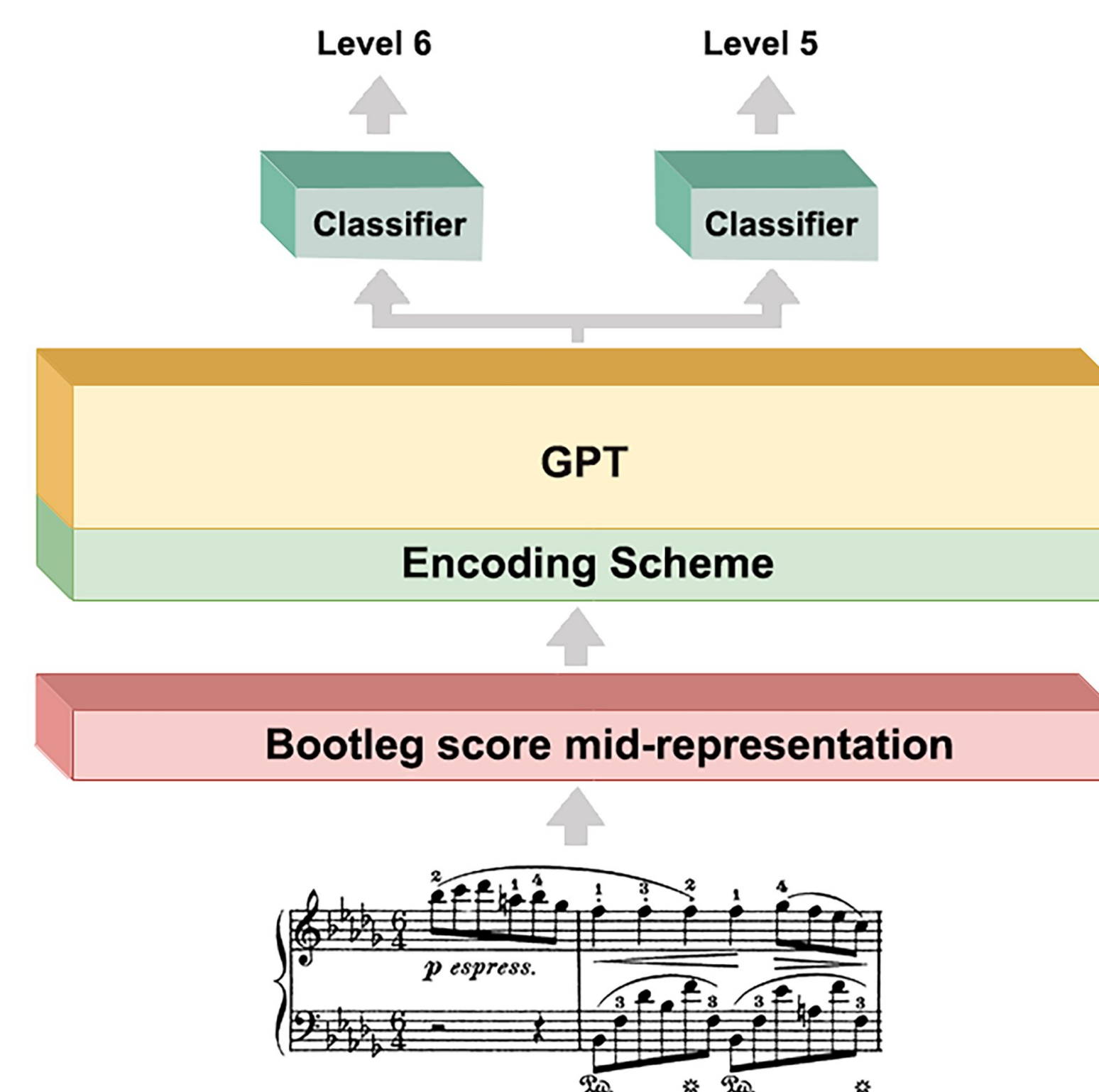


# PREDICTING PERFORMANCE DIFFICULTY FROM PIANO SHEET MUSIC IMAGES

Pedro Ramoneda<sup>1</sup>, Jose J. Valero-Mas<sup>2</sup>, Dasaem Jeong<sup>3</sup>, Xavier Serra<sup>1</sup>  
<sup>1</sup> Music Technology Group, Universitat Pompeu Fabra, Barcelona  
<sup>2</sup> University of Alicante, Alicante  
<sup>3</sup> MALer Lab, Department of Art & Technology, Sogang University, Seoul

Multi-task GPT-based framework to predict the performance difficulty associated to a piano score directly from sheet images.



## Motivation

- Sheet music images **expand** music selection, **preserve** untranscribed cultural heritage, and **offer practicality** due to scarce machine-readable scores.
- **Promoting diversity** in piano repertoires: enriching music education by highlights overlooked groups (e.g., female composers)

## Data

- Five datasets with **more than 7500 scores** with up to 9 difficulty levels.
- Two of them particularly compiled for this work.

Dataset	MK	CIPI	PS	FS	HV
Pieces	147	652	2816	4193	17

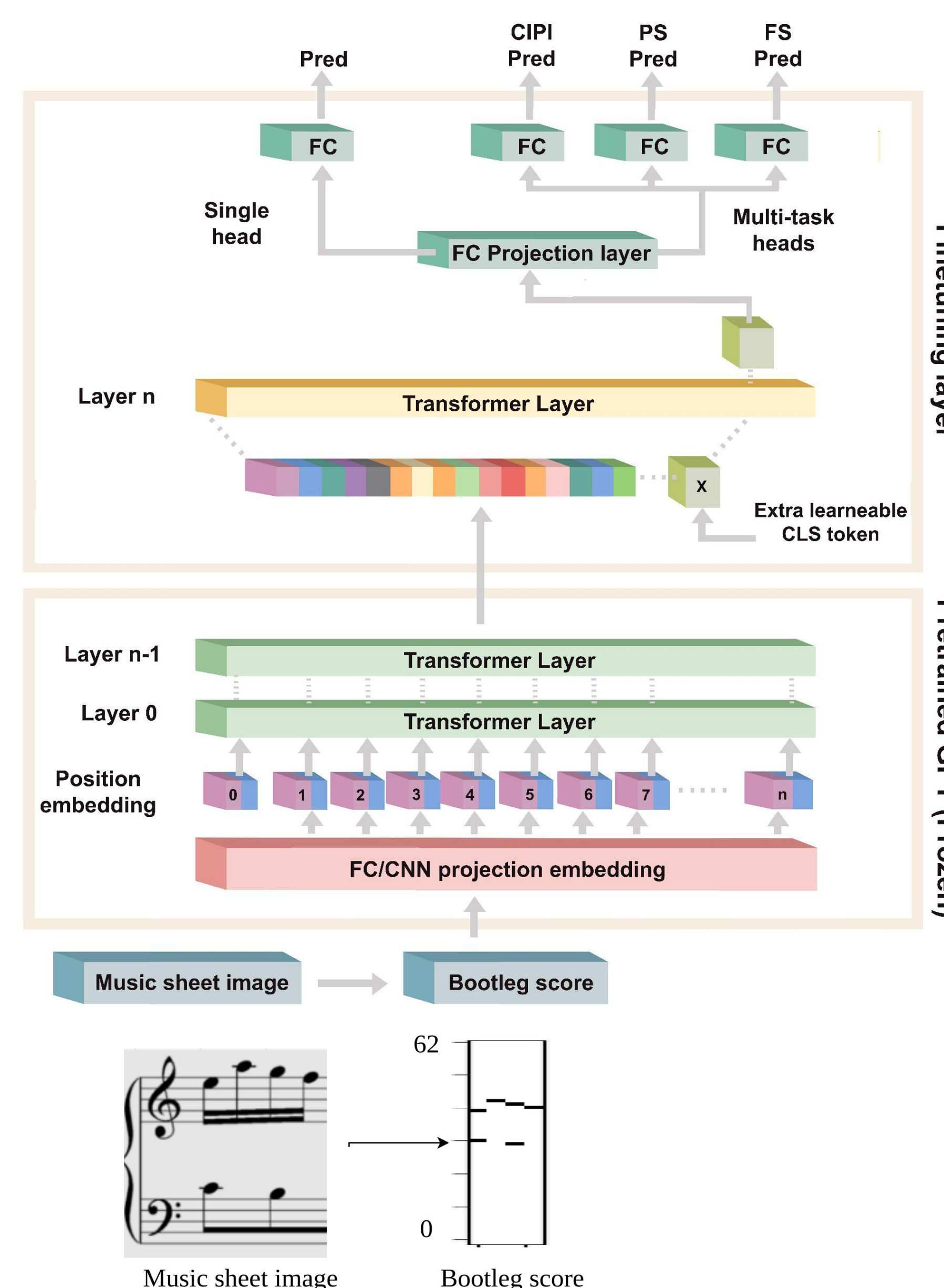
## Methodology

Building upon prior research, we utilize the GPT model combined with the bootleg score for sheet music analysis. We first pretrain a **GPT model on the IMSLP** collection to fine-tune it for difficulty recognition later.

We adapted the previous research [1] to the requirement of processing the whole music pieces:

- Encoding scheme that reduces the encoded sequence length to an eighth of the original size.
- Multi-hot optimization for GPT pretraining and transitioning from categorical encoding to causal convolutional or feedforward projection layers.

We also predict the multiple difficulty rankings simultaneously.



## Results

Single-task VS Multi-task comparison

Encoding	Acc <sub>0</sub> (%)	MSE
<b>GPT (FC) MULTI-TASK</b>		
CIPI	<b>40.3</b>	<b>1.3</b>
PS	<b>35.9</b>	<b>1.9</b>
FS	45.8	<b>0.8</b>
<b>GPT (FC) SINGLE-TASK</b>		
CIPI	34.3	1.6
PS	30.9	2.1
FS	<b>46.6</b>	0.8

Comparison with SoTA/other approaches

Case	Acc (%)	MSE
<i>Symbolic</i> [2]		
GRU+Att	39.5	<b>1.1</b>
<i>Tsat et al.</i> [1]		
GPT <sub>EMB</sub>	19.7	3.3
<i>Proposal</i>		
GPT <sub>FC</sub>	34.3	1.6
GPT <sub>CNN</sub>	36.2	1.4
GPT <sub>FC</sub> <sup>multi</sup>	<b>40.3</b>	1.3

- We observed a performance gap on HV when zero shooting in MK ad HV dataset.
- We must also acknowledge that most composers used for training are white males, and the HV results are significantly worse than the rest of the datasets.

PAPER



CODE



DATA



DEMO



[1] D. Yang and T. Tsai, **Composer classification with cross-modal transfer learning and musically-informed augmentation**, in Proceedings of the 22nd International Society for Music Information Retrieval Conference, ISMIR, Online, 2021, pp. 802–809.

[2] Pedro Ramoneda, Dasaem Jeong, Vsevolod Eremenko, Nazif Can Tamer, Marius Miron, Xavier Serra. **Combining piano performance dimensions for score difficulty classification**, Expert Systems with Applications, 2023.