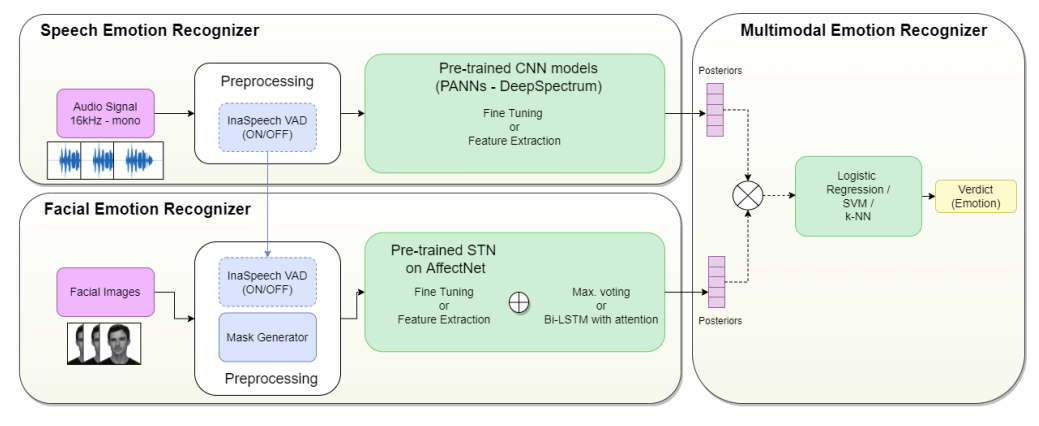
Multimodal Emotion Recognition on RAVDESS Dataset Using Transfer Learning 2021

link:https://www.mdpi.com/1424-8220/21/22/7665

Finally, from the combination of these two modalities with a late fusion strategy, we achieved 80.08%

accuracy on the RAVDESS dataset on a subject-wise 5-CV evaluation, classifying eight emotions.

-These modalities are combined employing two independent models connected by a late fusion strategy



-Facial Emotion Recognizer

two aapproaches

1-Feature-Extraction and Fine-Tuning with a pre-trained STN for Facial Emotion Recognition

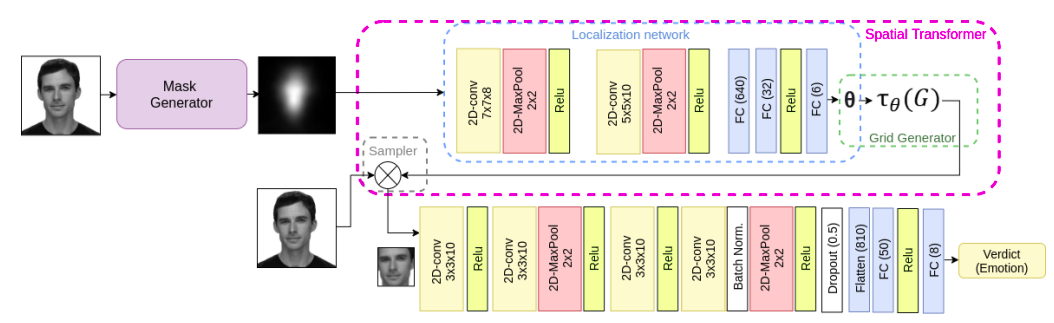
2-a static solution using a pooling (max. voting) and a dynamic solution using a bi-LSTM with an attention layer.

…………………

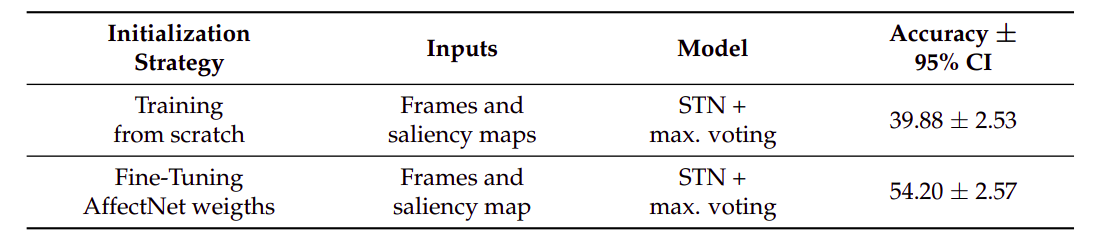
1-pre-trained STN for Facial Emotion Recognition

The pre-trained facial emotion recognition model was a modified STN that used saliency maps to capture the principal regions of a face, since the performance of these models improved if the STN had access to more processed information about the relevant figures and shapes that appear in an image

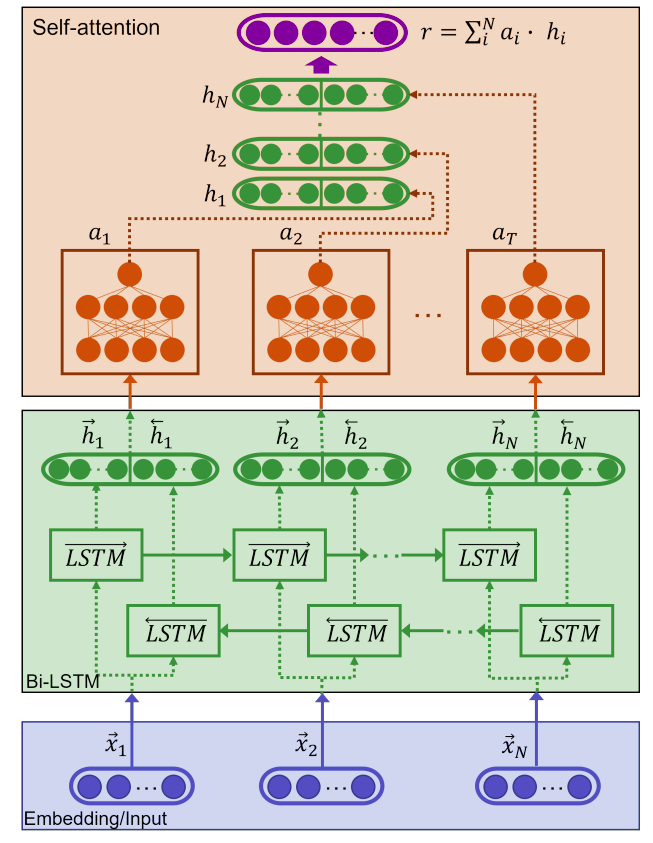
This model was trained on the AffecNet dataset and it received the facial images whose emotion has to be classified. Then, the Mask Generator created a saliency-based mask for each frame that passed to the localization network. The localization network learned the transformation parameters from this image and sended them to the sampler. The sampler received the θ parameters and the original image as inputs and returned the transformed version of the original image. This transformed image was passed to the second CNN to solve the facial emotion recognition problem

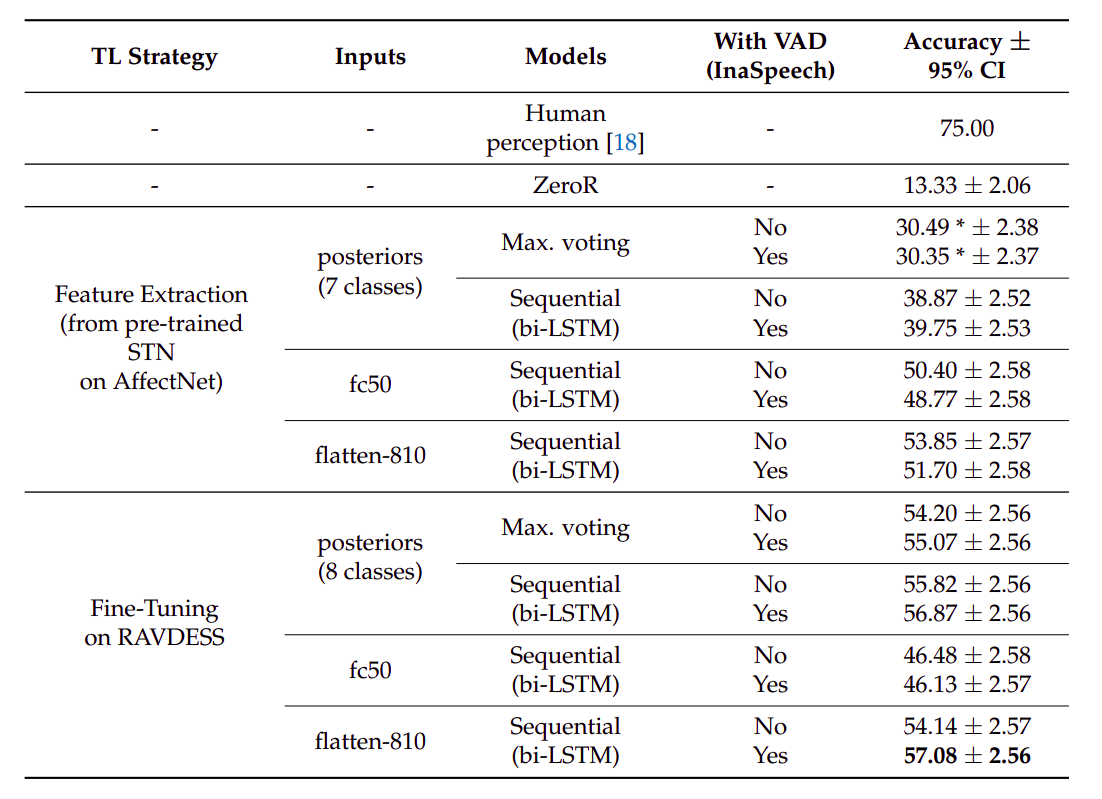


results:



2-For the sequential model, we employed an RNN on the posteriors of the seven-class pre-trained STN on AffectNet and the embeddings of the three-class pre-trained STN for valence recognition (positive, neutral, and negative)



Results: