

Table 2. Dataset summary

Format	Name	Modality			Applicable task	Size	Access
		Score	Perfor- mance	Audio			
MIDI	JSB Chorus	✓			Polyphonic	402 Bach four parts chorus	Music21toolkit[313]
	VGMIDI	✓	✓		Polyphonic with senti- ment	823 piano video game sound- tracks	Derived from [307]
	Lakh MIDI Dataset	✓	✓		Multi- instrumental	176,581MIDI files	http://colinraffel.com/projects/lmd/
	Projective Orches- tral Database	✓			Orchestral	392 MIDI files grouped in pairs containing a piano score and its orchestral version	https://qsdfio.github.io/LOP- /database
	e-Piano Competi- tion Dataset	✓	✓		Polyphonic & Performance	~1400 MIDI files of piano per- formance	http://www.piano-e- competition.com
	BitMidi	✓			Polyphonic	113,244 MIDI files curated by volunteers around the world	https://bitmidi.com/
	Classical Archives	✓			Polyphonic	Maximum number of MIDI files of free classical music	https://www.classical- archives.com/
	The largest MIDI dataset on the Inter- net	✓			Polyphonic & Style	About 130,000 pieces of mu- sic from 8 distinct genres (clas- sical, metal, folk, etc.)	http://stoneyroads.com/20- 15/06/behold-the-worlds- biggest-midicollection-on- the-internet/
	ADL Piano MIDI	✓	✓		Polyphonic	11,086 unique piano MIDI files	https://github.com/lucasnfe/- adl-piano-midi
	GiantMIDI-Piano	✓	✓		Polyphonic	10,854 MIDI files of classical piano, 1,237 hours in total	https://github.com/byte- dance/GiantMIDI-Piano
MusicXML	TheoryTab Data- base	✓			Polyphonic	16K lead sheet segments	https://www.hooktheory.- com/theorytab
	Hooktheory Lead Sheet dataset	✓			Polyphonic	11,329 lead sheet segments	Derived from [180]

Continued table 2

Format	Name	Modality			Applicable task	Size	Access
		Score	Performance	Audio			
	Wikifonia	✓			Polyphonic	2,252 western music lead sheets	http://marg.snu.ac.kr/chord_generation/(CSV format)
	MuseScore lead sheet dataset	✓	✓		Performance	lead sheet corresponding to Yamaha e-Competitions MIDI dataset	https://musescore.com
Pianoroll	Lakh Pianoroll Dataset	✓	✓		Multi-instrumental	Approximately equal to the size of LMD	https://salu133445.github.io/musegan/
Text	Nottingham Music Dataset	✓			Monophonic	About 1,000 folk songs	abc.sourceforge.net/NMD/
	ABC tune book of Henrik Norbeck	✓			Monophonic	More than 2,800 scores and lyrics in ABC format, mainly Irish and Swiss traditional music	http://www.norbeck.nu/abc/
	ABC version of FolkDB	✓			Monophonic	Unknown	https://thesession.org/
	KernScores	✓			Polyphonic	Over 700 million notes in 108,703 files	http://kern.humdrum.org
Audio	NSynth Dataset			✓	Music audio	306,043 notes	https://magenta.tensorflow.org/datasets/nsynth
	FMA dataset			✓	Music audio	106,574 tracks of 917GiB	Derived from cite212
	Minist musical sound dataset			✓	Music audio	50,912 notes	https://github.com/ejhumphrey/minst-dataset/
	GTZAN Dataset			✓	Music audio	1,000 30s music audios	http://marsyas.info/download/data_sets
	Studio On-Line (SOL)			✓	Music audio	120,000 sounds	Derived from [316]

Continued table 2

Format	Name	Modality			Applicable task	Size	Access
		Score	Perfor- mance	Audio			
	NUS Sung and Spoken Lyrics(NUS-48E) Corpus			✓	Sing Voice	169 minutes recordings of 48 English songs	Derived from [317]
Multi-modality	MusicNet Dataset	✓		✓	Fusion	330 recordings of classical music	https://homes.cs.washington.edu/~thickstn/musicnet.html
	MAESTRO Dataset	✓	✓	✓	Fusion	172 hours of virtuosic piano performances	https://g.co/magenta/-maestrodataset
	NES Music Database	✓		✓	Multi-instrumental	thousands of	Derived from [200]
	Piano-Midi	✓	✓	✓	Polyphonic & performance	332 classical piano pieces	www.piano-midi.de/
	Groove MIDI Dataset	✓	✓	✓	Drum	13.6 hours recordings, 1,150 MIDI files and over 22,000 measures of tempo-aligned expressive drumming	https://magenta.tensorflow.org/datasets/groove
	POP909	✓	✓	✓	Polyphonic	multiple versions of the piano arrangements of 909 popular songs	https://github.com/music-x-lab/POP909-Dataset
	ASAP	✓	✓	✓	Polyphonic Performance& Fusion	222 digital musical scores aligned with 1,068 performances	https://github.com/fosfrance-sco/asap-dataset
	Aligned lyrics-melody dataset	✓			Fusion	13,937 20-note sequences with 278,740 syllable-note pairs	https://github.com/yy1lab/Lyrics-Conditioned-Neural-Melody-Generation
	MTM Dataset	✓		✓	Fusion	Unknown	https://github.com/Morning-Books/MTM-Dataset