IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Spoken Language Recognition In Indian Multilingual Context: A Review And Proposed Approach

[1]Aryan Mangesh Joshi, [2]Shreyas Ganesh Raparti, [3]Vishwas Shivaji Bhosle, [4]Aniket Maruti Pawar, [5]Vaishnavi Sunil Shelke

[1234]Student, [5]Professor

[12345]Computer Engineering,

[12345]Zeal Polytechnic, Pune, India

Abstract: In Indian context due to cultural diversity and difference languages in use, it's very of important to understand the language being spoken. It becomes even important when different spoken language team's come together for common goal. To handle this challenge, Spoken Language Identification (SLID) helps to automatically determine the language without having any background of the specific language. This applies even for foreign languages. people from different countries are coming together virtually or face to face. In India and global scenario identification becomes very challenging due to linguistic diversity, diverse phonetics, un developed or limited used language, dialectal variations and code mixing. This study provides more insights on the various major approaches studied, followed and ongoing development on SLID Mainly in SLID classical machine learning methods such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forests (RF), Extreme Learning Machines (ELM) and contemporary deep learning methods, including Convolutional Neural Networks (CNN), Recurrent Neural Networks with Long Short-Term Memory (RNN-LSTM), Deep Neural Networks (DNN), Convolutional Recurrent Networks (CRNN), and ensemble learning architectures are used. We intend to analyze trends in feature extraction, including MFCC, Mel-spectrogram, i-vectors and self-supervised embedding like wav2vec. Challenges in real-world SLID, such as short utterances, noise robustness, and multilingual speech datasets. The proposed workflow integrates preprocessing, feature extraction, DL training, and classification, demonstrating potential for robust SLID in multilingual and noisy environments.

Index Terms - Spoken Language Identification, Deep Learning, Indian Languages, CNN, LSTM, ELM, CRNN

I.INTRODUCTION

Spoken Language Identification (SLID) technique plays very important role to determine the spoken language without relying on any lexical or textual information. It is basic and foundational component of multilingual speech processing system. This enables downstream applications like automatic speech recognition, speech-to-speech translation, human–computer interaction, and call-routing services so and so forth. SLID can have multiple / numerous applications. While SLID is matured enough however in Indian context it still need lot of work and ground to be covered. This is mainly due to multiple scheduled languages i.e. around 22 language spoken within single country, accent variability. From Indian perspective Indian speech / language has substantial phonetic similarities but languages has its own richness .due to overlapping and accent challenges language identifications becomes very challenging in Indian context . Due to reginal use of language scarcity of large annotated corpora for many Indian languages is another challenge. Most languages

are low-resource which poses challenges to the development of robust SLID models. Furthermore, real-world variability—including environmental noise, channel distortion, spontaneous speech, and frequent code-mixing—adds complexity to SLID tasks These constraints make Indian SLID a demanding research problem

It is found that, deep learning architectures have played a transformative role in improving LID accuracy. Convolutional Neural Networks (CNNs) have shown strong performance on spectrogram-based features , while sequential models such as LSTMs offer superior modelling of temporal speech patterns . End-to-end pipelines and deep spectrogram-based architectures further strengthen LID systems.

Traditional statistical model are based on methods like Gaussian Mixture Models (GMM), Hidden Markov Models (HMM), and N-gram textual models. These models are proven to be effective for controlled datasets. However these approaches does not work properly with noisy, low-resource, or short-duration speech patterns. Machine Learning (ML) and Deep Learning (DL) approaches also have been deployed and have improved the SLID performance to great extent. ML methods such as Support Vector Machines (SVM), K-Nearest Neighbours (KNN), Random Forests (RF), and Extreme Learning Machines (ELM) can classify languages based on extracted acoustic features [1,20]. DL approaches, including CNNs, LSTMs, DNNs, and hybrid architectures, can automatically learn hierarchical representations of speech signals, improving noise robustness and generalization [3,4,6,7,22].

Despite this progress, Indian LID research work still has challenges. This review consolidates and analyses a structured overview of current trends, methodological developments, challenges, and likely future directions. By integrating classical, deep learning, and self-supervised approaches, this paper aims to provide a comprehensive perspective to guide researchers and developers, working toward robust and scalable Indian LID systems.

The remaining portion of this paper is structured into different sections to provide a comprehensive understanding of spoken language identification specifically for Indian languages. Section 2 illustrates an extensive literature review covering machine learning, deep learning, multilingual modelling, self-supervised techniques, and dataset-focused studies. Section 3 illustrates the conceptual workflow of a typical LID system, highlighting each stage from speech input to final language prediction. Section 4 details the comparative summary of major works, examining the key approaches, datasets, and language coverage reported in earlier research. Section 5 discusses the major challenges, potential future directions associated with Indian LID, followed by an analysis of existing research gaps identified through the reviewed studies. Section 6 outlines, emphasizing trend analysis and practical considerations for real-world applications. Section 7 proposed work and section 8 present concluding remarks, summarizing the main insights and highlighting the importance of focused research in Indian SLID. Finally, Section 8 provides references supporting all concepts and results discussed in the paper.

II. LITERATURE REVIEW

This section reviews key research developments in spoken language identification (LID), highlighting how existing studies help advance solutions—especially for multilingual and low-resource Indian contexts.

2.1 Machine Learning Approaches

Machine learning approach remain relevant for SLID, particularly for **small datasets** and low-resource languages. ML-based SLID generally involves **feature extraction followed by supervised classification**.

1)Feature Extraction

Acoustic features such as **Mel-Frequency Cepstral Coefficients (MFCC)**, **Shifted Delta Cepstral (SDC)**, **i-vectors**, and spectral features are widely used [20,22]. These features capture phonetic and spectral properties essential for differentiating languages.

2)Classification Models

- **Support Vector Machines (SVM):** this method is very Efficient in high-dimensional feature spaces and for small datasets. SVMs construct hyperplanes that separate different language classes [1].
- **K-Nearest Neighbors (KNN):** Uses distance metrics to classify speech based on similarity to known samples. Works well for limited datasets but is sensitive to noise.

- **Random Forest (RF):** Ensemble of decision trees improves generalization, reduces overfitting, and handles non-linear feature patterns [1].
- **Extreme Learning Machines (ELM):** Single-layer feedforward networks with randomly assigned hidden layer weights. Albadr et al. [20] enhanced ELM using a genetic algorithm to optimize weights (OGA-ELM), achieving accuracies up to 100%.

Advantages:
ML models are computationally efficient and require smaller datasets.
Limitations: Depend heavily on handcrafted features and may not generalize well in noisy or multi-speaker environments [5,10].

2.2 Deep Learning Approaches

Deep Learning (DL) approaches removes the ML limitations by **automatically learning hierarchical feature representations** from raw audio or spectrograms.

- **CNNs:** Capture local spectral and temporal patterns [4,7].
- **RNNs/LSTMs:** Model sequential dependencies in speech for robust short-utterance recognition [3,22].
- **DNNs:** Multi-layer perceptron's capable of learning complex relationships between features [21].
- **Hybrid & Ensemble Models:** Combine CNN, LSTM, and DNN with semi-supervised GANs (SSGAN) for improved generalization [21].
-

Key Literature Findings

- **Kulkarni et al. [3]:** Employed DNN and RNN on IndicTTS and IIITH datasets, showing robustness for multiple Indian languages.
- **Arla et al. [4]:** CNN-based SLID for Bengali, Gujarati, Tamil, and Telugu achieved 88.82% accuracy using MFCC spectrograms.
- **Garain et al. [21]:** Proposed FuzzyGCP ensemble of DDMLP, DCNN, and SSGAN with Choquet integral fusion, achieving 98% F1-score on MaSS dataset.
- **Biswas et al. [22]:** Used MFCC-based time series features with FRESH-based feature selection, obtaining 99.93–99.94% accuracy on Indian language datasets.

Advantages of DL:
- Automatic feature extraction
- Robust to noise and short utterances
- Scalable to large multilingual datasets

Challenges:
- Requires large labeled datasets
- High computational cost
- Overfitting on small datasets

2.3 *Several important insights emerge from the literature*

I. **Feature Engineering:** MFCC, Mel-spectrogram, i-vectors, and self-supervised embeddings founds to be very crucial. Aggregating features improves performance [22].
II. **Ensemble DL Models:** Combining multiple DL models with bagging, boosting, or fusion strategies improves robustness and generalization [21].
III. **Noise Robustness:** Most models are evaluated on clean datasets; real-world noise handling remains a challenge [10,19].
IV. **Short Utterances:** Models struggle with <2s speech; attention mechanisms or temporal pooling can help [7,21].
V. **Low-Resource Languages:** Many Indian languages lack large corpora, limiting supervised DL training [5,9].

Based on the studies we will work on these observations and findings to improve accuracy and robustness related to Indian languages

III.SPOKEN LID WORKFLOW

The conceptual workflow of a Spoken Language Identification (LID) system in figure 1 illustrates the complete pipeline—from the moment speech is captured to the final prediction of the language. This diagram helps readers to understand how different components interact and how features flow through the system to produce accurate LID results.

Figure 1: Spoken LID Workflow

The workflow consists of the following major stages:

1. *Speech Input*

The process begins with raw speech collected through a microphone or an audio file. This speech may come from different environments—quiet rooms, noisy streets, or mobile devices. The goal of this stage is simply to capture the spoken signal in its natural form.

2. *Preprocessing*

In this stage, the raw audio is cleaned and prepared for analysis.
Key steps include:

- Noise Removal: Unwanted background disturbances are filtered out to improve clarity.
- Voice Activity Detection (VAD): Segments containing speech are separated from silence or background noise.
- Framing & Windowing: The continuous signal is divided into small frames (e.g., 20–30 ms) to capture its time-varying nature.

This stage ensures that the speech is normalized, stable, and ready for feature extraction.

3. *Feature Extraction*

This is one of the most important steps in LID. Models cannot directly interpret raw audio; therefore, informative features must be extracted.

Common approaches include:

- MFCCs / LPCCs
- Mel-Spectrograms

Self-supervised embeddings (wav2vec 2.0). These features capture the unique spectral and phonetic patterns of each language. The better the features, the higher the accuracy of the LID system.

4. *Deep Learning / Machine Learning Classifier*

The extracted features are passed into a classification model, which may include:

- CNNs (for spectrogram images)
- LSTMs / RNNs (for sequential speech patterns)
- Transformers or wav2vec 2.0 models (for end-to-end learning)

The classifier learns how each language "sounds" based on patterns in the training data and produces a probability score for each possible language.

5. *Language Decision Module*

Finally, the system selects the language with the highest probability.
This stage also includes:

- Confidence estimation
- Threshold-based decisions

The output of this process is prediction of the language. This prediction can be used for applications like multilingual speech assistants, call routing, and voice-based services, and for many more applications

IV.CHALLENGES, RESEARCH GAPS, AND FUTURE DIRECTIONS

Spoken Language Identification for Indian languages have many challenges which affect the accuracy and scalability. One of the major challenges is the vast accent and existence of dialect variability across India. In India even within a single language there are notable different pronunciations. And this makes acoustic modeling difficult [1][2]. In addition to this complexity there is scarcity of annotated corpora and also many Indian languages have low-resource which limits the development of robust supervised learning models [16]. Another major challenge is code-mixing, especially in forms such as Hinglish or Marathenglish, which occur frequently in natural conversations and often confuse LID models trained on pure-language datasets. Furthermore, environmental noise found in real-world recordings continues to degrade system robustness despite progress in noise-resistant architectures [23]. A final challenge lies in the limited availability of comprehensive benchmark datasets that cover all 22 scheduled Indian languages, restricting fair comparison and standardized evaluation across studies.

These challenges expose several research gaps that must be addressed for developing comprehensive model. There is a noticeable lack of studies exploring transformer-based architectures specifically for Indian LID. Research on code-mixed datasets remains limited, even though such speech patterns are increasingly dominant in urban communication. Additionally, strategies for low-resource data augmentation are still underdeveloped, which limit the progress for languages with minimal speech resources. Another gap is the minimal exploration of hybrid models that combine self-supervised fine-tuning which can significantly improve the performance in low-data scenarios. Finally, benchmarking efforts are insufficient. We found that very few works evaluate performance consistently across more than a handful of Indian languages, which resulting in less linguistic spectrum under exploration.

The work on several future directions can strengthen this field and address the shortcomings. Advancements in self-supervised pretraining offer promising solutions for low-resource Indian languages. This can allow models to learn rich acoustic representations without requiring extensive labeled datasets. Developing accent-invariant and speaker-invariant architectures can help mitigate variability across regions and individual speakers. The creation of unified multilingual transformer models tailored specifically for Indian languages may allow shared learning across linguistically related groups. There is also growing need for code-mixed LID systems capable of detecting language switches within a single utterance. For practical deployment, researchers must also focus on building real-time, lightweight, and edge-friendly LID models, making the technology accessible for mobile and embedded applications. Finally, expanding annotated speech corpora covering all major and minor Indian languages will be essential to ensure robust evaluation and accelerate research progress.

V. COMPARATIVE SUMMARY OF MAJOR WORKS

The comparative summary in table 1 highlights how major studies differ in their approaches, datasets, and language coverage.

Table1: Comparison of Key Approaches in Indian Language Identification

ef No.	Authors & Year	Method / Model	Dataset(s)	Key Features Used	Performance / Key Findings
1	Harinadh et al., 2025	ML: SVM, KNN, Random Forest	Tamil, Hindi, Marathi	N-gram text vectors	SVM achieved 95% accuracy; integrated Google Translator; data augmentation improved robustness
2	Dar & Pushparaj, 2025	Bi-directional LSTM	Kashmiri spoken numeral dataset (7,200 samples)	Mel-spectrogram	Best accuracy 85.28% with Mel-spectrogram; optimized frame/overlap sizes and Bi-LSTM layers
3	Kulkarni et al., 2022	DNN, RNN-LSTM, GMM	IndicTTS (13 langs), Open-source Multi-speaker (7 langs)	Resemblyzer embeddings	Evaluated 1.5s & 5s features; compared DNN, LSTM, GMM performances
4	Arla et al., 2020	CNN	4 Indian languages (Bengali, Gujarati, Tamil, Telugu)	MFCC spectrogram	Accuracy 88.82%; analyzed train/test duration effect
5	Dey et al., 2022	Review	Multiple Indian SLID datasets	MFCC, spectrogram, i-vector, DNN architectures	Highlights low-resource challenges, dataset descriptions, and future research directions
6	Julius et al., 2024	DNN	Hindi, Bengali, Tamil, English, Gujarati	MFCC	Deep architecture with batch norm; improved language recognition
7	Tomar et al., 2025	CNN	Hindi, Tamil, Malayalam	Spectrogram	Accuracy 98.9% for short-duration audio; robust to minimal input
8	Vuddagiri et al., 2018	i-vector, DNN, DNN with attention	IIITH-ILSC (23 Indian languages)	MFCC	DNN-WA best EER 15.18%; attention improves performance
9	Ingle & Mishra, 2025	ML + Transformer	250K sentences, 23 languages	Text embeddings	Outperformed pre-trained transformer models for LID; released dataset & code

10	Kumar et al., 2024	DNN with u-vector & WSSL	12 Indian languages, 10h each	wav2vec, data2vec, ccc-wav2vec	Efficient LID on real-world data; evaluated feature representations
11	Bam et al., 2022	Review: features + classifiers	Multiple datasets	Pitch, audio features	Survey on ML/DL for SLID; future research directions suggested
12	Lopez-Moreno et al., 2014	DNN	Google 5M LID corpus, NIST LRE 2009	Short-term acoustic features	Relative improvements 70% in Cavg vs i-vector baselines
13	Zissman, 1996	GMM, PRLM, parallel PRLM, PPR	Oregon Graduate Institute Multi-Language Telephone Corpus	Phone recognition	Best: parallel PRLM; error 2% (45s, 2 langs), 21% (10s, 11 langs)
14	Alemu et al., 2024	CNN with data augmentation	Ethio-Semitic languages: Amharic, Geez, Guragigna, Tigrigna	MFCC, mel-spectrogram, hybrid	Hybrid features + augmentation achieved 97–99.5% accuracy
15	Heracleous et al., 2018	DNN, CNN, i-vector	NIST 2015 i-vector Challenge (50 languages)	Acoustic embeddings	DNN EER 3.55%, CNN 3.48%, fused 3.3%; superior to SVM baseline
16	Aarti & Kopparapu, 2018	Review	Indian speech databases	MFCC, PLP	Comprehensive survey of Indian SLID; emphasized language-specific properties
17	Draghici et al., 2020	CNN, Convolutional RNN	Modified set of languages	Mel-spectrogram	Demonstrated ability to learn language-specific patterns effectively
18	Bartz et al., 2017	CRNN	Provided audio snippets	Spectrogram images	Robust to noise; can extend to unknown languages; accurate classification
19	Kotsakis et al., 2020	Supervised & unsupervised ML	Multilingual broadcast audio	Hierarchical discrimination	Proposed late integration for semi-automatic annotation; improved recognition scores
20	Albadr et al., 2019	Optimized GA + ELM	Standard LID datasets	MFCC, SDC, GMM, i-vector	Highest accuracies: 99.50–100% with optimized initial weights
21	Garain et al., 2021	FuzzyGCP (DDMLP + DCNN + SSGAN + Choquet ensemble)	4 benchmark datasets (2 Indic + 2 foreign)	Deep embeddings	F1-score: 98% (MaSS), 67% (VoxForge); ensemble improved precision and generalization

22	Biswas et al., 2023	ANN with MFCC-based time series + FRESH feature selection	IIT-M IndicTTS (6 langs), IIIT-H Indic (7 langs), VoxForge (8 langs)	MFCC time-series, aggregated macro features	Accuracy: 99.93%, 99.94%, 98.43%; robust to real-world noise

VI. ANALYSIS TRENDS

Recent research in Spoken Language Identification (SLID) has revealed several significant trends that are shaping the development of more accurate and robust systems as shown in figure 2.

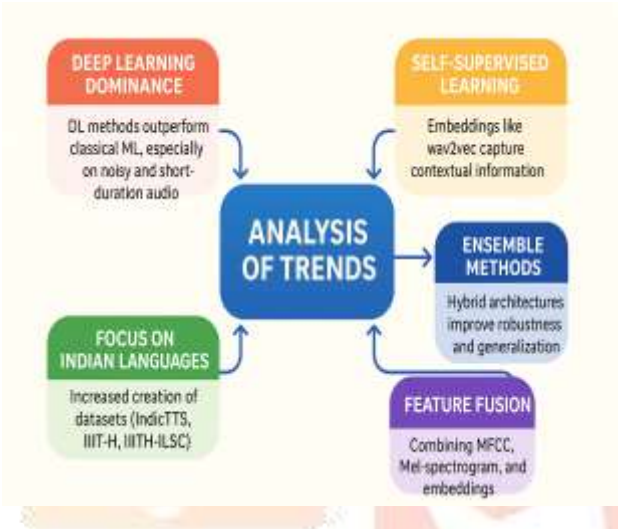

Figure 2: Analysis of Trends

One of the most notable trends is the **dominance of deep learning (DL) methods**. Compared to traditional machine learning approaches, DL models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) with LSTM units, and hybrid architectures consistently achieve higher accuracy, particularly on **noisy and short-duration audio samples**. Their ability to automatically learn hierarchical feature representations allows them to capture both spectral and temporal characteristics of speech, which are crucial for distinguishing between closely related languages [7,21].

Another important trend is the adoption of **self-supervised learning approaches**, such as wav2vec and data2vec. These models learn contextual embeddings directly from raw speech, thereby reducing the dependence on large labelled datasets, which are often scarce for low-resource languages. Self-supervised embeddings have proven highly effective in capturing nuanced phonetic and prosodic patterns, making them particularly suitable for Indian language SLID tasks [10].

Ensemble methods are also popular recent years. By combining multiple classifiers or integrating different architectures, hybrid ensemble approaches enhance system robustness and generalization. Techniques such as bagging, boosting, stacking, and Choquet integral fusion allow SLID systems to leverage complementary strengths of various models, improving performance across diverse recording conditions and speaker variations [21].

There has been a growing **focus on Indian languages**, driven by the creation of large-scale datasets such as IndicTTS, IIIT-H, and IIITH-ILSC. These resources facilitate research in low-resource language identification, enabling the development of models that can handle the rich linguistic diversity of India, including multiple dialects, accents, and phonetic variations [3,8,22].

Finally, a prominent trend in feature engineering is **feature fusion**, where multiple types of acoustic representations are combined to improve classification accuracy. Approaches that integrate MFCCs, Mel-spectrograms, and self-supervised embeddings allow models to exploit both traditional handcrafted features and learned contextual representations. This combination has been shown to significantly enhance performance, especially in challenging real-world conditions [14,22].

Overall, the current trends in SLID research reflect a shift towards **deep learning and self-supervised models**, with greater emphasis on Indian languages, hybrid ensembles, and multi-feature representations. These developments are enabling more reliable, accurate, and scalable SLID systems, even under noisy and low-resource scenarios.

VII. PROPOSED WORK

Inspired by Kulkarni et al. [3], we propose deep learning framework for robust Indian SLID. The proposed work is conceptually based on the framework presented by Kulkarni et al. [3], who developed a Spoken Language Identification (LID) system for native Indian languages using deep learning-oriented experimentation on multiple speech datasets as shown in figure 3. In their study, the authors systematically processed multilingual Indian speech corpora, segmented audio into different duration levels, and evaluated performance variations across datasets and temporal conditions. Their work demonstrated that language-identification accuracy is influenced not only by the model architecture but also by the characteristics of the speech corpus, language diversity, and utterance duration.

Inspired by this methodology, the present study focuses on developing an LID system using **IndicTTS Indian speech database**, ensuring uniformity in data characteristics and simplifying experimentation. The proposed work follows the essential steps adopted by Kulkarni et al., including preprocessing of speech data, segmentation into multiple time-based units, building an experimental environment, training LID systems, and comparing performance across segment durations. This design enables a controlled analysis of how different processing and learning strategies affect identification capability within the selected dataset.

Unlike Kulkarni et al., who compared multiple modelling techniques and feature approaches, the current work intentionally retains flexibility by not restricting the system to any particular feature representation or classifier. This allows the exploration of alternative, potentially more efficient methods tailored to the selected dataset. The overall aim is to examine how varying processing and learning configurations influence LID performance and to derive a streamlined approach that can be optimized for real-world multilingual Indian speech scenarios.

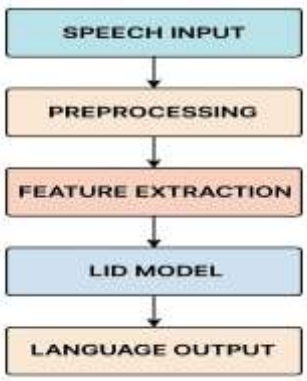

Figure 3: Framework of proposed system

VIII. CONCLUSION

Spoken Language Identification (SLID) for Indian languages continues to be a challenging research domain due to the country's vast linguistic diversity, scarcity of annotated data for low-resource languages and the difficulty of handling short, noisy, and code-mixed speech. Although both classical machine learning and deep learning techniques have demonstrated strong performance, existing systems still face limitations in robustness, scalability, and real-world adaptability.

Recent advancements indicate that approaches integrating multiple feature representations, attention mechanisms, and ensemble learning offer significant potential for improving recognition accuracy across

diverse Indian languages. These techniques enhance model sensitivity to important speech cues and provide better generalization across varied acoustic environments.

The proposed work aims to move toward a more flexible and scalable SLID framework capable of handling multilingual conditions, shorter utterances, and real-time requirements. Future research must focus on expanding real-world datasets, developing reliable code-mixed language identification techniques, and leveraging self-supervised learning to support low-resource Indian languages. Strengthening these areas will contribute to building practical, high-performance SLID systems suitable for widespread deployment across India's multilingual landscape.

IX. REFERENCE

[1] Harinadh, T., Medukonduri, A.V., Chowdary, A.V., et al., 2025. Enhancing Cross-Language Understanding: A Machine Learning-Based Approach to Multilingual Identification.

[2] Dar, M.A. and Pushparaj, J., 2025. Bi-directional LSTM-based isolated spoken word recognition for Kashmiri language utilizing Mel-spectrogram feature. *Applied Acoustics*, 231, p.110505.

[3] Kulkarni, R., Joshi, A., Kamble, M., Apte, S., 2022. Spoken language identification for native Indian languages using deep learning techniques. *Machine Learning and Autonomous Systems: Proceedings of ICMLAS 2021*, pp.75-97.

[4] Arla, L.R., Bonthu, S., Dayal, A., 2020. Multiclass spoken language identification for Indian Languages using deep learning. *IEEE Bombay Section Signature Conference (IBSSC)*, pp.42-45.

[5] Dey, S., Sahidullah, M., Saha, G., 2022. An overview of Indian spoken language recognition from machine learning perspective. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 21(6), pp.1-45.

[6] Julius, C.A., Vijayalakshmi, S., Palathara, T.S., 2024. Spoken Language Identification using Deep Learning. *2024 ICEECT*, Vol.1, pp.1-7.

[7] Tomar, H., Deshwal, D., Trivedi, N., 2025. Convolutional neural network based language identification system: A spectrogram based approach. *Multimedia Tools and Applications*, 84(24), pp.28951-28976.

[8] Vuddagiri, R.K., et al., 2018. IIITH-ILSC Speech Database for Indian Language Identification. *SLTU*, pp.56-60.

[9] Ingle, Y., Mishra, P., 2025. Ilid: Native script language identification for indian languages. *arXiv preprint arXiv:2507.11832*.

[10] Kumar, S., et al., 2024. Spoken Languages Identification for Indian Languages in Real World Condition. *IEEE Conference on Engineering Informatics*, pp.1-7.

[11] Bam, P., Degadwala, S., Upadhyay, R., Vyas, D., 2022. Spoken language recognization based on features and classification methods: A review. *ICAIS*, pp.868-873.

[12] Lopez-Moreno, I., et al., 2014. Automatic language identification using deep neural networks. *ICASSP*, pp.5337-5341.

[13] Zissman, M.A., 1996. Comparison of four approaches to automatic language identification of telephone speech. *IEEE Transactions on speech and audio processing*, 4(1), pp.31-39.

[14] Alemu, A.A., Melese, M.D., Salau, A.O., 2024. Ethio-Semitic language identification using CNN with data augmentation. *Multimedia Tools and Applications*, 83(12), pp.34499-34514.

[15] Heracleous, P., et al., 2018. Comparative study on spoken language identification based on deep learning. *EUSIPCO*, pp.2265-2269.

[16] Aarti, B., Kopparapu, S.K., 2018. Spoken Indian language identification: a review of features and databases. *Sādhanā*, 43(4), p.53.

[17] Draghici, A., Abeßer, J., Lukashevich, H., 2020. A study on spoken language identification using deep neural networks. *International Audio Mostly Conference*, pp.253-256.

[18] Bartz, C., Herold, T., Yang, H., Meinel, C., 2017. Language identification using deep convolutional recurrent neural networks. *ICNIP*, pp.880-889.

[19] Kotsakis, R., et al., 2020. Investigation of Spoken-Language detection and classification in broadcasted audio content. *Information*, 11(4), p.211.

[20] Albadr, M.A.A., Tiun, S., Ayob, M., AL-Dhief, F.T., 2019. Spoken language identification based on optimised genetic algorithm–extreme learning machine approach. *International Journal of Speech Technology*, 22(3), pp.711-727.

[21] Garain, A., Singh, P.K., Sarkar, R., 2021. FuzzyGCP: A deep learning architecture for automatic spoken language identification from speech signals. *Expert Systems with Applications*, 168, p.114416.

[22] Biswas, M., Rahaman, S., Ahmadian, A., Subari, K., Singh, P.K., 2023. Automatic spoken language identification using MFCC based time series features. *Multimedia Tools and Applications*, 82(7), pp.9565-9595.