CHAPTER 7
RESULTS AND DISCUSSIONS

7.1 RESULTS

The models explained in this research includes the CNN-YOLO model, ResNet-LSTM model, IUNet-CSWT model and Hierarchical M-CNN with LSTM model. All these models were trained and tested using UCSD dataset. More than thousands of video frames from dataset were used for training and 25 epochs were performed to reduce loss function and attained convergence by avoiding overfitting. Best results were obtained while using 80% data splits for training and 20% for testing. Table 7.1 presents the performance evaluation of these models, including Accuracy, Precision, Recall, F1 Score, AUC, PSNR and EER.

Table 7. 1 Performance Comparison of Four Models

Performance Metrics	CNN-YOLO	ResNet-LSTM	IUNet-CSWT	Hierarchical M – CNN with LSTM
Accuracy (%)	99.46	96.5	99	99.35
Precision (%)	99.24	96.5	97.3	99.35
Recall (%)	98.79	96.5	97.5	99.35
F1 Score (%)	99.26	96.5	97.5	99.35
AUC	0.9926	0.99	0.908	0.9985
PSNR (dB)	22.34	28.55	35.04	42.18
EER (%)	2.1	14.5	10.9	5.4

From the analysis, the Hierarchical M–CNN with LSTM model emerges as the most well-rounded solution, excelling in accuracy, image quality and classification robustness for VAD. The CNN–YOLO model demonstrates strong classification capabilities, attaining the maximum accuracy of 99.46% and lowest Equal Error Rate of 2.1%. However, its high accuracy is limited by its approach, as the model considers only 100 frames and does not undergo a video sequence evaluation, selecting a random frame for VAD. This frame-based evaluation restricts its ability to capture temporal dependencies, making it less effective in scenarios requiring context-aware anomaly detection. Additionally, the model exhibits poor image quality with the lowest PSNR of 22.34dB, indicating balance between classification accuracy and image fidelity.

The IUNet–CSWT model is ideal for tasks requiring high image quality, achieving a PSNR of 35.04dB, but slightly lags in classification metrics. While consistent, the ResNet–LSTM model requires further optimization to address its higher error rate, with an Equal Error Rate of 14.5% and its moderate overall performance.

Figure 7.1 presents a comparative analysis of Accuracy, Precision, Recall, and F1 Score for four models: CNN-YOLO, ResNet-LSTM, IUNet-CSWT, and Hierarchical M-CNN with LSTM, in comparison with the existing models EfficientNetB1 and EfficientNetB3 (Dilek & Dener, 2024).

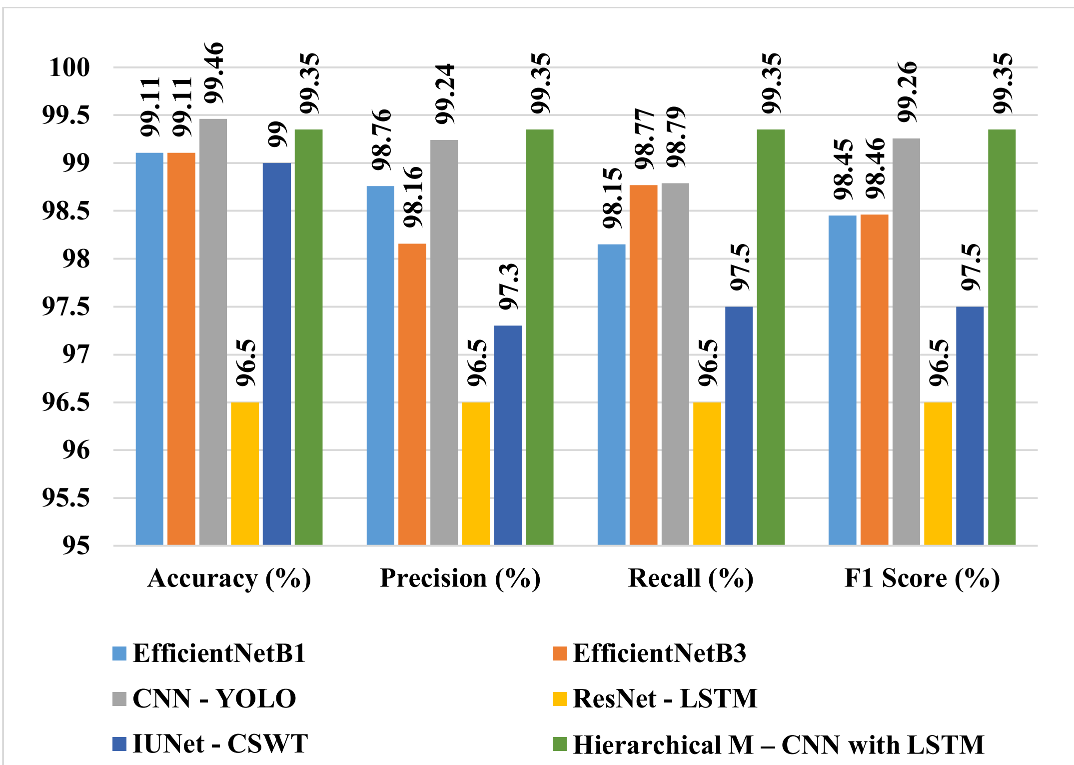

Figure 7.1 Comparison of Accuracy, Precision, Recall and F1 Score of Models

From the analysis of Figure 7.1, the performance of the proposed models in the research shows a better performance when compared with the existing models. The Hierarchical M-CNN with LSTM model presents the highest performance accuracy that shows the model's reliability. The higher Precision states the fewer false positive errors and increased recall states the model misses fewer true anomalies. The value of F1 Score conveys that the model exhibits a balance in performance.

Figure 7.2 presents a comparative analysis of AUC for four models: CNN-YOLO, ResNet-LSTM, IUNet-CSWT and Hierarchical M – CNN with LSTM, in comparison with the existing methods EfficientNetB1 and EfficientNetB3.

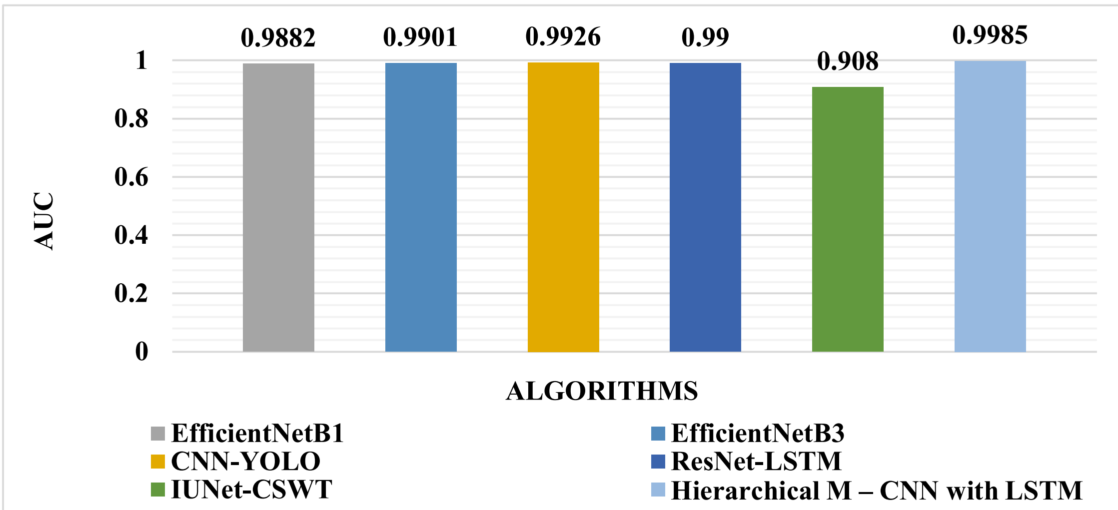

Figure 7.2 Comparison of AUC with Existing Models

The analysis of figure 7.2 shows the Hierarchical M-CNN with LSTM models present a high level of AUC that expresses the capability of correctly identifying the normal and anomalous events. From the AUC values taken from existing works the proposed models in the research finds a competing performance.

In addition to the four models, the Spatio Guard YOLO model was developed to process real-time video input datasets. This model integrates a Dynamic Contrast Enhancer (DCE) for improved visibility, YOLOv8 for robust object detection and a Spatially Expanded Neural Network (SENN) for detailed spatial feature extraction. Figure 7.2 shows the result of Spatio Guard YOLO model after preprocessing.

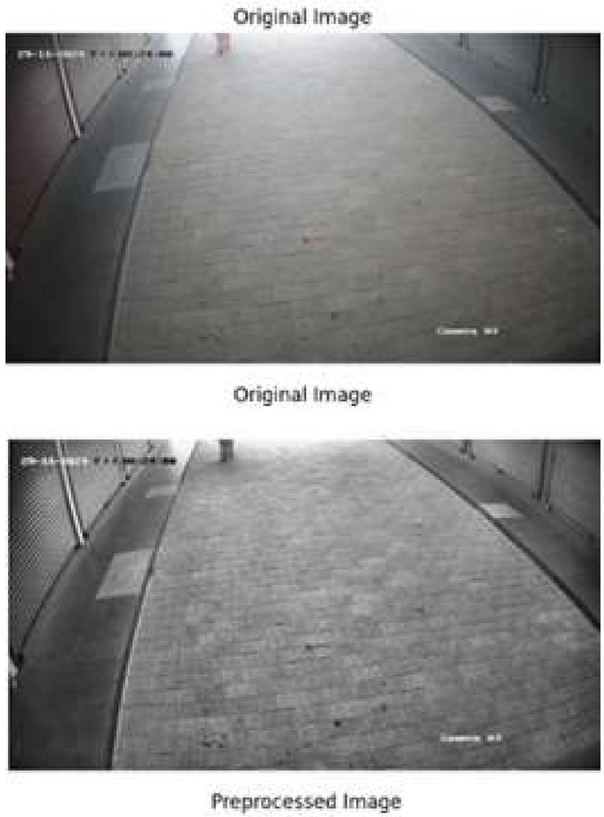

Figure 7.3 Original and Preprocessed Image for Real-Time Video

Figure 7.3 and Figure 7.4 provides the output of Spatio Guard YOLO for Normal and Anomalous events for Real-Time Video. It was evaluated on the UCSD and real-time video datasets, achieving superior performance in accuracy, AUC and EER.

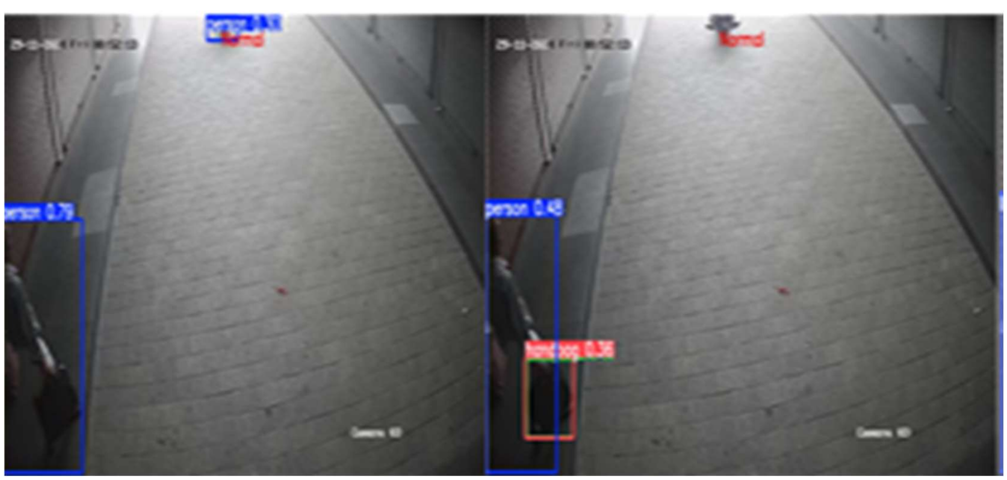

Figure 7.4 Normal Event Detection for Real-Time Video

Figure 7.5 Anomalous Event Detection for Real-Time Video

Table 7.2 exhibits a summary of the results of Spatio Guard YOLO Model for UCSD and Real-Time Video Data Set.

Table 7. 2 Performance Comparison of Spatio Guard YOLO Model for UCSD and Real-time Video Dataset

Performance Metrics	UCSD Dataset	Real-time Video Dataset
Accuracy (%)	96.5	99.81
Precision (%)	96.5	99.73
Recall (%)	96.5	99.73
F1 Score (%)	96.5	99.9
AUC	0.99	0.999
PSNR (dB)	23	69.17
EER (%)	12.8	2

Results of both datasets shows significantly improved performance for real-time dataset by achieving 99.81% Accuracy, 99.73% Recall, 99.73% Precision and 99.9% F1 Score, with a low EER of 2%. The essentiality of real-time VAD model with reliability and accuracy is observable from these results to ensure public safety.