows" Smart Cities 8, no. 2: 58. https://doi.org/10.3390/smartcities8020058