M Gmail   **Pratya Nuankaew <nuankaew.p@gmail.com>**

MIWAI 2024 notification for paper 119
1 message

MIWAI 2024 <miwai2024@easychair.org>   Wed, Sep 11, 2024 at 10:05 AM
To: Pratya Nuankaew <pratya.nu@up.ac.th>

Dear Pratya Nuankaew,

We are delighted to inform you that the following paper has been CONDITIONALLY accepted as a regular paper to be presented in THE 17TH MULTI-DISCIPLINARY INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE 2024 (MIWAI 2024):

Paper id: 119
Title: Speech Recognition Model for Confused Thai Lanna Vocabulary Using Deep Learning Techniques
Authors: Wongpanya S. Nuankaew, Pathapol Jomsawan, Thapanapong Sararat, Pratya Nuankaew

Congratulations!

The review process has been carefully and deliberately conducted to ensure only high quality papers are accepted to MIWAI 2024. Each paper has been assigned to at least two PC members to do double-blind review by themselves (or their sub-reviewers). In addition, there have been several rounds of discussions about the quality of each paper among MIWAI 2024 General and Program Co-Chairs in order to maintain high quality standard of the conference before making decisions. The reviews and comments can be found below (after the notification message). We request you do your best TO TAKE ALL OF THEIR COMMENTS INTO ACCOUNT TO REVISE AND POLISH YOUR PAPER. Once you have uploaded your revised version, we will confirm with you as soon as possible. The failure to comply with MIWAI high quality standard may result in the rejection of your paper. Your time and effort to revise and improve the paper will be profoundly appreciated.

To complete your registration for the paper, please complete the following:

1. The deadline for camera-ready version submission is September 21, 2024 @23:59UTC-12. The details about registration and all the required materials for camera-ready can be found in the Registration section of the website https://miwai24.miwai.org.

2. The page limit for each paper is 12 pages in the LNCS (LNAI) format (including everything, such as References and Appendices). Please do your best to comply with this condition.

3. At least one of the authors of each accepted paper needs to complete the author registration and make payment. The link for registration and payment will available from September 14, 2024. The registration and payment must be completed by September 21, 2024 @23:59UTC-12.

Looking forward to seeing you at MIWAI 2024.

Best regards,
General and Program Co-Chairs of MIWAI 2024

******************* LIST OF REVIEWS ******************

SUBMISSION: 119
TITLE: Speech Recognition Model for Confused Thai Lanna Vocabulary Using Deep Learning Techniques

----------------------- REVIEW 1 ---------------------
SUBMISSION: 119
TITLE: Speech Recognition Model for Confused Thai Lanna Vocabulary Using Deep Learning Techniques
AUTHORS: Wongpanya S. Nuankaew, Pathapol Jomsawan, Thapanapong Sararat and Pratya Nuankaew

----------- Overall evaluation -----------
SCORE: 2 (accept)
----- TEXT:
The authors propose a novel application to solve issues faced in local Thai dialect. The proposed methodology uses a combination of deep learning, transformer models, and speech recognition techniques to address the specific challenges of the Thai Lanna language vocabulary. The paper is technically very well written

Suggestion
1.   Replace figure1 with Fig 1 or Figure 1
2.   Justify the age distribution for the samples: 10-20 years old: 60.00%, 21-30 years old: 37.10%, and 41-50 years old: 2.90% . Why it is not balanced?
3.   In Page 9, one table seems to be missing. It is mentioned in the first paragraph in Page 9, that "Table X presents the parameter settings for speech recognition models: " The sentence is incomplete and Table "X" should be number.
4.   Suggest to bold the best results in Table 1
5.   Figure quality to be improved, especially the legends are not legible

----------------------- REVIEW 2 ---------------------
SUBMISSION: 119
TITLE: Speech Recognition Model for Confused Thai Lanna Vocabulary Using Deep Learning Techniques
AUTHORS: Wongpanya S. Nuankaew, Pathapol Jomsawan, Thapanapong Sararat and Pratya Nuankaew

----------- Overall evaluation -----------
SCORE: 1 (weak accept)
----- TEXT:
The paper addresses the issue of speech recognition in Thai Lanna dialect. This is a good effort in making progress on a low-resource language. However, there are notable limitations, particularly in the generalization of the models to new words and accent variants. These challenges highlight the need for further research and refinement, justifying the score.

Pros:
- Detailed data processing: The dataset includes authentic recordings from native speakers, ensuring high-quality data that enhances the robustness of the models. Authors also make sure the training data is split and balanced across speakers from several regions.
- Authors provide detailed comparison of several advanced ASR models (HuBERT, Wav2Vec2-TH, Wav2Vec2, WavLM), highlighting their performance in recognizing Thai Lanna vocabulary.
- Training and Validation loss are very aligned showcasing balanced training and no overfitting. Authors use of tokenizers like deepcut or newmm allows for both rule-based as well as ML based learning.

Cons:
- The testing sample size is relatively small (only 2 individuals), which might limit the statistical significance of the results.
- Evaluation result on unseen words is not very good. CER for all models vary a great degree (34% to 61%) which shows difficulty with untrained vocabs and not enough training data. even though authors make effort to make the training balanced.
- Some grammatical errors that should be fixed
- Provide mathematical equation or neural net diagrams in methodology section

Recommendations:
- Must: Fix grammatical errors and proof-read
- Optional: For generalization, consider expanding the dataset to include more diverse speakers and a broader range of vocabulary.
- Optional: In-depth Error Analysis: Provide a more detailed analysis of why certain models performed better in specific scenarios and discuss potential modifications to improve weaker models.

---------------------- REVIEW 3 ---------------------
SUBMISSION: 119
TITLE: Speech Recognition Model for Confused Thai Lanna Vocabulary Using Deep Learning Techniques
AUTHORS: Wongpanya S. Nuankaew, Pathapol Jomsawan, Thapanapong Sararat and Pratya Nuankaew

----------- Overall evaluation -----------
SCORE: 0 (borderline paper)
----- TEXT:
The paper presents an experimental study on applying deep learning models in the recognition of confused Thai Lanna words. Four deep-learning models were tried with different processing techniques on a self-collected dataset of 200 ambiguous words, and the results were presented and analyzed. The experimental results show the potential of deep learning in solving the complexities of regional language processing.

The following comments might help in improving the paper:

- The investigated models (HuBERT, Wav2Vec2-TH, Wav2Vec2, WavLM) should be described, analyzed, and compared clearly, especially the clarification of feature extraction/speech representation and recognition. The word segmentation methods (newmm and deepcut) should also be presented more adequately.

- The experiment data should be presented more clearly, e.g., the number of confused words (100, 200, or 300) and the number of words and utterances in the training, validation, and testing sets.

- The experiment results should be described more clearly. The reader could be confused among the terms "Training Words, "New Words" and "Rare Words", What are they and how many they are, using for which purposes? Conventionally, training performance has less meaning, and the authors should focus more on testing performance. The testing performance in recognizing confused words should also be presented and analyzed more adequately, e.g., the word errors and the confusion matrix.

- The presentation of the paper should be completed (e.g., "Table X presents the parameter settings") and carefully checked for error and presentation.