ISSN: 2633-4828 Vol. 5 No. S6, (Oct - Dec 2023)

International Journal of Applied Engineering & Technology

MEL FILTER BANK-BASED CEPSTRAL & WAVELET ANALYSIS OF CHHATTISGARHI DIALECTS IN SPEECH PATTERNS USING MACHINE LEARNING

[1]Madhuri Gupta, [2]Dr. Megha Mishra and [3]Dr. Jaspal Bagga

[1]Research Scholar, Department of Computer Science, Shri Shankaracharya Technical Campus, Bhiali, India

[2]Department of Computer Science, Shri Shankaracharya Technical Campus Bhilai, India

[3]Department of IT, Shri Shankaracharya Technical Campus Bhilai, India

ABSTRACT

Chhattisgarh the central part of India with diversity in religion, culture and geographic location has so many untouched fields of research. One of the areas is recognition of chhattisgarhi dialects using automatic speech recognition tools. Wavelet transformation is giving that base to extract the required features of different dialects and analyze the diversity of different dialects.

The Chhattisgarhi dialects have unique speech patterns, with this research work focusing on five specific dialects: Kankeri, Chamarii, Baheliya, Khairagarhi, and Devarboli. The main objective is to differentiate among these dialects through the application of advanced speech characterization methods, encompassing Mel Filter bank-based cepstral (MFCC) features and Wavelet Transform specifically Haar Wavelet Family. This paper portrays a summary about shared characteristics by examining filters, histograms, and statistical metrics, including standard deviation, mean, and mode, especially Mean. This Research gathers the mean value of different dialects such as Kharagarhi (-5.376019106),Baheliya(-3.379147377),Chamari(-3.5835),Devarboli (-13.9225969) and Kankeri(-7.163349057).

The fundamental principle of data analysis is by using wavelet transformation. Researchers found that wavelets initiate a transformative shift in information processing, fostering a creative mindset and perspective. This study contributes to the academic domain by offering profound insights into the distinctive qualities that define Chhattisgarhi dialects, however highlighting the efficacy of wavelet-based analysis in capturing their nuanced attributes.

This research also bridging the gap between conventional speech analysis methods and modern signal processing approaches, this study advances our understanding of language variation and makes indispensable contributions to the expansive field of speech recognition technology.

Keywords: Speech Recognition, Wavelet Transformation, Mel Filter bank-based cepstral (MFCC),Chhattisgarhi dialects,Feature Extraction,Automatic Speech Rrecognition(ASR).

1. INTRODUCTION

Speech is the form of interpersonal expression of thoughts, therefore study on the many layers of human expression and automatic speech recognition has been increasingly interesting. This curiosity has evolved over the previous fifty years from the mechanical realization of speech capabilities to the wish to automate tasks that require human-machine interaction.

Besides simply what is said, signals from speech carry additional information about what is being said, gender, emotional undertones, and speaker character throughout. Although recognition of speech works to understand spoken words, computerized speaker recognition systems concentrate on obtaining, characterization, and discovering characteristics pertinent to speaker identification from the sound flow.

Chhattisgarhi, an Indian dialect spoken in Chhattisgarh, is linguistically significant. Given the variety of dialects found in Chhattisgarhi, the current study focuses on 5 in particular: Kankeri, Chamarii, and Bilaspuri,Baheliya and Bhijwari. The method distinguishes between distinct dialects by applying speech characterisation addresses such as Mel Filter bank-based cepstral(MFCC) characteristics and Wavelet Transformation. The algorithm recognizes nuances and draws inferences about commonalities by examining various filters, histograms, and standard deviation, mean, and mode.

The application of wavelet transformation for the analysis of Chhattisgarhi dialects is founded on the fundamental concept of scrutinizing data with respect to scale. Scholars within the wavelet discipline assert that the utilization of wavelets engenders a paradigm shift, fostering a novel mindset and perspective in data processing. Wavelets, characterized as functions meeting specific mathematical criteria, involve a important criteria in the representation of data and other functions. The notion of approximating data through the superposition of functions has historical roots dating back to the early 1800s, notably observed when Joseph Fourier demonstrated the ability to superimpose sine and cosine functions to represent diverse functions.

In the realm of wavelet analysis, the selection of scale emerges as a critical factor in data examination. This underscores the significance of considering the scale employed when delving into wavelet-based data analysis. The Wavelet Algorithm, discussed in the context of various journal and international conferences.

Research gaps and problems are noted in Guo's (2022) study of wavelet analysis and its applications. The review offers insightful information about the uses of wavelet analysis, but it doesn't specifically state what the identified research gaps are. To fully comprehend the subtle features of these gaps—such as untapped application domains, methodological constraints, or new areas where wavelet analysis could be further optimized—more research is required. A concentrated investigation into these facets would lead to a more thorough comprehension of the possible constraints and unexplored prospects in wavelet analysis. After finding the gap the research focuses on providing a more robust solution for that.

Throughout a conference proceeding compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style.

This study focuses on literature survey of various studied done so far in this area, the data collection, analysis of MFFC and Wavelet Transformation and result and discussion of Chhattisgarhi dialects,

2. LITERATURE SURVEY

Multiple investigations utilizing wavelet transformation in speech recognition have produced encouraging results. By creatively fusing the wavelet transformation with the traditional CEP vector, Hao (2000) introduced a revolutionary feature that significantly enhanced system performance. Notwithstanding the capabilities of the MFCC extraction of features method that is typical, Kotnik (2003) suggested an effective method based on wavelet packet deconstruction for noise-robust speech feature extraction. For speaker-dependent, isolated word recognition, Krishnan (1994) investigated the effectiveness of Discrete Wavelet Transform when combined with Dynamic Time Warping, Vector Quantization, also known and Artificial Neural Networks, showing significant promise for feature extraction. Janer (1996) made a contribution to the field by proposing a technique that uses the Modulated Gaussian Wavelet Transform for non-uniform speech segmentation, which improved the system's recognition rate. Timoshevskaya and ei suggested scholarly work delves into the foundational theoretical principles and inherent properties of wavelet transforms. It addresses the intricacies of digital data processing grounded in wavelet transforms, emphasizing the analysis and manipulation of signals and functions exhibiting non-stationarity in time and inhomogeneity in space. The authors systematically introduce methodologies for progressive coefficient valuation, seamlessly integrating wavelet decomposition and quantization. The goal of effectively capturing the most significant data in a signal is a smart combination that serves the overriding [5].

Wavelet transform approaches have been thoroughly investigated in an array of recent years. Providing an in-depth investigation of a number of wavelet forms and their different uses by Rajini's seminal work from 2016 stands as a rigorous review. fingerprint image authentication, and biomedical image processing are a few notable uses by Radar technology. Study digs into the intricate areas of wavelet transform, with a special focus on signal analysis, image processing, and data compression by Tong-ying's 2004. Kumar's 1997 study addresses the use of wavelet analysis as a tool within geophysical domains, showing its effectiveness in evaluating nonstationary processes and discovering singularities. Furthermore, Saha's illuminating 2019 survey plunges into advanced wavelet

Continuous wavelet transformations (CWTs) offer a mechanism for comprehending a wide range of tidal phenomena that depart significantly from the presumed statistical stationarity or exact periodicity inherent in classic tidal methods. The use of wavelets enables the measurement of the degree of non-stationarity contained in time series that are often viewed as stationary, such as estuarine and shelf currents [10].

After studying all the literature in the literature survey, the study concludes the research gap in the following points:

Integration of Wavelet Techniques in Speech Recognition Systems: While the studies mentioned showcase the efficacy of wavelet transformations in speech recognition, there is a potential research gap in exploring the integration of various wavelet techniques into existing speech recognition systems. Investigating how different wavelet methods interact and complement each other could lead to enhanced performance.

Application of Wavelet Transform in Real-world Environments: The studies predominantly focus on controlled environments. A research gap exists in understanding the performance of wavelet-based speech recognition systems in real-world scenarios with background noise, varying acoustic conditions, and diverse speakers.

Comparative Analysis of Wavelet Types in Different Domains: Although Rajini's work provides a comprehensive review of diverse wavelet types, there is room for further research comparing the effectiveness of specific wavelet types in different application domains. This could help identify the most suitable wavelet for specific tasks.

Exploration of Wavelet Techniques in Tidal Phenomena: The text briefly mentions continuous wavelet transforms for understanding tidal phenomena. A research gap may exist in further exploring and evaluating the application of wavelet techniques in studying tidal dynamics, comparing their effectiveness with traditional methods.

After studying all the literature in the literature survey the study conclude the research gap in the following points:

Integration of Wavelet Techniques in Speech Recognition Systems: While the studies mentioned showcase the efficacy of wavelet transformations in speech recognition, there is a potential research gap in exploring the integration of various wavelet techniques into existing speech recognition systems. Investigating how different wavelet methods interact and complement each other could lead to enhanced performance.

Application of Wavelet Transform in Real-world Environments: The studies predominantly focus on controlled environments. A research gap exists in understanding the performance of wavelet-based speech recognition systems in real-world scenarios with background noise, varying acoustic conditions, and diverse speakers.

Comparative Analysis of Wavelet Types in Different Domains: Although Rajini's work provides a comprehensive review of diverse wavelet types, there is room for further research comparing the effectiveness of specific wavelet types in different application domains. This could help identify the most suitable wavelet for specific tasks.

Exploration of Wavelet Techniques in Tidal Phenomena: The text briefly mentions continuous wavelet transforms for understanding tidal phenomena. A research gap may exist in further exploring and evaluating the application of wavelet techniques in studying tidal dynamics, comparing their effectiveness with traditional methods.

Table 1: Literature Summary

S.No	Literature Review Summary		
	Author Name	**Year of Publication**	**Summary**
1.	Hao (2000)	2000	Introduced revolutionary feature by fusing wavelet with CEP vector, significantly enhancing system performance
2.	Kotnik (2003)	2003	Suggested effective method based on wavelet packet deconstruction for noise-robust speech feature extraction using MFCC
3.	Krishnan (1994)	1994	Investigated effectiveness of Discrete Wavelet Transform combined with Dynamic Time Warping, Vector Quantization, and Artificial Neural Networks for speaker-dependent, isolated word recognition
4.	Janer (1996)	1996	Proposed Modulated Gaussian Wavelet Transform for non-uniform speech segmentation, improving system's recognition rate
5.	Tomashevskaya and ei	Year not specified	Explored foundational theoretical principles of wavelet transforms for digital data processing, emphasizing non-stationarity and inhomogeneity
6.	Rajini (2016)	2016	Meticulous review of diverse wavelet types and applications, including radar technology, fingerprint image authentication, and biomedical image processing
7.	Tong-ying (2004)	2004	Explored wavelet transform theory and application in signal analysis, image processing, and data

			compression
8.	Kumar (1997)	1997	Focused on wavelet analysis in geophysical applications, highlighting its efficacy in analyzing nonstationary processes and detecting singularities
9.	Saha (2019)	2019	Survey on advanced wavelet transforms in image processing, discussing their significance, architecture, and applications

3. METHODOLOY

a. What exactly is wavelet transformation?

The wavelet transform is a strong tool for examining data since it divides it into multiple frequency components and studies each with a resolution proportional to its scale (Tabe, 2009). It is especially useful for distinguishing meaningful signals from noise and finding peak overlaps in chromatographic applications (Wahab, 2020). The technique, which is based on group theory and square integrable representations, allows for the decomposition of a signal into space and scale by applying analysis functions known as wavelets (Farge, 1992). It's used in signal processing, image coding, and numerical analysis (Farge, 1992). The wavelet transform is a relatively recent topic in signal processing, with ongoing study into its development and use (Bentley, 1994).

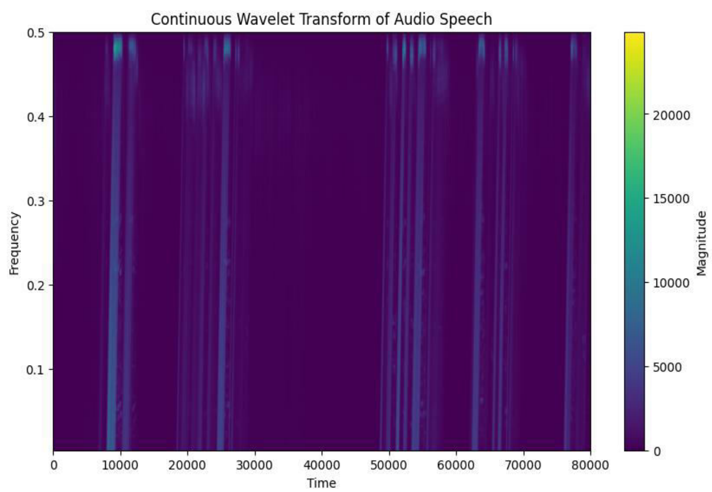

Fig 1: Continuous Wavelet Transformation of Audio: Chamari Dialects

The provided Python code employs the continuous wavelet transform (CWT) for analyzing audio speech signals. Utilizing the wavelets library, the code loads an audio file, performs CWT, and generates a spectrogram-like representation. This visualization displays time on the x-axis, frequency on the y-axis, and color to signify wavelet coefficients' magnitude. The choice of the 'cmor' wavelet and parameters can be adjusted based on specific signal characteristics. This code facilitates the exploration of signal features at varying scales, offering insights into the time-frequency domain structure of audio signals, crucial for tasks such as speech analysis or classification in machine learning as shown in Fig1.

b. Why do we need the wavelet transformation?

The wavelet transforms, a mathematical technique, offers unique advantages compared to the traditional Fourier transform. Various studies delve into its applications and benefits. A paper on "Application of Wavelet Transform and its Advantages Compared to Fourier Transform." The primary focus is on discerning the advantages of wavelet transform over Fourier transform. Wavelet Transforms and Their Applications," provides a historical introduction to wavelet transforms and emphasizes their relevance in applications, including Gabor transforms.

Wavelets provide simultaneous localization in both time and frequency domains, a crucial feature discussed in "Wavelet Transform Analysis to Applications in Electric." This simultaneous localization ability distinguishes wavelets, making them suitable for applications where precise temporal and frequency information is essential.

Furthermore, a comprehensive review in "A Review of Wavelet Analysis and Its Applications" [11] summarizes the advantages of wavelet analysis across various scenarios, offering insights into its diverse applications and comparative results. The benefits include higher efficiency in particular settings when compared to alternative approaches. In essence, the wavelet transform's unique ability for simultaneous time and frequency localization, in addition to its adaptability, presents it as an effective instrument with applications in an extensive variety of domains. Recordings of 12,000 utterances of numerals and 41,884 clips from YouTube videos, considering factors like background music, data from multiple environments, noise suppression avoidance, and the use of keywords from different languages in a single phrase is tested in the system[12]. focuses on analyzing the effect of frequency value characteristics utilized in the Fast Fourier Transform (FFT) process of the Mel Frequency Cepstral Coefficients (MFCC) feature extraction on the recognition accuracy of children's speech in noisy conditions in the paper [13]. These advantages make it preferable to standard Fourier-based approaches in some situations. VMD-based feature extraction algorithm for the fusion of MFCC and IMFCC speech emotions showed better emotion recognition rates compared to traditional speech emotion features in the proposed work [14].

4. DATA COLLECTION

Understanding the complexity of human communication is made possible in large part by speech analysis, particularly in linguistically diverse places like Chhattisgarh, India. In this work, modern wavelet analysis techniques are employed for differentiation and delineation.

Diverse dialects of Chhattisgarhi have diverse speech patterns. This study focuses on five distinct dialects: Kankeri, Chamarii, Baheliya, Khairagarhi, and Devarboli. For feature extraction, all dialects are first recorded in MP3 audio and subsequently converted to wav files. Sentences from everyday conversations, spoken by adult, child, female, and native speaker speakers, are included in this work.

5. WORKING FLOWCHART OF MFCC

In audio signal processing and speech recognition, the Mel-frequency cepstral coefficient (MFCC) feature extraction procedure is essential. It entails a number of procedures designed to extract pertinent data from audio signals for further examination and categorization. Below is a thorough breakdown of every step:

a) **Divide the Signal into Brief Frames:** The audio signal is separated into brief frames, each lasting between 20 and 40 milliseconds. This segmentation makes it possible to analyze the signal's properties over brief time periods and record the signal's temporal fluctuations. By taking into account variations in speech patterns and phonetic content, framing makes sure that the signal is conveyed in a way that makes it possible to extract pertinent features.

b) **Pre-emphasis:** High-frequency components, which are typically weaker than low-frequency components, are amplified in a signal by applying pre-emphasis. By increasing the higher frequencies in relation to the lower frequencies, this procedure raises the signal-to-noise ratio overall and boosts the efficiency of later processing stages.

c) **Windowing:** A window function is performed to each frame of the signal to attenuate discontinuities at the frame boundaries. The Hamming and Hanning windows are examples of common window functions. They minimize spectral leakage and reduce artifacts caused by abrupt transitions by tapering the frame's edges to zero. Windowing makes ensuring that the signal segments taper gently, which makes spectral analysis more precise and reduces

d) **Fast Fourier Transform (FFT):** The frequency-domain representations of the windowed frames are obtained by applying the Fast Fourier Transform (FFT). By converting the signal from the time domain to the frequency domain, this transformation reveals details about the signal's spectral properties. By breaking down every frame into its individual frequency components, FFT makes it possible to retrieve spectral properties like phase and magnitude information. Frequency domain theory (FFT) assists in detecting characteristic spectral patterns that are necessary for voice recognition.

e) **Mel Filter Bank:** A Mel filter bank divides the spectrum into several frequency bands in accordance with the Mel scale after the FFT-derived spectra are passed through it. This non-linear frequency scale highlights areas of higher perceptual relevance and is better in accordance with human auditory perception. The spectrum envelope of the signal is captured and significant spectral features are highlighted by the Mel filter bank, which computes the energy distribution across various frequency bands. The Mel filter bank improves the derived features' ability to discriminate by simulating the frequency sensitivity of the human auditory system.

The flowchart showing all the steps is shown in the Fig 2: MFCC working Flowchart

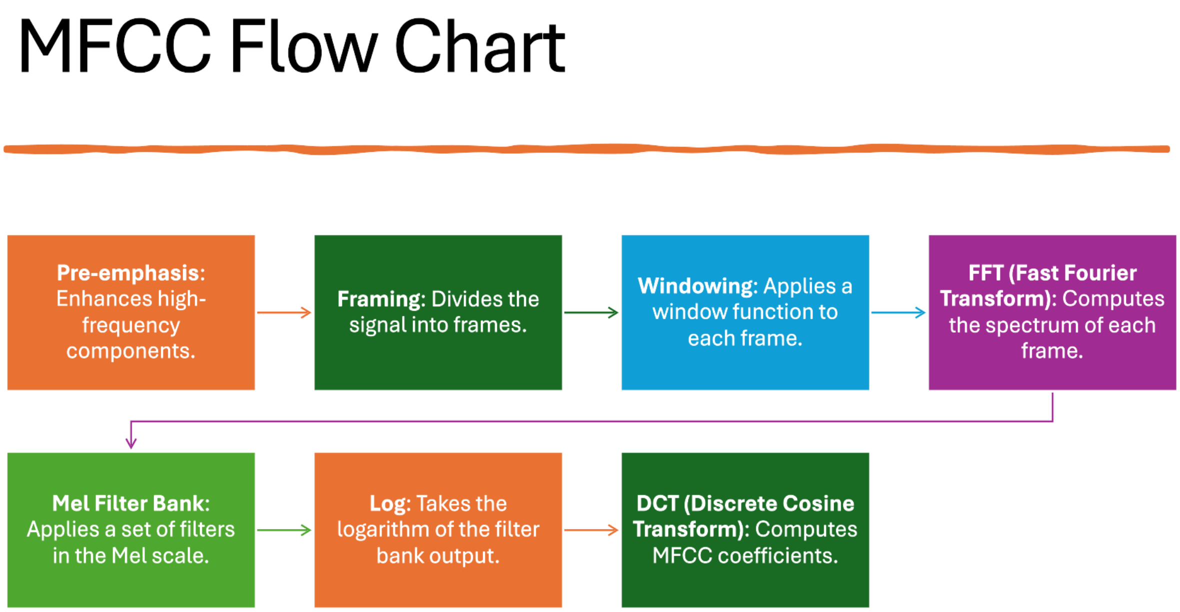

Fig 2: MCFF Working FlowChart

6. RESULT AND DISCUSSION

The research uses Python code for wavelet transform audio signal analysis. Here's a more detailed explanation by algorithm: As shown in the fig 3.

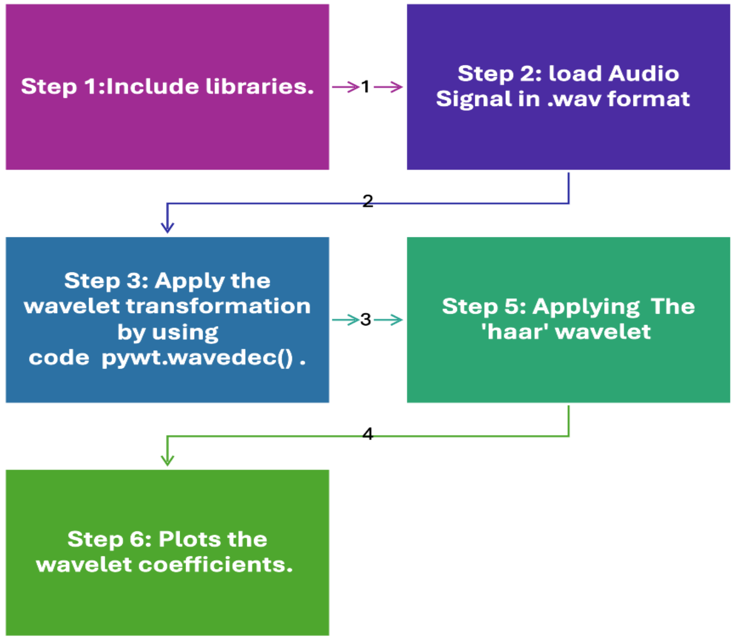

Fig 3: Algorithm for Model Creation

Step 1: Include libraries such as pywt for wavelet transform, numpy for numerical operations, matplotlib.pyplot for plotting, and scipy.io.wavfile for audio file reading.

Step 2: load Audio Signal in .wav format using wav.read() with the file location specified.

Step 3: Apply the wavelet transformation by using code pywt.wavedec() .

Step 5: Applying The 'haar' wavelet for breaking the data into multiple frequency components.

Step 6: Plots the wavelet coefficients. A subplot is formed for each level of the decomposition, and the coefficients are shown in figure 2.

For Histogram we have algorithm as:

Step 1: Apply plt.hist() to build a histogram of the original audio stream.

Step 2: Plot Histogram and this histogram depicts the distribution of amplitudes, providing insight into the signal's overall properties as shown in Figure 3.

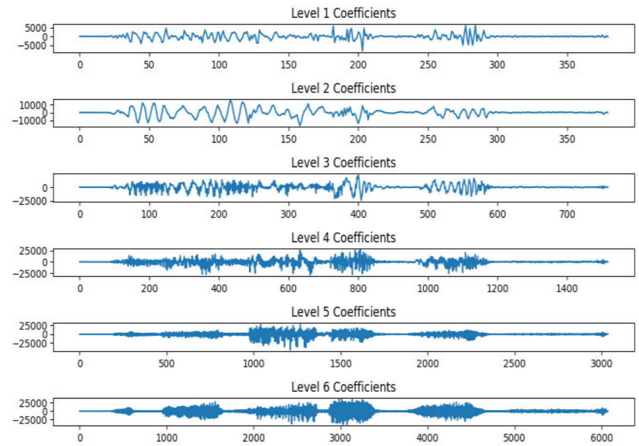

Fig 2: Signal analysis using the wavelet transform technique: Using Haar Wavelet

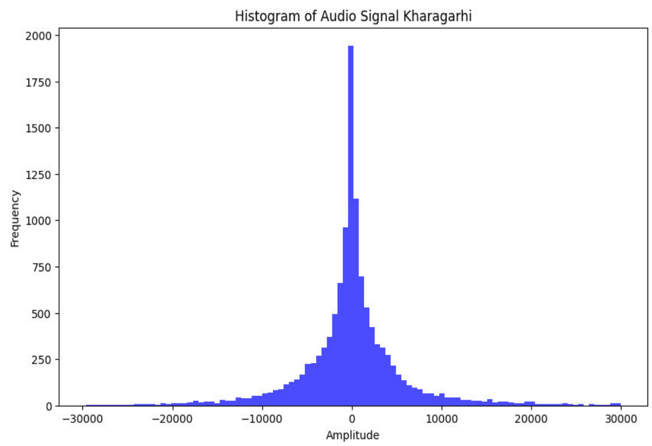

Fig 3: Histogram of Karagarhi Dialects

i. Mean Comparison of Different Dialects

In the below Bar Chart 1 which is Showing bar plot comparing mean values of different dialects. It visualizes a dataset with dialect names and corresponding mean numerical values.

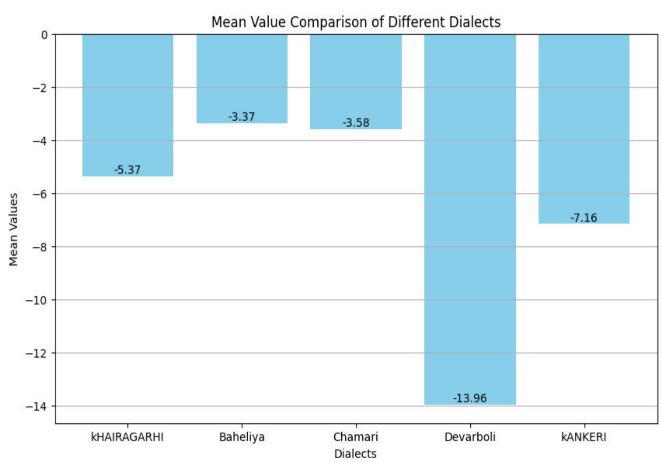

Chart 1: Mean Comparison of Different Dialects

7. CONCLUSION

Using advanced wavelet analysis constitutes a big step forward in understanding linguistic diversity in the examination into the speech patterns of Chhattisgarhi dialects. To reveal a dedication to recognizing the subtle grammatical variations present in these different expressions, the research focuses on the Kankeri, Chamarii, Bilaspuri, Baheliya, and Bhijwari dialects. By applying Mel Filter bank-based cepstral (MFCC) characteristics and Wavelet Transformation not only improves detection between various languages, but also enables a deep web of common linguistic qualities. The algorithm's ability to examine filters, histograms, and statistical metrics such as mean, and mode goes beyond simple word recognition, providing a comprehensive view of the complexity of human expression. Fostering a paradigm change in data processing by applying Wavelet transformation integration emerges as a cornerstone. Researchers agree that by using this transformative technique fosters an innovative mindset and also provides a unique lens through which linguistic nuances can be understood. This research work makes an essential contribution to the research and academic domain, enriching the understanding of the distinctive aspects of Chhattisgarhi dialects.

By smoothly spanning the gap among traditional speech analysis and modern digital signal processing approaches this study not only encourages academic discourse but also extends the extensive field of speech recognition technology.

REFERENCES

[1] Hao, Y., & Zhu, X. (2000). A new feature in speech recognition based on wavelet transform. WCC 2000 - ICSP 2000. 2000 5th International Conference on Signal Processing Proceedings. 16th World Computer Congress 2000, 3, 1526-1529 vol.3.

[2] Kotnik, B., Kacic, Z., & Horvat, B. (2003). The usage of wavelet packet transformation in automatic noisy speech recognition systems. The IEEE Region 8 EUROCON 2003. Computer as a Tool., 2, 131-134 vol.2.

[3] Krishnan, M., Neophytou, C.P., & Prescott, G.E. (1994). Wavelet Transform Speech Recognition Using Vector Quantization, Dynamic Time Warping And Artificial.

[4] Janer, L., Martí, J., Nadeu, C., & LLEIDA SOLANO, E. (1996). Wavelet transforms for non-uniform speech recognition systems. Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96, 4, 2348-2351 vol.4.

[5] Timoshevskaya, O., Londikov, V., Andreev, D., Samsonenkov, V., & Klets, T. (2021). DIGITAL DATA PROCESSING BASED ON WAVELET TRANSFORMS.

[6] Rajini, G.K. (2016). A COMPREHENSIVE REVIEW ON WAVELET TRANSFORM AND ITS APPLICATIONS.

[7] Tong-ying, G. (2004). Wavelet Transform Theory and Its Application Progress: a Review. Information & Computation.

[8] Kumar, P., & Foufoula-Georgiou, E. (1997). Wavelet analysis for geophysical applications. Reviews of Geophysics, 35, 385 - 412

[9] Saha, M., Naskar, M.K., & Chatterji, B.N. (2019). Advanced Wavelet Transform for Image Processing—A Survey. Lecture Notes in Networks and Systems.

[10] Flinchem, E.P., & Jay, D.A. (2000). An Introduction to Wavelet Transform Tidal Analysis Methods. Estuarine Coastal and Shelf Science, 51, 177-200.

[11] T. Guo, T. Zhang, E. Lim, M. López-Benítez, F. Ma and L. Yu, "A Review of Wavelet Analysis and Its Applications: Challenges and Opportunities," in IEEE Access, vol. 10, pp. 58869-58903, 2022, doi: 10.1109/ACCESS.2022.3179517.

[12] Himadri, Mukherjee., Sk, Md, Obaidullah., K., C., Santosh., Santanu, Phadikar., Kaushik, Roy. (2020). A lazy learning-based language identification from speech using MFCC-2 features. International Journal of Machine Learning and Cybernetics, 11(1):1-14. doi: 10.1007/S13042-019-00928-3

[13] Risanuri, Hidayat. (2022). Frequency Domain Analysis of MFCC Feature Extraction in Children's Speech Recognition System. Jurnal Infotel, 14(1):30-36. doi: 10.20895/infotel.v14i1.740

[14] (2023). Speech Emotion Feature Extraction Method Based on Improved MFCC and IMFCC Fusion Features. doi: 10.1109/eebda56825.2023.10090810