Multimodal Speech Emotion Recognition using Decoding-enhanced BERT with Disentangled Attention and Wav2Vec2.0

Răzvan-Gabriel Petec
Department of Computer Science
Babeș-Bolyai University
Cluj-Napoca, Romania
razvan.petec@stud.ubbcluj.ro

Andrei Popoviciu
Department of Computer Science
Babeș-Bolyai University
City, Romania
andrei.popoviciu@stud.ubbcluj.ro

***Abstract*—Speech emotion recognition (SER) is a challenging task that involves extracting information about the emotions of some speaker using speech signals. As discovered by S. Yoon et. al [1], combining the text and the audio data from a speech achieves better results in classifying the emotions of the speaker, so we propose a novel multimodal SER model, which combines the Decoding-enhanced BERT with disentangled attention (DeBERTa) [2] and Wav2Vec2.0 [3]. DeBERTa is used to extract contextual representations from language, while Wav2Vec2.0 is used to capture multi-scale contextual affective representations from various time scales. By combining those models, we achieve better results than both of them, which means that our model managed to understand how to complete informations extracted from one of the models with the ones extracted by the other.**

***Index Terms*—affective computing, speech emotion recognition, decoding-enhanced BERT with disentangled attention, wav2vec2.0**

I. Introduction

Human emotions are inherent and influence behavior, providing insights into underlying thought processes [14]. Therefore, the ability to comprehend and identify emotions is crucial for the development of AI technologies, including personal digital assistants, that engage directly with humans. In the midst of a conversation involving multiple individuals, there is a continuous fluctuation of emotions felt and conveyed by each participant. Multimodal emotion recognition aims to tackle the challenge of monitoring expressed emotions through various modalities, such as text (what the speaker is talking) and audio (how the speaker is talking), in diverse settings like conversations.

When coming to work with text data, Natural Language Processing (NLP) is a branch of Computer Science and Artificial Intelligence that focuses on the computational treatment of human language with the core intent of making machines understand and generate human languages. NLP's favored applications, such as translation systems, search engines, natural language assistants, sentiment, and opinion analysis, are resolving societal issues at an unprecedented rate [5].

The sequential nature of textual data, wherein the order and relationships among words are determinative of complete sentence meaning, underscores a fundamental challenge. Traditional unsupervised machine learning models, which disregard word order and relationships while being confined by fixed input sizes, prompt a paradigm shift toward computationally deep approaches for textual data analysis.

The Recurrent Neural Network (RNN), as a sequential model adept at handling sequential data [6], encounters limitations attributable to protracted training times and difficulties in accommodating long-range sequence dependencies. The Long Short-Term Memory (LSTM), a variant of RNN, addresses some of these challenges by offering a resolution to long-range sequence dependencies. However, the enhancement comes at the expense of neglecting parallel computations and exhibiting slower operational speeds compared to the conventional RNN [7].

Emotion Detection (ED) is a branch of sentiment analysis (SA) that aims to identify subtle emotional nuances from various sources, including speech, images, and text. Despite the abundance of textual data, accurately detecting emotions from text remains challenging [8]. This difficulty stems, in part, from the absence of contextual cues provided by voice modulation, facial expressions, and other non-verbal signals. Additionally, the lack of an effective context extraction method for text poses another obstacle. Moreover, the need to differentiate between emotion-conveying words and their actual emotional meanings presents a significant hurdle, as many texts convey multiple emotional expressions. Recent advancements in state-of-the-art (SOTA) results have been achieved using pre-trained transformer-based models in this field.

The speech emotion signal, existing as a continuous time-domain signal, encapsulates both emotional content and informational aspects. Speech features can be categorized as either local or global, contingent on the approach to feature extraction. Local features, referred to as segmental or short-term features, capture temporal variations within the signal. On the other hand, global attributes, also known as long-term or supra-segmental features, encapsulate the overall statistical characteristics of the signal. Within Speech Emotion Recognition (SER) systems, the analysis of local and global speech signal features is typically conducted across four categories: prosodic, spectral, voice-quality, and features derived from the Teager energy operator (TEO) [16].

Support Vector Machine (SVM), Random Forest (RF), Gaussian Mixture Model (GMM), K-nearest neighbor (KNN), Hidden Markov Model (HMM), Decision Tree, and Dynamic Time Warping are among the frequently utilized classifiers in Speech Emotion Recognition (SER). However, traditional machine learning-based approaches have demonstrated suboptimal performance due to their reliance on manually crafted features, inadequate feature representation, and limitations in handling intricate and extensive datasets [17].

In contrast, deep learning-based methods have emerged as more effective solutions for SER, showcasing superior capabilities in feature representation, adeptness in handling complex features, capacity to learn from unlabeled data, and scalability to manage larger datasets. Various deep learning algorithms, including Convolutional Neural Network (CNN), Deep Neural Network (DNN), and Long Short-Term Memory (LSTM), have proven successful in automating Speech Emotion Recognition [18].

Thus, the objective of this paper is:

- to review DeBERTa, one of the most precise transformer-based models at the moment, specifcally in text-based emotion detection.
- to train Wav2Vec2.0 to predict different emotions in the audio data of the used dataset.
- to design and to test a novel architecture that tries to combine the data from the previously mentioned models, to improve the performance of the last two models.

II. Related Work

A. Speech Emotion Recognition

Speech emotion recognition (SER) has gained significant traction in recent years, with researchers exploring various approaches to accurately classify emotions expressed through spoken language. A common approach involves extracting acoustic features from speech signals and employing machine learning algorithms to identify the corresponding emotion. For instance, in the study by Bhangale and Kothandaraman [15], multiple acoustic features, including mel-frequency cepstral coefficients (MFCCs), pitch, and energy, were extracted from speech signals and fed into a 1D deep convolutional neural network (DCNN) for emotion classification. This approach achieved an accuracy of 93.1% on the EMODB dataset, with 7 predicted emotions.

B. CNN and LTSM based Emotion Recognition Text Models

Yoon et. al [9] proposed 3 text models using 2 different approaches for the IEMOCAP dataset: Text_Model1 which used 1D convolutions of kernel size 3 each, with 256, 128, 64 and 32 filters using Relu as Activation and Dropout of 0.2 probability, followed by 256 dimension fully connected layer and Relu, feeding to 4 output neurons with Softmax; Text_Model2 which used two stacked LSTM layers with 512 and 256 units followed by a Dense layer with 512 units and Relu Activation; and Text_Model3 which was simillar to Text_Model2, but they used Randomized initialization with 128 dimensions, the first two models being initialized with Glove Embeddings based word-vectors. Text_Model3 obtained similar performance as Text_Model2. The LSTM based models use Adadelta and Convolution based models use Adam as optimizers. The best of the three models was Text_Model3, obtaining an accuracy of 64.78%, using 4 emotions from the IEMOCAP dataset.

C. BERT based Emotion Recognition Text Models

Emile et. al [10] proposed a hierarchical transformer-based encoder tailored for spoken dialog. They extend two well-known pre-training objectives to adapt them to a hierarchical setting. They used BERT [1] through the pytorch implementation provided by the Hugging Face transformers library [11]. The pre-trained model was fed with a concatenation of the utterances. Formally given an input context Ck = (u1, ... uT), the concatenation [u1, . . . , uT] was fed to BERT. With this approach, they obtained an accuracy of 45.0% on the IEMOCAP dataset with 6 labels.

D. Multimodal Text and Speech Emotion Recognition

Multimodal approaches, combining acoustic and textual features, have shown promising results in enhancing the accuracy of SER systems. Yoon et al. [9] proposed a multimodal SER system that utilizes acoustic features extracted from speech signals using a DCNN and text features extracted from transcripts using a bidirectional long short-term memory (BLSTM) network. The extracted features are then fused and classified using a support vector machine (SVM). This system achieved an accuracy of 91.2% on the IEMOCAP dataset, using 4 emotions, a lot better than the 64.78% they obtain by only using the text features.

III. Approach

A. IEMOCAP Dataset

Interactive EMOtional dyadic motion CAPture database (IEMOCAP) [4] is a multimodal and multi-speaker database that contains approximately 12 hours of audiovisual data, including video, speech, motion capture of the face, and text transcriptions. The database is annotated by multiple annotators into categorical labels: anger, excited, fear, sad, surprised, frustrated, happy, disappointed and neutral. From all of this data, the text and the labels were extracted to fit through the proposed model.

Because the dataset isn't so balanced (Fig. 1), we decided to work only with the following emotions: Angry, Frustrated, Neutral, Sadness, Excited and Happy. Also, there isn't an official split of the dataset, so we splitted it in 80% training and 20% validation, and setted all the seeds to 0, so the results could be consistent between experiments.

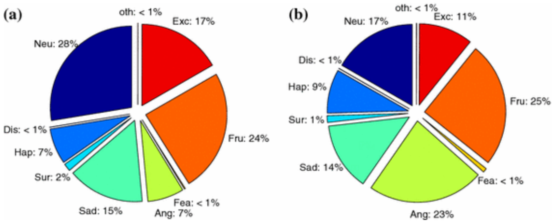


Fig. 1. IEMOCAP data distribution

B. Data Preprocessing

To be able to make more experiments, we prepreprocessed the data, so the dataloader wouldn't need to process all the things at runtime. Firstly, the IEMOCAP dataset, consists in 5 directories, one for each session. Each of those contains 2 directories: dialog and sentences. We used the data from the EmoEvaluation directory to match the audio files with some emotion, based on their evaluation. After that, we used the transcriptions directory to match the audio files to what the actors were saying, and then exported the table with the audio file, the text and the emotion.

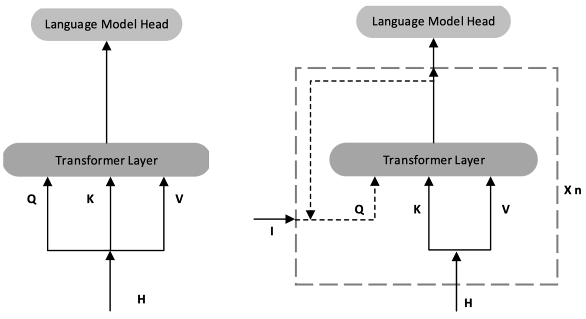
Furthermore, we processed the raw audio files by passing the raw audio data to the Wav2Vec2 processor correspondent to the model we used (facebook/wav2vec2-base-960h), which wraps a Wav2Vec2 feature extractor and a Wav2Vec2 CTC tokenizer into a single processor and exported the table containing the output of the audio processor, the raw text data and the emotion label. Lastly, we processed the raw text data by passing it to the DeBERTa AutoTokenizer, receiving some tokens, which we exported in a table along the processed audio data and the emotion label.

Because we wanted to train the model on mini-batches to improve generality [13], at these steps, we padded the processed audio and text data to match the one with the longest length. As a compromise, because some of the biggest audio data were way too big (546220), and we didn't have the required computing power, we truncated the data a 10th of its length, so some of the longer audio data wouldn't be very accurate, but most of them aren't in this category, so it shouldn't drastically affect our results.

C. Text Model Transfer Learning

Similarly to Emile et. al [10], the DeBERTaV3-Small model published by Microsoft [12] from the transforms library provided by Hugging Face [11] was used to try to achieve some good resulst on the text data of IEMOCAP. This version of the model has 44M backbone parameters, having a vocabulary containing 128K tokens.

The decision to use this model instead of other BERT models, such as BERT or RoBERTa, was taken because DeBERTa improves the other two models by using disentangled attention and enhanced mask decoder (Fig. 2). With those two improvements, DeBERTa out perform RoBERTa on a majority of Natural Language Understanding (NLU) tasks with 80GB training data.



(a) BERT decoding layer

(b) Enhanced Mask Decoder

Fig. 2. Comparison between the BERT's and DeBERTa's decoding layer.

As a further improvement, DeBERTa V3 used ELECTRA-Style pre-training with Gradient Disentangled Embedding Sharing. Compared to DeBERTa, Microsoft's V3 version significantly improves the model performance on downstream tasks.

D. Audio Model Transfer Learning

The Wav2Vec2.0 [3] model (Fig. 3) (Wav2Vec2-Base, published by Facebook, trained on 960 hours of LibriSpeech 16kHz sampled speech audio, having 94.4M parameters) is notable for its approach in self-supervised learning from speech audio. It significantly outperforms other semi-supervised methods in simplicity and effectiveness. The model, trained using Connectionist Temporal Classification (CTC), requires decoding of its output using a specific tokenizer. It excels in scenarios with limited labeled data, demonstrating its robustness and versatility across various speech recognition contexts. When training this model on our data, we used the crossentropy loss.

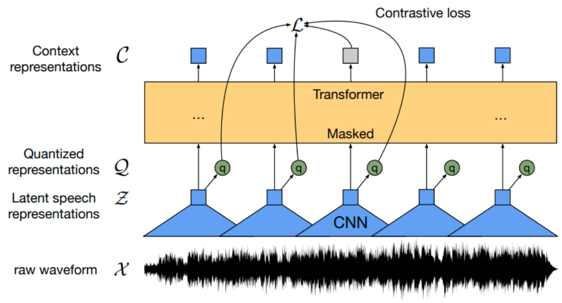


Fig. 3. The architecture of the Wav2Vec2.0 model.

E. Fusioning the models

We firstly fine-tuned on the IEMOCAP dataset both of the audio model (Wav2Vec2.0, pipeline 1) and the text model (DeBERTa, pipeline 2). After that, when needed to train the Fusion Model (Fig. 4, pipeline 3), we frozen all the weights of both the audio model and the text model, then took the features and the classification of the audio model, the features and the classification of the text model, flattened them, and passed that data through an MLP (Multi-Layer Perceptron) with 4 layers, and then optimized the output of this MLP to match the emotion, so only learnable weights in this model would be only the weights of the MLP.

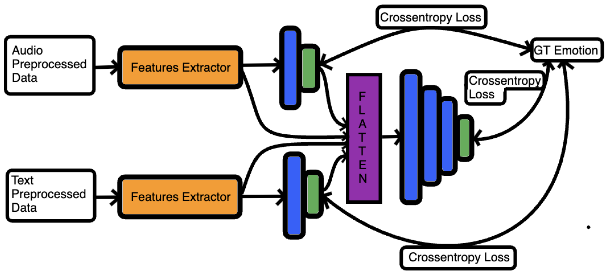


Fig. 4. The architecture of our Fusion Model.

F. Training

a) Loss function: Having the labeled IEMOCAP dataset and taking into the consideration that the output of our model was equal to the number of possible classes in the dataset, the cross entropy loss function was used:

$$-\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$$

b) Optimizer: To minimize our loss and for a faster convergence, the weighted Adam optimizer was used, having the following equations:

- $m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$
- $v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$
- $\hat{m}_t = \frac{m_t}{1-\beta_1^t}$
- $\hat{v}_t = \frac{v_t}{1-\beta_2^t}$
- $\theta_t = \theta_{t-1} - \frac{\text{lr} \cdot \hat{m}_t}{\sqrt{\hat{v}_t}+\epsilon} - \text{lr} \cdot \text{weight\_decay} \cdot \theta_{t-1}$

where:

- m_t is the first moment estimate.
- v_t is the second moment estimate.
- β_1 and β_2 are the exponential decay rates for the moment estimates.
- g_t is the gradient at time step t.
- $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected moment estimates.
- lr is the learning rate.
- ϵ is a small constant to prevent division by zero.
- weight_decay is the weight decay term.

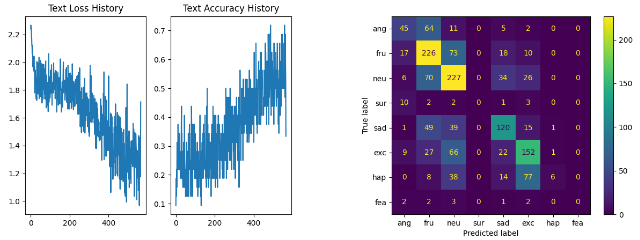
c) Metrics: The metric we choose to validate our models was accuracy (number of right decisions divided by the total number of decisions, this is one of the most popular metric in classification, so we could easily compare our results with others), paired with the confusion matrix (for each class which prediction the model gave versus the ground truth) to see where our model would screw up.

IV. Results

The DeBERTa model was trained for 9 epochs with all the emotions in the IEMOCAP database, the training time for an epoch was approximately 45 minutes on a M2 Pro with 12 CPU cores and 19 GPU cores, 32GB RAM. Different metrics were took throughout the training for each batch.

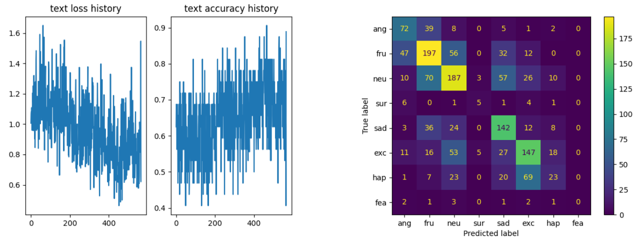
TABLE I
DeBERTa Evolution Across Epochs

Epoch Start	Epoch End	Validation Accuracy
1	3	51.493%
4	6	51.293%
7	9	51.957%



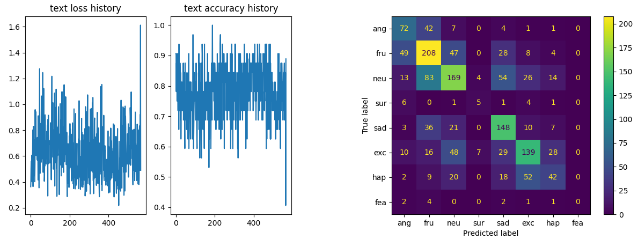
	ang	fru	neu	sur	sad	exc	hap	fea
ang	45	64	11	0	5	2	0	0
fru	17	226	73	0	18	10	0	0
neu	6	70	227	0	34	26	0	0
sur	10	2	2	0	1	3	0	0
sad	1	49	39	0	120	15	1	0
exc	9	27	66	0	22	152	1	0
hap	0	8	38	0	14	77	6	0
fea	2	2	3	0	1	2	0	0

Fig. 5. Loss and accuracy evolution for the training set and confusion matrix for the evaluation set for DeBERTa, epochs 1-3.



	ang	fru	neu	sur	sad	exc	hap	fea
ang	72	39	8	0	5	1	2	0
fru	47	197	56	0	32	12	0	0
neu	10	70	187	3	57	26	10	0
sur	6	0	1	5	1	4	1	0
sad	3	36	24	0	142	12	8	0
exc	11	16	53	5	27	147	18	0
hap	1	7	23	0	20	69	23	0
fea	2	1	3	0	1	2	1	0

Fig. 6. Loss and accuracy evolution for the training set and confusion matrix for the evaluation set for DeBERTa, epochs 4-6.



	ang	fru	neu	sur	sad	exc	hap	fea
ang	72	42	7	0	4	1	1	0
fru	49	208	47	0	28	8	4	0
neu	13	83	169	4	54	26	14	0
sur	6	0	1	5	1	4	1	0
sad	3	36	21	0	148	10	7	0
exc	10	16	48	7	29	139	28	0
hap	2	9	20	0	18	52	42	0
fea	2	4	0	0	2	1	1	0

Fig. 7. Loss and accuracy evolution for the training set and confusion matrix for the evaluation set for DeBERTa, epochs 7-9.

We can see that the loss function converges very well for the first epochs (Fig. 5, Fig. 6), and for the last ones it seems like it converged and there is no need for further training (Fig. 7), but the accuracies on the validation dataset throughout the epochs doesn't seem to change, meaning that it could be harder for the model to generalize. Seeing the confusion matrix, it's visible that many of the emotions the model is confusing are similar, so it could be harder for the model to distinguish between similar emotions.

TABLE II
TEXT MODEL PERFORMANCE COMPARISON

Model	*Accuracy*	*Metric*	*Labels*
bc-LSTM + Attention (2017)	58.540%	Unweighted Accuracy	4
Attention-BLSTM (2018)	62.900%	Unweighted Accuracy	4
Dialogue-CRN + RoBERTa (2021)	67.530%	Unweighted Accuracy	4
EmoCAPS (2022)	71.770%	Unweighted Accuracy	4
DeBERTa (ours)	**51.957%**	**Unweighted Accuracy**	**9**

The Wav2Vec2.0 model was also trained for 9 epochs, but with only the 6 labels mentioned in the IEMOCAP Dataset section, the training time for an epoch was approximately 25 minutes on a M2 Pro with 12 CPU cores and 19 GPU cores, 32GB RAM.

TABLE III
WAV2VEC2.0 EVOLUTION ACROSS EPOCHS

Epoch Start	Epoch End	Validation Accuracy
1	3	35.501%
4	6	45.461%
7	9	43.970%

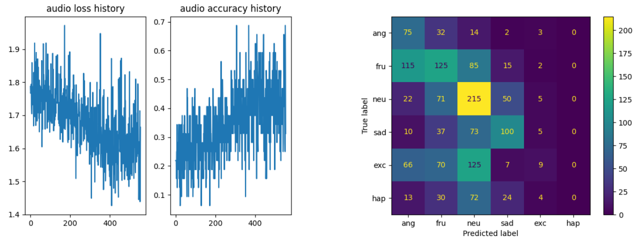


Fig. 8. Loss and accuracy evolution for the training set and confusion matrix for the evaluation set for Wav2Vec2.0, epochs 1-3.

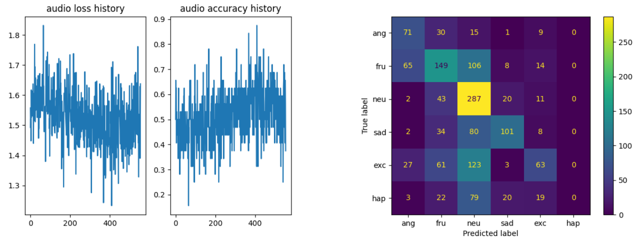


Fig. 9. Loss and accuracy evolution for the training set and confusion matrix for the evaluation set for Wav2Vec2.0, epochs 4-6.

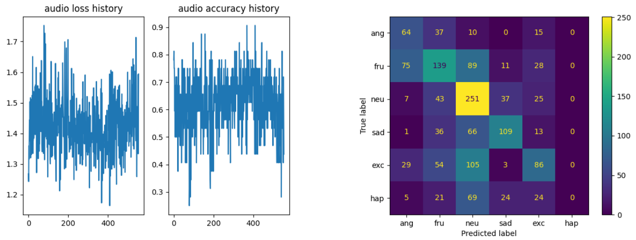


Fig. 10. Loss and accuracy evolution for the training set and confusion matrix for the evaluation set for Wav2Vec2.0, epochs 7-9.

After the 4 – 6 training (Fig. 9), the model seemed to be converged, remaining in that area, so we didn't train it any longer. Another problem that can be seen in the confusion matrices is that the model doesn't guess the happy emotion, which can be a problem on our side, but the accuracy seemed good enough to continue our experiments.

TABLE IV
AUDIO MODEL PERFORMANCE COMPARISON

Model	Accuracy	Metric	Labels
CNN + LSTM (2018)	61.7%	Unweighted Accuracy	4
SYSCOMB (2020)	74.0%	Unweighted Accuracy	4
SER with MTL (2021)	78.9%	Unweighted Accuracy	4
MMER (2022)	81.7%	Unweighted Accuracy	4
Wav2Vec2.0 (ours)	**45.461%**	**Unweighted Accuracy**	**6**

The fusion model was trained for 6 epochs with the 6 labels mentioned before, the training time for an epoch was approximately 15 minutes on a 16GB P5000 GPU with 8 CPU cores and 30GB RAM.

TABLE V
FUSION EVOLUTION ACROSS EPOCHS

Epoch Start	Epoch End	Validation Accuracy
1	3	58.943%
4	6	59.078%

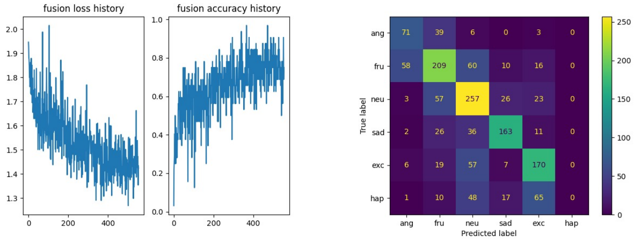


Fig. 11. Loss and accuracy evolution for the training set and confusion matrix for the evaluation set for the Fusion model, epochs 1-3.

We can see that the model pretty much converged after those epochs (Fig. 12. The results were greater than the ones

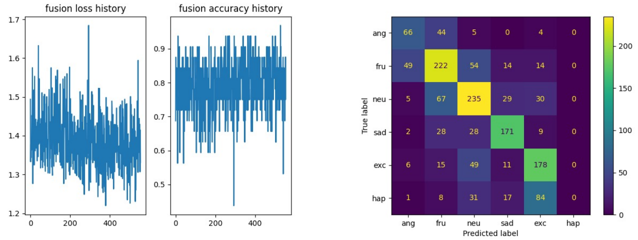


Fig. 12. Loss and accuracy evolution for the training set and confusion matrix for the evaluation set for the Fusion model, epochs 4-6.

obtained from the models individually, and that's probably because the MLP learned how to combine the data in such a way when there were some uncertainty from one of the models, maybe the other one could help. From the confusion matrix, we can see that the same problem as before occurred again (the last column not being guessed by the model), but whatsoever, we proved that in this way, we managed to increase the performance the performance of both the models.

TABLE VI
FUSION MODEL PERFORMANCE COMPARISON

Model	Accuracy	Metric	Labels
CHFusion (2018)	75.9%	Unweighted Average	4
Self-attention weighted correction (2022)	76.8%	Unweighted Average	4
COGMEN (2022)	68.2%	Unweighted Average	6
DeBERTA + W2V2.0 (ours)	**59.078%**	**Unweighted Average**	**6**

V. CONCLUSIONS AND FUTURE WORK

This paper demonstrates the effectiveness of applying DeBERTa, a pre-trained language model architecture that integrates disentangled attention and enhanced mask decoder into the BERT framework, and Wav2Vec2.0, a pre-trained audio model architecture, in a multimodal fashion for the task of speech emotion recognition.

In future work, some more hyperparameter tuning could be done to try to achieve better results, as not much was done in this paper. Also, the experiments were done using the small version of DeBERTa and the base version of Wav2Vec2.0 due to hardware limitations, but bigger versions of the same models could be used to achieve better results.

REFERENCES

[1] Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” North American Chapter of the Association for Computational Linguistics. (2019)

[2] He, Pengcheng, Xiaodong Liu, Jianfeng Gao and Weizhu Chen. DeBERTa: Decoding-enhanced BERT with Disentangled Attention. (2020)

[3] Baevski, Alexei, Henry Zhou, Abdel-rahman Mohamed and Michael Auli. “wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations.” ArXiv abs/2006.11477 (2020): n. pag.

[4] C. Busso, M. Bulut, C.C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J.N. Chang, S. Lee, and S.S. Narayanan, ”IEMOCAP: Interactive emotional dyadic motion capture database,” Journal of Language Resources and Evaluation, vol. 42, no. 4, pp. 335-359. (2008)

[5] Yue L, Chen W, Li X, Zuo W, Yin M. A survey of sentiment analysis in social media. Knowl Inf Sys 60(2):617–663. (2019)

[6] Du K-L, Swamy MN. Neural networks and statistical learning. Springer Science & Business Media, Berlin. (2013)

[7] Sundermeyer M, Schlüter R, Ney H. Lstm neural networks for language modeling. In: Thirteenth annual conference of the international speech communication association, p 4. (2012)

[8] Acheampong FA, Wenyu C, Nunoo-Mensah H. Text-based emotion detection: Advances, challenges, and opportunities. Engineering Reports e12189. (2020)

[9] S. Yoon, S. Byun and K. Jung, ”Multimodal Speech Emotion Recognition Using Audio and Text” IEEE Spoken Language Technology Workshop (SLT), pp. 112-118. (2018)

[10] Emile Chapuis, Pierre Colombo, Matteo Manica, Matthieu Labeau, and Chloé Clavel. Hierarchical Pre-training for Sequence Labelling in Spoken Dialog. In Findings of the Association for Computational Linguistics: EMNLP, pages 2636–2648. (2020)

[11] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, R’emi Louf, Morgan Funtowicz, and Jamie Brew. Huggingface’s transformers: State-of-the-art natural language processing. (2019)

[12] He, Pengcheng & Gao, Jianfeng & Chen, Weizhu. DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing. (2021)

[13] Geiping, Jonas & Goldblum, Micah & Pope, Phillip & Moeller, Michael & Goldstein, Tom. Stochastic Training is Not Necessary for Generalization. (2021)

[14] Marvin Minsky. The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. SIMON & SCHUSTER. (2007)

[15] Bhangale, Kishor, and Mohanaprasad Kothandaraman. Speech Emotion Recognition Based on Multiple Acoustic Features and Deep Convolutional Neural Network. Electronics 12 (2023)

[16] Michalis P., Spyrou E., Giannakopoulos T., Siantikos G., Sgouropoulos D., Mylonas P., Makedon, F. Deep visual attributes vs. hand-crafted audio features on multidomain speech emotion recognition. Computation (2017)

[17] Majid W.T., Gunawan T.S., Qadri S.A.A., Kartiwi M., Ambikairajah E. A comprehensive review of speech emotion recognition systems. IEEE Access (2021)

[18] Rashid J., WahTeh Y., Hanif F., Mujtaba, G. Deep learning approaches for speech emotion recognition: State of the art and research challenges. Multimed. Tools Appl. (2021)

[19] Joshi, A., Bhat, A., Jain, A., Singh, A., & Modi, A. COGMEN: COntextualized GNN based Multimodal Emotion recognitioN. ArXiv, abs/2205.02455. (2022)