Shashi Kumar

📞 +91-834-041-6933
✉ shashiiitg5572@gmail.com
in /shashikumar4
G Google Scholar
🌐 Personal Home Page

🎓 Education

Indian Institute of Technology, Guwahati   2013 – 2017
Bachelor of Technology, Electronics and Communication Engineering   *Guwahati, Assam*

📖 Publications

> S. Kumar, Shakti P. Rath and A. Pandey, "**Improved far-field speech recognition using Joint Variational Autoencoder**," arXiv preprint arXiv:2204.11286, 2022

> S. Kumar, Shakti P. Rath and A. Pandey, "**Speaker Normalization Using Joint Variational Autoencoder**," Proc. Interspeech 2021, pp. 1289-1293.

> V. P. Singh*, S. Kumar*, R. S. Jha*, and A. Pandey, "**SRIB Submission to Interspeech 2021 DiCOVA Challenge**," arXiv preprint arXiv:2106.07972, 2021

> V. Agrawal, S. Kumar and Shakti P. Rath, "**Whisper Speech Enhancement Using Joint Variational Autoencoder for Improved Speech Recognition**," Proc. Interspeech 2021, pp. 2706-2710.

> S. Kumar and Shakti P. Rath, "**Far-field speech enhancement using heteroscedastic autoencoder for improved speech recognition**," Proc. Interspeech 2019, pp. 446–450.

> M. K. Chelimilla, S. Kumar, and Shakti P. Rath, "**Joint distribution learning in the framework of variational autoencoders for far-field speech enhancement**," in 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). IEEE, 2019, pp. 245–251

> S. Kumar, Y. Iwahori, and M. K. Bhuyan, "**PCB defect classification using logical combination of segmented copper and non-copper part**," In Proceedings of International Conference on Computer Vision and Image Processing, pp. 523-532. Springer, Singapore, 2017.

🕐 Experience

Level AI   May 2022 – Present
Lead ML Engineer   *Oct. 2022 – Present*
Senior ML Engineer   *May 2022 – Sep. 2022*

> Improving Wav2Vec2 - CTC based ASR system for low-resource scenario. Focusing on self-supervised, semi-supervised and augmentation techniques along with better loss formulation.

> Working on Conformer transducer based ASR with improved robustness. Improving Named Entity Recognition (NER) using cold and shallow fusion.

> Developed Voice Activity Detection (VAD) and Diarization models with significant improvement in WER and DER on our internal test sets.

> Our ASR system is currently serving more than 100K hours of audios per month.

Samsung Research Institute   Apr. 2019 – Apr. 2022
Lead Software Engineer (Research)   *Feb. 2022 – Apr. 2022*
Senior Software Engineer (Research)   *Apr. 2019 – Jan. 2022*

> Tranformer based End2End ASR model training for UK accented English and targeted context-based words recognition improvements. Focused on generic improvements in each release cycle.

> Worked on novel speaker normalization and speech enhancement methods for far-field and whisper speech using Joint Variational Autoencoders (VAEs).

> Worked on ASR for far-field speech in home condition.

Software Engineer (Research) June 2017 – Mar. 2019
Bangalore, KA

> Developed Hybrid ASR models for UK and US accent English.
> Worked on different techniques like Heteroscedastic Autoencoder, complex Ideal Ratio Mask (cIRM) based speech enhancement.
> Worked on discriminative training (sMBR) of acoustic models.

FeltSo
May 2016 – July 2016

Remote Internship *Hyderabad, TL*

> Review text classification for aspect and sentiment analysis.
> Trained Word2Vec and Doc2Vec on similar text corpus as Embedding layer. Trained ML models like RF, SVM etc to improve performance.

Chubu University
May 2015 – July 2015

Research Internship *Nagoya, Japan*

> Developed a robust algorithm to detect and classify different types of defects present in PCBs.
> Proposed algorithm achieved state of the art results in non-referential approaches.
> A research paper discussing the algorithm is accepted at International Conference on Computer Vision and Image Processing, 2016.

⚗ Projects

Interspeech DiCOVA Challenge 2021
Feb. 2021 – Mar. 2021

Special session at Interspeech 2021 *Samsung Bangalore*

> Aimed to use cough sound recordings to detect COVID-19. On final leaderboard, our team was ranked 5th out of 29 teams.
> Our system yielded an AUC of 83.93 with specificity 70.83 at 80% sensitivity against baseline with AUC 68.54 and 53.65 specificity. Report Link: https://arxiv.org/abs/2106.07972
> Leaderboard link: https://competitions.codalab.org/competitions/29640#results

Aesthetic and Emotion analysis in Natural images using Deep Learning techniques
Aug. 2016 – Apr. 2017

Bachelor Thesis project under Dr. Amit Sethi *IIT Guwahati*

> Project aimed to classify natural images based on their aesthetics beauty and induced emotions.
> There was no large scale dataset available, so we crawled images from Flickr based on six emotion tags. These six emotions are Happy, Sad, Motivation, Fear, Anger and Solitude.
> We verified variance of this dataset in terms of concepts captured for each type of emotion. Finalized around 300,000 images. Trained ResNet-34 on this dataset gives around 55% top-1 accuracy.

HER2 scoring in breast cancer histology images
May 2016 – Jun. 2016

Challenge participation under Dr. Amit Sethi *IIT Guwahati*

> A research competition organized by University of Warwick, UK for automated scoring of HER2 in whole-slide images of breast cancer histology slides. We achieved overall 15th rank worldwide.
> Trained clusters using K-Means algorithm on a subset of evenly distributed foreground patches. Trained a modified VGG16 net using cluster number from K-Means as ground truth. On top of the CNN, we trained a Logistic Classifier for final prediction.

⚙ Skills

Prog. Lang.   Python, C++, C

ML libs/tools   PyTorch, Kaldi, SRILM, tensor2tensor, ESPnet, scikit-learn, fairseq, NeMo, SpeechBrain