

Subset names for extended GiagMIDI Dataset	Link for each subset	Number of files	
Music Gateway MIDI drum loops	https://www.musicgateway.com/blog/how-to/best-free-midi-drum-loops/	336	
The Drum Percussion Midi Archive	https://www.reddit.com/r/WeAreTheMusicMakers/comments/9kz8qg/the_drum_percussion_midi_archive/	800,000	
Shredder Odd Groove Drum Loops	https://oddgrooves.com/free-metal-drum-loops/	36	
Metal-MIDI	https://drive.google.com/drive/u/0/folders/1ePDxWM	133	
Groove MIDI Dataset	https://magenta.tensorflow.org/datasets/groove#drum	45538	
MetaMIDI Dataset	https://zenodo.org/records/5142664#.YQN3c5NKgV	433,527	
Drum space	https://github.com/allsoph/drum_space	33000	
ABA MIDI Pack V002 Drum	https://drive.google.com/file/d/1EWQ3LadzV-piAA/	50000	
ABA MIDI Pack V003 Dance Music Drum	https://drive.google.com/file/d/13YM8aFgZYpWcPl	50000	
ABA Melody MIDI package V001	https://drive.google.com/file/d/1s4rwbT5W-TeGcw	50000	
Axel's Midi Collection	https://drive.google.com/file/d/1tm-CIKCEhtLcLMVC	2350	
MIDI Files with curated style metadata by a music researcher at doctoral level	Dr. Cale Plut	29598	
Expanded Groove MIDI	https://magenta.tensorflow.org/datasets/e-gmd	45,537	
MAESTRO Dataset	https://magenta.tensorflow.org/datasets/maestro	1276	
YM2413-MDB-v1.0.2-Game	https://github.com/jech2/ym2413-mdb	669	
NES-MDB	https://github.com/chrisdonahue/nsmdb	5281	
ComMU	https://pozalabs.github.io/ComMU/	11144	
ATEPP	https://zenodo.org/record/7182820#.ZA_1-C8r1bU	11742	
Melody Track Identification	https://github.com/maxichu/MelodyTrackIdentification	11625	
Form-NN	https://github.com/danielathome19/Form-NN	200	
ASAP Piano Dataset	https://github.com/fofrancesco/asap-dataset	1067	
Lo-fi Hip-hop MIDI	https://www.kaggle.com/datasets/zakari/lofi-hip-hop	93	
POP909	https://github.com/music-x-lab/POP909-Dataset	1000	
EMOPIA	https://zenodo.org/record/5257995	1,078	
SymphonyNet_Dataset	https://symphonynet.github.io/	46,360	
Symbolic orchestral database	https://qsdf0.github.io/LOP/database.html	5,711	
DadaGP	https://github.com/dada-bots/dadaGP	22,677	
Doug Jazz Piano McKenzie	https://bushgrafts.com/midi/	308	
Saarland Music Data	https://resources.mpi-inf.mpg.de/SMD/SMD_MIDI-A	51	
Pacific Coast Colombian Corpus provided by a music researcher in Colombia	Santiago Quesada	23	
Giant-piano MIDI	https://github.com/bytedance/GiantMIDI-Piano	10,855	
IsoVAT	Cale Plut (https://summit.sfu.ca/_flysystem/fedora/2)	93	
URMP Dataset	https://docs.google.com/forms/d/e/1FAIpQLSdGQx4	44	
Orchset	https://zenodo.org/records/1289786#.WzNu2hyxXM	65	
PHENICX-Anechoic	https://zenodo.org/records/1289821#.WzNuXRYxXl	22	
MID-FILD	https://pozalabs.github.io/MID-FILD_demo/	4,422	
PDMX	https://zenodo.org/records/13763756	445,640	222820 (Non-expressive) + 222820 (Expressive, augmented)
IrishMAN	https://huggingface.co/datasets/sander-wood/irishman	216,284	
JAZZVAR	https://zenodo.org/records/11264676	502	
SymMV	https://github.com/zhuole1025/SymMV/tree/main/data	1,141	
XMIDI	https://github.com/xmusic-project/XMIDI_Dataset	108,023	
PJAMA	https://zenodo.org/records/8354955	2,777	
PIAST	https://github.com/Hayeonbang/PIAST	11,696	
PercePiano	https://github.com/JonghoKimSNU/PercePiano?tab=readme	1,202	
Total (Before deduplication test)		2,463,126	
Total (After deduplication test)		2,136,218	