CHAPTER 8

CONCLUSION AND FUTURE WORKS

8.1 SUMMARY

The first contribution of this work was implementation and evaluation of CNN-YOLO hybrid model for Video Anomaly Detection (VAD). This model acquired peak accuracy of 99.46% and minimal Equal Error Rate (EER) of 2.1% making suitable it for faster and precision applications. Even though the model has significant classification accuracy but the low value of Peak Signal to Noise Ratio (PSNR) 23.34dB indicates poor image quality. This small model handles only a few image frames and does not process a full video sequence. Random frame selection of the model limits the capability to obtain temporal dependencies which inhibits the context aware anomaly detection.

The second contribution relayed on the hybrid ResNet-LSTM model, which was able to identify the spatio-temporal relationships among continuous video input frames. The ResNet-50 architecture efficiently captures the features and the LSTM architecture helps to efficiently identify the temporal dependencies from the extracted features. The model has a moderate classification performance, but has the probability of misclassification. Even though this model has consistent performance, higher EER of 14.5% and moderate PSNR reflects the requirement of further modifications.

The third part of this research focused on reducing overfitting by using an Improved UNet's segmentation capability. The Convolutional LSTM architecture deployed in the UNet helps to capture the spatio-temporal feature extraction effectively. The Wiener filter enables noise reduction of image frames which improves the model's performance. The hybrid IUNet-CSWT model yields good image quality with a PSNR of 35.04dB, making it apt for high-resolution applications. While the model has good classification performance, the Precision and Recall figures slightly lag characteristics.

The fourth contribution proposed is the Hierarchical Multiscale-CNN with LSTM, an excellent, efficient model for VAD. The model scores 99.35% Accuracy, Precision, Recall and F1 Score reveals its superior characteristics namely classification, image quality reconstruction, differentiating ability and reliability. The model recorded a PSNR of 42.18dB, shows high-quality video reconstruction capability and an Area Under the Curve (AUC) of 99.85%, conforming its stringent capacity to differentiate normal and anomalous events. The reduced EER of 5.4% enhances its reliability in minimizing false positives and false negatives.

The increased PSNR value expresses the model's ability to preserve or restore the original frame without distortion or noise. The improved AUC reflects the classification accuracy in distinguishing normal and anomalous events. The reduced value of EER indicates a better balance between false acceptance and false rejection rates. That helps in fewer misclassifications and improves the robustness of the system. These performance measures indicate that the hybrid models are well-suited for real-time anomaly detection and provide a remarkable contribution to VAD in surveillance systems.

The research provides a sound discussion of VAD which has a strong potential in social, urban, public safety and emergency applications. Automated monitoring enhanced safety and security, providing proactive measures for anomalous incident prevention. The improvements proposed in this work have vital importance in accurate, precise and real-time recognition of events for authorities.

8.2 POTENTIAL CHALLENGES

The potential challenges that might arise during the real-time implementation of the research:

The Hierarchical M-CNN fusion with LSTM model is effective for detecting video anomalies by learning both spatial and temporal features. However, several challenges may arise during its real-time implementation. A primary concern is the model's computational complexity. Due to its deep architecture and the requirement to process high-dimensional video data continuously, the model demands significant processing power and memory, limiting its applicability on low-power or legacy surveillance hardware.

The synchronization between spatial and temporal features can be problematic. Misalignment between frame-level spatial patterns and longer temporal dependencies may lead to inconsistencies in detection, especially during fast scene transitions or occlusions.

Another critical concern is the model's sensitivity to variations in real-world conditions. While the multiscale approach enhances spatial generalization, unexpected changes such as occlusions and unusual viewpoints may result in misclassification or detection failure.

Models with complex architectures, such as CSWT, may introduce processing delays, which are critical in real-time anomaly detection. Furthermore, integration into existing surveillance infrastructures can provide compatibility issues. Many traditional surveillance systems are not designed to support AI-based models, lacking the necessary computational resources or interface support required for seamless deployment.

8.3 LIMITATIONS OF THE RESEARCH

Major limitations of research:

1. **Real-Time Scalability:** Large-scale, multi-camera deployments require optimization to overcome computational challenges and ensure efficient real-time performance.
2. **Limited Real-world Testing:** The performance under real-world conditions like varying camera angles, lighting changes and occlusions is not fully understood.
3. **Sensitivity to Video Quality:** Despite using noise reduction techniques, the models are still affected by video issues like motion blur, low frame rates and poor resolution in CCTV systems.
4. **High Computational Requirements:** Deep architectures require significant computing power, which may limit it's use in low-power surveillance systems.
5. **Ethical Concerns:** The Privacy implication was not extensively addressed, highlighting the need for compliance frameworks and ethical guidelines.

8.4 FUTURE WORK

Future studies can be extended to explore on improving the accuracy and consistency of Anomaly Detection (AD) models by incorporating additional contextual information. Utilizing a versatile dataset, such as a combination of audio, video and sensor inputs, can offer an insight of anomalies in depth.

Expanding datasets to encompass a wider space of scenarios, activities and environmental conditions, such as night-time surveillance combined with progressive data augmentation techniques, can significantly improve model generalizability.

Research may also prioritize real-time scalability, explainability through Explainable AI (XAI) and privacy-compliant frameworks. Leveraging generative AI models, such as GANs or diffusion models, can help generate training data to address class imbalance and enhance model generalization.

Reducing error rates in diverse conditions is vital to enhancing Anomaly Detection reliability and advancing its applications in public safety, traffic monitoring and critical infrastructure protection.