Findings in Tamil Dialect Speech Recognition and Classification

B. Bharathi[1], **Bharathi Raja Chakravarthi**[2],
Shunmuga Priya Muthusamy Chinnan[2], **S. Saranya**[3], **S. Suhasini**[4]
[1]Sri Sivasubramaniya Nadar College of Engineering, Tamil Nadu, India
[2]Unit for Inclusive AI, Data Science Institute, University of Galway, Ireland
[3] St. Joseph's Institute of Technology, Tamil Nadu, India
[4]Saveetha Engineering College, Tamil Nadu India
bharathib@ssn.edu.in, bharathi.raja@universityofgalway.ie

Abstract

As part of DravidianLangTech-2026, we provide a overview of Shared Task on Dialect-based Speech Recognition and Classification in Tamil. Creating reliable system for Tamil dialect identification from audio signals and dialect-aware Automatic Speech Recognition (ASR) is the main goal of the joint work. Dialect-based Tamil Speech Recognition and Tamil Dialect Classification from Speech are the two subtasks that make up the task. 5,134 audio recordings in four Tamil dialects: Southern, Northern, Western, and Central-spanning 9 hours and 22 minutes make up the training dataset. There are 579 audio samples in the test set, totaling almost two hours in length. The shared task involved 17 teams in total. For speech recognition and dialect classification, the top-performing system obtained a Word Error Rate (WER) of 0.51 and a macro F1-score of 0.79, respectively. The findings emphasize the difficulties in understanding Tamil speech due to dialectal diversity and set solid foundations for further study on low-resource dialect-aware ASR systems.

1 Introduction

Dialect-aware Tamil Speech Recognition and Dialect Classification, a shared task initially established to evaluate intra-language dialect modeling in a low-resource environment, is presented in this work. In order to close a significant gap in dialect-aware speech processing, the work attempts to methodically assess the robustness of automated speech recognition (ASR) systems and dialect identification approaches across four important Tamil dialects.

Over 75 million people worldwide speak Tamil, a traditional Dravidian language that makes an interesting case study for dialect-aware speech processing. Despite having a standardized literary form, Tamil's spoken dialects which include Southern, Northern, Western, and Central dialect groups differ significantly depending on the locale. These varieties of speech differ in lexical substitution patterns, prosodic contours, vowel length realization, and consonant articulation. In both ASR and dialect identification tasks, this variance raises inter-dialect confusability and generates auditory ambiguity.

Tamil is characterized by a significant amount of phonetic, lexical and prosodic variation between the geographical areas. These differences have a considerable impact on the speech recognition systems and dialect identification systems (Nanmalar et al., 2019). Although there are standard Tamil ASR systems, there are few dialect-sensitive benchmarks and evaluation platforms.

The past decades have seen considerable improvement in the performance of ASR in different languages due to the development of deep neural networks and self-supervised learning of speech representation (Radford et al., 2023). The advancements have made it possible to have the large- scale multilingual models that can cope with the different acoustical conditions (Ardila et al., 2020). Most current systems however just consider languages as acoustically homogenous objects, ignoring systematic intra-language variation like regional dialects. That the dialectal variation can significantly impair the ASR accuracy, particularly when the training data is imbalanced or has no phonetic variation in the region.

The development of deep learning and massive pre-trained speech models has allowed the advancement of Automatic Speech Recognition (ASR) systems (Baevski et al., 2020). Nonetheless, dialectal diversity has been one of the greatest challenges especially to the morphologically rich and low-resource languages like Tamil language(Nanmalar et al., 2022).

Although the core ASR activities have been extensively studied in Indian languages including Tamil, dialect level benchmarking is a mostly

Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 71–79

under-researched area (Srivastava et al., 2018). Available datasets normally address standard or broadcast speech and are not usually dialect annotated. Moreover, speech-based dialect identification in a monolingual environment is also an unsolved task.

This shared task solves these shortcomings by providing a curated benchmark data set of four Tamil dialects and assessing two complementary tasks: dialect-aware speech recognition and dialect classification. The task will contribute to improving the research in dialect-robust ASR and intra-language speech modeling by offering standardized metrics of evaluation and the competitive leaderboard environment. The task's objectives are:

- Promote dialect-aware Tamil ASR research
- Present a standardized dataset for assessment.
- Develop both baseline and cutting-edge performance standards.
- Encourage the study of dialect modeling with limited resources.

2 Related Work

Automatic Speech Recognition (ASR) for Tamil has evolved from the Hidden Markov Model (HMM)-Gaussian Mixture Model (GMM) frameworks to current end-to-end neural networks. Early systems used acoustic characteristics which were handcrafted like the MFCCs with statistical language models. The advent of deep learning saw the development of recurrent neural networks and attention-based encoder-decoder models that achieved much higher levels of recognition accuracy especially on morphologically rich languages.

Latest developments in self-supervised learning have also changed the speech processing. Representation learning using raw audio is made possible through such models as Wav2Vec 2.0 (Baevski et al., 2020; Hsu et al., 2021; Chen et al., 2022), which use large-scale unlabeled audio data to train. The pretrained models have been particularly useful in the low resource languages such as Tamil as they do not rely on large labeled corpora. Most more recently, Whisper, a large multilingual encoder decoder model, had been trained with weak supervision, showing good zero-shot and fine-tuning results on a wide range of languages, and among different dialects within identical languages.

Identifying dialect in the speech is even more difficult than regular language identification because the phonetic, prosodic and lexical differences in the same language are sometimes very slight. Previously used methods were based on statistical classifiers and manual crafted acoustic features, but newer techniques use deep neural networks and speaker embedding models as ECAPA-TDNN do (Desplanques et al., 2020). Both text-based and acoustic-based approaches have been examined and transformer-based encoders are always more effective than shallow approaches in the ability to capture dialect-specific features.

Dialect-rich languages Low-resource speech modeling is a developing field of study, especially when modeling low-resource languages. Multilingual pretraining and transfer learning have demonstrated a high possibility of enhancing generalization between dialects through the use of common acoustic representations. More computationally efficient fine-tuning models like LoRA(Hu et al., 2022) also impose lower demands on computational resources and achieve good performance, which is why they are also considered in dialect-aware adaptation.

According to (Srivastava et al., 2018), the development of ASR systems in Indian languages in limited data conditions is difficult. Likewise, (Ismail, 2020) surveyed language and dialect identification systems by focusing on the problem of acoustic similarity and the issue of class imbalance.

Studies that have been conducted on multilingual and transfer learning strategies have revealed major advancements in the low-resource context. (Yadav and Sitaram, 2022) conducted a review of multilingual ASR models and have shown that self-supervised pretrained models are effective. (Madhavaraj and Ramakrishnan, 2019) also demonstrated that data pooling and multi-task learning are more effective when it comes to the performance with a variety of low-resource languages.

The creation of the powerful ASR systems has been enabled by large multilingual corpora like Common Voice (Ardila et al., 2020). In the case of Tamil in particular, a model of ASR, end-to-end, was proposed by (R et al., 2023; Changrampadi et al., 2022) suggested a multilingual Indian speech corpus, which can be used in future studies.

By combining speech recognition with higher-level semantic processing, these approaches highlight the importance of robust acoustic modeling for dialectal speech in real-world scenarios. How-

ever, existing real-time systems primarily focus on end-to-end application development rather than systematic benchmarking across multiple dialects (Saranya et al., 2025).

Although recent methods of self-supervised and multilingual ASR models have been developed, the speech processing in Tamil dialect is not well explored. The current systems are mostly based on standard Tamil and scanty consideration on systematic benchmarking among dialects has been done. The dialect classification and ASR interaction are not thoroughly considered in unified conditions and publicly available dialect-specific datasets are still rare.

To reduce such gaps, the given shared task offers a common standard of dialect-conscious ASR and speech-based dialect recognition in Tamil that would allow systematic comparative analysis of the modeling approaches under equalized assessment conditions.

3 Task Description

The two complementing subtasks of the shared task, which focuses on dialect-aware speech processing for Tamil, are intended to assess transcription robustness and dialect discrimination capability. In a low-resource environment, the activities are designed to promote the creation of models that can manage intra-language variance. The speech-based dialect classification is significantly more difficult than a Text-based dialect identification, because the models have to account for the subtle variations in the acoustic, phonetic and prosodic aspects directly from the speech signal. Speech-level dialect recognition presents another challenge besides transcribed text, as it involves the recognition of variations in speaker, pronunciation, and recording context.

Input: Participants receive the following: 16 kHz sampling rate, variable-duration utterances, raw speech audio files in WAV format, and dialect labels that are exclusive to the training data.

3.1 Subtask 1: Dialect-Aware Tamil Speech Classification

Automatic dialect classification from speech signals is the main objective of Subtask 1. Systems are required to categorize a Tamil audio recording into one of four predetermined dialect groups. The dataset includes four major dialect groups:

- Southern_Dialect
- Northern_Dialect
- Western _Dialect
- Central_Dialect

Classification of these dialects is a challenging auditory discriminating challenge due to their systematic phonetic and prosodic variance.

Evaluation Metrics Systems are evaluated using:

- Macro F1-score is the thing we look at

We picked Macro F1 because it is fair to all dialect classes. This is especially important when some classes have a lot things in them than others. Macro F1 makes sure that each dialect class is treated the same so the classes with a lot of things in them do not get all the attention. This way Evaluation Metrics like Macro F1 are really good, for systems and dialect classes.

3.2 Subtask 2: Dialect-Aware Tamil Speech Recognition

Subtask 2 is to predict the corresponding transcription from a Tamil audio recording. In contrast to traditional ASR benchmarks, which presume standardized speech, this task specifically assesses recognition ability across a variety of dialects.

The test data does not include transcriptions.

Evaluation Metrics Performance is evaluated using: Word Error Rate (WER) as the primary metric WER = (Substitutions + Insertions + Deletions) / Total Reference Words

Word Error Rate is the number of mistakes, which are substitutions, which is when we use the wrong word, insertions, which is when we add an extra word, deletions, which is when we leave out a word. The purpose of the two subtasks is to compliment each other: The robustness of transcription under dialect variation is assessed in Subtask 1. The explicit modeling of dialect identity is assessed in Subtask 2.

Figure 1 presents the classification of the dialects and sample sentences considered in this task. They offer a thorough evaluation of dialect-aware speech modeling when combined. Although dialect cues may be implicitly learned by ASR systems, discriminative modeling of fine-grained audio variations is necessary for dialect classification. Deeper understanding of dialect confusability and representation learning is made possible by joint analysis of the two tasks.

S.No.	Name of the Dialect	Region	Example sentence
1	Northern_Dialect	Madras, Chengalpet and North Arcot Districts	சாப்டுறதுக்கு எதனா வாங்கின்னு வா. (Saapdrathukku edhana vaanginnu vaa.)
2	Southern_Dialect	Madurai, Ramnad, Tirunelveli and Kanyakumari	ஏல எப்பல வந்த? (yela eppala vantha)
3	Western_Dialect	Salem, Dharmapuri, Coimbatore and Nilgiris Districts	இன்னைக்கு என்னங்க ஒரே உப்பசமா இருக்குதுங்க. (Innaikku ennanga ore uppasama irukkudunga.)
4	Central_Dialect	Tiruchirappalli, Thanjavur and South Arcot Districts	ரெண்டு பேரும் அடிச்சுக்காதீங்க செத்த செவனேனு இருங்க (Rendu perum adichukkadheenga seththa sevanenu irunga.)

Figure 1: Dialect classifications with sample sentences

4 Dataset Description

Table 1: Training Data Statistics

Dialect	Audio_Files	Duration
Southern_Dialect	1427	2:44:30
Northern_Dialect	1696	3:29:15
Western_Dialect	1126	1:59:59
Central_Dialect	885	1:08:18
Total	5134	9:22:02

The training dataset consists of 5,134 audio recordings spanning 9 hours and 22 minutes across four Tamil dialects[(Bharathi et al., 2025).

Table 2: Test Data Statistics

Dialect	Audio_Files	Duration
Test Data	579	02:04:05

All audio files were sampled at 16 kHz and provided in WAV format. Table 1 and 2 summarizes the dataset distribution across dialects.

5 Participation

In the Dialect based Tamil speech recognition and classification shared task, 17 teams from academic institutions participated. The increasing interest in dialect-aware speech processing and low-resource ASR benchmarking is reflected in the participants' variety. Teams that took part used a variety of modeling techniques, including lightweight classification frameworks, hybrid acoustic models, and pretrained transformer-based architectures.

The competition followed a common evaluation procedure, with training set and test set of 579 audio examples (about 2 hours of speech). The final system submissions were tested on the test set without lable to avoid any kind of overfitting.

5.1 Modeling Approaches

Participants explored several modeling paradigms:

The majority of the best-performing systems utilized large pretrained models like Whisper encoder-decoder models or wav2vec/XLSR fine-tuning pipelines. These models had the following advantages:

- Self-supervised pretraining on large multilingual datasets
- Strong contextualized speech representations
- Fine-tuning on dialect-specific Tamil datasets

Transformer models had better generalization capabilities, especially in Subtask 1 (ASR).

Team CHMOD_777 fine-tuned Wav2Vec 2.0 Base for dialect classification directly from speech.

For ASR, they fine-tuned Whisper Small. Their approach relies on supervised adaptation of pretrained multilingual speech encoders.

Team TamilVoiceLab used Whisper Base for ASR and extracted embeddings from Wav2Vec 2.0 for dialect classification, followed by a softmax classification layer. Their system focuses on transfer learning under limited dialectal data.

Team Dialectmind experimented with Whisper Small and Whisper Medium models for ASR. For dialect identification, they trained a text classifier on ASR-generated transcripts using transformer-based encoders.

Team CUET_InferX fine-tuned Whisper Small for transcription and used MuRIL (Multilingual BERT for Indian Languages) for dialect classification from text outputs. Their pipeline separates acoustic and linguistic modeling.

Team Violet_Vortex used Whisper Base for ASR and applied an LSTM-based classifier over acoustic embeddings for dialect classification. Their approach evaluates sequential neural models for dialect discrimination.

Team AITamilDialect adopted a multi-task learning framework built on Wav2Vec 2.0, jointly optimizing ASR and dialect classification objectives with task-specific output heads. **Team HAG Signals** fine-tuned Whisper Small for ASR and experimented with beam search decoding strategies. For dialect classification, they used a fully connected classifier on speech embeddings.

Team DLRG employed parameter-efficient fine-tuning (LoRA) on Whisper Base for ASR. For dialect classification, they used Wav2Vec 2.0 embeddings with a linear classifier.

Team IITK_SpeechScape used Conformer-based ASR architecture initialized with pretrained acoustic representations. For dialect classification, they fine-tuned WavLM.

Team Sync2 focused exclusively on dialect classification using ECAPA-TDNN speaker embedding architecture, adapted to capture dialectal acoustic variations.

Team CODERS implemented a unified framework using Wav2Vec 2.0 as a shared encoder backbone, with separate heads for ASR and dialect classification.

Team Wave2Word used Whisper Small for ASR and a transformer-based classifier for dialect prediction from transcriptions. In their second submission, Wave2Word experimented with WavLM for dialect classification and compared performance against their Whisper-based pipeline.

Team Wise adopted a lightweight dialect classification model built on Wav2Vec 2.0 Base, followed by a fully connected neural classifier.

Team SpokenRoots used HuBERT Base for acoustic feature extraction and trained a supervised classifier for dialect identification.

Team GigitAI employed an ECAPA-TDNN embedding model, originally designed for speaker recognition, adapted for dialect classification tasks.

Team Azrael developed a computationally efficient dialect classifier based on Wav2Vec 2.0 Small, optimized for low-resource deployment environments.

The 17 participating teams' architectures and training methods are compiled in Table 3. We note that the most popular backbone models were Whisper and Wav2Vec 2.0, whereas only a small number of teams investigated multi-task learning and parameter-efficient fine-tuning methods.

6 Results Overview

The official results of the evaluation of both subtasks of the shared task are presented (i) Dialect Classification and (ii) Dialect-aware Automatic Speech Recognition (ASR). There were 17 teams in the competition that completed at least one subtask and handed systems that were rated on a held-out test set. The evaluation of Subtask 1 was conducted with Macro F1-score to consider the imbalance in classes among the dialects, The evaluation for Subtask 1 was based on macro F1-score since there was no uniform distribution of the dialect classes. Unlike Micro-F1, Macro F1-score measures the performance separately for each class and then averages the results, thus assigning equal weights to all dialect categories irrespective of their size, whereas in Subtask 2 the evaluation of the performance was by Word Error Rate (WER) where smaller numbers showed better performance.

In Subtask 1, Wise scored the highest Macro F1-score of 0.79, a score that is far ahead of the one that second-ranked system scored. The performance distribution indicates that there is a significant difference between the highest performance system and the rest of the submissions, which implies a difference in the ability to model dialectal acoustic variation.

Table 3: Overview of modeling approaches used by the 17 participating teams.

Team	Whisper Small	Whisper Base	Whisper Medium	Wav2Vec 2.0	WavLM	HuBERT	ECAPA-TDNN	MuRIL	LoRA	Multi-task	Data Aug.
CHMOD_777(Karunanidhi and Arumugam, 2026)	✓			✓							✓
TamilVoiceLab(Priya and Bharathi, 2026)		✓		✓							
Dialectmind(Gayathri and Bharathi, 2026)	✓		✓					✓			
CUET_InferX(Semon et al., 2026)	✓							✓			
Violet_Vortex		✓		✓							✓
AITamilDialect(Varalakshmi and Bharathi, 2026)				✓						✓	
HAG_Signals	✓			✓							✓
DLRG_1(Abhinav et al., 2026)		✓		✓					✓		
IITK_SpeechScape(Sekar et al., 2026)					✓						
Sync2							✓				
CODERS				✓						✓	
Wave2Word_1 (Naswan and Ahmad, 2026)	✓										
Wave2Word_2(Naswan and Ahmad, 2026)					✓						
Wise(Ganesh Sundhar et al., 2026)				✓							
SpokenRoots						✓					
GigitAI							✓				
Azrael(J et al., 2026)				✓							

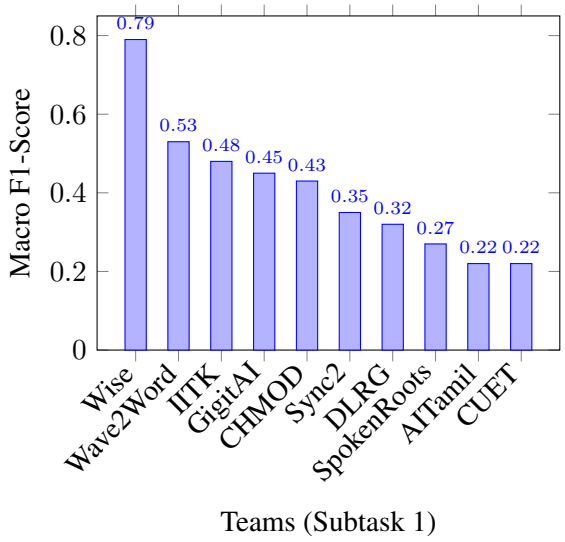

Figure 2: Performance trend of top teams in Subtask 1 (Dialect Classification).

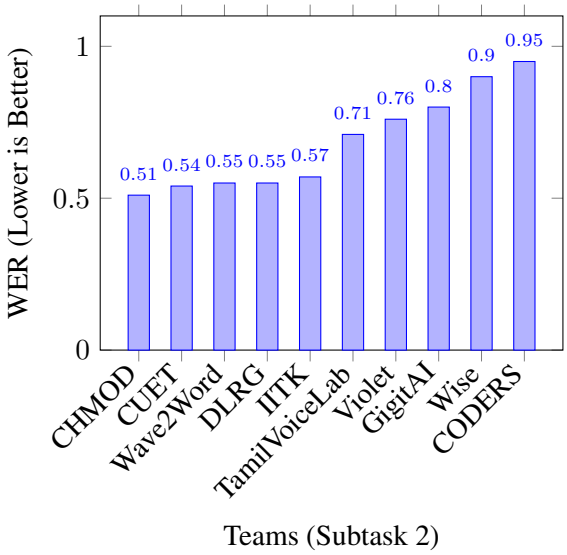

Figure 3: Performance distribution in Subtask 2 (Dialect-aware ASR).

In the case of Subtask 2, CHMOD_777 had the lowest WER of 0.51 with CUET_InferX (0.54) and Wave2Word (0.55) being close by. The most successful systems were mainly based on highly tuned systems using transformer-based ASR models, specifically Whisper models. Nevertheless, some submissions obtained a WER value exceeding 0.90, which means that there are still significant difficulties with speech recognition with dialect awareness in low-resource.

In Subtask 1, Wise obtained the highest Macro F1-score, as indicated in Table 4.

The official WER ranking for Subtask 2 is shown in Table 5.

Figures 2 and 3 present the performance trends for both subtasks. The study of the model architectures shows that there is a definite correlation between the pretrained transformer-based speech models and the increased performance. Whisper variants based systems were always dominant in Subtask 2 and CHMOD_777 was the lowest WER

Table 4: Official ranking of Subtask 1 (Dialect Classification) on the test set.

Rank	Team Name	Run	Precision	Recall	F1_Score
1	Wise	Run2	0.89	0.76	**0.79**
2	Wave2Word	Run1	0.54	0.54	0.53
3	IITK_SpeechScape	Run2	0.54	0.50	0.48
4	GigitAI	Run1	0.46	0.49	0.45
5	CHMOD_777	Run1	0.41	0.47	0.43
6	Sync2	Run3	0.40	0.39	0.35
7	DLRG	Run1	0.33	0.37	0.32
8	SpokenRoots	Run2	0.30	0.28	0.27
9	AITamilDialect	–	0.23	0.22	0.22
9	CUET_InferX	Run2	0.23	0.30	0.22
10	CODERS	Run1	0.21	0.22	0.18
11	Azrael	Run1	0.08	0.25	0.12

Table 5: Official ranking of Subtask 2 (Dialect-aware ASR) based on Word Error Rate (WER) on the test set

Rank	Team Name	WER
1	CHMOD_777_Run1	0.51
2	CUET_InferX_Run1	0.54
3	Wave2Word_Run1	0.55
3	DLRG_Run1	0.55
4	IITK_SpeechScape_Run1	0.57
5	TamilVoiceLab_Run1	0.71
6	VIOLET_VORTEX_Run1	0.76
7	GigitAI_Run3	0.80
8	Wise_Run2	0.90
9	CODERS_Run1	0.95
10	AITamilDialect	1.00
10	HAG_Signals	1.00
10	DialectMind_Run1	1.00

(0.51). Likewise, encoders based on transformers (including Wav2Vec 2.0 and WavLM) were also able to achieve competitive performances in dialect classification.

However, this trend was not observed in systems based only on lightweight or shallow classifiers, and in this case, Macro F1-scores were significantly lower, indicating that dialect discrimination is indeed greatly improved with the help of deep contextual acoustic representations. Surprisingly, the teams which implemented common encoder backbone on both subtasks (e.g. Wave2Word and DLRG) exhibited equal classification and ASR performance, which suggests that a strong acoustic modeling can generalize to similar speech tasks.

Interestingly, the best performing system in Subtask 1 did not appear to perform competitively in Sub task 2 and vice versa. It indicates that the modeling attributes of dialect classification and dialect-aware ASR are different. Discriminative acoustic representations with good capturing of dialect difference is the major advantage of dialect classification, while accurate lexical and phonetic transcription ability is the key to ASR performance. As a result, models learnt from one task may not necessarily perform well on the other task.

In general, the findings re-establish the roles of the large pretrained multilingual speech models in low-resource dialectal settings.

7 Limitations

Although, this common task can be regarded as a good source of insight into dialect-conscious speech recognition and classification on the Tamil language, there are still a number of limitations.

To begin with, the size of the dataset is quite small in comparison with the large-scale benchmarks of ASR. Though several dialects were involved, the quantity of speech data of each dialect might not be large enough to comprehensively reflect the intra-dialect variation, the speaker heterogeneity, and the environmental noise factors.

Second, the task was performed on one held-out test set, which might not necessarily be applicable in the real world of deployment. The spontaneous conversation speech can have dialect variations that may introduce new challenges outside of the edited dataset.

Third, the majority of systems involved in it were heavily dependent on the pretrained multilingual transformer models, including Whisper and Wav2Vec 2.0. These models showed good performance but their computational needs can confine its use in low resource or on-device deployment environments.

Lastly, this combined task concentrated solely

on acoustic as well as transcription-based modelling. Future research may look into multimodal cues, language modeling that is dialect-aware and fairness-aware across demographic boundaries.

Although these are the constraints, the activity forms a brilliant foundation to further studies of dialect finely-sensitive speech technologies in Tamil.

8 Conclusion

This shared task on Dialect-based Speech Recognition and Classification in Tamil highlights both the progress and remaining challenges in dialect-aware speech technologies. In 17 competing teams, transformer-based pretrained models, and specifically Whisper and Wav2Vec 2.0, were the most widespread architectures and they showed impressive generalization in low-resource dialectal settings. Although there is a top-of-class competitive performance, there is a significant difference between the highest-ranked systems and the lowest-ranked submissions, especially in the territory of ASR, wherein various models appeared incapable of achieving WER to less than 0.70. This means that the dialectal variation remains a major challenge to the acoustic modeling and adaptation.

The research directions of the future are the investigation of multi-task learning models, which will optimize transcription and dialect classification, parameter-efficient fine-tuning strategies, and dialect-aware decoding models. The assignment highlights the need to further expand datasets and enhance the representation learning to enhance inclusivity in the underrepresented dialects speech technologies.

Acknowledgments

The authors, Bharathi Raja Chakravarthi was funded by a research grant from Research Ireland under grant number SFI/12/RC/2289_P2 (Insight_2).

References

Gulisetty Abhinav, Tanisha Nanda, Ramesh Kannan R, and Ratnavel Rajalakshmi. 2026. Dlrg@dravidianlangtech 2026: Dual-purpose whisper adaptation for tamil dialect identification and dialectal speech recognition. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Rosana Ardila, Megan Branson, Kelly Davis, Michael Kohler, Josh Meyer, Michael Henretty, Reuben Morais, Lindsay Saunders, Francis Tyers, and Gregor Weber. 2020. Common voice: A massively-multilingual speech corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4218–4222, Marseille, France. European Language Resources Association.

Alexei Baevski, Henry Zhou, Abdelrahman Mohamed, and Michael Auli. 2020. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:12449–12460.

B Bharathi, S Saranya, P Vijayalakshmi, and T Nagarajan. 2025. Multi-dialect speech corpus creation for enhancing tamil automatic speech recognition. *Circuits, Systems, and Signal Processing*, pages 1–19.

Mohamed Hashim Changrampadi, A. Shahina, M. Badri Narayanan, and A. Nayeemulla Khan. 2022. End-to-end speech recognition of Tamil language. *Intelligent Automation & Soft Computing*, 32(2):1309–1323.

Sanyuan Chen, Chengyi Wang, Zhengyang Chen, Yu Wu, Shujie Liu, Zhuo Chen, Jinyu Li, Yao Qian, and Furu Wei. 2022. Wavlm: Large-scale self-supervised pre-training for full stack speech processing. In *Proceedings of Interspeech*, pages 1503–1507.

Brecht Desplanques, Jenthe Thienpondt, and Kris Demuynck. 2020. Ecapa-tdnn: Emphasized channel attention, propagation and aggregation in tdnn based speaker verification. In *Proceedings of Interspeech*, pages 3830–3834.

S Ganesh Sundhar, N Hari Krishnan, G Gnanasabesan, KP Suriya, and G Jyothish Lal. 2026. Wise@dravidianlangtech 2026: Dialect-aware tamil speech classification and recognition via cross-pipeline embedding transfer. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

K Gayathri and B Bharathi. 2026. Dialect-mind@dravidianlangtech 2026: Zero-shot dialectal tamil automatic speech recognition using a large pre-trained conformer model. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, and Abdelrahman Mohamed. 2021. Hubert: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3451–3460.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*.

Tanvira Ismail. 2020. A survey of language and dialect identification systems. *Adalya Journal*, 9(1).

Janish Andrin J, Mohammed Sahil S, and S Saranya. 2026. Azrael@dravidianlangtech 2026: Dialect-sensitive automatic speech recognition and classification for tamil. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Arunaggiri Pandian Karunanidhi and Prabalakshmi Arumugam. 2026. Chmod_777@dravidianlangtech 2026: Tamil-adapted whisper and mms for dialect speech recognition and classification. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

A Madhavaraj and A G Ramakrishnan. 2019. Data-pooling and multi-task learning for enhanced performance of speech recognition systems in multiple low resourced languages. In *2019 National Conference on Communications (NCC)*, pages 1–5.

M. Nanmalar, P. Vijayalakshmi, and T. Nagarajan. 2019. Literary and colloquial dialect identification for Tamil using acoustic features. In *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, pages 1303–1306.

M. Nanmalar, P. Vijayalakshmi, and T. Nagarajan. 2022. Literary and colloquial tamil dialect identification. *Circuits Syst. Signal Process.*, 41(7):4004–4027.

Ruwad Naswan and Shadab Tanjeed Ahmad. 2026. Wave2word@dravidianlangtech 2026: Whistam: A unified framework for dialect based tamil speech recognition and classification. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

S. B. Priya and B. Bharathi. 2026. Tamilvoice-lab@dravidianlangtech 2026: Investigating whisper tamil large-v2 for dialectal tamil speech recognition. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Nithya R, Malavika S, Jordan F, Arjun Gangwar, Metilda N J, S Umesh, Rithik Sarab, Akhilesh Kumar Dubey, Govind Divakaran, Samudra Vijaya K, and Suryakanth V Gangashetty. 2023. Spring-inx: A multilingual indian language speech corpus by spring lab, iit madras. *Preprint*, arXiv:2310.14654.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. *Proceedings of the 40th International Conference on Machine Learning (ICML)*.

S Saranya, B Bharathi, S Gomathy Dhanya, and Aishwarya Krishnakumar. 2025. Real-time continuous tamil dialect speech recognition and summarization. *Circuits, Systems, and Signal Processing*, 44(4):2855–2881.

G Srishtik Sekar, Harissh Ragav Dhamodaran, Kishore Shankar S, Balasubramanian Palani, and R. Tharaniya Sairaj. 2026. Iiitk_speechscape@dravidianlangtech 2026: Dialect based speech recognition and classification using speech foundation models and deep learning techniques. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Md. Ashraful Islam Semon, Jihadul Islam, Ratnajit Dhar, and Hasan Murad. 2026. Cuet_inferx@dravidianlangtech 2026: Shared task on dialect based speech recognition and classification in tamil. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Brij Mohan Lal Srivastava, Sunayana Sitaram, Rupesh Kumar Mehta, Krishna Doss Mohan, Pallavi Matani, Sandeepkumar Satpal, Kalika Bali, Radhakrishnan Srikanth, and Niranjan Nayak. 2018. Interspeech 2018 Low Resource Automatic Speech Recognition Challenge for Indian Languages. In *Proc. 6th Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU 2018)*, pages 11–14.

K Varalakshmi and B Bharathi. 2026. Aitamil-dialect@dravidianlangtech 2026: Zero-shot whisper and wav2vec2 embedding-based tamil speech recognition and dialect classification. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Hemant Yadav and Sunayana Sitaram. 2022. A survey of multilingual models for automatic speech recognition. *Preprint*, arXiv:2202.12576.