DEEP FUNCTIONAL MULTIPLE INDEX MODELS WITH AN APPLICATION TO SER

Saumard Matthieu
VISION-AD Team, Labisen
ISEN Yncréa Ouest
Brest, France
matthieu.saumard@isen-ouest.yncrea.fr

El Haj Abir
VISION-AD Team, Labisen
ISEN Yncréa Ouest
Caen, France
abir.el-haj@isen-ouest.yncrea.fr

Napoleon Thibault
VISION-AD Team, Labisen
ISEN Yncréa Ouest
Brest, France
thibault.napoleon@isen-ouest.yncrea.fr

***Abstract*—Speech Emotion Recognition (SER) plays a crucial role in advancing human-computer interaction and speech processing capabilities. We introduce a novel deep-learning architecture designed specifically for the functional data model known as the multiple-index functional model. Our key innovation lies in integrating adaptive basis layers and an automated data transformation search within the deep learning framework. Simulations for this new model show good performances. This allows us to extract features tailored for chunk-level SER, based on Mel Frequency Cepstral Coefficients (MFCCs). We demonstrate the effectiveness of our approach on the benchmark IEMOCAP database, achieving good performance compared to existing methods.**

***Index Terms*—speech emotion recognition, human-computer interaction, functional data analysis, deep learning**

I. Introduction

Emotion recognition is an essential aspect of human-robot interaction, as it allows for a more natural and effective means of communication. Humans communicate their emotions through various modalities, including facial expressions, body language, and speech. However, among these modalities, speech plays a crucial role in emotion recognition, as it is one of the most reliable and informative ways to convey emotional information. Through speech, individuals can convey not only the content of their message but also the underlying emotional state, including tone, pitch, and intonation. Moreover, speech can also provide insights into the speaker's personality, cognitive state, and overall well-being. As such, developing accurate and efficient methods for speech emotion recognition is critical for creating more effective and responsive human-robot interfaces.

Support Vector Machines have been shown to be effective for speech emotion recognition. They are able to learn complex decision boundaries and can handle high-dimensional data, such as speech. Other types of classifiers that have been used for speech emotion recognition include decision trees, k-nearest neighbors, and neural networks. Speech emotion recognition has made rapid progress in recent years with the use of deep learning and convolutional neural networks (CNN) [1], transformer, attention and self-supervised learning [2]–[5] methods.

Since the pioneer monographs [6] and [7], functional data analysis (FDA) has become a vibrant field of research in the statistical community due to it vast applications in various sciences, including astronomy, chemo-metrics, health and finance [8]–[10]. The core objects of FDA are curves or functions of a separable Hilbert space like $L^2[0,1]$. However, statistical study of random variables of more general functions space like Banach space [11] or curves on manifold [12] falls within the spectrum of FDA. The fundamental frequency curve of an utterance has already been considered as a functional object [13], [14]. Recently, [15] analysed dialect sound variations across Great Britain using a spatial modeling approach that employs MFCC. In this article, we extend this point of view to speech emotion recognition. Each coefficient of MFCC can be interpreted as a functional data variable. With a collection of coefficients, it is therefore natural to establish a correspondence between a speech recording and a multivariate functional data object. The number of covariates of the multivariate functional data object is the number of coefficients of the MFCC used. A transformation of the multivariate functional object is employed to serve as the final multivariate object, which is considered in a functional multiple-index model.

In our work, we propose a novel approach for speech emotion recognition that involves treating Mel Frequency Cepstral Coefficients (MFCCs) as a functional data object. By doing so, we can represent each coefficient as a function of time, and thus extract information from the speech signal. However, to compare functional data objects between samples with different duration, we need to preprocess the MFCC. This is achieved by splitting the MFCC in chunks, which allows us to represent each sample as a multivariate functional object. An abundant literature on chunk-level SER exists, see [16], [17] and [18] for example. This multivariate object can then be transformed into another multivariate object using a suitable transformation, enabling us to use the functional multiple-index model for multivariate functional covariate to classify the emotions of the speaker.

This novel approach is particularly advantageous as it allows us to consider each MFCC as a functional variable, which captures the dynamic nature of speech and its relationship to emotions. By using a multivariate functional object, we can compare it across samples with different duration. Moreover, the functional multiple-index model allows us to consider the interdependence between the different coefficients of

the MFCC, providing a more accurate and comprehensive representation of the speech signal. Overall, our approach shows interesting perspectives for improving speech emotion recognition.

The paper is organized as follows. We highlight the previous work in the following section. Next, we describe the method in details with our contributions. We propose to evaluate our method on IEMOCAP database described in the following fourth section along with the results and comparison with other methods. We conclude and discuss the results in the final section.

II. Related work

A. On functional data models

The functional single index models have been studied both from a theoretical and practical point of view in [19]–[21]. Some authors [22] and [23] use functional additive models that can be more stable than functional multiple-index models. The authors of [24] use deep neural network strategy to learn the parameter of the functional linear model by using a new functional neuron that can learn the functional representation with functional inputs. The article of [25] proposes a novel neural network that learns the best basis functions for supervised learning tasks with functional inputs. They called it AdaFNN. The AdaFNN network parameterizes each basis node with a micro neural network that outputs a score of the input function $X(t)$, which is the inner product between the basis function $\theta(t)$ and the input function:

$$c = \langle \theta, X \rangle = \int \theta(t)X(t)\mathrm{d}t.$$

They introduce neural networks that employ a new Basis Layer whose hidden units are each basis functions themselves implemented as a micro neural network. On the link between deep learning and FDA, there exist other approach based on a functional layer developed by [26] investigated in details by [27] and [24]. We present in the next section our model and a simulation comparison with [24]. The advantage of the approach in [25] is that we can use all the deep learning tools without rewriting the back-propagation algorithm which is necessary in [24].

Let us introduce the functional single index model in the regression context. Let $(Y_i, X_i)_{i=1,\dots,n}$ be the set of data where X_i represent the i-th functional data variable in $L^2[0,1]$, the space of square integrable function on $[0,1]$, and $Y_i \in \mathbb{R}$.

$$Y_i = g\left(\langle X_i, \theta \rangle\right) + \varepsilon_i, \tag{1}$$

where ε_i are the error terms, $\langle ., . \rangle$ is the inner product in $L^2[0,1]$. g and θ are unknown function to be estimated. Generally, g is estimated from a nonparametric framework with a kernel and θ is estimated from a basis decomposition or a functional principal component analysis. We propose to extend this definition to a functional multiple-index model for multivariate functional data covariate, see section 3.3.

B. On speech emotion recognition

There is a rich literature on speech emotion recognition using various deep learning architectures. We can cite [28]–[31] which use self-attention, Bayesian neural networks and positional encoding. The article of [32] proposes to make silence representations of speech. The article of [33] makes progress by creating self-supervised learning for SER tasks.

III. Method

A. MFCC

Let us recall the method to calculate the MFCC. With at hand a raw audio signal data representing by a time series $s(t)$ for $t = 1, \dots, T$. We can consider that s is defined for $t \in \mathbb{Z}$ by adding 0 to non-value. Let $w_M(t)$ for $t \in \mathbb{Z}$ be a window function of width M, we can define the spectrogram of the audio signal $s(t)$ by

$$\text{Spec}(t, \omega) = |\sum_{u=1}^{T} s(t-u) w_M(u) \exp(-i\omega u)|, \tag{2}$$

for $t = 1, \dots, T, \omega \in [0, 2\pi]$.

Hence, we can define the Mel spectrogram, which is a filtered version of the spectrogram to represent the human ear auditory system:

$$\text{MelSpec}(t, f) = \sum_{k=0}^{N-1} \text{Spec}\left(t, \frac{2k\pi}{N}\right) b_{f,k} \tag{3}$$

for $f = 0, \dots, F$, with $b_{f,k}$ representing the set of the Mel-scale filter bank. Recall that the Mel scale is $m = 2595 \log_{10}(1 + \frac{f}{700})$. The MFCC are then:

$$\text{MFCC}(t, m) = \frac{1}{F} \sum_{f=0}^{F} \log\left(\text{MelSpec(t,f)}\right) \exp(i(2\pi \frac{m-1}{F+1})f), \tag{4}$$

for $m = 1, \dots, n_{MFCC}$.

Note that there exist variations of the definition of MFCC in the literature. For example, we can use a Discrete Cosinus Transform (DCT) to return to the time scale.

B. Functional multiple index models

For extracting new features based on MFCC, we employ a deep functional multiple index model. Let $(Y_i, X_i)_{i=1,\dots,n}$ be the set of data where X_i represent the i-th multivariate functional object and Y_i is its associated label. Let us introduce the model.

$$Z_i = T(X_i) \tag{5}$$

$$Y_i = g\left(\langle Z_i^1, \theta_1 \rangle, \dots, \langle Z_i^p, \theta_K \rangle\right) + \varepsilon_i, \tag{6}$$

where $T : (L^2[0,1])^p \to (L^2[0,1])^p$ is a transformation of the multivariate functional data, θ_j are the indexes (functions of $L^2[0,1]$), $g : \mathbb{R}^{p \times K} \to \{0, \dots, C-1\}$ is the link function, C is the number of classes, $\langle ., . \rangle$ is the inner product of

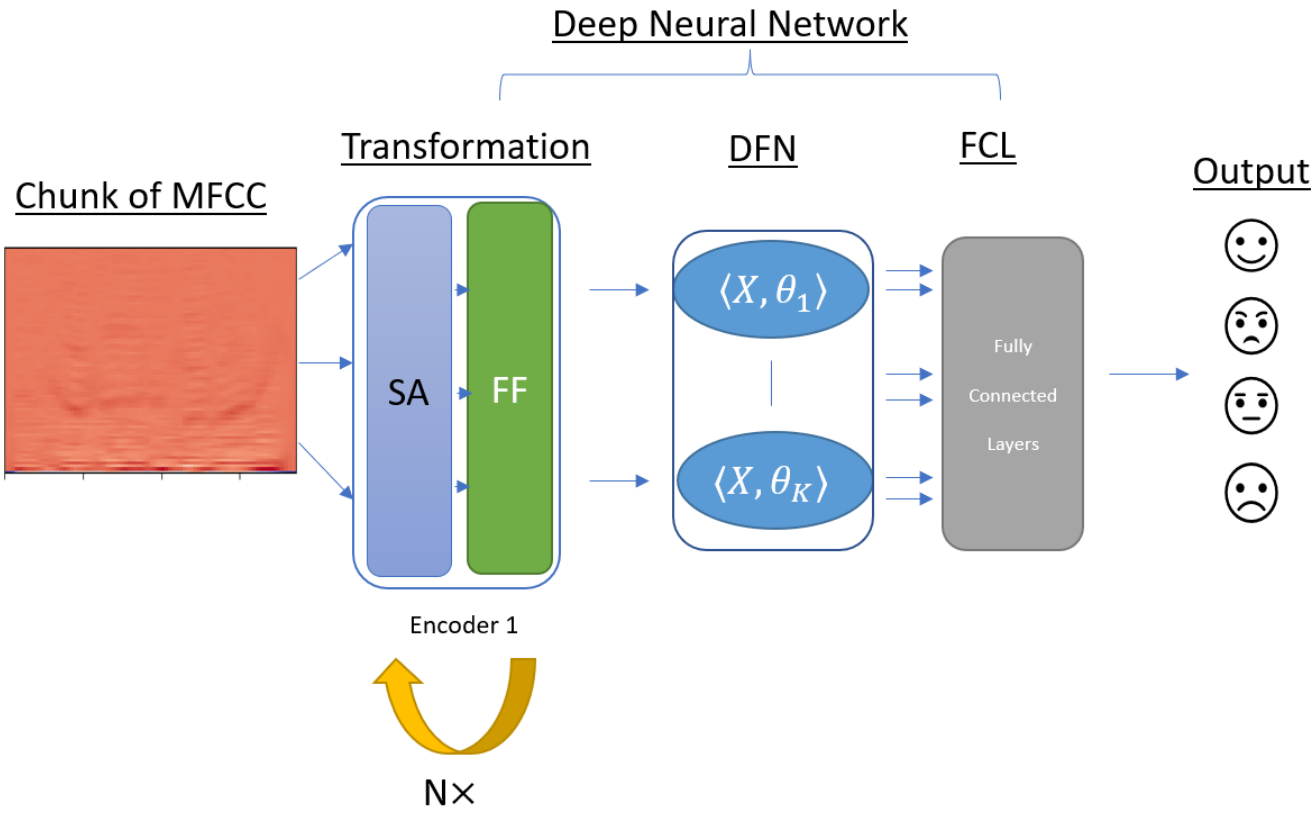

Fig. 1. Proposed method. SA: Self-Attention, FF: Feed-Forward network, DFN: Deep Functional Network, FCL: Fully Connected Layers.

$L^2[0,1]$ functions and the power j of the variable Z_i is the j-th component function of Z_i as Z_i is a multivariate functional data.

The unknown functions are the transformation T, the link function g and the indexes θ_j, $j = 1, \ldots, p \times K$. The transformation T is inferred by a transformer encoder architecture. Next, we apply an AdaFNN network to each functional variable, the output represents new features extracted from the MFCCs. Thus, the θ_j are estimated by a deep functional network (DFN) based on a concatenation of the AdaFNN module. Then, we apply a fully connected layer to estimate the link function g.

C. Deep neural network of the model

With a speech recording, we calculate the MFCC associated and cut it with an overlapping method. Then, we can feed the Deep Neural Network with this chunk of MFFC seen as a multivariate functional object. First, we make a transformation of the MFCC chunk. There is undoubtedly a link between the size of MFCC and the information about the emotion of the speech. But we do not know how is this link. So we apply a transformation T on the resized MFCC to accurately predict the emotion contained in the speech. The transformation module is made of a stack of N transformer encoders with self-attention and Feed-Forward network. The second module is the neural network of the paper [25] which is generalized to adapt to the multivariate context. In few words, the paper of [25] proposes to integrate the data by an adaptive function by a numerical scheme . So, the outputs of this second module are simply the L^2 product between each component of the transformed MFCC and an adaptive function. The third module of our network is a classical fully connected layer. The proposed method is represented globally in Figure 1.

The three modules of our new network architecture are:

1) The transformation module made of N transformer encoders.
2) The Deep Functional Network that outputs the scores of multivariate functional variable.
3) A Fully Connected Layers to classify the emotions.

D. Simulations

In order to show the ability of our proposed method, we study three different scenarios. For completing the study on other potential models using adaptive layer, we refer to the simulation section of [25]. We simulate four functional covariates $X^{(j)}$ $j = 1, 2, 3, 4$ coming from four different processes: exponential variogram ($j = 1$), Brownian ($j = 2$), Fractional Brownian ($j = 3$) and Gaussian process with Matérn covariance function. These curves are evaluated at 30 equally-spaced time points from $[0; 1]$. The unknown parameter functions are $\beta_1(t) = 5\sin(2\pi t)$, $\beta_2(t) = 5\sin(3\pi t)$, $\beta_3(t) = 3\cos(2\pi t)$ and $\beta_4(t) = 3\cos(3\pi t)$. A Gaussian error of variance 0.04

has been added to the regression term. Let us introduce the three different scenarios:

$$(S1)\, Y_i = g\left(\sum_{j=1}^{4}\langle\beta_1, X_i^{(j)}\rangle, \sum_{j=1}^{4}\langle\beta_2, X_i^{(j)}\rangle\right) + \varepsilon_i$$

with

$$g(a,b) = a^2 + b^2.$$

$$(S2)\, Y_i = g\left(\sum_{j=1}^{4}\langle\beta_1, X_i^{(j)}\rangle, \cdots, \sum_{j=1}^{4}\langle\beta_4, X_i^{(j)}\rangle\right) + \varepsilon_i$$

with

$$g(a,b) = a^2 + b^2 + c^2 + d^2.$$

$$\begin{aligned}(S3)\, Y_i =&(\langle\beta_1, \tilde{X_i}^{(1)}\rangle + \langle\beta_2, \tilde{X_i}^{(2)}\rangle + \\ &\langle\beta_3, X_i^{(3)}\rangle \times \langle\beta_4, X_i^{(4)}\rangle)^2 + \\ &(\langle\beta_1, \tilde{X_i}^{(1)}\rangle \times \langle\beta_2, \tilde{X_i}^{(2)} + \langle\beta_3, X_i^{(3)}\rangle + \langle\beta_4, X_i^{(4)}\rangle)^2 + \varepsilon_i\end{aligned}$$

TABLE I
PERFORMANCE ON SIMULATIONS. WE REPORT RMSE.

Scenario	RMSE
$S(1)$	0.085
$S(2)$	0.074
$S(3)$	0.031

To make an evaluation of the transformation part, we change $X^{(1)}$ by $\tilde{X}^{(1)} = (X^{(1)})^2$ and take $\tilde{X}^{(2)} = |X^{(2)}|$ in the last scenario. We do not have the code of the paper [24], so we could not compare to their method. In table I, the simulations reflect a good behaviour of our approach, even with complex behaviour and transformations in variables.

IV. APPLICATION TO SER

A. Database

The dataset IEMOCAP [34] contains approximately 12 hours of speech from 10 speakers. The literature selects a total of 5531 utterances that are labeled with one of five categories: Happy, Angry, Neutral, Sad, Exited. This set is reduced to four categories by merging Exited and Happy into a single category.

For evaluating our method, we use the protocol designed in [35]. Namely, we perform a 10-fold Speaker Independent cross-validation with one speaker as test, eight as training and one as validation set.

B. Implementation and hyperparameters

The code is written in python and use the pytorch library including its modules. We use the Adam optimizer with a learning rate of 3×10^{-4} and a focal loss with the L^2 penalty on the basis parameter. We set the batch size to 32 with 15 epoch and return the best model on the validation set. We do not tune the number of basis K, we use the best result of [25], namely 4 basis functions. We set the number of MFCC p to 40. And, we make chunks of the MFCC with a duration of 64. So, finally, the size of the final MFCC is $(64, 40)$. We use $N = 2$ transformer encoder layers. We use in the DFN of the basis layers an hidden FF of three connected layers of 128 each. The subsequent FCL network is two dense layer with an $\tanh$ activation function, dropout of 0.2, follow up by a projection to the 4 classes.

C. Results

We calculate two metrics namely WA for weighted accuracy (overall accuracy) and UA for unweigthed accuracy (average of the recall). We compare with [2], [36] which are the best results on IEMOCAP with four emotions and speech only, see Table II. It is worth mentioning that [36] do not only use MFCC but also Zero-crossing rate (ZCR), root mean square (RMS), Mel vector, chroma. And [2] use MFCC and first and second order frame-to-frame difference.

TABLE II
PERFORMANCE OF DIFFERENT APPROACHES ON THE IEMOCAP TESTING SET. WE REPORT WA AND UA.

Method	WA	UA
GRU [36]	64.95	.
AUDIO-CNN [2]	66.6	68.4
DFN (ours)	**57.27**	**56.43**

The results cannot be compared directly because we make our analysis on chunk-level. We think that the performance of our approach with the protocol of [35] can be improve by considering all the chunks of the audio record considered.

V. DISCUSSION

In this method, we choose to make chunks with an overlap of 25% of the duration of the chunk. Recently, [16] and [18] choose dynamically the percentage of overlapping with a significant improvement in accuracy. We can enhance our method by dynamically choosing the chunk overlapping by using the method in [16] and [18]. Moreover, the features can be passed to a recurrent neural network (RNN) (like LSTM, Bi-LSTM, GRU) to make a global decision on the whole audio input. Adding a RNN on top of the model may enhance the results, and that can be done in a two-stage or end-to-end learning.

VI. CONCLUSION

In conclusion, the article presents a promising advancement in SER using a novel model. The results, validated through simulations and tested on the IEMOCAP dataset at the

chunk-level, demonstrate satisfactory performance. The model leverages new features extracted from MFCC and relies on functional data, showcasing an innovative approach to emotion detection in speech. This development not only contributes to the ongoing evolution of SER methodologies but also highlights the potential for further exploration of functional data in enhancing emotion recognition systems.

REFERENCES

[1] B. F. P. Dossou and Y. K. S. Gbenou, "Fser: Deep convolutional neural networks for speech emotion recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, October 2021, pp. 3533–3538.

[2] Z. Peng, Y. Lu, S. Pan, and Y. Liu, "Efficient speech emotion recognition using multi-scale CNN and attention," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, jun 2021.

[3] W. Zhu and X. Li, "Speech emotion recognition with global-aware fusion on multi-scale feature representation," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, may 2022.

[4] L. Goncalves and C. Busso, "Improving speech emotion recognition using self-supervised learning with domain-specific audiovisual tasks," *Proc. Interspeech 2022*, pp. 1168–1172, 2022.

[5] E. Morais, R. Hoory, W. Zhu, I. Gat, M. Damasceno, and H. Aronowitz, "Speech emotion recognition using self-supervised features," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, may 2022.

[6] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*. Springer New York, 2005.

[7] F. Ferraty and P. Vieu, *Nonparametric functional data analysis: theory and practice*. Springer Science; Business Media, 2006.

[8] S. Robbiano, M. Saumard, and M. Curé, "Improving prediction performance of stellar parameters using functional models," *Journal of Applied Statistics*, vol. 43, no. 8, pp. 1465–1476, 2016.

[9] W. Saeys, B. De Ketelaere, and P. Darius, "Potential applications of functional data analysis in chemometrics," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 22, no. 5, pp. 335–344, 2008.

[10] R. Cao, L. Horváth, Z. Liu, and Y. Zhao, "A study of data-driven momentum and disposition effects in the chinese stock market by functional data analysis," *Review of Quantitative Finance and Accounting*, vol. 54, no. 1, pp. 335–358, 2020.

[11] D. Bosq, "Estimation of mean and covariance operator of autoregressive processes in banach spaces," *Statistical Inference for Stochastic Processes*, vol. 5, no. 3, pp. 287–306, 2002.

[12] D. Chen and H.-G. Müller, "Nonlinear manifold representations for functional data," *The Annals of Statistics*, vol. 40, no. 1, pp. 1–29, 2012.

[13] J. P. Arias, C. Busso, and N. B. Yoma, "Energy and F0 contour modeling with functional data analysis for emotional speech detection," in *Proc. Interspeech 2013*, 2013, pp. 2871–2875.

[14] ——, "Shape-based modeling of the fundamental frequency contour for emotion detection in speech," *Computer Speech & Language*, vol. 28, no. 1, pp. 278–294, jan 2014.

[15] S. Tavakoli, D. Pigoli, J. A. D. Aston, and J. S. Coleman, "A spatial modeling approach for linguistic object data: Analyzing dialect sound variations across great britain," *Journal of the American Statistical Association*, vol. 114, no. 527, pp. 1081–1096, jul 2019.

[16] W.-C. Lin and C. Busso, "An efficient temporal modeling approach for speech emotion recognition by mapping varied duration sentences into fixed number of chunks," *Interspeech 2020*, 2020.

[17] P. Kumawat and A. Routray, "Applying tdnn architectures for analyzing duration dependencies on speech emotion recognition." in *Interspeech*, 2021, pp. 3410–3414.

[18] W.-C. Lin and C. Busso, "Chunk-level speech emotion recognition: A general framework of sequence-to-one dynamic temporal modeling," *IEEE Transactions on Affective Computing*, vol. 14, no. 2, pp. 1215–1227, 2023.

[19] F. Jiang, S. Baek, J. Cao, and Y. Ma, "A functional single-index model," *Statistica sinica*, vol. 30, no. 1, pp. 303–324, 2020.

[20] C.-R. Jiang and J.-L. Wang, "Functional single index models for longitudinal data," *The Annals of Statistics*, vol. 39, no. 1, pp. 362–388, 2011. [Online]. Available: http://www.jstor.org/stable/29783641

[21] F. Ferraty, J. Park, and P. Vieu, "Estimation of a functional single index model," in *Recent advances in functional data analysis and related topics*. Springer, 2011, pp. 111–116.

[22] H.-G. Müller and F. Yao, "Functional additive models," *Journal of the American Statistical Association*, vol. 103, no. 484, pp. 1534–1544, 2008.

[23] R. K. Wong, Y. Li, and Z. Zhu, "Partially linear functional additive models for multivariate functional data," *Journal of the American Statistical Association*, vol. 114, no. 525, pp. 406–418, 2019.

[24] A. R. Rao and M. Reimherr, "Nonlinear functional modeling using neural networks," *Journal of Computational and Graphical Statistics*, pp. 1–10, 2023.

[25] J. Yao, J. Mueller, and J.-L. Wang, "Deep learning for functional data analysis with adaptive basis layers," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 11 898–11 908. [Online]. Available: https://proceedings.mlr.press/v139/yao21c.html

[26] F. Rossi, B. Conan-Guez, and F. Fleuret, "Functional data analysis with multi layer perceptrons," in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, vol. 3. IEEE, 2002, pp. 2843–2848.

[27] Q. Wang, S. Zheng, A. Farahat, S. Serita, T. Saeki, and C. Gupta, "Multilayer perceptron for sparse functional data," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–10.

[28] J. Kim, Y. An, and J. Kim, "Improving speech emotion recognition through focus and calibration attention mechanisms," in *Proc. Interspeech 2022*, 2022, pp. 136–140.

[29] N. R. Prabhu, G. Carbajal, N. Lehmann-Willenbrock, and T. Gerkmann, "End-to-end label uncertainty modeling for speech-based arousal recognition using bayesian neural networks," in *Proc. Interspeech 2022*, 2022, pp. 151–155.

[30] H. Dhamyal, B. Raj, and R. Singh, "Positional encoding for capturing modality specific cadence for emotion detection," in *Proc. Interspeech 2022*, 2022, pp. 166–170.

[31] M. Perez, M. Jaiswal, M. Niu, C. Gorrostieta, M. Roddy, K. Taylor, R. Lotfian, J. Kane, and E. M. Provost, "Mind the gap: On the value of silence representations to lexical-based speech emotion recognition," in *Proc. Interspeech 2022*, 2022, pp. 156–160.

[32] E. Vaaras, M. Airaksinen, and O. Räsänen, "Analysis of self-supervised learning and dimensionality reduction methods in clustering-based active learning for speech emotion recognition," in *Proc. Interspeech 2022*, 2022, pp. 1143–1147.

[33] M. Baruah and B. Banerjee, "Speech emotion recognition via generation using an attention-based variational recurrent neural network," in *Proc. Interspeech 2022*, 2022, pp. 4710–4714.

[34] 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," *Language resources and evaluation*, vol. 42, no. 4, pp. 335–359, 2008.

[35] N. Antoniou, A. Katsamanis, T. Giannakopoulos, and S. Narayanan, "Designing and evaluating speech emotion recognition systems: A reality check case study with iemocap," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[36] J.-K. H. Hyun-Sam Shin, "Performance analysis of a chunk-based speech emotion recognition model using rnn," *Intelligent Automation & Soft Computing*, vol. 36, no. 1, pp. 235–248, 2023. [Online]. Available: http://www.techscience.com/iasc/v36n1/50034