IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

Text-Independent Automatic Dialect Recognition of Marathi Language using Spectro-Temporal Characteristics of Voice

Miss. Palwe Priyanka Dharmnath and Prof. Mahajan Jagruti R
Department of Computer
Adsul Technical Campus, Chas, Ahilyanagar, India
priya.palwe19@gmail.com and mahajanjagruti@gmail.com

Abstract: *India's linguistic landscape is rich with regional variations, making dialect recognition an important component in speech-based technologies. This study presents a text-independent automatic dialect recognition system for the Marathi language, focusing on four regional dialects: Marathwada, Puneri, Vidarbha, and Goan Marathi. Unlike conventional speech recognition systems that depend on fixed phrases, the proposed system operates independently of textual content and instead relies on spectro-temporal features extracted from voice signals. The system utilizes the LDC-IL Marathi speech corpus, comprising 7555 audio samples recorded at 48 kHz. Fifteen acoustic features—including Mel-Frequency Cepstral Coefficients (MFCCs), spectral centroid, chroma energy, spectral contrast, and tempo—are extracted using the Librosa library. To improve classification efficiency and reduce feature redundancy, feature selection techniques such as Chi-Square, Mutual Information, and ANOVA f-test are applied. Six machine learning classifiers—K-Nearest Neighbors (KNN), Naïve Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Stochastic Gradient Descent (SGD), and Ridge Classifier (RC)—are trained on these features. Experimental results reveal that the SGD classifier, when paired with Chi-Square-selected features, achieves the highest accuracy of 84.64%. This demonstrates the system's effectiveness in handling subtle dialectal variations without relying on complex deep learning models, making it suitable for deployment in resource-constrained environments. The study contributes to the development of inclusive voice technologies and sets a foundation for further research in dialect-aware automatic speech systems, especially for underrepresented Indian languages.*

Keywords: Marathi Dialect Recognition, Spectro-Temporal Features, MFCC, Feature Selection, Machine Learning, Chi-Square Test, LDC-IL Corpus, Text-Independent Speech Processing, Regional Language Technology, Classifier Comparison

I. INTRODUCTION

The exponential growth of voice-based interfaces has revolutionized how humans interact with machines. From virtual assistants like Siri and Alexa to automated call centers and voice-based search engines, speech technology has become an integral part of modern digital ecosystems. At the core of these systems lies the ability to accurately understand and process spoken language, which includes recognizing not just words and sentences but also speaker characteristics such as emotion, gender, age, and dialect. Among these, dialect recognition is particularly crucial in multilingual societies like India, where a single language can have multiple dialects differing significantly in phonology, prosody, and accent.

Dialect recognition systems aim to classify the regional variation of a spoken language by analyzing the unique characteristics embedded in speech signals. These systems are essential for improving the performance of downstream speech applications such as Automatic Speech Recognition (ASR), Text-to-Speech (TTS) systems, and speaker verification systems. Moreover, dialect recognition contributes to preserving linguistic diversity, promoting inclusive technology design, and enabling more personalized and culturally aware voice services.

DOI: 10.48175/IJARSCT-26239

The Marathi language, spoken predominantly in the Indian state of Maharashtra and parts of neighboring regions, exhibits rich dialectal variation. Major dialects include Puneri, Marathwada, Vidarbha, and Goan Marathi. Despite their mutual intelligibility, these dialects differ in intonation, pronunciation, stress patterns, and phoneme usage. Such variation often poses a challenge to standard speech processing systems that are typically trained on standardized forms of the language. Therefore, building dialect-aware systems tailored to these variations is critical for enhancing the reliability and accessibility of speech technologies for Marathi speakers.

In the context of speech processing, dialect recognition can be approached using two primary methods: text-dependent and text-independent systems. Text-dependent systems require speakers to utter specific predetermined phrases, which limits their applicability in real-world scenarios where spontaneous speech is more common. In contrast, text-independent systems operate without constraints on linguistic content, making them more versatile and practical for deployment in natural communication settings. This research focuses on developing a text-independent dialect recognition system for the Marathi language using acoustic and temporal features of speech.

The fundamental idea behind text-independent dialect recognition lies in the analysis of the spectral and temporal properties of the voice signal. These features capture essential information about a speaker's pronunciation, rhythm, pitch, and speaking style, which collectively contribute to dialectal identity. Among the most widely used features in speech processing are Mel Frequency Cepstral Coefficients (MFCCs), which simulate the human auditory system's response and are effective in representing phonetic content. Additionally, features such as spectral centroid, spectral contrast, chroma energy, zero-crossing rate, and tempo offer valuable insights into the prosodic and tonal characteristics of speech.

This study employs the Linguistic Data Consortium for Indian Languages (LDC-IL) Marathi Speech Corpus, a high-quality dataset consisting of 7555 audio samples recorded at 48.0 kHz. The dataset includes recordings from diverse age groups and both genders, ensuring a representative sample of the Marathi-speaking population. Each audio sample is labeled with its corresponding dialect, making it suitable for supervised machine learning approaches. The dataset's richness in terms of linguistic and demographic diversity makes it an excellent resource for building and evaluating dialect recognition models.

The research methodology involves multiple stages, starting with preprocessing the raw audio data to ensure consistency and quality. Preprocessing steps include trimming silence, normalizing amplitude, resampling to a uniform rate, and converting stereo to mono. These steps help standardize the input and reduce variability due to recording conditions. Once the data is preprocessed, a comprehensive set of acoustic and temporal features is extracted using the Librosa Python library. The feature set includes 15 descriptors that encapsulate both short-term and long-term characteristics of speech signals.

Given the high dimensionality of the extracted features, feature selection becomes a critical step to enhance model performance and computational efficiency. This study explores three statistical techniques for feature selection: Chi-Square test, Mutual Information, and ANOVA F-test. These methods identify the most informative features by evaluating their relevance to the target labels (i.e., dialect classes). By reducing the feature space, the system minimizes overfitting, speeds up training, and improves generalization on unseen data.

For the classification task, six machine learning algorithms are employed: K-Nearest Neighbors (KNN), Naïve Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Stochastic Gradient Descent (SGD), and Ridge Classifier (RC). These classifiers are chosen for their simplicity, interpretability, and proven effectiveness in speech classification tasks. The models are trained on 70% of the dataset and tested on the remaining 30%. Performance is evaluated using metrics such as accuracy, mean absolute error (MAE), and mean squared error (MSE).

Among the classifiers, the combination of SGD with Chi-Square-selected features yields the highest accuracy of 84.64%. This result highlights the effectiveness of feature selection in enhancing classifier performance and underscores the potential of lightweight models for dialect recognition tasks. Unlike deep learning models, which require substantial computational resources and large datasets, traditional machine learning classifiers can achieve competitive performance when paired with appropriate feature engineering.

The contributions of this research are manifold. First, it addresses the underexplored area of dialect recognition for the Marathi language, thereby filling a significant research gap. Second, it demonstrates the feasibility of building accurate

DOI: 10.48175/IJARSCT-26239

ISSN
2581-9429
IJARSCT

and efficient text-independent dialect recognition systems using classical machine learning techniques. Third, the study provides a benchmark for future research in regional language processing and encourages the development of more inclusive speech technologies.

The practical applications of this research are extensive. In voice-based user interfaces, dialect recognition can enable more natural and culturally adaptive interactions. In educational technology, it can support pronunciation training and dialect-aware learning tools. In forensic linguistics, it can assist in speaker profiling and identification. Furthermore, in the context of digital governance and rural outreach, dialect-sensitive speech systems can enhance accessibility and user engagement.

Despite its promising results, the study also acknowledges certain limitations. The dataset, while diverse, is relatively small compared to large-scale speech corpora used in modern deep learning systems. Additionally, the system's performance in noisy or real-time environments has not been evaluated. These limitations point to directions for future research, such as integrating noise-robust feature extraction, exploring deep learning architectures, and expanding the dataset through data augmentation or crowdsourcing.

In conclusion, this study presents a robust and interpretable framework for text-independent dialect recognition in the Marathi language. By leveraging spectro-temporal features, statistical feature selection, and classical machine learning algorithms, the system achieves high accuracy while maintaining low computational overhead. This research not only advances the field of dialect recognition but also contributes to the broader goal of building inclusive and linguistically aware speech technologies. The findings set a strong foundation for future work in multilingual and dialect-rich language processing, ultimately supporting the vision of truly accessible and intelligent voice-enabled systems.

II. LITERATURE SURVEY

Dialect identification has been an essential area in speech processing and has contributed significantly to linguistic studies, automatic speech recognition, and forensic purposes. Techniques have been developed by researchers to facilitate dialect classification based on spectral, prosodic, and phonetic characteristics. Machine learning approaches were employed in early studies for improving classification performance, including SVM, decision trees, and ensemble. For instance, Devi and Thaoroijam [3] tested vowel-based dialect identification for Meeteilon by using spectral features such as formant frequencies and combining these with prosodic features, such as pitch, energy, and duration. Application of a Random Forest classifier to discriminate between Imphal, Kakching, and Sekmai dialects gave them an accuracy of 61.57%, bringing out the success of vowel-based features as methods of dialect identification.

Similarly, Chittaragi and Koolagudi [4] studied the recognition of Kannada dialects by a study of vowel sounds considering formant frequencies (F1–F3), energy, pitch (F0), and duration. They found through their experiments that ensemble classifiers such as Random Forest and Extreme Gradient Boosting were better than standard classifiers, and with an accuracy rate of 76%, proved that global and local features played an important role in the differentiation between dialects.

In another similar work, Ciobanu et al. [5] employed an ensemble of SVM classifiers with character and word-level n-gram features for German dialect classification. With data transcripts from four Swiss-German dialects (Basel, Bern, Lucerne, and Zurich), their system placed third in the German Dialect Identification shared task with a 62.03% F1 score, thereby proving the utility of ensemble methods in handling intricate dialectal variations.

Recent work in word-level dialect identification by Chittaragi and Koolagudi [6] involved extracting prosodic and spectral features from the IViE corpus where ensemble classifiers such as Extreme Gradient Boosting performed better than SVM. Later, Chittaragi et al. [7] followed this up by adding a greater number of features like Mel Frequency Cepstral Coefficients (MFCCs), spectral flux, entropy, pitch, energy, and duration to carry out dialect identification among nine British English dialects. Their comparative study demonstrated that although the Random Forest classifier was at a rate of 78.5%, Gradient Boosting outperformed it by 81.2%, again proving the efficacy of ensemble methods for dialect identification.

Honnavalli et al. (2019) [8] conducted accent detection through feature extraction models on the VCTK corpus, which contains a number of English regions. The authors contrasted classifiers such as Neural Networks, Logistic Regression, KNN, SVM, and GMM. Results showed that Neural Networks and Logistic Regression achieved 95% accuracy, followed by KNN at 91%, SVM at 54%, and GMM at 43%. It highlighted the capacity of deep learning models for accent identification to learn even slight nuances of phonetic objects. From a performance perspective, the paper focuses on feature engineering and prioritizes the utilization of MFCC features for improved performance. The two primary issues so far have been scalability and data variability. This research also proposed a hybrid strategy employing the standard model with some elements from the deep learning aspect in order to realize optimal performance without computationally wasteful. It did a lot in facilitating the recognition of accent by providing the analysis of strengths and weaknesses of the classifiers, along with matters concerning applicability in multilingual settings.

Singh et al. (2024) [9] analyzed Telugu accents, Telangana, Rayalaseema, and Coastal Andhra, by feature extraction with MFCC and classification with SVM and KNN. The accent challenge of the region was solved, and classification of native speech samples provided encouraging results. The authors explained that the technique of extraction needs to be adjusted to particular linguistic contexts because Telugu has regional and culturally conditioned different phonetic patterns. Singh et al. discovered that the models face specific training issues: accent overlap and variability of datasets. To prevent such issues, they suggested methods that involve data augmentation and sophisticated preprocessing techniques. This study was to demonstrate the viability of machine learning efficiency in constructing the localized speech recognition system and made thoughtful contributions towards its usage across linguistically diverse areas.

Pratik Kurzekar et al. (2021) [10] developed a continuous speech recognition system with a farm application for Marathi. This employed MFCC and Fast Fourier Transform as the feature extractor and attained 92.5% accuracy. The research indicated that phonetically dense datasets were in significant dearth for Marathi, an important limitation. The authors suggested a large-scale dataset of heterogeneous phonetic and dialectal variation to enhance the generality of the system. Future research would include further experimentation with neural network structures for further acoustic modelling as well as injecting context-specific information beneficial for use in agriculture. The research has demonstrated the viability of domain-dependent speech recognition systems in practical real-world applications, particularly with rural or under-developed communities in mind.

Despite remarkable progress, existing research in dialect identification is limited by a number of key gaps. Most work has been confined to particular languages or local accents, making the results non-generalizable to various linguistic environments. While ensemble methods and conventional machine learning have achieved encouraging results, the combination of sophisticated neural architectures is not adequately investigated. Additionally, ongoing issues like accent overlap, fine-grained dialectal differences, and the variable quality of existing datasets worsen the challenges involved in proper dialect classification. This problem is especially acute for underrepresented languages like Marathi, where the lack of large, phonetically diverse, and rich datasets further erodes model robustness. Overcoming these challenges is crucial to the creation of scalable and stable dialect identification systems that can capture the subtle linguistic characteristics embedded in natural speech.

III. METHODOLOGY

The methodology adopted in this research revolves around designing an efficient and lightweight dialect recognition system that functions independently of the lexical content of speech. The process is broken down into six systematic stages: (1) Dataset Acquisition, (2) Preprocessing, (3) Feature Extraction, (4) Feature Selection, (5) Model Training and Classification, and (6) Evaluation.

Dataset Acquisition

The dataset used in this study is the Linguistic Data Consortium for Indian Languages (LDC-IL) Marathi speech corpus, which contains audio samples representative of various dialects spoken across Maharashtra and Goa. It is an ideal resource for this study because it contains carefully recorded and annotated speech data from different regions, making it suitable for training machine learning models in dialect recognition.

- Total samples: 7555 audio clips
- Dialect categories: Marathwada, Puneri, Vidarbha, Goan Marathi
- Sampling rate: 48.0 kHz

DOI: 10.48175/IJARSCT-26239

ISSN
2581-9429
IJARSCT

- Speaker variability: Male and female speakers from various age groups
- Speech type: Spontaneous and read speech, allowing for text-independent recognition

This balanced dataset enables generalization across different speaking styles and demographics, which is critical for building a robust system.

Preprocessing

To ensure consistency and enhance the quality of features extracted, all audio samples undergo a series of preprocessing steps using Python and Librosa:

- Resampling: All audio files are resampled from 48.0 kHz to a fixed 22.05 kHz to reduce processing time while retaining sufficient spectral detail.
- Mono Conversion: Stereo audio files are converted to mono to simplify feature extraction and modeling.
- Normalization: Volume normalization is performed to bring all audio signals to a common amplitude range.
- Noise Trimming: Silence and background noise are removed using energy-based thresholding to retain only the informative segments of speech.

These steps prepare the dataset for reliable and consistent feature extraction, which is crucial in text-independent tasks.

Feature Extraction

Feature extraction is the cornerstone of speech-based machine learning systems. The goal is to capture the distinct spectro-temporal characteristics of dialectal speech that can be fed into classifiers.

Acoustic Features Used

Using the Librosa library, 15 acoustic features were extracted from each audio file:

- MFCC (Mel Frequency Cepstral Coefficients): 13 coefficients capturing the vocal tract characteristics
- Zero-Crossing Rate: Measures frequency content changes
- Spectral Centroid: Indicates the center of mass of the spectrum
- Spectral Bandwidth: Measures the spread of the spectrum
- Spectral Contrast: Difference in energy between peaks and valleys
- Spectral Rolloff: Frequency below which a certain percentage of the total spectral energy lies
- Chroma Features: 12 chroma vectors that represent the pitch class
- RMS Energy: Measures the energy in the signal
- Spectral Flatness: Indicates noisiness of a sound
- Tempo: Beat-related rhythmic characteristic
- Root Mean Square Value
- Onset Strength
- Tonnetz
- Short-time Fourier Transform (STFT) magnitude
- Mean and Standard Deviation across frames

These features are well known for capturing both *static* (e.g., pitch and tone) and *dynamic* (e.g., rhythm and rate) characteristics of speech which vary across dialects.

Feature Vector Construction

Each audio file is processed frame-by-frame (usually 25 ms windows with 10 ms overlap). From these, statistics like mean, median, and standard deviation are computed for each feature across the time dimension. This ensures that each sample is represented as a fixed-length feature vector, enabling use in classical machine learning algorithms.

DOI: 10.48175/IJARSCT-26239

Feature Selection

Due to the high dimensionality of the extracted feature vectors, not all features contribute equally to dialect differentiation. Feature selection helps improve classifier performance by removing irrelevant or redundant attributes. Techniques Used Three statistical methods are applied:

- Chi-Square Test: Measures dependence between features and target labels
- Mutual Information: Measures shared information between features and labels
- ANOVA F-test: Compares variance between and within groups to test significance

Each method ranks features based on their discriminative power. The top-k features (where k is tuned) are selected and used for training classifiers.

Importance of Feature Selection

- Reduces overfitting
- Lowers training time
- Improves classification accuracy
- Enhances model interpretability

Experiments revealed that the Chi-Square selection technique yielded the best performance in combination with the Stochastic Gradient Descent (SGD) classifier.

Classification Models

The system evaluates six machine learning models using the selected features:

K-Nearest Neighbors (KNN)

- Non-parametric, instance-based learning
- Distance metric: Euclidean
- Tends to overfit on noisy features without selection

Naïve Bayes (NB)

- Probabilistic classifier based on Bayes theorem
- Assumes independence among features
- Fast but may suffer from accuracy drop with correlated features

Support Vector Machine (SVM)

- Constructs optimal separating hyperplanes
- Kernel used: Linear
- Performs well on high-dimensional spaces

Decision Tree (DT)

- Rule-based model using feature thresholds
- Easy to interpret but prone to overfitting

Stochastic Gradient Descent (SGD)

- Optimizes a linear model using gradient descent
- Suitable for large-scale and sparse datasets
- Shows highest accuracy in this study

Ridge Classifier (RC)

- Regularized linear regression for classification
- Balances model complexity and performance
- Each classifier is trained using an 80:20 train-test split and evaluated using cross-validation to ensure generalization.

Evaluation Metrics

The performance of each classifier is evaluated using the following metrics:

- Accuracy: Proportion of correctly predicted dialects over the total predictions
- Mean Absolute Error (MAE): Average absolute difference between predicted and actual labels
- Mean Squared Error (MSE): Average squared difference between predicted and actual labels

The following observations were made:

- SGD + Chi-Square: Best combination with 84.64% accuracy
- SVM and Ridge: Competitive performance but higher computational time
- KNN and DT: Sensitive to feature scaling and prone to overfitting without selection
- Naïve Bayes: Fast but slightly lower accuracy due to strong independence assumption

IV. CHALLENGES FACED

Developing a text-independent automatic dialect recognition system for the Marathi language posed several technical, linguistic, and computational challenges. These challenges span from dataset limitations to algorithmic constraints, and each impacted the system's accuracy, generalizability, and scalability. Addressing these challenges was essential for building a robust model and for setting directions for future enhancements.

Limited Availability of Annotated Dialect Data

One of the most significant challenges was the limited availability of high-quality, annotated speech datasets for Marathi dialects. Unlike major global languages, Marathi lacks extensive open-source corpora with dialect-level labeling. Although the LDC-IL Marathi corpus provided a structured dataset, it was still modest in size compared to those available for English or Hindi. The relatively small sample count (7555 audio files across four dialects) imposed constraints on training complex models like deep neural networks and limited the ability to generalize across speaker variability.

Dialectal Overlap and Mutual Intelligibility

The high degree of mutual intelligibility among Marathi dialects posed a challenge in designing discriminative models. While differences in tone, prosody, and pronunciation exist across dialects such as Marathwada, Puneri, Vidarbha, and Goan Marathi, the acoustic distinctions can be subtle, making it difficult for standard classifiers to identify patterns without extensive feature engineering. The variations are often contextual and influenced by speaker habits, requiring more advanced techniques like deep embeddings or prosodic modeling to capture accurately.

Background Noise and Recording Quality

Although preprocessing steps were implemented, some audio files still contained ambient noise, speaker hesitation, overlapping speech, or inconsistent loudness, which affected feature extraction. Such distortions are especially problematic in text-independent systems, where the acoustic environment has a stronger impact due to the absence of lexical cues. While noise trimming helped, it was not always possible to perfectly isolate clean speech segments, particularly from longer or spontaneous recordings.

Feature Redundancy and Dimensionality

The extraction of 15 distinct spectro-temporal features from each audio file, each with multiple statistical variations (mean, variance, etc.), led to a high-dimensional feature space. Without proper feature selection, this redundancy could lead to overfitting, increased computational time, and degraded model performance. Identifying the most relevant features using methods like Chi-Square and ANOVA f-test was essential, but it required iterative experimentation and fine-tuning.

Class Imbalance in Dialect Representation

In the dataset, some dialects—such as Puneri and Vidarbha—had more samples than others, particularly Goan Marathi. This class imbalance could bias classifiers toward over-represented dialects, resulting in reduced accuracy for under-represented dialects. Though techniques like stratified sampling and data augmentation were considered, the intrinsic distribution of the dataset remained a constraint in maintaining uniform classification performance across all dialects.

DOI: 10.48175/IJARSCT-26239

Generalization to Unseen Speakers

Text-independent dialect recognition systems must generalize well across speakers, genders, and age groups. However, due to limited diversity in the dataset, especially among elderly or very young speakers, the model occasionally misclassified dialects when exposed to speaker variations not seen during training. This highlights the challenge of speaker dependency, even in models designed to be speaker-agnostic.

Lack of Deep Learning Integration

While classical machine learning algorithms such as SGD and SVM performed well with selected features, modern deep learning techniques like CNNs or LSTMs—which have shown high performance in speech-related tasks—were not integrated in this study due to dataset size and computational constraints. This limited the system's capacity to learn deeper temporal dependencies and latent dialectal patterns.

Evaluation in Real-Time or Noisy Environments

The system was tested in a controlled offline environment, using preprocessed audio files. However, real-time dialect recognition involves handling variability in microphones, background environments, and spontaneous speech flow. This real-world applicability is yet to be tested and remains a challenge for future deployment.

Tool and Library Constraints

Most audio processing was conducted using Python and the Librosa library. While Librosa is versatile, it occasionally faced limitations in terms of feature resolution, memory handling of large audio files, and limited prosodic feature support (e.g., pitch contour, stress modeling). These limitations restricted the exploration of more complex acoustic features that might be useful in dialect classification.

Applications Observed

The development of a text-independent Marathi dialect recognition system opens up a wide range of practical applications across multiple domains, particularly in a linguistically diverse country like India. Voice-based technologies, including virtual assistants and speech-to-text systems, are increasingly being integrated into everyday life. By recognizing dialects, these systems can deliver more accurate and personalized interactions. When dialect-aware models are incorporated into voice assistants like Alexa, Siri, or Google Assistant, users experience improved recognition accuracy and culturally relevant responses. This adaptation enhances user satisfaction and makes voice-based human-computer interaction more inclusive, especially for non-standard speech patterns.

In the field of education, dialect recognition has significant potential in the development of intelligent tutoring systems and language learning tools. Speech-based educational applications can offer feedback based on a learner's dialect, helping them gradually align with standardized pronunciation without disregarding their native speech traits. This is especially beneficial for school children in rural areas, where regional accents often differ significantly from the textbook language. Additionally, such systems can support speech therapy, accent correction, and interactive language labs for Marathi learners.

Another important application lies in automatic speech recognition (ASR) systems. Integrating dialect recognition into the ASR pipeline allows for the dynamic selection of dialect-specific models. When the system first identifies the dialect of the input speech, it can then load the appropriate ASR engine, leading to significant improvements in transcription accuracy. This approach is especially useful in large-scale applications like call centers, audio transcribers, and voice-controlled software platforms.

The forensic domain also stands to benefit from dialect recognition technologies. In forensic linguistics and speaker profiling, recognizing the dialect of a speaker can provide insights into their geographic origin. This can be vital in cases involving anonymous threats, hoax calls, or unidentified voice recordings. Law enforcement agencies can utilize dialect classification to narrow down suspect pools based on speech evidence, adding another layer to speaker verification systems.

In media and broadcasting, dialect recognition offers a means to personalize content and advertisements. Broadcasters can automatically categorize and tag regional content based on dialect, which is useful for creating region-specific playlists or archives. Advertising agencies can also use dialect cues to target users with localized commercials, increasing engagement and relatability. Moreover, news platforms and streaming services can tailor subtitles and voiceovers to match the audience's dialect, enhancing comprehension and viewer satisfaction.

DOI: 10.48175/IJARSCT-26239

ISSN 2581-9429 IJARSCT

E-governance platforms increasingly rely on voice interfaces to provide services to rural citizens. Dialect recognition ensures that these systems can understand and respond appropriately to users speaking in their native regional varieties of Marathi. This helps bridge the gap between rural populations and government services, promoting digital inclusion and efficient service delivery. Farmers, for instance, could receive weather updates, market prices, or agricultural advice in their dialect, making such services more accessible and useful.

In the domain of telemedicine and healthcare, voice-enabled systems equipped with dialect recognition can enhance communication between patients and healthcare providers. Many patients in remote or semi-urban areas prefer to speak in their regional dialects. If the system can recognize and adapt to these dialects, it becomes easier to collect accurate patient data and provide clear instructions. This is particularly valuable in mobile health applications where voice is the primary mode of interaction.

Smart devices and Internet of Things (IoT) applications also benefit from dialect recognition. Devices such as smart home assistants, agricultural tools, or wearable technology can incorporate voice control features that respond better to commands in the user's dialect. For example, a Marathi-speaking farmer using a smart irrigation device would be more comfortable giving voice commands in his native dialect, and the device’s recognition capability would be crucial to ensure correct operation.

Culturally, dialect recognition contributes to the documentation and preservation of linguistic diversity. As more systems begin to understand and classify dialectal speech, it becomes easier to archive oral traditions, folk songs, and dialect-specific narratives. Linguists and researchers can use such systems to study dialect evolution and maintain records of endangered speech patterns. This helps in preserving the rich cultural heritage of the Marathi-speaking community.

From a business perspective, dialect recognition enables market segmentation and localized service delivery. Businesses can analyze customer voice inputs to determine their regional preferences and offer products, advertisements, or services accordingly. Regional customization based on dialect enhances customer engagement, builds brand loyalty, and helps companies reach wider audiences across linguistic boundaries.

In summary, the applications of Marathi dialect recognition span a wide array of fields—education, voice assistance, healthcare, governance, security, media, smart technology, and business. Each application benefits from improved personalization, better comprehension, and more inclusive communication. The system developed in this research lays a strong foundation for deploying dialect-aware solutions in real-world contexts, making technology more adaptive and accessible for Marathi speakers across different regions.

V. CONCLUSION

This study has presented a text-independent, machine learning-based approach to dialect recognition for the Marathi language, focusing on the classification of four major dialects: Puneri, Marathwada, Vidarbha, and Goan Marathi. By leveraging spectro-temporal features—such as Mel Frequency Cepstral Coefficients (MFCC), spectral centroid, zero-crossing rate, chroma energy, and other acoustic attributes—extracted from a publicly available speech corpus (LDC-IL), the system successfully identifies dialectal variations without relying on any specific textual content. This makes the system highly applicable in real-world, spontaneous speech environments where speakers may not follow a script or pre-defined sentence structure. The study used statistical feature selection techniques, including Chi-Square, Mutual Information, and ANOVA f-test, to refine the extracted features and reduce dimensionality, ensuring better generalization and faster training times. A comparative analysis of six classical machine learning classifiers—KNN, Naïve Bayes, SVM, Decision Tree, SGD, and Ridge Classifier—was conducted. Among these, the Stochastic Gradient Descent (SGD) classifier paired with Chi-Square feature selection emerged as the most effective, achieving an accuracy of 84.64%. This result demonstrates that even with lightweight, interpretable algorithms, high accuracy in dialect classification is attainable when supported by meaningful feature engineering and appropriate preprocessing.

Beyond its technical performance, the system holds great potential for integration into various applications, including speech-to-text platforms, virtual assistants, educational tools, forensic linguistics, and e-governance systems. Its ability to recognize and adapt to regional dialects can significantly enhance user experiences and ensure that voice-based technologies are inclusive and accessible to diverse speaker communities. The research also contributes to the growing body of work aimed at supporting under-resourced languages and dialects in the field of computational linguistics. However, the study also faced several challenges, such as limited dataset size, class imbalance, speaker variability, and the absence of deep learning integration due to computational constraints. Despite these limitations, the system lays a solid foundation for future work. Expanding the dataset, incorporating deep neural architectures, integrating real-time capabilities, and improving noise robustness are logical next steps that can elevate the system's performance and scalability.

In conclusion, the research successfully demonstrates that text-independent dialect recognition using classical machine learning techniques is both feasible and effective for a linguistically rich and diverse language like Marathi. The findings support the broader vision of building dialect-aware, inclusive speech systems and contribute meaningfully to digital linguistics, voice interaction design, and regional language technology development. With further refinement and expansion, such systems can bridge the linguistic gap in speech technologies and promote equitable access to AI-powered voice interfaces.

VI. FUTURE OBSERVATIONAL SCOPE

The current study lays a strong foundation for Marathi dialect recognition; however, several areas offer scope for future enhancement. Integrating deep learning models like CNNs or Transformers can improve performance by learning complex speech patterns. Expanding the dataset to include more speakers, spontaneous speech, and noisy environments will boost the system's generalizability. Additionally, optimizing the model for real-time and edge deployment using compression techniques can enable use on mobile or embedded devices. Future work may also explore recognition of finer dialectal distinctions, code-switching behavior, and adaptation for other Indian languages. Incorporating user feedback mechanisms and active learning can make the system adaptive over time. Overall, these improvements will strengthen the practical usability and linguistic value of dialect recognition systems in real-world applications.

REFERENCES

[1]. T. C. Devi and K. Thaoroijam, "Vowel-based Meeteilon dialect identification using a random forest classifier," 2020.

[2]. N. B. Chittaragi and S. G. Koolagudi, "Acoustic-phonetic feature-based Kannada dialect identification from vowel sounds," International Journal of Speech Technology, 2019.

[3]. M. Ciobanu, S. Malmasi, and L. P. Dinu, "German dialect identification using classifier ensembles," Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects, 2018.

[4]. N. B. Chittaragi and S. G. Koolagudi, "Acoustic features-based word-level dialect classification using SVM and ensemble methods," 2018.

[5]. N. B. Chittaragi, A. Prakash, and S. G. Koolagudi, "Dialect identification using spectral and prosodic features on single and

[6]. Honnavalli, Dweepa & S S, Shylaja. (2021). Supervised Machine Learning Model for Accent Recognition in English Speech Using Sequential MFCC Features. 10.1007/978-981-15-3514-7_5.

[7]. M. K. Singh, D. Kishore, and R. A. Kumar, "Accent recognition of speech signal using MFCC-SVM and k-NN technique," Evergreen, vol. 11, no. 2, pp. 1305–1312, Jun. 2024.

[8]. Kurzekar, Pratik & Deshmukh, Ratnadeep & Waghmare, Dr. Vishal & Shrishrimal, Pukhraj. (2014). CONTINUOUS SPEECH RECOGNITION SYSTEM: A REVIEW. Asian Journal Computer Science & Information Technology. 4. 62-66. 10.15520/ajcsit.v4i6.3

DOI: 10.48175/IJARSCT-26239
ISSN 2581-9429 IJARSCT