

# Findings of the Shared Task on Speech Recognition for Vulnerable Individuals in Tamil

B. Bharathi<sup>1</sup>, Bharathi Raja Chakravarthi<sup>2</sup>,  
Subalalitha Chinnaudayar Navaneethakrishnan<sup>3</sup>, N. Sripriya<sup>1</sup>, Arunaggiri Pandian<sup>4</sup>, Swetha Valli<sup>4</sup>

<sup>1</sup>SSN College of Engineering

<sup>2</sup>National University of Ireland Galway

<sup>3</sup>SRM Institute Of Science And Technology

<sup>4</sup>Thiagarajar College of Engineering

bharathib@ssn.edu.in, bharathiraja.akr@gmail.com

## Abstract

This paper illustrates the overview of the shared task on automatic speech recognition in the Tamil language. In the shared task, spontaneous Tamil speech data gathered from elderly and transgender people was given for recognition and evaluation. These utterances were collected from people when they communicated in the public locations such as hospitals, markets, vegetable shop, etc. The speech corpus includes utterances of male, female, and transgender and was split into training and testing data. The given task was evaluated using WER (Word Error Rate). The participants used the transformer-based model for automatic speech recognition. Different results using different pre-trained transformer models are discussed in this overview paper.

**Keywords:** Automatic Speech Recognition, Word Error Rate, Tamil speech corpus, Transformer model, Pre-trained model.

## 1 Introduction

There have been tremendous developments in smart technologies that continue to evolve and enhance human-machine interaction (Chakravarthi et al., 2020; Sampath et al., 2022; Ravikiran et al., 2022; Chakravarthi et al., 2022; Priyadharshini et al., 2022). One such recent technology is Automatic Speech Recognition (ASR) which has paved the way to a lot of voiced-based interfaces to many automated systems. Many elderly and transgender people are unaware of the technologies available that are facilitated to aid people in public places such as banks, hospitals and administrative offices. Hence, speech is the only medium that could assist them in satisfying their needs (Hämäläinen et al., 2015). However, the usage of these ASR systems by the elderly, transgender and less educated people are limited. The reason is most of the existing automated systems are enabled with voiced-based interfaces that are in English language. Old aged

and people in rural areas only feel comfortable to interact in their regional language. If the systems developed to aid people in public places are enabled with speech interfaces in the regional language, the aiding systems are benefited by all people. The spontaneous speech data in Tamil language is gathered from old-aged and transgender people, who are bereft of using these facilities to their advantage. This task is organized to find an efficient ASR model to handle the elderly people speech corpus. The speech corpus creation is represented in Fig. 1.

The earliest Old Tamil documents are small inscriptions in Adichanallur dating from 905 BC to 696 BC. Tamil has the oldest ancient non-Sanskritic Indian literature of any Indian language. Tamil uses agglutinative grammar, which uses suffixes to indicate noun class, number, case, verb tense, and other grammatical categories. Tamil's standard metalinguistic terminology and scholarly vocabulary is itself Tamil, as opposed to the Sanskrit that is standard for most Aryan languages. Tamil has many forms, in addition to dialects: a classical literary style based on the ancient language (cankattami), a modern literary and formal style (centami), and a current colloquial form (kotuntami) (Sakuntharaj and Mahesan, 2021, 2017, 2016; Thavareesan and Mahesan, 2019, 2020a,b, 2021). These styles blend into one another, creating a stylistic continuity. It is conceivable, for example, to write centami using cankattami vocabulary, or to utilize forms connected with one of the other varieties while speaking kotuntami (Subalalitha, 2019; Srinivasan and Subalalitha, 2019; Narasimhan et al., 2018). Tamil words are made up of a lexical root and one or more affixes. The majority of Tamil affixes are suffixes. Tamil suffixes are either derivational suffixes, which modify the part of speech or meaning of the word, or inflectional suffixes, which designate categories like as person, number, mood, tense, and so on. There is no ultimate limit to the length and

scope of agglutination, which might result in large words with several suffixes, requiring many words or a sentence in English (Anita and Subalalitha, 2019b,a; Subalalitha and Poovammal, 2018).

Initially, an ASR system will extract the features from speech signals. Further, acoustic models will be created with the features. Finally, the language model will be created which captures the linguistic information from the text (Das et al., 2011). The performance of the ASR systems has to be evaluated for it to be used in real time applications. On large scale automatic speech recognition (ASR) tasks, an end-to-end speech recognition system has showed promising performance, making it competitive with traditional hybrid systems. The end-to-end system includes an acoustic model, lexicon, and language model that turns acoustic data into tag labels immediately (Zeng et al., 2021; Pérez-Espinosa et al., 2017). In the field of end-to-end speech recognition, two common frameworks are used. One is distinguished by frame synchronous prediction, which means that one target label is assigned to each input frame (Miao et al., 2020; Xue et al., 2021; Miao et al., 2019; Watanabe et al., 2017). The performance can also be measured in terms of phoneme recognition with different test feature vectors and different model parameters. The use of acoustic models for speech recognition, which are created using the voices of younger adults, may be a significant factor in the recognition of elderly speech (Fukuda et al., 2020; Zeng et al., 2020; Iribe et al., 2015). There are few acoustic models created to carry out the speech recognition task. Some of the acoustic models are Japanese Newspaper Article Sentences (JNAS), Japanese Newspaper Article Sentences Read Speech Corpus of the Aged (S-JNAS) and Corpus of Spontaneous Japanese (CSJ). In the literature, all the acoustic models are compared and found that the CSJ model achieves the lowest WER only after the adaptation of the elderly voices (Fukuda et al., 2020). Similarly, dialect adaptation is also required so as to improve recognition accuracy (Fukuda et al., 2019). Due to recent developments in large vocabulary continuous speech recognition (LVCSR) technologies, speech recognition systems have become widely used in a variety of fields (Xue et al., 2021). Acoustic differences between speakers are thought to be one of the primary causes of the decline in speech recognition rates. For elderly speakers to use speech recognition systems trained



Figure 1: Speech corpus collected from vulnerable individuals in Tamil language

using normal adult speech data, the acoustic difference between the speech of elderly speaker and that of a typical adult should be analyzed and adapted accordingly. Instead, an acoustic model trained on the utterances of elderly speakers can reduce this degradation, as confirmed by a document retrieval system. High recognition accuracy can be obtained for speech reading a written text or similar by using cutting-edge speech recognition technology; however, the accuracy degrades for freely spoken spontaneous speech. The main reason for this issue is that acoustic and linguistic models used in speech recognition have been developed primarily using written language text or read speech. However, spontaneous speech and written language differ significantly both acoustically and linguistically (Zeng et al., 2020). Nowadays, developing ASR systems recognizing elderly people speech data has become more common. Due to the ageing population in modern society and the growth of smart devices, there is a need to improve speech recognition in smart devices so that information can be freely accessible to the elderly as well as the younger people (Kwon et al., 2016; Vacher et al., 2015; Hosain et al., 2017; Teixeira et al., 2014). Due to the impacts of speech articulation and speaking style, speech recognition systems are often optimised for an average adult’s voice and have a lower accuracy rate when recognising an elderly person’s voice. Adapting the currently available speech recognition systems for handling the speech of senior users is certain to incur additional costs (Kwon et al., 2016).

## 2 Task Description

This shared task tackles a difficult problem in Automatic Speech Recognition: vulnerable elderly and

transgender individuals in Tamil. People in their senior years go to primary places such as banks, hospitals, and administrative offices to meet their daily needs. Many elderly persons are unsure of how to use the devices provided to assist them. Similarly, because transgender persons are denied access to primary education as a result of societal discrimination, speech is the only channel via which they may meet their needs. The data on spontaneous speech is collected from elderly and transgender people who are unable to take advantage of these services. For the training set, a speech corpus containing 5.5 hours of transcribed speech will be released, as well as 2 hours of speech data for testing test.

### 3 Related Work

When a model is fine-tuned on many languages at the same time, a single multilingual speech recognition model can be built that can compete with models that are fine-tuned on individual language speech corpus. Speech2Vec expands the text-based Word2Vec model to learn word embeddings directly from speech by combining an RNN Encoder-Decoder framework with skipgrams or cbow for training. Acoustic models are designed at phoneme/syllable level to carry out the speech recognition task. Initially, the acoustic models were created with JNAS, S-JNAS and CSJ speech corpus (Lin and Yu, 2015; Iribe et al., 2015). Later, the models were trained/fine-tuned with different speech corpus. To get a better performance and accuracy, backpropagation using the transfer learning was attempted in the literature. Similar work was performed for other languages like Bengali, Japanese, etc. Also, more speech corpus is collected from the young people for many languages (Zeng et al., 2020; Lee et al., 2021). However, speaker fluctuation, environmental noise, and transmission channel noise all degrade ASR performance. As the shared task is given with a separate training data set, an effective model has to be created during the training. Therefore, hierarchical transformer based model for large context end to end ASR can be used (Masumura et al., 2021). In the recent era, the environment is changing with smart systems and is identified that there is a need for ASR systems that are capable of handling speech of elderly people spoken in their native languages. To overcome this problem, the shared task is proposed for the research community to build an efficient model for recognizing the

speech of elderly people and transgenders in Tamil language.

### 4 Data-set Description

The dataset given to this shared task is an Tamil conversational speech recorded from the elderly people whose average age is around 61 for male, 59 for female and 30 for transgender people which are tabulated in Table 1. A total of 6 hours and 42 minutes is collected from the elderly people. 46 audio files were recorded and each audio file is split into many subsets as transformer model does not support the large audio files. The speech is recorded with a sampling rate of 16KHZ. The audio files from Audio - 1 to Audio - 36 are used for training (duration is approximately 5.5 hours) and Audio - 37 to Audio - 47 are used for testing (duration is approximately 2 hours).

### 5 Methodology

The methodology used by the participants in shared task of speech recognition for vulnerable individuals in Tamil is discussed in this section. Different types of pre-trained transformer models used by the participants in this shared task are

Amrrs/wav2vec2-large-xlsr-53-tamil<sup>1</sup> (Dhanya et al., 2022)

akashsivanandan/wav2vec2-large-xlsr-300m-tamil-colab-final<sup>2</sup> (Dhanya et al., 2022)

nikhil6041/wav2vec2-large-xlsr-tamil-commonvoice<sup>3</sup> (Dhanya et al., 2022)

Rajaram1996/wav2vec-large-xlsr-53-tamil<sup>4</sup> (Suhasini and Bharathi, 2022)

The above mentioned models are fine tuned on facebook/wav2vec-large-xlsr-53<sup>5</sup> pre-trained model using multilingual common voice dataset. To fine-tune the model, they had a classifier representing the downstreams task's output vocabulary on top of it and train it with a Connectionist Temporal Classification (CTC) loss on the labelled data. The models used are based on XLSR wav2vec model, this XLSR model is capable of learning cross-lingual speech data, where the raw speech

<sup>1</sup><https://huggingface.co/Amrrs/wav2vec2-large-xlsr-53-tamil>

<sup>2</sup><https://huggingface.co/akashsivanandan/wav2vec2-large-xls-r-300m-tamil-colab-final>

<sup>3</sup><https://huggingface.co/nikhil6041/wav2vec2-large-xlsr-tamil-commonvoice>

<sup>4</sup><https://huggingface.co/Rajaram1996/wav2vec2-large-xlsr-53-tamil>

<sup>5</sup><https://huggingface.co/facebook/wav2vec2-large-xlsr-53>

Table 1: Age, gender and duration of the utterances in speech corpus

S.No	Filename	Gender	Age	Duration(in secs)
1	Audio - 1	M	72	10
2	Audio - 2	F	61	9
3	Audio - 3	F	71	11
4	Audio - 4	M	68	8
5	Audio - 5	F	59	14
6	Audio - 6	F	67	9
7	Audio - 7	M	54	8
8	Audio - 8	F	65	16
9	Audio - 9	F	55	3
10	Audio - 10	M	60	13
11	Audio - 11	F	55	17
12	Audio - 12	F	52	6
13	Audio - 13	F	53	11
14	Audio - 14	F	61	9
15	Audio - 15	F	54	1
16	Audio - 16	F	56	6
17	Audio - 17	F	52	12
18	Audio - 18	F	54	6
19	Audio - 19	F	52	8
20	Audio - 20	F	52	9
21	Audio - 21	F	62	13
22	Audio - 22	F	52	12
23	Audio - 23	F	62	13
24	Audio - 24	F	53	4
25	Audio - 25	F	65	3
26	Audio - 26	F	64	8
27	Audio - 27	F	54	6
28	Audio - 28	M	62	8
29	Audio - 29	M	54	16
30	Audio - 30	F	76	9
31	Audio - 31	F	55	9
32	Audio - 32	M	50	6
33	Audio - 33	F	63	6
34	Audio - 34	M	84	6
35	Audio - 35	F	70	6
36	Audio - 36	F	50	6
37	Audio - 37	M	53	6
38	Audio - 38	F	55	6
39	Audio - 39	M	62	6
40	Audio - 40	T	24	6
41	Audio - 41	T	22	7
42	Audio - 42	T	40	8
43	Audio - 43	T	25	11
44	Audio - 44	T	29	10
45	Audio - 45	T	35	9
46	Audio - 46	T	33	16

S. No	Model Name	WER (in %)
1	SSN_NLP Submission 1 (Dhanya et al., 2022)	39.4512
2	SUH_ASR (Sahasini and Bharathi, 2022)	39.6487
3	SSN_NLP Submission 2 (Dhanya et al., 2022)	39.6834
4	SSN_NLP Submission 3 (Dhanya et al., 2022)	39.9982

Table 2: Results of the participating systems in Word Error Rate

waveform is converted to multiple languages by pre-training a single model.

## 6 Evaluation of Results

The results submitted by the participants are evaluated based on the WER computed between the ASR hypotheses submitted by the participants and the ground truth of human speech transcription.

$$\text{WER (Word Error Rate)} = (S + D + I) / N$$

where,

S = No. of substitutions

D = No. of deletions

I = No. of insertions

N = No. of words in the reference transcription

As discussed in the methodology, different average word error rate are measured using various pre-trained transformer based models.

Performance of the ASR submitted by the participants are tabulated in Table 2. From Table 2, the Amrrs/wav2vec2-large-xlsr-53-tamil<sup>6</sup> model produces less WER compared to other models.

## 7 Conclusion

This overview paper discusses the shared task for vulnerable speech recognition in Tamil, where the speech corpus shared for this task is recorded from the elderly people. Recognizing the speech elderly people with better accuracy is a challenging task. Therefore, the collected speech corpus has been shared to participants to address the problem with their method to increase the accuracy and performance in recognizing the elderly people speech. Totally, there were two participants who took part in this shared task and submitted the results as transcripts of the given data. The team has compared the result with the human transcripts and calculated the WER. Both the participants have used different

transformer based model for building their recognition systems. Finally, the word error rates of the two participants are 39.4512 & 39.6487 respectively. Based on the observations, it is suggested that the transformer based model can be trained with given speech corpus which could give a better accuracy than the pre-trained model, as the transformer based model used are trained with common voice dataset. Also, a separate language model can also be created for this corpus.

## References

- R Anita and CN Subalalitha. 2019a. An approach to cluster Tamil literatures using discourse connectives. In *2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP)*, pages 1–4. IEEE.
- R Anita and CN Subalalitha. 2019b. Building discourse parser for Thirukkural. In *Proceedings of the 16th International Conference on Natural Language Processing*, pages 18–25.
- Bharathi Raja Chakravarthi, Ruba Priyadharshini, Thenmozhi Durairaj, John Phillip McCrae, Paul Buitaleer, Prasanna Kumar Kumaresan, and Rahul Ponnusamy. 2022. Findings of the shared task on Homophobia Transphobia Detection in Social Media Comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.
- Bharathi Raja Chakravarthi, Ruba Priyadharshini, Vigneshwaran Muralidaran, Shardul Suryawanshi, Navya Jose, Elizabeth Sherly, and John P McCrae. 2020. Overview of the track on sentiment analysis for Dravidian languages in code-mixed text. In *Forum for Information Retrieval Evaluation*, pages 21–24.
- Biswajit Das, Sandipan Mandal, and Pabitra Mitra. 2011. Bengali speech corpus for continuous automatic speech recognition system. In *2011 International conference on speech database and assessments (Oriental COCOSA)*, pages 51–55. IEEE.
- Srinivasan Dhanya, B Bharathi, D Thenmozhi, and Senthil Kumar. 2022. Ssnscse\_nlp@lt-edi-acl2022: Speech recognition for vulnerable individuals in Tamil using pre-trained Xlsr models. In *Proceedings of the Second Workshop on Speech and Language*

<sup>6</sup><https://huggingface.co/Amrrs/wav2vec2-large-xlsr-53-tamil>



- Technologies for Dravidian Languages*. Association for Computational Linguistics.
- Meiko Fukuda, Ryota Nishimura, Hiromitsu Nishizaki, Yurie Iribe, and Norihide Kitaoka. 2019. A new corpus of elderly japanese speech for acoustic modeling, and a preliminary investigation of dialect-dependent speech recognition. In *2019 22nd Conference of the Oriental COCODA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, pages 1–6. IEEE.
- Meiko Fukuda, Hiromitsu Nishizaki, Yurie Iribe, Ryota Nishimura, and Norihide Kitaoka. 2020. Improving speech recognition for the elderly: A new corpus of elderly japanese speech and investigation of acoustic modeling for speech recognition. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6578–6585.
- Annika Hämäläinen, António Teixeira, Nuno Almeida, Hugo Meinedo, Tibor Fegyó, and Miguel Sales Dias. 2015. Multilingual speech recognition for the elderly: The aalfred personal life assistant. *Procedia Computer Science*, 67:283–292.
- M Shamim Hossain, Md Abdur Rahman, and Ghulam Muhammad. 2017. Cyber–physical cloud-oriented multi-sensory smart home framework for elderly people: An energy efficiency perspective. *Journal of Parallel and Distributed Computing*, 103:11–21.
- Yurie Iribe, Norihide Kitaoka, and Shuhei Segawa. 2015. Development of new speech corpus for elderly japanese speech recognition. In *2015 International Conference Oriental COCODA held jointly with 2015 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE)*, pages 27–31. IEEE.
- Soonil Kwon, Sung-Jae Kim, and Joon Yeon Choeh. 2016. Preprocessing for elderly speech recognition of smart devices. *Computer Speech & Language*, 36:110–121.
- Taewoo Lee, Min-Joong Lee, Tae Gyoon Kang, Seokyeoung Jung, Minseok Kwon, Yeona Hong, Jungin Lee, Kyoung-Gu Woo, Ho-Gyeong Kim, Jiseung Jeong, et al. 2021. Adaptable multi-domain language model for transformer asr. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7358–7362. IEEE.
- Hui Lin and Yibiao Yu. 2015. Acoustic feature analysis and conversion of age speech. In *IET Conference Proceedings*. The Institution of Engineering & Technology.
- Ryo Masumura, Naoki Makishima, Mana Ihori, Akihiko Takashima, Tomohiro Tanaka, and Shota Orihashi. 2021. Hierarchical transformer-based large-context end-to-end asr with large-context knowledge distillation. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5879–5883. IEEE.
- Haoran Miao, Gaofeng Cheng, Changfeng Gao, Pengyuan Zhang, and Yonghong Yan. 2020. Transformer-based online ctc/attention end-to-end speech recognition architecture. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6084–6088. IEEE.
- Haoran Miao, Gaofeng Cheng, Pengyuan Zhang, Ta Li, and Yonghong Yan. 2019. Online hybrid ctc/attention architecture for end-to-end speech recognition. In *Interspeech*, pages 2623–2627.
- Anitha Narasimhan, Aarthi Anandan, Madhan Karky, and CN Subalalitha. 2018. Porul: Option generation and selection and scoring algorithms for a tamil flash card game. *International Journal of Cognitive and Language Sciences*, 12(2):225–228.
- Humberto Pérez-Espinosa, Juan Martínez-Miranda, Ismael Espinosa-Curiel, Josefina Rodríguez-Jacobo, and Himer Avila-George. 2017. Using acoustic paralinguistic information to assess the interaction quality in speech-based systems for elderly users. *International Journal of Human-Computer Studies*, 98:1–13.
- Ruba Priyadarshini, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, Thenmozhi Durairaj, Malliga Subramanian, Kogilavani Shanmugavadivel, Siddhant U Hegde, and Prasanna Kumar Kumaresan. 2022. Findings of the shared task on Abusive Comment Detection in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.
- Manikandan Ravikiran, Bharathi Raja Chakravarthi, Anand Kumar Madasamy, Sangeetha Sivanesan, Ratnavel Rajalakshmi, Sajeetha Thavareesan, Rahul Ponnusamy, and Shankar Mahadevan. 2022. Findings of the shared task on Offensive Span Identification in code-mixed Tamil-English comments. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.
- Ratnasingam Sakuntharaj and Sinnathamby Mahesan. 2016. A novel hybrid approach to detect and correct spelling in Tamil text. In *2016 IEEE International Conference on Information and Automation for Sustainability (ICIAFS)*, pages 1–6.
- Ratnasingam Sakuntharaj and Sinnathamby Mahesan. 2017. Use of a novel hash-table for speeding-up suggestions for misspelt Tamil words. In *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, pages 1–5.
- Ratnasingam Sakuntharaj and Sinnathamby Mahesan. 2021. Missing word detection and correction based on context of Tamil sentences using n-grams. In

- 2021 10th International Conference on Information and Automation for Sustainability (ICIAfS), pages 42–47.
- Anbukkarasi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, Ruba Priyadharshini, Subalalitha Chinnaudayar Navaneethakrishnan, Kogilavani Shanmugavadivel, Sajeetha Thavareesan, Sathiyaraj Thangasamy, Parameswari Krishnamurthy, Adeep Hande, Sean Benhur, Kishor Kumar Ponnusamy, and Santhiya Pandiyan. 2022. Findings of the shared task on Emotion Analysis in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.
- R Srinivasan and CN Subalalitha. 2019. Automated named entity recognition from tamil documents. In *2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP)*, pages 1–5. IEEE.
- C. N. Subalalitha. 2019. Information extraction framework for Kurunthogai. *Sādhanā*, 44(7):156.
- CN Subalalitha and E Poovammal. 2018. Automatic bilingual dictionary construction for Tirukural. *Applied Artificial Intelligence*, 32(6):558–567.
- S Suhasini and B Bharathi. 2022. Suh\_asr@lt-ediac12022: Transformer based approach for speech recognition for Vulnerable Individuals in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.
- António Teixeira, Annika Hämäläinen, Jairo Avelar, Nuno Almeida, Géza Németh, Tibor Fegyó, Csaba Zainkó, Tamás Csapó, Bálint Tóth, André Oliveira, et al. 2014. Speech-centric multimodal interaction for easy-to-access online services—a personal life assistant for the elderly. *Procedia computer science*, 27:389–397.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2019. Sentiment analysis in Tamil texts: A study on machine learning techniques and feature representation. In *2019 14th Conference on Industrial and Information Systems (ICIIS)*, pages 320–325.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2020a. Sentiment lexicon expansion using Word2vec and fastText for sentiment prediction in Tamil texts. In *2020 Moratuwa Engineering Research Conference (MERCon)*, pages 272–276.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2020b. Word embedding-based part of speech tagging in Tamil texts. In *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, pages 478–482.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2021. Sentiment analysis in Tamil texts using k-means and k-nearest neighbour. In *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 48–53.
- Michel Vacher, Frédéric Aman, Solange Rossato, and François Portet. 2015. Development of automatic speech recognition techniques for elderly home support: Applications and challenges. In *International Conference on Human Aspects of IT for the Aged Population*, pages 341–353. Springer.
- Shinji Watanabe, Takaaki Hori, Suyoun Kim, John R Hershey, and Tomoki Hayashi. 2017. Hybrid ctc/attention architecture for end-to-end speech recognition. *IEEE Journal of Selected Topics in Signal Processing*, 11(8):1240–1253.
- Jiabin Xue, Tieran Zheng, and Jiqing Han. 2021. Exploring attention mechanisms based on summary information for end-to-end automatic speech recognition. *Neurocomputing*, 465:514–524.
- Jiazhong Zeng, Jianxin Peng, and Yuezhe Zhao. 2020. Comparison of speech intelligibility of elderly aged 60–69 years and young adults in the noisy and reverberant environment. *Applied Acoustics*, 159:107096.
- Zhiping Zeng, Haihua Xu, Yerbolat Khassanov, Eng Siong Chng, Chongjia Ni, Bin Ma, et al. 2021. Leveraging text data using hybrid transformer-lstm based end-to-end asr in transfer learning. In *2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pages 1–5. IEEE.