Applications and Challenges of Deep Learning and Non-Deep Learning Techniques in Video Compression Approaches

Mr.K. Siva Kumar [1] , Dr. P. Bindhu Madhavi [2], Dr.K.Janaki [3]

[1] Assistant Professor, CSE Department, R.V.R & J.C College of Engineering,

Guntur, Andhra Pradesh, India. kommerlasivakumar@gmail.com

[2] Professor & HOD, Department of AI & ML, The Oxford College of Engineering,

Bengaluru, Karnataka. bindhu.prathipati@gmail.com

[3] Associate Professor, Department of CSE-AI, Faculty of Engineering, Jain Deemed to be University,

RamNagar District, Karnataka. karur.janaki@gmail.com

Abstract
A detailed survey, applications and challenges of video encoding-decoding systems is discussed in this paper. A novel architecture has also been set aside for future work in the same direction. The literature reviews span the years 1960 to the present, highlighting the benchmark methods proposed by notable academics in the field of video compression. The timeline used to illustrate the review is divided into three sections. Classical methods, conventional heuristic methods, and current deep learning algorithms are all used for video compression in these categories. The milestone contributions are discussed for each category. The methods are summarized in various tables, along with their benefits and drawbacks. The summary also includes some comments regarding specific approaches. Existing studies' shortcomings are thoroughly described, allowing potential researchers to plot a course for future research. Finally, a closing note is made, as well as future work in the same direction.

Keywords:
Video, Encoding, Decoding, Deep learning.

1. Introduction

A video is a series of bitmap images that are shown in fast succession. This fast-paced presentation creates the illusion of moving elements in a scenario. The importance of digital video may be seen in a wide range of applications. It has applications in communication, medical science, sports, advertising, cosmic imagery, and more, starting with a television show. The television and entertainment market is currently swamped with a plethora of products, including HD and UHD display devices. The introduction of modern broadcasting tools such as YouTube, Netflix, and Prime Videos, among others. These service aids are gaining popularity since they provide video on demand. This approach has added to the commercial diversity of the entertainment sector.

The number of users in this area is continuously increasing day by day. More than one billion unique users use you tube and Netflix each month. The days of telephonic discussion are long gone. Because of recent technological breakthroughs, people all across the world are engaging in long-term video interactions. As a result, the research community has given it more attention. For example, research funding in the United Kingdom has surpassed two billion pounds. As the demand for digital video and its applications grows, so does the need for more efficient storage solutions. The biggest problem so far in this regard has been efficiently storing and delivering digital video data.

Many academics have proposed video compression algorithms that allow for the storage of digital video data with minimum or no quality degradation. Since the 1980s, some video compression techniques have been proposed ([1]-[2]). These methods are based on the idea of presenting and transferring digital video data in a robustly compressed format for faster storage and transmission, which can lead to a variety of benefits such as video interactions and interactive gaming. The MPEG (motion picture expert groups) and H.26.version are the two most common digital coding standards ([3]). The objective behind a brute force technique to video compression is to remove unnecessary pixel data from successive video frames. This redundancy could be of various types. The compression approaches varies depending on the nature of the redundancies and their individual properties. This leads to removal of highly redundant pixels information between those frames and thus reduces the storage involvement.

Manuscript received June 5, 2023
Manuscript revised June 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.6.17

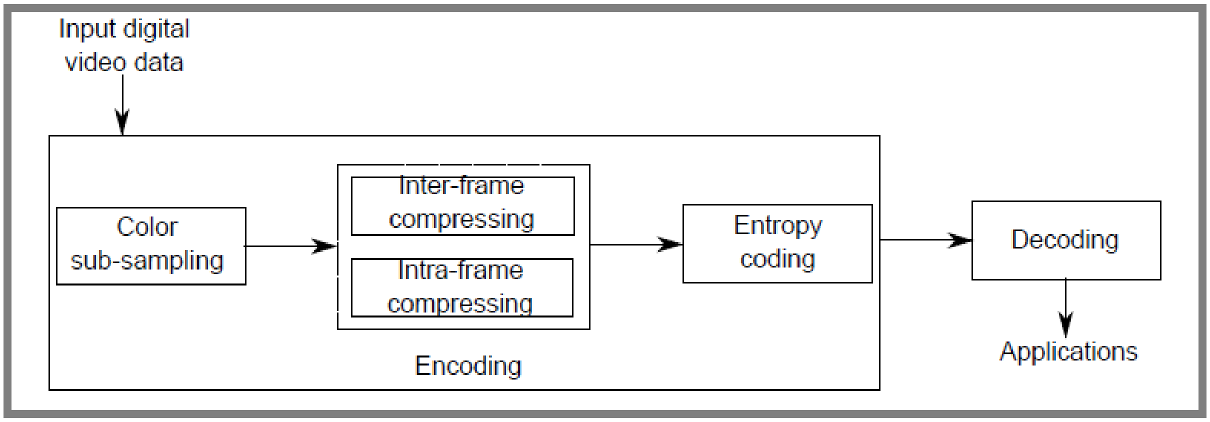

Fig. 1. General block diagram of a video compression Procedure.

In this paper, we explore various video compression techniques which are based on the latest deep learning mechanisms. We begin the literature study with classical approach, followed by individual heuristics and finally an in-depth analysis of deep learning techniques used for efficient video compression. The paper is thus organized as follows. Three sub-sections are presented under the Section 2 (literature review). These are dubbed as classical, mid-era, and the modern age. Section 3 discusses about the pros and cons of these schemes along with the advantageous features of deep learning over the others. A novel Comparative analysis is made in previous sections which are followed by concluding remark at Section 4.

2. Liturature Survey:

Starting from the 80's, the video compression techniques came to existence and their importance became significant. The number of techniques discovered for the purpose during 1980 - 2000 can be termed as the classic era of video compression where generic and straight-forward approaches were adopted for the purpose. This duration is followed by the time line between 2001 - 2010, whereby the researchers and industries presented advanced video compression mechanisms. The time starting from 2011 - till date can be viewed as the era of deep learning. In this era, efficient and robust deep learning mechanisms are introduced for the purpose of video encoding and decoding.

2.1 The Classical Era ([4]-[13]):

The classical works on video compression presented during the early period fall under this category. The actual work related to compression of digital video data is reported starting the 80's. However, the concept of compression of information in the analog mode is reported back in the 60's. For instance, the works reported in [4]–[7] can be taken as good cases in this regard. In [4], the authors have experimented on the audio part of a complete video for the purpose of compression. They have maintained the synchronization between the analog audio and video information even after the compression of audio-only data. Successful compression of data over physical media is observed in this work. In [5], a mixture of different methods is used to achieve video compression in the analog form. Curve fitting and dynamic aperture technique are the major approaches being used in this work. They have successfully experimented on pictures, aerial photographs, and ranger photos. The said work could reduce bandwidth up to the extent of six. In this Noise cancellation not available. So that the Presence of noisy signal may impact the efficiency. In [6], a technique called as TBCF based on time buffering is proposed. It is meant for coarser level of compression. Pulse code modulation with six bits of pulse can be reproduced using the scheme. They used the work for noisy identification and elimination of data with run length coding. In this Delayed time of Computation.so that it is not suitable for live video transmission.

TABLE I: Significant Contribution towards Compressed Video Transmission with Classical Methods

Literature	Year	Approach	Pros	Cons	Remark
Kliger et al. [7]	1957	Signal compression using input coupling in wave tubes.	Good signal gain output (27 dB).	Hardware level implementation.	Further improvement.
Hochman et al. [8]	1967	Polynomial approximation of frames in sequence.	bandwidth minimization (2-6 units)	Partially implementable in practical.	Mostly suitable for B/W samples.
Kutz et al. [9]	1968	Prediction based noise removal and adaptive compression.	Quantitative And qualitative results.	Partially implementable in practical.	Overheads during implementation.
Seitzer et al. [10]	1969	Multiplexing.	Simple but effective multiplexing technique used.	Limited to B/W samples only.	Worst case time complexity for large size videos.
Oliver et al. [11]	1973	2-bit differential PCM.	Adaptive method.	Higher time complexity.	Remarkable contribution with scalable level procedure.
Scheinberg et al. [12]	1984	Color carrier; delta modulation; NTSC.	Real time adaptation.	Dependent on hardware chip size.	Needs specific hardware support?
Prabhu et al. [13]	1985	Adaptive switch-based prediction.	Implemented on natural scene.	Higher time complexity.	Poor response time during processing.

2.2 The era of Generic Heuristics ([14]-[27]):

Researchers began to provide novel heuristics that surpassed traditional works in terms of speed and storage. They concentrated on strategies that might be applied in a generic and scalable manner. With the introduction of modern digital technology for storing and processing images and videos in 1980, genuine attempts on video compression began. Because storage is a difficult requirement for maintaining and transmitting digital video information, encoding and decoding methods are in high demand. Many researchers have contributed useful video compressing techniques. During this time, numerous standard video formats with defined definitions are also launched.

Encoding bandwidth as an early step toward digital video compression was done in [14]. For this goal, mono-bit quantization on tiny space blocks is adapted. Along with motion frame correction, the overall frame registration is taken into account. The imagery collected from a low-altitude aero plane is used as a sample in this study. This work is also a great contribution to the field of remote sensing. Modified CAQ in relation to quantized area bandwidth was studied in [15]. Seventy percent of the mean squared error (MSE) is reduced. The two-phase modification approach starts with the CAQ and then moves on to line-by-line data repetition. The first step does not require hardware; however, the second phase is concerned with implementations at the hardware level. As of now, standard codecs such as JPEG and JPEG-2000 ([16]) as well as BPG are available. The HEVC standard mechanism is used in these schemes ([17]). The encoding and decoding pipelines were manually tuned. Other common formats include h.261, mpeg-4, and h.264/AVC ([18]). This standard can be used to create a solution for high-definition video processing and compression (HD-videos). These techniques, however, are not yet optimal. Table II shows some of the benchmark conventional encoding techniques.

2.3 The era of modern techniques and Deep Learning ([28]-[38]):

Deep learning algorithms are currently conquering the world of information technology. These schemes are extremely resilient and adaptable. They have a significant amount of intelligence processing capability. As a result, deep learning algorithms are commonly employed for video encoding and decoding. Although other traditional approaches use auto-encoding techniques to reduce attribute dimensions, deep learning techniques leverage the same for improved performance on compression challenges.

([28]– [29]) report on some of the most efficient research on video compression utilizing deep learning techniques. These works mostly employ CNN and RNN ideas, with the most recent developments in the field being reported. For this, the generative adversarial search technique is widely utilized. In [30], an interpolation neural network (INN) was effectively employed. The positive aspect of adopting deep learning algorithms is that they adapt inputs quickly, allowing for effective compression while maintaining video quality (HD/UHD). Learning in either a supervised or unsupervised setting can be used to auto-encode.

They may not use wavelet or Fourier transforms, which are common image alteration mechanisms. With highly adaptive learning modules, they are capable of handling the rate of distortion. Most of these approaches are based on the concept of inter-prediction. Table III shows some of the benchmark video encoding techniques that use deep learning algorithms.

TABLE II: The Mid – Era Depicting Definition Specific Heuristics for Video Encding

Literature	Year	Approach	Pros	Cons	Remark
Kim et al. [19]	2000	3-D tree; Rendering.	Less memory use; suitable for 30-60 frame rate; lesser bits; scalable model.	Robust scheme for images, but, not suitable for videos; only suitable for constant traffic.	No need for external bit allocation.
Cheng et al. [20]	2000	Part encoding; Quad tree.	Time efficient; part-encoding.	Not adaptable.	Little overheads during processing.
Sun et al. [21]	2006	Flex SP; DCT.	Compatible with MPEG4.	Always needs bit switching.	May introduce noise.
Tiwari et al. [22]	2008	Short term and long term correlation; simulated annealing.	Dual-frame consideration adds to faster computation.	Fixed window size; device specific.	Need to redefine for various display devices.
Gao et al. [23]	2009	Bit-depth scaling; Advance H.264 encoding.	Higher profiling with less bit rate.	Need SVC software tool.	Needs specific hardware support.
Chen et al. [24]	2009	Bayer-pattern; chroma sampling.	Improved PSNR (1-1.6dB).	Needs intensive preprocessing. (demossaicing)	Although it is faster but pre-processing consumes additional time of computation.
Zhang et al. [25]	2010	Lagrange equation; Motion estimation.	Generic and robust.	Need JM 12.0 tool	Requires specific h/w support.
Nam et al. [26]	2010	DCT; Motion array.	Robust on noise and fluctuations.	Involves more buffer for execution.	Needs large capacity h/w memory.
Peng et al. [27]	2010	H.264 intra-framing; DFT.	Smoother 2-D processing and Compression.	Not scalable due to sample specificity.	Little overheads during processing.

TABLE III: Benchmark Encoding Schemes for the Current Decade

Literature	Year	Approach	Pros	Cons	Remark
Van et al. [31]	2015	M/L hybrid schemes; correlation.	Adaptive; higher gain in transmission rate.	Only for off-line samples.	Not applicable for video conferencing and instant on-line streaming.
Zhang et al. [32]	2015	DCT; motion Array.	Robust on noise and fluctuations.	Involves more buffer for execution.	Requires large size hardware memory.
Duanmu et al. [33]	2016	Rapid-mode computation.	Reduced computational cost.	Iterative execution sometimes leading to long-time looping.	May enter system deadlock at some time.
Chen et al. [34]	2017	CNN; Huffman coding; Quantization.	Time efficient; robust;	Only suitable for basic color schemes.	Not suitable for high end color schemes.
Jiang et al. [35]	2018	M/L techniques; Gradient aspects.	Suits the H.264 standard.	Not applicable for UHD samples.	May be enhanced using optimization schemes.

Zhu et al. [36]	2018	Distortion rate learning; SVG.	Higher rate of accuracy for frame prediction.	Needs huge training samples.	Require more time for training the samples.
Liu et al. [37]	2019	CNN.	Lower resolution video compression.	Not implemented at scalable level.	There is less need for compression for low resolution files.
Gao et al. [38]	2019	Transmission schema based coding rate.	Efficient for video conferencing.	No steps mentioned about offline video storage	Efficient scheme for rapid transmission.

3. Discussion

A thorough examination of the literature over the last five decades reveals that numerous video compression strategies have been presented. These methods have provided a large number of efficient and reliable processes. These programs offer a plethora of benefits. However, more progress is needed in this area, particularly in relation to the constraints listed below.

- Due to the processing of large-scale videos, compression techniques are more complicated.
- Compression, transmission, and streaming errors are all possible.
- The video quality appears to be deteriorating during the decompression procedure.
- Improvised Deep Learning Algorithms still have a way to go in terms of displaying higher-quality metrics.

We are currently experiencing innovations in video capturing, storing, and display technologies every year. Thus, matching the speed of innovations is also a considerable challenge for video codecs. Every user wants to have a great experience at their end. Still, they do not know the efforts video codecs have to put in behind delivering it, as well the complexity of the operation of producing, storing, and providing the data to the user. **The following are the challenges faced by today's video codecs**.

- Faster encoders do not guarantee potential compression efficiency: Most codecs try their best to compress, but they do not promise us a potential compression efficiency. Although few can perform good compression levels, they are slower than older codecs. It may be because of the variety and complexity of data being generated by devices today. The other reason may be the data formats of the data; they are changing very quickly. HFR, HDR, 4K, 6K, 8K, 3D, and 360-degree videos are newly evolved challenging formats.

- Encoder search problem: Finding an efficient encoder for data compression is challenging. There are several hurdles in the path that must cross by the encoder. Currently, ML algorithms are extensively being used to reduce the complexity of the encoder. However, we must admit that ML has its advantages and disadvantages.

- Many software encoders can support lower resolutions: Compression is becoming more difficult because of changes in the resolution of the data. It is becoming tough to find redundancy between data and further compression.

- Further compression is more complex: Obtaining an efficient output depends on the changes made in the ML model and the hardware required to run that model. They are time consuming and costly. Input given to the data is not really in the programmer's hands; it will change in the future, so we need to develop a system that will adapt to the changes and help us do different level compression.

- Deep learning methods are very successful for applications such as image classification. However, the system was found to be very instable when it comes to image restoration. A tiny change occurring on the image results in losing artifacts or important features from the image. It also led to a degradation of the quality of the image. This may occur because of changes in resolution, faulty source equipment, use of inappropriate method for processing the image, etc. This instability makes us think about how much we should rely of these deep learning-based methods.

4. Conclusion

Over the course of five decades, three main eras in the growth of video compression techniques have been studied. As a result, a literature review is presented chronologically. The timeline of evolution is divided into three scenarios: the classic era, the age of heuristic and standard schemes, and the era of deep learning frameworks. Beginning in 1960, the timeline is displayed, and the description continues until today. For the years 1960-1980, 1980-2000, and 2001-till-date, the three categories are assigned. The key concepts behind the most widely utilized schemes are illustrated in chronological sequence. The advantages and disadvantages are also briefly summarized in order to gain a better knowledge of how to plan future research projects. Suitable remarks are made in opposition to the initiatives. It should be mentioned that deep learning techniques have so far outperformed other traditional systems in terms of storage, computation time, and accuracy rates.

5. References

[1] Image and Video Compression for Multimedia Engineering (Fundamentals, Algorithms, and Standards), Yun Q. Shi New Jersey Institute of Technology Newark, NJ, Huifang Sun Mitsubishi Electric Information Technology Center America Advanced Television Laboratory New Providence, NJ ,by CRC Press LLC, ©2000.

[2] A Practical Guide to Video and Audio Compression from Sprockets and Raster's to Macro blocks, Cliff Wootton, Copyright © Elsevier Inc, 2005.

[3] H.264 and MPEG-4 Video Compression Video Coding for Next-generation Multimedia, Iain E. G. Richardson, The Robert Gordon University, Aberdeen, UK, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England,2003.

[4] J. Loper, "A method for time compression of instructional television materials," Journal of the SMPTE, vol. 73, no. 9, pp. 753–755, 1964.

[5] P. E. Drapkin, "Video data compression," IEEE Transactions on Aerospace and Electronic Systems, vol. AES-2, no. 4, pp. 392–400, 1966.

[6] G. P. Richards and W. T. Bisignani, "Redundancy reduction applied to coarse-fine encoded video," Proceedings of the IEEE, vol. 55, no. 10, pp. 1707–1717, 1967.

[7] .J. Kliger and J. Munushian, "Signal compression traveling-wave tube devices," in 1957 International Electron Devices Meeting, 1957, pp. 113–113.

[8] D. Hochman, "Application of redundancy reduction to television bandwidth compression," Proceedings of the IEEE, vol. 55, no. 3, pp. 263–266, 1967.

[9] R. Kutz and J. Sciulli, "The performance of an adaptive image data compression system in the presence of noise," IEEE Transactions on Information Theory, vol. 14, no. 2, pp. 273–279, 1968.

[10] D. Seitzer, "An experimental approach to video bandwidth compression by multiplexing," IEEE Transactions on Communication Technology, vol. 17, no. 5, pp. 564–568, 1969.

[11] M. Oliver, "An adaptive delta modulator with overshoot suppression for video signals," IEEE Transactions on Communications, vol. 21, no. 3, pp. 243–247, 1973.

[12] N. Scheinberg, "A composite ntsc color video bandwidth compressor," IEEE Transactions on Communications, vol. 32, no. 12, pp. 1331–1335, 1984.

[13] K. Prabhu, "A predictor switching scheme for dpcm coding of video signals," IEEE Transactions on Communications, vol. 33, no. 4, pp. 373–379, 1985.

[14] D. Healy, "Digital video bandwidth compression using block truncation coding," IEEE Transactions on Communications, vol. 29, no. 12, pp. 1809–1817, 1981.

[15] J. Arnold and M. Cavenor, "Improvements to the constant area quantization bandwidth compression scheme," IEEE Trans on Communications, vol. 29, no. 12, pp. 1818–1823, 1981.

[16] R. J. Majid Rabbani, "An overview of the jpeg2000 still image compression standard," ELSEVIER Signal Processing: Image Communication, vol. 17, no. 1, pp. 3–48, and 2002.

[17] W. H. G. J. Sullivan, "Overview of the high efficiency video coding (hevc) standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649–1668–48, 2012.

[18] G. B. A. L. T. Wiegand, "Overview of the h.264/avc video coding standard," IEEE Trans on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560–576, 2003.

[19] Beong-Jo Kim, Zixiang Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-d set partitioning in hierarchical trees (3-d spiht)," IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, no. 8, pp. 1374–1387, 2000.

[20] H. Cheng, "Partial encryption of compressed images and videos," IEEE Transactions on Signal Processing, vol. 48, no. 8, pp. 2439–2451, 2000.

[21] Xiaoyan Sun, "Drift-free switching of compressed video bit streams at predictive frames," IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 5, 2006.

[22] M. Tiwari , "Selection of long-term reference frames in dual-frame video coding using simulated annealing," IEEE Signal Processing Letters, vol. 15, pp. 249–252, 2008.

[23] Y. Gao, "H.264/advanced video coding (avc) backward-compatible bit-depth scalable coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, no. 4, pp. 500–510, 2009.

[24] H.Chen, "Compression of Bayer-pattern video sequences using adjusted chroma sub sampling," IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, no. 12, pp. 1891–1896, 2009.

[25] J. Zhang, "Context adaptive Lagrange multiplier (calm) for rate-distortion optimal motion estimation in video coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 6, pp. 820–828, 2010.

[26] H. Nam, "Low complexity content-aware video retargeting for mobile devices," IEEE Transactions on Consumer Electronics, vol. 56, no. 1, pp. 182–189, 2010.

[27] X. Peng, "Directional filtering transform for image/intra-frame compression," IEEE Transactions on Image Processing, vol. 19, no. 11, pp. 2935–2946, 2010.

[28] Alan Jacob, "Deep Learning Approach to Video Compression", 2019 (IBSSC).

[29] Guo Lu1, "DVC: An End-to-end Deep Video Compression Framework", Proceedings of the IEEE,2019.

[30] P. K. C-Y W, "Video compression through image interpolation," in 15th European Conference on Computer Vision, 2018, pp. 8–14.

[31] G. V. W. J. D. C. L. P. Van, "Performance analysis of machine learning for arbitrary downsizing of pre-encoded hevc video," IEEE Transactions on Consumer Electronics, vol. 61, no. 4, pp. 507–5115, 2015.

[32] X. W. H. Y. Z. P. Y. Zhang, "Machine learning-based coding unit depth decisions for flexible complexity allocation in high efficiency video coding," IEEE Transactions on Image Processing, vol. 24, no. 7, pp. 2225–2238, 2015.
[33] Z. M. F. Duanmu and Y. Wang, "Fast mode and partition decision using machine learning for intra-frame coding in hevc screen content coding extension," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 6, no. 4, pp. 517–531, 2016.
[34] Chen, "Deep coder: A deep neural network-based video compression," in 2017 IEEE - VCIP, 2017, pp. 1–4.
[35] S. L. J. R. X. G. F. Jiang, W. Tao and D. Zhao, "An end-to-end compression framework based on convolutional neural networks," IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 10, pp. 3007–3018, 2018.
[36] S. K. X. W. L. Zhu, "Fuzzy svm based coding unit decision in hevc," IEEE Transactions on Broadcasting, 2018.
[37] Li, "Learning a CNN for image compact-resolution," IEEE Trans on Image Processing, vol. 28, no. 3, 2019.
[38] Q. J. C. F. P. H. W. W. W. Gao, "Data-driven rate control for rate-distortion optimization in hevc based on simplified effective initial qp learning," IEEE Transactions on Broadcasting, vol. 65, no. 1, pp. 94–108, 2019.

Mr. K. Siva Kumar
Assistant Professor,
Department of CSE, R.V.R & J.C College of Engineering.
Guntur, Andhra Pradesh.

Qualifications:
Ph.D. in CSE [Pursuing] - Visvesvaraya Technological University, Karnataka.
M.Tech in CSE, JNTUK, Andhra Pradesh.
B. Tech in CSE, JNTUH, Andhra Pradesh.

Dr. P . Bindhu Madhavi
Professor & HOD, Department of AI & ML,
The Oxford College of Engineering, Bengaluru, Karnataka.

Qualifications:
Ph.D. in CSE, SKU, Anantapur, Andhra Pradesh.
M.Tech in CSE, ANU, Andhra Pradesh.
B. Tech in CSE, RGM, JNTUH, Andhra Pradesh.

Dr.K. Janaki
Associate Professor, Department of CSE-AI, Faculty of Engineering, Jain Deemed to be University, RamNagar District, Karnataka.

Qualifications:
Ph.D. in CS - Anna University, Chennai.
M.E in CSE - Anna University, Chennai.
B.E in CSE - Madurai Kamaraj University, Madurai.