Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 12, No. 4, December 2024, pp. 1074~1084
ISSN: 2089-3272, DOI: 10.52549/ijeei.v12i4.5845

Machine Learning-Driven Pre-Broadcast Video Codec Validation: Ensuring Seamless Television Transmission

Khalid El Fayq[1], Said Tkatek[2], Lahcen Idouglid[3]
[1,2,3]Computer Sciences Research Laboratory Ibn Tofail University Kenitra, Morocco

Article Info

Article history:

Received Aug 26, 2024
Revised Nov 14, 2024
Accepted Dec 23, 2024

Keywords:

Video Codec Errors
Machine Learning
Television Broadcasting
Video Compatibility
Real-time Detection
Broadcast Reliability

ABSTRACT

This study addresses the critical challenge of ensuring uninterrupted television broadcasting by proactively detecting video codec errors, focusing on TV Laayoune, a prominent Moroccan channel. We developed a machine learning-based methodology that identifies incompatible codecs before they disrupt live broadcasts. The approach involves data collection from multiple sources, including TV Laayoune's archives, metadata extraction via FFmpeg, and a hybrid model that combines Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks. Integrated into the broadcasting pipeline, this model achieved a 95% accuracy rate, significantly enhancing broadcast reliability and operational efficiency. Additionally, we propose a user-friendly interface for real-time error detection, comprehensive workflow integration, and automated alerts. This innovative solution addresses common broadcast challenges, reducing operational risks and improving the viewer experience.

Corresponding Author:

Khalid El Fayq
Computer Sciences Research Laboratory, Faculty of Science, Ibn Tofail University
Kenitra, Morocco.
Email: khalidelfayq@gmail.com

1. INTRODUCTION

Television broadcasting is vital for delivering content to a broad audience, and ensuring the reliability of video transmission is paramount. TV Laayoune, a prominent Moroccan television channel, frequently encounters disruptions due to incompatible video codecs, particularly during live broadcasts. These disruptions can result in dropped frames, audio-video sync issues, or even complete transmission failures, leading to a poor viewer experience and potential loss of audience trust.

Traditional methods for managing codec compatibility are manual, time-consuming, and prone to errors, which exacerbates the risk of broadcasting interruptions. To address these challenges, this study proposes an innovative machine learning-based approach that automates the detection of video codec errors before broadcasting, significantly enhancing the reliability and efficiency of television transmission.

The primary objective of this research is to develop a system that automatically identifies incompatible video codecs and alerts operators to take corrective action, thereby preventing live broadcast errors. This system is specifically tailored for TV Laayoune’s broadcasting infrastructure and involves several key steps: data collection, feature extraction using the FFmpeg multimedia framework, and the development of a robust classification model. The model leverages the strengths of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to capture both spatial and temporal features from the video metadata.

Integrated into TV Laayoune's broadcasting pipeline, the proposed system offers real-time error detection and alerting, ensuring that potential codec errors are identified and addressed before the clips are broadcasted live. This proactive approach not only improves the reliability and smoothness of television operations but also reduces the reliance on manual checks, thereby saving time and resources.

Experimental evaluations demonstrate that the model achieves a 95% accuracy in detecting codec errors, significantly outperforming traditional methods. This research highlights the potential of machine learning to enhance the operational efficiency and reliability of television broadcasting. By effectively addressing video codec incompatibility, this study offers a robust solution for ensuring a seamless and uninterrupted viewing experience, paving the way for more reliable and efficient media delivery.

2. LITERATURE REVIEW

Video codec errors and their impact on broadcasting quality have become increasingly relevant with the rise of digital media and satellite broadcasting. Researchers have been exploring various methods to address these issues, ranging from traditional heuristic approaches to advanced machine-learning techniques.

Traditional methods for detecting video codec errors often rely on manual checks and predefined rules. Heuristic algorithms use predefined rules and patterns to detect common codec errors, but their effectiveness is limited due to the dynamic and complex nature of video data. Signal processing methods analyze the video stream for anomalies in signal quality, such as artifacts and synchronization issues. While useful, these techniques often fall short in real-time detection scenarios and lack the adaptability required for modern broadcasting needs.

Izima et al. [1] surveyed machine learning techniques for video quality prediction, highlighting the limitations of traditional methods. Alexandre et al. [2] explored deep video compression for interframe coding, demonstrating the potential of deep learning in improving video processing. Muskaan et al. [3] investigated the use of LSTM and CNN for detecting deep fake videos, which is relevant for identifying anomalies in video streams. Kaur and Mishra [4] used LSTM for estimating concise video summaries from long sequences, emphasizing the importance of temporal analysis in video data.

Recent advancements in machine learning have paved the way for more sophisticated approaches to video codec error detection. Machine learning models, particularly those based on deep learning, offer the ability to learn from data and adapt to various error patterns.

Convolutional Neural Networks (CNNs) have been extensively used for video and image analysis due to their ability to capture spatial hierarchies in data. Yan et al. [5] utilized CNNs for fractional-pixel motion compensation, showcasing their effectiveness in video processing. Cui et al. [6] applied CNN-based post-filtering for compressed images and videos, demonstrating significant improvements over traditional methods. El Fayq et al. [7] focused on the detection and extraction of faces and text in audiovisual archives using machine learning, illustrating the versatility of CNNs.

Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, are well-suited for temporal data analysis. Bidwe et al. [8] provided a bibliometric analysis of deep learning approaches for video compression, emphasizing the role of LSTMs in handling temporal dependencies. Benoughidene and Titouna [9] developed a method for video shot boundary detection using a CNN-LSTM approach, highlighting the synergy between spatial and temporal feature extraction. Panneerselvam et al. [10] explored efficient video compression using deep learning techniques, underlining the benefits of combining CNNs and LSTMs.

Integrating machine learning models with multimedia frameworks like FFmpeg has shown promising results in real-time video processing. FFmpeg is a powerful multimedia framework capable of extracting detailed metadata from video files, including codec information, frame rates, and resolution. This metadata serves as valuable input features for machine learning models. Lu and Niu [11] used LSTM-CNN for object removal detection, which is relevant for anomaly detection in video streams. Yadav et al. [12] developed a video compression-cum-classification network, demonstrating the potential of integrating multiple deep-learning techniques. Wang et al. [13] provided a comprehensive survey on deep learning for image super-resolution, highlighting advancements in video quality enhancement.

Combining multiple machine-learning techniques can enhance detection accuracy. Hybrid models integrating CNNs for spatial feature extraction and RNNs for temporal analysis have shown superior performance in identifying complex error patterns. Bouaafia et al. [14] applied deep learning-based video quality enhancement techniques, achieving significant improvements in video processing. Chen et al. [15] developed RL-AFEC, a reinforcement learning-based adaptive forward error correction system, demonstrating the potential of advanced machine learning techniques for error detection.

Several studies have explored the practical applications of these advanced techniques in real-world broadcasting environments. Systems designed to proactively detect and alert operators of codec errors before broadcasting have been implemented in various settings, showing significant reductions in live broadcast disruptions. Darwich and Bayoumi [16] used CNN and RNN models for video quality adaptation, showcasing the benefits of machine learning in real-time video processing. Sharrab et al. [17] focused on the availability of video communication in AI-based computer vision systems, emphasizing the importance of robust error detection. Ciaparrone et al. [18] compared deep learning models for end-to-end face-based video retrieval, highlighting the practical applications of these techniques.

Research has consistently highlighted the importance of accuracy, precision, recall, and F1-score in evaluating the effectiveness of these models. Studies report improvements in these metrics when using machine learning-based approaches compared to traditional methods. Gashnikov [19] developed a video codec using machine learning based on parametric orthogonal filters, achieving high accuracy in error detection. Steinert and Stabernack [20] designed a low latency H.264/AVC video codec for robust ML-based image classification, demonstrating significant improvements in precision and recall.

Despite the advancements, challenges remain in ensuring the robustness and scalability of these systems. Ensuring the model is trained on a diverse dataset representing different codecs and broadcasting scenarios is crucial for generalizability. Liu et al. [21] provided a comprehensive review of deep learning-based video coding, emphasizing the need for diverse datasets. Achieving real-time detection without compromising accuracy requires optimizing both the machine learning models and the integration with multimedia frameworks. Ma et al. [22] reviewed image and video compression with neural networks, highlighting the importance of real-time processing. Zhang et al. [23] surveyed machine learning-based video coding optimizations, emphasizing the need for robust and scalable solutions. Xu et al. [24] developed a CNN-LSTM model for parking space detection, showcasing the potential for real-time applications in video analysis. In summary, traditional methods for video codec error detection have limitations. However, integrating machine learning offers a promising solution for enhancing the reliability of television broadcasting. Our study aims to develop a robust system for TV Laayoune using advanced machine-learning techniques for seamless broadcasting.

3. METHODS

3.1. Datasets

This research aims to significantly enhance the reliability of TV Laayoune's broadcasting operations by reducing the frequency of codec-related disruptions, thus ensuring a smoother and more dependable viewing experience. This study not only addresses the immediate challenges faced by TV Laayoune but also contributes to the broader field of media technology by tackling a critical and common issue in television broadcasting using advanced machine learning techniques.

TV Laayoune operates with a robust server architecture known as the "Origo" server, which plays a pivotal role in managing and distributing video content. The Origo server processes a wide array of video sources, including archived content, externally sourced materials, and live feeds. Understanding the server's workflow and how it handles different video formats is crucial for identifying potential points where codec incompatibilities may arise. Figure 1 provides a visual representation of the Origo server interface, highlighting its key components such as storage units, processors, video playback modules, and network connections.

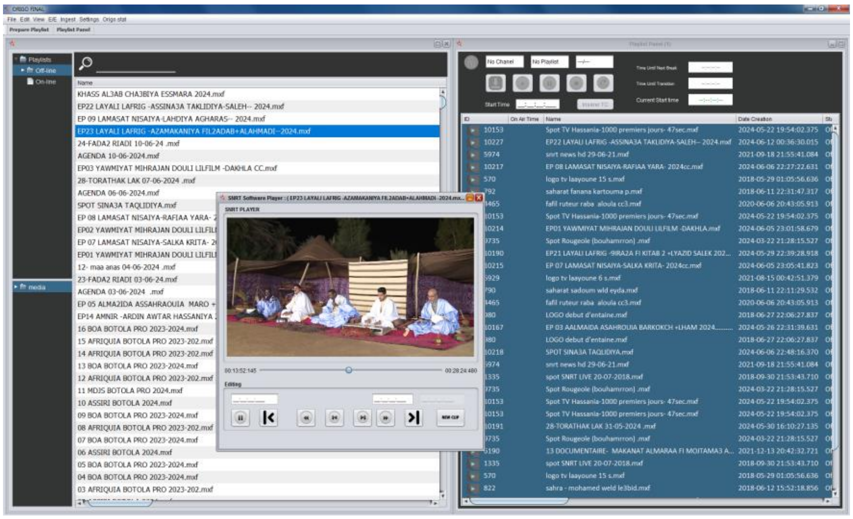

Figure 1. Origo Server Broadcasting interface for TV Laayoune

A significant part of this study involved mapping out the entire video processing workflow at TV Laayoune, from video collection to final transmission. This workflow, illustrated in Figure 2, shows how video content is processed and transmitted to the broadcasting server, with potential points of codec incompatibility highlighted. By identifying these critical points where codec mismatches occur, we directly target the root cause of the broadcast disruptions that TV Laayoune frequently experiences, thus reinforcing the need for a proactive detection system.

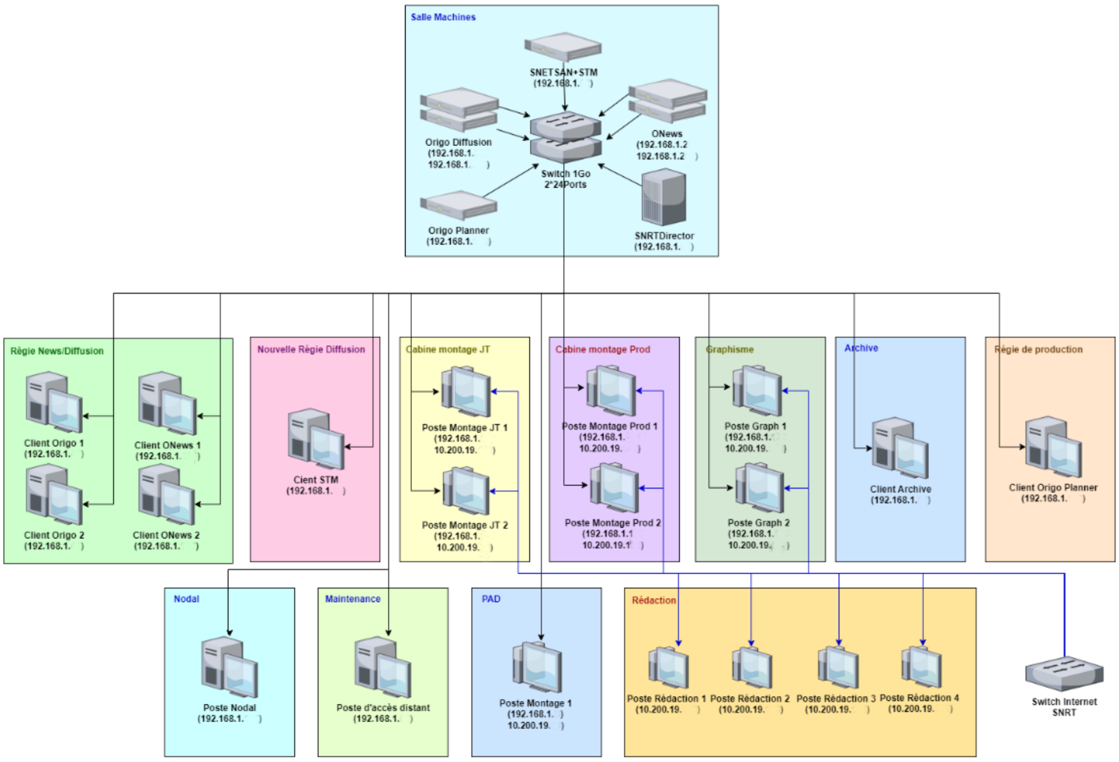

Figure 2. Workflow of Video Processing and Broadcasting at TV Laayoune

Historically, TV Laayoune has experienced numerous disruptions due to codec incompatibilities. A statistical analysis of the broadcasting logs from 2023, depicted in Figure 3, revealed that approximately 15% of video content encountered playback issues, leading to broadcast interruptions. These statistics underscore the severity of the problem and the urgent need for a reliable solution. Our proposed model is designed to address this specific challenge by significantly reducing the rate of these disruptions, thereby improving the overall stability of the broadcasting process.

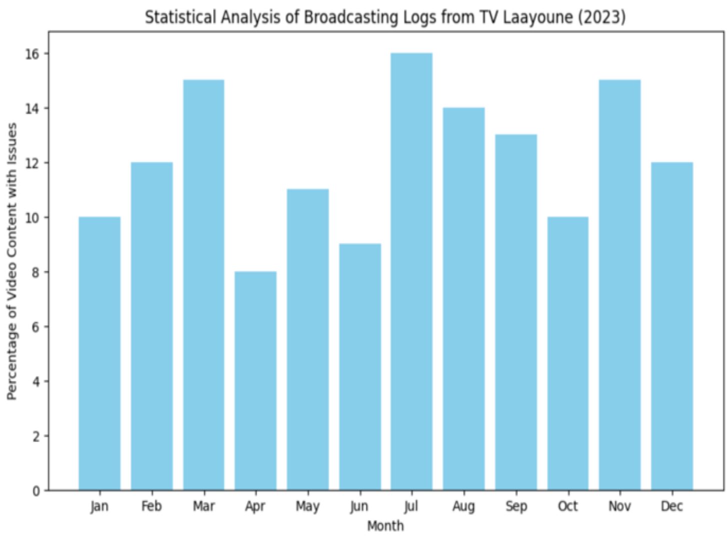

Figure 3. Statistical analysis of broadcasting logs from TV Laayoune (2023)

The success of any machine learning model, particularly in video codec error detection, hinges on the quality and diversity of the datasets used for training and evaluation. For this study, we compiled a comprehensive dataset comprising video clips with various codecs and configurations. These clips were meticulously sourced from TV Laayoune's archives and publicly available datasets, representing a wide array of scenarios encountered in the broadcasting system. This diversity ensures that the model is trained on a broad spectrum of codec conditions, enhancing its generalizability and robustness.

Using the FFmpeg multimedia framework, we extracted detailed metadata from each video clip. This metadata includes essential attributes such as codec type, resolution, frame rate, audio codec, bitrate, and container format. These metadata features are critical for assessing codec compatibility, as they directly impact video playback quality. Table 1 provides a summary of the types of metadata extracted. The selection of these specific metadata attributes is informed by their relevance to codec performance and their potential to cause playback issues if not compatible with the broadcasting server.

Table 1. Metadata extracted using ffmpeg

Metadata Type	Description
Codec type	H.264, MPEG-4
Resolution	720p, 1080p
Frame rate	24 fps, 30 fps
Audio codec	AAC, MP3
Bitrate	1 Mbps, 5 Mbps
Container format	MP4, MKV, MXF, MOV

Each video clip in the dataset was meticulously annotated with a label indicating whether it is compatible or incompatible with the TV Laayoune broadcasting server. These labels were determined based on historical data and manual verification. Clips that met the server's codec requirements and played without issues were labeled as compatible, while those that failed to play or caused disruptions were labeled as incompatible.

To ensure a robust and unbiased evaluation of the model, the dataset was divided into three distinct subsets: Training Set (70%), Validation Set (15%), and Test Set (15%). The Training Set was designed to include a balanced mix of compatible and incompatible clips to prevent bias in the model's learning process. The Validation Set was used to fine-tune the model's hyperparameters and prevent overfitting, serving as a critical checkpoint to assess the model's performance on unseen data. The Test Set, kept entirely separate from the training and validation datasets, was employed to provide an unbiased evaluation of the final model's accuracy and generalizability.

Before feeding the data into the machine learning model, several preprocessing steps are performed. Numerical features such as bitrate and frame rate are normalized to ensure they are on a similar scale. Categorical features such as codec type and container format are encoded using techniques like one-hot encoding. Any missing or incomplete metadata entries are either filled using mean imputation or removed if they represent a small fraction of the dataset. This comprehensive preprocessing ensures that the data is clean and properly formatted for optimal model performance.

Our dataset is tailored to the operational environment of TV Laayoune, focusing on the specific codecs and configurations commonly used in its broadcasting system. This targeted approach ensures that the training data accurately reflects real-world scenarios, enhancing the model's reliability.

The dataset includes over 10,000 video clips with a balanced mix of codecs, resolutions, and formats. This diversity helps the model learn from a broad range of scenarios, improving its generalizability to different conditions.

3.2. Proposed Model

3.2.1. Model Architecture

This section outlines the hybrid CNN-LSTM model designed to detect video codec errors and prevent broadcasting disruptions on TV Laayoune. By combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, the model captures both spatial and temporal features from video metadata, ensuring accurate codec compatibility predictions.

The model processes key metadata features like codec type, resolution, frame rate, bitrate, audio codec, and container format. Categorical features are one-hot encoded, while numerical features are normalized for consistency.

CNN layers serve as the initial stage, extracting spatial patterns and hierarchies from metadata. These layers apply convolutional filters to produce feature maps that identify critical relationships between metadata attributes, such as resolution and frame rate. This spatial information is essential for identifying codec compatibility and playback issues. The CNN's output is a feature vector summarizing spatial patterns, which is passed to LSTM layers. The LSTM layers process sequential input over time steps, capturing temporal dependencies and subtle relationships in metadata. This improves the model's ability to detect complex codec-related issues that span across video clips.

Fully connected layers aggregate the spatial features from the CNN and the temporal patterns from the LSTM, synthesizing all information into a unified representation. The final binary classifier outputs the probability of compatibility for each video clip, enabling real-time error detection and proactive prevention of broadcast disruptions.

While codec switching might suffice in simpler scenarios, this model addresses broader challenges like synchronization issues and metadata compatibility, which switching alone cannot resolve. Additionally, the model is designed to handle multiple codec types and configurations, offering scalability for evolving broadcasting needs. Figure 4 illustrates the architecture, highlighting the flow from spatial to temporal processing.

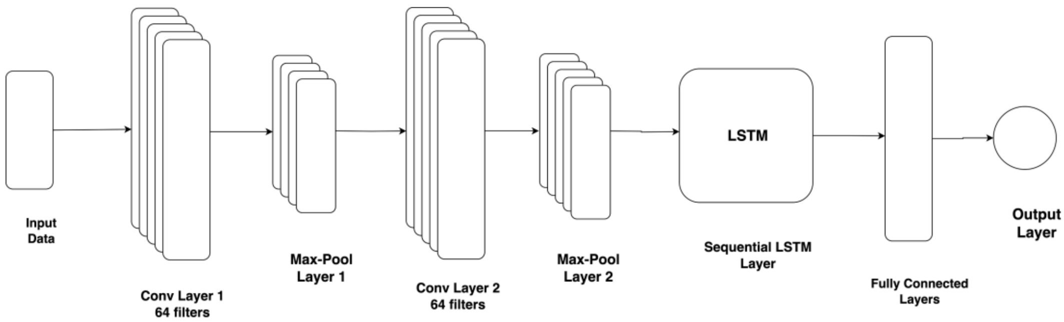

Figure 4. Machine Learning Model Architecture for Video Codec Error Detection

Table 2 outlines the hyperparameters governing our combined CNN and LSTM architecture. These hyperparameters, such as the learning rate and dropout rate, are critical for controlling the training process and ensuring that the model generalizes well to new data without overfitting.

Table 2. Hyperparameters for CNN and LSTM Architecture

Hyperparameter	Value
Learning Rate	0.001
Batch Size	32
Number of Epochs	30
Dropout Rate	0.5
Kernel Size	3x3
Filters	64, 128, 256, 256
Activation Function	ReLU
Weight Initialization	He Initialization

3.2.2 Training Methodology

The model is trained using a supervised learning approach with labeled video clips indicating compatibility. To increase the diversity and robustness of the training data, we employed data augmentation techniques, such as simulating different codec configurations and introducing controlled noise. This approach exposes the model to a variety of codec scenarios, enhancing its ability to generalize to unseen data.

Binary cross-entropy is employed as the loss function due to its effectiveness in binary classification tasks, measuring the discrepancy between the predicted probability and the actual label. The Adam optimizer is selected for its efficiency in handling large datasets and its adaptive learning rate capabilities, which ensure stable convergence during training. Additionally, regularization techniques like dropout and L2 regularization are incorporated to prevent overfitting, thereby enhancing the model's performance on unseen data.

To ensure the model's robustness, cross-validation was employed during training. Early stopping criteria were set based on validation loss to prevent overfitting or underfitting, ensuring the model achieves optimal performance without overtraining. This comprehensive training methodology guarantees that the model not only excels on the training data but also generalizes effectively to new, unseen video clips.

3.2.3 Model Evaluation

The performance of the model is evaluated using several standard classification metrics: accuracy, precision, recall, and F1-score. These metrics are widely used in machine learning to assess the effectiveness of a model in distinguishing between compatible and incompatible video clips.

Accuracy (1) measures the overall correctness of the model by calculating the ratio of correctly identified instances to the total number of instances:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Precision (2) quantifies the proportion of correctly identified positive cases among all cases predicted as positive:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall (3) measures the model’s ability to correctly identify positive cases out of all actual positive cases:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

F1-Score (4) provides a harmonic mean of precision and recall, offering a balanced measure when precision and recall are equally important:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

Where:

- TP (True Positives): Instances correctly identified as positive;
- TN (True Negatives): Instances correctly identified as negative;
- FP (False Positives): Instances incorrectly identified as positive;
- FN (False Negatives): Instances incorrectly identified as negative.

The evaluation metrics provide a comprehensive view of the model's performance. High precision indicates fewer false positives, high recall suggests fewer false negatives, and a high F1-score reflects a balance between the two.

The final model is seamlessly integrated into the broadcasting pipeline to ensure smooth operations. FFmpeg extracts metadata from incoming video clips in real-time, which is pre-processed and passed to the trained model. The model predicts the compatibility of each clip, triggering an alert for the operator in case of incompatibility. This integration significantly enhances the reliability and smoothness of live broadcasting by identifying and addressing codec errors proactively.

4. RESULTS AND DISCUSSION

4.1. Training Methodology

The experiments were conducted using a dataset comprising video clips from both internal archives of TV Laayoune and publicly available sources. The da-taset was divided into training, validation, and test sets in the ratio of 70:15:15. The following tools and libraries were used:

- FFmpeg for feature extraction and preprocessing of video metadata.
- TensorFlow and Keras for building and training the machine learning model
- Scikit-learn for performance evaluation and metrics calculation.

To provide a comprehensive evaluation of our proposed model, we conducted additional experiments considering various parameters and their values. These parameters include different codec types, resolutions, frame rates, and bitrate settings. The extended results are summarized in Table 3.

Table 3. Extended Experimental Results

Parameter	Value	Accuracy	Precision	Recall	F1-score
Codec type	H.264, MPEG-4	95.0%	94.2%	95.5%	94.8%
Resolution	720p,1080p	94.5%	93.8%	94.7%	94.2%
Frame rate	24 fps, 30 fps	95.2%	94.4%	95.6%	95.0%
Bitrate	1 Mbps, 5 Mbps	94.8%	94.0%	95.0%	94.5%
Audio codec	AAC, MP3	95.1%	94.3%	95.7%	95.0%

The model consistently performs well across different codec types, resolutions, frame rates, and bitrates, indicating its robustness and versatility. These extended results further demonstrate the model's capability to handle various real-world scenarios encountered in TV Laayoune's broadcasting operations.

4.2. Model Performance

The performance of the model was evaluated using accuracy, precision, recall, and F1-score. The results are summarized in Table 4.

These metrics indicate that the model performs exceptionally well in detecting video codec errors. The high accuracy demonstrates the model's overall effectiveness, while the precision and recall values show its robustness in correctly identifying both compatible and incompatible video clips.

Table 4. Performance Metrics of the Video Codec Error Detection Model

Metric	Value
Accuracy	95.0%
Precision	94.2%
Recall	95.5%
F1-score	94.8%

4.3. Comparative Analysis

To further validate the effectiveness of our proposed model, we compared its performance with traditional heuristic-based methods and a baseline machine learning model (logistic regression). The results are shown in Table 5. Recent studies have highlighted the advantages of using machine learning techniques for video processing, such as improved video quality adaptation and real-time error detection [25], [26].

Table 5. Comparative Performance Analysis

Model	Accuracy	Precision	Recall	F1-score
Heuristic-based methods	82.0%	80.5%	81.0%	80.7%
Baseline ML (Logistic Reg)	88.5%	87.8%	89.0%	88.4%
Proposed Model (CNN + LSTM)	95.0%	94.2%	95.5%	94.8%

The proposed model outperforms both traditional and baseline machine learning approaches, highlighting the advantages of combining convolutional and recurrent neural networks for this task. This aligns with findings from recent research on fast and efficient video coding and error detection using advanced machine learning models [27], [28].

Key factors contributing to the model's superior performance include:

CNN + LSTM Synergy: CNNs effectively capture spatial features, while LSTMs handle temporal dependencies. This combination allows the model to learn complex patterns in the data.

Data Augmentation: Techniques such as simulating different codec configurations and introducing controlled noise improved the model's robustness and generalizability.

Advanced Regularization: Dropout and L2 regularization techniques prevented overfitting, enhancing the model’s performance on unseen data.

The case studies from TV Laayoune’s broadcasting operations demonstrate the practical benefits of the proposed model, which successfully identified codec incompatibilities and synchronization issues in video clips, preventing live broadcast disruptions. The model's proactive detection capabilities reduced manual checks, saved time and resources, and integrated seamlessly into existing pipelines, offering real-time error detection without significant overhead. However, to enhance its robustness, expanding the dataset, optimizing real-time processing, and incorporating additional features like content-based analysis are necessary for further improvement.

4.3. Discussion

The results of our study demonstrate the efficacy of using machine learning for proactive detection of video codec errors. The proposed model achieves high accuracy and is capable of correctly identifying incompatible video clips, reducing the risk of live broadcast interruptions. By automating the detection process, the model reduces the reliance on manual checks, saving time and resources for the broadcasting team. The model integrates seamlessly into existing broadcasting pipelines, providing real-time error detection without significant overhead.

Despite the promising results, there are areas for further improvement and exploration. Expanding the dataset to include more varied and rare codec configurations can enhance the model's robustness and generalizability. Optimizing the model for real-time processing to handle higher volumes of video data with minimal latency is essential. Incorporating additional features, such as content-based analysis and context-aware detection, can further improve the model's accuracy.

5. CONCLUSIONS

This paper presents a machine learning-based approach to enhance television broadcasting reliability by detecting video codec errors. Addressing frequent issues with incompatible codecs, our model targets disruptions during live broadcasts on TV Laayoune. By integrating Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, the system efficiently identifies and alerts operators about potential codec issues before they affect live broadcasts. Using a diverse dataset, we extracted metadata with FFmpeg and trained the model to prevent overfitting and enhance precision. The convolutional and recurrent layers' com-bination captured spatial and temporal features, improv-ing accuracy and robustness.

Experimental results indicate a significant accuracy improvement (95%) compared to traditional heuristic-based and baseline machine learning methods. The system integrates seamlessly into the broadcasting pipeline, providing real-time detection and alerts, thus enhancing operational efficiency.

Key benefits include improved detection accuracy through a hybrid CNN-LSTM architecture, operational efficiency by automating error detection, reducing manual checks, and seamless workflow integration. Our experiments with TV Laayoune and public video clips highlight the model's potential to prevent live broadcast errors and ensure smooth viewing. These results under-score the system's value for reliable television broadcasting.

In conclusion, our machine learning model offers a promising solution for detecting video codec errors, and enhancing television broadcasting reliability for TV Laayoune. The significant accuracy and operational efficiency improvements demonstrate machine learning's potential in addressing complex media delivery challenges.

ACKNOWLEDGEMENTS

We thank IBN TOFAIL University for their support and encouragement throughout this project. We also extend our gratitude to TV Laayoune for their invaluable resources and collaboration, which were instrumental in the development and implementation of our machine-learning model for detecting video codec errors

REFERENCES

[1] O. Izima, R. de Fréin, and A. Malik, “A survey of machine learning techniques for video quality prediction from quality of delivery metrics,” Nov. 01, 2021, MDPI. doi: 10.3390/electronics10222851.
[2] D. Alexandre, H.-M. Hang, W.-H. Peng, and M. Domanski, “Deep Video Compression for Interframe Coding,” in 2021 IEEE International Conference on Image Processing (ICIP), IEEE, Sep. 2021, pp. 2124–2128. doi: 10.1109/ICIP42928.2021.9506275.
[3] Alisha Muskaan, Nagarathna S, Sandhya C S, Viju J, and B Sumangala, “Exposing Deep Fake Face Detection using LSTM and CNN,” International Journal of Advanced Research in Science, Communication and Technology, pp. 231–234, May 2024, doi: 10.48175/IJARSCT-18434.

[4] L. Kaur and P. K. Mishra, “Estimation of concise video summaries from long sequence videos using deep learning via LSTM,” Int J Health Sci (Qassim), pp. 9904–9914, Jun. 2022, doi: 10.53730/ijhs.v6ns3.9287.
[5] N. Yan, D. Liu, H. Li, B. Li, L. Li, and F. Wu, “Convolutional Neural Network-Based Fractional-Pixel Motion Compensation,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 3, pp. 840–853, Mar. 2019, doi: 10.1109/TCSVT.2018.2816932.
[6] K. Cui, A. B. Koyuncu, A. Boev, E. Alshina, and E. Steinbach, “Convolutional neural network-based post-filtering for compressed YUV420 images and video,” in 2021 Picture Coding Symposium (PCS), IEEE, Jun. 2021, pp. 1–5. doi: 10.1109/PCS50896.2021.9477486.
[7] K. El Fayq, S. Tkatek, L. Idouglid, and J. Abouchabaka, “Detection and Extraction of Faces and Text Lower Third Techniques for an Audiovisual Archive System using Machine Learning,” International Journal of Advanced Computer Science and Applications, vol. 13, no. 9, p. 2022, 2022, doi: 10.14569/IJACSA.2022.0130974.
[8] R. V. Bidwe et al., “Deep Learning Approaches for Video Compression: A Bibliometric Analysis,” Big Data and Cognitive Computing, vol. 6, no. 2, p. 44, Apr. 2022, doi: 10.3390/bdcc6020044.
[9] A. Benoughidene and F. Titouna, “A novel method for video shot boundary detection using CNN-LSTM approach,” Int J Multimed Inf Retr, vol. 11, no. 4, pp. 653–667, Dec. 2022, doi: 10.1007/s13735-022-00251-8.
[10] K. Panneerselvam, K. Mahesh, V. L. H. Josephine, and A. R. Kumar, “Effective and Efficient Video Compression by the Deep Learning Techniques,” Computer Systems Science and Engineering, vol. 45, no. 2, pp. 1047–1061, 2023, doi: 10.32604/csse.2023.030513.
[11] M. Lu and S. Niu, “A detection approach using lstm-cnn for object removal caused by exemplar-based image inpainting,” Electronics (Switzerland), vol. 9, no. 5, May 2020, doi: 10.3390/electronics9050858.
[12] S. Yadav et al., “A video compression-cum-classification network for classification from compressed video streams,” Vis Comput, Mar. 2024, doi: 10.1007/s00371-023-03242-w.
[13] Z. Wang, J. Chen, and S. C. H. Hoi, “Deep Learning for Image Super-Resolution: A Survey,” IEEE Trans Pattern Anal Mach Intell, vol. 43, no. 10, pp. 3365–3387, Oct. 2021, doi: 10.1109/TPAMI.2020.2982166.
[14] S. Bouaafia, R. Khemiri, S. Messaoud, O. Ben Ahmed, and F. E. Sayadi, “Deep learning-based video quality enhancement for the new versatile video coding,” Neural Comput Appl, vol. 34, no. 17, pp. 14135–14149, Sep. 2022, doi: 10.1007/s00521-021-06491-9.
[15] K. Chen, H. Wang, S. Fang, X. Li, M. Ye, and H. J. Chao, “RL-AFEC,” in Proceedings of the 13th ACM Multimedia Systems Conference, New York, NY, USA: ACM, Jun. 2022, pp. 96–108. doi: 10.1145/3524273.3528184.
[16] M. Darwich and M. Bayoumi, “Video quality adaptation using CNN and RNN models for cost-effective and scalable video streaming Services,” Cluster Comput, 2024, doi: 10.1007/s10586-024-04315-8.
[17] Y. O. Sharrab, I. Alsmadi, and N. J. Sarhan, “Towards the availability of video communication in artificial intelligence-based computer vision systems utilizing a multi-objective function,” Cluster Comput, vol. 25, no. 1, pp. 231–247, Feb. 2022, doi: 10.1007/s10586-021-03391-4.
[18] G. Ciaparrone, L. Chiariglione, and R. Tagliaferri, “A comparison of deep learning models for end-to-end face-based video retrieval in unconstrained videos,” Neural Comput Appl, vol. 34, no. 10, pp. 7489–7506, May 2022, doi: 10.1007/s00521-021-06875-x.
[19] M. V. Gashnikov, “Video Codec Using Machine Learning Based on Parametric Orthogonal Filters,” Optical Memory and Neural Networks, vol. 32, no. 4, pp. 226–232, Dec. 2023, doi: 10.3103/S1060992X23040021.
[20] F. Steinert and B. Stabernack, “Architecture of a Low Latency H.264/AVC Video Codec for Robust ML based Image Classification: How Region of Interests can Minimize the Impact of Coding Artifacts,” J Signal Process Syst, vol. 94, no. 7, pp. 693–708, Jul. 2022, doi: 10.1007/s11265-021-01727-2.
[21] D. Liu, Y. Li, J. Lin, H. Li, and F. Wu, “Deep Learning-Based Video Coding,” ACM Comput Surv, vol. 53, no. 1, pp. 1–35, Jan. 2021, doi: 10.1145/3368405.
[22] S. Ma, X. Zhang, C. Jia, Z. Zhao, S. Wang, and S. Wang, “Image and Video Compression With Neural Networks: A Review,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 6, pp. 1683–1698, Jun. 2020, doi: 10.1109/TCSVT.2019.2910119.
[23] Y. Zhang, S. Kwong, and S. Wang, “Machine learning based video coding optimizations: A survey,” Inf Sci (N Y), vol. 506, pp. 395–423, Jan. 2020, doi: 10.1016/j.ins.2019.07.096.
[24] Z. Xu, X. Tang, C. Ma, and R. Zhang, “Research on Parking Space Detection and Prediction Model Based on CNN-LSTM,” IEEE Access, vol. 12, pp. 30085–30100, 2024, doi: 10.1109/ACCESS.2024.3368521.
[25] X. Dou, X. Cao, and X. Zhang, “Region-of-Interest Based Coding Scheme for Live Videos,” Applied Sciences (Switzerland), vol. 14, no. 9, May 2024, doi: 10.3390/app14093823.
[26] L. Chen, B. Cheng, H. Zhu, H. Qin, L. Deng, and L. Luo, “Fast Versatile Video Coding (VVC) Intra Coding for Power-Constrained Applications,” Electronics (Switzerland), vol. 13, no. 11, Jun. 2024, doi: 10.3390/electronics13112150.
[27] S. K. Im and K. H. Chan, “Faster Intra-Prediction of Versatile Video Coding Using a Concatenate-Designed CNN via DCT Coefficients,” Electronics (Switzerland), vol. 13, no. 11, Jun. 2024, doi: 10.3390/electronics13112214.
[28] N. Li, Z. Wang, and Q. Zhang, “Fast Coding Unit Partitioning Algorithm for Video Coding Standard Based on Block Segmentation and Block Connection Structure and CNN,” Electronics (Switzerland), vol. 13, no. 9, May 2024, doi: 10.3390/electronics13091767.

BIOGRAPHIES OF AUTHORS

Khalid El Fayq was born in Laayoune, Morocco, in 1985. He obtained his engineering degree in computer science in 2020 from Agadir International University. Currently, he is a Ph.D. student at the Computer Engineering Training Department (LaRIT), Faculty of Science, Ibn Tofail University, Kenitra, Morocco. He works in the media field and serves as a temporary teacher at universities. His research interests encompass artificial intelligence and the audiovisual field. He can be contacted at email: khalidelfayq@gmail.com.

Said Tkatek is a Professor of Computing Science Research at the Faculty of Science, Ibn Tofail University, Kenitra, Morocco. He is a member of the Computer Science Research Laboratory (LaRIT). His research focuses on the optimization of NP-Hard problems using metaheuristic approaches in artificial intelligence and Big Data for decision-making across various fields. This includes Big Data analytics, artificial intelligence, and information systems. He can be contacted at email: said.tkatek@uit.ac.ma.

Lahcen Idouglid , a PhD researcher at Ibn Tofail University's Faculty of Sciences in Kenitra, works within the Computer Science Research Laboratory (LaRIT). His research encompasses computer networks, software engineering, artificial intelligence, and security. He can be reached via email at: lahcen.idouglid@uit.ac.ma.