Chapter - 2

Literature Review

CHAPTER - 2

LITERATURE REVIEW

2.1 Introduction

The literature review of digital audio watermarking is elaborated in this section using the research articles from 2019 to 2023. The articles were observed and analyzed to elaborate on the recent achievements as well as the developed models in digital audio watermarking. The major contribution of the chapter is depicted as follows,

- The models available in the digital audio watermarking research are analyzed with several standard articles to address the data resources that are available throughout the article.
- The methods utilized in the researches are analyzed comprehensively to elaborate the frameworks, techniques, and so on. The review includes information on the methods that aid in achieving better information for future endeavors.
- The research gaps found in the researches are added as they aids the future researches in the digital audio watermarking research area.

2.2 Review of Literature

This section elaborates the discussion of the existing researches on the digital audio watermarking. Based on the chosen methods, the taxonomy is formed and illustrated in Figure 2.1, which aids in achieving the better understanding of the technical growth.

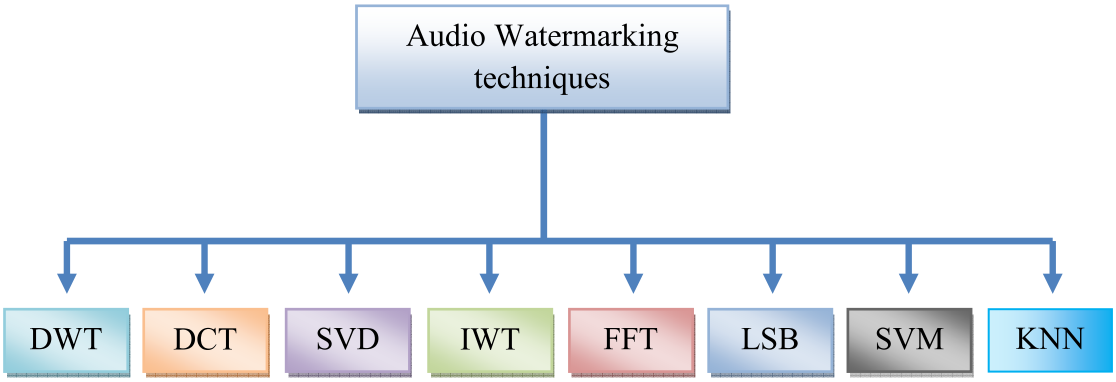

Figure 2.1: Taxonomy of the Digital Audio Watermarking techniques

In 2019, the digital audio watermarking achieved the better way of development through several researches that worked to improve the performance of the watermarking in all possible directions. The Huda Karajeh *et al.,* [15] utilized the hybrid Discrete Wavelet Transform (DWT) and Schur decomposition methods in the research of the digital audio watermarking. The watermarking scheme included the embedding and extraction process with the proposed and the transpose of the proposed method respectively, which showed the robust performance against various attacks such as MP3 compression, echo and so on. Though the proposed method enhanced the balance performance in terms of the subjective and the objective tests, the method could only embed the minimal sized image with the insecure scheme of watermarking that remained the major challenge of the research. Salma Masmoudi *et al*., [16] modeled the Huffman data to extract and embed the watermarking into the audio, where the source of the audio should be in the decompressed form. The model of Huffman data included the big value region as well as the calibration that achieved the data privacy along with the blind retrieval of the embedded watermark. The big value region utilized provided the better resistance with the attacks of the decompression and recompression which improves the performance of the model. The model included only the decompressed source of audio that remain the major challenge as the recompression model was ignored from the source. In [17], Chur-Jen Chen elaborated the optimized DWT with the minimum amplitude scaling. The model included the KKT theorem as well as the minimum length that added advantage over the low-frequency amplitude. The model worked only with the certain audio quality as well as the embedding capacity that remain as the major disadvantage of the model. Mahdi Mosleh *et al*., in [18] that introduced the fuzzy inference system with the Discrete Cosine Transform (DCT) domain. The model introduced fuzzy logic, the Singular value decomposition (SVD) algorithm, and the Fibonacci sequence that enhance the robustness, transparency, and so on. The model failed to work with the neural network that could enhance the efficiency. The SVD was utilized as the audio watermarking technique especially in the Fractional Fourier Transform (FRT) domain and the model was explained in [19], which was proposed by the Khaled M. Abdelwahab. The model enforced better stability that made the model more robust to certain attacks and in addition, the audio signals were embedded with the encrypted images in order to save quality of the signals. The better performance of the segment-based implementation was attained specifically in the correlation-based domain. The utilized method observed the features with linear relationship and failed to look at the complex relations. The model is sensitive to the scale of the features, which made the detection accuracy poor. Gulivindala Suresh *et al* suggested the SVD model in audio watermarking with the Integer Wavelet Transform (IWT) in [20]. The False positive problem (FPP) is addressed with the utilized model along with the principal component, which enhances the robustness, imperceptibility, and payload capacity. Even though the model had certain advantages, the model was computationally complex to suit the real-world applications and find robustness only with counterfeit attack. In [21], Seyed Mostafa Pourhashemi *et al*., utilized the Ensemble watermark detector that included the K-Nearest Neighbor (KNN), Support Vector Machine (SVM) along with the DWT. The drawbacks of the conventional methods in audio watermarking were recovered, the effects of certain watermarking attacks were observed, and the high security learning was emerged with the Ensemble model. The model failed to involve certain high-performance classifier along with bio-inspired optimizations that may increase the detection accuracy. Hwai-Tsu Hu and Tung-Tsun Lee in [22] discussed the digital audio watermarking with the 3-level Lifting Wavelet Transform (LWT) ensemble model, which consist of the RDM-MV, VN-AQIM, and DCT-AQIM. The model provided the robust watermarking with the three level approximations and able to withstand all common attacks that enhanced the tracking synchronization code, which made the frame alignment possible. The major challenge observed in the model was the disability to trace the frame boundaries with the self recovery of the watermarking scheme, which was unmanageable. In [23], Seyed Mostafa Pourhashemi *et al*., introduced the Lucas sequence as well as the Fast Fourier Transform (FFT), in which the Lucas Sequence helped in achieving the high payload capacity. The model provided the audio signal in the discrete frequency steps that made the model loss energy in terms of the adjacent frequencies. In [24], Venkata Lalitha Narla *et al*., implemented the BCH code-based watermarking and hash generation-based tamper recovery. The model included the hash bits that utilized in the tampering detection enhanced the robustness, imperceptibility, and the payload values. Though the proposed model provided better performance, the model could not withstand the de-synchronization attacks that resulted in the less robust performance of the model. Md. Shohidul Islam *et al*., suggested the dual domain two-fold audio watermarking with the dual-tree complex wavelet transforms (DTCWT), Sort-time Fourier Transform (STFT), and SVD in [25]. This ensemble model resisted the signal processing attacks that usually degrade the robustness of the model. The better watermark detecting wavelets such as the types and the nature of the primary and the mother wavelet was not considered that degraded the efficiency of the model. Wenhuan Lu *et al*., in [26] introduced the robust feature point scheme algorithm (RFPS) to achieve the robust digital audio watermarking. The model eliminated the host signal interference as well as the de synchronization attacks. The accuracy of the detection through the RFPS method was decreased, due to the zoom level utilized in the model. Uma Nair and Gajanan K. Birajdar in [27] provided the DWT-SVD-AES model that acquired high-capacity audio watermarking along with the chaotic watermarking algorithm, which enhanced the encryption. The advanced encryption standard (AES) encryption involved in the model used the chaotic baker map, which was considered as the strategy depending permutation. Though the encryption method added advantage to the research, the model was hard to implement and follow the same simple structure of encryption that can be easily broken. Hwai-Tsu Hu and Tung-Tsun Lee in [28] introduced the Unified FFT framework with adaptive vector norm modulation (AVNM). The models are robust against the signal processing attacks and have the better composite measures that even created the trade-off between the payload as well as the model robustness. The model faced challenges due to the unrecoverable loss, when setting the cutoff frequency lower. Mohsen yoosefi Nejad *et al*., [29] modeled the digital audio watermarking with the Least Significant Bit (LSB) based quantum NEQR. The model reduced the circuital complexity by utilizing the quantum computing and had the better term of transparency. The schemes implemented in the model required the tedious process that resulted in the complex computation. Mahdi Mosleh *et al*., [30] involved the LU decomposition as well as the DCT with the Genetic algorithm (GA). Though the utilized GA finds optimal frequency that improved the trade of transparency, capacity and robustness, the model has high payload capacity that in turn decreased the transparency. Hence the model can apply the fuzzy logic as well as the intelligent watermark detector to increase the robustness of the research. In [31], Mohsen Yoosefi Nejad *et al*., explained the quantum audio watermarking with the DCT. The model worked with the frequency domains that resulted in efficient quantum circuits and robustness of the model in audio watermarking. The way of dividing the frequency domain signal remained the major challenge as it was derived only for the sake of reducing quantum grates. Hence, the model should choose the more appropriate medium frequency component that increase transparency. The author Shiqiang Wu *et al*., [32] achieved higher embedding capacity as well as the robustness with the codebook spread spectrum with the diversity reception (DR).The model did not consider the attacks encountered in the watermarked audio and Gaussian distribution, which were used to model the HIS that failed to provide the theoretical support to the DR. In addition, the model faced the white boxer attack that developed the security issues. K. Vivekananda Bhat *et al*., [33] developed the SVD model to perform the digital audio watermarking. The model has certain advantages such as the imperceptibility, good robustness against the stirmark attacks, but failed to include the attacks other than the stirmark attack. In [34], Juren Qin *et al*., introduced the Lattice based Meet-in-the-middle Embedding (MME) that aided in rectifying the lattice quantization errors that are properly scaled to achieve the better digital audio watermarking. In addition, the model achieved the higher robustness with the less distortion, but showed some higher probability of the decoding errors. The Pranab Kumar Dahr *et al*., [35] suggested the Parametric Slant-Hadamard Transform (PSHT) with the Hessenberg Decomposition (HD). The model was computationally fast and had high payload capacity in even resisting almost every attacks. The outcomes of the model were not compared based on the SOTA methods, which were the standard way of the comparison. Kehan Chen *et al*., [36] presented the Dual quantum DCT model in the digital audio watermarking. Though the model had the capability to withstand the attacks of the watermarked audio content, the trade-off between the imperceptibility and the robustness remained the major challenge of the model.

2.3 Analysis and Discussion

2.3.1 Summary of the Related works

The below tabulated section summarizes the related works that are considered for the digital audio watermarking research. This analysis includes the methods, advantages, disadvantages, achievements and the future scope of the 22 research articles tabulated in Table 2.1.

Table 2.1: Summary of the related works

Citation	Model	Advantages	Disadvantages	Achievements	Future scope
[15]	Hybrid DWT and Schur decomposition	The model was robust against several attacks and had high payload capacity	The embedding process of the model could include only small size of the watermark image.	SNR-81.43 dB, ODG-0.18, SDG-4.78, payload rate- 319.29 bps.	The model can focus on enhancing the watermark image size and to implement the technique in the image and video signals
[16]	Huffman data in the MP3 compression	The model showed the better performance with the de compressed MP3 source.	The model worked only with the decompressed sources that lack the adaptability characteristics of the model.	ODG-0.49	The model can include the robust methods to achieve the better watermarking in the MP3 files.
[17]	Minimum amplitude scaling on optimized	The algorithm utilized in research achieved the robust performance by minimizing the	The process included various parameters that increased the computational	BER – 0.25%	The model can decrease the number of parameters utilized to achieve the better

	DWT.	optimization problems.	complexity of the model.		efficiency of the research.
[18]	Fuzzy interference with SVD and DCT.	The model is robust against several attacks and had no imperceptibility, and the audible quality is stable.	The utilized fuzzy system required a proper defined function, which was difficult to implement.	SNR- 49,80 dB, ODG-0.18, capacity-598.34bps	In future, the watermark detector based on the neural network to enhance the performance.
[19]	SVD and FRT	The better performance was achieved in the segment-based implementation in terms of the correlation-based domains.	SVD failed to capture the complex relationships that led the model to achieve the poor accuracy.	Phase angle - $\frac{5\pi}{4}$.	The model should include the complex relationships with the features to attain the maximum detection accuracy.
[20]	Integer wavelet transform	The model was robust against counter fit attack.	The proposed model had high computational complex.	SNR-32.89, ODG-0.52, capacity-655.36 bps	Soft computing techniques can be implemented to mitigate the computational complexity.

[21]	Ensemble model of SVM, KNN and DWT.	The model enhanced the accuracy and created the robust model.	The model failed to involve certain high-performance classifier	SNR-43.82dB, ODG-0.354	The ensemble model could include certain high-performance classifier along with the optimization algorithms to achieve the maximum accuracy.
[22]	3-level LWT ensemble with RDM-MV, VN-AQIM, DCT-AQIM.	The model provides the robust watermarking with the three level approximations and able to withstand the attacks. Themodel has the better tracking synchronization code.	The self recovery of the watermarking scheme is unable to handle.	SNR- 19.17 to 20.07 dB, ODG- -0.56 to -0.79, payload capacity- 1523.9 bps.	The model can attach the self recovery of the watermarking scheme.
[23]	Lucas sequence and FFT	The parameter tuning process in the research adds effectiveness to	The model provided the audio signal in the discrete frequency steps.	Payload capacity- 1 to 8 bps,SNR- 33 to 58 dB,	The model can utilize the intelligent extraction method based on machine learning

		accomplish the compromise between capacity, imperceptibility, and robustness.		ODG--0.35 to -1.57.	algorithms
[24]	BCH code-based watermarking and hash generation-based tamper recovery.	The hash bits utilized in the tampering detection enhance the robustness, imperceptibility, and the payload values.	The proposed model cannot withstand the de-synchronization attacks.	SNR-36.73 dB, Payload capacity-409.6 bps, Detection accuracy -93%.	The model can include the synchronization code to overcome the de-synchronization attacks.
[25]	Ensemble model with DTCWT, STFT, and SVD.	The model can resist the signal processing attacks,	The better watermark detecting effect can be included to improve the efficiency.	Payload capacity-10,270 bps,	The better watermarking techniques can be included to improve the efficiency.
[26]	Feature points scheme algorithm.	The model eliminates the host signal interference as well as the de	The accuracy of the detection through the feature point extraction method decreases, due to	BER-12.5%	The steady features should be extracted to ensure the change in feature points and to attain the embedding watermark

		synchronization attacks.	the zoom level utilized.		sequence that has high embedding capacity.
[27]	DWT-SVD-AES	The model acquired high capacity audio watermarking along with the chaotic watermarking algorithm that enhances the encryption.	The encryption model followed simple and same structure of the encryption.	SNR-55.48dB, payload-3010 bps	The model should include certain complex encryption model to enhance privacy of the data.
[28]	Unified FFT framework with AVNM.	The models were robust against the signal processing attacks and have the better composite measures.	The model caused unrecoverable loss, when setting the cutoff frequency lower.	Payload capacity-344.53 to 1033.59, SNR-29.98 dB, ODG--0.05.	The model should include techniques to eliminate the loss.
[29]	LSB based quantum NEQR	The model reduced the circuital complexity by utilizing the quantum computing and had the	The provided schemes required to implement the various processes that resulted in the complex	SNR-61.51 dB	The process of the schemes should be minimized to achieve the better efficiency.

		better term of transparency.	computation.		
[30]	LU decomposition and DCT transform with GA.	The utilized GA finds optimal frequency that improves the trade of transparency, capacity and robustness.	The model has high payload capacity that in turn decreases the transparency.	SNR- 33.41 dB, ODG--0.32, SDG- 4.67, payload capacity -2002 bps.	The model can apply fuzzy logic as well as the intelligent watermark detector to increase the robustness of the research.
[31]	Quantum audio watermarking with DCT	The model worked with the frequency domains that resulted in efficient quantum circuits and robustness.	The way of dividing the frequency domain signal remains the major challenge as it was derived only for the sake of reducing quantum grates.	SNR-44.32 dB	Selecting the better techniques to choose the more appropriate medium frequency component that increase transparency.
[32]	Extended codebook spread spectrum with DR	The model achieved higher embedding capacity as well as the	The model did not consider the attacks encountered in the watermarked audio and Gaussian distribution used	SWR -24.55	The model should include the algorithm that should resist the white boxer attack that

		robustness.	to model the HSI lack to provide the theoretical support.		leads to the security issues.
[33]	SVD	The model provides the high imperceptibility, good robustness against the stirmark attacks.	The model worked only with the stirmark attack and ignored several other attacks.	Payload capacity- 196 bps, SDG- -0.2, ODG- -0.28.	The model can include the de-synchronization attack algorithms that enhance thes efficiency of the model.
[34]	Lattice based MM embedding	The model achieves higher robustness with less distortion.	The model showed certain probability of decoding errors.	SWR- 33.24 dB, SNR-25dB, ODG- -1.5.	The model should utilize the various techniques to improve the accuracy of the research model.
[35]	PSHT-HD	The model worked resisting the almost every attack, computationally faster and has high payload.	The result obtained from the proposed method was not evaluated against the SOTA methods, which were considered as the	Payload capacity- 172.39 bps, SNR- 46.56 dB, ODG- -.39	The model should be compared with several SOTA methods

			comparison standards.		
[36]	Dual quantum - DCT	The model had the capability to withstand the attacks of the watermarked audio content.	The trade-off between the imperceptibility and the robustness remains the issue.	SNR-36 dB, 22dB.	The efficient metrics should be designed to estimate the congruity between the two or more audio contents.

2.3.2 Analysis based on the methods

The analysis based on the methods provides the better knowledge to the researchers to understand the utilized algorithms and techniques on the digital audio watermarking. The methods such as DWT, DCT, and so on are utilized and are tabulated in Table 2.2.

Table 2.2: Analysis based on methods

Methods	References
DWT	[15], [17], [22]
DCT	[18], [27], [30], [36]
SVD	[18], [19], [22], [27], [33]
IWT	[20]
FFT	[23], [28]
LSB	[29]
SVM-KNN	[21]

2.3.3 Performance Evaluation metrics

Signal-to-Noise Ratio (SNR):

SNR is the objective metric that computes the ratio of the noise to the original audio that represents the better imperceptibility of the digital audio watermarking. The SNR acts as the statistical measure introduced by the IFPI standards and is represented as,

$$SNR = 10\log_{10} \frac{\sum W^2(s)}{\sum_s \left[W(s) - W'(s)\right]^2}, \tag{2.1}$$

Where, $W(s)$, represents the original audio and $W'(s)$, represents the watermarked audio with the particular s, sample. The IFPI standards suggest that above 20 dB indicates the watermarked audio remains imperceptible.

Objective Difference Grade (ODG):

ODG is the objective grade obtained from the perceptual evaluation of the audio quality (PEAQ). The model evaluates the dissimilarities of the original as well as the audio watermarked signals. The usual range of the ODG varies from -4 to 0, where -4 represents the annoying watermarked audio and the 0 represents the signals with no difference. The database that utilizes the human auditory system can have the score that is greater than zero and that can be estimated with the open-source software called EAQUAL, which is elaborated as Evaluation of the Audio Quality.

Payload Capacity:

The total number of bits included in the signal especially in the watermarked signal at the certain period is known as the payload capacity. The measure of the signal as the bits per second is represented as,

$$Payload = \frac{Bits}{Time(\text{sec})}(bps). \tag{2.2}$$

Usually, the payload capacity carries the unit of bits per second (bps). The minimum payload capacity of the watermarked signal should be 20bps.

Subjective Difference Grade (SDG):

SDG calculates the impairment degree of the original and the watermarked audio sources. As ODG, SDG also exhibits the range of -4 to 0, where -4 indicates the terrible impairment as well as 0 indicates the no impairment.

Bit-error-Rate (BER):

BER measures the error rate in terms of the bit within the original and the watermarked signal. If the watermarked and the original signal are identical, then the error rate remains zero or neutral. In general, the BER is the measure used to address the robustness of the proposed models against the several auditory attacks. The BER is estimated as,

$$BER = \frac{\sum_{s=1}^{N} W(s) \oplus W'(s)}{N}, \tag{2.3}$$

Where, N, represents the total samples including the sample s.

Detection accuracy:

The detection accuracy is utilized to detect the tampers or the watermarks included in the original audio signal. The higher accuracy obtains when the accurate detection is made in the proposed model. The detection accuracy is estimated as,

$$Detection\ Accuracy = \frac{No.of\ samples\ watermarked}{Total\ samples}. \tag{2.4}$$

2.3.4 Analysis based on the Performance metrics

The performance metrics acts as the measure of the qualitative characteristics of the digital audio watermarking techniques. The several performance metrics that are included in the literatures are analyzed and illustrated in Figure 2.2.

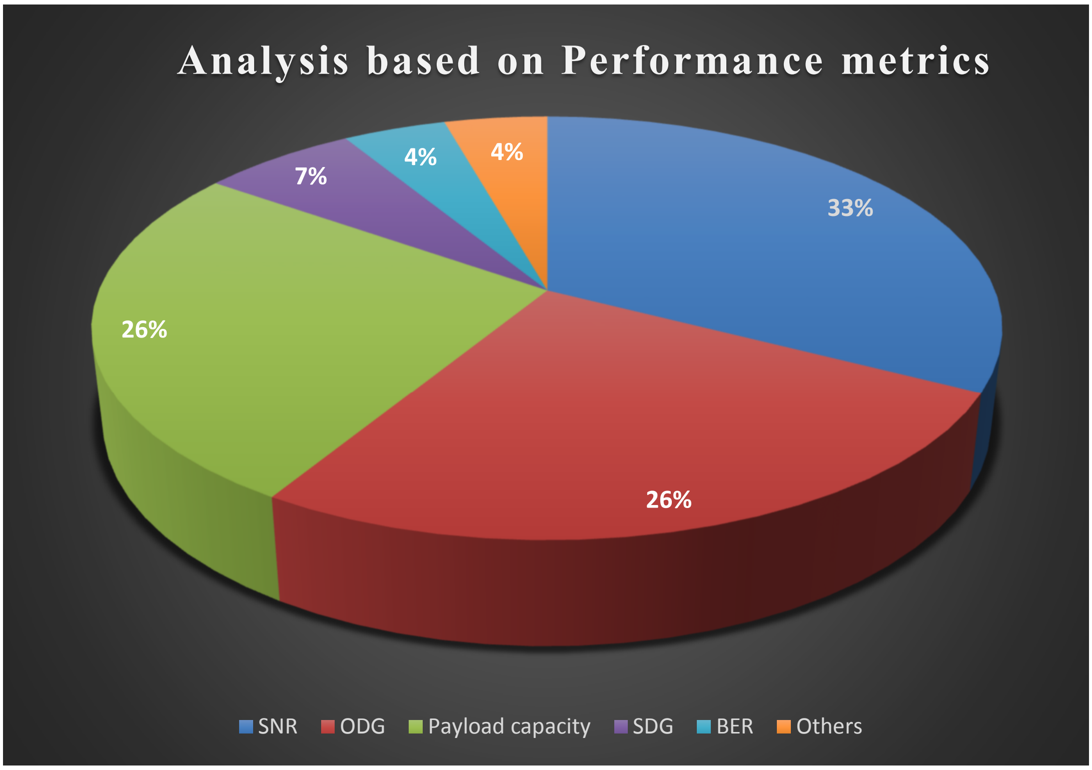

Figure 2.2: Analysis based on Performance metrics

2.4 Research gaps

- A robust watermark should withstand common signal processing operations and attacks while remaining imperceptible. However, ensuring robustness without compromising security is a delicate task, as some robust watermarking methods might be more susceptible to malicious attempts at removal.

- Maintaining perceptual transparency, where the presence of the watermark is imperceptible to human listeners, is crucial for user acceptance. Striking a balance between embedding a watermark robust enough for detection and ensuring it doesn't degrade the audio quality poses a significant challenge.

- Embedding additional information as a watermark reduces the capacity available for the audio content itself. Finding an optimal trade-off between the amount of watermark information embedded and the preservation of audio fidelity is a constant challenge in digital audio watermarking.

- ❖ Digital audio exists in various formats and may undergo compression using different codes. Designing watermarks that are resilient across diverse audio formats and compression methods is a challenge, as each format may introduce unique challenges to the watermarking process.

- ❖ Digital audio may undergo various signal processing operations during its lifecycle, such as equalization, filtering, or compression. Adapting watermarking techniques to handle these real-world signal processing challenges, which can vary widely, is a complex task that researchers need to address for practical applicability.

2.5 Summary

The discussed literature review chapter elaborated the information on the digital audio watermarking technique including the models, contribution, their achievements and so on. The added research articles are classified based on the models that are utilized to achieve the better accuracy in the digital audio watermarking. The research models are elaborated on the drawbacks and the respective future scope that can address the future research ideas for the improvement of audio watermarking. The analysis included in the chapter discusses the models as well as the performance metrics that shows the majority of metrics included in the audio watermarking researches beforehand.