Supplementary Material B: Feature Descriptions

Feature	Description
Duration	Length of piece in seconds, from end time of last note (`pretty_midi.get_end_time`).
Mean tempo	Mean of all tempo values recovered from tempo-change events (`get_tempo_changes`).
Note count	Total number of note events, summed across all instrument tracks.
Track count	Number of instrument tracks (`pretty_midi` instruments).
Unique instrument count	Number of distinct MIDI program numbers, excluding drum tracks.
Note density	Note count divided by duration, giving notes per second.
Mean polyphony	Mean number of simultaneously sounding notes, computed by sweeping note on/off events and averaging the running count of active notes.
IOI std	Standard deviation of inter-onset intervals (the time gaps between successive note onsets).
Mean pitch	Mean MIDI pitch number across all notes.
Pitch range	Difference between the highest and lowest MIDI pitch.
Pitch entropy	Shannon entropy (bits) of the 12-bin pitch-class histogram, measuring how evenly the twelve pitch classes are used.
Velocity mean	Mean MIDI note velocity (loudness), on a 0–127 scale.
Velocity std	Standard deviation of note velocities.
Velocity entropy	Shannon entropy (bits) of the 128-bin velocity histogram, measuring the spread of dynamics.

Table S5: Symbolic (MIDI) features, extracted with `pretty_midi`.

Feature	Description
Duration	Length of the audio in seconds (number of samples divided by the sample rate).
RMS energy (mean, std)	Mean and standard deviation of the frame-level root-mean-square energy, a measure of loudness (`librosa.feature.rms`).
Spectral centroid (mean, std)	Mean and standard deviation of the spectral centroid, the "centre of mass" of the spectrum and a correlate of perceived brightness (`librosa.feature.spectral_centroid`).
Spectral bandwidth (mean, std)	Mean and standard deviation of the spectral bandwidth, the spread of the spectrum about its centroid (`librosa.feature.spectral_bandwidth`).
Zero-crossing rate (mean)	Mean rate at which the waveform changes sign, higher for noisier or more percussive material (`librosa.feature.zero_crossing_rate`).
MFCCs 1–13	Mean of each of the first 13 Mel-frequency cepstral coefficients over all frames, summarising spectral timbre (`librosa.feature.mfcc`, `n_mfcc=13`).

Table S6: Audio (original recording) features, extracted with `librosa`. Audio was decoded to mono and resampled before extraction.

Feature	Description
Aspect ratio	Image width divided by height.
Mean/std hue	Mean and standard deviation of the hue channel in HSV colour space, describing dominant colour and colour variability.
Mean/std saturation	Mean and standard deviation of the saturation channel, describing colour vividness and its spread.
Mean/std brightness	Mean and standard deviation of the value (brightness) channel.
Colourfulness	Perceived colourfulness computed with the Hasler and Süsstrunk (2003)† metric, which combines the mean and spread of the opponent-colour channels (red-green and yellow-blue) into a single score; higher values indicate more vivid, varied colour.
Contrast	RMS contrast, the standard deviation of pixel intensities in greyscale.
Image entropy	Shannon entropy (bits) of the greyscale intensity histogram, higher for busier, more detailed images.
Edge density	Proportion of pixels flagged as edges by the `cv2.Canny` detector, a measure of visual clutter or line work.

Table S7: Visual (box art) features, extracted with OpenCV. Each image was resized to 300 px width before extraction to make scale-dependent metrics comparable across scans.

† Hasler, D., & Suesstrunk, S. E. (2003). Measuring colorfulness in natural images. In B. E. Rogowitz & T. N. Pappas (Eds), *Human Vision and Electronic Imaging VIII*. doi:10.1117/12.477378