LARGE-SCALE MEDIA ANALYTICS

Master of Science in Signal Theory and Communications
TRACK: Signal Processing and Machine Learning for Big Data
Large-Scale Media Analytics

Fernando Marcos Macías

Víctor Gutiérrez García

Contents

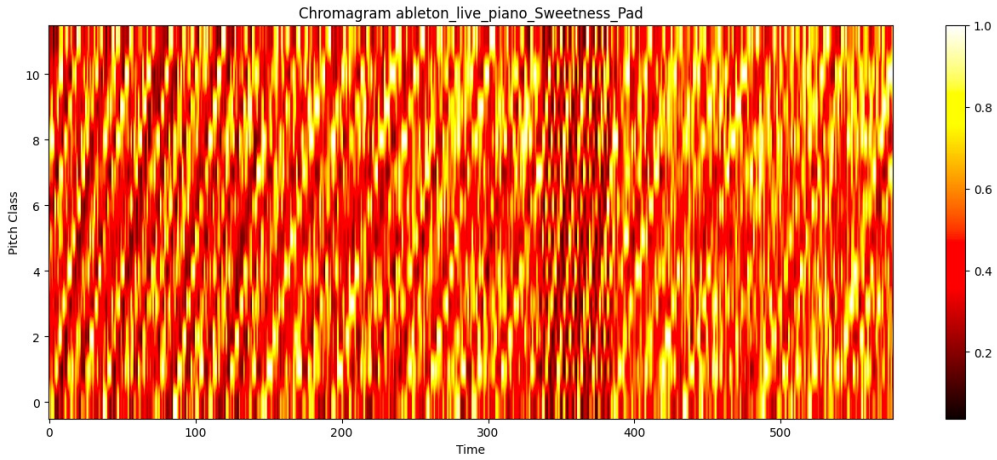
1. Preprocessing changes (In search of computational relief)
2. Vanilla classification
3. Classification – Machine Learning approach
4. Deep Learning approach (SOTA)
5. Chordifier MLP v1.0 (model)

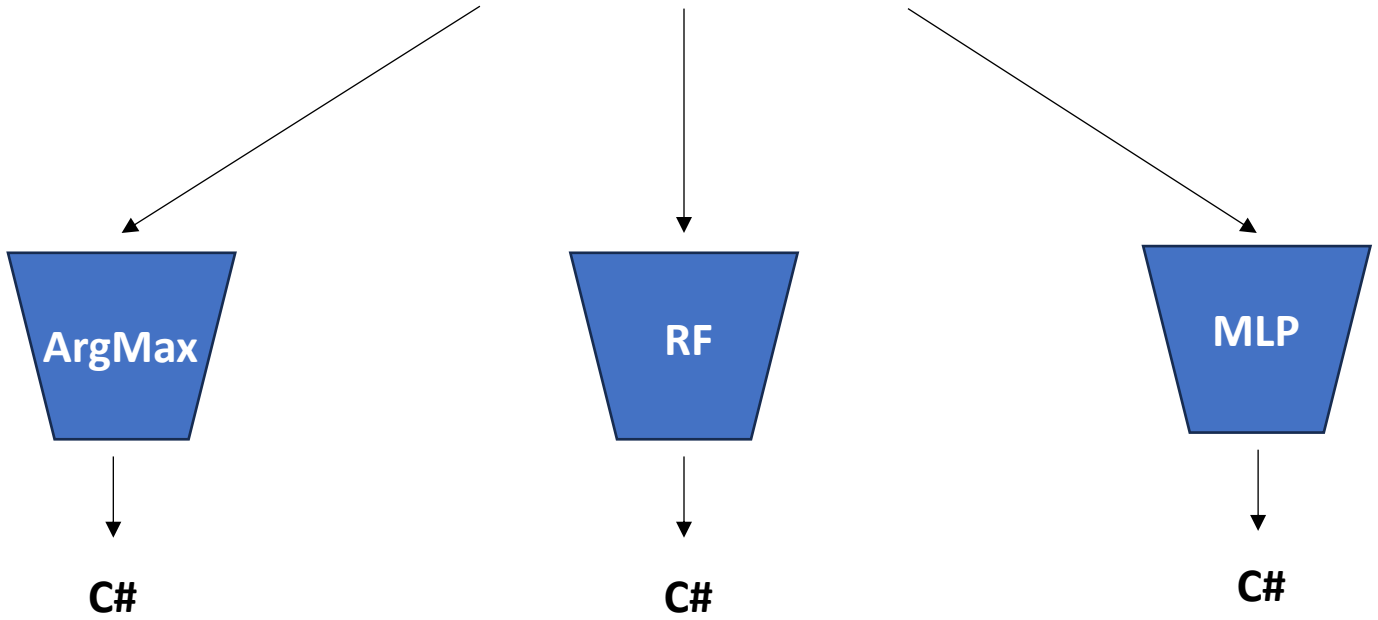
Preprocessing changes (In search of computational relief)

```
file_name,root_note,chord_type,instrument,0:C,1:C#,2:D,3:D#,4:E,5:F,6:F#,7:G,8:G#,9:A,10:A#,11:B
0000.wav,3,maj,non_guitar,0.904923,0.4398673,0.8172096,0.8267878,1.50597,0.48736066,0.65353477,2.0,0.78777236,0.45348868,0.55726534,1.0577536
0001.wav,3,maj/3,non_guitar,1.6517758,0.6860993,0.95392,0.92192984,1.8668052,0.7529041,0.7708566,2.0,0.7530966,0.4401064,0.48154235,1.1232809
0002.wav,3,maj/5,non_guitar,1.7701682,0.6261766,1.0550392,0.71275896,1.4523592,0.5825397,0.6273421,2.0,0.7860852,0.45420018,0.46564275,0.86736023
0003.wav,4,maj,non_guitar,0.9013666,0.901221,0.50261724,0.52798724,0.466196,1.3841572,0.7371292,0.68431073,2.0,0.8083938,0.42222685,0.44017124
0004.wav,4,maj/3,non_guitar,1.4220532,1.4964783,0.6301332,0.6197506,0.72147334,1.8278862,0.7115764,0.94909674,2.0,0.6106889,0.4562219,0.64709747
0005.wav,4,maj/5,non_guitar,0.92317677,1.6240267,0.6419565,0.6736475,0.5921629,1.5613351,0.6375462,0.5593946,2.0,0.9563987,0.50441194,0.52228177
0006.wav,5,maj,non_guitar,0.6246598,1.105669,1.2053002,0.6230696,0.9705212,0.9220522,1.7968769,0.7509885,0.78538066,2.0,1.1025949,0.6147498
0007.wav,5,maj/3,non_guitar,0.5644562,1.1153225,1.6481037,0.7102196,0.91029847,0.9265121,1.995942,0.71109635,0.7616801,1.9183404,0.6897918,0.41157863
0008.wav,5,maj/5,non_guitar,0.33601302,0.93069553,1.5698633,0.74803704,0.78145206,0.90817714,1.3852493,0.6011673,0.9646503,2.0,0.5469792,0.30162102
0009.wav,6,maj,non_guitar,0.40867317,0.47555226,1.0594431,1.3931407,0.7492863,0.86156034,0.7708069,1.8378762,0.77803093,0.7367356,1.9697325,0.84797
0010.wav,6,maj/3,non_guitar,0.35501137,0.4902497,1.0801834,1.5557756,0.8182808,1.0647993,0.99195266,2.0,0.9034375,0.7380421,1.8722719,0.7447258
0011.wav,6,maj/5,non_guitar,0.28565347,0.39132595,0.8467652,1.7040002,0.7176605,0.9545558,0.6531738,1.4612813,0.8841637,0.9919944,2.0,0.5862005
0012.wav,7,maj,non_guitar,0.62502134,0.31444705,0.3788839,0.9352263,1.9431742,1.018309,0.8927579,0.8416824,1.8922156,0.81661505,0.80637544,1.9673306
0013.wav,7,maj/3,non_guitar,0.7658702,0.3921622,0.425983,0.75483537,1.3915832,0.7040162,0.7301383,0.7366327,1.9662039,0.70304143,0.59069216,1.9770436
0014.wav,7,maj/5,non_guitar,0.7058693,0.45434034,0.5294031,1.061533,1.5841589,0.64182526,0.72633606,0.7073674,1.3034003,0.63075423,0.82638216,2.0
0015.wav,8,maj,non_guitar,2.0,0.7166803,0.38926315,0.39790806,0.96515167,1.3728478,0.6248592,0.5994758,0.7023922,1.3498169,0.48805326,0.67879236
```

- Precompute the features outside the dataloader.
- Added Chromagram values to each audio.
- Label enconding for root notes.

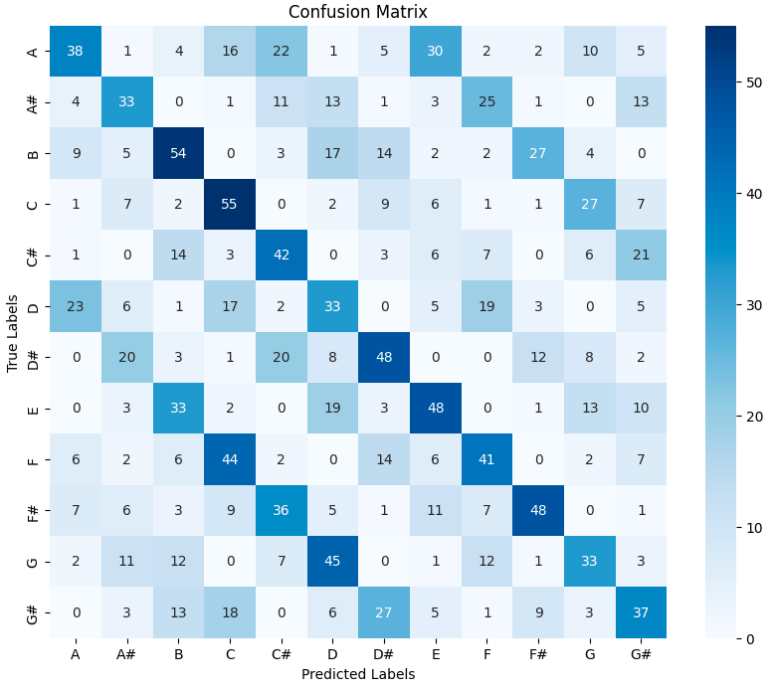
First features – Chromagram



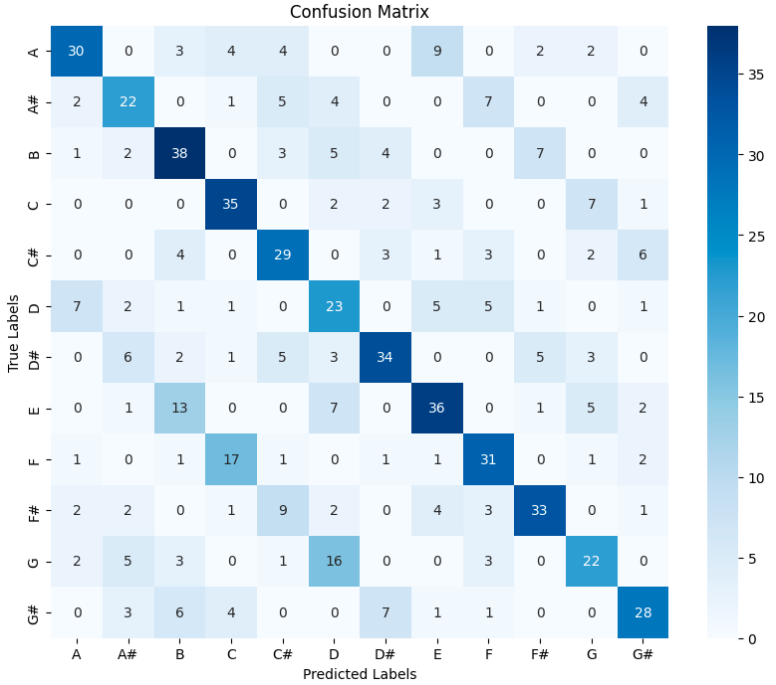


ArgMax – Vanilla Approach

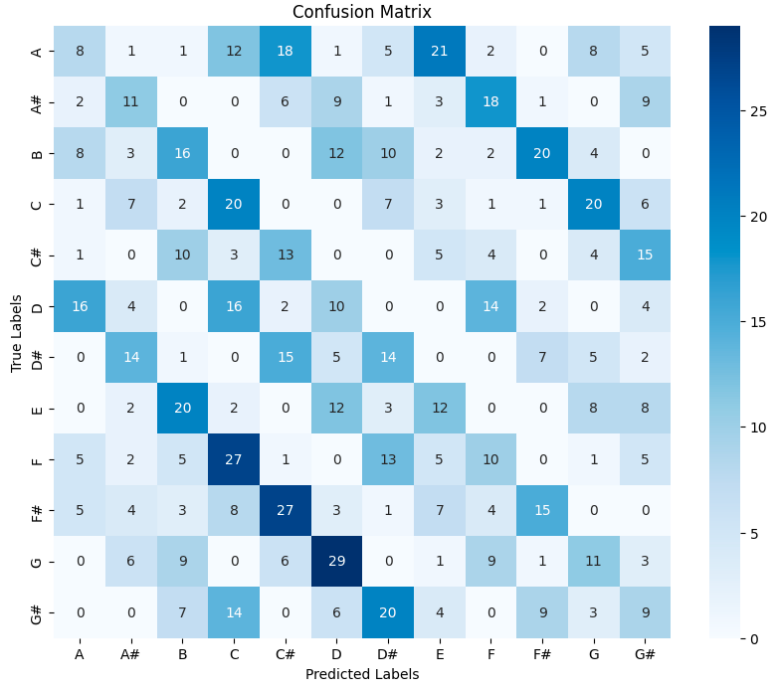
Total Test Set



Root notes Test Set



Inversions Test Set



ArgMax – Vanilla Approach

Total Test Set

	precision	recall	f1-score	support
0	0.42	0.28	0.33	136
1	0.34	0.31	0.33	105
2	0.37	0.39	0.38	137
3	0.33	0.47	0.39	118
4	0.29	0.41	0.34	103
5	0.22	0.29	0.25	114
6	0.38	0.39	0.39	122
7	0.39	0.36	0.38	132
8	0.35	0.32	0.33	130
9	0.46	0.36	0.40	134
10	0.31	0.26	0.28	127
11	0.33	0.30	0.32	122
accuracy			0.34	1480
macro avg	0.35	0.35	0.34	1480
weighted avg	0.35	0.34	0.34	1480

Root notes Test Set

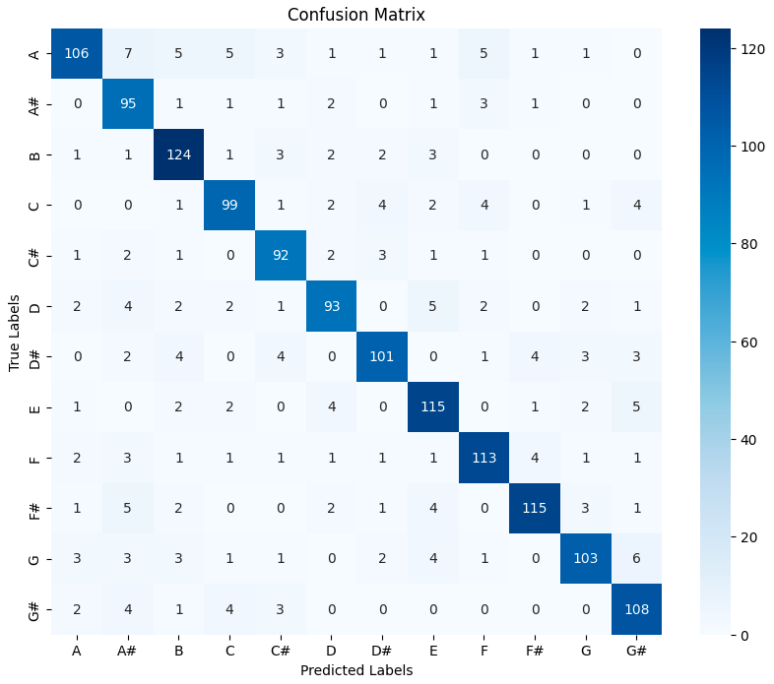
	precision	recall	f1-score	support
0	0.67	0.56	0.61	54
1	0.51	0.49	0.50	45
2	0.54	0.63	0.58	60
3	0.55	0.70	0.61	50
4	0.51	0.60	0.55	48
5	0.37	0.50	0.43	46
6	0.67	0.58	0.62	59
7	0.60	0.55	0.58	65
8	0.58	0.55	0.57	56
9	0.67	0.58	0.62	57
10	0.52	0.42	0.47	52
11	0.62	0.56	0.59	50
accuracy			0.56	642
macro avg	0.57	0.56	0.56	642
weighted avg	0.57	0.56	0.56	642

Inversions Test Set

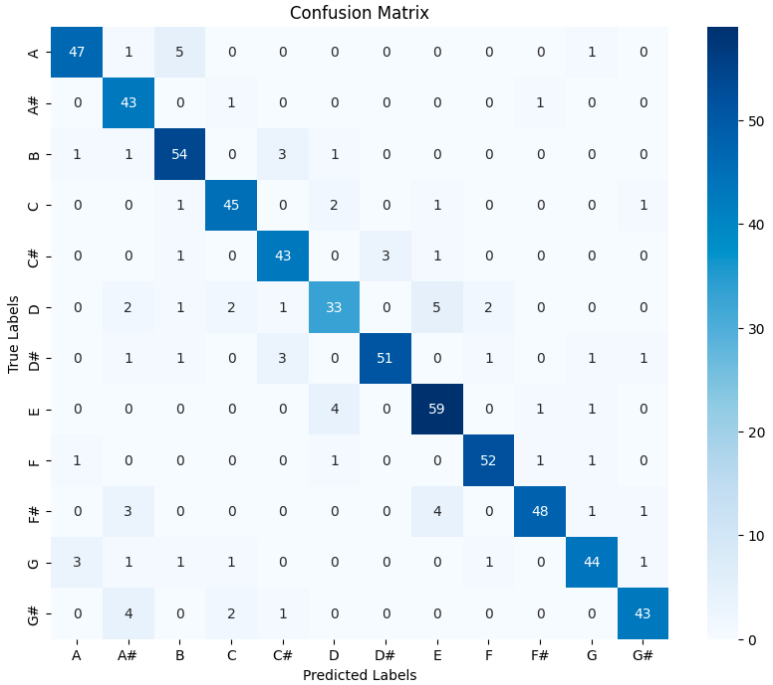
	precision	recall	f1-score	support
0	0.17	0.10	0.13	82
1	0.20	0.18	0.19	60
2	0.22	0.21	0.21	77
3	0.20	0.29	0.24	68
4	0.15	0.24	0.18	55
5	0.11	0.15	0.13	68
6	0.19	0.22	0.20	63
7	0.19	0.18	0.18	67
8	0.16	0.14	0.14	74
9	0.27	0.19	0.23	77
10	0.17	0.15	0.16	75
11	0.14	0.12	0.13	72
accuracy			0.18	838
macro avg	0.18	0.18	0.18	838
weighted avg	0.18	0.18	0.18	838

Random Forest – Machine Learning Approach

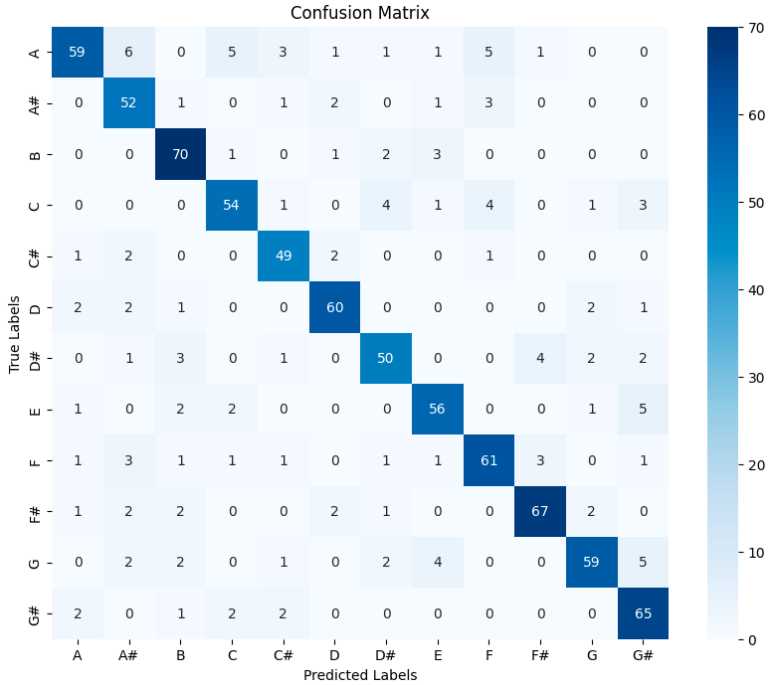
Total Test Set



Root notes Test Set



Inversions Test Set



Random Forest – Machine Learning Approach

Total Test Set

	precision	recall	f1-score	support
0	0.89	0.78	0.83	136
1	0.75	0.90	0.82	105
2	0.84	0.91	0.87	137
3	0.85	0.84	0.85	118
4	0.84	0.89	0.86	103
5	0.85	0.82	0.83	114
6	0.88	0.83	0.85	122
7	0.84	0.87	0.86	132
8	0.87	0.87	0.87	130
9	0.91	0.86	0.88	134
10	0.89	0.81	0.85	127
11	0.84	0.89	0.86	122
accuracy			0.85	1480
macro avg	0.85	0.86	0.85	1480
weighted avg	0.86	0.85	0.85	1480

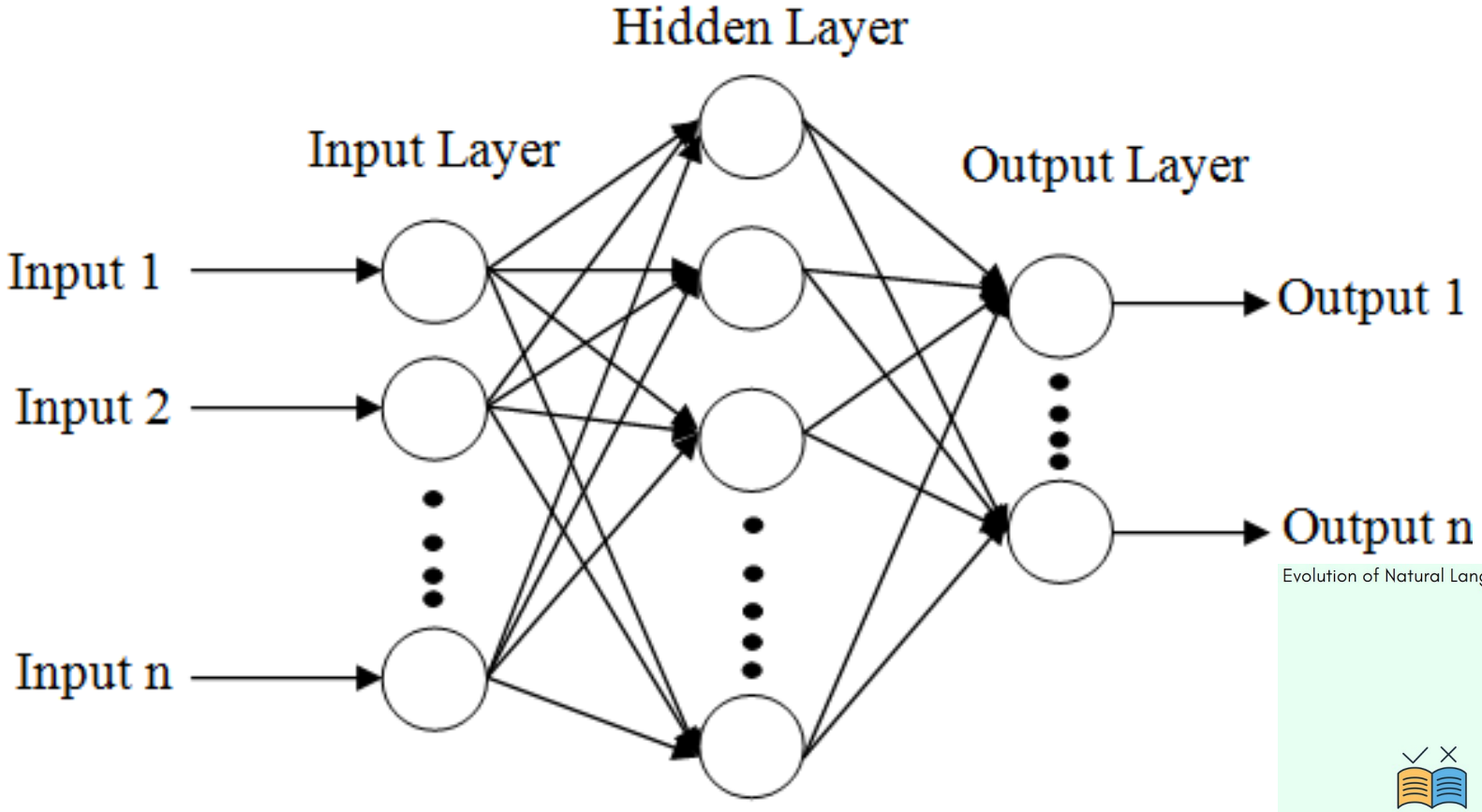
Normal chords notes Test Set

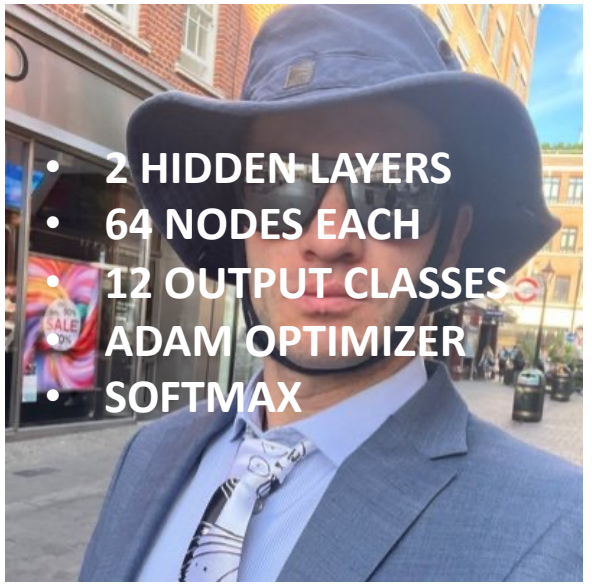
	precision	recall	f1-score	support
0	0.90	0.87	0.89	54
1	0.77	0.96	0.85	45
2	0.84	0.90	0.87	60
3	0.88	0.90	0.89	50
4	0.84	0.90	0.87	48
5	0.80	0.72	0.76	46
6	0.94	0.86	0.90	59
7	0.84	0.91	0.87	65
8	0.93	0.93	0.93	56
9	0.94	0.84	0.89	57
10	0.90	0.85	0.87	52
11	0.91	0.86	0.89	50
accuracy			0.88	642
macro avg	0.88	0.87	0.87	642
weighted avg	0.88	0.88	0.88	642

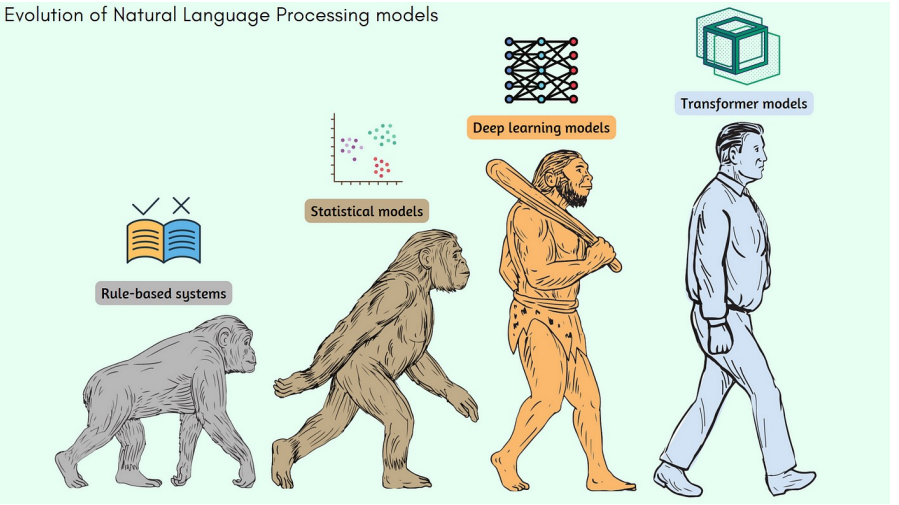
Inversions Test Set

	precision	recall	f1-score	support
0	0.88	0.72	0.79	82
1	0.74	0.87	0.80	60
2	0.84	0.91	0.88	77
3	0.83	0.79	0.81	68
4	0.83	0.89	0.86	55
5	0.88	0.88	0.88	68
6	0.82	0.79	0.81	63
7	0.84	0.84	0.84	67
8	0.82	0.82	0.82	74
9	0.89	0.87	0.88	77
10	0.88	0.79	0.83	75
11	0.79	0.90	0.84	72
accuracy			0.84	838
macro avg	0.84	0.84	0.84	838
weighted avg	0.84	0.84	0.84	838

Chordifier MLP alphaV1.0

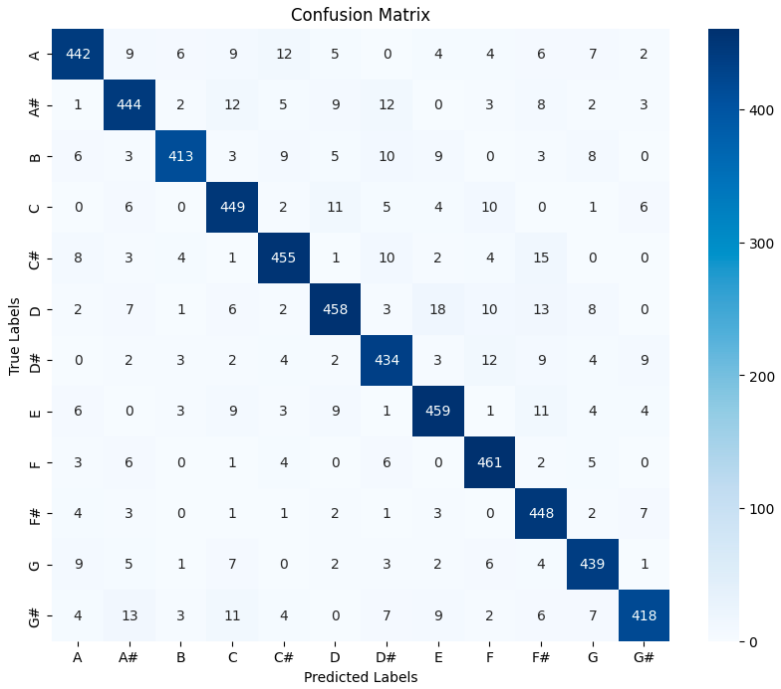




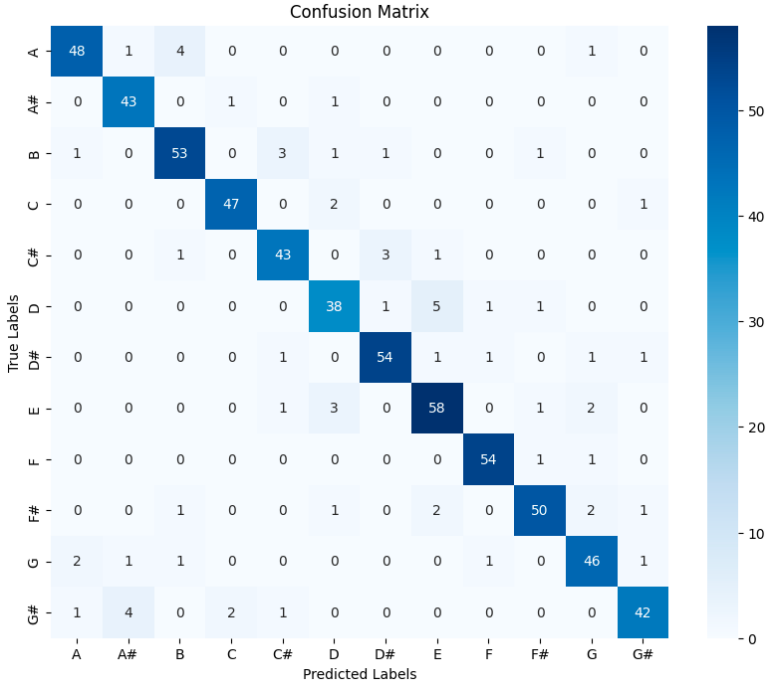


MLP – Deep Learning Approach

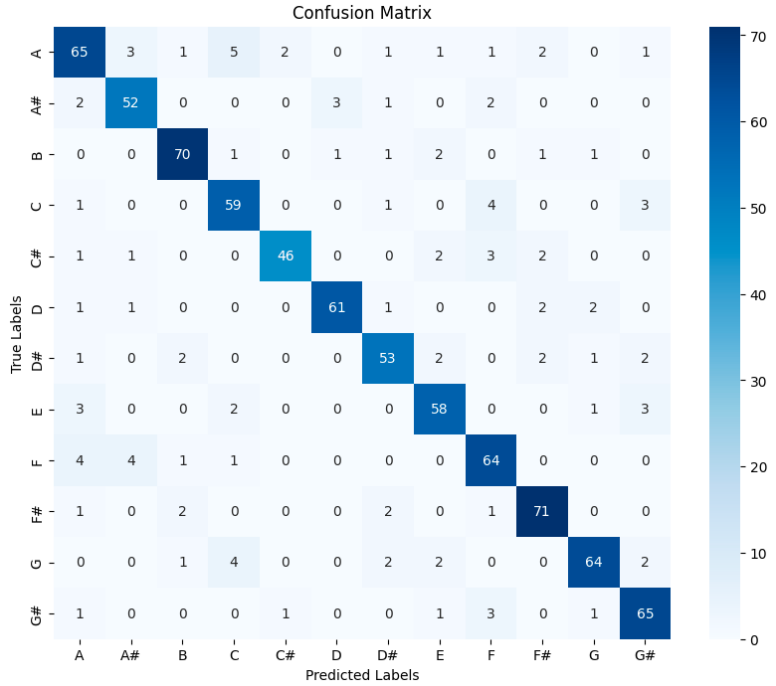
Total Test Set



Root notes Test Set



Inversions Test Set



MLP – Deep Learning Approach

Total Test Set

Classification Report:

	precision	recall	f1-score	support
0	0.91	0.87	0.89	506
1	0.89	0.89	0.89	501
2	0.95	0.88	0.91	469
3	0.88	0.91	0.89	494
4	0.91	0.90	0.91	503
5	0.91	0.87	0.89	528
6	0.88	0.90	0.89	484
7	0.89	0.90	0.90	510
8	0.90	0.94	0.92	488
9	0.85	0.95	0.90	472
10	0.90	0.92	0.91	479
11	0.93	0.86	0.90	484
accuracy			0.90	5918
macro avg	0.90	0.90	0.90	5918
weighted avg	0.90	0.90	0.90	5918

Normal chords notes Test Set

Classification Report:

	precision	recall	f1-score	support
0	0.92	0.89	0.91	54
1	0.88	0.96	0.91	45
2	0.88	0.88	0.88	60
3	0.94	0.94	0.94	50
4	0.88	0.90	0.89	48
5	0.83	0.83	0.83	46
6	0.92	0.92	0.92	59
7	0.87	0.89	0.88	65
8	0.95	0.96	0.96	56
9	0.93	0.88	0.90	57
10	0.87	0.88	0.88	52
11	0.91	0.84	0.87	50
accuracy			0.90	642
macro avg	0.90	0.90	0.90	642
weighted avg	0.90	0.90	0.90	642

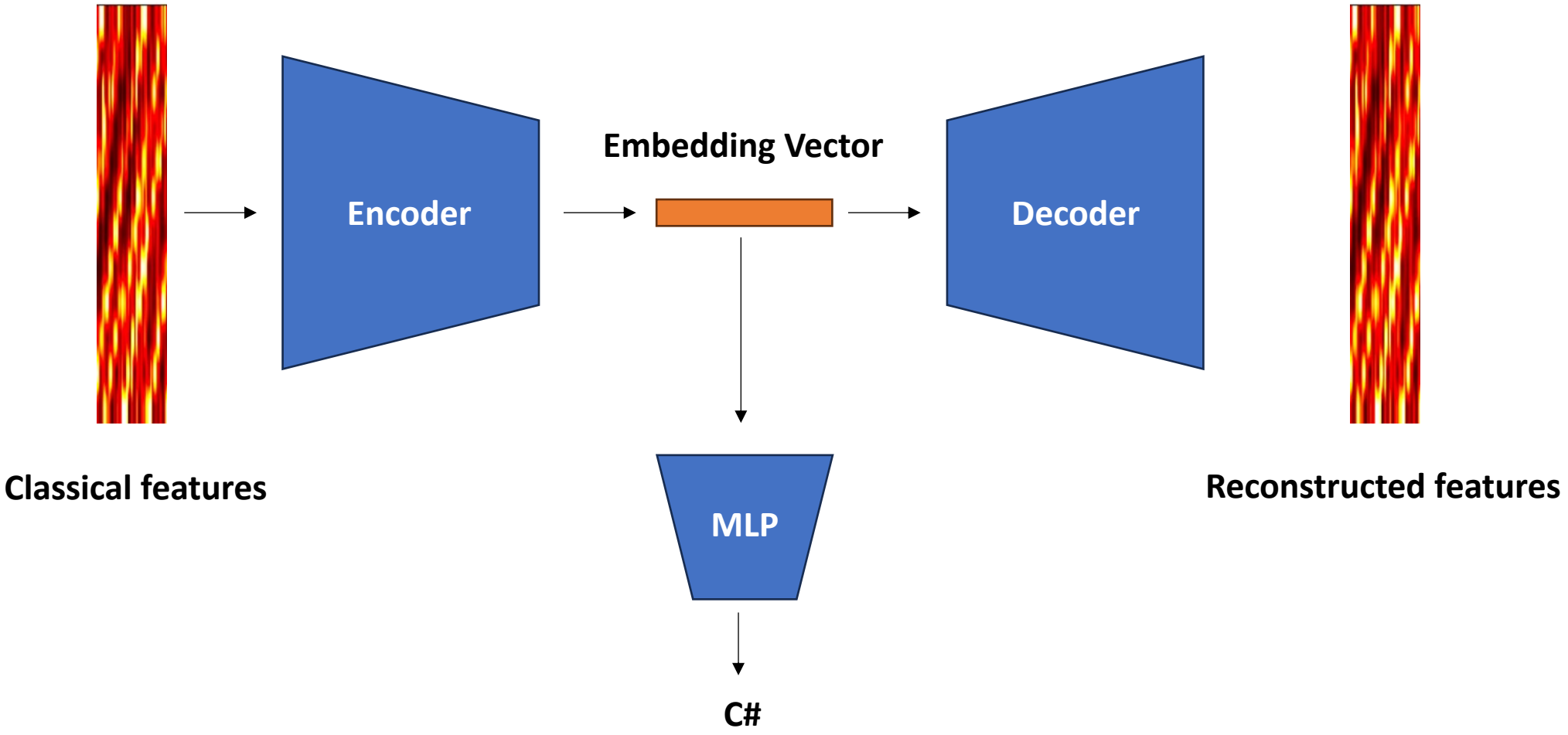
Inversions Test Set

Classification Report:

	precision	recall	f1-score	support
0	0.81	0.79	0.80	82
1	0.85	0.87	0.86	60
2	0.91	0.91	0.91	77
3	0.82	0.87	0.84	68
4	0.94	0.84	0.88	55
5	0.94	0.90	0.92	68
6	0.85	0.84	0.85	63
7	0.85	0.87	0.86	67
8	0.82	0.86	0.84	74
9	0.89	0.92	0.90	77
10	0.91	0.85	0.88	75
11	0.86	0.90	0.88	72
accuracy			0.87	838
macro avg	0.87	0.87	0.87	838
weighted avg	0.87	0.87	0.87	838

Chordifier MLP v1.0

Architecture proposal:



Chordifier MLP v1.0

PyTorch Lightning

- High-level management of **pytorch** functionalities
- **Dataset class** incorporates feature extraction in a memory-efficient way

```python
class CustomDataset(Dataset):
    def __init__(self, dataframe, root_dir, feature = 'Chromagram',transform=None):
        self.dataframe = dataframe['file_name']
        self.root_dir = root_dir
        self.feature = feature
        self.transform = transform
        self.label = dataframe['root_note']

    def __len__(self):
        return len(self.dataframe)

    def __getitem__(self, idx):
        audio_file = os.path.join(self.root_dir, self.dataframe.iloc[idx, 0])

        if self.transform:
            audio_data = self.transform(audio_data)
        x, sr = librosa.load(audio_file, sr=None)

        return self.label[idx], self.get_features(x, sr, self.feature)

    def get_features(self, x, sr, feature='Chromagram'):

        rerturned_feature = np.empty((0, 0))
        hop_length = int(44.1e3*2)
        if feature == 'Chromagram':
            n_chroma = 12
            n_octaves =7
            rerturned_feature = librosa.feature.chroma_cqt(y=x, sr=sr, n_chroma=n_chroma, n_octaves=n_octaves, hop_length=hop_length)

        elif feature == 'Mel Spectrogram':
            n_mels = 128
            n_fft = hop_length
            rerturned_feature = librosa.feature.melspectrogram(y=x, sr=sr, n_mels=n_mels, n_fft=n_fft, hop_length=hop_length)

        else: pass # Implement other features

        return rerturned_feature
```

SSR

Chordifier MLP v1.0

DEMO VERSION ON GRADIO NEXT WEEK

Questions?