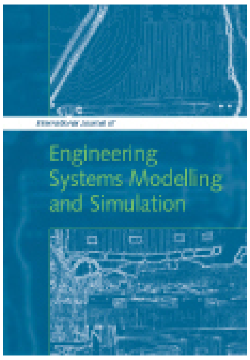


International Journal of Engineering Systems Modelling and Simulation

ISSN online: 1755-9766 - ISSN print: 1755-9758
https://www.inderscience.com/ijesms

Deep learning-based video coding optimisation of H.265

C. Karthikeyan, Tammineedi Venkata Satya Vivek, S. Lakshmi Narayanan, S. Markkandan, D. Vijendra Babu, Shilpa Laddha

DOI: 10.1504/IJESMS.2021.10043077

Article History:

Received:	31 July 2021
Accepted:	02 September 2021
Published online:	03 December 2022

Deep learning-based video coding optimisation of H.265

C. Karthikeyan*

Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation (Deemed to be University),
Vaddeswaram, Guntur, Andhra Pradesh, India
Email: ckarthik2k@rediffmail.com
*Corresponding author

Tammineedi Venkata Satya Vivek

International School of Technology and Sciences (For Women),
Rajanagaram, Rajahmundry, Andhra Pradesh-533294, India
Email: tvsvivek1990@gmail.com

S. Lakshmi Narayanan

Department of ECE,
Gojan School of Business and Technology,
Chennai, India
Email: s.lakshminarayanan@gojaneducation.com

S. Markkandan

Department of ECE,
SRM TRP Engineering College,
Tamil Nadu, India
Email: markkandan.s@trp.srmtrichy.edu.in

D. Vijendra Babu

Department of Electronics & Communication Engineering,
Aarupadai Veedu Institute of Technology,
Vinayaka Mission's Research Foundation,
Paiyanoor 603-104, Tamil Nadu, India
Email: drdvijendrababu@gmail.com

Shilpa Laddha

Department of Information Technology,
Government College of Engineering,
Aurangabad, Maharashtra, India
Email: kabrageca@gmail.com

Abstract: Today's multi-media applications need high video quality with low bitrates. However, it is restricted in its capacity to provide higher quality than earlier coding methods. Deep learning (DL) approaches for video coding have shown compression capacities equal to or better than traditional methods, including high-efficiency video coding (HEVC) methods. The trade-off between compression efficiency and encoding/decoding complexity, optimisation for perceptual nature of semantic dependability, specialisation, and universality, the federalised layout of various deep toolkits, etc. remains unclear. HEVC encoding is more efficient than previous standards. Improved efficiency is driven by intra image prediction, which incorporates more prior directions (35 modes) than previous standards. Its high efficiency comes from balancing encoder complexity and dependability. This article presents DL, which uses a convolutional neural network to predict the best model with the least rate-distortion (RD) and further promotes study into deep learning video coding (DLVC).

Keywords: deep learning video coding; DLVC; high-efficiency video coding; HEVC/H.264; rate-distortion; rate-distortion optimisation; RDO.

Reference to this paper should be made as follows: Karthikeyan, C., Vivek, T.V.S., Narayanan, S.L., Markkandan, S., Babu, D.V. and Laddha, S. (2023) 'Deep learning-based video coding optimisation of H.265', *Int. J. Engineering Systems Modelling and Simulation*, Vol. 14, No. 1, pp.52–57.

Biographical notes: C. Karthikeyan has been awarded a PhD in Computer Science and Engineering from the Jawaharlal Nehru Technological University (JNTU) Hyderabad. He has more than 20 years of teaching experience in computer science and engineering background in India and abroad. He has published more than 25 papers in SCI/SCOPUS/WoS journals and book chapters. He is also the author of the two books titled, *Hands on Data Science*, *Doorstep Python* and five book chapters. He has filed three patents and got granted one patent. He is a reviewer in Springer and Scopus indexed journals. He organised various seminars, workshops, symposiums, and conferences in India and abroad. He has a professional membership in IEEE, IACSIT, ACEEE, ISTE, SDIWC and IAENG Associations.

Tammineedi Venkata Satya Vivek received his Bachelor's in Computer Science and Engineering from the Jawaharlal Nehru Technological University affiliated college way back in 2012. Thereafter, he did his Masters of Technology (MTech) from the KL University, Vijayawada under the specialisation of Computer Networks and Security. He completed his PhD thesis from the Dr. M.G.R Educational and Research Institute (Deemed to be University), Chennai. Presently, he is working as an Associate Professor under the Department of Computer Science and Engineering from the International School of Technology and Sciences (For Women), Rajahmundry, Andhra Pradesh, India. To his credit, he is having more than 15 research papers from Scopus indexing, UGC care journals and one Indian patent. His research areas include applications of cryptography, data hiding in images and machine learning.

S. Lakshmi Narayanan is currently working as an Associate Professor at the Gojan School of Buisness and Technology, Chennai, Tamil Nādu, India. He has more than 17years of teaching experience in reputed institutes. His interest areas are Dat warehousing, signal and image processing and low power VLSI. He has 20 international / national research papers to his credit.

S. Markkandan is currently working as an Assistant Professor at the SRM TRP Engineering College. He has completed his PhD in ECE from the Anna University. His areas of interests are MIMO, communication and machine learning.

D. Vijendra Babu obtained is BE from the University of Madras, India, MTech from the SASTRA, Tanjore, India and PhD from the Jawaharlal Nehru Technological University, Hyderabad, India. He is currently designated as the Vice Principal and Professor in the Department of Electronics and Communication Engineering at Aarupadai Veedu Institute of Technology, Vinayaka Mission's Research Foundation, Deemed to be University. He has 22 years of experience in the field of education, research and administration at various levels. He has served as HoD for 13 years. He has obtained grant for three Australian patents, one Indian design patent and published five Indian patents. He has published 90+ articles in referred international/national journals and conferences and also a reviewer in various leading journals/conferences.

Shilpa Laddha is Assistant Professor in the Department of Information Technology at Government College of Engineering, Aurangabad (Maharashtra). She has a doctorate (PhD) in Computer Science and Engineering. Her area of interest includes artificial intelligence, machine learning, information retrieval, semantic web mining and ontology and many more. She has a profound expertise in taking the full depth training of engineering students. She has three patents and two copyrights, to her credit and her many research papers are published in prominent international journals.

1 Introduction

Because of the accessibility of greater network bandwidth, quicker processor speed, innovative capture as well as display systems, applications that have advantage in precise video recording, effective depiction as well as coding, error-free transmitting, and entirely subjective optimisation display is growing over the decades. According to current research, coded video information is rapidly has become the majority of user internet activity, with such a forecasted proportion of 90% before 2019. Contemporaneous video conferences, streaming video via broadband services, with digital television broadcasts are among the foremost widely utilised application. Most today's mobile hand-held gadgets have a video camera that can collect and encode a streaming video in a standardised way. Video clients that can decode as well as playback video are also included in such devices. Every one of the changes needs the advancement of highly effective video coding methods capable of reducing bitrate independent of losing video quality or allowing for an improvement in video quality independent of increasing bitrate. The most recent response to this customer requirement is high-efficiency video coding (HEVC), commonly referred to as H.265.

The initial stage in parametric dependent video CODEC optimisation is to identify the coding parameters which exhibit a substantial influence on important attributes including bandwidth, image, or video resolution, as well as CPU cycles, among others. Once the video content is familiar, an authorised professional of a video CODEC may predict those parameters with certain precision, but to properly determine the data values with minimal subjective error, a rigorous scientific technique is required. After obtaining these values, the main features of the video CODEC mentioned above may be modelled using the important parameters. Such approaches may subsequently be utilised to improve the video codec's efficiency while operating under real-world restrictions, rendering parameter-based characterisation, and modelling realistic.

2 Review of the state of art

On the video coding optimisation techniques, a substantial number of researches are done in the prior and published in academia. Algorithm-dependent (Zhihai et al., 2008; Jaemoon et al., 2011; Markkandan et al., 2021a), and parametric-based (Vanam et al., 2009; Bharath Kumar et al., 2021) optimisation techniques seem to be the types of optimisation algorithms. The study focuses on parameter-based optimisation methods. As a result, the review would be restricted to them. Algorithmic optimisation techniques give crucial information on how to build a video CODEC that performs optimally, particularly when various operational restrictions are present. To overcome the issue, they offered a comprehensive analysis of the relevance of multiple objective optimisation architecture in Al-Abri (2010) and Jana et al. (2021), as well as the technique given in Al-Abri et al. (2009) and Sangeetha et al. (2021). The proposed methodology was designed to create a combined complexity-memory-rate-distortion (CM-RD) optimisation architecture for an H.264 video coding, but it may be expanded to include a set of requirements and utilised in either form of video CODEC.

An encoder as well as decoder, the architecture considers the optimisation of various goals. The functional goal for rate distortion and CPU was created using SPSS category regression. Those functionalities were later put into a genetic algorithms-dependent optimisation (NSGA-II) in MATLAB, together with the quantifiable values of objective features, to produce a collection of parameters which resulted in the CODEC's best efficiency under a variety of restrictions. HEVC is the ITU-T video coding specialist's community as well as the ISO/IEC Moving Image Specialists Community's newer video coding specification (Sullivan et al., 2012). Weiwei et al. (2014) demonstrated that HEVC offers considerably enhanced compression effectiveness, i.e., a 50% reduction in bit rate in comparison with the highest current video coding methods while maintaining a similar visual quality. The study suggested a technique of HEVC intra-coding especially for rate-distortion optimisation (RDO) that is hardware compatible. The research's findings revealed that suggested RD cost functional reduces space by 85.8% and improves throughput by 1,260% in system architecture, with just a small loss of bitrate as well as PSNR, making it ideal for contemporaneous encoder applications.

Grois et al. (2013) and Markkandan et al. (2021b) provided a comparison study of VP9, H.265, as well as H.264 encoders. Then, H.265/MPEG-HEVC had been demonstrated to continue providing important mean bit-rate investments of 43.4% and 39.5% to VP9 as well as H.264/MPEG-AVC, in both, thus accordingly experimental data was acquired for an entire testing dataset of video sequences utilising identical encoding setups for examined representative encoders. (Bross et al., 2013; Markkandan et al., 2021c) demonstrated that code improvements, involving the significant need for singular instruction multiple-data (SIMD), are able to allow HEVC software decoding for full precision to HD (1,920 × 1,080) in numerous applications. Single-threaded operation including code optimisation was also demonstrated to be insufficient while decoding (3,840 × 2,160) UHD video.

3 Preliminaries

In this work, the coding techniques are explored for realistic image/video, for example, image/video captured by everyday mobile phones or cameras as perceived by humans. Although the approaches are usually relevant, these are created with realistic image/video in mind but do not work with other types (e.g., remote sensing, biomedical).

Almost all realistic image/video is now stored digitally. $x \in D^{m \times n}$ represents the domain of a unique image element in a digitalised greyscale image, wherein the number of rows is represented as m and columns of the image is represented as n. D represents the unique image pixel. $D = \{0, 1, \ldots, 255\}$ is a typical choice, wherein $|D| = 256 = 28$ and therefore the pixel value may be defined via an 8-bit integer; consequently, digitalised an uncompressed greyscale picture contains 8 bits-per-pixel (BPP), whereas compressed bits have significantly fewer.

Conventional lossless coding techniques may produce a compression ratio of 1.6 to 3 for realistic images that are far less than what is required. As a result, lossy coding is used to compress data more efficiently but at the expense of quality. The differential among the actual and reconstructed pictures may be used to calculate the loss, for example, utilising mean-squared-error (MSE) for digitalised greyscale images:

$$MSE = \frac{\left\| x - x_{rec} \right\|^2}{m \times n} \tag{1}$$

As a result, the peak signal-to-noise ratio (PSNR) may be used to assess the quality of the processed picture in comparison to the actual image:

$$PSNR = 10 \times \log_{10} \frac{(\max(D))^2}{MSE} \tag{2}$$

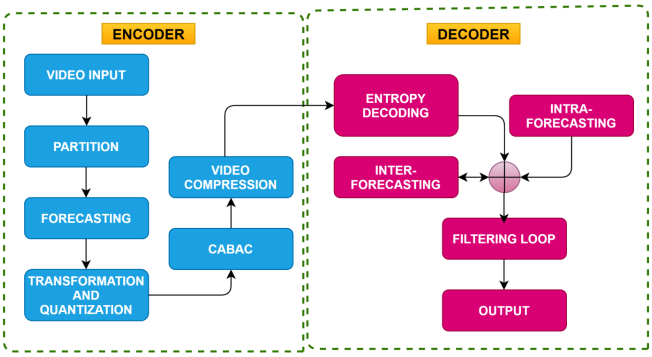
3.1 *High-efficiency video coding*

Figure 1 shows the structural block diagram of H.265. In this video coding layer, then HEVC uses a hybrid coding technique that includes intra-image forecasting, inter-image forecasting, and a two-dimensional transformation. The following is the encoding algorithmic approach: Every image is divided into coding tree modules (CTMs), which are block-shaped areas. The CTM is a coding module that is comparable to the macroblocks utilised in older video coding specifications (AVC/H.264). It's broken down even more into coding module (CM). Two types of predictions are supported by HEVC.

- Intra: each module is calculated based on neighbouring image data available in the present image.
- Inter: makes use of image information from several other images as a basis.

Intra-image forecasting is used to forecast the initial clip in a video sequence since there exist neither frame to contrast (reference frames). Inter-image forecasting coding modes have been used to identify the remaining images in a sequence or among any random moments.

Figure 1 H.265 structural block (see online version for colours)



The motion information including the chosen reference image as well as the motion vector (MV) is utilised for forecasting each block sample are chosen during the encoding processes for inter-image forecasting. The encoder is provided frame sequences, which are divided as a quad-tree design and provided to the encoder individually to determine whether the block has been compressed either of both I-frame and P-frame. An encoder, as well as a decoder utilising MC, creates equivalent inter-image forecasting utilising mode decision with the MV. To calculate the variation among the frames in H.265, the actual block frame image is deducted from the predicted block frame image else reference frame. Error in the image block or remaining image block is the name given to the resulting block. The spatial transform transforms this intra/inter residual block. These residual transformation coefficients are scaled, quantised, and CABAC before being sent with the forecast data.

The encoder duplicates the decoder's working loops, resulting in similar forecasts for future information from both. The coefficients of quantised transform are then rebuilt using inverse scaling as well as modify to replicate the decoded signal's approximations. Both the incorrect image and also the anticipated image blocks are combined and sent through in-loop filters to eliminate quantised artefacts. The completed image will be stored in the decoded image buffer that would aid with inter-image forecasting by allowing the forecast block to be compared to reference images.

4 Proposed architecture for performance modelling

The proposed system for multi-objective optimisation (MOO) is designed to find the best coding factors for an H.265 video if several restrictions are present. The MOO architecture aims to reduce the complexity in CPU use, increase the bit rate, as well as improve the compressed stream video quality. The suggested MOO architecture is completed through the following process.

1. An encoder as well as decoder analysis test has been conducted to identify whether coding settings had a substantial influence on every distortion, rate, and CPU usage objectives/constraints. It is accomplished by calculating the influence of each factor (as it has been modified) on each of the aforementioned factors.
2. Employing an appropriate regression method, the goal function is developed for each objective/constraint depending on the relevant factors.
3. These goal functions are utilised to find optimal parameters in a genetic algorithm (GA) depending on the MOO architecture.

In the first two stages above, defining the relevant coding factors and constructing the appropriate functional goal are more important. These two phases allow for performance modelling and are then utilised to optimise the CODEC. The video is captured by the devices initially encodes it, and sends it across networks to some other device, which decodes then displays the material to a user in a realistic multi-media application domain. Considering that network having bandwidth restrictions and the device includes the encoder installation has computing power limitations, but that possible content users may require low-quality levels, the suggested MOO architecture may be utilised in this circumstance.

A large variety of encoder settings may be specified to regulate the encoder's efficiency, bit rate, and computing power needs, ensuring that the encoder's functionality is optimum under the provided multiple restrictions. Moreover, due to the high number of configurable encoder settings, this necessitates modelling of the encoder's quality, bit rate, and CPU consumption. A conventional technique to optimisation may be utilised if a numerical-functional goal can be generated. The identification of the relevant coding parameters, which is the main emphasis of the study given here, is required to derive function's goal, for instance via mathematical regression. Similar reasoning may be used to the decoder parameter choosing which leads to optimum decoder performance. Let us consider that the data network transmission is flawless in the proposed framework, with bit losses, no delays, or mistakes. As a result, the bitstream of the encoder is delivered to the decoder in the contemporaneous scenario, independent of any loss or modification. The experimental procedure used to find the important coding parameters in the case of both the encoder as well as the decoder is described in the next section.

The functional goal for the three criteria in the case of cactus video has been obtained using the linear regression method's outcomes used as stated previously (1). These functions allow you to go through the importance of every parameter as well as determine how it affects the bitrate, PSNR, and CPU encoder time in a detailed manner. The analysis, in specific, analyses every test film independently and evaluates the influence of every coding parameter considering the provided features of every video's data.

$$f(1)_{Bitrate} = -22.4664 * x(1) - 386.2482 * x(3) + 18{,}066.616$$

$$f(2)_{PSNR} = -0.0039 * x(1) - 0.4404 * x(3) + 48.873$$

$$f(3)_{Enc\_Time} = 3.9537 * x(1) + 2.5501 * x(2) - 36.2174 * x(3) + 545.1239 * x(4) + 2{,}768.025$$

5 Data analysis

5.1 Encoder analysis

The encoder functional goal generated because of the experimental approach allows for discussion of the importance of every coding setting. The models acquired for every video sequence have been listed below, with *f*(1) indicating PSNR, *f*(2) indicating rate, and *f*(3) indicating CPU encoder time. The study of the linear regression expressions derived for the video sequence of cactus reveals that all four factors, notably: intra-period, search-limit, quantisation parameter, speed encoding, have a substantial influence on CPU use.

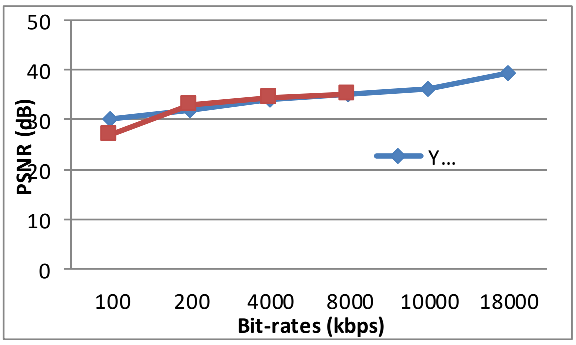
The functional goal produced for all examined video sequences regarding CPU encoding time suggests that the parameter with the most important influence on CPU is the Speedy decisional encoder, according to the results of the CPU Usability. The smaller intra-period outcomes in a longer processing time when selecting the intra-period. The quantisation feature will have the next biggest influence. Search range (SR), as well as intra period (IP), has a minor influence. Encoding time may rise somewhat as the SR expands. These tests have no discernible effect on the video's quality. Encoding time would be somewhat longer if FEN is disabled. Moreover, this has no significant effect on quality.

- PSNR analysis: QP parameter which has the greatest influence on PSNR. The PSNR data in Table 1 show that the two movies with the fewest movements/changes, cactus well as Yacht ride, had the highest correlation coefficients. It is predicted owing to the CODEC's reliability while encoding every individual frame of sequential coding.
- Bit-rate analysis: the QP parameter that has the most influence. As seen in Figure 2, lower quantises lead to higher bitrate and superior visual quality. The significant influence of QP on the H.265 video coding. The search-limit has no effect on PSNR or bit-rate in cacti. It's true because optimal matching would not be identified rapidly, i.e. without scanning the full movie, for recordings with fast-moving elements. A constant component appears in all functional, suggesting that there is a set computing expense for encoding that is irrespective of the coding factors used. Considering the processes that occur are autonomous of the coding settings, this is to be assumed.

Table 1 Correlation coefficient of the encoder

Video	*Yacht ride*	*Cactus*
Enc_Time	0.9840	0.999
PSNR	0.9985	0.9987
Bit-rate	0.9536	0.9547

Figure 2 PSNR versus bit-rate at QP 27, 37, and 45 (see online version for colours)



5.1.1 Decoder analysis

The decoder's study is confined to decoder settings which have a substantial impact on the computing difficulty of the decoder. The PSNR, as well as bit-rate, are defined by the encoder, therefore the decoder parameters being free from effect. The decoder receives the identical quality whereas the bit-rate is fed as an encoder outcome in the proposed architecture. The decoder's computing complexity is assessed utilising a similar technique as the encoder.

Figure 3 PSNR vs. bit-rate (see online version for colours)

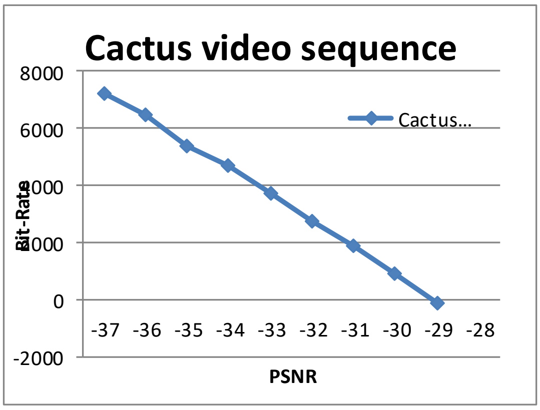


Figure 3 shows the outcomes of the analysis of a Pareto set using MOO. The collection of non-dominated alternatives or Pareto-front for PSNR and Bit-Rate is shown in Figure 3. The Pareto-front depicted is restricted in range, indicating that points seem to be a straight line. When the testing range is expanded, the curve takes on the form of a Pareto curve where this curve helps you to choose the highest performance points plus, therefore, the coding factors which gave you the best functional goal values for video coding.

6 Conclusions

We formally present CNN-based categorisation for forecasting the optimal directed modes for H.265 in this paper. The proposed neural network receives the actual video is provided as an input. The challenge of deciding on the optimal intra-image prediction mode is presented as a classification issue. With a decrease of up to 0.522% over H.265, the Bjntegaad-Delta loss rate becomes negligible. CNN offers the appropriate video encodes. It is most useful in circumstances such as streaming video, which involves the generation and retrieving of digital information via the internet at the same time. As a result, every encoder could use a single classification for the best intra-image prediction mode.

We utilised multi-variate regression evaluation specific to define a functional goal for CPU usage, the bitrate, and PSNR of a video CODEC while encoding/decoding a specific movie. The construction of the functional goals is present by using available data about the content as well as motion in the test films. These regression expressions are demonstrated to be capable of modelling the performance of a common H.265 video coding.

References

Al-Abri, F. (2010) *Multi-Objective Optimization of Video Coding and Transcoding, in Computer Science*, Loughborough University.

Al-Abri, F. et al. (2009) 'A novel framework for multi-objective optimization of video CODECs', in *International Conference on CyberWorlds, CW '09*.

Bharath Kumar, R., Suresh, P. and Hemakumar, V.S. (2021) 'Automated anesthesia level controller using fuzzy logic for injectable type', *Journal of Applied Science and Engineering*, Vol. 24, No. 5, pp.749–755.

Bross, B. et al. (2013) 'HEVC real-time decoding', in *SPIE Optical Engineering+ Applications*, International Society for Optics and Photonics.

Grois, D. et al. (2013) 'Performance comparison of H.265/ MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders' in *Picture Coding Symposium (PCS)*.

Jaemoon, K. et al. (2011) 'Power-rate-distortion modeling for energy minimization of portable video encoding devices', in *2011 IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS)*.

Jana, S., Begum, A.R., Selvaganesan, S. and Suresh, P. (2021) 'Deep belief network-based disease detection in pepper leaf for farming sector', *Turkish Journal of Physiotherapy and Rehabilitation*, Vol. 32, No. 2, pp.994–1002.

Markkandan, S., Logeshwaran, R. and Venkateswaran, N. (2021a) 'Analysis of precoder decomposition algorithms for MIMO system design', *IETE Journal of Research*, pp.1–8, https://doi.org/10.1080/03772063.2021.1920848.

Markkandan, S., Narayanan, L., Theivadas, J.R. and Suresh, P. (2021b) 'Edge based interpolation with refinement algorithm using EDGE strength filter for digital camera images', *Turkish Journal of Physiotherapy and Rehabilitation*, Vol. 32, No. 2, pp.981–993.

Markkandan, S., Sharma, A., Singh, S.P. et al. (2021c) 'SVM-based compliance discrepancies detection using remote sensing for organic farms', *Arab Journal of Geosciences*, Vol. 14, Nos. 1334, pp.1–8, https://doi.org/10.1007/s12517-021-07700-4.

Sangeetha, M., Nimala, K., Saveetha, D., Suresh, P. and Sivaperumal, S. (2021) 'Autonomous self-driving vehicle using deep Q-learning', *Turkish Journal of Physiotherapy and Rehabilitation*, Vol. 32, No. 2, pp.1003–1012.

Sullivan, G.J. et al. (2012) 'Overview of the high efficiency video coding (HEVC) standard', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, No. 12, pp.1649–1668.

Vanam, R., Riskin, E.A. and Ladner, R.E. (2009) 'H.264/MPEG-4 AVC encoder parameter selection algorithms for complexity distortion tradeoff', in *Data Compression Conference, DCC '09*.

Weiwei, S. et al. (2014) 'A hardware-friendly method for rate-distortion optimization of HEVC intra coding, in VLSI Design', *2014 International Symposium on Automation and Test (VLSI-DAT)*.

Zhihai, H., Wenye, C. and Xi, C. (2008) 'Energy minimization of portable video communication devices based on power-rate-distortion optimization', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 18, No. 5, pp.596–608.