Journal of the Brazilian Computer Society, 2025, 31:1, doi: 10.5753/jbcs.2025.5464

MLISP: Machine-Learning-based ISP Decision Scheme for VVC Encoders

Larissa Araújo [**Federal University of Pelotas (UFPel)** | ***ldaaraujo@inf.ufpel.edu.br***]
Adson Duarte [**Federal University of Pelotas (UFPel)** | ***airduarte@inf.ufpel.edu.br***]
Bruno Zatt [**Federal University of Pelotas (UFPel)** | ***zatt@inf.ufpel.edu.br***]
Guilherme Correa [**Federal University of Pelotas (UFPel)** | ***gcorrea@inf.ufpel.edu.br***]
Daniel Palomino [**Federal University of Pelotas (UFPel)** | ***dpalomino@inf.ufpel.edu.br***]

Video Technology Research Group (ViTech), Graduate Program in Computer Science (PPGC), Federal University of Pelotas (UFPel), Rua Gomes Carneiro, 1, Centro, Pelotas, RS, 96010-610, Brazil.

Received: 21 January 2025 • **Accepted:** 30 July 2025 • **Published:** 27 October 2025

Abstract The Versatile Video Coding (VVC) standard achieves high compression rates by introducing new encoding tools, such as the Intra Subpartition Prediction (ISP). However, the ISP increases the computational effort necessary to perform the mode decision in the intra-prediction step. In this paper, we propose the **MLISP**, a machine learning-based ISP decision scheme for VVC encoders where two solutions are adopted to accelerate the intra-mode decision process for the ISP tool. The first solution, named **ISP Skip Decision**, utilizes a Decision Tree trained with image features that predicts whether the evaluation of the ISP tool is necessary, resulting in an average time saving of 8.53% with only 0.22% of coding efficiency loss. The second solution called **ISP Mode Decision**, uses a Decision Tree trained with encoding features to predict the optimal class of intra modes between Planar/DC and Angular to be evaluated with the ISP tool, obtaining an average time saving of 7.01% with only 0.19% of coding efficiency loss. By combining these solutions, MLISP achieves an average time saving of 10.97% with only 0.32% loss in coding efficiency, demonstrating its effectiveness in reducing encoding time with minimal impact on compression performance. Compared with related works, MLISP achieves competitive results and introduces a novel approach for optimizing the ISP decision.

Keywords: VVC, Intra Prediction, ISP, Machine Learning

1 Introduction

Digital videos have been fundamental in many areas, from entertainment and communication to surveillance applications and live broadcasts. A study reveals that during the third quarter of 2022, live streams featuring gaming-related content accumulated approximately 7.2 billion hours of content watched across leading streaming platforms [Ceci, 2023]. In this context, video coding standards such as the Versatile Video Coding (VVC) [Bross *et al.*, 2021] play a crucial role in enabling applications to manipulate high-definition videos for storage and transmission.

The VVC [Bross *et al.*, 2021] is one of the latest and most advanced video coding standards. It offers superior bit-rate reduction without compromising visual quality compared to its predecessor, the High Efficiency Video Coding (HEVC) standard [Siqueira *et al.*, 2020]. This is possible due to several new encoding tools introduced in the standard, especially in the intra-prediction step of VVC. While VVC maintains the Planar, DC, and Angular directional modes from HEVC, it extends the number of Angular modes from 33 to 65. VVC also introduces the Matrix-weighted Intra Prediction (MIP) [Schäfer *et al.*, 2019] and the Intra Subpartition Prediction (ISP) [De-Luxán-Hernández *et al.*, 2019] to improve prediction accuracy. Combined with the Planar, DC, and Angular modes, the ISP tool enhances prediction granularity by processing a block through subpartitions in the horizontal or vertical directions. Despite the coding efficiency improvements of VVC, typical encoder implementations of VVC, such as VVC Test Model (VTM) [Bossen *et al.*, 2018], exhibit a trade-off in encoding time, making them up to 34 times slower than typical HEVC encoder implementations like HM [Mercat *et al.*, 2021]. Therefore, it is essential to develop solutions to improve the encoding time by targeting the new intra-prediction tools in VVC, such as the ISP tool.

Some works propose solutions to save time in the intra-mode decision in VVC, specifically focusing on the ISP tool. The main idea of these works is to avoid the costly evaluation of the Rate-Distortion Optimization (RDO) [Sullivan and Wiegand, 1998] process for ISP modes that are less likely to be optimal. The works usually use heuristics or machine-learning solutions to predict the most promising modes. For example, in [Park *et al.*, 2022], a machine-learning model is trained with a key feature computed over the image known as the Mean Absolute Sum of Transform coefficients. The model predicts whether the evaluation for each ISP mode is necessary or can be skipped. Another approach, presented in [Saldanha *et al.*, 2021], involves setting a threshold based on an image characteristic: the variance of the block. The solution avoids evaluating all ISP candidates whenever the variance exceeds the threshold. In [Liu *et al.*, 2021], the texture complexity of the block is obtained through an image feature called Mean Absolute Deviation. Then, this feature is used to decide whether the evaluation of ISP candidates can be skipped. A heuristic is proposed in [Park *et al.*, 2020], where the list of ISP candidates is pruned according to the shape of the block and the ISP subpartition direction.

Although prior works report time-saving results for the ISP mode decision, most of them, including [Saldanha *et al.*, 2021] and [Liu *et al.*, 2021], rely on computing image features over the entire block. However, since the ISP tool performs the intra prediction on subpartitions within the block, features extracted at the subpartition level are also important. Notably, the approach in [Park *et al.*, 2022] is the only one that computes image features specifically for subpartitions. Yet, these features are significantly more complex than those in [Saldanha *et al.*, 2021] and [Liu *et al.*, 2021], as they involve applying a transform to the block. In addition, none of the related works leverage features inherently available during the encoding process, such as rate-distortion costs and the best intra modes of neighboring blocks, which we refer to as "encoding features". These gaps highlight the potential for developing solutions that explore both encoding and simple image features computed at block and subpartition levels, addressing the limitations of prior approaches while maintaining computational efficiency.

This paper proposes MLISP: a Machine-Learning-based ISP Decision Scheme, which combines two solutions to accelerate the ISP mode decision process in the VVC intra prediction. The first solution, named **ISP Skip Decision**, employs a Decision Tree trained on image features computed over the entire block and its subpartitions to predict whether the evaluation of ISP is necessary, classifying the blocks into two categories: (i) Non-ISP and (ii) ISP. When the ISP Skip Decision solution determines that ISP modes should be evaluated, the second solution, named **ISP Mode Decision**, utilizes a Decision Tree trained on encoding features to predict the most suitable class of intra modes to be evaluated with the ISP tool, classifying the blocks into two categories: (i) ISP Planar/DC and (ii) ISP Angular. By integrating these two complementary solutions, MLISP effectively reduces the overall encoding time while maintaining negligible losses in coding efficiency.

The ISP Mode Decision solution was previously published in [Araújo *et al.*, 2024]. However, in this work, we enhanced the training methodology of the Decision Tree by introducing a feature selection step, which led to further improvements in the encoding time reductions achieved by the solution. Additionally, we propose a novel solution called ISP Skip Decision and the comprehensive MLISP Decision Scheme, combining both solutions to maximize the encoding time reductions.

2 VVC Intra Subpartition Prediction

VVC introduced several innovations in the intra prediction. Firstly, it introduced the possibility of performing intra prediction on rectangular blocks Huang *et al.* [2020], a feature not available in previous standards such as HEVC, where rectangular blocks were only used in the transform stage. This enhancement, combined with the support for square blocks, results in a total of 17 possible block sizes for intra prediction. Considering the intra modes, the Planar and DC modes were preserved from the previous HEVC standard, while VVC expanded the Angular modes from 33 to 65, which are indicated by the red arrows in Figure 1. VVC also introduced a novel family of intra modes called MIP [Schäfer *et al.*, 2019]. Alongside these, new tools that can be combined with the Planar, DC, and Angular modes were incorporated, such as the Multiple Reference Line (MRL) [Chang *et al.*, 2019], which extends the number of available reference samples for intra prediction and is showcased in Figure 2, and the ISP [De-Luxán-Hernández *et al.*, 2019], shown in Figure 3, which is the primary focus of this work.

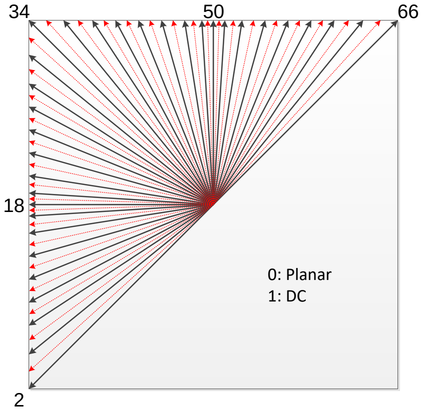


Figure 1. Intra modes in VVC.

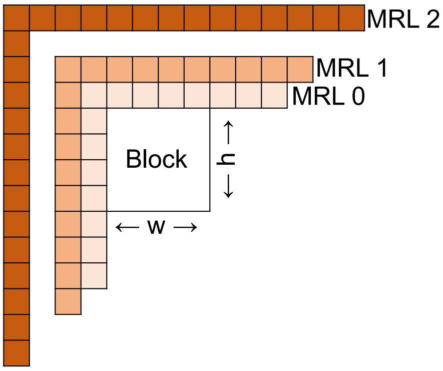
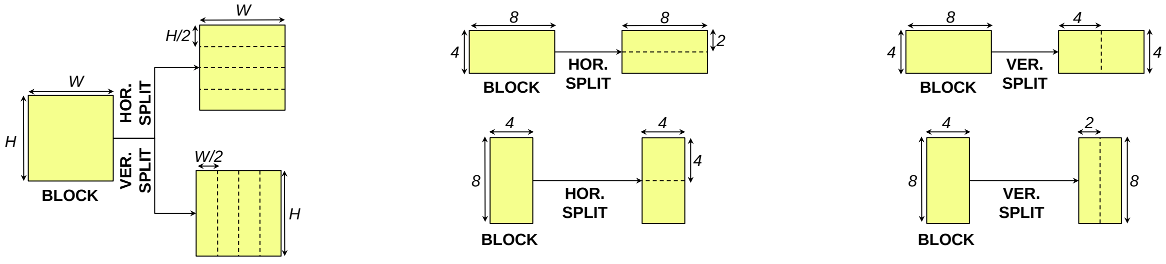


Figure 2. Multiple Reference Line tool.

The ISP tool enables more granular block prediction. The ISP tool horizontally or vertically divides the original image block into two or four subpartitions, depending on the block size, as illustrated in Figure 3. For blocks sized 8x4 and 4x8, the ISP tool generates only two subpartitions, either in the horizontal or vertical direction, as shown in Figure 3(b) and Figure 3(c), respectively. This restriction ensures that each subpartition contains at least 16 samples.

For other block shapes, the ISP tool generates four subpartitions in either the horizontal or vertical direction, as shown in Figure 3(a). The prediction process for each subpartition occurs sequentially, and samples generated from the prediction of one subpartition are used as reference samples for the prediction of the next subpartitions.

While predictions are conducted individually for each subpartition, they must converge to the same intra mode, such as Planar, DC, or one of the Angular modes. This is done to improve the prediction accuracy of each subpartition and also to save the bits necessary from signaling the intra modes in the bitstream since only one mode will be signaled. The



(a) Four subpartitions. **(b)** Two horizontal subpartitions. **(c)** Two vertical subpartitions.

Figure 3. Intra Subpartitions Prediction Tool.

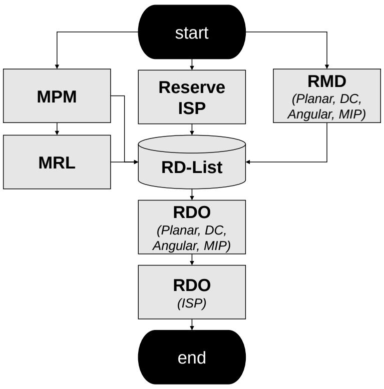


Figure 4. VTM standard intra mode decision.

addition of the ISP can improve coding efficiency by approximately 0.57% with a 12% increase in encoding time [De-Luxán-Hernández *et al.*, 2019]. This increase in encoding time occurs because the ISP tool introduces an additional step in the intra-mode decision process of VVC Test Model (VTM) [Bossen *et al.*, 2018], the reference software that implements VVC. In this process, the encoder evaluates the Planar, DC, and Angular modes twice. First, a list of these modes is evaluated for the entire block by the RDO, and then, they are evaluated again for each possible ISP subpartition.

The intra-mode decision process determines the best intra mode for each block by evaluating several possible combinations through the Rate-Distortion Optimization (RDO) process [Sullivan and Wiegand, 1998]. However, this process is computationally intensive since the encoder must evaluate many intra-mode candidates. For each one of these modes, the rate-distortion cost must be computed by the intra-mode decision. This cost is available only after the encoding steps, including prediction, direct and inverse transformation and quantization, and entropy coding. To handle this complexity, VTM [Bossen *et al.*, 2018] incorporates the Rough Mode Decision (RMD) [Zhao *et al.*, 2011] and the Most Probable Modes (MPM) list [Pfaff *et al.*, 2021]. The main idea behind the RMD and MPM steps is to generate the RD-List, a subset of the most promising modes. Only this subset is evaluated by the RDO process, avoiding evaluating modes less likely to be optimal.

VTM performs the intra-mode decision following Figure 4 steps. The RMD, MPM, and MRL steps jointly select up to eight intra modes to compose the RD-List. Both the RMD and MRL steps compute fast rate-distortion costs for the intra modes, selecting the six best ones. However, while the RMD evaluates the Planar, DC, Angular, and MIP modes with MRL set to zero, as shown in Figure 2, the MRL step evaluates the MPM modes twice. One with the MRL set to one and then with MRL set to two. The MPM step generates a list of six intra modes likely optimal for the current block based on the best intra modes in neighboring blocks [Pfaff *et al.*, 2021]. The first MPM mode is always the Planar mode, while the remaining ones can be the DC or one of the Angular modes. If the first two MPM modes are not already in the RD-List, the MPM step adds them. At the end of the RD-List, the encoder reserves 16 positions evenly distributed between the horizontal and vertical subpartitions for ISP. VTM includes the same intra modes obtained from the RMD and MPM steps in the horizontal and vertical reserved positions, excluding the MIP modes. In addition, three Angular modes with the lowest costs during the RMD step, excluding the ones already in the RD-List, are also selected for the ISP evaluation. Then, the RDO evaluates all non-ISP modes (Planar, DC, Angular, and MIP modes) and only then starts the evaluation of the ISP modes.

While the RD-List is a subset of the most promising intra modes, only one will yield the best result in terms of coding efficiency. Even when we consider only the subset of ISP candidates present in the RD-List, the RDO must evaluate up to 16 modes for a single block to decide the best ISP mode. In this context, there is a need for solutions that target reducing the number of modes to be evaluated by the RDO process. The solutions must accurately predict the most promising ISP modes to reduce the computational effort required for the ISP mode decision with minimal loss of compression efficiency.

3 Related Works

To reduce the number of intra modes evaluated during the RDO process, many works adopt heuristic-based solutions, most of which target decisions regarding the Planar, DC, or Angular modes, as described below. For instance, [Yang *et al.*, 2020] employs a gradient descent search to propose a fast intra-mode decision for Planar, DC, and Angular modes. The authors in [Chen *et al.*, 2020] propose a linear model that maps fast RMD costs to full RDO costs, allowing modes with estimated high RDO costs to be skipped. In [Zhang *et al.*, 2020], the texture direction of the block is used to reduce the number of Angular modes evaluated during the RMD step. Additionally, during the RDO step, any intra mode with an RMD cost higher than one of the MPMs is excluded from evaluation. Although these works achieve good results, they are all developed for earlier versions of VTM, which do not include the ISP modes.

There are solutions focusing on ISP modes, which is also the focus of this work, and they can be proposed specifically for ISP or for a broader set of modes, including ISP. For example, [Saldanha *et al.*, 2021] introduces three solutions to reduce the number of RDO evaluations for Angular, MIP, and ISP modes. The first and second solutions employ Decision Trees trained with encoding features to predict whether RDO evaluation for Angular and MIP modes can be skipped, respectively. The third solution uses the block variance predict whether the evaluation of ISP candidates can be skipped according. If the variance exceeds a certain threshold, all ISP candidate evaluations are skipped.

In [Liu *et al.*, 2021], block texture complexity is estimated through the Mean Absolute Deviation (MAD), an image feature used to decide whether ISP candidate evaluation can be skipped. A heuristic is proposed in [Park *et al.*, 2020], where the list of ISP candidates is pruned based on the block shape and ISP sub-partition direction.

The solution in [Park *et al.*, 2022] predicts whether RDO evaluation is required for each ISP mode by employing a lightweight Gradient Boosting model trained with an image feature called the Mean Absolute Sum of Transform coefficients. This feature captures block characteristics, enabling the model to decide if the evaluation of each ISP mode can be avoided.

The authors in [Dong *et al.*, 2022] propose a scheme containing three solutions. First, Decision Trees are trained with image and encoding features, such as texture complexity and best neighboring intra modes, to predict whether the evaluation of ISP and Intra Block Copy (IBC) tools in VVC is necessary. Second, a probability model is used to reorder the RD list so that modes with the highest probability of being optimal are evaluated first. A threshold is then applied to early terminate the RDO evaluation once the probability of the optimal mode being found is high enough. Finally, the authors propose an early termination solution for block splitting, where models trained based on the best mode at the current depth predict whether further depth evaluations are required.

In [Liu *et al.*, 2023], the authors introduce a deep learning-based solution employing a Convolutional Neural Network (CNN) similar to ResNet. The CNN is trained using image blocks and their reference samples to predict probabilities for each intra mode in VVC. Based on these probabilities, only the most promising intra modes undergo evaluation in both RMD and RDO steps, significantly reducing the number of mode evaluations.

It is worth noting that we found only six works that are proposing solutions targeting the ISP modes. This reflects the fact that research efforts focusing on the ISP tool remain relatively scarce, especially when compared to the broader set of works addressing other intra prediction modes in VVC. For works specifically targeting ISP, most of the proposed approaches rely exclusively on image features to guide the decision process, without exploring encoding features in the context of an ISP mode decision solution, which could provide additional context regarding the RD behavior of the modes. Furthermore, these image features are generally computed over the entire block, rather than at the level of ISP subpartitions, which limits the granularity of the information exploited during the decision process. This sparsity of research targeting ISP, combined with the limitations observed in existing approaches, further motivates the development of new solutions such as the one proposed in this work.

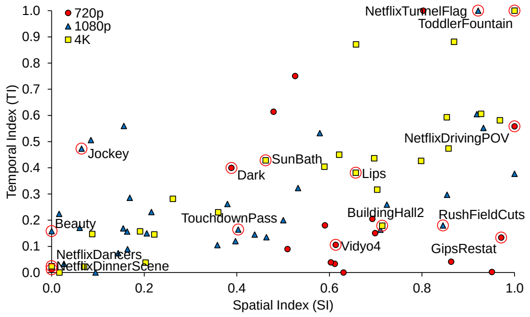


Figure 5. Selected videos according to their SI and TI values.

4 ISP Occurrence Rate Analyses

We conducted three analyses of the ISP tool in VVC. In the first analysis, we evaluated the overall occurrence rate of ISP, i.e., how often the final intra-mode decision of VTM selects an ISP mode as the best choice for a given block. To this end, we grouped all VVC intra modes into two classes: **(i) Non-ISP**, including samples where Planar, DC, Angular, or MIP modes are selected, and **(ii) ISP**, including samples where an ISP mode is selected. The ISP class includes modes with horizontal or vertical subpartitions, combined with Planar, DC, or any of the Angular modes.

In the second and third analyses, we focused on the occurrence rates of specific intra modes during the ISP step and on the number of evaluations performed for these modes. In these analyses, the ISP modes were further grouped into two classes: ***(i) ISP Planar/DC*** and ***(ii) ISP Angular***. This classification is based on the distinct nature of these modes: Planar and DC modes are better suited for homogeneous textures, while Angular modes are more effective for directional textures.

In the first two analyses, we computed the occurrence rates of each class, organized by block size, to identify scenarios where specific mode evaluations could be avoided. In the third analysis, we focused on the number of evaluations performed for the ISP Angular class to understand the frequency of evaluating these modes during the ISP step.

For all three analyses, we selected the same 15 video sequences used in [Duarte *et al.*, 2023], which are highlighted with red circles in Figure 5. These videos were selected for their diversity in motion and texture, based on the Spatial Information (SI) and Temporal Information (TI) metrics [ITU, 2023]. Each video was encoded using VTM 18.0 [Bossen *et al.*, 2018] with the *All Intra* configuration and four Quantization Parameters (QP): 22, 27, 32, and 37. To collect the necessary data, we modified the VTM code, inserting routines to extract the final intra-mode decisions and the number of ISP candidates in the RD-List for each block.

Figure 6 presents the occurrence rate at which the ***(i) Non-ISP*** and ***(ii) ISP*** classes yield the best rate-distortion result, categorized by block size. We observe that, for all block sizes, the ***(i) Non-ISP*** class consistently has a higher occurrence rate compared to the ***(ii) ISP*** class. For blocks of size 64x64, the ***(ii) ISP*** class achieves a more competitive occurrence rate,

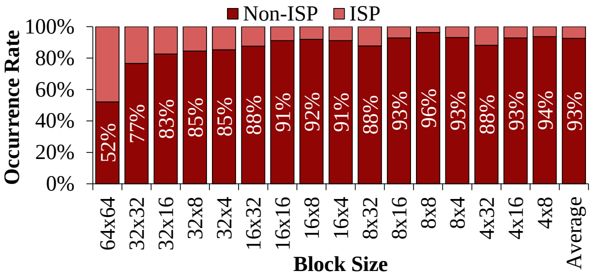


Figure 6. Occurrence rate of Non-ISP and ISP classes by block size.

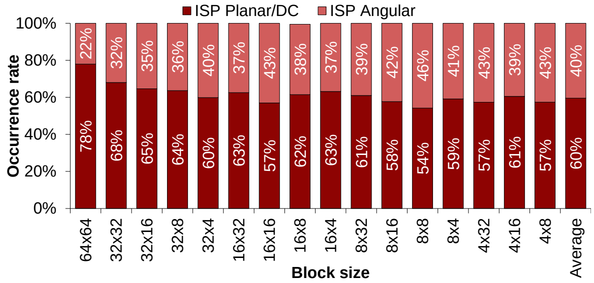


Figure 7. Occurrence rate of ISP Planar/DC and ISP Angular classes by block size.

approaching that of the ***(i) Non-ISP*** class. However, as the block size decreases, the occurrence rate of the ***(ii) ISP*** class also decreases. This suggests that, generally, when VTM performs the intra-mode decision for a given block, a Non-ISP mode is more likely to achieve the best rate-distortion result than an ISP mode. On average, the ***(i) Non-ISP*** class achieves an occurrence rate of 93%, while the ***(ii) ISP*** class achieves only 7%. In other words, there is a high probability that a Non-ISP mode will yield the best rate-distortion result, and the evaluation of ISP modes could potentially be avoided 93% of the time, saving encoding time with minimal impact on coding efficiency.

The second analysis illustrated in Figure 7 presents the occurrence rate in which the ***(i) ISP Planar/DC*** and ***(ii) ISP Angular*** classes yield the best rate-distortion result by block size. We can observe that the ***(i) ISP Planar/DC*** class has a higher occurrence rate when compared to the ***(ii) ISP Angular*** class for larger block sizes, particularly for 64x64 and 32x32. Furthermore, the ***(i) ISP Planar/DC*** class maintains a higher occurrence rate across all block sizes, obtaining 60% of the occurrence rate on average. In other words, when the ISP is chosen for prediction, the Planar and DC modes are more likely to be chosen.

In the third analysis in Figure 8, we can see the frequency of ISP candidates associated with Angular modes. We can observe that the number of ISP candidates associated with angular modes is always even. That happens because the VTM software evaluates the same intra modes for horizontal and vertical subpartitions in the ISP. Besides, it is possible to notice that in most cases, there are 8, 10, and 12 ISP candidates associated with angular modes. This means that, in most cases, VTM will evaluate 8 to 12 ISP candidates associated with angular modes for a single block. This finding highlights the potential for reducing the computational effort of the intra mode decision in VVC by early predicting when the encoder can avoid evaluating the ISP candidates associated with the Angular modes.

In summary, considering the high occurrence rate of the ***(i) Non-ISP class***, machine-learning-based solutions designed to predict when the evaluation of the ISP modes can be avoided are promising since the ISP modes are always evaluated even though in most of the cases they do not achieve the best rate-distortion result. Furthermore, considering the high occurrence rate of the ***(i) ISP Planar/DC*** class and the high frequency of the Angular ISP candidates in the RD-List, it is possible to reduce the computational effort of the ISP mode decision if an accurate machine learning model is employed to predict when the best ISP mode is Planar or DC. When this happens, the RDO evaluation of many Angular ISP candidates can be skipped to save encoding time while minimizing the final coding efficiency loss.

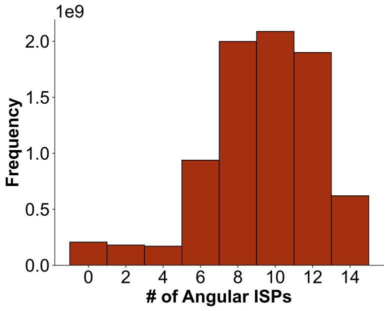


Figure 8. Frequency of angular ISPs in the RD-List.

5 MLISP: Machine-Learning-Based ISP Decision Scheme

This paper proposes MLISP, a machine-learning-based decision scheme designed to enhance the intra-mode decision process for the ISP tool in VVC encoders. MLISP comprises two solutions, developed based on the analysis results presented earlier, to address specific subproblems within the ISP decision process and achieve significant time savings.

The first solution, **ISP Skip Decision**, employs a Decision Tree model trained on image features to classify blocks into two classes: ***(i) Non-ISP***, representing blocks where the best rate-distortion cost is achieved by Planar, DC, Angular, or MIP modes, and ***(ii) ISP***, representing blocks where an ISP mode yields the best rate-distortion cost. This classification enables **ISP Skip Decision** to determine whether evaluating the ISP modes is necessary for each block, effectively skipping unnecessary RDO evaluations when the ***Non-ISP*** class is predicted.

If **ISP Skip Decision** predicts the ***ISP*** class, indicating that ISP evaluation is required, the second solution, **ISP Mode Decision**, is activated. This solution uses a Decision Tree trained with encoding features to classify blocks into two additional classes: ***(i) ISP Planar/DC***, where ISP combined with Planar or DC yields the best rate-distortion cost, and ***(ii) ISP Angular***, where ISP combined with an Angular mode provides the best cost. For blocks classified as ***ISP Planar/DC***, only the Planar and DC modes are evaluated using the ISP tool, reducing the computational effort in the RDO process for the ISP modes. We decided never to remove the ISP candidates associated with Planar/DC modes to bring a balance between the time reduction and the coding efficiency loss, given the high occurrence rate of the ISP Planar/DC class presented in Figure 7.

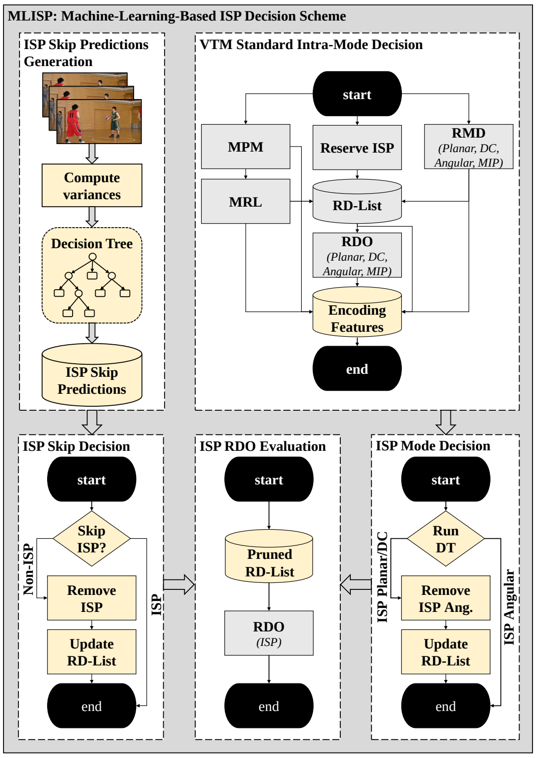


Figure 9. MLISP: Machine-learning-based ISP Decision Scheme solution.

Figure 9 illustrates the integration of these solutions into the standard VTM intra-mode decision process. Initially, the intra-mode decision follows the standard procedure, progressing through the RDO evaluation for Planar, DC, Angular, and MIP modes. During this step, encoding features—such as rate-distortion costs and the best neighboring intra modes—are extracted to serve as inputs for the Decision Tree of **ISP Mode Decision**. Simultaneously, raw YUV video frames are processed independently to generate luminance blocks of all possible sizes defined by the VVC intra prediction. For each block, image features are calculated across the entire block and its horizontal and vertical subpartitions, as depicted in Figure 3. These image features are inputs for the Decision Tree of **ISP Skip Decision**, which generates predictions for each block within the frame, deciding whether ISP evaluation is needed.

When applied during encoding, the two solutions interact as follows: **ISP Skip Decision** first predicts whether ISP modes should be considered for the current block. If the prediction is ***Non-ISP***, all ISP modes are excluded from the RD-List, skipping the ISP evaluation entirely. Conversely, if the prediction is ***ISP***, the RD-List retains the ISP modes, and **ISP Mode Decision** determines the next steps. If **ISP Mode Decision** predicts the ***ISP Planar/DC*** class, ISP candidates associated with Angular modes are removed from the RD-List, restricting the RDO evaluation to Planar and DC modes only. If the prediction is ***ISP Angular***, the RD-List remains unchanged, and all ISP candidates proceed to evaluation.

By combining these two solutions, the MLISP scheme optimizes the RD-List during the RDO stage for ISP modes, either by excluding all ISP modes or by eliminating Angular ISP candidates when deemed unnecessary. This effective pruning of the RD-List significantly reduces the computational effort involved in ISP RDO evaluation while maintaining minimal losses in coding efficiency.

The following sections will detail the dataset construction process for these solutions, the methodology for feature selection, the behavior of features with the highest information gains, the training procedures for the Decision Trees, and the accuracy results achieved by the proposed models.

5.1 Feature Extraction and Dataset Generation

To create the datasets required for training the Decision Trees, we encoded the same 15 videos used in the analysis described in Section 4. The encodings were performed using VTM version 18.0 with the *All Intra* configuration and four QP values: 22, 27, 32, and 37. For each processed block, we extracted the encoding features listed in Table 1 along with the corresponding final intra-mode decision. Subsequently, we generated balanced datasets for each proposed solution by grouping the samples into the desired classes according to the final intra-mode decision. Then, we performed a subsampling step to ensure we had datasets balanced by video, QP, block size, and class. It is important to acknowledge that the dataset used in this work may inherit some bias from the VTM encoder configuration, as VTM employs fast tools such as RMD and MPM (see Section 2), which prune certain intra modes during the RDO process and do not evaluate all possible intra modes. Consequently, for some samples, the final intra-mode decision in the dataset might differ if all intra modes were exhaustively tested. However, it should be noted that all related works use datasets constructed under the same conditions, making comparisons fair within this context.

5.1.1 ISP Skip Decision Dataset

The ISP tool performs the prediction over subpartitions, utilizing the reference samples generated after predicting one subpartition to refine the prediction of the next. This approach aims to improve prediction efficiency in cases where Planar, DC, Angular, or MIP modes applied to the entire block may struggle to accurately capture variations within the block. Since the ISP tool operates at the subpartition level, the variance of the entire block and its subpartitions can indicate the texture complexity of the block [Saldanha *et al.*, 2021] [Dong *et al.*, 2022], which may help determine whether the ISP tool will likely improve the prediction for a given block.

Therefore, inspired by the work of [Saldanha *et al.*, 2021], the dataset for the **ISP Skip Decision** solution includes image features based on the variance of the original luminance blocks, as a measure to indicate the block texture. However, unlike [Saldanha *et al.*, 2021], which only considers the variance of the entire block, we also compute the variances of the horizontal and vertical subpartitions defined by the ISP modes. This provides additional information that may indicate whether applying ISP is appropriate. Furthermore, our approach seeks a balance between the simplicity of the variance-based feature from [Saldanha *et al.*, 2021] and the more complex approach proposed in [Park *et al.*, 2022], which also analyzes the subpartitions but requires computing the Mean Absolute Sum of Transform coefficients in the frequency domain. By using the variance of the subpartitions, we

Table 1. Encoding features extracted from VTM.

Name	Description	Step	Type	Min	Max	# of values
BlockPosition	The *x* and *y* block positions.	RMD	Integer	0	4096	2
BestAngular	Best angular mode.	RMD	Integer	2	66	1
BestMIP	Best MIP mode.	RMD	Integer	0	7	1
MRLAngular	MRL reference line from the best angular mode.	RMD	Integer	0	2	1
MRLDC	MRL reference line from the best DC mode.	RMD	Integer	0	2	1
ModesMPM	MPM modes excluding Planar.	MPM	Integer	1	66	5
ModesPosition	First occurrence position of each intra mode type.	RD-List	Integer	1	12	4
FirstAngular	First angular mode in the RD-List.	RD-List	Integer	2	66	1
FirstMIP	First MIP mode in the RD-List.	RD-List	Integer	0	7	1
NeighborMode	Best intra mode number in neighboring blocks.	MPM	Integer	0	66	2
NeighborType	Best intra mode type in neighboring blocks.	MPM	Boolean	0	1	8
DCMPM	DC mode is an MPM.	MPM	Boolean	0	1	1
SAD	Sum of Absolute Differences.	RMD	Decimal	0.63	1390.38	4
SATD	Sum of Absolute Transformed Differences.	RMD	Decimal	0.52	497.81	4
FracBits	Estimated number of bits.	RMD	Decimal	1.19	15669.28	4
RMD Cost	Fast rate-distortion cost.	RMD	Decimal	0.65	502.01	4
RDO Cost	Complete rate-distortion cost.	RDO	Decimal	182.14	1881715.88	4
					Total	**48**

obtain additional spatial information about the block structure with a much lower computational cost.

The block variance is calculated using Equation 1, where B represents the block, x_i denotes the sample values, μ is the mean of the samples, and N is the total number of samples in the block.

$$Var(B) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2 \qquad (1)$$

For 4x8 and 8x4 blocks, there are fewer subpartitions compared to other block sizes (only two subpartitions in the horizontal and vertical directions, as opposed to four in the remaining block sizes). Consequently, there are five variance features for 4x8 and 8x4 blocks: the variance of the entire block, two variances for the two horizontal subpartitions, and two variances for the two vertical subpartitions. In contrast, for the other block sizes, there are nine variance features: the variance of the entire block, four variances for the horizontal subpartitions, and four variances for the vertical subpartitions. To account for this difference, one dataset is created for 4x8 and 8x4 block sizes, and another dataset is created for the remaining block sizes. Thus, the **ISP Skip Decision** solution utilizes two separate Decision Trees. Both datasets are balanced by video, QP, block size, and class; however, the 4x8 and 8x4 block sizes dataset contains about 100,000 samples, while the dataset for the other block sizes contains around 700,000 samples. In addition to the variance features, we include the width, height, and QP of the block as features, given that the datasets contain a mix of block sizes and QP values.

5.1.2 ISP Mode Decision Dataset

The dataset for the **ISP Mode Decision** solution includes the encoding features listed in Table 1, which were extracted from the VTM encodings as described earlier in this section. Each feature is presented with its name, description, the stage of the encoding process where it is extracted (RMD, MPM, RD-List, or RDO), its type (Boolean, Decimal, or Integer), the *min* and *max* values observed, and the number of values the feature provides.

We chose to use such encoding features because similar features have been successfully adopted in previous works, but for different objectives. For example, [Duarte *et al.*, 2022] uses encoding features to perform early prediction of affine motion estimation in inter prediction, while [Saldanha *et al.*, 2021] applies them to early predict the Angular and MIP modes in intra prediction. In this work, we apply the same principle to a different context, proposing a solution that uses encoding features to improve the decision process specifically for ISP modes.

RMD computes fast rate-distortion costs for the Planar, DC, Angular, and MIP modes. These fast costs hint at whether the ISP Planar/DC or ISP Angular classes will produce the best cost. This way, considering the best costs in RMD for the Planar, DC, Angular, and MIP modes, we extract the Sum of Absolute Differences (SAD), the Sum of Absolute Transformed Differences (SATD), the estimated number of bits, and the fast rate-distortion costs. Besides that, we also extract the *x* and *y* positions of the block, the best angular and MIP modes, and the best angular and DC MRL [Chang *et al.*, 2019] numbers.

The MPM provides a list containing six intra modes. The first one is always the Planar; the remaining modes will be the DC or one of the Angular modes. These six intra modes are likely the best ones since they were the best for the left and upper neighboring blocks. Therefore, from the MPM, we extract the number of the five selected intra modes, excluding the Planar as it is constant, the number of the best intra modes in the left and upper neighboring blocks, eight boolean values indicating whether the best intra mode for the left and upper neighboring blocks is Planar, DC, Angular, or MIP, and a boolean value indicating if the DC is an MPM.

The VTM software distributes the non-ISP modes in the RD-List in ascending order according to their fast rate-distortion costs obtained from the RMD step. This way, the modes distribution order in the RD-List can also hint if the

ISP Planar/DC or ISP Angular classes will likely provide the best rate-distortion cost. Therefore, from the RD-List, we extract the position of the first occurrence of Planar, DC, Angular, and MIP modes and also the mode number of the first Angular and MIP modes.

Finally, since the RDO evaluation for non-ISP modes occurs before evaluating the ISP modes, all the complete rate-distortion costs computed by the RDO for non-ISP modes are available. This way, we extract the best rate-distortion costs obtained by the RDO step for the Planar, DC, Angular, and MIP modes. It is essential to highlight that despite the high number of encoding features, there is no need for additional computations since they are all available during the encoding process.

In the **ISP Mode Decision** solution, we have a single dataset for all block sizes, given that the number of features is always the same. This dataset contains approximately 800,000 examples, balanced by video, QP, block size, and class. However, we applied normalization to the rate-distortion cost-related features according to Equation (2). In this equation, X represents the set of rate-distortion cost-related features, which include SAD, SATD, FracBits, RMD Cost, and RDO Cost listed in Table 1. Additionally, w and h refer to the width and height of the block, respectively.

$$x_{\text{norm}} = \frac{x}{w \cdot h}, \quad x \in X, \quad w, h \in \{4, 8, 16, 32, 64\} \quad (2)$$

5.2 Feature Selection and Analysis

We conducted a feature selection step in all of our three datasets (two for the **ISP Skip Decision** and one for the **ISP Mode Decision**), using Recursive Feature Elimination with Cross-Validation (RFECV) [Pedregosa *et al.*, 2011]. The main idea is to retain only the most relevant features, discarding features that are less informative and by consequence reducing the complexity and improving the generalization of the model.

RFECV starts by training a Decision Tree model using all available features with 5-fold cross-validation. From this, the features are ranked according to their importance, which is determined by the trained model. RFECV removes the least important feature using this ranking and retrains the Decision Tree model with *N-1* features. This process is repeated iteratively until only one feature remains. In the end, the features selected are those that, when used in the model, resulted in the best average F1 score across the folds in the cross-validation.

5.2.1 ISP Skip Decision Feature Analysis

For the ISP Skip Decision datasets, the feature set remained unchanged after RFECV. Specifically, for the dataset containing 4x8 and 8x4 block sizes, all eight features were retained, which include the variance of the entire block, the two variances for the horizontal subpartitions, the two variances for the vertical subpartitions, as well as the width, height, and QP of the block. For the dataset containing the remaining block sizes, all 12 features were kept, including the variance of the entire block, the four variances for the horizontal subpartitions, the four variances for the vertical subpartitions, and the width, height, and QP of the block.

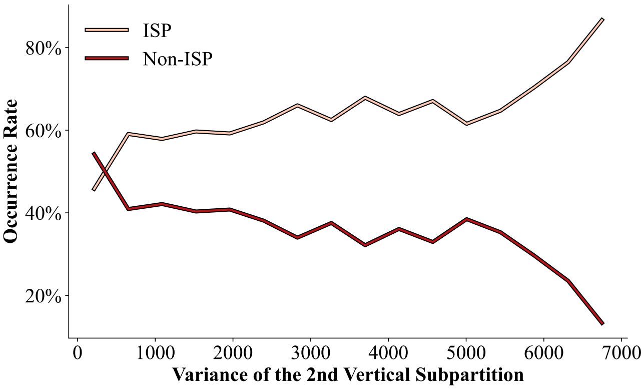


Figure 10. Occurrence rate of Non-ISP and ISP classes according to the variance of the second vertical subpartition.

To analyze the behavior of the image features, we selected the feature with the highest information gain for each dataset: one for the 4x8 and 8x4 block sizes and another for the remaining block sizes. For the 4x8 and 8x4 block sizes, the feature with the highest information gain was the variance of the second vertical subpartition, while for the remaining block sizes, it was the variance of the entire block. To examine each feature, we divided its values into 20 equal intervals based on their minimum and maximum observed values. Then, within each interval, we calculated the occurrence rates of the **Non-ISP** and **ISP** classes.

Figure 10 shows how the variance values of the second vertical subpartition may influence the ISP modes resulting in the best rate-distortion cost. It can be observed that only for very low variance values in the second vertical subpartition, the **Non-ISP** class achieves a higher occurrence rate compared to the **ISP** class. However, as the variance increases, the occurrence rate of the **ISP** class also rises, peaking at over 80% for very high variance values. This suggests that high variance in the second vertical subpartition indicates scenarios where ISP modes are better suited, as they are more effective at modeling the texture complexity of the block in such cases.

Figure 11 shows how the variance values of the entire block may influence the ISP modes achieving the best rate-distortion cost. Similar to the variance of the second vertical subpartition, the **Non-ISP** class exhibits a higher occurrence rate only for very small variance values across the entire block. As the variance increases, the occurrence rate of the **ISP** class also rises, reaching nearly 100% for very high variance values. This analysis underscores that, by leveraging the variances of the entire block and its subpartitions, a Decision Tree model can learn to predict whether the evaluation of ISP modes is necessary based on the texture complexity indicated by these variances.

5.2.2 ISP Mode Decision Feature Analysis

For the ISP Mode Decision dataset, RFECV reduced the number of features from 48 to 28. The selected features include the *x* and *y* positions of the block. From the RMD step, the retained features are the SAD, SATD, estimated number of bits, and fast rate-distortion costs for the Planar, DC, Angular, and MIP modes (excluding the fast rate-distortion cost for the Angular modes). Additionally, the MRL number and the mode number corresponding to the best Angular mode in the RMD step were included. The mode number in the second

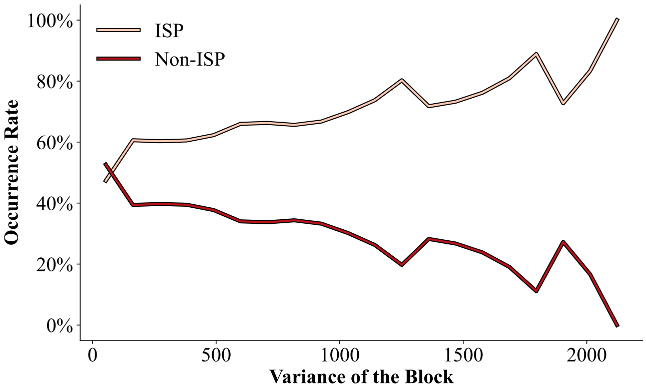


Figure 11. Occurrence rate of Non-ISP and ISP classes according to the variance of the entire block.

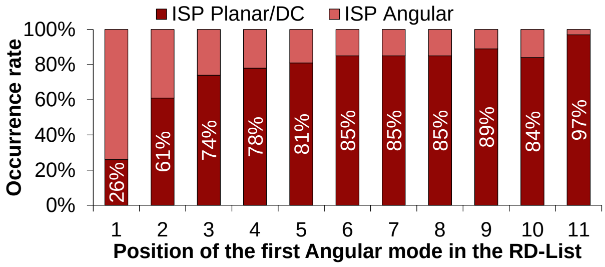


Figure 12. Occurrence rate of ISP Planar/DC and ISP Angular classes according to the position of the first Angular mode in the RD-List.

position was selected from the MPM list. From the RD-List, the retained features are the position of the first occurrence of the Planar, DC, Angular, and MIP modes, along with the number of the Angular modes that appear first. Finally, the complete rate-distortion costs for the Planar, DC, and Angular modes were included from the RDO step for Non-ISP modes.

We selected the two features with the highest information gains to analyze, where the first one is the position of the first Angular mode in the RD-List, and the second one is the intra mode in the second position of the MPM list. The encoding feature analysis was performed by computing the occurrence rates of the ISP Planar/DC and ISP Angular classes based on their respective values.

Figure 12 illustrates the occurrence rate of the ISP Planar/DC and ISP Angular classes according to the position of the first Angular mode in the RD-List. One can notice that the first Angular mode occurs from the first to the eleventh position in the RD-List. When the first Angular mode occurs in the first position of the RD-List, the ISP Planar/DC exhibits an occurrence rate of 26%, indicating that in 74% of the cases, the ISP Angular class results in the best rate-distortion cost. However, as the position of the first Angular mode in the RD-List increases, the ISP Planar/DC class occurrence rate also increases. For instance, when the first Angular mode in the RD-List occurs from the second to the eleventh position, the ISP Planar/DC class consistently achieves a higher occurrence rate, peaking at 97% when the first Angular mode occurs in the eleventh position of the RD-List. In other words, as the position of the first Angular mode in the RD-List increases, it also increases the cases where VTM can avoid the RDO evaluation of the ISP Angular modes.

In Figure 13, the occurrence rate of the ISP Planar/DC and ISP Angular classes is shown according to the intra mode present in the second position of the MPM List. Since the second position of the MPM list can contain any intra mode among DC and Angular, we grouped the values of this feature into two categories: DC, with cases where the second position of the MPM list has the DC mode, and Angular, with cases where the second position of the MPM list has one of the Angular modes (modes 2 to 66 in Figure 1).

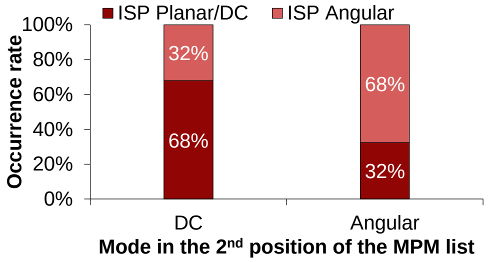


Figure 13. Occurrence rate of ISP Planar/DC and ISP Angular classes according to the intra mode in the second position on the MPM list.

From Figure 13, one can see that when the second position of the MPM list has the DC mode, the ISP Planar/DC class achieves a 68% occurrence rate. This means that in only 32% of the cases, the ISP Angular class achieves the best rate-distortion cost. In contrast, when the second position of the MPM list has one of the Angular modes, the opposite occurs, and the ISP Angular class achieves a higher occurrence rate of 68%. Therefore, the analysis of this feature reveals that when the DC mode is the second in the MPM list, the ISP Planar/DC class has a higher chance of containing the best rate-distortion cost. As a result, when the DC mode is the second in the MPM list, VTM can avoid the RDO evaluation of the ISP Angular modes in 68% of the cases.

In summary, the analysis of these two features reveals the importance of encoding features in predicting when the RDO evaluation of the ISP Angular modes can be avoided by VTM. For instance, Figure 12 shows that when the first Angular mode in the RD-List occurs from the second to the eleventh position, VTM can avoid the RDO evaluation of the ISP Angular modes most of the time, given the higher occurrence rate of the ISP Planar/DC class. Specifically, when the first Angular mode occurs in the ninth, tenth, or eleventh position, VTM can avoid the RDO evaluation of the ISP Angular modes in 89%, 84%, and 97% of the cases, respectively. Similarly, Figure 13 demonstrates that when the DC mode is in the second position of the MPM list, VTM can avoid the RDO evaluation of the ISP Angular modes in 62% of the cases.

5.3 Decision Tree Training

Using the Scikit-learn library [Pedregosa *et al.*, 2011], the datasets were split into 75% for training and validation, reserving the remaining 25% for testing. This division ensured that the models never saw 25% of the data throughout the training and validation stages. We performed training and validation in two steps: a Random Search and a Grid Search. The Random Search [Bergstra and Bengio, 2012] step involved evaluating 1,000 random combinations across a wide search space of the hyperparameters *criterion*, *min samples split*, *min samples leaf*, *max depth*, *max leaf nodes*, and *max features*. Each combination was assessed using a 5-fold cross-validation approach over 75% of the data reserved for training and validation. Subsequently, we used the Random Search

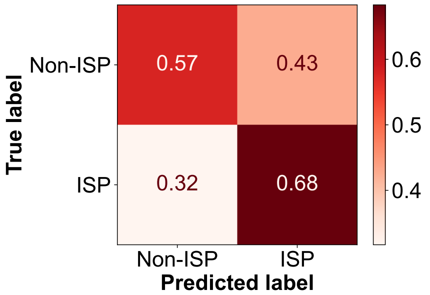


Figure 14. Confusion matrix for the final decision-tree model classifying Non-ISP and ISP classes for 4x8 and 8x4 block sizes in the test set.

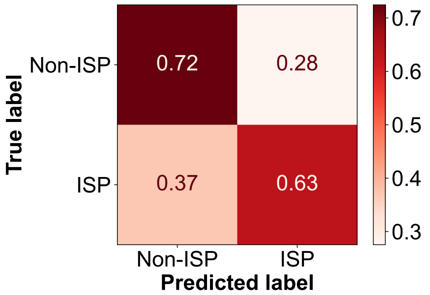


Figure 15. Confusion matrix for the final decision-tree model classifying Non-ISP and ISP classes for remaining block sizes in the test set.

results to compute the Pearson Correlation between each hyperparameter and the F1-score.

The *max leaf nodes* and *max depth* were the two hyperparameters most strongly correlated with an increased F1-score for the datasets of the **ISP Skip Decision** solution, whereas *max leaf nodes* and *max features* showed the highest correlation with an increased F1-score for the dataset of the **ISP Mode Decision** solution. The two hyperparameters selected for each dataset were further refined in the Grid Search step, while the remaining hyperparameters were set to their default values or the best values found during the Random Search. The Grid Search step also evaluated combinations through a 5-fold cross-validation over the same 75% of the data used for training and validation. The final models were derived from the hyperparameter combination that achieved the best F1-score in the Grid Search.

5.3.1 ISP Skip Decision Training Results

Figure 14 shows the confusion matrix for the final Decision Tree model classifying the Non-ISP and ISP classes for the 4x8 and 8x4 block sizes, evaluated on the 25% of the data reserved for testing, which was not used during the Random Search and Grid Search steps. The main diagonal represents the correct predictions of the model, while the entries above and below indicate incorrect predictions. The final model for the 4x8 and 8x4 block sizes achieved accuracies of 57% and 68% for the Non-ISP and ISP classes, respectively.

In Figure 15, the confusion matrix for the final Decision Tree model that classifies between the Non-ISP and ISP classes for the remaining blocks is presented, also evaluated under the test set. One can notice that the final model for the remaining blocks achieves accuracies of 72% and 63% for the Non-ISP and ISP classes, respectively.

Since the Decision Tree models are applied in the context of solutions for reducing encoding time in video coding, the incorrect predictions of the models can be categorized into two types: **time errors** and **coding efficiency errors**. For the **ISP Skip Decision** solution, a time error occurs when the model predicts the **ISP** class, but the sample actually belongs to the **Non-ISP** class. This results in a time error because the solution decides that the RDO process for the ISP modes is necessary, even though they could have been skipped. Conversely, a coding efficiency error occurs when the model predicts the **Non-ISP** class, but the sample actually belongs to the **ISP** class. In this case, the solution excludes the ISP modes from the RD-List, reducing encoding time but possibly introducing a loss in coding efficiency.

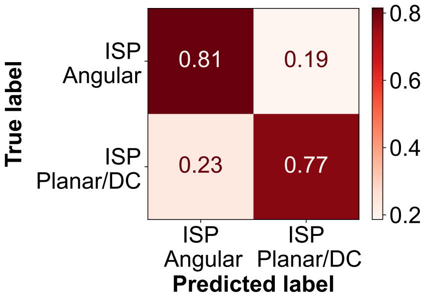


Figure 16. Confusion matrix for the final decision-tree model classifying ISP Planar/DC and ISP Angular classes in the test set.

By analyzing the accuracy results for the model designed for the 4x8 and 8x4 block sizes in Figure 14, we observe that time errors occur in 43% of cases while coding efficiency errors occur in 32% of cases. In contrast, the model designed for the remaining block sizes exhibits an opposite behavior, with time errors occurring in only 28% of cases while coding efficiency errors are observed in 37% of cases.

5.3.2 ISP Mode Decision Training Results

Figure 16 presents the confusion matrix obtained by the final Decision Tree model trained to predict between the ISP Planar/DC and ISP Angular classes when evaluated under the 25% of the data reserved for testing purposes. The final model obtained accuracies of 77% and 81% for the ISP Planar/DC and ISP Angular classes, respectively.

In the context of the **ISP Mode Decision** solution, a **time error** happens when the model misclassifies an example of the ISP Planar/DC class in the ISP Angular class. There is no coding efficiency loss when that happens because the solution will not remove the ISP candidates associated with the Planar/DC modes from the RD-List. However, a **time error** occurs since the RDO evaluation for the ISP modes could be performed exclusively for the ISP candidates associated with the Planar/DC modes. On the other hand, a **coding efficiency error** happens when the model misclassifies an example belonging to the ISP Angular class in the ISP Planar/DC class. Then, the solution removes the ISP candidates associated with the Angular modes, reducing the encoding time but providing a loss of coding efficiency. By looking at Figure 16, **time errors** occur 23% of the time, while **coding efficiency errors** occur 19% of the time.

6 Experimental Results

To evaluate the performance of MLISP, we followed the Common Test Conditions (CTC) [Bossen *et al.*, 2020] of

Table 2. Time saving and coding efficiency results for All Intra configuration. Anchor: VTM 18.0 with ISP enabled.

		ISP Skip Decision		**ISP Mode Decision**		**MLISP Scheme**		**ISP Disabled**	
Class	**Video**	**Time Saving**	**BDBR**	**Time Saving**	**BDBR**	**Time Saving**	**BDBR**	**Time Saving**	**BDBR**
A1	Tango2	9.51%	0.09%	6.98%	0.08%	10.86%	0.10%	14.47%	0.10%
	FoodMarket4	7.85%	0.05%	6.41%	0.05%	9.84%	0.09%	13.00%	0.06%
	Campfire	10.60%	0.03%	8.61%	0.05%	12.05%	0.08%	15.02%	0.13%
A2	CatRobot	8.43%	0.16%	6.71%	0.15%	11.09%	0.23%	15.03%	0.34%
	DaylightRoad2	10.49%	0.23%	7.82%	0.12%	12.65%	0.26%	17.17%	0.37%
	ParkRunning3	6.49%	0.03%	5.89%	0.04%	8.84%	0.05%	11.25%	0.05%
B	MarketPlace	10.36%	0.08%	10.21%	0.09%	13.19%	0.11%	15.98%	0.11%
	RitualDance	7.92%	0.17%	6.98%	0.21%	10.20%	0.25%	14.81%	0.30%
	Cactus	9.19%	0.21%	8.15%	0.25%	12.28%	0.34%	16.89%	0.50%
	BasketballDrive	9.48%	0.38%	6.38%	0.20%	11.33%	0.46%	16.58%	0.71%
	BQTerrace	7.39%	0.22%	4.35%	0.12%	9.86%	0.28%	16.23%	0.51%
C	RaceHorsesC	9.16%	0.22%	8.89%	0.20%	11.96%	0.29%	16.51%	0.40%
	BQMall	8.30%	0.37%	7.10%	0.36%	9.29%	0.59%	15.49%	0.97%
	PartyScene	8.48%	0.27%	7.73%	0.20%	11.84%	0.36%	18.38%	0.53%
	BasketballDrill	8.81%	0.64%	6.50%	0.37%	11.04%	0.67%	18.29%	1.01%
D	RaceHorses	9.09%	0.13%	9.20%	0.13%	11.20%	0.25%	16.43%	0.39%
	BQSquare	6.13%	0.21%	6.49%	0.22%	10.39%	0.35%	17.61%	0.67%
	BlowingBubbles	8.92%	0.35%	7.24%	0.27%	12.14%	0.49%	17.96%	0.65%
	BasketballPass	7.85%	0.34%	6.27%	0.27%	10.05%	0.53%	15.45%	0.73%
E	FourPeople	7.43%	0.30%	6.29%	0.31%	10.13%	0.48%	15.85%	0.71%
	Johnny	7.79%	0.26%	4.19%	0.30%	10.69%	0.41%	14.67%	0.70%
	KristenAndSara	7.96%	0.18%	5.76%	0.22%	10.34%	0.36%	16.00%	0.86%
	Average	**8.53%**	**0.22%**	**7.01%**	**0.19%**	**10.97%**	**0.32%**	**15.87%**	**0.49%**

VVC under the **All Intra** configuration, encoding 22 video sequences with QP values of 22, 27, 32, and 37. The evaluation was conducted both with the anchor VTM 18.0 and with the modified VTM 18.0 implementing the proposed MLISP scheme. The videos were encoded sequentially on a dedicated server with an Intel® Core™ i7-8700K processor and 16GB of RAM. **None** of the videos from the VVC CTC were used to train the Decision Trees for MLISP. Consequently, the proposed scheme was evaluated using videos that the models had never seen.

To assess the performance of MLISP, we calculate two metrics: time saving (TS), which compares the encoding time of the anchor and the MLISP scheme, and coding efficiency, measured in terms of BDBR [Bjontegaard, 2001], which evaluates bit-rate variation between two encoders maintaining the same visual quality. We also evaluated each solution's performance to understand its potential for reducing encoding time, allowing us to analyze the contribution of each solution within the complete MLISP scheme.

Table 2 presents the results for the **ISP Skip Decision**, for the **ISP Mode Decision**, for the entire **MLISP** scheme, and for **ISP Disabled** regarding Class, Video, TS, and BDBR. The classes are defined according to the Common Test Conditions (CTC) of VVC [Bossen *et al*., 2020] and indicate the resolution of the videos. Videos in classes A1 and A2 have 4K resolution, class B videos have 1080p resolution, class C videos have 480p resolution, class D videos have 240p resolution, and class E videos have 720p resolution.

6.1 ISP Skip Decision Results

The **ISP Skip Decision** achieves an average time saving of 8.53% with only a 0.22% loss in coding efficiency. The best time-saving result is observed for the *Campfire* video, achieving a time saving of 10.60% with just a 0.23% coding efficiency loss. We believe that this best time-saving result is due to the **ISP Skip Decision** predicting the **Non-ISP** class more frequently, which enhances the time-saving potential.

On the other hand, the worst time-saving result for the **ISP Skip Decision** is seen for the *BQSquare* video, with a time saving of 6.13% and a 0.21% coding efficiency loss. This result is due to the **ISP Skip Decision** predicting the **ISP** class more often, indicating that the evaluation of ISP modes is required more frequently for this video sequence, which reduces the time-saving potential.

6.2 ISP Mode Decision Results

The **ISP Mode Decision** achieves an average time saving of 7.01% with only a 0.19% loss in coding efficiency. The best time-saving result is observed for the *MarketPlace* video, reaching a time saving of 10.21% with just a 0.09% loss in coding efficiency. This video achieves the best time-saving result because its texture is well-suited for the Planar/DC modes, leading the **ISP Mode Decision** to predict the **ISP Planar/DC** class more frequently, thereby improving the time-saving potential.

In contrast, the worst time-saving result is observed for the *Johnny* video, where the solution achieves a time saving of 4.19% with a 0.30% loss in coding efficiency. The lower time saving for this video may be attributed to its texture being more suitable for the Angular modes, which leads the **ISP Mode Decision** to predict the **ISP Angular** class more frequently, thereby reducing the time-saving potential.

6.3 MLISP Scheme Results

The MLISP combines both solutions to maximize the time-saving potential. On average, MLISP achieved a time saving of 10.97% with only a 0.32% loss in coding efficiency. The best time-saving result was observed for the *MarketPlace* video, with a 13.19% time saving and just a 0.11% loss in coding efficiency. In contrast, the worst time-saving result occurred with the *ParkRunning3* video, where MLISP achieved an 8.84% time saving with a 0.05% loss in coding efficiency.

As shown in Table 2, MLISP achieves time savings exceeding 10% for most videos, demonstrating its ability to reduce encoding time with minimal loss in coding efficiency across video sequences with varying resolutions and texture characteristics. For high-definition videos, such as those in classes A1, A2, and B, MLISP achieves not only substantial time savings but also introduces negligible losses in coding efficiency, remaining below 0.30% for most videos in these classes. For example, for the *FoodMarket4* and *Campfire* videos, MLISP incurred only 0.09% and 0.08% losses in coding efficiency, respectively. These results are particularly significant because high-definition videos require the longest encoding times. In other words, our solution effectively reduces encoding time where it is most critical while introducing only minimal losses in coding efficiency.

To demonstrate that combining both solutions effectively maximizes the time-saving potential and that this improvement is statistically significant, we conducted a statistical significance analysis using the Student's *t*-test. We compared the distributions of time saving and BDBR results between the MLISP scheme and each individual solution, namely ISP Skip Decision and ISP Mode Decision.

When comparing MLISP to the ISP Skip Decision, we obtained *p*-values of 3.76×10^{-13} for time saving and 4.10×10^{-7} for BDBR. In the comparison with the ISP Mode Decision, the *p*-values were 4.64×10^{-14} for time saving and 6.66×10^{-7} for BDBR. All *p*-values are well below the commonly accepted threshold of 0.05, indicating that the observed improvements in time saving, as well as the increases in BDBR, are statistically significant.

It is important to highlight, however, that although the increase in BDBR is statistically significant, its absolute magnitude remains minimal: on average, only $+0.10\%$ compared to the ISP Skip Decision and $+0.13\%$ compared to the ISP Mode Decision. This confirms that the MLISP scheme significantly reduces encoding time while maintaining a negligible impact on coding efficiency.

6.4 Comparison with ISP Disabled

To evaluate whether it is more advantageous to use the proposed MLISP scheme or simply disable the ISP modes, we performed an ablation study in which all encodings were carried out with the ISP modes disabled.

As observed, simply disabling the ISP modes results in an average time saving of 15.87%, at the cost of a 0.49% increase in BDBR. It can also be noted that for the Class A1 sequences, disabling the ISP modes yields very similar BDBR results when compared to the proposed MLISP scheme. We believe this occurs because, for these sequences, the ISP modes are rarely used, and thus disabling them provides a notable time saving with minimal impact on coding efficiency.

However, for 18 out of the 21 test sequences, disabling the ISP modes consistently results in greater losses in coding efficiency. This effect is even more pronounced in sequences where the ISP modes are likely to be used more frequently, such as KristenAndSara, BQMall, BasketballDrill, and BQSquare, where disabling the ISP modes leads to BDBR increases of 0.50%, 0.38%, 0.34%, and 0.32%, respectively, when compared to the proposed MLISP scheme. In particular, while our MLISP solution incurs a maximum BDBR loss of 0.67% for the BasketballDrill sequence, simply disabling the ISP modes results in a maximum BDBR loss of 1.01% for the same sequence.

In summary, although simply disabling the ISP modes offers significant time saving, it comes at the expense of higher losses in coding efficiency. In contrast, the proposed MLISP scheme achieves comparable time saving while introducing considerably smaller degradations in coding efficiency.

6.5 Comparison with Related Works

Table 3 compares the proposed MLISP scheme with related works in terms of time saving (TS), BDBR, and the TS/BDBR ratio, which represents the trade-off between time saving and coding efficiency loss.

Although [Dong *et al.*, 2022] and [Liu *et al.*, 2023] also propose solutions targeting the ISP modes, we do not compare our results with these works because, in both cases, it is not possible to isolate the results related exclusively to ISP modes. Specifically, [Liu *et al.*, 2023] presents a solution that targets not only the ISP modes but also the Planar, DC, Angular, and MIP modes, reporting results only for the complete approach. Similarly, [Dong *et al.*, 2022] proposes a method targeting both ISP modes and Intra Block Copy (IBC), but does not present separate results for ISP modes alone.

We observe in Table 3 that MLISP achieves the best time-saving results compared to the works of [Park *et al.*, 2022], [Saldanha *et al.*, 2021], and [Liu *et al.*, 2021]. To confirm the relevance of these improvements, we conducted statistical significance tests by comparing the time-saving results of MLISP with those of each related work through a paired Student's t-test. The results indicate that the improvements provided by MLISP are statistically significant, with *p*-values of 1.60×10^{-6}, 4.13×10^{-7}, and 1.60×10^{-5} when compared to the works of [Park *et al.*, 2022], [Saldanha *et al.*, 2021], and [Liu *et al.*, 2021], respectively.

This is expected since these works focus solely on identifying when the evaluation of ISP modes can be skipped. In contrast, the MLISP scheme also includes a solution to identify the most promising class of intra modes to be evaluated with ISP among **ISP Planar/DC** or **ISP Angular**. By combining both the **ISP Skip Decision** and **ISP Mode Decision** solutions, **MLISP** improves the time-saving potential by addressing two subproblems within the intra-mode decision process of the ISP.

Although the work in [Park *et al.*, 2020] achieves a higher time saving than MLISP, it also introduces a higher loss in coding efficiency. To further investigate whether this difference in BDBR is statistically significant, we performed

Table 3. Comparison with related works.

Solution	TS	BDBR	TS/BDBR
MLISP (Ours)	10.97%	0.32%	34.28
[Park *et al.*, 2022]	7.20%	**0.08%**	**90.00**
[Saldanha *et al.*, 2021]	8.32%	0.31%	26.84
[Liu *et al.*, 2021]	7.00%	0.09%	77.78
[Park *et al.*, 2020]	**12.11%**	0.43%	28.16

a paired Student's t-test comparing the BDBR distributions of both solutions. The results indicate that there is no statistically significant difference in BDBR between the two approaches. Therefore, although the average values suggest a trade-off advantage for MLISP, from a statistical perspective, both solutions present similar coding efficiency.

Another observation is that the MLISP scheme, along with the solution proposed in [Park *et al.*, 2022], are the only approaches that compute image features not only for the entire block but also for its subpartitions. This aspect is particularly relevant since the ISP tool performs predictions at the subpartition level rather than for the entire block. However, while [Park *et al.*, 2022] computes features for subpartitions, its feature—the mean absolute sum of transform coefficients—is considerably more complex than our feature (the block variance). The feature in [Park *et al.*, 2022] requires the summation of all transformed coefficients within the block. In contrast, MLISP achieves superior time-saving results while employing simpler image features, such as block variance, and maintaining competitive BDBR and TS/BDBR metrics.

7 Conclusion

This paper presented MLISP: a Machine-Learning-Based ISP Scheme Decision for VVC encoders. We conducted an indepth analysis of the ISP tool in VVC, revealing that it often does not achieve the best rate-distortion cost. Furthermore, when the ISP tool achieves the best cost, the Planar and DC modes are frequently used, despite the high number of Angular modes being evaluated. Based on these findings, we proposed two complementary solutions: the ISP Skip Decision, which employs Decision Trees trained on image features to predict when the evaluation of ISP modes can be skipped, and the ISP Mode Decision, which leverages a Decision Tree trained on encoding features to determine the most suitable class of intra modes to evaluate with the ISP tool, selecting between Planar/DC and Angular classes. The experimental results demonstrate the effectiveness of both individual solutions and the complete MLISP scheme, which combines these two approaches to achieve the maximum potential in reducing encoding time while maintaining minimal losses in coding efficiency. Compared with related works, the MLISP scheme achieves competitive results in the trade-off between encoding time reduction and coding efficiency loss while using simple yet effective image features alongside encoding features in one of its solutions.

Declarations

Acknowledgements

The authors thank FAPERGS, CNPq and CAPES (Finance Code 001) Brazilian research support agencies that financed this investigation.

Authors' Contributions

Larissa Araújo and Adson Duarte contributed to the conceptualization, methodology, data curation, and research, and wrote the original draft of the manuscript. Daniel Palomino supervised the project, contributed to the conceptualization, and reviewed and edited the manuscript. Bruno Zatt and Guilherme Correa contributed to the methodology, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Data can be made available upon request.

References

Araújo, L., Duarte, A., Zatt, B., Correa, G., and Palomino, D. (2024). Fast isp mode decision for the versatile video coding intra prediction using machine learning. In *Proceedings of the 30th Brazilian Symposium on Multimedia and the Web*, pages 162–170, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/webmedia.2024.241692.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2). Available at:`https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf`.

Bjontegaard, G. (2001). Calculation of average psnr differences between rd-curves. Available at: `https://www.itu.int/wftp3/av-arch/video-site/0104_Aus/VCEG-M33.doc`.

Bossen, F., Boyce, J., Sühring, K., Li, X., and Seregin, V. (2020). Vtm common test conditions and software reference configurations for sdr video. Available at:`https://jvet-experts.org/doc_end_user/current_document.php?id=10545`, .

Bossen, F., Suehring, K., and Li, X. (2018). Vtm reference software for vvc. Available at:`https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM`.

Bross, B., Wang, Y.-K., Ye, Y., Liu, S., Chen, J., Sullivan, G. J., and Ohm, J.-R. (2021). Overview of the versatile video coding (vvc) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10):3736–3764. DOI: 10.1109/TCSVT.2021.3101953.

Ceci, L. (2023). Live streaming - statistics & facts. Available at:`https://www.statista.com/topics/8906/live-streaming/#topicOverview`.

Chang, Y.-J., Jhu, H.-J., Jiang, H.-Y., Zhao, L., Zhao, X., Li, X., Liu, S., Bross, B., Keydel, P., Schwarz, H., Marpe, D., and Wiegand, T. (2019). Multiple reference line coding for most probable modes in intra prediction. In *2019 Data Compression Conference (DCC)*, pages 559–559, Snowbird, UT, USA. IEEE. DOI: 10.1109/DCC.2019.00071.

Chen, Y., Yu, L., Wang, H., Li, T., and Wang, S. (2020). A novel fast intra mode decision for versatile video coding. *Journal of Visual Communication and Image Representation*, 71:102849. DOI: 10.1016/j.jvcir.2020.102849.

De-Luxán-Hernández, S., George, V., Ma, J., Nguyen, T., Schwarz, H., Marpe, D., and Wiegand, T. (2019). An intra subpartition coding mode for vvc. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1203–1207, Taipei, Taiwan. IEEE. DOI: 10.1109/ICIP.2019.8803777.

Dong, X., Shen, L., Yu, M., and Yang, H. (2022). Fast intra mode decision algorithm for versatile video coding. *IEEE Transactions on Multimedia*, 24:400–414. DOI: 10.1109/TMM.2021.3052348.

Duarte, A., Gonçalves, P., Agostini, L., Zatt, B., Correa, G., Porto, M., and Palomino, D. (2022). Fast affine motion estimation for vvc using machine-learning-based early search termination. In *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. DOI: 10.1109/ISCAS48785.2022.9937973.

Duarte, A., Zatt, B., Correa, G., and Palomino, D. (2023). Fast intra mode decision using machine learning for the versatile video coding standard. In *2023 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, Monterey, CA, USA. IEEE. DOI: 10.1109/ISCAS46773.2023.10181769.

Huang, Y.-W., Hsu, C.-W., Chen, C.-Y., Chuang, T.-D., Hsiang, S.-T., Chen, C.-C., Chiang, M.-S., Lai, C.-Y., Tsai, C.-M., Su, Y.-C., Lin, Z.-Y., Hsiao, Y.-L., Chubach, O., Lin, Y.-C., and Lei, S.-M. (2020). A vvc proposal with quaternary tree plus binary-ternary tree coding block structure and advanced coding techniques. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(5):1311–1325. DOI: 10.1109/TCSVT.2019.2945048.

ITU (2023). Subjective video quality assessment methods for multimedia applications. Available at:`https://www.itu.int/rec/T-REC-P.910-202310-I/en`.

Liu, Z., Dong, M., Guan, X., Zhang, M., and Wang, R. (2021). Fast isp coding mode optimization algorithm based on cu texture complexity for vvc. *EURASIP Journal on Image and Video Processing*, 2021. DOI: 10.1186/s13640-021-00564-4.

Liu, Z., Li, T., Chen, Y., Wei, K., Xu, M., and Qi, H. (2023). Deep multi-task learning based fast intra-mode decision for versatile video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(10):6101–6116. DOI: 10.1109/TCSVT.2023.3262733.

Mercat, A., Mäkinen, A., Sainio, J., Lemmetti, A., Viitanen, M., and Vanne, J. (2021). Comparative rate-distortion-complexity analysis of vvc and hevc video codecs. *IEEE Access*, 9:67813–67828. DOI: 10.1109/ACCESS.2021.3077116.

Park, J., Kim, B., and Jeon, B. (2020). Fast VVC intra prediction mode decision based on block shapes. In *Applications of Digital Image Processing XLIII*, volume 11510, page 115102H, Basel, Switzerland. SPIE. DOI: 10.1117/12.2567919.

Park, J., Kim, B., Lee, J., and Jeon, B. (2022). Machine learning-based early skip decision for intra subpartition prediction in vvc. *IEEE Access*, 10:111052–111065. DOI: 10.1109/ACCESS.2022.3215163.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Édouard Duchesnay (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830. Available at:`https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?source=post_page`.

Pfaff, J., Filippov, A., Liu, S., Zhao, X., Chen, J., De-Luxán-Hernández, S., Wiegand, T., Rufitskiy, V., Ramasubramonian, A. K., and Van der Auwera, G. (2021). Intra prediction and mode coding in vvc. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10):3834–3847. DOI: 10.1109/TCSVT.2021.3072430.

Saldanha, M., Sanchez, G., Marcon, C., and Agostini, L. (2021). Learning-based complexity reduction scheme for vvc intra-frame prediction. In *2021 International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–5, Munich, Germany. IEEE. DOI: 10.1109/VCIP53242.2021.9675394.

Schäfer, M., Stallenberger, B., Pfaff, J., Helle, P., Schwarz, H., Marpe, D., and Wiegand, T. (2019). An affine-linear intra prediction with complexity constraints. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1089–1093, Taipei, Taiwan. IEEE. DOI: 10.1109/ICIP.2019.8803724.

Siqueira, I., Correa, G., and Grellert, M. (2020). Rate-distortion and complexity comparison of hevc and vvc video encoders. In *2020 IEEE 11th Latin American Symposium on Circuits & Systems (LASCAS)*, pages 1–4. DOI: 10.1109/LASCAS45839.2020.9069036.

Sullivan, G. and Wiegand, T. (1998). Rate-distortion optimization for video compression. *IEEE Signal Processing Magazine*, 15(6):74–90. DOI: 10.1109/79.733497.

Yang, H., Shen, L., Dong, X., Ding, Q., An, P., and Jiang, G. (2020). Low-complexity ctu partition structure decision and fast intra mode decision for versatile video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(6):1668–1682. DOI: 10.1109/TCSVT.2019.2904198.

Zhang, Q., Wang, Y., Huang, L., and Jiang, B. (2020). Fast cu partition and intra mode decision method for h.266/vvc. *IEEE Access*, 8:117539–117550. DOI: 10.1109/ACCESS.2020.3004580.

Zhao, L., Zhang, L., Ma, S., and Zhao, D. (2011). Fast mode decision algorithm for intra prediction in hevc. In *2011 Visual Communications and Image Processing (VCIP)*, pages 1–4, Tainan, Taiwan. IEEE. DOI: 10.1109/VCIP.2011.6115979.