WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI For Smart Surveillance: Enhancing Security with Real-Time Human Detection

Rupalika Kumari

Department of Computer Science Engineering, Student of Computer Science Engineering, Arya College of Engineering and IT, Kukas, Jaipur.

ABSTRACT :

A more efficient technique to guarantee safety and security in a variety of settings is through video surveillance, also known as closed circuit television (CCTV). It is frequently employed in strategic sectors, including security at home, public transportation, banks, and ATMs' hubs, commercial districts, airports, and public roadways, and it is crucial for safeguarding crucial infrastructures. Due to the numerous uses, human detection in surveillance system video scenes has therefore grown in prominence in recent years. Objects of interest should be able to be found, categorized, and tracked by a real-time video surveillance system. This study provides an in-depth analysis of such video surveillance systems and presents a full assessment of methods and data sets utilized in human (object) detection. The most significant analyses of these systems are provided along with the employed architectures. To provide a clearer image and a comprehensive overview of the system, existing surveillance systems were compared in terms of their features, advantages, and challenges. These comparisons are summarized in this document. Future trends are also examined, laying the groundwork for new study avenues.

Keywords: Video Surveillance, Safety and Security, Real-Time Monitoring, Object Detection

1. INTRODUCTION

Video/Image surveillance systems has gained popularity and increased tremendously in recent years. Due to fast expanding population, growing incidents of theft, domestic crime, and acts of terrorism security concerns is also growing. As a result of this, residential societies, commercial and public spaces, and government and private organizations are using CCTV systems to monitor various malicious activities for security and safety purpose. The definition of surveillance is "Sur" means "from above", and "veiller" means

"to keep an eye on", Surveillance is the monitoring of people's movements, activities, and behaviours to manage, control, and protect them. Remote and continuous monitoring is beneficial to surveillance systems. Closedcircuit television (CCTV) technology is used to view events live and later monitor activities in any location. Because of the increasing number of thefts and criminal activities, CCTV cameras must be used.

Non-IP (Internet Protocol) CCTV cameras, IP CCTV cameras, and wireless CCTV cameras are the three types of CCTV cameras. Because of advanced technological features like flexibility, usability, and affordability, the use of IP-based wireless CCTV cameras is becoming popular in the current scenario. In general, in conventional video surveillance system, there is a need for human resources to monitor CCTV cameras 24 h a day, making ongoing monitoring costly and human monitoring may remain unaware of many events or give false conclusion.

The advantage of IoT and artificial intelligence technology is that it provides new vision to emerging real-time video surveillance systems.

Surveillance cameras can record the video at the moment in real time and display it afterwards for error prone operations, violence detection, intruders, and robbery. It is much desired that cameras automatically discover the ill activities and ration the same to various buildings for immediate required action.

The detection of motion and objects is a vital step in any video surveillance system. There have been four generations of video surveillance systems, which can be described in a

Figure 1.1

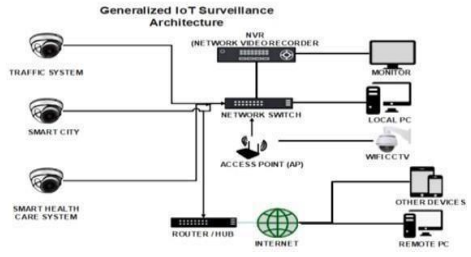


Figure 1.1 : Generalized IOT Surveillance Architecture

BACKGROUND

Video surveillance systems, commonly referred to as closed-circuit television (CCTV), have seen a rapid evolution over the decades, driven by advancements in technology and increasing security demands. The core function of these systems is to enhance safety and security by enabling real-time monitoring, detection, and recording of activities in various environments. The utility of CCTV has become indispensable in combating issues such as theft, vandalism, terrorism, and other malicious activities.

Evolution of Video Surveillance Systems

The progression of video surveillance can be categorized into four distinct generations, each marked by technological advancements that have significantly improved their efficiency and functionality.

Current Trends and Technologies:

Modern surveillance systems are shifting towards intelligent automation with an emphasis on real-time threat detection and minimal human intervention. The following innovations have transformed traditional CCTV systems into advanced security solutions

Benefits of Modern Surveillance Systems:

1.Reduce Human Intervention: Automated detection minimizes the need for continuous human monitoring, lowering operational costs.
2.Real-Time Monitoring and Alerts: Immediate notification of unusual events allows for prompt action, enhancing security.
3.Scalability and Flexibility: Wireless IP cameras and cloud storage make it easier to expand and customize surveillance systems based on specific needs.
4. Enhanced Accuracy: Advanced algorithms reduce false positives and improve the reliability of alerts and analysis.

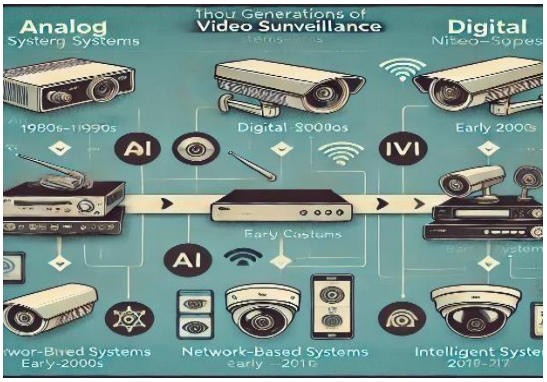


Figure2.1

2.1 Challenges and Contribution/Solution

1. **Data Loss**: In today's life as the surveillance systems are so heavily relied upon, data loss due to malfunction or other external events may have negative effects. Security camera recordings must be well documented and secure for both residential and business use to prevent information loss.

Solution: Of course, backing up the data on a cloud-based system is the most straightforward solution to this.
problem. There will be a backup copy for law enforcement to examine if the camera or backup disc is compromised. With the use of a cloud backup service, data loss was avoided.

2. **Operator issues:** Because the people in charge of monitoring video surveillance are just human, there is a chance that criminals may sneak into their blind spots or attack when operators are not paying attention. In certain situations, video surveillance may be unable to reliably record movement and monitor crimes as they take place.

Solution: Using AI technology may find solution for this. According to Forbes, artificial intelligence is shaping surveillance controls. AI technology can map and track events. When combined with self-learning capabilities, AI can improve human operations and analysis video footage in real time, even alerting concerned authorities or police if suspicious occurrences are detected.

3. **Power failure**: when power cut happens, the surveillance systems can work on WIFI, battery backup for many hours, but there are chances the captured video may not be clear and have to be compromised and may fail due to operational in efficiencies.

Solution: Keeping a security camera connected to a backup network or another power source can keep it operational in the event of a power outage. This problem can be resolved in some cases by connecting to electrical or solar power or by gaining access to LTE coverage.

2.2 Advantages of Video Surveillance System:

Continuous real-time monitoring: Enables an authorized individual to continuously monitor and see the activities of a person's important areas, restricted areas, or suspicious zones. **Remote video monitoring:** You can monitor the activity on your surveillance feeds from anywhere in the world if you have an Internet connection and a highly integrated digital video surveillance system. It also enables to observe on site behaviours, monitoring and managing the flow of human or vehicle traffic in remote areas, and also helps to visualize remote geographical regions, so that medical facilities can be provided in remote areas.
Increase security and safety: The use of surveillance systems in specific locations or restricted regions helps to monitor suspicious situations/activities, unwanted visitors/vehicles, and trace such individuals. It might be advantageous to stop crimes and break-ins.
Prevent dishonest claims: To refute claims of fraudulent information regarding loans for real estate, crimes, and inappropriate behaviour, surveillance system is the best option to decrease the fraud rates in a region.
Sort out variance: By providing clear visual evidence, surveillance system can resolve disputes in a smooth manner.
Visual evidence for investigations: Video/footage acts as a secondary proof for investigation to collect and utilize the invaluable visual evidence for investigations of a criminal activity.
Improve storage and accessibility: When compared to analogy video system, it uses stapes to store video footage where the amount of video data storage is decreased. However, the current video surveillance system uses digital video recorder (DVD) to store video data in large storage and can store multiple videos which can be viewed from any location and it is easy in accessible also. In today's current video surveillance system, the recorded video can be stored on cloud, network servers for easy access.
Reduce costs and scale more easily: The current system for video surveillance is cloud embedded, so the cloud infrastructure easily provides plenty of resources. As a result, the cost of processing/service is lower in cloudbased video surveillance systems, and scalability is increased more easily.

Pitfalls of Surveillance System

In real time, there are many kinds of risk factors where the video surveillance system needs to undergo like the following:
Static: Video monitoring system is static which leads to blind spot condition where the system cannot capture the footage, so the video cameras have to place in strategic location, so that it will cover the maximum portion of the location.
Invades the privacy: when the cameras are installed everywhere it means people are continuously watched which limits them from being themselves.

2. Related Work

Various methods have been used in tracking and detecting humans in videos. Some of the popular methods include HOG (Histogram of Oriented Gradients as feature extractor) with salience windowed frames of the video to the HOG, HOG application example of ATM Video Surveillance system with SVM Classifier, and NMS algorithm results better performance for human detection of accuracy 97% , a few more methods like CNN were used in real-time human detection and actions using optical flow along with CNN classifiers which results in success rate of detecting humans around 95–90%, one more method which had many updated versions is RCNN was Fast R-CNN and SPP-net which provided detailed experiments with results to improve the human detection quality, YOLO was one more method used in human and object detection in real videos and images, the further developments had a great impact in object detection fled, SSD is a brand-new, cutting-edge object detector that uses a sliding window approach rather than generated recommendations to categories all boxes at once. To find items of varying sizes in a single forward pass, SSD builds a scale pyramid. The SSD technique may be one of the most effective algorithms for object detection because of its speed and excellent accuracy. In one of the algorithms, Blitzen, at the same time plays object detection and semantic segmentation in a single step pass, permitting real-time computations. It suggests that object detection and semantic segmentations advantage from every other in terms of accuracy.

3. Object Detection

Human object detection is one of the most important tasks and is necessary in real video surveillance system as it is practically impossible for an individual to track the video clips in real time continuously for object tracking, human gait analysis, criminal behaviour, etc. The ability to identify moving objects in video streams is necessary for any video surveillance systems. Because of the increase in theft, terrorism, and other crime, human detection is an important field of study. It is beneficial to distinguish between moving objects as human or nonhuman in most applications, for example, security monitoring. The real detection process involves two stages: the first is the detection of a moving object, and the second is classification. Background subtraction, optical flow, or spatial temporal filtering methods may be employed for detecting objects, which may be classified depending on their motion, texture Ans shape as indicated in Fig. 2. 1.

Background subtraction: A common technique for identifying moving objects in

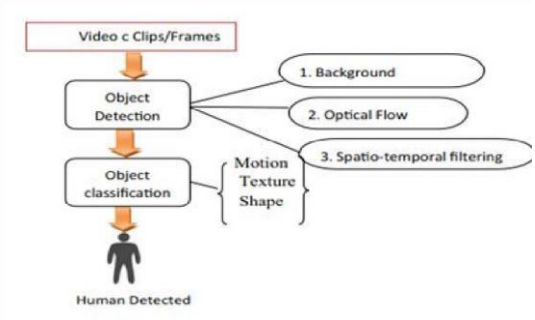


Figure 3.1

from static cameras is background removal. The fundamental idea behind this method is to identify moving objects by measuring the change between the current frame and the reference frame, often known as the "background image" or "background model". By isolating the foreground objects and comparing them to the frame with no objects present, this technique can be used to locate foreground objects. It will identify the differences between the objects and provide a distance matrix. In essence, it compares the threshold value to the difference in values of two frames, one without an object and the other with objects to count. Using the first few frames of the video, the threshold value is already established. Therefore, the result is tagged as a moving item (object detected) identified, if the difference between the values of two frames is greater than the threshold value that has been defined. Applications including object tracking, traffic monitoring, and human action detection systems all make use of the background subtraction technique (BSM).

The threshold value selection is important when using the background subtraction approach. There are numerous approaches to selecting a threshold value. Automatic Thresholding is the process of choosing a threshold value by hand. The BSM's primary flaw is its inability to adapt to abrupt changes in lighting and illumination. Therefore, attention should be used when choosing the threshold parameter.

Optical flow: This technique makes it possible to fully understand how an object moves and can be used to distinguish a moving target from the background. Even when the camera is moving, optical flowbased techniques can be utilized to find independently moving objects.

Optical flow offers a thorough description of the moving areas of the image as well as their speed.

Optical flow can be represented as 2D vector which represents a points displacement from first frame to second frame. Consider the above figure; it shows a moving ball in five consecutive frames and the arrow represents the displacement vector.

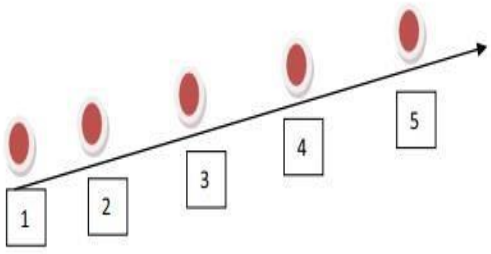


Figure 3.2

Spatio-temporal filtering: Spatiotemporal filtering approach to object detection is a method of perceiving the additional dimension of motion object information in a given space and even can provide more cues regarding a scene's structure, contents, and other high- or lowlevel information. This method is highly adjustable to varying scenes, employing the pixel-by-pixel difference of two or three consecutive frames within a video sequence, and temporal differencing attempts to detect moving object precisely. Human Gait motion spatial and temporal data can be captured better with spatiotemporal motion-based approach. Spatiotemporal-based methods are more precise

in the noisy areas. The methods are computationally efficient and yield promising results in applications where detection of rare events is involved.3.1 Object Detection

The goal of object detection is to develop computational models that provide the most fundamental information needed by video surveillance Applications [18, 19]. Human detection in a variant of object detection is used to detect human in video frames. For an object in motion to be recognized as a human, there are three main classification techniques, such as shape-based approaches, motion-based methods, and texture-based methods. These methods are discussed below:

1. **Motion-Based Method:** Based on the properties and patterns of object motion, motion may be used to identify people as well as distinguish between different forms of human movements including walking, sprinting, and skipping. Motion can be utilized as a metric to identify the object. This approach of categorization is predicated on the motion that the properties and patterns of item motion are distinctive enough to identify between objects [20]. Motion-based techniques leverage the periodic quality of the collected images directly to distinguish between people and other moving objects.

2. **Texture-Based Method:** Texture is an inherited property of all surfaces [21]. It offers crucial details regarding the structural configuration of surfaces and how they interact with their surroundings. Since the textural characteristics of visuals seem to contain useful information, features have long been calculated for textures for discriminating purposes. One of the methods used to detect texture in videos is Local Binary Pattern which quantifies intensity patterns in the neighbourhood of the pixel of a frame.

3. **Shape-Based Method:** Shape-based classification only focuses on a geometric of the object, not its structural analysis. Objects can be categorized based on the shape of the extracted portions, such as boxes, blobs, etc., which contain motion. Generally, for capturing articulation of human poses, human body is represented with six-part regions like head, pair of upper legs, pair of lower legs, and torso in geometric shapes; heads and torso are vertical rectangles; likewise, humans are detected in videos based on shapes. **3.2 Object Detection Methodology**

3.2.1 Histogram of Oriented Gradient (HOG)

One of the earliest techniques for object detection is the histogram of oriented gradients. A feature extractor is used by HOG to locate object in a given input image. The feature descriptor employed in HOG is a representation of a portion of an image from which we only take the most important details into account and ignore everything else. The feature descriptor's job is to transform the image's total size into an array or feature vector.
To locate the most important areas of an image in HOG, we employ the gradient orientation technique. Here is how HOG functions before we comprehend the overall architecture of it. The histogram of the gradient for a specific pixel in an image is computed by taking into account the vertical and horizontal values to get the feature vectors. The stages that HOG will take to get an image's gradient that is histogram oriented are as follows: **First,** it will consider a segment of an image with a particular size, and then, it divides the entire image into 8*8 cells. Now, 64 gradient vectors have been proved; from this, a split process for each cell happens into angular bins and the resultant histogram for a particular area of image is achieved.

3.2.2 R-CNN

R-CNN is an abbreviation for region-based Convolutional Neural Network. The R-CNN series is built around the concept of region proposals. To locate objects within an image, region proposals are used. We use selective search to extract only 2000 image regions. As a result, instead of attempting to classify a large number of regions, you can now work with only 2000 regions. The selective search algorithm described below is used to generate these 2000 region proposals. Selective Search:

1. Generate initial sub-segmentation and candidate regions.
2. Recursively combine similar regions into larger ones using the greedy algorithm.
3. Create the final candidate region proposal using the generated regions.

AS seen in image, R-CNN takes input image using selective search algorithm which selects only Region of interest (ROI) in rectangle form and then given to CNN networks to produce output features, and here, support vector machine classifiers are applied to decide which type of object is there within the region. Thus, object is detected successfully. One of the updating of RCNN object detection algorithm is Faster RCNN. "Fast R-CNN" is faster than R-CNN, because input to the convolutional neural network does not have to be 2000 region proposals every time.
However, the convolution operation is performed only once per image and a feature map is generated as a result. Furthermore, Faster R-CNN is an optimized form of R-CNN that is designed to improve computation speed (run RCNN much faster).

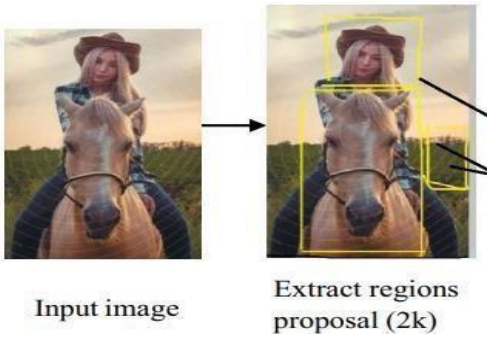


Figure 3.2.2

3.2.3 Spatial Pyramid Pooling (SPP-net)

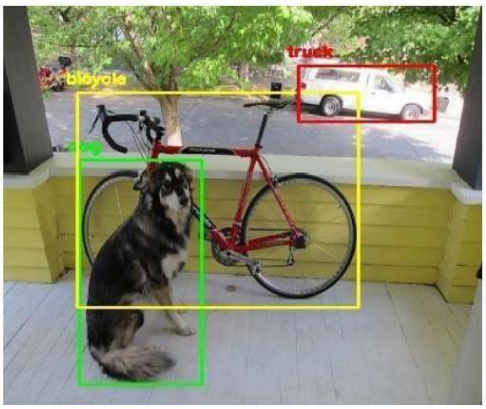
SPP-net is a network structure that can generate a fixed length representation regardless of image size/scale. Researchers can use SPP-net to compute feature maps from the entire image only once, and then pool features from arbitrary regions (sub-images) to generate fixed-length representations for training the detectors.
This method avoids computing the convolutional features repeatedly (Fig. 5). AS shown in the figure, before the fully connected layers, the SPPnet adds a spatial pyramid pooling layer. This spatial pyramid pooling layer pools areas separated by different scale grids to transform an arbitrary size feature map into a fixed size input. SPP-net, like R-CNN, requires multistep training on feature extraction and SVM classification to classify detected objects as human or nonhuman. The condition of having a fixed input image size from the R-CNN was lifted as an important feature of SPP-net! The network worked flawlessly regardless of image size, making

4. YOLO

YOLO is an abbreviation for the term 'You Only Look Once'. This is an algorithm that detects and recognizes various objects in a picture (in real time) (Fig. 6). There are many versions which have developed by scholars which are summarized here after the basic model YOLO, the next was YOLOV2 released on 2017, which was better and faster than YOLO, and it had better normalization and high resolution when worked large input images and small input images and accuracy in detecting the images were good.

Then, YOLO3 was introduced in 2018 which was enhanced using multi-scale features for detecting object real time. It uses 106 fully conventional under laying architecture, it uses feature graphs of three scales meaning that objects are detected by applying 1*1 kernels on feature maps of three different sizes at three

Finally, YOLOv7 (released in July 2022) is the fastest and most accurate real-time object detection model, providing great improved real time object detection accuracy without increasing the inference cost, resulting in high detection accuracy when compared to all YOLO versions.



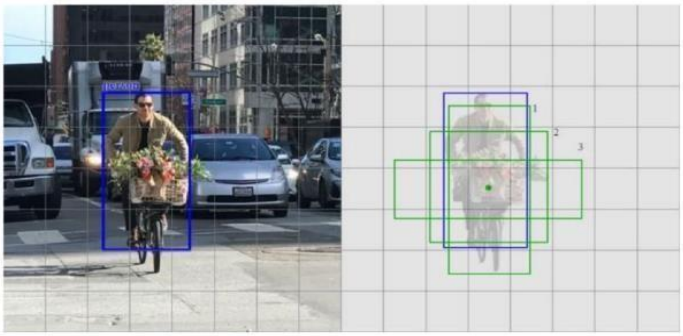
Figure

5. BLITZNET

A real-time network called Blitznet recognizes objects and semantically separates them. It works as an encoder and decoder network, with the encoder–decoder module processing the input image and creating feature maps in both the detecting objects of varying sizes.

Blitznet includes a multi fusion layer that aids in semantic segmentation of the given input image. Blitznet provides real-time segmentation and object bounding boxes. The computational cost is reduced by utilizing a single network to address both issues, and the two tasks object detection and segmentation produce good accurate results. **5.1 Single-Shot Detector (SSD)**

In, it is suitable for real-time object detection and fastest object detection algorithm; here, multiple objects are denoted of an single image as it is having property of multi-scale features; the frame per second is five times as much as the R-CNN, thus making the predication process of the detection object faster and it increases the speed and accuracy.



Crowd Management: Human detection can help manage public events, protests, or emergencies by tracking crowd densities and identifying risk situations.

2.IoT and Edge Computing Integration

IoT Ecosystems: Surveillance systems will interact with IoT devices to communicate and make decisions such as opening doors or setting off alarms if unauthorized access is detected.

Edge Computing: Deployment of models on edge devices, such as cameras, will reduce latency and diminish reliance on centralized servers. This will increase the scalability and responsiveness of the systems.

3. Improved Threat Detection

Behavioural Analysis: State-of-the-art models will predict threatening actions, including aggressive or loitering, before such actions present as threats.
Biometric Integration: Bringing face and gait recognition in conjunction with human detection enhances identity confirmation and threat appraisal.

6. Future Prospects of AI-Driven Surveillance with Real-Time Human Detection

AI-driven surveillance systems have a potential revolution in all walks of life and future prospects, much more beyond security. It is based on the progress in AI, sensor technology, and ethical frameworks. Some key future prospects are discussed below.

1.Smart Cities and Urban Management

Intelligent Traffic Monitoring: AI-based surveillance will automatically identify traffic violation patterns, congestion patterns, and pedestrian movement patterns in urban areas for optimized flow.

Retail Analytics: The use of these systems can be done by retailers in analysing customer footfall, shopping patterns, and dwell times to optimize layouts and marketing.

Agriculture: Human detection with activity recognition can be used in monitoring workforce efficiency and safety protocols compliance. **5.Legal and Ethical Evolution**

Data Privacy Frameworks: There will probably be a better adherence to privacy laws as the surveillance systems would have stricter requirements, and anonymization, among other things, will become a norm.

Ethical AI: The efforts will be put into developing AI systems that are unbiased, explainable, and in line with societal norms to earn public trust and fair deployment.

6. CONCLUSIONS:

In conclusion, AI-driven surveillance systems with real-time human detection signify a revolution in security technology and its applications. They may reshape how safety and monitoring are approached, with higher precision, scalability, and efficiency compared to traditional surveillance methods. Such a level of integration of AI into IoT with advanced tools like edge computing enables real-time, proactive solutions within a wide range of applications, from public safety to healthcare, retail analytics, and smart city management.

As these systems evolve, they are expected to expand into areas beyond traditional security, like traffic management, behavioural analysis, personalized home security, and autonomous surveillance. Their detection and prediction of threats, together with improved response times, make them indispensable in high-risk environments, critical infrastructure, and urban spaces. How ever the long-term success of AI driven surveillance is contingent upon the resolution of several challenges.

showing how there is no real difference in definition since being initially coined in 2012. It is also evident that some research wrongly identifies as models and shadows. Across the literature, there are examples of small-scale projects, but a lack of large-scale projects. One reason for this is the lack of domain knowledge on successfully scaling up larger. Papers concerning use in manufacturing identify a range of publications with particular growth in the health of the machines and predictive maintenance for healthcare draws on similar themes in terms of health status and monitoring, with a number of papers investigating use for predictive analytics of human users. The paper also highlights the advancements in remote surgery and the importance of researching data fusion, mainly due to the nature of sensitive data used in healthcare. Research for smart cities is limited, but the potential to investigate for traffic management systems and smart city developments is on the rise.

AI is becoming a component and exploring where these algorithms can be applied is another avenue of open research. The effects of AI combined with are topics amongst the publications but on a small scale. The exciting and inevitable future research will explore scaling up smaller successful and AI projects. An important finding is the lack of standardisation and misconceptions with definitions for. Addressing the challenges with standardisation ensures future developments are actually and not wrongly defined concepts.

This next section discusses the open research questions for in a healthcare setting. Some of the research cites the potential for adapting Digital technology for humans. An example is a of a person to monitor day-to-day health and wellbeing giving the potential for a human twin for simulating what positive and negative lifestyle changes could have on the physical huma.

REFERENCS:

1. Ashwin Karale,,"The Challenges of IoT Addressing Security, Ethics, Privacy, and Laws", Internet of Things, Volume 15, 2021, 100420, ISSN2542-6605,
2. M. O. Osifeko, G. P. Hancke and A. M. AbuMahfouz, "SurveilNet: in IEEE Sensors Journal, vol. 21, no. 22, pp. 25293-25306, 15 Nov.15, 2021
3. A. Singh and B. Sikdar, "Adversarialand Defence Strategies for Deep-LearningBased IoT Device Classification Techniques," in IEEE Internet of Things Journal, vol. 9, no. 4, pp. 260210.1109/JIOT.2021.3138541.
4. Wang, X., Zhang, D., & Guo, J. (2021). Deepfake Detection Using Reinforcement Learning. IEEE Transactions on Circuits and Systems for Video Technology.
5. Liu, Y., Xie, L., & Yang, Z. (2020). A Deep Reinforcement Learning Approach for Deepfake Video Detection. arXiv preprint arXiv:2012.07550.
6. Zhao, W., Chellappa, R., Phillips, P.Rosenfeld, A. (2003). Face recognition: A literature survey. ACM Computing Surveys (CSUR), 35(4), 399-458.
7. Jain, A. K., Ross, A. A., & Nandakumar, K. (2011). Introduction to biometrics. Springer Science & Business Media.
8. Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. British Machine Vision Conference (BMVC).

9. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. Proceedings of the IEEE
10. Li, Y., Yang, X., Sun, P., Qi, H., Lyu, S., & Wu, W. (2020). Celeb-DF: A New Dataset for DeepFake Forensics. arXiv preprint arXiv:1909.12962.
11. Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2020). Learning Rich Features for Image Manipulation Detection. Proceedings of the IEEE/CVF
12. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems,
13. Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks.
14. Korshunov, P., & Marcel, S. (2018). Deepfakes: a new threat to face recognition? Assessment and detection.
15. Liu, Y., Xie, L., & Yang, Z. (2020). A Deep Reinforcement Learning Approach for Deepfake Video Detection. arXiv preprint arXiv:2012.07550.
16. A. Singh and B. Sikdar, "Adversarial Attack and Defence Strategies for Deep-LearningBased IoT Device Classification Techniques," in IEEEInternet of Things Journal, vol. 9, no.