

Recent Implementation of Deep Learning Methods for Shoplifting Detection in Retail Environments Using Surveillance Videos

Chioma Virginia Anikwe-Onyiwa ¹; Henry Friday Nweke ^{1,2}; and Kingsley Nnabuike Omeje ¹

¹Computer Science Department, David Umahi Federal University of Health Sciences, P.M.B 211, Uburu, Ebonyi State, Nigeria. ²Institute for Machine Learning, Robotics and Artificial Intelligence Research, David Umahi Federal University of Health Sciences, P.M.B 211, Uburu, Ebonyi State, Nigeria.

Corresponding author: nwekeh@dufuhs.edu.ng

DOI: <https://doi.org/10.62154/ajastr.2025.021.01012>

Abstract

Shoplifting is a problem that continues to face retail businesses and is associated with a serious loss of revenue and security issues for retailers. The recent developments in deep learning have created new possibilities to develop intelligent surveillance systems that detect suspicious behaviour automatically and with high precision. This paper systematically reviews recent literature between 2018-2025 on deep learning methods for shoplifting detection in retail environments. It discusses the applications of deep learning mechanisms in the detection of shoplifting in retail settings. In addition, the paper presents methods that utilise Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid deep neural networks to detect shoplifting. Further discussions were presented on privacy-preserving architectures, behaviour analysis, and anomaly detection using surveillance videos. To ensure continued implementation by researchers, the paper outlines research gaps such as low levels of real-time use, poor general applicability in a variety of retail environments, low compliance with privacy, and ethical issues that require further research. The findings reveal that deep learning approaches are important artificial and machine learning techniques for shoplifting surveillance and detection of suspicious activities in retail settings for prompt prevention of crimes and thefts by business owners. Also, the study will add to the contained improvements to the existing body of knowledge in the field of computer vision, deep learning, and cybersecurity.

Keywords: Shoplifting Detection, Deep Learning, Retail Surveillance, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN).

Introduction

Shoplifting is an ongoing threat that has a big effect on communities, businesses, and the economy. Undiscovered thefts cause significant financial losses, operational inefficiencies, and security issues for retailers (Ansari and Singh, 2022). Retail-related theft has led to a loss of revenue amounting to \$112.1 billion in 2022 and an estimated \$121.1 billion in 2023 in the United States alone, and the loss is projected to exceed \$143 billion in subsequent

years (Capital One Shopping, 2024). However, with only about 2% of shoplifters being apprehended, current security measures appear largely ineffective despite the growing scale of the problem (Capital One Shopping, 2024).

Artificial intelligence-powered shoplifting detection systems have garnered a lot of attention due to the shortcomings of conventional security measures (Ansari and Singh, 2023; Arroyo et al., 2015). Despite being widely used, video surveillance systems produce enormous volumes of data that security guards are unable to process in real time, opening the door for automated detection systems (Arroyo et al., 2015; Muneer et al., 2023; Pazho et al., 2023). Real-time theft detection, staff warnings, and actionable information, including pinpointing high-risk locations and peak shoplifting periods, might all be made possible by AI systems coupled with current security infrastructures (Kirichenko et al., 2022a).

Three major obstacles that persist in vision-based shoplifting detection studies, despite advancements in AI-based computer vision, include a lack of real-world data, tagged data, and the privacy of shoppers. Research on shoplifting detection is hampered by a scarcity of real-world datasets. The intricacies of real shoplifting episodes are not adequately captured by existing datasets (Ansari and Singh, 2022; Arroyo et al., 2015; Muneer et al., 2023), which frequently rely on data compiled from internet sources or staged situations using actors (Sultani et al., 2018). The application of these datasets in real-world contexts is limited since they lack the contextual details unique to each retail venue.

In addition, progress is further hampered by the intricacy and paucity of tagged data. Because shoplifting incidents are uncommon, unexpected, and challenging to document in real-world settings, supervised learning techniques that depend on labelled data are less successful (Pazho et al., 2023). Moreover, theft habits vary greatly because of variations in product categories, store layouts, and customer interactions, which makes it more difficult to identify general trends.

Moreover, privacy and prejudice issues are important factors to take into account when determining shoplifting behaviours (Noghre et al., 2023; Hirschorn and Avidan, 2023). Using raw video footage raises privacy concerns since it may contain identifying customer information. Additionally, detection models may incorporate biases based on gender, race, or clothing, which might lead to unfair results (Benthall and Haynes, 2019; Buet-Golfouse and Utyagulov, 2022; Nagpal et al., 2019; Noghre et al., 2024).

Deep learning, a kind of machine learning that uses multi-layered neural networks, has demonstrated remarkable achievements in computer vision applications, including object detection, action recognition, and anomaly detection. When analysing CCTV footage in retail settings, where it's critical to spot subtle or hidden behaviours, these characteristics are quite helpful. To find patterns of behaviour linked to shoplifting, such as hiding objects or taking odd movements, models such as Convolutional Neural Networks (CNNs), Long Short-Term Memory Networks (LSTMs), and hybrid architectures have been used (Jabur et al., 2024; Nazir et al., 2023; Muneer et al., 2023; Rashvand et al., 2025). As these technologies advance, their use in the retail industry has the potential to completely

transform loss prevention tactics and provide safer, more secure, and more productive shop environments. This study advances the state-of-the-art in shoplifting detection systems and establishes the groundwork for future research in privacy-preserving retail security.

In this review, the study comprehensively discusses categories of conventional machine learning and deep learning models for shoplifting detection, outlining their strengths and weaknesses of each model. Furthermore, the review enumerates limitations that hinder the effective implementation of these models and discusses further research directions to improve the implementation. This study bridges the gap in understanding how deep learning architecture can enhance automated shoplifting detection, focusing on model accuracy, real-time implementation, and privacy preservation during implementation in retail environments.

The remainder of the paper is organized as follows: Section 2 explains machine learning and its types. Section 3 discusses deep learning methods for shoplifting detection. Section 4 outlines the challenges hampering deep learning implementation for shoplifting detection and solutions. Section 5 explains the open research directions, while Section 6 concludes the research and outlines the main findings.

Methodology

This study utilized a comprehensive and systematic literature review to investigate artificial intelligence methods for shoplifting detection in retail environments using surveillance videos. To collect and analyze appropriate research studies to achieve the aim of the study, we followed the steps outlined below. These steps include literature search and review, inclusion and exclusion criteria, synthesis, and analysis of data.

- i. We performed a comprehensive search of different databases such as Scopus, IEEE Xplore, Google Scholar, ACM Digital Library, and Web of Science to retrieve relevant peer-reviewed and published articles between 2018 and 2025. We used a combination of different search strategies, keywords, and Boolean operators ("machine learning" OR "deep learning" OR "Computer vision" OR "recurrent neural networks", OR "Long short-term memory" OR "convolutional neural networks") AND ("shoplifting" OR "detection" OR "image analysis") AND ("retail environment")
- ii. The articles selected for review were selected using the criteria: the article must be published in English between 2018 and 2025, and provide technical insights into shoplifting detection in retail environments using deep learning models.
- iii. The selected articles were synthesized and analysed to identify studies related to deep learning models for shoplifting detection, strengths, weaknesses, and open research directions (Kirichencko et al, 2022; Alsaiani and Ilyas, 2025).

Machine Learning for Shoplifting Detection

Machine Learning (ML) is an artificial intelligence subfield that enables computer systems to identify patterns and relationships in unstructured data (He et al., 2024). Through

repetitive training, the ML models can enhance their performance over time by adapting to new data and refining their predictions. The most common types of ML are supervised, unsupervised, semi-supervised, and reinforcement learning models, and the highest level of methods include deep learning and explainable AI (Ragnoli et al., 2023). They are based on high volumes of data to identify underlying structures, correlations, and trends automatically and thus allow systems to make data-driven decisions and predictions with limited human intervention.

Object detection and video analysis are among the most important applications of machine learning and are crucial for automated shoplifting detection. ML systems can analyze surveillance camera footage to spot suspicious activities or unusual behavior by using computer vision algorithms and deep learning models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. ML models are trained with labelled video data to identify specific objects (e.g., merchandise, human figures) and track their movement across frames. These capabilities allow theft-related behaviors to be detected in real time, reducing the need for human oversight and enhancing the accuracy of security systems (Dramsch et al. 2020).

Machine learning has a number of benefits in such areas as shoplifting detection. It facilitates the automation of complicated processes and helps security systems to work 24/7 and always without weakening or bias (Perumal and Ganesh, 2025). ML models are able to process and rapidly analyze large quantities of visual information, detect intricate patterns that might otherwise go unnoticed by human users, and do so as they are trained on new data (Umut et al., 2025). Moreover, they enhance decision-making accuracy, reduce false alarms, and allow for scalable deployment across multiple locations (Rashvand et al., 2025). These benefits make machine learning a powerful and cost-effective tool for improving security and loss prevention in retail environments.

Supervised ML Algorithms

Supervised machine learning is a form of machine learning in which models are trained on labeled data, i.e., each training instance consists of the input data and the answer data. The model is trained to give an output as patterns and relationships are detected in the data (Manasa, 2025). When trained, it can provide the prediction of the results of new, unknown data, which is why it is very useful in the real world. It is one of the most widespread and functional paradigms of machine learning and is therefore commonly used in various tasks, including classification (placing inputs into categories), regression (forecasting continuous values), and so forth (Khaldi et al., 2025; Osisanwo et al., 2020). Decision trees, support vector machines, and neural networks are classification algorithms that are usually used in the fields of healthcare, finance, and transportation because they are effective in prediction and can be adapted to different needs.

Osisanwo et al. (2017) and Singh et al. (2016) claim that supervised learning is a subject that covers a variety of categories of problems and methods, such as classification, regression, approaches based on neural networks, ensemble methods, techniques based on

optimization, object detection, filtering features, and dimensionality reduction. Common supervised algorithms are Decision Trees, Support Vector Machines (SVM), Random Forests, k-Nearest Neighbors (k-NN), Bayesian algorithms, boosting algorithms (including Random Under-Sampling Boosting (RUSBoost) and Adaptive Boosting [AdaBoost]), and regression models. RUSBoost uses random under-sampling to deal with class imbalance, and AdaBoost uses multiple weak learners to enhance general model performance.

Unsupervised Machine Learning

When analyzing data without labels or prediction goals, unsupervised algorithms are used. These algorithms, which can be categorized into areas such as dimension reduction, clustering algorithms, and set theory-based optimization, seek patterns or structures within the data without relying on external variables. To identify new data categories, the models utilize previously learned characteristics (Chervonenkis et al., 2013; Tyagi et al., 2022).

Clustering algorithms, such as Cluster Analysis (CA), the Iterative Self-organising Data Analysis Technique (ISODATA), and Cluster Confusion Normalized Mutual Information (CC-NMI), are a set of methods used to group related items based on specific similarity or dissimilarity criteria. The Positive-Unlabelled (PU) learning algorithm is a notable example of a semi-supervised learning method. To enhance model performance, this approach utilizes a combined dataset that includes both labeled and unlabeled data. It operates on the assumption that both positive and negative examples may be present in the unlabeled data pool. This learning method is especially valuable when there is a large amount of unlabeled data and a small amount of positively labeled data.

Finally, policy gradient methods, Q-learning, State-Action-Reward-State-Action (SARSA), and Deep Q-Networks (DQNs) are examples of machine learning paradigms that use reinforcement learning, where an agent learns to act in an environment while receiving feedback in the form of rewards or penalties to maximize a specific goal (Binetti et al., 2024).

Reinforcement Learning

Reinforcement Learning (RL) is a form of machine learning in which an agent is allowed to train in order to make a decision and obtain feedback as a reward or penalty based on its interactions with the environment (Kushwaha et al., 2025). In contrast to supervised learning, RL does not include the presence of labels of possible outputs. Rather, the agent tries out various behaviors, sees the consequences, and over time learns an optimal policy that will optimize cumulative rewards (Waradpande et al., 2020). The RL problems are frequently described as Markov Decision Processes (MDPs), where the agent makes decisions according to the current state and tries to improve its strategy of decision-making by trial and error (Pacchiano et al., 2021).

Among the key benefits of reinforcement learning, there is the possibility to cope with the complex decision-making problems in dynamic environments and uncertainties. It also

helps in continuous improvement via self-education, enabling the system to be adjusted to changing circumstances without the clear reprogramming (Liang et al., 2024). The RL can find new approaches that humans may not be expecting and can perform best in cases where a definition of what is right behaviourally is hard or unfeasible (Liu et al., 2024). This is appropriate in situations that involve long-term planning, sequential decision making, and optimization in the presence of unpredictability (Woods, 2021).

Reinforcement learning has been effectively used in many fields, such as robotics (autonomous navigation and manipulation), game playing (Chess and Go), finance (autonomous algorithmic trading and portfolio optimization), industrial process control, personalized recommendations, and autonomous vehicles. The application of RL can be applied in security and surveillance tasks to optimise camera positions, efficiently allocate computational resources, or dynamically change alert levels in intelligent surveillance (Issa and Aldair, 2022).

Limitations of Conventional Machine Learning Models for Shoplifting Detection

Conventional machine learning methods have played critical roles in the detection of suspicious activities in surveillance videos. These machine learning models utilized handcrafted features extracted through feature engineering approaches such as histogram of oriented gradients (HOG), Scale-invariant feature transformation (SIFT), local binary pattern (LBP), etc and then model using machine learning models, including support vector machine, k-Nearest Neighbors, decision tree, random forest, and artificial neural networks (Islam et al, 2020). However, conventional machine learning methods depend solely on handcrafted features, which makes it difficult to generalize to new application areas. In addition, conventional machine learning approaches have little capability for temporal activity detection, are highly sensitive to noise, illumination, and background variation (Kumar et al, 2017; Bhuiyan et al, 2022; Jia et al, 2025). Here, we outline the challenges of implementing conventional machine learning for shoplifting detection and the reasons for the adoption of recent implementations of deep learning methods.

- i. *Lack of generalization and scalability to complex video scenes*: Detection of shoplifting in surveillance video requires the collection of a large volume of crime scenes, which is difficult for conventional machine learning methods to handle. These models produce results that cannot scale and generalize to unseen data when implemented with a small dataset. Nonetheless, it exhibits poor generalization with multiple videos (Choudhry et al, 2023). Conventional machine learning methods utilized static videos with limited targets, which makes it difficult to scale learning from large datasets.
- ii. *Inability to capture temporal and contextual information from video*: Conventional machine learning methods lack the capacity to capture long-range dependencies, contextual information, and suspicious behaviours from video frames as a result of processing single frames from videos at a time and narrow temporal scope (Kang et al, 2024).

- iii. *Overdependence on handcrafted feature representation*: The performance of conventional machine learning methods in shoplifting detection tasks and video analysis depends on the knowledge of the experts who implement the models. Therefore, the types of feature extraction techniques, such as histogram of oriented gradients, scale-invariant feature transform, etc, and the performance of the machine learning models are a result of experience of the expert's to extracting relevant features and tuning the features for specific environmental conditions.
- iv. *Inability to handle high-dimensional feature vectors and unstructured data*: Another challenge faced by conventional feature engineering methods for shoplifting detection is the inability of the algorithms to handle high-dimensional feature vectors and unstructured video data (Jia et al, 2025). Shoplifting video scenes contain high-dimensional video with complex visual and temporal segmentation. Therefore, detecting suspicious shopping activities requires reducing the video scene to the required dimension, which leads to a loss of relevant information for detecting shoplifting. In addition, the higher the dimensionality of the video data, the lower the classification accuracy produced by conventional machine learning models (Adnan and Akbar, 2019).
- v. *Highly sensitive to background variation, noise, and light illumination*: Shoplifting detection models implemented using conventional machine learning models often suffer from background changes, noise, and light illumination. These include the changes in environmental conditions during video data collections, lighting, camera jitter, and noise variations. The presence of these challenges might lead to false positives, misclassification, and low performance accuracy.

Table 1 outlines the challenges of implementing conventional machine learning methods for shoplifting detection and video analysis, and how deep learning methods are deployed to resolve the issues.

Table 1. Challenges of Conventional Machine learning and Solutions

Conventional Machine Learning Challenges	Solutions
Lack of generalization and scalability	Deep learning has higher scalability to learning from large datasets, and the use of transfer learning and pre-trained architectures for massive datasets
Inability to capture temporal and contextual information	The use of various deep learning models, such as recurrent neural networks, convolutional neural networks, and transformer models, to capture long-term dependence from videos
Over-reliance on manually engineered features	Implementation of deep learning methods for automatic feature representation that captures temporal and spatial patterns from an image without relying on expert knowledge.

Conventional Machine Learning Challenges	Solutions
Inability to handle high-dimensional and unstructured data	Deep learning methods are designed to automatically handle high-dimensional data by automatically detecting patterns and correlations in data.
Sensitive to background variation, noise, and light illumination	Utilize hierarchical feature extractions to extract different feature levels, such as appearance, motion cues, and variations from data.

Deep Learning Methods for Shoplifting Detection

Deep learning is a subfield of artificial intelligence that uses representational learning to automatically discover salient features in videos and images (Lecun, 2015; Szegedy et al., 2015). Due to its ability to extract features from data with minimal human intervention, deep learning has found extensive applications in image recognition, speech recognition, natural language processing, drug discovery, mobile sensor-based activity recognition, and recently for anomaly and shoplifting detections (Nweke et al, 2018; Kirichencko et al, 2022). Various deep learning methods have been developed over the years for efficient feature representation, and include Restricted Boltzmann machine, deep autoencoder, convolutional neural networks, and recurrent neural networks. However, convolutional neural networks and recurrent neural networks are widely implemented for shoplifting detection due to their ability to automatically extract spatial and temporal features from crime videos (Abdolrasol et al., 2021; Nazir et al., 2023). Therefore, subsequent sections will discuss the implementation of convolutional neural networks and recurrent neural networks for shoplifting detection, both individually and in a fusion model. Figure 1 depicts the structure of a deep learning model for shoplifting detection.

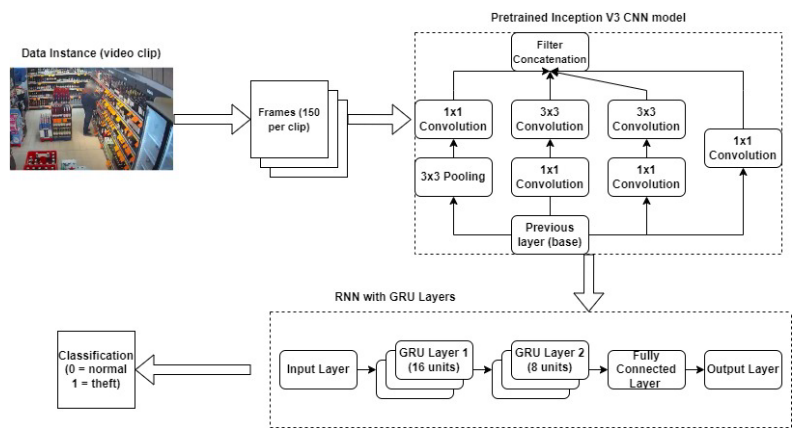


Figure 1. Architecture of deep learning for shoplifting detection (Pham et al, 2024)

Convolutional Neural Networks (CNNs)

A convolutional neural network is a deep learning method made up of interconnected deep neural networks that perform convolution operations on shoplifting videos and images to automatically learn spatial features and recognize crime patterns (Nweke et al, 2018). CNN is composed of a convolutional layer, a pooling layer, and a fully connected layer. These layers work in sequence to extract needed spatial information from data for detection activities. A deep neural network is designed to process image information. The convolutional layer scans input images and videos to identify spatial features such as edges, textures, and shapes, and combines them successively to identify complex shoplifting scenes. The pooling layer minimizes the feature map extracted from videos and makes them translational invariant to changes and distortions. The fully connected layer integrates Softmax classification algorithm, or other classification algorithms to detect shoplifting incidents (Martínez-Mascorro et al, 2020). The main function of a convolutional neural network (CNN) in shoplifting and anomaly detection is to extract spatial features from video segments collected during shoplifting incidents. Recent studies have implemented convolutional neural networks for feature extraction tasks. For instance, Nazir et al. (2023) researched the detection of suspicious behaviors by using temporal feature extraction and time-series classification to prevent shoplifting. They transformed a video classification task into a time-series classification problem by applying object tracking techniques, offering a new approach for identifying suspicious activity in videos for crime prevention. The method was tested on the well-known UCF Crime dataset and compared to the state-of-the-art Robust Temporal Feature Magnitude (RTFM) method, which employs the Inflated 3D ConvNet (I3D) preprocessing approach. Their proposed method uses YOLOv5 for object detection combined with Deep Sort for tracking people across video frames, using the resulting bounding box coordinates as temporal features. Results show an impressive 8.45-fold increase in detection inference speed compared to RTFM, along with an F1 score of 92%, which surpasses RTFM by 3%.

Martínez-Mascorro et al. (2021) researched the application of 3D CNN for early detection of criminal intentions and shoplifting. In this work, features were extracted and classified using a 3DCNN. Typically employed for spatiotemporal analysis, the 3DCNN has shown remarkable effectiveness in processing films across several domains, such as gesture, movement, and motion object detection. It was also used in more challenging situations, such as searching for patterns in video samples that don't display any illegal or questionable visual activity. The model consists of four Conv3D layers, two pairs of successive convolutional layers for capturing long-term dependencies, two max-pooling layers, and two fully connected layers. For these experiments, a certain number of videos from the UCF-Crimes dataset were used. According to the study's findings, the model created in this work achieves accuracy and recall values of 0.8571 and 1, respectively, in one of the test cases.

Martínez-Mascorro et al. (2020) employed 3D CNNs for the detection of suspicious behavior and the prevention of shoplifting crimes. By examining unique circumstances that the

general public might view as normal, but which could result in criminal activity, the concept focuses on shoplifting offenses that might be viewed as normal but could lead to criminal activity. The model's architecture comprises four Conv3D layers, two max-pooling layers, and two fully connected layers. By default, the model applies 32 filters to the first pair of Conv3D layers and 64 filters to the second pair. Every kernel is $3 \times 3 \times 3$, and the model calculates losses using cross-entropy and an Adam optimiser. Two thick layers of 512 and 2 neurones, respectively, are present at the conclusion of the model. The result is binary, with 0 denoting normal behaviour and 1 denoting suspicious behaviour. The study's findings indicate that the proposed design achieves 90% accuracy in both scenarios. The confusion matrices demonstrate that the suggested design obtains a minimal number of false positives from normal-behavior samples and accurately classifies 100% of suspicious samples for both balanced and unbalanced datasets.

Recurrent Neural Networks (RNNs)

The recurrent neural networks are deep learning methods developed to extract temporal and action sequences in shoplifting scenes and other classification tasks. It integrates a temporal layer to capture and learn complex changes and dependencies within video sequences. RNNs can be applied in shoplifting detection to look at the way the behavior develops over time and to determine the difference between normal and suspicious behavior patterns. However, as a result of vanishing gradients that limit its application for modeling long-range video sequences, variations of recurrent neural networks have been developed by researchers. These include the long short-term memory and gated recurrent unit (Kirichencko et al, 2022).

- *Long Short-Term memory (LSTM)*: it incorporates a memory cell to capture long-term dependencies in video sequences. The memory cell holds information over a long period of time to eliminate vanishing gradient. This is achieved through the introduction of three main gates, viz input gate, the forget gate, and the output gate. The input gate controls the amount of information that gets into the memory cell. The forget gate checks the information in the memory to determine what information to remove from the memory cell. Finally, the output gate determines the amount of information produced from the memory cell. The incorporation of these gates enables LSTM to learn long-term dependencies and extract temporal features from long video sequences (Guan & Ploetz, 2017).
- *Gated Recurrent Unit (GRU)*: Even though Long Short-Term Memory enables recurrent neural networks to learn global features, it is affected by too many parameter optimizations to achieve optimal performance results, thereby increasing computational time, especially in video processing (Nweke et al, 2018). Therefore, the gated recurrent unit was introduced to resolve the issue of too many parameters to update and computational complexity. A gated recurrent unit has a few parameters to update and is easy to implement. (Kirichencko et al, 2022).

Moreover, the model updates its hidden state based on the time needed to keep the information in the memory cell, while LSTM updates information based on constant summation operations (Nweke et al, 2018).

Recent advances in information processing have seen the implementation of variations of recurrent neural networks, such as long short-term memory and gated recurrent unit, for shoplifting detection.

For instance, Ansari and Singh (2023) present an optimized parameter tuning approach for the recurrent learning process, which can be applied for the classification of shoplifting activity. The suggested technique mines the salient characteristics from video clips using an InceptionV3 architecture that is based on deep convolutional learning. These characteristics are utilised to distinguish between human thieving behaviours in video sequences using a proprietary Long Short-Term Memory (LSTM) network. One real addition of this study is the optimisation of the recurrent learning classifier utilising various modelling parameters, such as batch size and sequence length. The studies show that the suggested system surpasses other current approaches, with an accuracy of 89.36% on the synthesised dataset.

Kirichenko et al. (2022b) applied a hybrid network for the detection of shoplifting in a video scene. Both convolutional and recurrent neural networks were included in the hybrid neural network. The video frames' features were extracted using a convolutional network. The recurrent network categorised the video fragments by processing the time sequence of the video frame characteristics. Gated recurrent units were used as the recurrent network in this work. The training and test datasets were derived from the popular UCF-Crime dataset. The results of the classification indicated a remarkable 93% accuracy.

Hybrid Deep Learning Methods

Hybrid deep learning algorithms use several model architectures to exploit their complementary capabilities. As an illustration, CNN-LSTM hybrid can be applied wherein the spatial characteristics of the video frames (obtained with the help of CNN) are initially obtained, and the temporal patterns within the sequence of the frames are subsequently analyzed (using LSTM) (Pandey et al., 2024). It is particularly effective in the context of shoplifting detection since such a combination captures both what happens in each frame (object, people) and how behavior changes over time (behavioral patterns) (Muneer et al., 2023). Hybrid models are better in real-time surveillance tasks, as they can give a more accurate and context-aware prediction and achieve higher scores in benchmark assessment compared to traditional CNN or LSTM-only models (Alsaiani and Ilyas, 2025). Recent studies have explored the effectiveness of combining different deep learning models for shoplifting detection.

Muneer et al. (2023) utilized a benchmark dataset developed for creating a shoplifting detection system. This study presents a large benchmark dataset of 900 instances with hand-annotated data based on various shoplifting tactics, comprising 450 cases involving theft and 450 instances not involving theft. Furthermore, the benchmarked dataset is

analyzed using baseline deep learning approaches, such as 2D CNN and 3D CNN, as well as the proposed method. After implementation and comparison with alternative techniques, the proposed method, which combines BiLSTM and Inception V3, achieved an accuracy of 81% and outperformed other baseline methods.

Rashvand et al. (2025) used a real-world shoplifting dataset and benchmark to explore pose-based anomaly detection for retail security. To address issues such as data scarcity, privacy concerns, and model biases, the study presents PoseLift, a privacy-preserving dataset specifically designed for the detection of shoplifting. PoseLift uses anonymised human posture data from real-world situations and was developed in partnership with a retail establishment. PoseLift strikes an anonymizing identity while maintaining a crucial balance between privacy and usefulness by preserving crucial behavioral data. In the USA, the film records were captured at a nearby retail establishment. All of the films were shot at a frame rate of 15 frames per second and from high-angle perspectives with a resolution of 1920 x 1080. Temporal-Spatial Graph Anomaly Detection (TSGAD), Graph Embedding with Pose Clustering (GEPC), and Spatio-Temporal Graph Normalising Flows (STG-NF) were all used in the study. These methods are all learnt in an unsupervised learning environment. According to the study's findings, STG-NF continuously performs better than the other models, obtaining the greatest Area Under the Curve-Receiver Operator Characteristics (AUC-ROC) of 67.46%, AUC-PR of 84.06%, and EER of 0.39. In terms of AUC-ROC performance, TSGAD comes in second, while GEPC has the lowest AUC-ROC at 60.61% and a middle-range AUC-PR performance.

Kirichenko et al. (2022c) integrated a CNN-RNN model for the recognition of shoplifting activity from CCTV video footage. The original and processed video fragments (a histogram of orientated gradients and optical flow) make up the two parallel streams of the hybrid architecture. While the recurrent network uses the temporal information from frame sequences as features to categorise the activity, the convolutional neural network model collects features from each frame of the video segment. The UCF-Crime dataset was used as the training dataset, featuring videos illustrating shoplifting occurrences. While the video data set was purposefully increased, the primary class was undersampled to solve the dataset's imbalance problem. A CNN that had already been trained was used for additional tests. 92% classification accuracy was demonstrated by the classifier.

Mohite et al. (2022) present a study on the application of AI for a surveillance system to detect theft activities. The suggested solution uses a feature-matching-based, extremely basic decision-making algorithm. As the reference image, the initial picture taken will be kept in the database. Using the feature matching approach, objects identified by the new current picture are compared to the database values. The suggested system will provide the user/owner an alert message and the recorded picture in comparison to the huge original movie if the current image frame and the reference image frame saved in the dataset match. The procedure proceeds from the picture acquisition phases if there is no match between the two frames. The output of these modules is then combined to create an input for a different machine learning model that determines who should receive the alert

message. The work used a variety of classification techniques, such as logistic regression and Support Vector Machine (SVM), among others, to create this model. Nevertheless, the research did not report on the system's performance.

Jebur et al. (2024) researched the detection of anomalies in surveillance videos using deep learning frameworks. This study presents a novel deep learning framework that consists of three essential elements: multi-task classification to generalise the classifier across multiple tasks without training from scratch when a new task is introduced; model fusion to improve feature representation; and transfer learning to improve feature generalisation. The primary benefit of the framework is that it can generalise without needing to be retrained for every new activity. The efficacy of the framework is demonstrated empirically, with a single classifier without retraining reaching an accuracy of 97.99% on the RLVS dataset (violence detection), 83.59% on the UCF dataset (shoplifting detection), and 88.37% across both datasets. Furthermore, the framework demonstrated an accuracy of 87.25% when evaluated on an unknown dataset. Table 2 summarizes recent studies for shoplifting detection using deep learning.

Table 2. Recent studies on shoplifting detection, descriptions, strengths, and weaknesses

Methods	Descriptions	Strengths	Weaknesses	Results obtained	Reference
BILSTM + Inception V3	Developed a shoplifting dataset and tested the performance of the dataset using the integration of Inception network V3 and bidirectional long short-term memory (BiLSTM)	The method provides a mechanism to analyze videos from security cameras to detect instant theft. Generate alerts and notifications to security personnel for prompt intervention.	Dataset limited to 900 instances; potential for larger-scale testing	Accuracy: 81%;	Muneer et al. (2023)
PoseLift dataset and unsupervised anomaly detection approach.	An anomaly detection problem that focused on the identification of deviations from typical shopping patterns	Incorrect flagging of legitimate frames as shoplifting incidents leads to high false positives.	Moderate AUC-ROC values; room for improved accuracy	STG-NF: AUC-ROC 67.46%, AUC-PR 84.06%, EER 0.39	Rashvand et al. (2025)
CNN-RNN hybrid; two parallel streams using HOG and optical flow	A hybrid model to extract local features from video segments and process the temporal sequence of the video	Enable the extraction of temporal and spatial features from video segments.	Limited dataset variety; relies on UCF-Crime only	Accuracy: 92%	Kirichenko et al., (2022)
Generalized deep learning using transfer learning and multi-task transfer learning	Developed a transfer learning approach to improve feature generalization and	Deep learning fusion algorithms that extract robust features and are trained on multiple data samples.	Complexity may limit real-time deployment	Accuracy: RLVS 97.99%, UCF 83.59%, Combined 88.37%, Unknown dataset 87.25%	Jebur et al., (2024)

Methods	Descriptions	Strengths	Weaknesses	Results obtained	Reference
YOLOv5 + Deep Sort; temporal feature extraction	A deep convolutional neural network model to extract temporal features from shoplifting incidents.	The proposed approach can track individuals across multiple frames in video sequences	Focused on inference speed; generalization across datasets unclear	F1 score 92%; 8.45× faster than RTFM	Nazir et al. (2023)
3D CNN (similar architecture as above)	Implement three-dimensional (3D) convolutional neural networks to extract behavioural features from videos to detect suspicious behaviour.	The method provides a mechanism to detect shoplifting crime before it occurs using behavioural features	Binary classification may not handle multi-class anomalies	Accuracy >90%; low false positives	Martínez-Mascorro et al., (2020)
InceptionV3 for feature extraction + optimized LSTM	Developed a dual-stream fusion deep neural network to extract temporal features from shoplifting actions	Integration of Inception V3 and Long Short-Term Memory to extract spatial and salient motion information that provides accurate actions with the least loss	Tested on a synthesized dataset; real-world performance uncertain	Accuracy 89.36%	Ansari and Singh (2023)
CNN + GRU hybrid	Developed a convolutional neural network and a gated recurrent unit to extract sequences of video segments	The approach is fast with provision for real-time implementation.	Dataset limited to UCF-Crime; generalization not reported.	Accuracy 93%	Kirichenko et al. (2022b)

Challenges of Deep Learning Methods for Shoplifting Detection and Solutions

It was found that the review of the available literature demonstrated that deep learning methods have contributed immensely to the formation of automated shoplifting detection frameworks in stores. There exist different models, such as CNNs, RNNs, CNN-LSTM models, 3D CNNs, and privacy-aware pose-based approaches that have proved to be successful in the detection of suspicious activities in surveillance video. Although some studies displayed high accuracy and detection rates, including benchmark databases such as UCF-Crime, interesting differences arise in the implementation of the methods, the datasets considered, and the system's performance. The majority of the methods are interested in video-based analysis based on the use of spatial and temporal aspects, and a

certain part with the use of tracking algorithms and anomaly detection procedures. Nonetheless, various limitations remain among the studies, mainly regarding dataset variety, real-time usefulness, model interpretability, and information protection, and there is more innovation to be done to make such systems practical, ethical, and efficient in the retail environment. These challenges and possible solutions are summarized in Table 3.

Stand-Alone Deep Learning Models

The implementation of single deep learning models such as convolutional neural networks, long short-term memory, and gated recurrent units individually comes with their challenges (Nweke et al, 2018). These include computational efficiencies, lack of generalizations, inability to handle spatial data, regular parameter updates, and vanishing gradients. Therefore, studies have begun to use hybrid models by integrating Convolutional Neural Networks (CNN) with Recurrent Neural Networks (RNN) models, such as LSTMs and GRUs. The implementation has continued to perform better when compared to standalone models. BiLSTM, CNN-GRU, and CNN-LSTM are examples of hybrid models that combine spatial and temporal information in order to improve the accuracy of detecting shoplifting by approximately 93 or 81% (Kirichenko et al., 2022b).

Modelling Temporal Feature Extraction

Modelling of the time with time-series and features of tracking has become one of the crucial issues in the process of identifying the suspicious behaviour. Methods that included object detection together with Deep Sort tracking (Nazir et al., 2023) and 3D CNNs demonstrated enhanced detection skills that were attributed to the ability to capture mouth and behavioural patterns through time.

Privacy-Preserving Real-World Datasets

Deep learning models implicitly memorize details of the training data, thereby leading to privacy issues. Privacy-preserving approach that implements privacy-safe data, such as PoseLift, which anonymizes pose information to maintain the behavioural cues, was recently developed (Rashvand et al., 2025). This answers ethical issues and regulatory demands and permits training of models on in-the-wild surveillance videos.

Shoplifting Detection Real-Time Deployment

Most of these models have high accuracy on test benchmarks, but it has been observed that, few of them handle the challenges of trying to implement such models in real-time scenarios, like inference speed, latency, and scalability of a system. A recent study by Nazir et al. (2023) is promising, as they demonstrate an impressive increase in speed (8.45x faster than previous methods), and this field is still insufficiently studied.

Generalization and Adaptability of the Deep Learning Models

Many of the deep learning models learned on one particular dataset (e.g., UCF-Crime) and are poorly generalized to other unseen environments. To address it, Jebur et al. (2024) suggested a multi-task learning system capable of re-learning to perform new surveillance tasks without requiring complete re-training, which points to one of the future trends in terms of more versatile and transferable solutions.

Limited Use of Explainability and Human-in-the-Loop Systems

Irrespective of the complexity of models implemented, most of the studies are not explainable or have machinery to help introduce human oversight to pass them through verification or interventions. Since the detection of shoplifting may be legally and ethically problematic, the involvement of explainable AI and user feedback may help increase confidence and responsibility.

Need for Standardization in Evaluation

In the literature, other databases (UCF-Crime, synthetic datasets, PoseLift), measures (accuracy, F1-score, AUC-ROC), and experimental conditions are used, so it is not possible to make a straightforward comparison across them. Standardization of benchmarking is required, along with publicly available databases specifically designed to measure the retail industry.

Table 3. challenges and solutions to deep learning for shoplifting detection

S/N	Challenges	Solutions	Reference
1	Computation inefficiencies, regular parameter updates, and the inability to handle spatial datasets by a single model.	Implementation of fusion models by integrating deep learning methods such as convolutional neural networks, long short-term memory, and gated recurrent units.	Ansari and Singh (2023)
2	Challenges of tracking temporal features for easy identification of suspicious behaviours	Capturing mouth and behavioural patterns in real-time	(Nazir et al., 2023)
3	Real-world data privacy	Introduction of PoseLift, which anonymizes pose information to maintain the behavioural cues.	(Rashvand et al., 2025)
4	Difficulty implementing real-time model deployment	Implementation of deep sort for temporal feature extraction	(Nazir et al., 2023)
6	Lack of generalization and adaptability of shoplifting	Implementation of a multi-task learning system capable	Jebur et al. (2024)

S/N	Challenges	Solutions	Reference
	detection-based deep learning models	of re-learning to perform new surveillance tasks without requiring complete re-training	
7	Limited Use of Explainability and Human-in-the-Loop Systems	Involvement of explainable AI and user feedback may help increase confidence and responsibility	Hoftman et al., 2024
8	Lack of benchmarked and standardized performance evaluation methods	Collections of more shoplifting scenes and videos, and making them public for the implementation of a shoplifting detection system.	(Nazir et al., 2023) (Rashvand et al., 2025)

Open Research Directions

Following the huge achievements made in the use of deep learning in retail surveillance, many research gaps still need to be filled. Many of the open research directions are related to model generalization, privacy, and model explainability.

- a. *Deep Learning generalization*: The greatest loss is the inability of existing models to be generalizable and real-time. The majority of reviewed systems are trained and tested on limited datasets, such as UCF-Crime or artificial video recordings, and they do not cover the variety of layouts, lighting conditions, and customer behavior in real-life retail settings. This restricts their flexibility in a dynamic real-life environment. Moreover, although hybrid architectures, such as CNN-LSTM or 3D CNNs, demonstrated high accuracies, they are either not optimized to be efficient in practice (i.e., real-time detection). Scaling and efficient, lightweight architectures and real-time inference pipelines will be required to be practical in surveillance systems installed in shopping stores.
- b. *Model explainability and privacy*: The other notable gaps are minimal attention given to privacy, explainability, and multimodal combination. Few works, e.g., PoseLift, including privacy-preserving data methods, exist to answer the question of the ethical consequences of full-time video surveillance. Most of the models are still black boxes and therefore highly accurate but not interpretable, which is a fatal problem when one expects the systems to assist legal or managerial decisions. They also fail to integrate with other complementary data sources, which the system can use to increase the accuracy and reliability of shoplifting being detected, such as RFID, audio cues, or transactional logs.
- c. *Dataset benchmarking and real-time systems*: In addition, methods of evaluating shoplifting systems are diverse, hence it is a challenge to compare practices or even standardize points of benchmarks. The way forward must be based on making these gaps through the development of transparent, privacy-aware, and context-

rich deep learning solutions, which are robust to deploy in the retail setting in real-time.

- d. *Interdisciplinary collaborations*: The development of efficient and operational deep learning-based shoplifting detection requires interdisciplinary collaboration between the key stakeholders, such as computer scientists, ethicists, and retail practitioners. These stakeholders build the neural network architecture, address potential privacy and algorithmic bias, and define the operational requirements, respectively (Lepri et al, 2021; Yogi et al, 2024). Further research areas could look into how the collaboration of these stakeholders will improve user privacy, the collection of ground truth data, and the implementation of more explainable artificial intelligence techniques for shoplifting detection.

Conclusion

Deep learning techniques have proven to be promising when it comes to monitoring shoplifters in a retail setting, and models like the CNNs, RNNs, and even hybrid models have produced high accuracy in detecting suspicious behaviours in surveillance videos. Reviewed studies demonstrate the possibilities of temporal feature extraction, pose estimation, and optimized neural networks in improving the performance of systems of shoplifting detection. These innovations have made it possible to create automation, intelligent surveillance systems that may help to curb cases of theft in the retail industry, enhance security in shops, and help curtail profits. Moreover, the increased popularity of creating benchmark datasets and the use of video analytics has created a solid base in terms of further progress in the area. Nonetheless, there are also numerous gaps identified in the review that should be filled in so that the implementation of these systems could be used successfully in practice in real-life retail surroundings.

One of the major setbacks hindering the pursuit of real-time detection includes challenges such as unmet performance expectations, privacy concerns, lack of explainability, and disagreement over evaluation standards. These remain some of the most persistent barriers. Future research should focus on developing lightweight, explainable models that preserve privacy, are computationally efficient, and are capable of utilizing multimodal data sources, making them suitable for deployment on edge devices. Moreover, the development of uniform, open sets of data and measures of evaluation will be required as a means of uniform benchmarking and a fair comparison of the proposed solutions. By closing such gaps, not only is the technical robustness of shoplifting detection systems improved, but the levels of trust and acceptance of all involved actors (both retailers and customers) will be elevated too. As deep learning methods are increasingly being implemented for the security of retail environments, practical implications of this implementation include prevention of theft, fraud reduction, and efficient monitoring of retail environments.

Conflict of Interest

The author declares that there is no conflict of interest regarding the publication of this manuscript.

References

- Abdolrasol, M. G. M., Suhail Hussain, S. M., Ustun, T. S., Sarker, M. R., Hannan, M. A., Mohamed, R., Ali, J. A., Mekhilef, S., & Milad, A. (2021). Artificial neural networks-based optimization techniques: A review. *Electronics*, 10(10), 2689. <https://doi.org/10.3390/electronics10102689>
- Adnan, K., & Akbar, R. (2019). Limitations of information extraction methods and techniques for heterogeneous unstructured big data. *International Journal of Engineering Business Management*, 11, 1847979019890771.
- Alsaiani, A., & Ilyas, M. (2025). A hybrid CNN-LSTM deep learning model for intrusion detection in smart grid. *arXiv preprint arXiv:2509.07208*. <https://arxiv.org/abs/2509.07208>
- Ansari, M. A., & Singh, D. K. (2022). An expert video surveillance system to identify and mitigate shoplifting in megastores. *Multimedia Tools and Applications*, 81(16), 22497–22525.
- Ansari, M. A., & Singh, D. K. (2023). Optimized parameter tuning in a recurrent learning process for shoplifting activity classification. *Cybernetics and Information Technologies*, 23(1), 141–160. <https://doi.org/10.2478/cait-2023-0008>
- Arroyo, R., Yebes, J. J., Bergasa, L. M., Daza, I. G., & Almazán, J. (2015). Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls. *Expert Systems with Applications*, 42(21), 7991–8005.
- Benthall, S., & Haynes, B. D. (2019). Racial categories in machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 289–298).
- Bhuiyan, M. R., Abdullah, J., Hashim, N., & Al Farid, F. (2022). Video analytics using deep learning for crowd analysis: a review. *Multimedia Tools and Applications*, 81(19), 27895–27922.
- Binetti, M. S., Massarelli, C., & Uricchio, V. F. (2024). Machine Learning in Geosciences: A Review of Complex Environmental Monitoring Applications. *Machine Learning and Knowledge Extraction*, 6(2), 1263–1280. <https://doi.org/10.3390/make6020059>
- Buet-Golfouse, F., & Utyagulov, I. (2022). Towards fair unsupervised learning. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1399–1409).
- Capital One Shopping. (2024). <https://www.capitalone.com/> Accessed: November 20, 2024.
- Chervonenkis, A. Y. (2013). Early history of support vector machines. In *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik* (pp. 13–20). https://doi.org/10.1007/978-3-642-41136-6_2
- Choudhry, N., Abawajy, J., Huda, S., & Rao, I. (2023). A comprehensive survey of machine learning methods for surveillance videos anomaly detection. *IEEE Access*, 11, 114680–114713.
- Dramsch, J. S. (2020). 70 years of machine learning in geoscience in review. *Advances in Geophysics*, 61, 1–55. <https://doi.org/10.1016/bs.agph.2020.05.001>
- Guan, Y., & Plötz, T. (2017). Ensembles of deep lstm learners for activity recognition using wearables. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 1(2), 1–28.
- He, H., Ming, Z., Zhang, J., Wang, L., Yang, R., Chen, T., & Zhou, F. (2024). Robust estimation of landslide displacement from multitemporal UAV photogrammetry-derived point clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 6627–6641. <https://doi.org/10.1109/JSTARS.2024.XXXXX>
- Hirschorn, O., & Avidan, S. (2023). Normalizing Flows for Human Pose Anomaly Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 13545–13554).
- Hofman, O., Giloni, A., Hayun, Y., Morikawa, I., Shimizu, T., Elovici, Y., & Shabtai, A. (2024).

- X-detect: Explainable adversarial patch detection for object detectors in retail. *Machine Learning*, 113(9), 6273–6292.
- Islam, M. S., Sultana, S., Kumar Roy, U., & Al Mahmud, J. (2020). A review on video classification with methods, findings, performance, challenges, limitations and future work. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, 6(2), 47–57.
- Issa, A., & Aldair, A. (2022). Learning the quadruped robot by reinforcement learning (RL). *Iraqi Journal for Electrical and Electronic Engineering*, 18(1), 1–8.
- Jabur, S. A., Alzubaidi, L., Hussein, K. A., Hoomod, H. K., Saihood, A. A., & Gu, Y. T. (2024). A scalable and generalized deep learning framework for anomaly detection in surveillance videos. Department of Computer Sciences, University of Technology, Baghdad, Iraq.
- Jia, Y., Anida Abdullah, N., Eliza, H., Lu, Q., Si, D., Guo, H., & Wang, W. (2025). A narrative review of deep learning applications in sports performance analysis: current practices, challenges, and future directions. *BMC Sports Science, Medicine and Rehabilitation*, 17(1), 249.
- Kang, B., Chen, X., Lai, S., Liu, Y., Liu, Y., & Wang, D. (2025, April). Exploring enhanced contextual information for video-level object tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 4, pp. 4194–4202).
- Khalidi, M. I., Erraissi, A., Hain, M., & Banane, M. (2025). Comparative analysis of supervised machine learning classification models. In *Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment* (pp. 321–326). Springer. https://link.springer.com/chapter/10.1007/978-3-031-88304-0_44
- Kirichenko, L., Pichugina, O., Sydorenko, B., & Yakovlev, S. (2022a). Recognition of shoplifting activities in CCTV footage using the combined CNN-RNN model. In *Progress in Polish Artificial Intelligence Research* (Vol. 4). <https://doi.org/10.34658/9788366741928.8>
- Kirichenko, L., Radivilova, T., Sydorenko, B., & Yakovlev, S. (2022b). Detection of shoplifting on video using a hybrid network. *Computation*, 10(11), 199.
- Kirichenko, L., Sydorenko, B., Radivilova, T., & Zinchenko, P. (2022c). Video surveillance shoplifting recognition based on a hybrid neural network. In *2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT)* (pp. 44–47). IEEE.
- Kumar, N. (2017, December). Machine intelligence prospective for large scale video based visual activities analysis. In *2017 Ninth International Conference on Advanced Computing (ICoAC)* (pp. 29–34). IEEE.
- Kushwaha, A., Ravish, K., Lamba, P., & Kumar, P. (2025). A survey of safe reinforcement learning and constrained MDPs. *arXiv preprint arXiv:2505.17342*. <https://arxiv.org/pdf/2505.17342>
- Liang, X., Ouyang, Y., Yang, H., & Zhang, Y. (2024). RL-OPC: Mask optimization with deep reinforcement learning. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 43(5), 1123–1135.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436–444.
- Liu, S., Yuan, H., Hu, M., & Zhang, J. (2024). RL-GPT: Integrating reinforcement learning and code-as-policy. *arXiv preprint arXiv:2403.11245*.
- Manasa, S. (2025). A comparative study of supervised learning algorithms for real-time classification tasks. *International Journal of Future Machine Research*, 4(49854), 1–12. <https://www.ijfmr.com/papers/2025/4/49854.pdf>
- Martínez-Mascorro, G. A., Abreu-Pederzini, J. R., Ortiz-Bayliss, J. C., Garcia-Collantes, A., & Terashima-Marín, H. (2021). Criminal intention detection at early stages of shoplifting cases by using 3D convolutional neural networks. *Computation*, 9(2), 24. <https://doi.org/10.3390/computation9020024>

- Martínez-Mascorro, G. A., Abreu-Pederzini, J. R., Ortiz-Bayliss, J. C., & Terashima-Marín, H. (2020). Suspicious behavior detection on shoplifting cases for crime prevention by using 3D convolutional neural networks. arXiv preprint arXiv:2005.02142. <https://arxiv.org/abs/2005.02142>
- Muneer, I., Saddique, M., Habib, Z., & Mohamed, H. G. (2023). Shoplifting detection using a hybrid neural network CNN-BiLSTM and development of a benchmark dataset. *Applied Sciences*, 13(14), 8341. <https://doi.org/10.3390/app13148341>
- Nagpal, S., Singh, M., Singh, R., & Vatsa, M. (2019). Deep learning for face recognition: Pride or prejudice? arXiv preprint arXiv:1904.01219.
- Nazir, A., Mitra, R., Sulieman, H., & Kamalov, F. (2023). Suspicious behavior detection with temporal feature extraction and time-series classification for shoplifting crime prevention. *Sensors*, 23(13), 5811. <https://doi.org/10.3390/s23135811>
- Noghre, G. A., Pazho, A. D., Katariya, V., & Tabkhi, H. (2023). Understanding the challenges and opportunities of pose-based anomaly detection. In *Proceedings of the 8th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence* (pp. 1–9).
- Noghre, G. A., Yao, S., Pazho, A. D., Ardabili, B. R., Katariya, V., & Tabkhi, H. (2024). PHEVA: A privacy-preserving human-centric video anomaly detection dataset. arXiv preprint arXiv:2408.14329.
- Nweke, H. F., Teh, Y. W., Al-Garadi, M. A., & Alo, U. R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications*, 105, 233–261.
- Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: Classification and comparison. *International Journal of Computer Trends and Technology*, 48, 128–138. <https://doi.org/10.14445/22312803/IJCTT-V48P126>
- Pacchiano, A., Saha, A., & Lee, J. (2021). Dueling RL: Reinforcement learning with trajectory preferences. *Proceedings of the 38th International Conference on Machine Learning*, 139, 8325–8335.
- Pandey, A., Mannepalli, P. K., Gupta, M., Dangi, R., & Choudhary, G. (2024). A deep learning-based hybrid CNN-LSTM model for location-aware web service recommendation. *Neural Processing Letters*, 56, Article 234. <https://doi.org/10.1007/s11063-024-11687-w>
- Pazho, A. D., Neff, C., Noghre, G. A., Ardabili, B. R., Yao, S., Baharani, M., & Tabkhi, H. (2023). Ancilia: Scalable intelligent video surveillance for the artificial intelligence of things. *IEEE Internet of Things Journal*, 10(17), 14940–14951.
- Perumal, S., & Ganesh, M. K. (2025). Shoplifting detection system using YOLOv5, MediaPipe pose estimation, and machine learning-based behavior analysis. *International Journal of Advance Research in Multidisciplinary*, 3(2), 297–301. <https://doi.org/10.5281/zenodo.15590304>
- Trung Pham, D., & Lukowski, F. (2024). BobDontRob: Hybrid Neural Network for Shoplifting Detection: Detecting Theft from Camera Footage in Grocery Retail Stores.
- Ragnoli, M., Esposito, P., Stornelli, V., Barile, G., Santis, E. D., & Sciarra, N. (2023, September 22–24). A LoRa-based wireless sensor network monitoring system for urban areas subjected to landslide. In *Proceedings of the 2023 8th International Conference on Cloud Computing and Internet of Things* (pp. 91–97). Okinawa, Japan. <https://doi.org/10.1109/CCIT.2023.XXXXX>
- Rashvand, N., Noghre, G. A., Pazho, A. D., Yao, S., & Tabkhi, H. (2025). Exploring Pose-Based Anomaly Detection for Retail Security: A Real-World Shoplifting Dataset and Benchmark. In *Proceedings of the Winter Conference on Applications of Computer Vision* (pp. 1123–1131). <https://doi.org/10.1109/WACVW65960.2025.00125>
- Singh, A., Thakur, N., & Sharma, A. (2016, March 16–18). A review of supervised machine learning algorithms. In *Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. New Delhi, India.

- Sultani, W., Chen, C., & Shah, M. (2018). Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6479–6488).
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
- Tyagi, K., Rane, C., Sriram, R., & Manry, M. (2022). Unsupervised learning. In *Artificial Intelligence, Machine Learning and Edge Computing* (pp. 33–52). <https://doi.org/10.1016/B978-0-323-89832-2.00002-6>
- Umut, Ö., Atasoy, Y., Seskir, A., & Ateş, H. F. (2025). Shoplifting detection from customer behavior using deep learning. *Signal, Image and Video Processing*, 19, Article 988. <https://link.springer.com/article/10.1007/s11760-025-04543-4>
- Waradpande, V., Kudenko, D., & Khosla, M. (2020). Graph-based state representation for deep reinforcement learning. *Proceedings of the 16th International Workshop on Mining and Learning with Graphs (MLG '20)*, 1–8.
- Woods, W. (2021). RL-GRIT: Reinforcement learning for grammar inference. *IEEE Security and Privacy Workshops (SPW)*, 2021, 1–6.
- Yogi, M. K., & Mundru, Y. (2024). Illuminating Evidence: Explainable AI in Cyber Crime Investigations. In *Redefining Security With Cyber AI* (pp. 195-229). IGI Global.