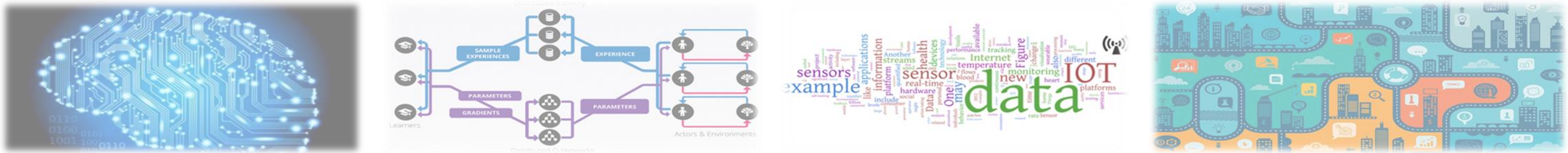


Musical Genre Classification

AutoML 연구실
석사 과정 조성현
21.12.28

To Do List

- Last Week
  - Evaluation metrics

- This Week
  - 코드 분석

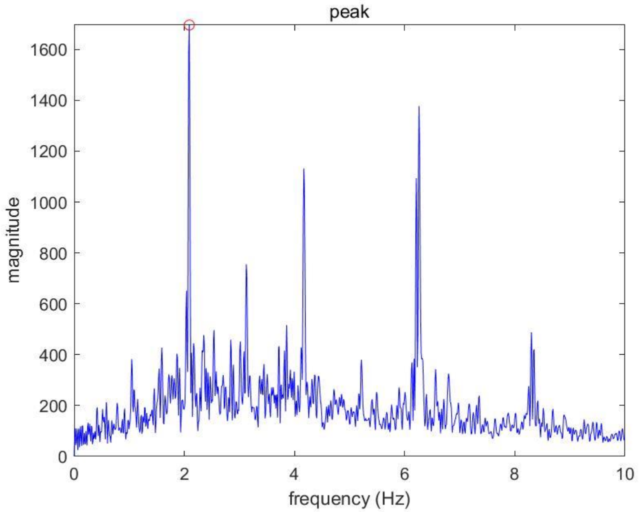
The standard statistics of the datasets

Author	Year	Neural Network	# of employed datasets	Employed Dataset	Input Type	Evaluation Metric	Ref.
Massoudi	2021	CNN	1	Urban Sound Classification Challenge	MFCC	Accuracy	[1]
Ghildiyal	2020	CNN	1	GTZAN	Mel-Spectrogram	Accuracy	[2]
Palanisamy	2020	CNN	3	ESC-50, GTZAN, UrbanSound8K	Log Mel-Spectrogram	Accuracy	[3]
Yu	2019	CNN	2	GTZAN, Extended Ballroom	STFT Spectrogram	Accuracy	[4]
Won	2020	CNN	3	MagnaTagATune, Million Song Dataset, MTG-Jamendo	Mel-Spectrogram	ROC-AUC, PR-AUC	[5]
Bisharad	2019	CNN	1	GTZAN	Mel-Spectrogram	Precision, Recall, F1 score	[6]
Zhang	2019	RNN	1	GTZAN	MFCC, chromagram	Accuracy	[7]
Chillara	2019	CNN	1	FMA-SMALL	Spectrogram	Accuracy	[8]
Zhang	2016	CNN	1	GTZAN	STFT Spectrogram	Accuracy	[9]
Choi	2017	CRNN	1	Million Song Dataset	Log Mel-Spectrogram	ROC-AUC	[10]
Su	2019	CNN	1	UrbanSound8k	Log Mel-Spectrogram, MFCC	Accuracy	[11]
Choi	2016	CNN	2	MagnaTagATune, Million Song Dataset	Mel-Spectrogram, MFCC, STFT	ROC-AUC	[12]
Li	2018	CNN	3	ESC-10, ESC-50, UrbanSound8k	Log Mel-Spectrogram	Accuracy	[13]

MIRToolbox

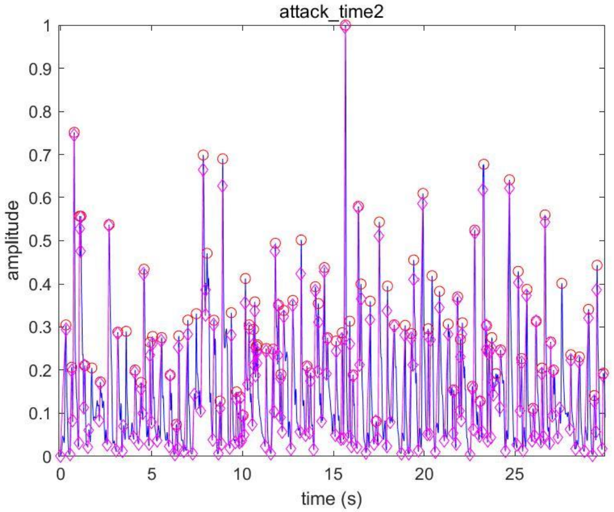
X_axis = frequency

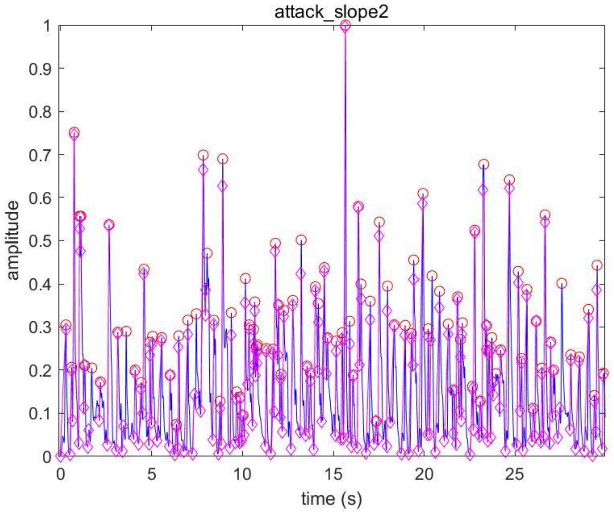
Fluctuation



X_axis = time

Rhythm

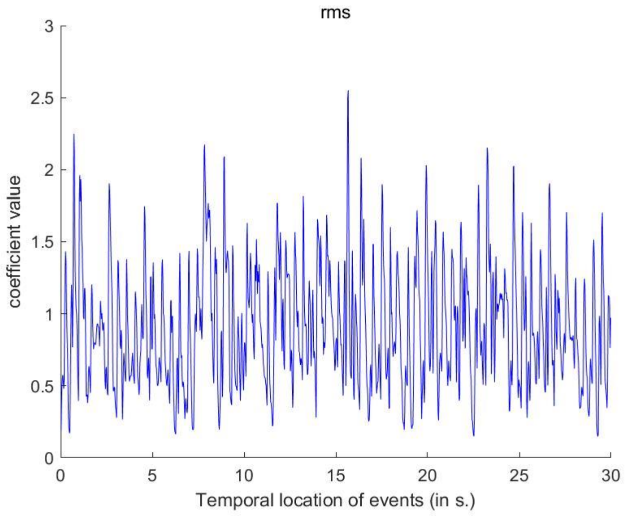




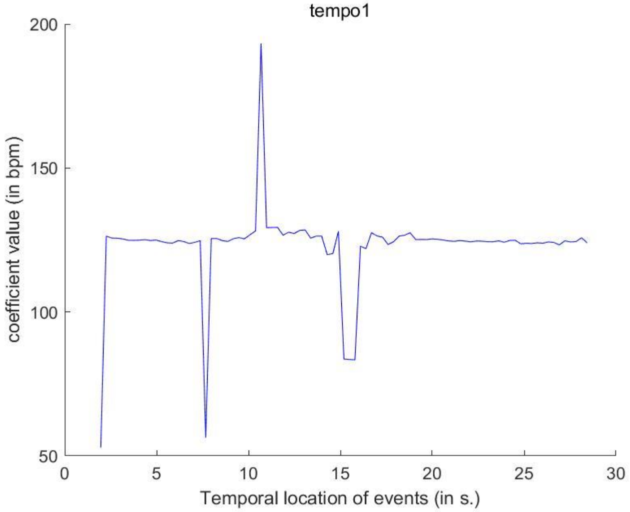
MIRToolbox

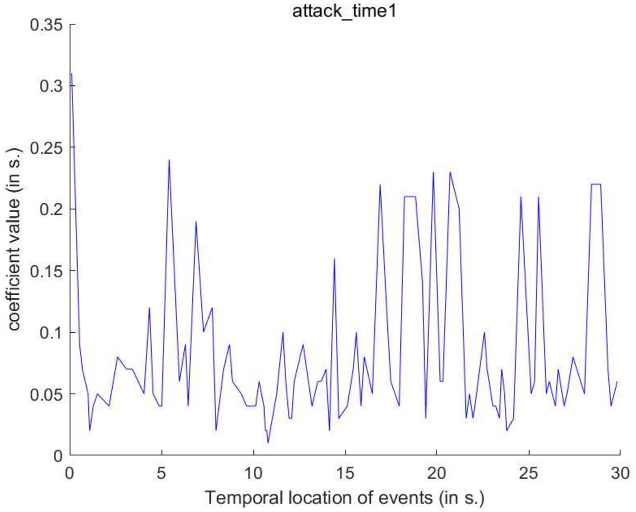
X_axis = Temporal location of events (in s.)

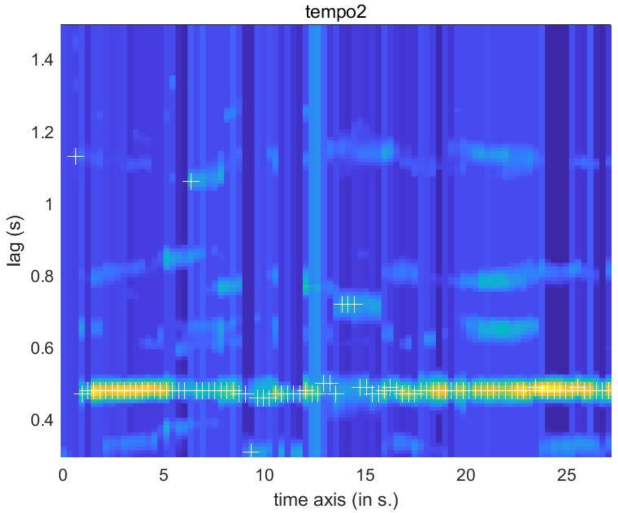
Dynamics

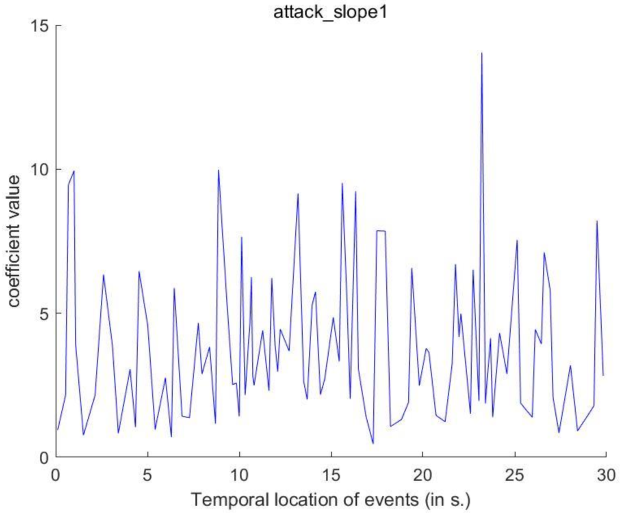


Rhythm





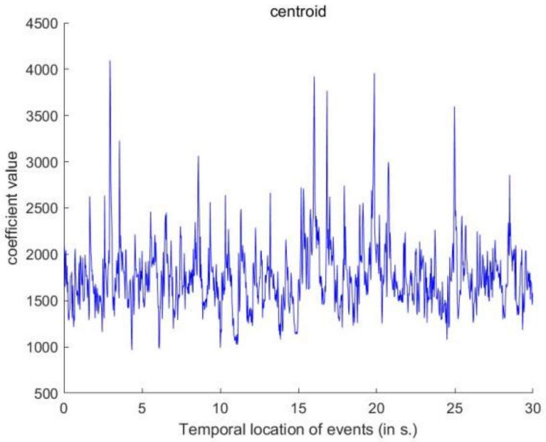


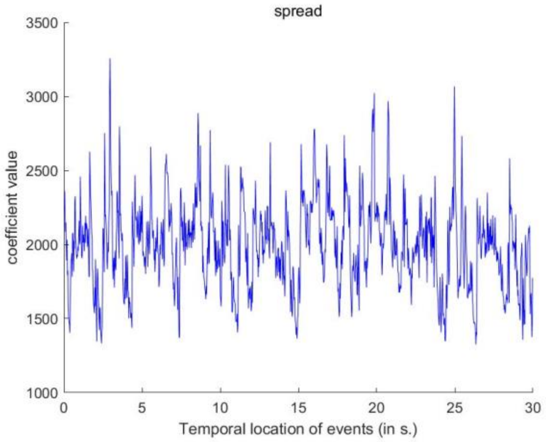


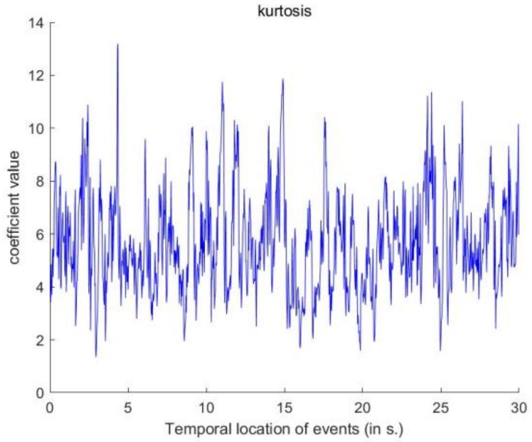
MIRToolbox

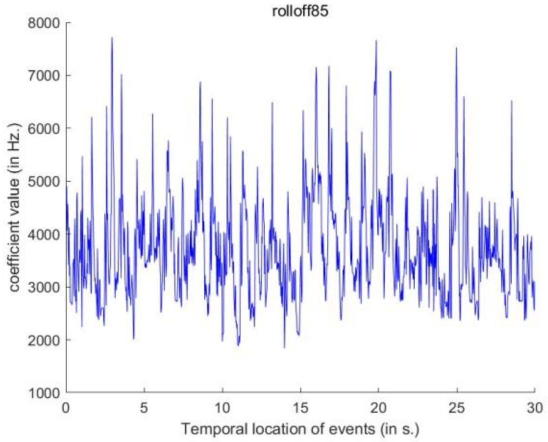
X_axis = Temporal location of events (in s.)

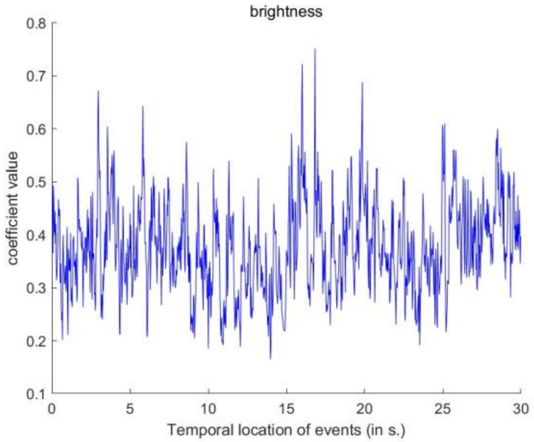
Spectral

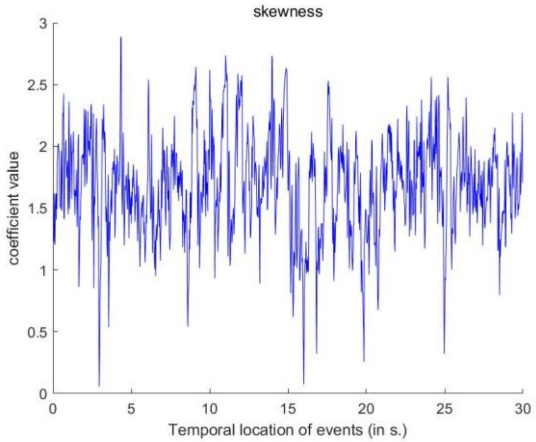


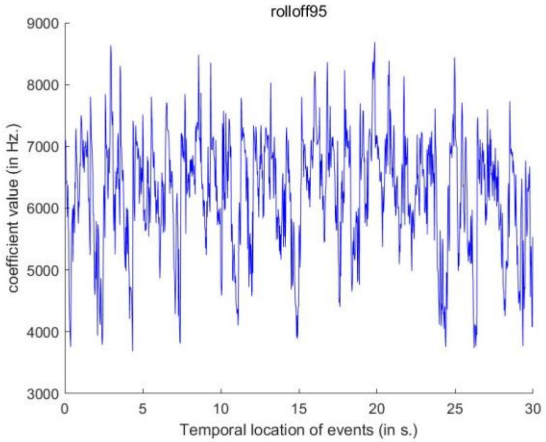








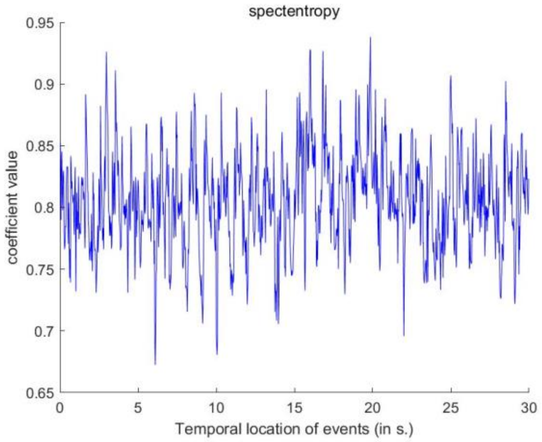


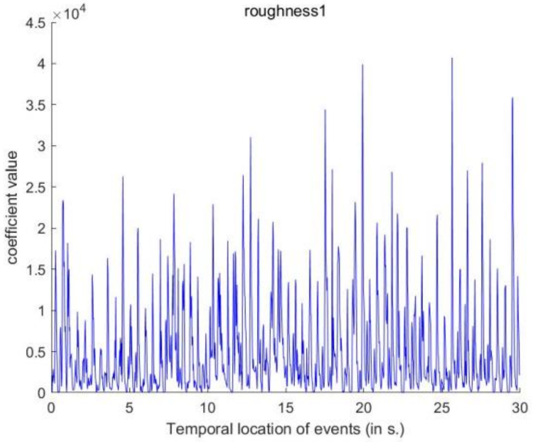


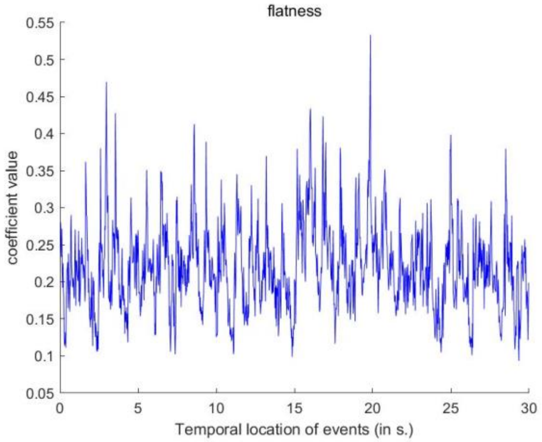
MIRToolbox

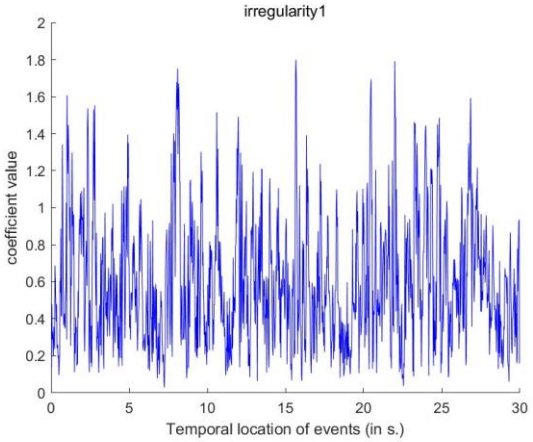
X_axis = Temporal location of events (in s.)

Spectral

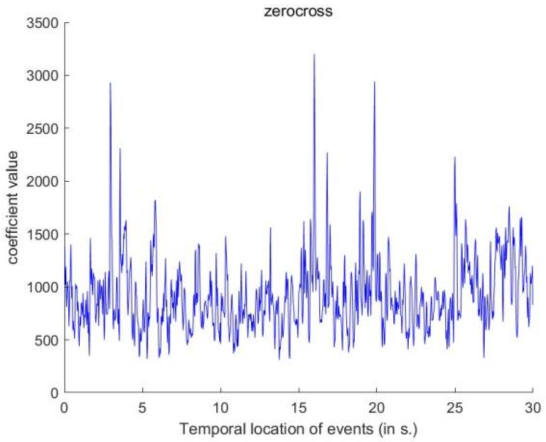


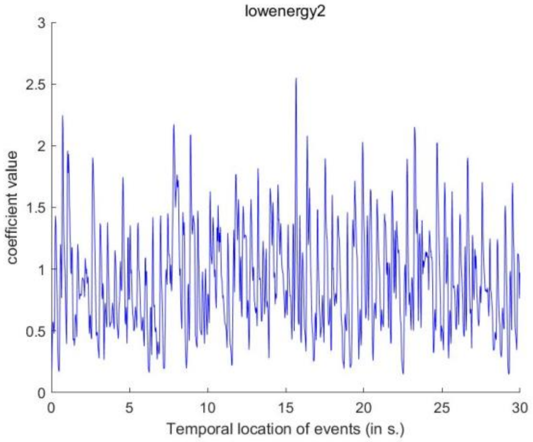


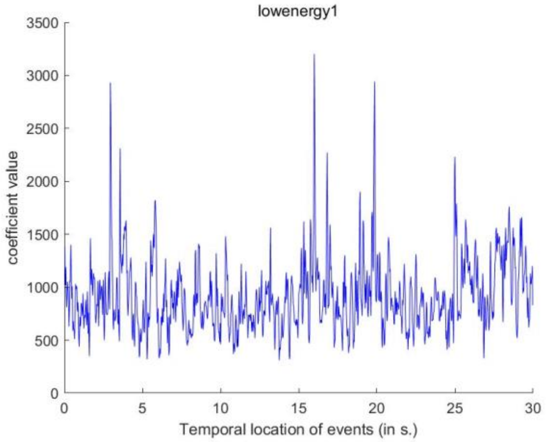


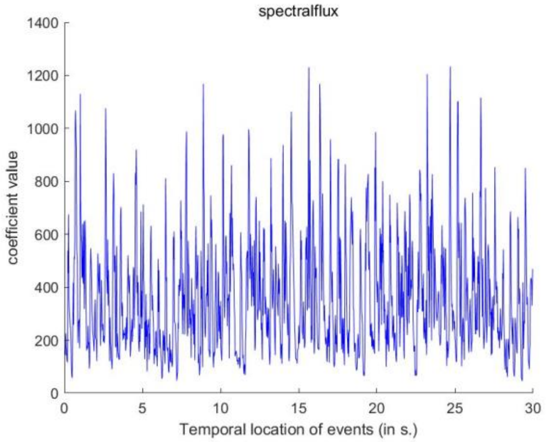


Timbre





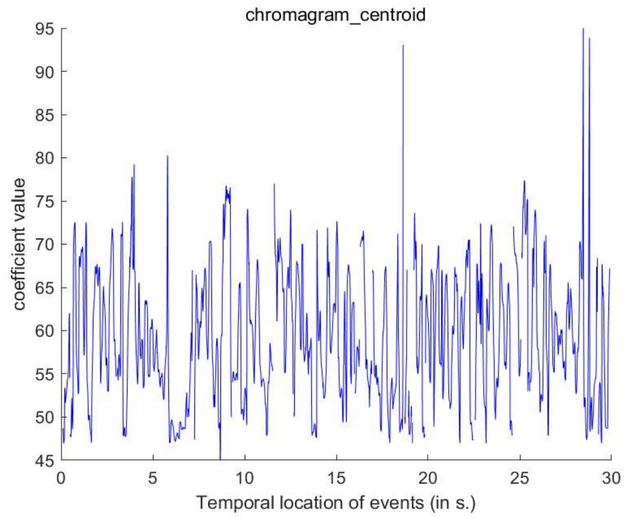


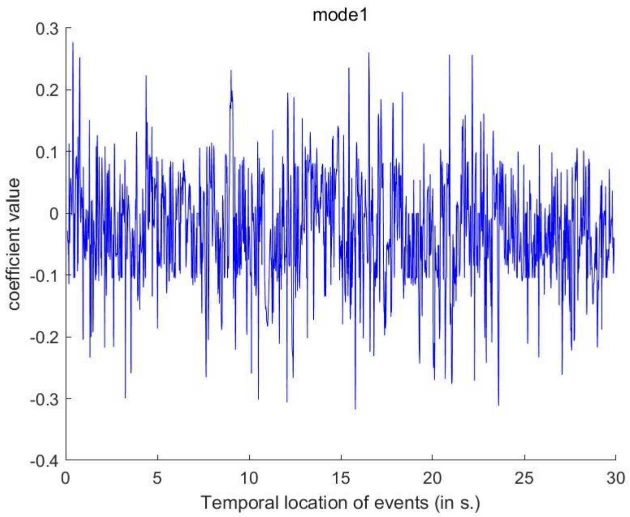


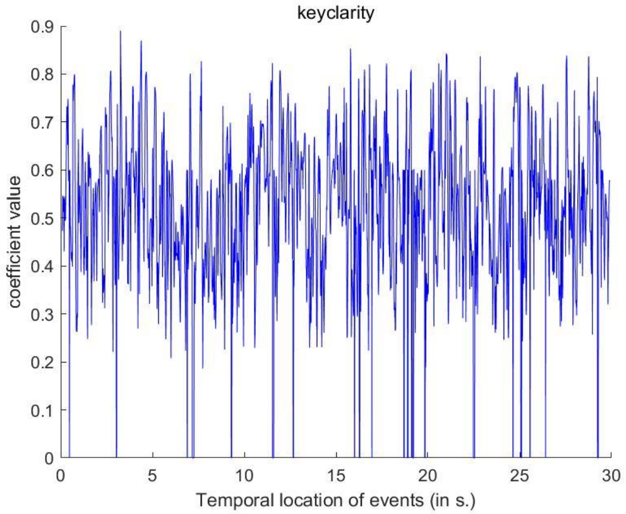
MIRToolbox

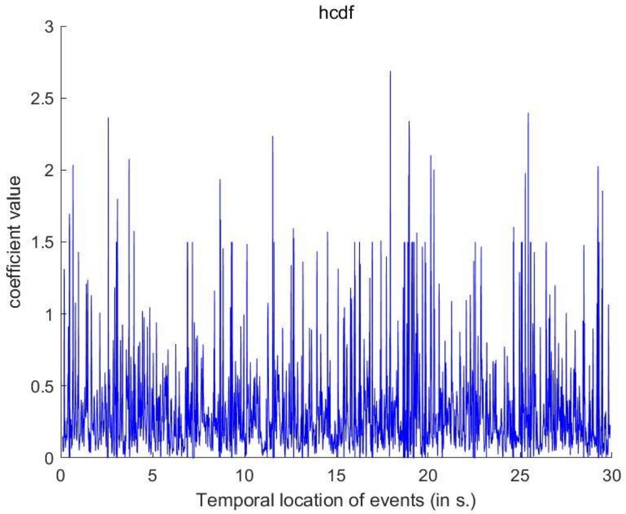
X_axis = Temporal location of events (in s.)

Tonal





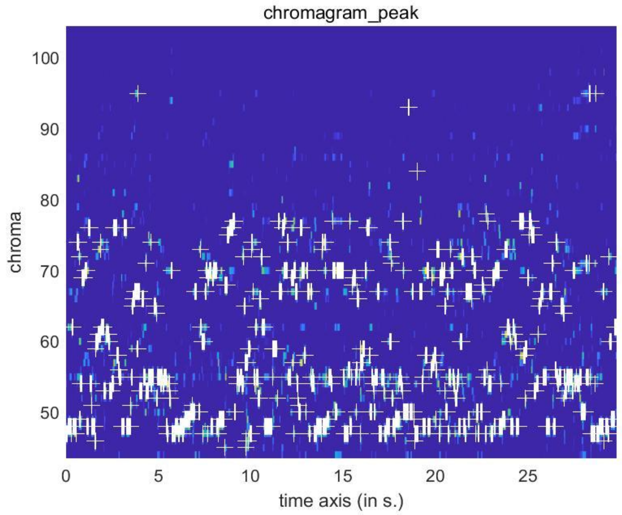


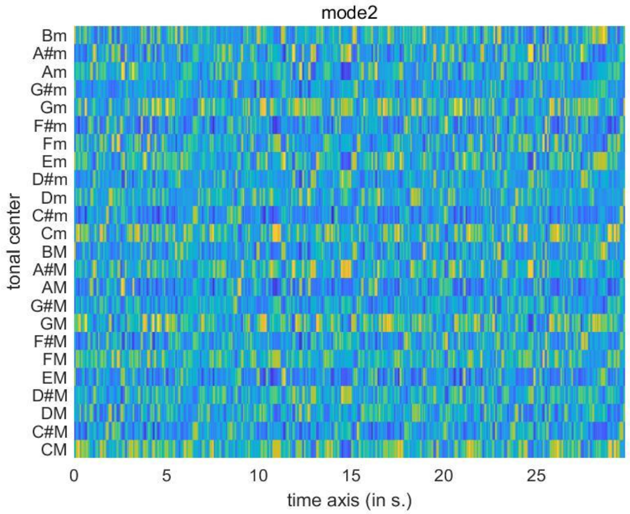


MIRToolbox

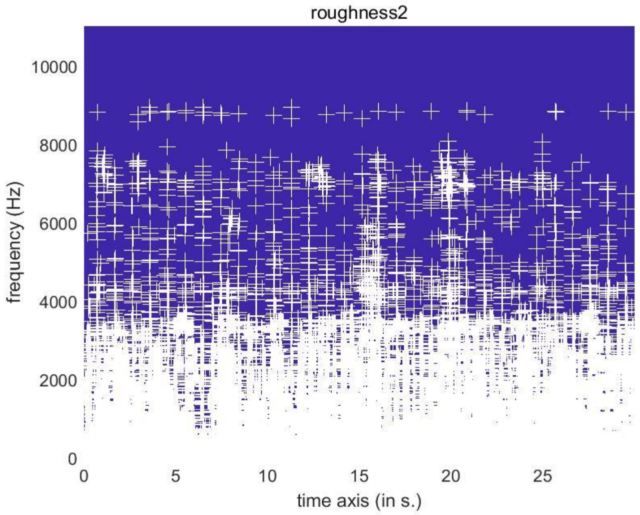
X_axis = time axis (in s.)

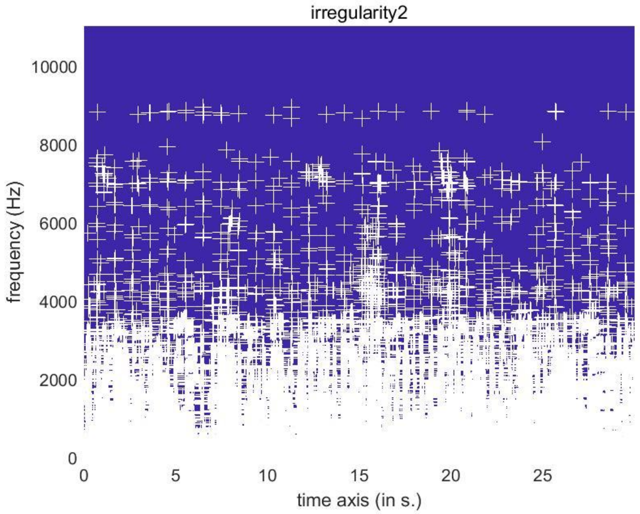
Tonal





Spectral

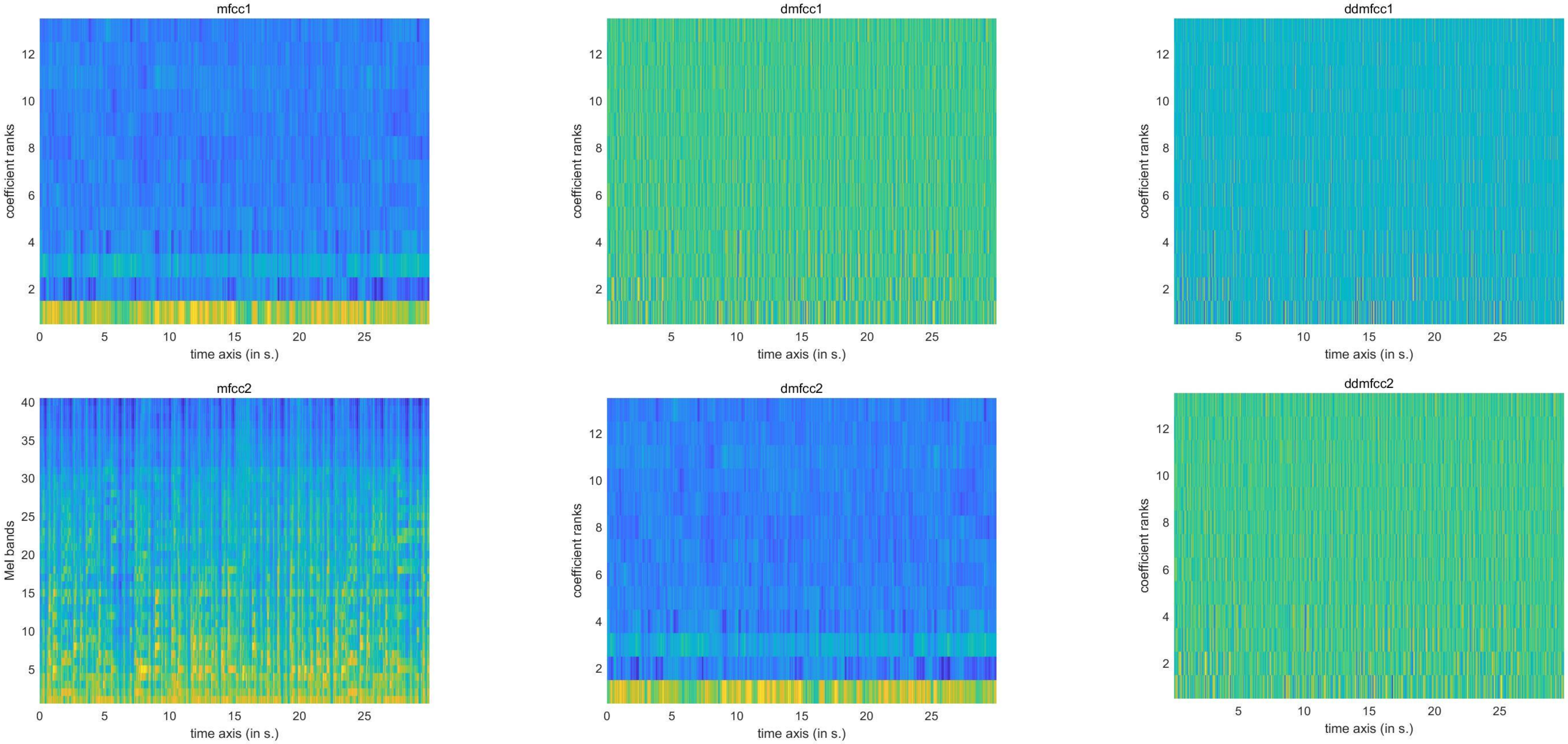




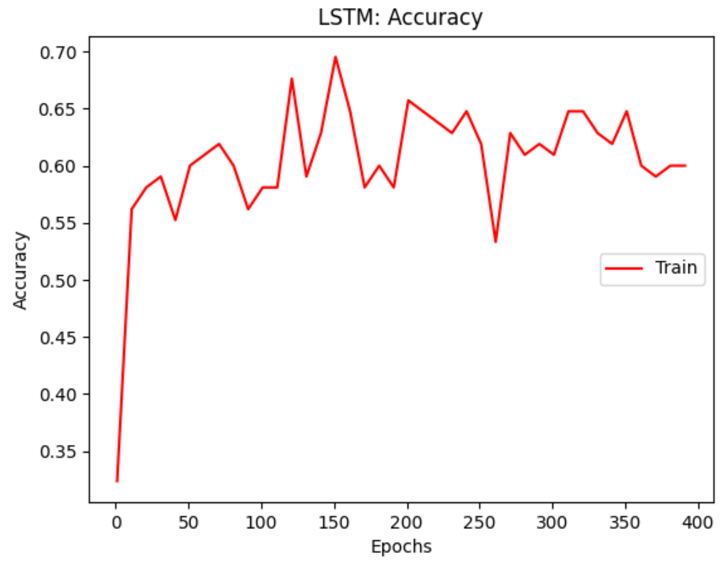
MIRToolbox

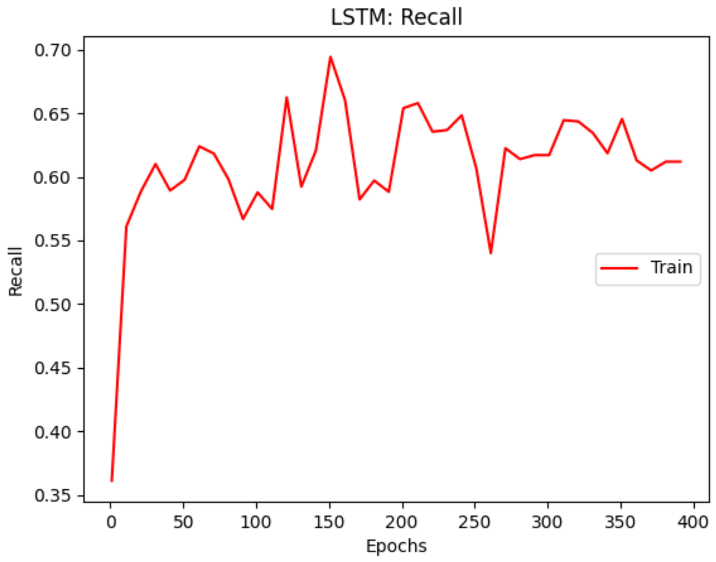
X_axis = time axis (in s.)

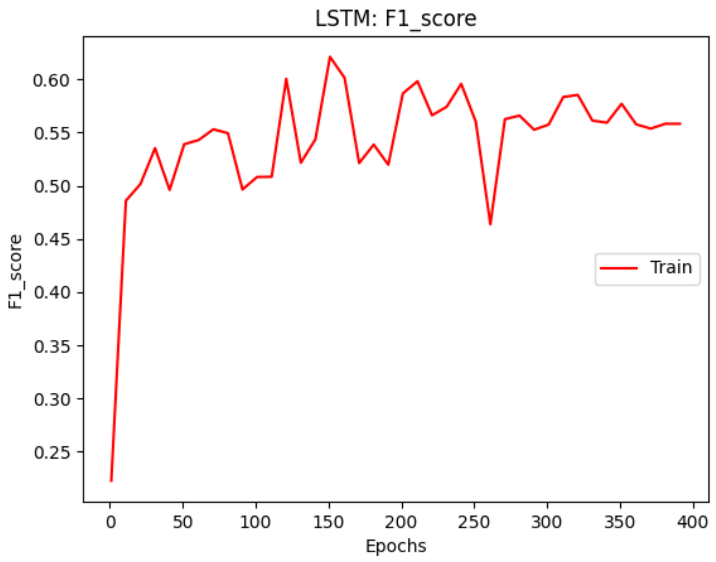
Spectral

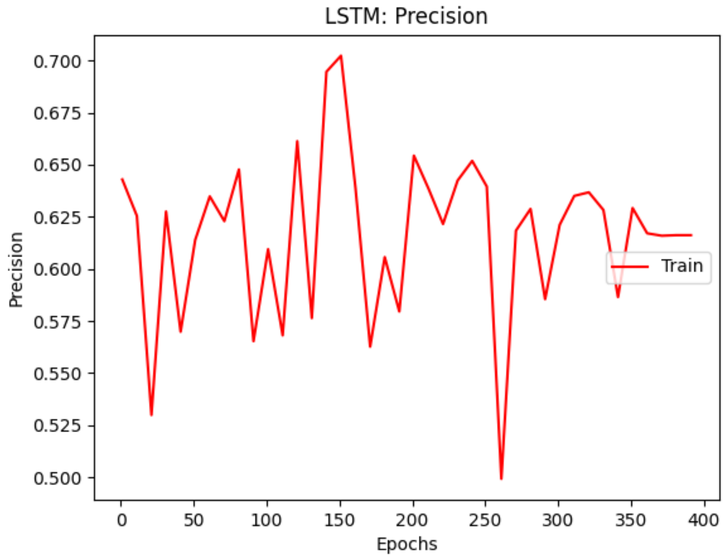


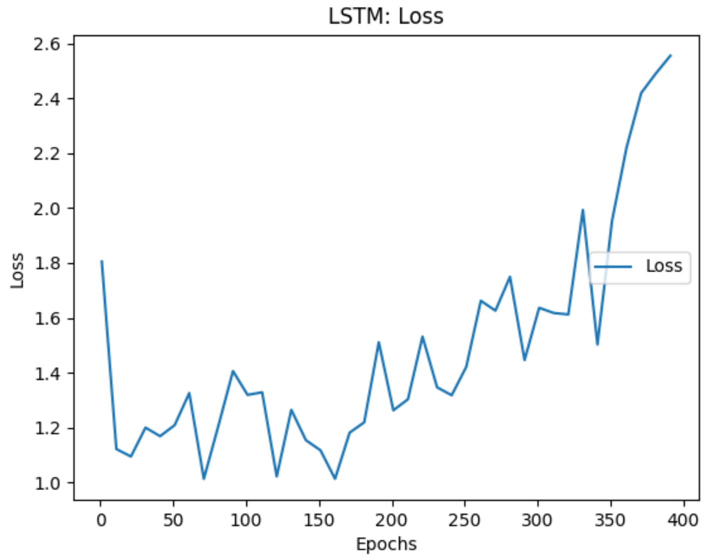
Evaluation Metrics











Experimental Settings

❖ Input settings
- MFCC [0:13]
- Spectral_centroid [13:14]
- Chroma_stft [14:26]
- Spectral_contrast [26:33]

❖ Cross validation
- Train : Validation : Test = 7 : 2 : 1

❖ Timeseries_length = 128

→ Train = [420,128,33]
Validation = [120,128,33]
Test = [60,128,33]

❖ Experimental settings
- Optimizer = Adam
- Learning rate = 0.001
- Batch size = 35
- Epoch = 400
- Loss function = NLLLoss
- Stateful = False

Stateful vs Stateless

상호 작용 상태가 얼마나 오래 기록되는지, 해당 정보가 어떤 식으로 저장되는지를 기준으로 구별

	전달 여부
stateful	O
stateless	X

Functional

- NLLLoss (Negative Log Likelihood Loss)
  - Classification에 유용
  - Neural Network에서 log-probabilities를 얻으려면 마지막 layer에 Log Softmax layer를 추가

- Softmax
  - Multicalss 분류에서 사용
  - 분류될 클래스가 n개라고 할 때, n차원의 벡터를 입력 받아 각 클래스에 속할 확률 추정
  - Log-Softmax = Softmax함수에 log를 취한 것
  - Softmax에 의한 결과는 NLL Loss에 직접적으로 사용할 수 없어서 log-softmax 사용

- CrossEntropyLoss
  - LogSoftmax + NLLLoss가 하나의 single class에 합쳐져 있음

Reference

[1] Massoudi, M., Verma, S., & Jain, R. (2021, January). Urban Sound Classification using CNN. In *2021 6th International Conference on Inventive Computation Technologies (ICICT)* (pp. 583-589). IEEE.
[2]Ghildiyal, A., Singh, K., & Sharma, S. (2020, November). Music genre classification using machine learning. In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1368-1372). IEEE.
[3] Palanisamy, K., Singhania, D., & Yao, A. (2020). Rethinking cnn models for audio classification. *arXiv preprint arXiv:2007.11154*.
[4] Yu, Y., Luo, S., Liu, S., Qiao, H., Liu, Y., & Feng, L. (2020). Deep attention based music genre classification. *Neurocomputing*, *372*, 84-91.
[5] Won, M., Ferraro, A., Bogdanov, D., & Serra, X. (2020). Evaluation of cnn-based automatic music tagging models. *arXiv preprint arXiv:2006.00751*.
[6] Bisharad, D., & Laskar, R. H. (2019, October). Music Genre Recognition Using Residual Neural Networks. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)* (pp. 2063-2068). IEEE.
[7] Zhang, S., Gu, H., & Li, R. (2019). MUSIC GENRE CLASSIFICATION: NEAR-REALTIME VS SEQUENTIAL APPROACH.
[8] Chillara, S., Kavitha, A. S., Neginhal, S. A., Haldia, S., & Vidyullatha, K. S. (2019). Music genre classification using machine learning algorithms: a comparison. *Int Res J Eng Technol*, *6*(5), 851-858.
[9] Zhang, W., Lei, W., Xu, X., & Xing, X. (2016, September). Improved music genre classification with convolutional neural networks. In *Interspeech* (pp. 3304-3308).
[10] Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017, March). Convolutional recurrent neural networks for music classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2392-2396). IEEE.
[11] Su, Y., Zhang, K., Wang, J., & Madani, K. (2019). Environment sound classification using a two-stream CNN based on decision-level fusion. *Sensors*, *19*(7), 1733.
[12] Choi, K., Fazekas, G., & Sandler, M. (2016). Automatic tagging using deep convolutional neural networks. *arXiv preprint arXiv:1606.00298*.
[13] Li, S., Yao, Y., Hu, J., Liu, G., Yao, X., & Hu, J. (2018). An ensemble stacked convolutional neural network model for environmental event sound recognition. *Applied Sciences*, *8*(7), 1152.