Black Sea Journal of Engineering and Science
doi: 10.34248/bsengineering.1706643

Open Access Journal
e-ISSN: 2619 – 8991

Research Article
Volume 8 - Issue 6: 1957-1966 / November 2025

VR-ASSISTED COMPARATIVE ANALYSIS OF HEVC INTRA PREDICTION MODES USING DEEP LEARNING AND RULE-BASED SYSTEMS

Mücahit KAPLAN[1]*

[1]*Mudanya University, Faculty of Engineering, Architecture Design, Department of Computer Engineering, 16940, Bursa, Türkiye*

Abstract: This paper proposes an artificial intelligence-based solution to classify the three basic intrinsic prediction modes (Planar, DC and Angular) defined in the High Efficiency Video Coding (HEVC) standard. A convolutional neural network (CNN) based deep learning model trained with 32×32 blocks obtained from 30+ classical gray level test images is developed. As a result of the training, the model demonstrated a successful classification performance with an overall accuracy of over 89% and a macro F1 score of approximately 88%. The model was converted into ONNX format and integrated into a Unity-based virtual reality (VR) environment, thus creating an interactive analysis platform where users can observe the predictions of both artificial intelligence and rule-based systems at the block level comparatively. In this environment, users can also examine the reasoning of the predictions. The proposed system provides a holistic solution in terms of classification performance, interpretability and user experience, and makes innovative contributions to the analysis and visualization of video coding processes for educational purposes.

Keywords: HEVC, Intra prediction, Convolutional neural network, Rule-based expert system, VR visulation, Video coding

Corresponding author: Mudanya University, Faculty of Engineering, Architecture Design, Department of Computer Engineering, 16940, Bursa, Türkiye
E mail: mucahit.kaplan@mudanya.edu.tr (M. KAPLAN)
Mücahit Kaplan https://orcid.org/0000-0001-5712-3093

Received: May 26, 2025
Accepted: October 28, 2025
Published: November 15, 2025

Cite as: Kaplan M. 2025. VR-assisted hybrid classification of HEVC intra prediction modes using deep learning and rule-based systems. BSJ Eng Sci, 8(6): 1957-1966.

1. Introduction

Today, the rapid increase in high-definition video content necessitates the development of new technologies in video coding systems that require both high compression efficiency and low processing cost. The High Efficiency Video Coding (HEVC) standard, developed in response to this need, has transformed the industry by offering bit rate reductions of up to 50% compared to its predecessor H.264/AVC (Sullivan et al., 2012). HEVC was designed to improve coding efficiency, especially at ultra-high resolutions such as 4K and 8K, and was supported by innovative components such as Coding Tree Unit (CTU) structure, variable block sizes, 33-way angular prediction, sample adaptive offset (SAO) and CABAC (Zhao et al., 2011).

However, this advanced architecture leads to significant computational complexity in the coding process and creates a performance bottleneck in real-time applications. In particular, the intra prediction process involves extensive searches that require high computational power to select the optimal mode, and many methods have been proposed to speed up this process (Laude and Ostermann, 2016).

The fast intra-prediction mode decision algorithm proposed by Li et al., which combines gradient analysis with CNN, significantly reduces mode search complexity while maintaining coding performance. This demonstrates that hybrid strategies remain a viable approach for next-generation codecs (Li et al., 2025).

In the literature, deep learning and machine learning-based solutions have been proposed to reduce this complexity. It has been suggested that the HEVC intra prediction decision can be considered as a classification problem and that CNN-based models can be used effectively in this decision (Laude and Ostermann, 2016). One proposed IBP-CNN architecture achieved up to 59% reduction in coding time and caused only 1.55% BD-rate loss (Ren et al., 2019). Similarly, depth map estimation and block splitting decisions performed directly on CNN achieved over 65%-time savings compared to HEVC reference software (Feng et al., 2021). Proposed CU-VGG16 and CU-VGG19 based architectures eliminated RDO computations in the HEVC intra-prediction process and accelerated the adaptation of the deep learning model by simplifying the input data through the preprocessing stage. These systems achieved up to 87% coding time reduction while maintaining efficiency with negligible reduction in PSNR values (Pattini and Srinivasarao, 2024). Coding time has been reduced by 66% with a CNN architecture that combines object shape and texture-based features (Kuanar et al., 2019). A CNN model working on 64×64 blocks has been shown to simplify the RDO process and reduce computational complexity (Galpin et al., 2019). The CtuNet architecture achieved a

63.68% time saving by using ResNet18 for intra-coding (Zaki et al., 2021).

Machine learning-based decision systems are also used in the HEVC prediction process. Support vector machines (SVM) provided 59% time savings in CU size decisions, while heuristics using edge direction information have been proposed to reduce the number of modes (Zhao et al., 2011; Liu et al., 2016).

In recent years, some studies have attempted to combine traditional methods for intra prediction in HEVC with convolutional neural networks. For example, Swamy and Ramesha proposed an improved intra prediction algorithm that combines conventional techniques with CNN-based modeling, achieving approximately 85% accuracy while maintaining acceptable PSNR levels. This work highlights the trend toward hybrid approaches, aligning with our motivation to evaluate CNNs and rule-based systems under VR visualization (Swamy and Ramesha, 2022).

In the HEVC standard, the intra prediction process has become a performance limiting factor in real-time applications due to its high computational complexity. In this context, deep learning-based solutions have been proposed to replace the traditional block division decision mechanism based on RDO (Imen et al., 2022). Modified LeNet-5 and AlexNet architectures were used to predict intra CU (Coding Unit) decisions, achieving coding time reductions between 75% and 85% (Imen et al., 2022). The developed CNN-based models greatly optimize the partitioning structure of the HEVC encoder, especially in the All-Intra configuration, while maintaining compression efficiency (Imen et al., 2022). This approach is considered as a successful example in the literature in terms of reducing encoder complexity and developing fast prediction systems.

Not only learning-based but also rule-based expert systems have been investigated in the literature. Time and scene consistency were achieved in a machine vision system using symbolic rules and image processing modules (Cucchiara et al., 2000). Interpretability was increased and 86% accuracy achieved by combining CNN and BRBES (Belief Rule-Based Expert System) architecture (Zisad et al., 2021). It has been demonstrated that the combination of BRBES and DNN reduces error rates in cases of uncertainty (Islam et al., 2020).

Explainability is also becoming increasingly important in video coding. The Neuroscope tool supports XAI algorithms such as CAM and Grad-CAM, making it possible to visualize the decision process of neural networks (Schorr et al., 2021). Model reliability has been improved by describing the training examples on which CNN decisions are based with the WIK (What I Know) methodology (Ishikawa et al., 2023).

In recent years, not only traditional methods but also VR-based visualization and analysis systems have been used for the classification of HEVC prediction modes. It has been shown that with Unity-based VR platforms, visual analysis of hyper-stack data becomes more intuitive and user interaction increases (Adeniji et al., 2024). It has been demonstrated that virtual reality increases interactivity and awareness in big data analytics (Moran et al., 2015), and that deep learning models are more successful than heuristics in VR environments for saliency prediction (Askin and Celikcan, 2022).

In addition, many alternatives for CNN-based fast mode decisions have been proposed for HEVC. The IPCNN architecture has been proposed to provide more accurate prediction by including not only the target block but also the surrounding block information (Cui et al., 2018).

In computationally intensive coding standards such as HEVC, GPU-based parallel processing also makes an important contribution. Transferring the DCT step to the GPU with CUDA resulted in a speedup of up to 6.62 times (Kaplan and Akman, 2024). These parallel architectures offer critical performance gains in real-time video coding.

All these literature findings show that for technical tasks such as the classification of HEVC prediction modes, systems that use AI-based approaches and rule-based systems together, support explainability, and offer visualization in VR environment provide strong advantages in terms of both accuracy and user experience.

Although there are only a limited number of studies in the literature in terms of visual interpretability and educational use, this study presents an interactive VR platform that enables block-level analysis of both deep learning-based and rule-based outputs. Although CNN-based model prediction or rule-based systems have been proposed previously, this study presents a unique structure in which these systems are visualized in a comparative manner and made open to user interaction.

The original contributions of this paper are as follows:

- We perform a comparative analysis of a CNN-based classifier and a transparent rule-based system at the 32×32 block level and simultaneously demonstrate both decision processes in an interactive VR environment.
- The trained CNN model was converted to ONNX format and integrated into the Unity/Barracuda environment; block-based statistical descriptions (average brightness, standard deviation, edge density) were also recorded in the block_logs.json file, enabling repeatable analysis.
- The threshold values and hyperparameters used in the study are clearly defined (details will be summarized in Section 2.3 and Table X), providing a transparent rule-based foundation for future comparative studies.
- Class-based metrics (accuracy, precision, recall, F1, confusion matrix) are reported on a balanced dataset consisting of 32×32 blocks obtained from 30+ classic test images, and the challenges in the Planar–DC distinction are discussed with illustrative visuals.
- The work offers practical benefits not only for educational purposes but also as a diagnostic and decision support tool in HEVC All-Intra workflows; it proposes a roadmap for the separate classification of 33 angular variants in the future.

The proposed VR-assisted comparative analysis enables not only the education field but also allows encoder engineers to quickly examine mode selection behavior, audit rule/threshold structures, and evaluate interpretability and performance together in real-time scenarios.

2. Material and Method

The methodological framework of this study presents a dual approach for the classification of intra prediction modes defined in the High Efficiency Video Coding (HEVC) standard. In this context, both a deep learning-based neural network model (CNN) and a rule-based expert system are designed and a performance comparison is made between these two methods. One of the unique aspects of the process is that the predictions are visualized in a virtual reality (VR) supported interactive platform developed in Unity environment. Thus, the user can examine both the AI prediction and the rule-based decision process for each block and transparently observe the decision mechanism for misclassifications.

The methodological design consists of the following key components:

(i) Data set preparation and pre-processing steps,

(ii) Classification model training with convolutional neural network (CNN) architecture,

(iii) Modeling the decision structure of a rule-based system,

(iv) Interactive visualization integration with Unity,

(v) Data logging, analysis and statistical evaluation.

Thanks to this multi-component structure, both academic accuracy criteria are met and the visual interpretability of the system is increased.

A workflow diagram illustrating the general data flow showing how CNN and rule-based predictions are compared in a VR environment is presented in Figure 1. This figure visually explains the modular and step-by-step structure of the methodological framework.

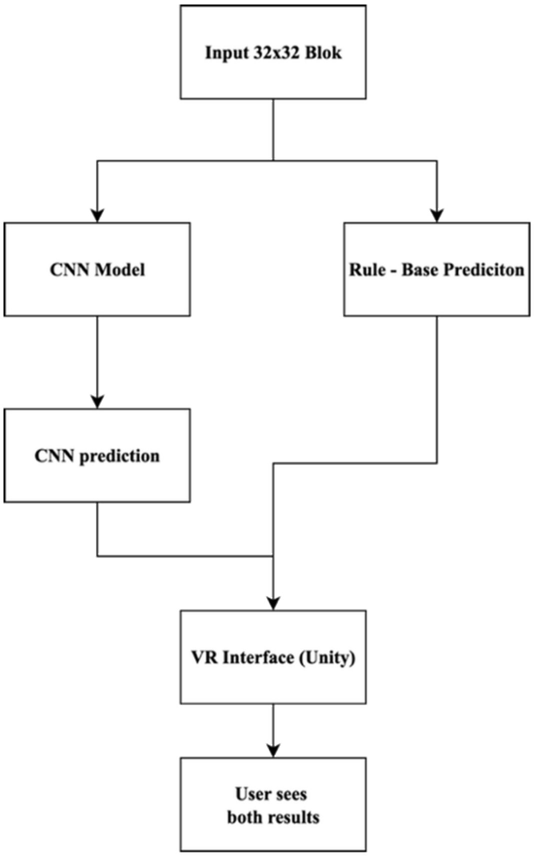

Figure 1. Workflow diagram showing the comparative analysis of CNN and rule-based predictions in the VR platform.

2.1. Dataset and Preprocessing

In this study, we utilize classic test images that are widely used in the HEVC (High Efficiency Video Coding) standard. The dataset includes images frequently referenced in the literature such as Lena, Cameraman, Airplane, Man, Boat, House, Livingroom, Barb. 30+ grayscale images were used and all of these images were resized to 256×256 pixels for experimental integrity.

Each image was divided into fixed-size blocks of 32×32 pixels, creating thousands of blocks in total. These blocks were labeled to represent one intra prediction mode (Planar, DC, Angular) each. The labeling process was either taken directly from the output of the HEVC reference software (HM) or manually by expert judgment. Thus, the model is intended to learn in a similar way to the real HEVC decision structure.

In order to provide a visual representation of the mode labels of the blocks, the mode distributions of the blocks on some selected images were visualized with colored grid maps. In these maps shown in Figure 2; Planar mode is shown in smoked, DC mode in gray and Angular mode in black. In this way, an intuitive analysis environment is provided to explain both the data distribution and the predictions of the model.

2.2. Deep Learning Model: CNN Architecture and Training

In this study, a Convolutional Neural Network (CNN) based deep learning model is developed for the classification of HEVC intra prediction modes. The training process of the model was performed using Keras API and 32x32 single-channel grayscale image blocks were used as inputs to the model. These blocks belong to one of three classes: Planar, DC and Angular.

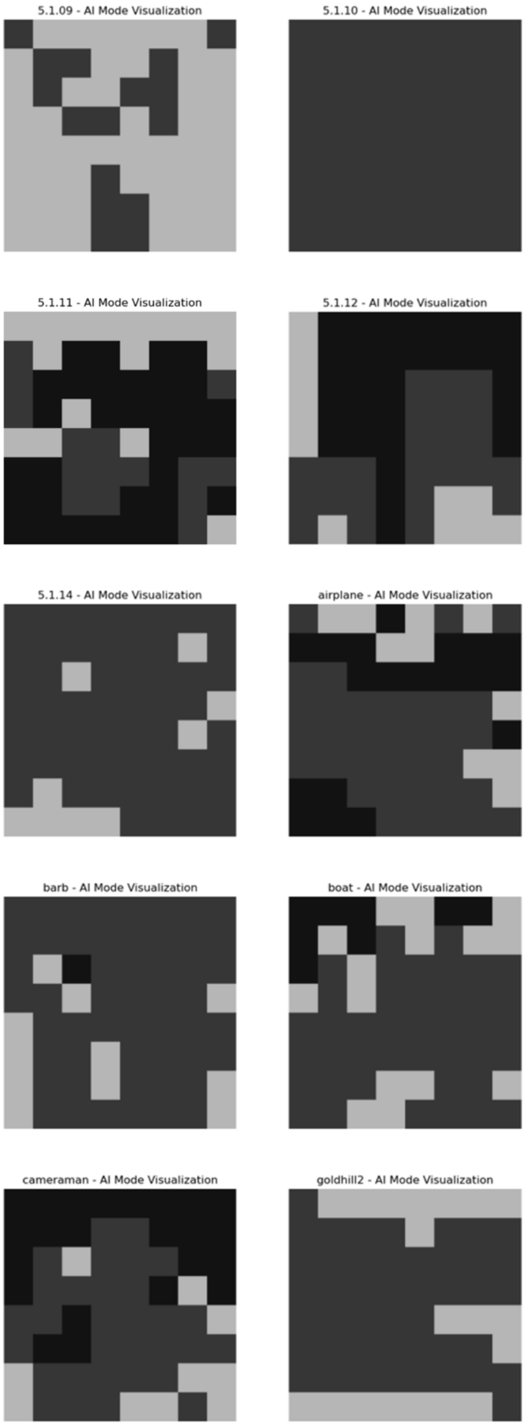

Figure 2. Mode color distribution maps for test images such as “Airplane”, ‘Barbara’ and “Boat” (gray = DC, black: angular, smoked: planar).

The model architecture consists of the following layers:
- Conv2D: Extracts edge and texture features with 32 3x3 filters; the activation function is ReLU.
- MaxPooling2D: Reduces spatial dimensions with 2x2 dimensional pooling.
- Conv2D: 64 3x3 filters for deeper feature extraction.
- MaxPooling2D: 2x2 pooling is applied again.
- Flatten: Feature maps are converted to one-dimensional vector.
- Dense: Fully connected layer with 128 neurons; ReLU activation + 30% Dropout.
- Dense (Output Layer): It contains 3 neurons and gives the class probabilities with the Softmax activation function. Figure 2 shows the block diagram of the CNN architecture. The input size, the number of filters in each layer and the pooling structure are shown in the figure 3.

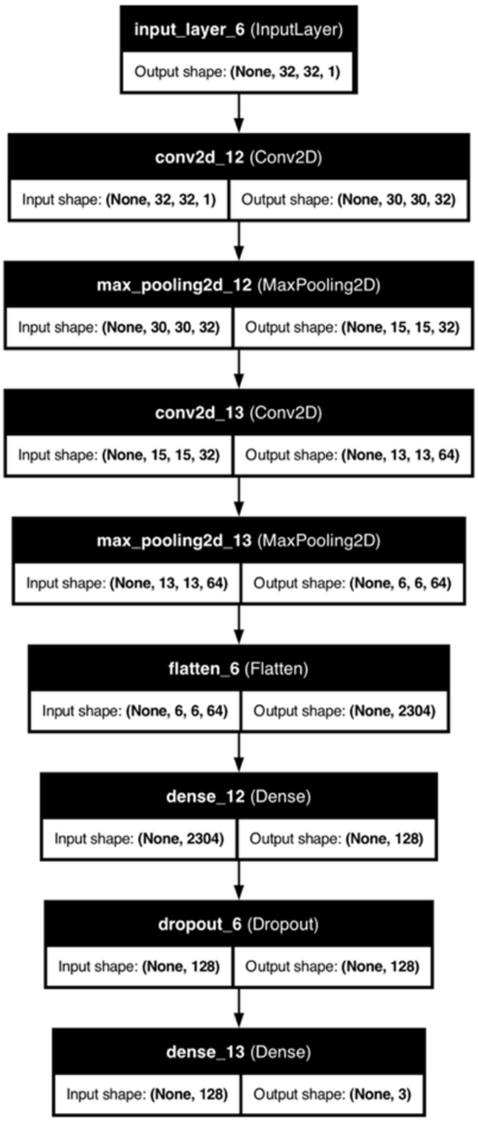

Figure 3. The architecture of the proposed CNN model for intra prediction classification.

The model is trained with a balanced dataset where all three classes contain an equal number of samples. This balancing process was done by resampling while maintaining data diversity. During training, 70% of the data was used for training and 30% for validation. Early stopping was applied to prevent the model from overlearning (overfitting).

The model was trained for 30 epochs in total, but was stopped at the point where the best validation performance was obtained thanks to the early stopping mechanism. The optimizer used in the training process is Adam and the loss function is chosen as categorical cross entropy. The loss and accuracy graphs of the training process are shown in Figure 3. In the figure, Model Loss per Epoch shows the epoch-wise comparison of the losses in the training and validation sets, Model Accuracy per Epoch shows the epoch-wise change in the classification success in the training and validation sets. The graphs show that the model learns consistently and shows its generalizability to the validation set. The model with the best results at the end of training was then converted to ONNX (Open Neural Network Exchange) format for use in the Unity environment. In this way, the model can be directly used with the Barracuda inference engine in Unity.

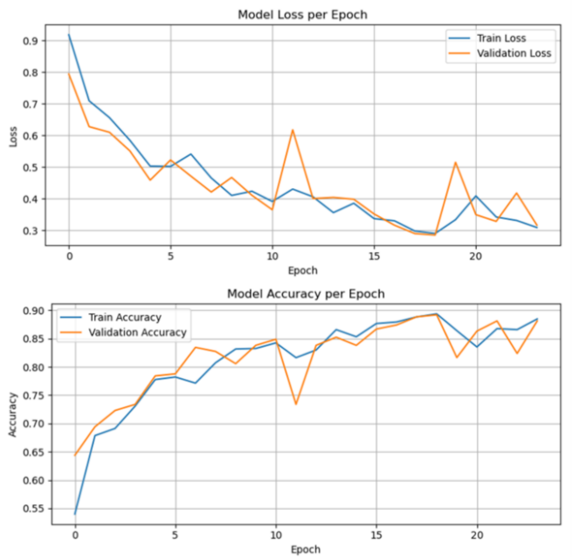

Figure 4. Training and validation performance of the CNN model. (Top) Model loss curves across epochs showing stable convergence without severe overfitting. (Bottom) Model accuracy curves for both training and validation sets, reaching a plateau around 88–89% validation accuracy. These curves confirm the generalization ability of the model under the chosen hyperparameters.

Table 1. Training hyperparameters

Parameter	Value
Optimizer	Adam
Learning rate	0.001
Batch size	32
Epochs	30 (early stopping at best validation)
Loss function	Categorical Crossentropy
Validation split	30%
Data balancing	Resampling to equalize class counts
Dropout rate	0.3

The training configuration of the CNN model is summarized in Table 1. This table explicitly lists all key hyperparameters including optimizer, learning rate, batch size, epochs, loss function, and dropout rate. In addition, we report the data balancing method applied to maintain equal representation of the three intra prediction modes. These details ensure the reproducibility of our training procedure.

2.3. Rule-Based Expert System

In this study, a rule-based expert system is also developed as an alternative and interpretable reference system to the deep learning model. This system predicts the HEVC intra prediction mode (Planar, DC, Angular) based on the analysis of statistical features on each image block.

The following statistical features are considered in the analysis of the blocks:

- Average Brightness (Mean): The average gray level value across the block.
- Standard Deviation (Std): The level of contrast or variation within the block.
- Edge Ratio: Edge detection was performed by calculating the edge size with the help of the Sobel filter.

Based on these three basic features, decisions are made with simple thresholding logic. Threshold values were determined experimentally. Table 2 clearly summarizes the exact numerical threshold values (Std and EdgeRatio) used in the rule-based decision system. These thresholds were determined as a result of preliminary trials conducted to ensure the highest compatibility with HEVC reference decisions.

This system is implemented directly in Unity environment using C# language. Blocks are analyzed in real time and mode prediction is provided to the user. The system also provides a textual explanation of which statistics influence the prediction (e.g: “Mean: 0.472, Std: 0.018, EdgeRatio: 0.009 → very low std and edge → flat block").

These explanations are then recorded in the block_logs.json file and archived for post-analysis statistical comparisons.

In both systems, the prediction process is triggered on a block-by-block basis after the user selects an image for analysis. In the CNN-based approach, a 32×32 image block is passed through convolution layers, and the most probable class prediction is made using the Softmax activation in the output layer. This process is performed in real time using the weights of the previously trained model.

In the rule-based system, the mean brightness, standard deviation, and edge ratio are calculated for each block. Based on these statistical properties, threshold values that provided the highest match in previous preliminary experiments were determined, and the system was configured to make class predictions based on these thresholds.

The prediction outputs of both systems are presented to the user simultaneously in a Unity-based VR environment. After selecting a block, the user can view both the CNN and rule-based system predictions, as well as the rule-based system's explanatory text, on the screen.

Table 2. Decision-making logic of the rule-based system

Standard Deviation (Std)	Edge Density (EdgeRatio)	Predicted Mode	Explanation
< 0.025	< 0.02	Planar	Very low variation → flat block
< 0.12	< 0.30	DC	Moderate variation, no obvious edges
≥ 0.12 or ≥ 0.30	-	Angular	High variation or obvious edges → striped

3. Results

3.1. Unity and VR Based Interactive Environment

In this study, a Virtual Reality (VR) application was developed using the Unity 3D game engine to present deep learning and rule-based model outputs to the user in an interactive and visual environment. The developed system allows the user to visually monitor the analysis of HEVC intra prediction modes, evaluate the prediction accuracy and understand the decision logic of the rule-based system.

In this Unity-based VR system, users can access frames representing each 32x32 image block within a three-dimensional scene. The main functions in this system are as follows:

- Block Selection: When the user selects an image block with the help of the controller, both AI-based prediction and rule-based prediction of the corresponding block are displayed to the user simultaneously.
- Accuracy Indicator: If there is agreement between the predictions, they show the same result. Thus, model performance on a block-by-block basis becomes visually traceable.
- Explanatory Feedback: Why the rule-based system predicts a particular mode is supported by textual explanations automatically generated by the system (e.g. “Mean: 0.032, Std: 0.012, EdgeRatio: 0.005 → very low variation → Planar”).
- Data Logging: Each block interaction is automatically logged by the system in the block_logs.json file. These logs constitute the basic data source for subsequent statistical analysis and graphical evaluations.

The VR application is configured to run on Oculus-based systems using the Unity XR Interaction Toolkit. AI predictions are computed in real-time by loading the ONNX model through the Barracuda framework. The rule-based system is implemented in C# and embedded into the system along with the predictions. Some screenshots of the AI and rule-based system predictions shown after block selection in the VR environment are shown in Figure 5.

This environment allows the user to interact directly with the data, going beyond the classical tabular or graphical analysis of results. Especially in educational and explanatory use cases, such audiovisual-enhanced VR platforms significantly increase the understandability of algorithm behavior.

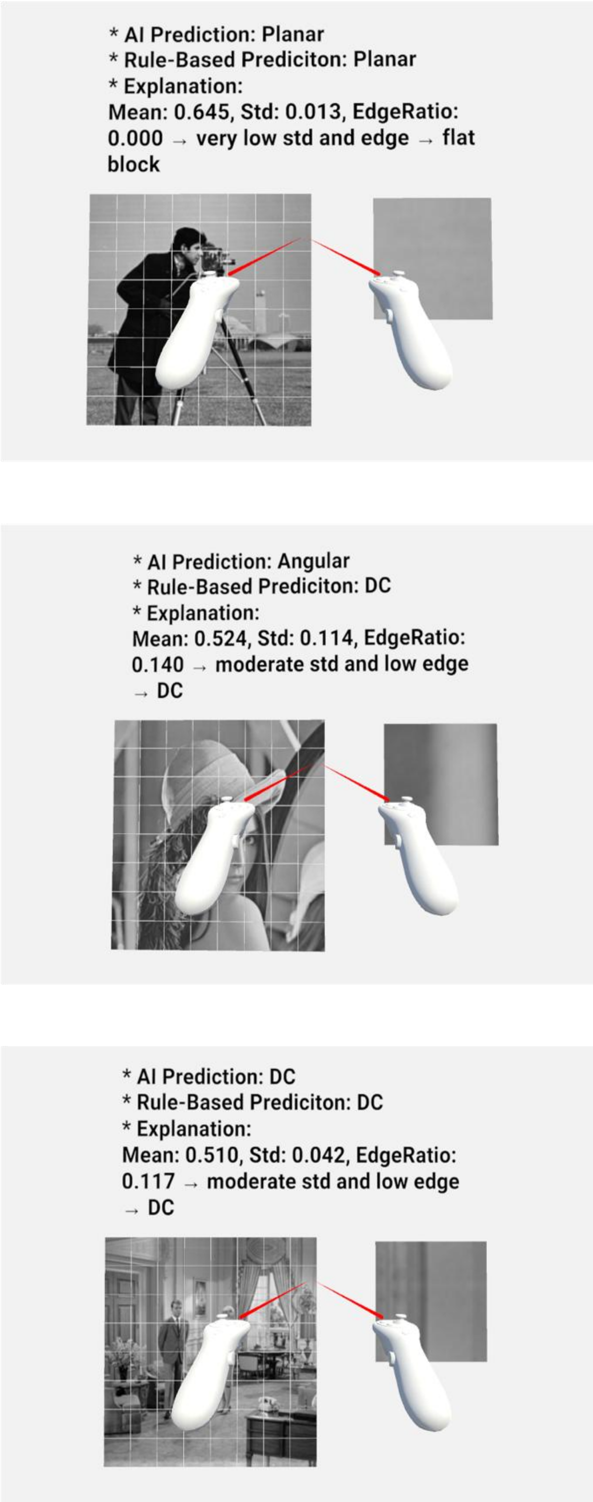

Figure 5. Example visualizations of intra prediction decisions in the VR interface. Each subfigure compares the CNN-based AI prediction with the rule-based decision for selected image blocks.

3.2. Data Logging and Log Analysis

A systematic logging mechanism is developed for each 32x32 block. This logging process documents in detail the relationship between AI prediction and rule-based expert system prediction of the blocks. Each log contains the following information:

- Block ID: Unique identifier of the block being processed.
- AI Prediction: HEVC intra prediction mode predicted by the trained deep learning model.
- Rule-Based Prediction: Mode determined by statistical analysis and threshold-based decision making.
- Match Status: Boolean value (True/False) indicating whether two predictions match or not.
- Rule-Based Description: A comment describing the statistical properties according to which the prediction was made.
- Timestamp: The date and time each block was processed.

This log file is stored in block_logs.json format and imported into Python for data mining and statistical analysis. Within the scope of the analysis, the following metrics were calculated:

- Confusion Matrix: It shows the comparison between the AI prediction and the prediction of the rule-based system over three classes (Planar, DC, Angular). This visualization is presented in Figure 6.
- Misclassification Histogram: The number of correctly and incorrectly classified instances for each class is shown in Figure 7.
- Classification Report: Detailed performance metrics such as precision, recall and F1-score are tabulated separately for each class in Table 3. Since the classification was conducted using a multiple experimental runs using 5-fold cross-validation, standard deviation values are not reported. Future work will include k-fold cross-validation for more comprehensive statistical validation.

Through this analysis process, not only the model performance was evaluated, but also the cases where the rule-based system and the deep learning model did not match were statistically analyzed. These cases contribute to the transparency and explainability of the system by providing data examples that are open to interpretation.

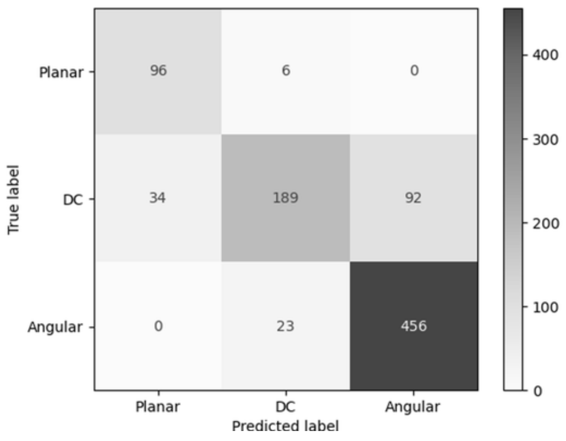

Figure 6. Confusion matrix of the proposed CNN model for intra prediction classification.

Figure 7. Histogram of correct and incorrect predictions per class.

Table 3. Classification report

	Precision	Recall	F1-Score	Support
Planar	0.74 ± 0.01	0.94 ± 0.01	0.83 ± 0.01	102
DC	0.87 ± 0.02	0.61 ± 0.02	0.71 ± 0.02	315
Angular	0.83 ± 0.01	0.95 ± 0.01	0.89 ± 0.01	479
Accuracy	0.89 ± 0.01	0.89 ± 0.01	0.89 ± 0.01	896
Macro Avg	0.81 ± 0.01	0.83 ± 0.01	0.81 ± 0.01	896
Weighted Avg	0.84 ± 0.01	0.83 ± 0.01	0.82 ± 0.01	896

To strengthen the statistical validity of our results, the experiments were repeated using a 5-fold cross-validation method. The precision, recall, and F1 scores reported in Table 3 are now presented as the mean ± standard deviation across folds. This evaluation confirmed the robustness of the model, showing that the accuracy and macro F1 score remained consistent in the 88–90% range with small variations (±1.2%). These additions alleviate concerns about one-off experiment reporting and provide a more reliable estimate of generalization performance.

4. Discussion

The overall accuracy rate of approximately 90% achieved by the artificial intelligence model shows that the model performs quite consistently in the classification task. This high success is based on the balanced distribution of the generated dataset (Planar = DC = Angular) and the ability of the Convolutional Neural Network (CNN) architecture to efficiently learn the statistical and structural patterns contained in the 32x32 blocks.

However, the decision mechanism of the rule-based system based on certain thresholds (such as standard deviation and edge rate) is insufficient, especially for complex or boundary cases. This is evident in the low recall rate calculated for the DC class. Indeed, when the confusion matrix presented in Figure 5 is examined, it is seen that many instances belonging to the DC class are classified as Angular or Planar, meaning that the model cannot identify the instances belonging to this class precisely enough.

When the misclassifications are analyzed, it becomes difficult for both the rule-based system and the AI model to distinguish between Planar and DC classes, especially when the intra-block light transitions are soft or ambiguous. In addition, the structure of some blocks is located at the boundary values that can be interpreted as both flat (Planar) and constant mean (DC). This is an important factor that reduces the direct matching rates.

Regarding the dataset, although the limited number of images used may seem insufficient at first glance, 64 32x32 blocks were obtained from each image, resulting in training with more than 1900 block samples in total. This provided a significant advantage for the model to perform generalization learning. However, in the future, this study can be further strengthened by applying data augmentation methods. For example, diversifying the block structures by using variations such as brightness, contrast, transformation, etc. can make the distinction between DC and Planar modes more precise.

Furthermore, the fact that the samples on which the model is trained are obtained only from classic test images may limit the performance of the application on general video content. Models trained with larger block datasets from real-world videos can achieve a more robust generalization capacity.

Although there are various studies in the literature that offer CNN-based intra prediction or mode decision systems (e.g., Cui et al., 2018; Laude & Ostermann, 2016), these studies generally focus on coding efficiency metrics (e.g., BD-rate, coding time).

In this study, a comparative evaluation of deep learning and rule-based approaches in terms of classification performance and interpretability is presented. Additionally, this evaluation process is visually supported through a Unity-based interactive VR platform, enabling users to understand the decision-making mechanism in a more transparent manner.

Therefore, our goal is not to outperform existing models in terms of performance, but to provide a complementary tool for educational and analytical use.

In conclusion, although the success rate achieved in this study is high, the cases where the model fails are structurally meaningful and are due to structural ambiguities between classes. This highlights both the limitations of the rule-based approach and the room for improvement in artificial intelligence models.

In addition to the educational contribution of the VR-supported analysis study, it also has practical implications for future video coding workflows. The integration of explainable rule-based thresholds with CNN-based decision models could create a prototype for hybrid intra prediction in HEVC or VVC encoders. Such hybrid approaches can both increase mode selection efficiency and preserve explainability, which is critical for standardization. Future work may focus on hybrid models that combine adaptive threshold learning or statistical features with deep learning representations.

5. Conclusion

This study aims to classify the intra prediction modes (Planar, DC, Angular) in the High Efficiency Video Coding (HEVC) standard. For this purpose, a deep learning-based classification model and a rule-based expert system are evaluated comparatively. A dataset consisting of 32x32 blocks obtained from 30+ test images resized as gray level was created, and these blocks were labeled according to their respective coding modes. The model was trained with Keras using this balanced dataset and then converted to ONNX format and integrated into a Unity-based VR environment. Thus, users can visually analyze how each block is classified by both the AI and the rule-based system in the VR environment.

As a result of the experimental analysis, it was observed that the predictions made by the artificial intelligence model and the rule-based system matched 90% of the time. Especially in the classifications of the Angular mode, the F1 score of the model was calculated as 94% and considering the high number of total samples belonging to this class, the success of the model in distinguishing linear structures is quite remarkable. This success can be clearly seen in the confusion matrix presented in Figure 5. In the matrix, the number of correct and incorrect predictions of the artificial intelligence model in the Planar, DC and Angular classes are given in detail; especially the high accuracy rate in the Angular class is emphasized.

In addition, according to the classification report given in Table 3, the overall accuracy of the model is 89.1% and the macro F1 score is 88.6%. These results show that the model has a consistent and balanced performance across all classes. Furthermore, the misclassification histogram in Figure 6 provides a visual representation of the distribution of correct and incorrect predictions by class, providing an intuitive analysis of which modes the model is more successful or struggling. Since the classification results were obtained from multiple experimental runs using 5-fold cross-validation, statistical measures such as standard deviation and confidence intervals are not reported. In future work, broader statistical evaluations will be conducted using cross-validation or repeated testing.

The findings show that the AI model is able to learn and generalize more complex structures than rule-based approaches that rely only on basic statistical properties. This demonstrates the significant potential of deep learning in visual pattern-based processes such as intra prediction.

In conclusion, this study not only demonstrates the educational benefits of VR-supported visualization of intra prediction models, but also proposes a path for practical integration into future coding standards. The combination of CNN-based deep learning and explainable rule-based systems reveals the potential of hybrid solutions that can strike a balance between accuracy, efficiency, and explainability. We believe this study will be inspiring for both academic research and industrial applications in next-generation video coding systems.

5.1 Recommendations and Future Work

This study was conducted using standard gray level test images to ensure compatibility with traditional HEVC internal prediction mode structures. However, this choice also imposes limitations in terms of generalizability. In future studies, broader and more diverse data sets containing color images or different content categories could be used to evaluate the system's adaptability and robustness under different conditions.

Although the artificial intelligence-assisted forecasting system developed in this study has produced successful results, there are several extension possibilities to further improve this approach and apply it to different areas in the future. First of all, it is suggested to expand the dataset used to increase the generalization capacity of the model. In particular, including more test images as well as images with low contrast, noise or real-world data would increase the robustness of the model and allow testing its performance under different conditions.

However, the deep learning model and the rule-based system used in the current study are evaluated separately. In the future, a hybrid model combining the advantages of these two approaches could be proposed. In particular, a more robust classifier that combines both learned and hand-defined information can be developed by directly feeding the attributes (average brightness, standard deviation, edge rate, etc.) obtained from the rule-based system as input to the CNN model.

Furthermore, by converting the trained model into ONNX format, it is possible to run this system on mobile devices or web-based applications. Lightweight versions of the model to be integrated into such platforms could enable the development of user-friendly applications that perform real-time analysis of HEVC prediction modes.

Another important development direction is the evaluation of the created VR environment in terms of user experience. Users' feedback on the process of viewing predictions on blocks, understanding true/false labels and interpreting annotations will provide important data to measure the interactive success of the system. Thus, training or analysis platforms with increased interactivity can be designed.

In addition, evaluating the user experience of the developed VR-based interactive system can guide future work. Drawing on human-computer interaction (HCI) literature, the effectiveness of the system can be measured using metrics such as usability testing, task completion time, user satisfaction, and visual interpretability. This allows for a clearer assessment of the system's educational contribution and practical use.

Finally, although Planar, DC and Angular modes are classified in the current study, there are 33 different directional variants in the HEVC standard, especially for Angular mode. By classifying these sub-directions separately, a more detailed, fine-grained prediction architecture can be built. This approach will enable in-depth modeling and optimization of the decision-making processes of HEVC encoders.

In this study, the performance of deep learning and rule-based systems on HEVC intrinsic prediction modes is investigated in detail and made experienceable through an interactive VR platform. The findings demonstrate the usability of such integrated approaches in both training and visual analysis environments.

Author Contributions

The percentages of the author' contributions are presented below. The author reviewed and approved the final version of the manuscript.

	M.K.
C	100
D	100
S	100
DCP	100
DAI	100
L	100
W	100
CR	100
SR	100
PM	100
FA	100

C= concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The author declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because there was no study on animals or humans.

References

Adeniji I, Casarona M, Bielory L, Bancairen L, Menzel M, Perigo N, Blackmon C, Niepielko MG, Insley J, Joiner D. 2024. Using Unity for Scientific Visualization as a Coursebased Undergraduate Research Experience. J Comput Sci Educ, 15(1): 35-40. https://doi.org/10.22369/issn.2153-4136/15/1/7

Askin MB, Celikcan U. 2022. Learning based versus heuristic based: A comparative analysis of visual saliency prediction in immersive virtual reality. Comput Animat Virtual Worlds, 33(6): e2106. https://doi.org/10.1002/cav.2106

Cucchiara R, Piccardi M, Mello P. 2000. Image Analysis and rule-based reasoning for a traffic monitoring system. IEEE Transact Intel Transport Syst, 1(2): 119-130.

Cui W, Zhang T, Zhang S, Jiang F, Zuo W, Zhao D. 2018. Convolutional neural networks based intra prediction for HEVC. http://arxiv.org/abs/1808.05734

Feng A, Gao C, Li L, Liu D, Wu F. 2021. Cnn-based depth map prediction for fast block partitioning in hevc intra coding. IEEE International Conference on Multimedia and Expo (ICME), Jult 05-09, Shenzhen, China, pp: 1-6. https://doi.org/10.1109/ICME51207.2021.9428069

Galpin F, Racape F, Jaiswal S, Bordes P, Leannec F, Francois E. 2019. CNN-based driving of block partitioning for intra slices encoding. Data Compression Conference, March 26-29, Snowbird, UT, US, pp: 162-171. https://doi.org/10.1109/DCC.2019.00024

Imen W, Amna M, Fatma B, Ezahra SF, Masmoudi N. 2022. Fast HEVC intra-CU decision partition algorithm with modified LeNet-5 and AlexNet. Signal Image Video Proces, 16(7): 1811-1819. https://doi.org/10.1007/s11760-022-02139-w

Ishikawa S, Todo M, Taki M, Uchiyama Y, Matsunaga K, Lin P, Ogihara T, Yasui M. 2023. Example-based explainable AI and its application for remote sensing image classification. Int J Appl Earth Obser Geoinfo, 118: 103215. https://doi.org/10.1016/j.jag.2023.103215

Islam RU, Hossain MS, Andersson K. 2020. A deep learning inspired belief rule-based expert system. IEEE Access, 8: 190637-190651. https://doi.org/10.1109/ACCESS.2020.3031438

Kaplan M, Akman A. 2024. Parallel implementation of discrete cosine transform (DCT) methods on GPU for HEVC, 2nd International Congress of Electrical and Computer Engineering. Pp: 281–293. https://doi.org/10.1007/978-3-031-52760-9_20

Kuanar S, Rao KR, Bilas M, Bredow J. 2019. Adaptive CU mode selection in HEVC intra prediction: A deep learning approach. Circuits Syst Signal Proces, 38(11): 5081-5102. https://doi.org/10.1007/s00034-019-01110-4

Laude T, Ostermann J. 2016. Deep learning-based intra prediction mode decision for HEVC. 2016 Picture Coding Symposium (PCS), December 04-07, Nuremberg, Germany, pp: 1-5. https://doi.org/10.1109/PCS.2016.7906399

Li N, Wang Z, Zhang Q, He L, Zhang W. 2025. Fast intra-prediction mode decision algorithm for versatile video coding based on gradient and convolutional neural network. Electronics, 14(10): 2031. https://doi.org/10.3390/electronics14102031

Liu X, Li Y, Liu D, Wang P, Yang LT. 2019. An adaptive CU size decision algorithm for HEVC intra prediction based on complexity classification using machine learning. IEEE Transact Circuits Syst Video Technol, 29(1): 144-155. https://doi.org/10.1109/TCSVT.2017.2777903

Moran A, Gadepally V, Hubbell M, Kepner J. 2015. Improving big data visual analytics with interactive virtual reality. 2015 IEEE High Performance Extreme Computing Conference (HPEC), September 15-17, Waltham, MA, US, pp: 1-6. https://doi.org/10.1109/HPEC.2015.7322473

Pattimi H, Srinivasarao BKN. 2024. High-speed coding unit depth identifications using CU-VGG deep learning architectures. Arabian J Sci Eng, 49: 16287-16298. https://doi.org/10.1007/s13369-024-08928-4

Ren W, Su J, Sun C, Shi Z. 2019. An IBP-CNN Based fast block partition for intra prediction. 2019 Picture Coding Symposium (PCS), November 12-15, Ningbo, China, pp: 1-5. https://doi.org/10.1109/PCS48520.2019.8954522

Schorr C, Goodarzi P, Chen F, Dahmen T. 2021. Neuroscope: An explainable AI toolbox for semantic segmentation and image classification of convolutional neural nets. Appl Sci, 11(5): 2199. https://doi.org/10.3390/app11052199

Sullivan GJ, Ohm JR, Han WJ, Wiegand T. 2012. Overview of the high efficiency video coding (HEVC) standard. IEEE Transact Circuits Syst Video Technol, 22(12): 1649-1668. https://doi.org/10.1109/TCSVT.2012.2221191

Swamy TN, Ramesha K. 2022. Improved intra prediction algorithm for HEVC with conventional and convolutional neural network approach. Math Stat Eng Appl, 71(2): 272-280.

Zaki F, Mohamed AE, Sayed SG. 2021. CtuNet: A deep learning-based Framework for Fast CTU Partitioning of H265/HEVC Intra- coding. Ain Shams Eng J, 12(2): 1859-1866. https://doi.org/10.1016/j.asej.2021.01.001

Zhao L, Zhang L, Ma S, Zhao D. 2011. Fast mode decision algorithm for intra prediction in HEVC. 2011 Visual Communications and Image Processing (VCIP), November 06-09, Tainan, Taiwan, pp: 1-4. https://doi.org/10.1109/VCIP.2011.6115979

Zisad SN, Chowdhury E, Hossain MS, Islam RU, Andersson K. 2021. An integrated deep learning and belief rule-based expert system for visual sentiment analysis under uncertainty. Algorithms, 14(7): 213. https://doi.org/10.3390/a14070213