Birdclef 2021

Birdcall Identification

Tobias Priesholm Gårdhus & Kaare Endrup Iversen

0:00

Motivation

Technical

- Working with "complex" mixed data-sources including audio
- "Real world" problem

Ethical

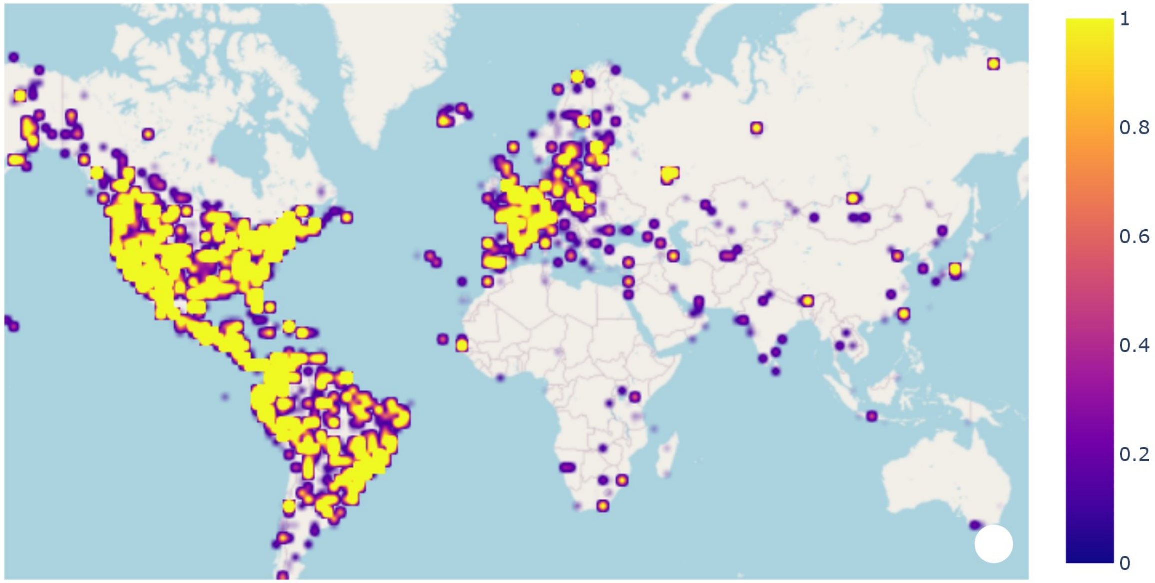
- Pros: Contributing to wild-life monitoring and preservation
- Cons: Enables automatic survaillance

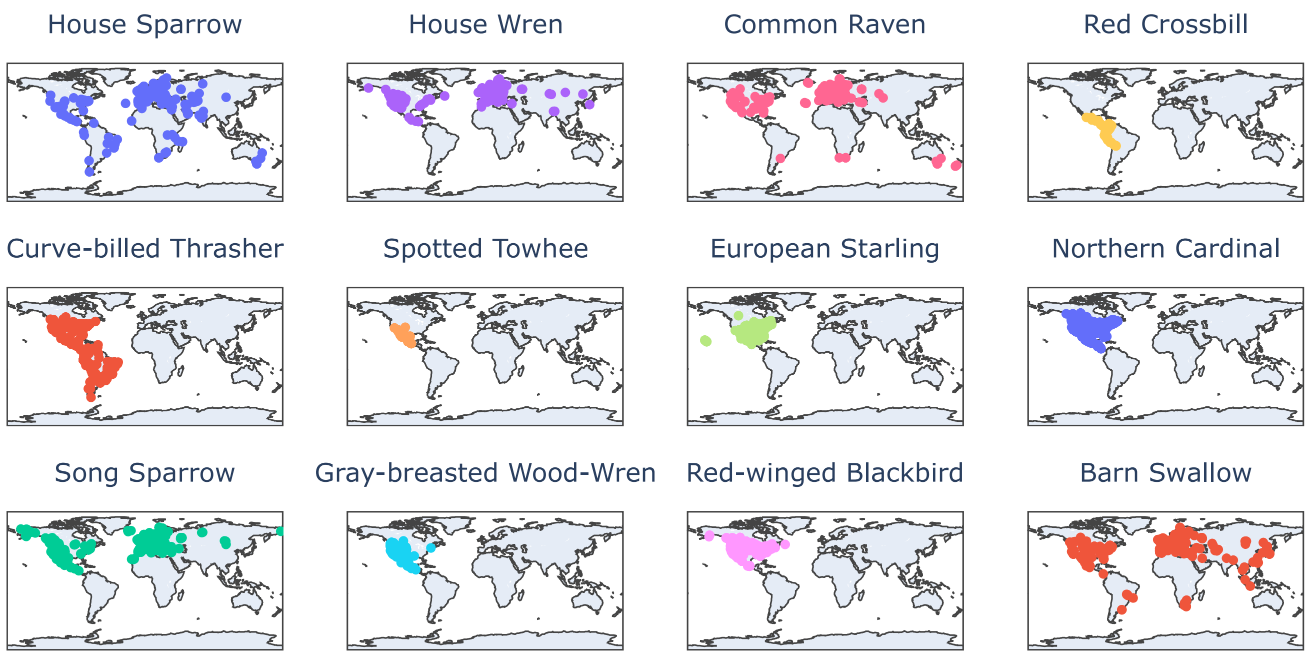
Fun

- Working in a field new to both of us

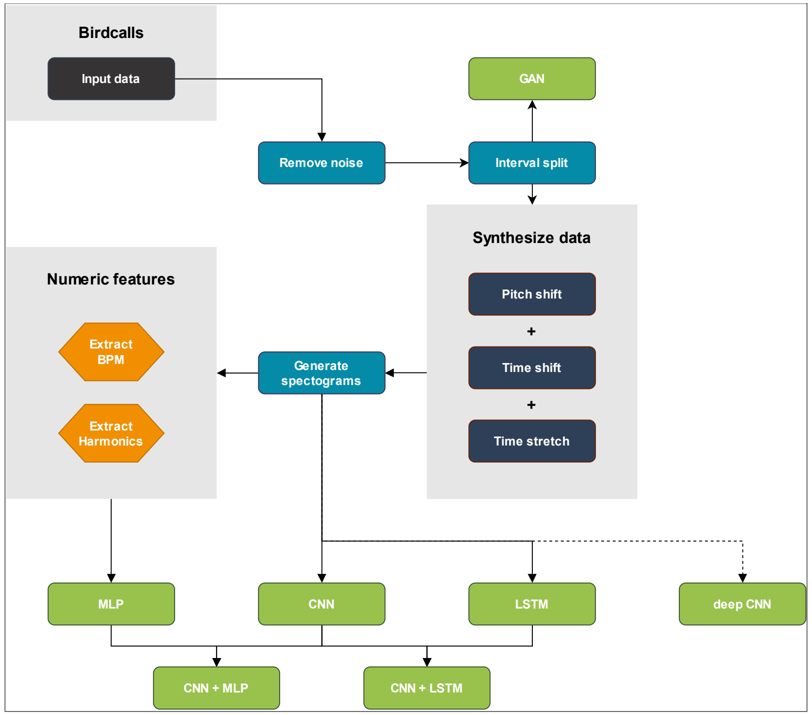
Data

8548 Audio files containing 27 different species



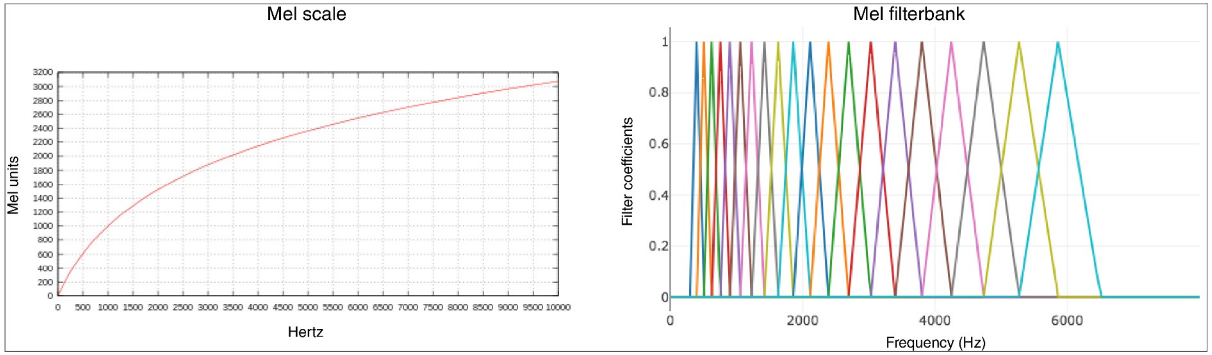

House Sparrow
House Wren
Common Raven
Red Crossbill
Curve-billed Thrasher
Spotted Towhee
European Starling
Northern Cardinal
Song Sparrow
Gray-breasted Wood-Wren
Red-winged Blackbird
Barn Swallow

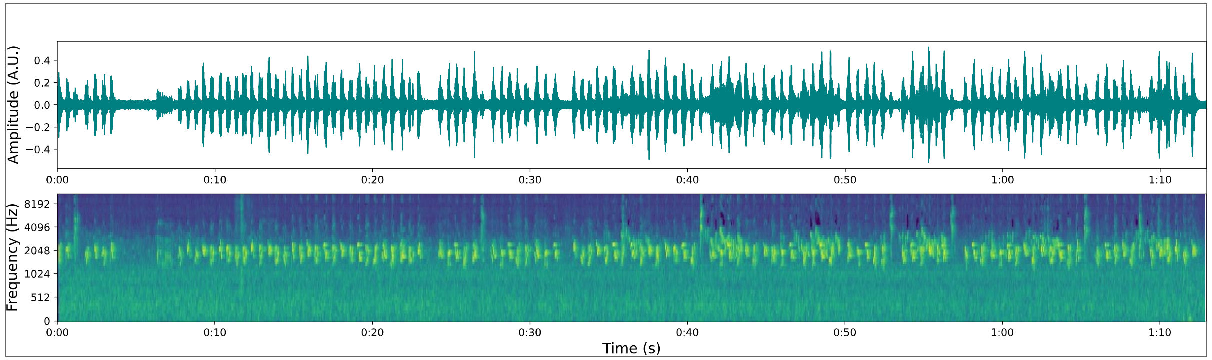
Procedure



Birdcalls
Input data
GAN
Remove noise
Interval split
Synthesize data
Numeric features
Pitch shift
+
Extract
BPM
Generate
spectograms
Time shift
+
Extract
Harmonics
Time stretch
MLP
CNN
LSTM
deep CNN
CNN + MLP
CNN + LSTM

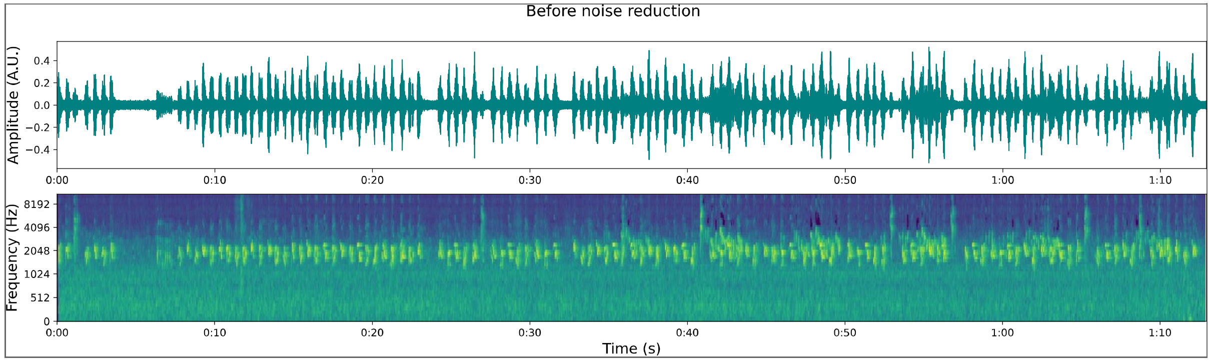
Mel Spectograms

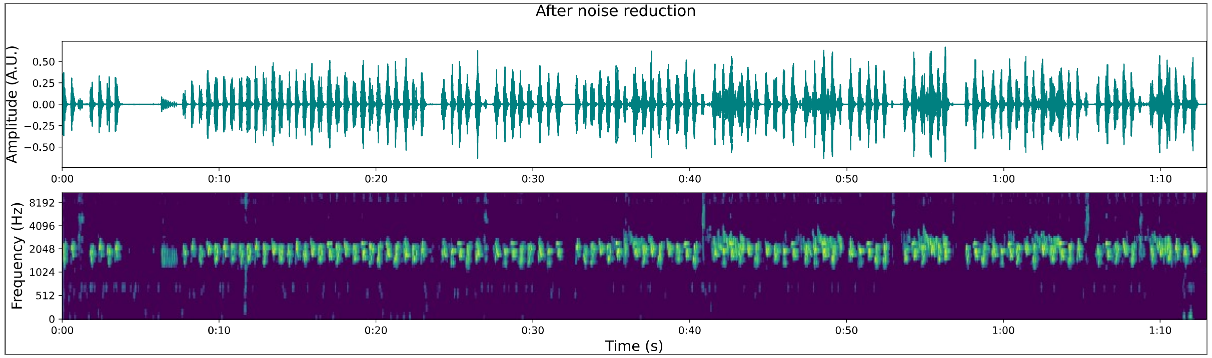




Feature Engineering

Noise reduction

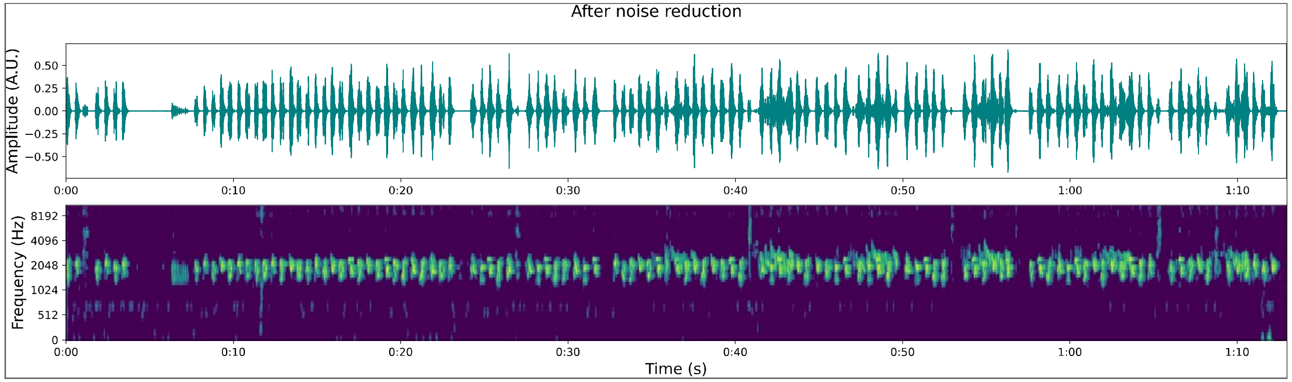




Feature Engineering

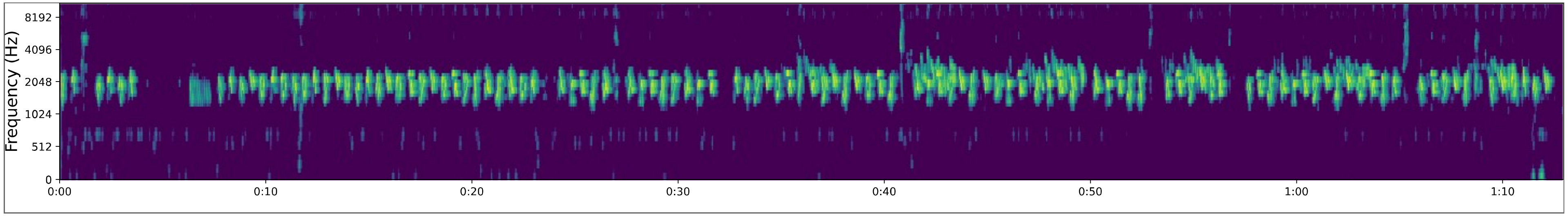
Noise reduction

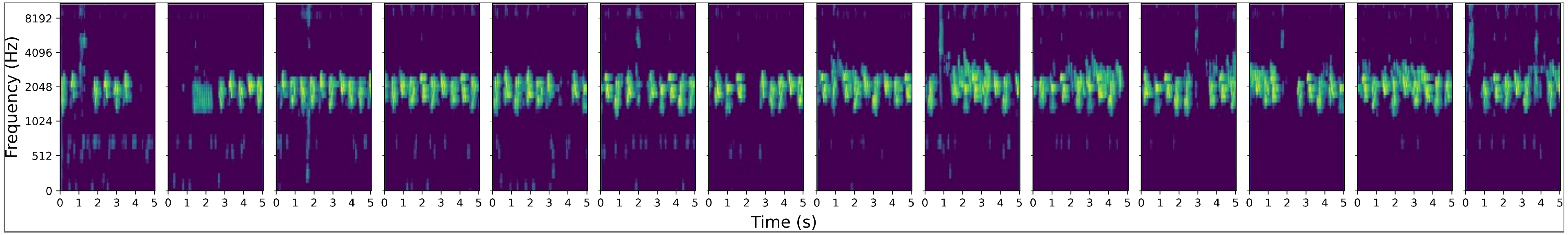
0:00



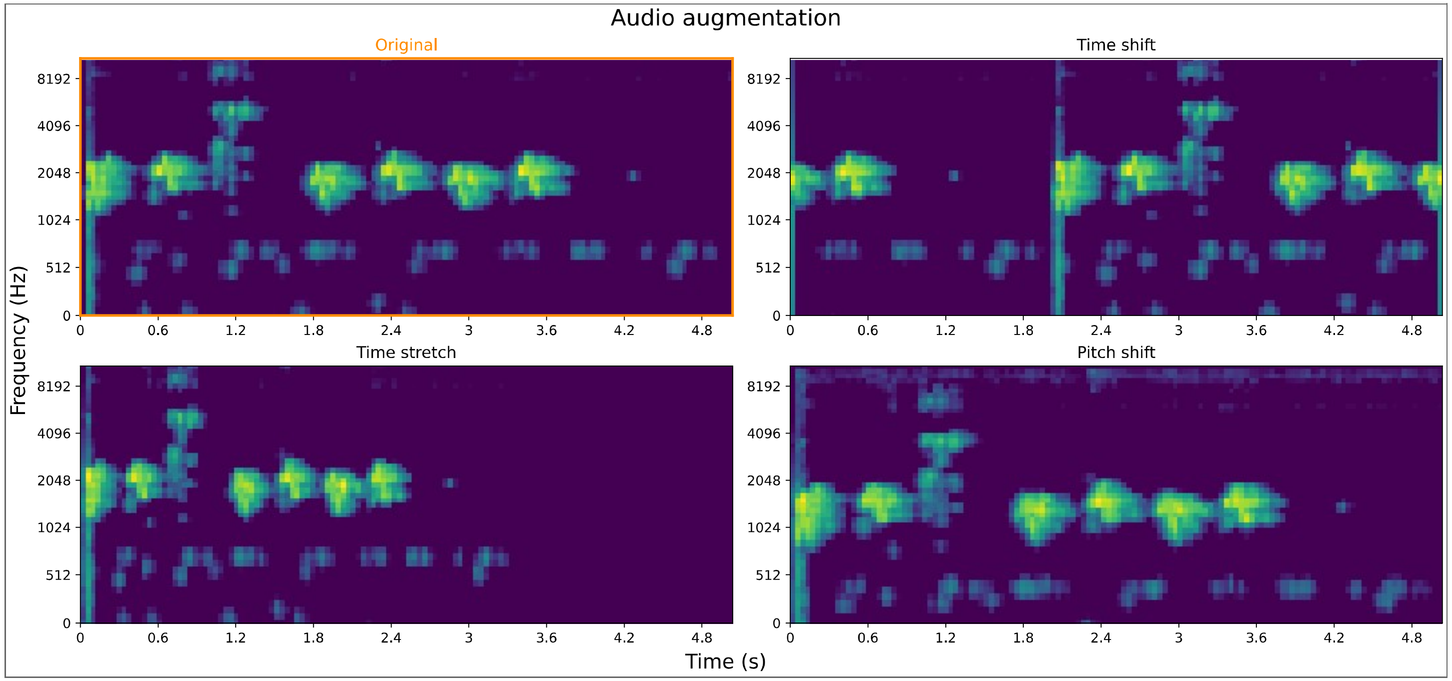
0:00:00 / 12:25:39

Split

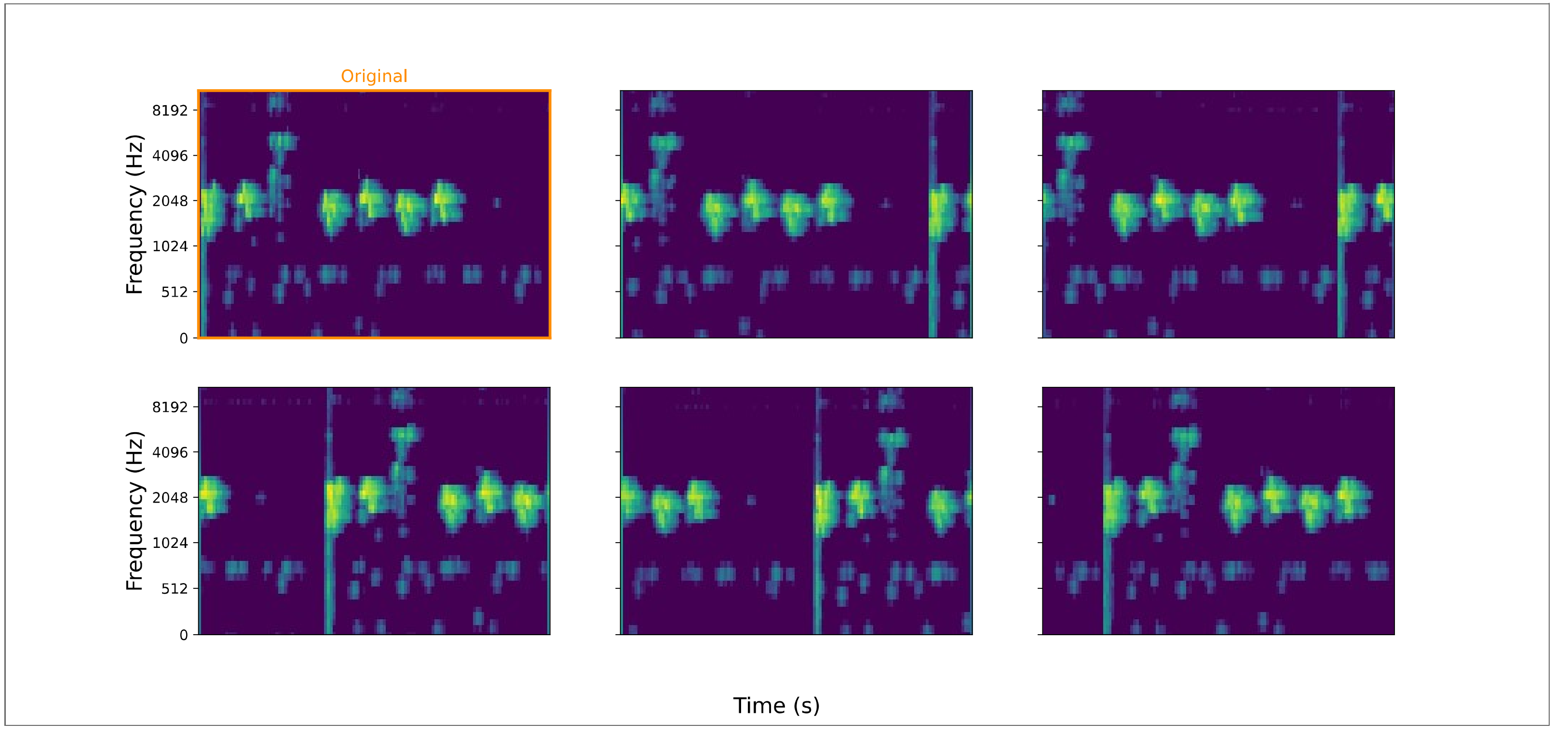

Frequency (Hz)
8192
4096
2048
1024
512
0
0:00
0:10
0:20
0:30
0:40
0:50
1:00
1:10


Frequency (Hz)
8192
4096
2048
1024
512
0
0 1 2 3 4 5
Time (s)

Audio Augmentation

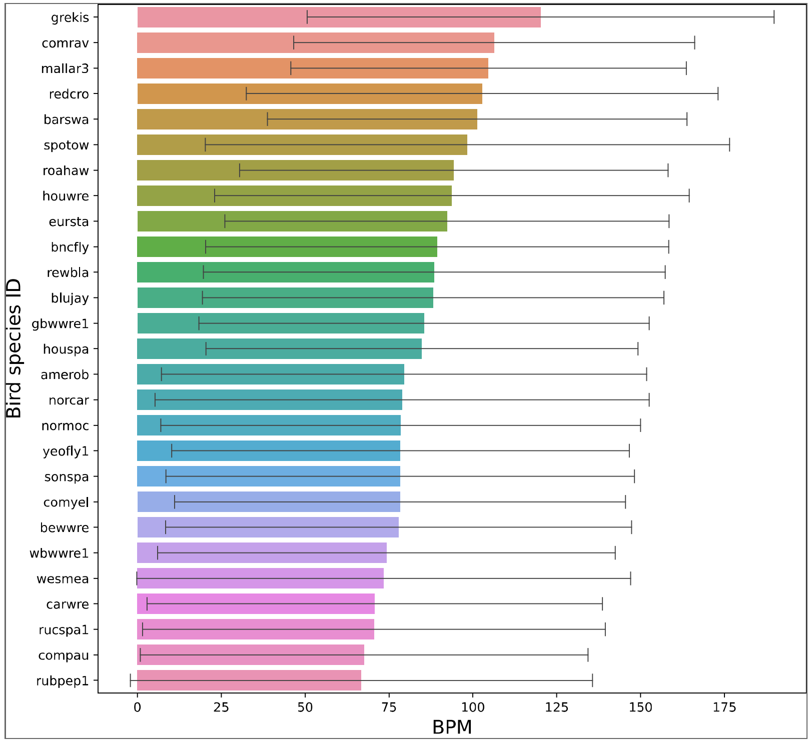


Audio Augmentation



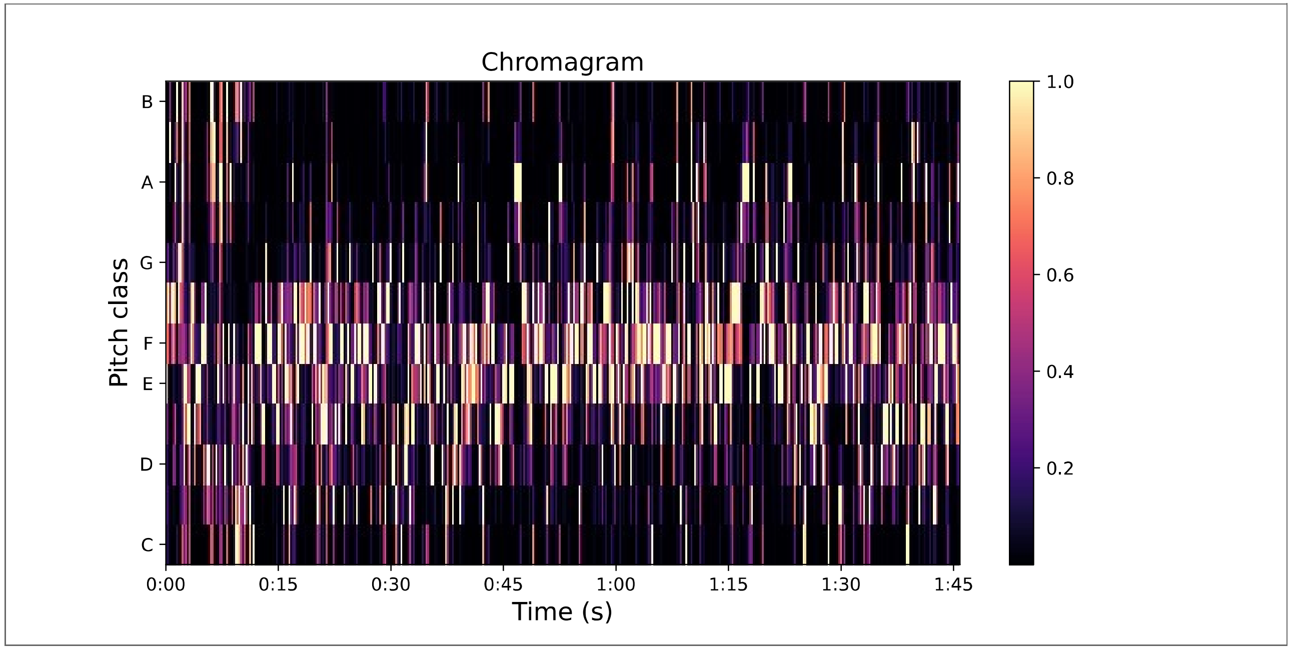
Numeric features

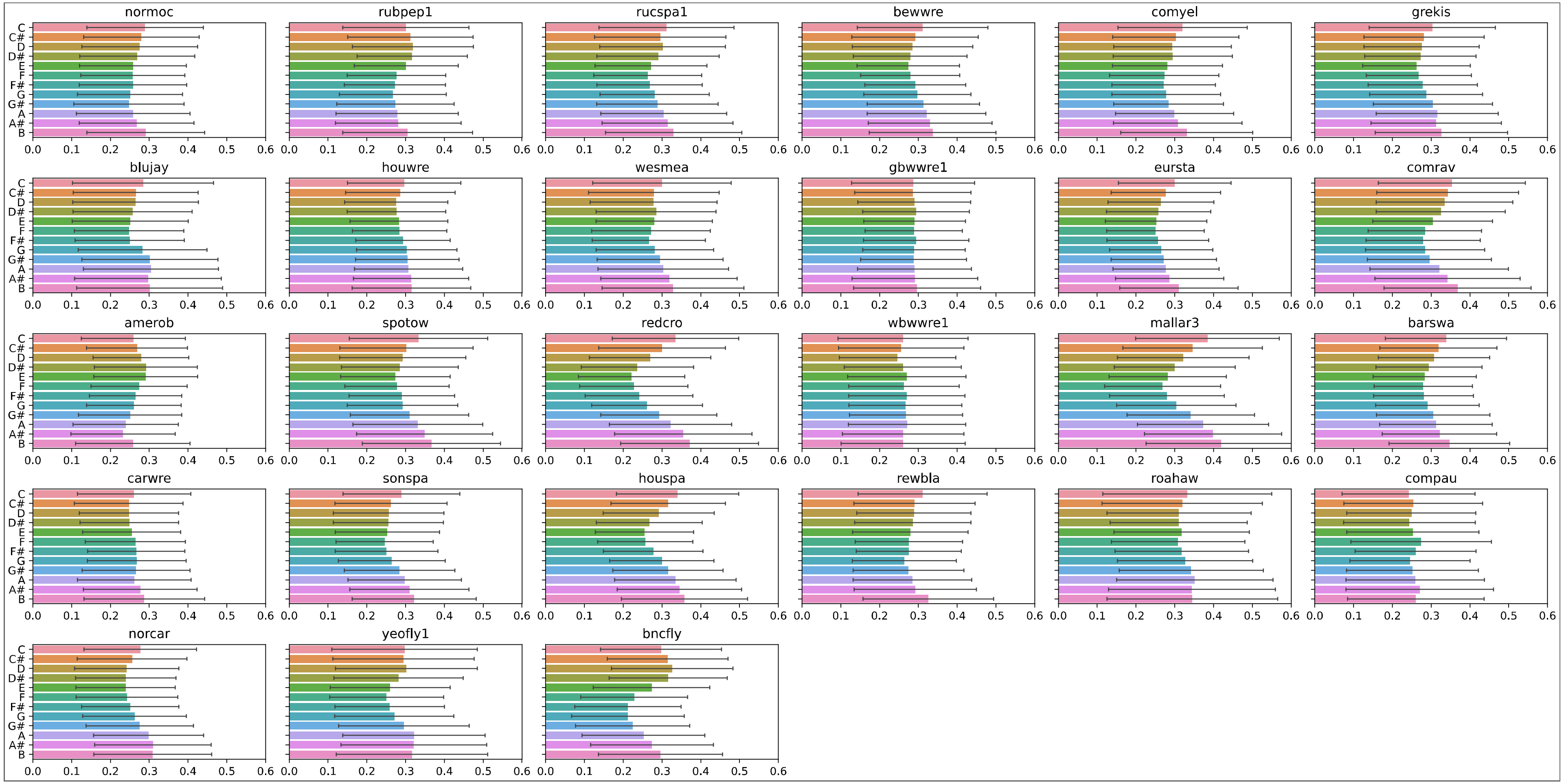
BPM



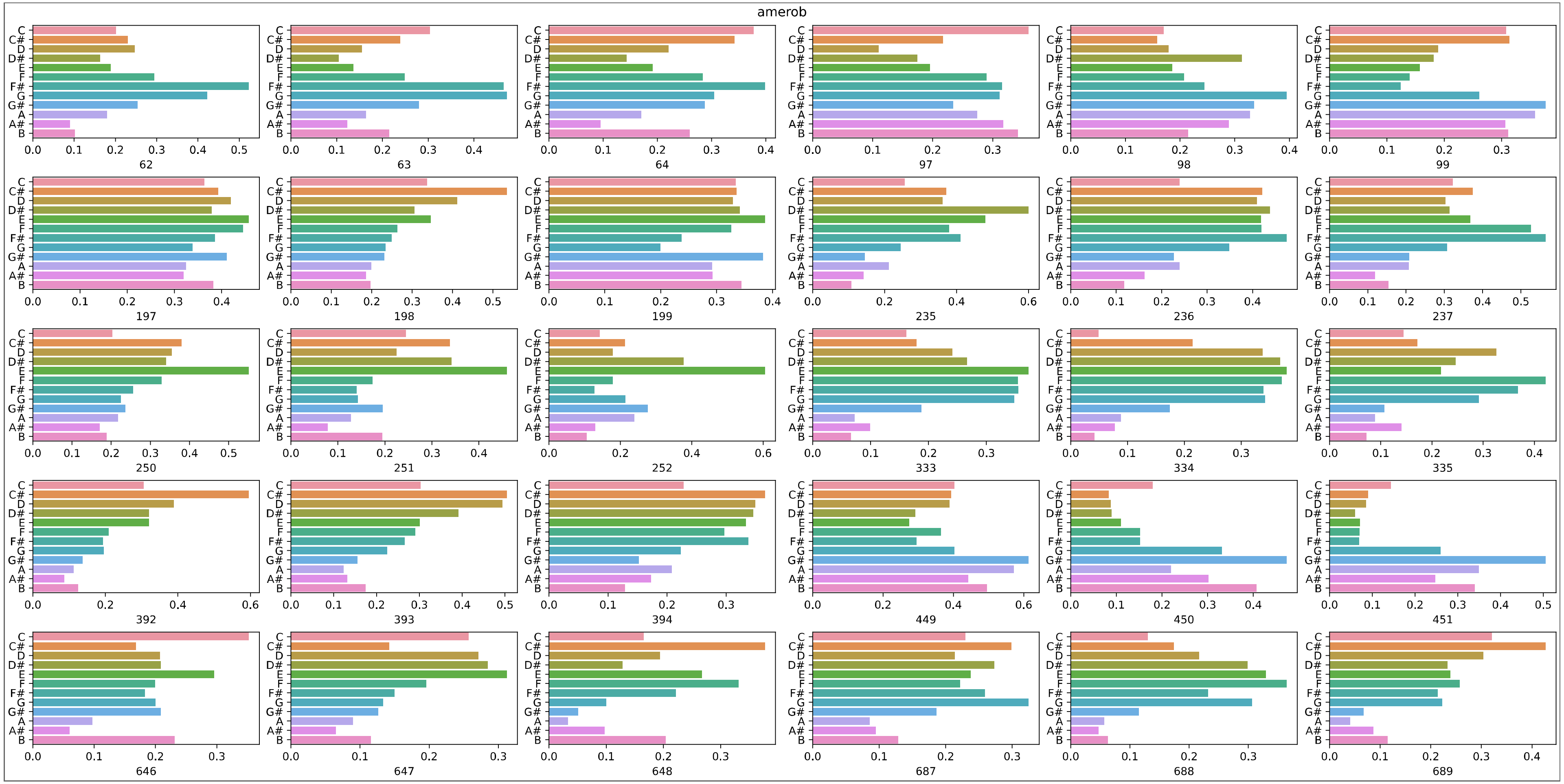
Numeric features

Harmonics





normoc
rubpep1
rucspa1
bewwre
comyel
grekis
blujay
houwre
wesmea
gbwwre1
eursta
comrav
amerob
spotow
redcro
wbwwre1
mallar3
barswa
carwre
sonspa
houspa
rewbla
roahaw
compau
norcar
yeofly1
bncfly
C
C#
D
D#
E
F
F#
G
G#
A
A#
B
0.0 0.1 0.2 0.3 0.4 0.5 0.6



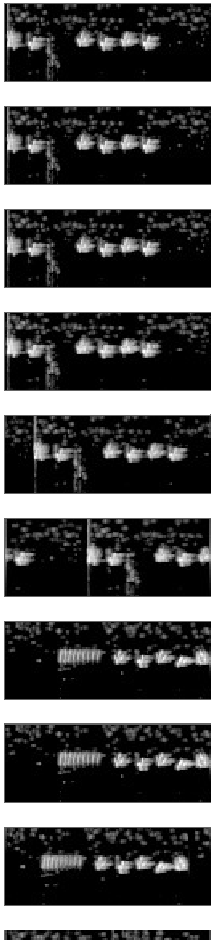
amerob
62
63
64
97
98
99
197
198
199
235
236
237
250
251
252
333
334
335
392
393
394
449
450
451
646
647
648
687
688
689
C
C#
D
D#
E
F
F#
G
G#
A
A#
B

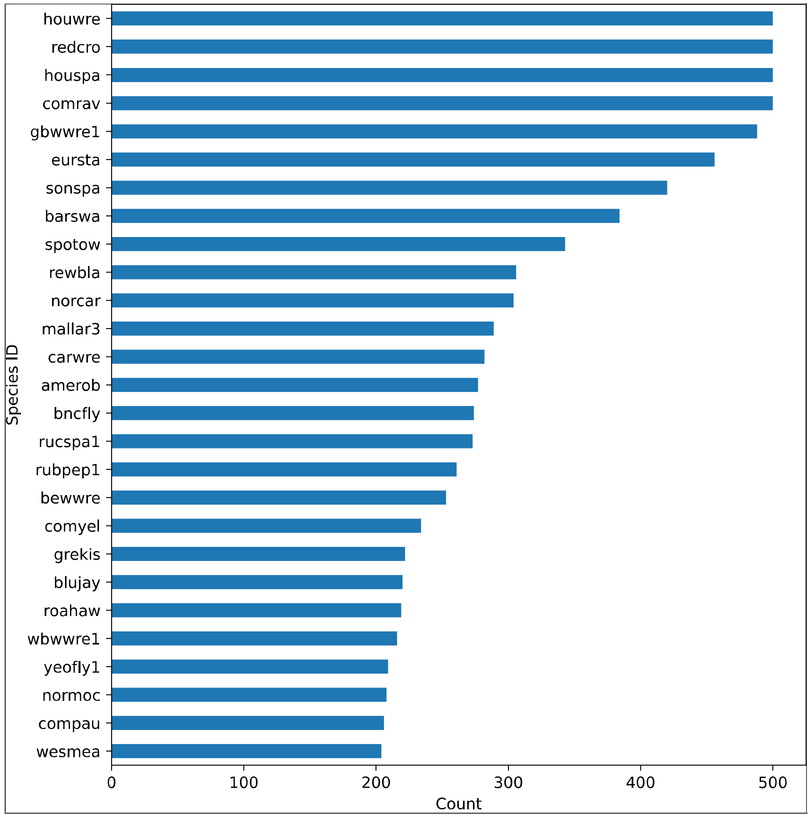
Input data set

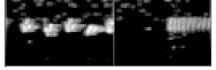
Train samples: 113568 (of which 5/6th [94640] is augmented data and 1/6th [18928] is original data)

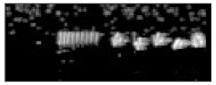
Validation samples: 2366

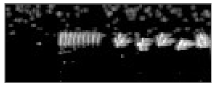
Test samples: 2367



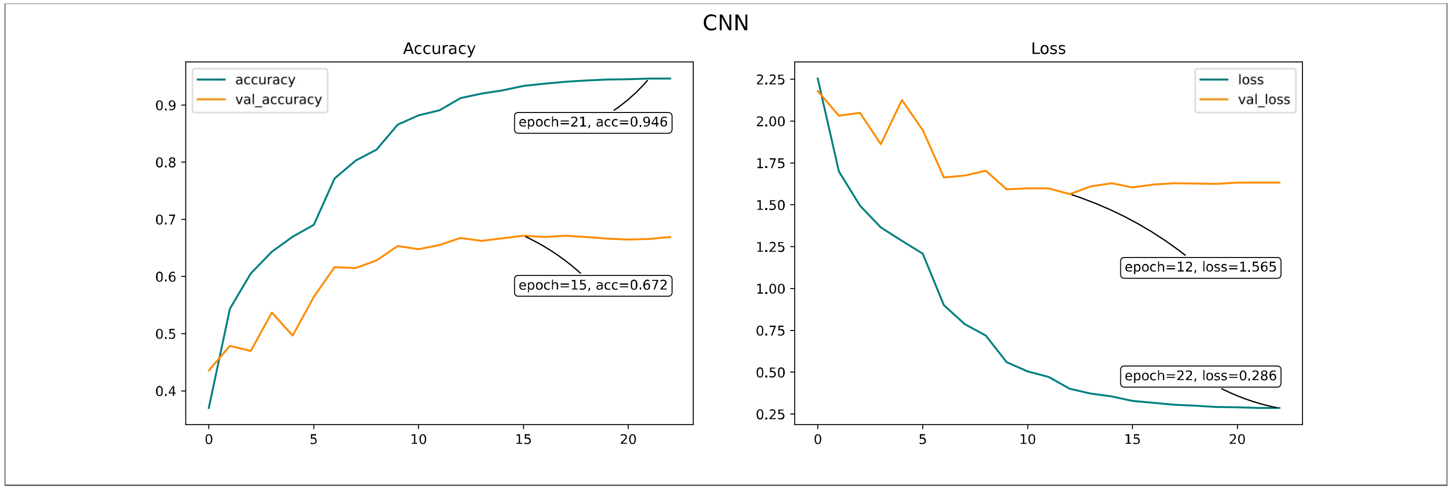




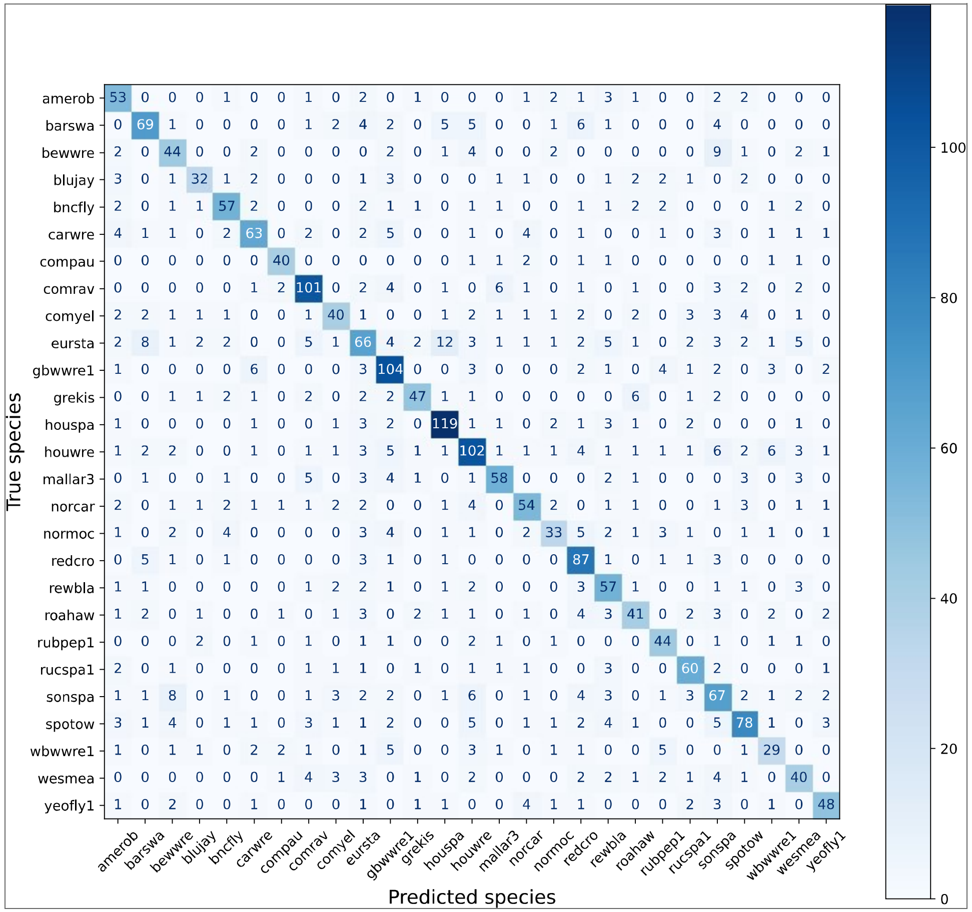




ML Models - CNN



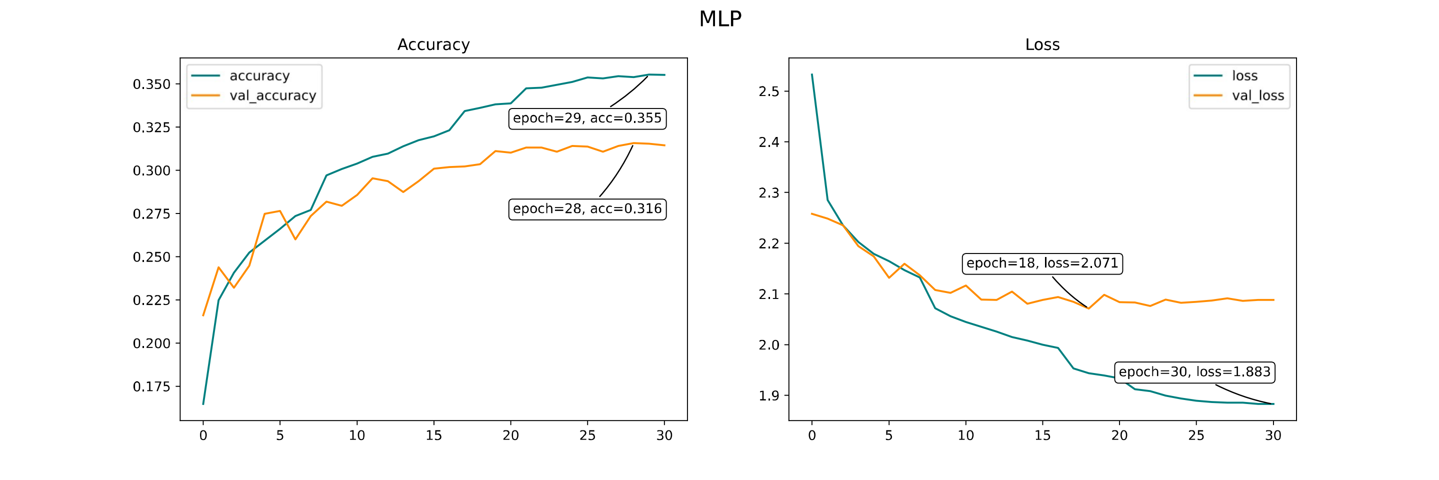
ML Models - CNN



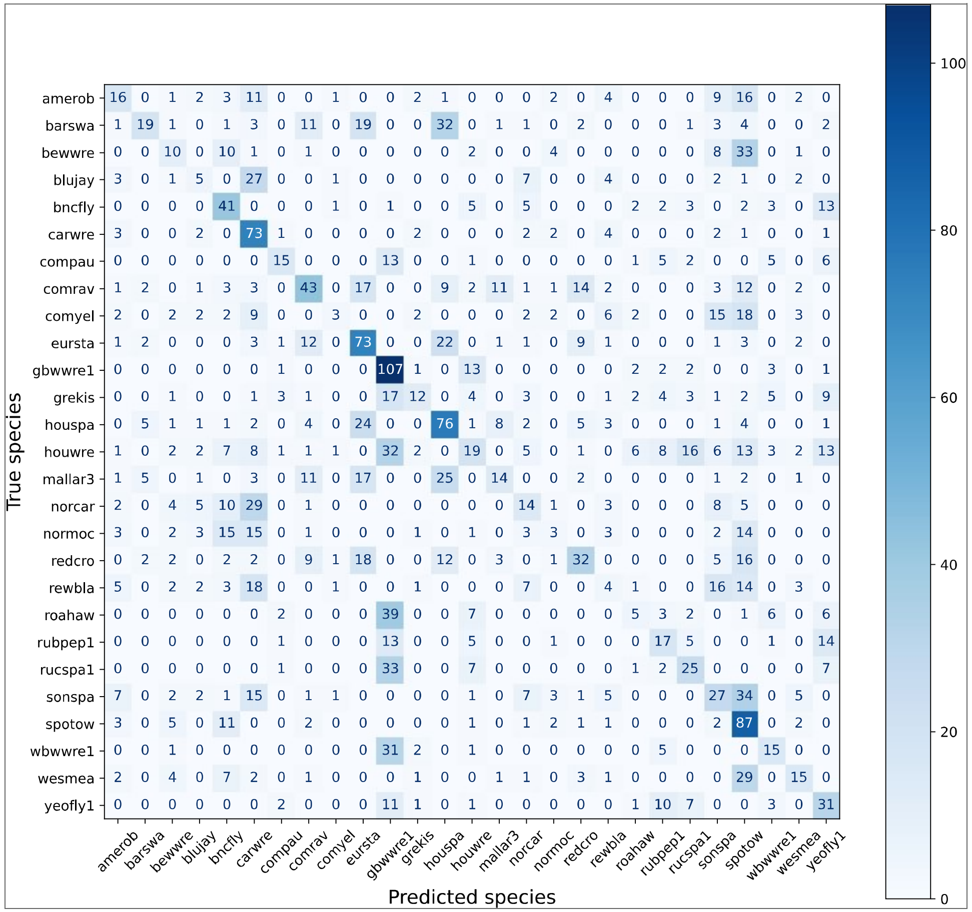
Classification report

	precision	recall	f1-score	support
amerob	0.63	0.76	0.69	70
barswa	0.74	0.68	0.71	101
bewwre	0.60	0.63	0.62	70
blujay	0.76	0.60	0.67	53
bncfly	0.75	0.73	0.74	78
carwre	0.74	0.68	0.71	93
compau	0.85	0.83	0.84	48
comrav	0.77	0.80	0.78	127
comyel	0.69	0.57	0.62	70
eursta	0.56	0.50	0.53	132
gbwwre1	0.68	0.79	0.73	132
grekis	0.81	0.68	0.74	69
houspa	0.80	0.86	0.83	139
houwre	0.67	0.68	0.68	149
mallar3	0.78	0.70	0.74	83
norcar	0.71	0.66	0.68	82
normoc	0.67	0.50	0.57	66
redcro	0.66	0.83	0.74	105
rewbla	0.60	0.74	0.66	77
roahaw	0.64	0.58	0.61	71
rubpep1	0.67	0.77	0.72	57
rucspa1	0.74	0.79	0.76	76
sonspa	0.53	0.60	0.56	112
spotow	0.74	0.66	0.70	118
wbwwre1	0.59	0.53	0.56	55
wesmea	0.59	0.60	0.59	67
yeofly1	0.76	0.72	0.74	67
accuracy			0.69	2367
macro avg	0.69	0.68	0.69	2367
weighted avg	0.69	0.69	0.69	2367

ML Models - MLP



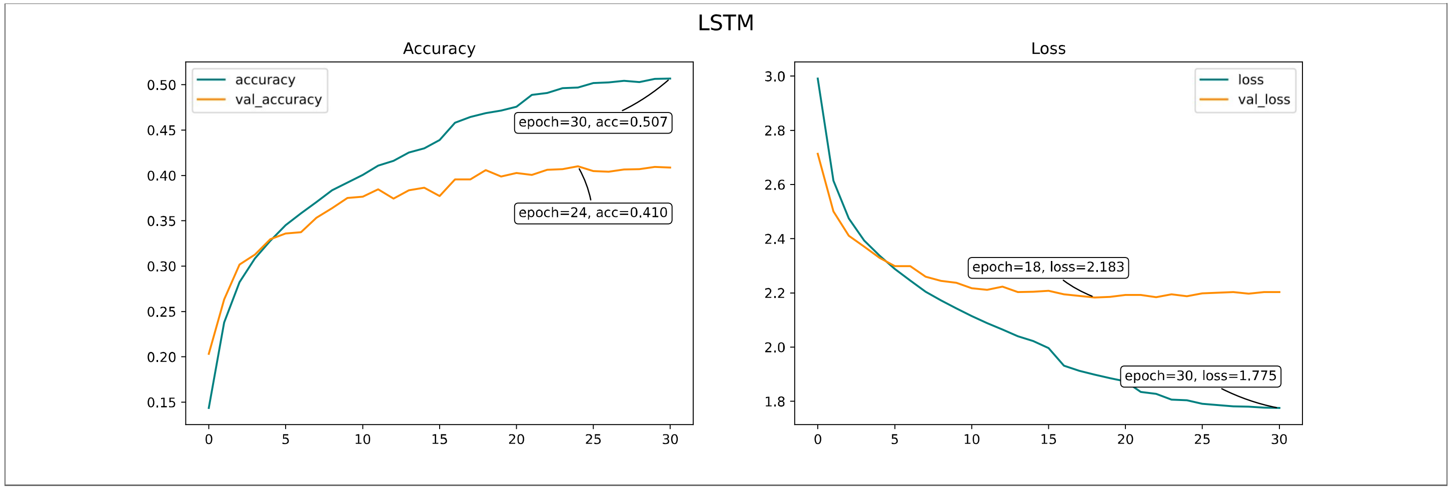
ML Models - MLP



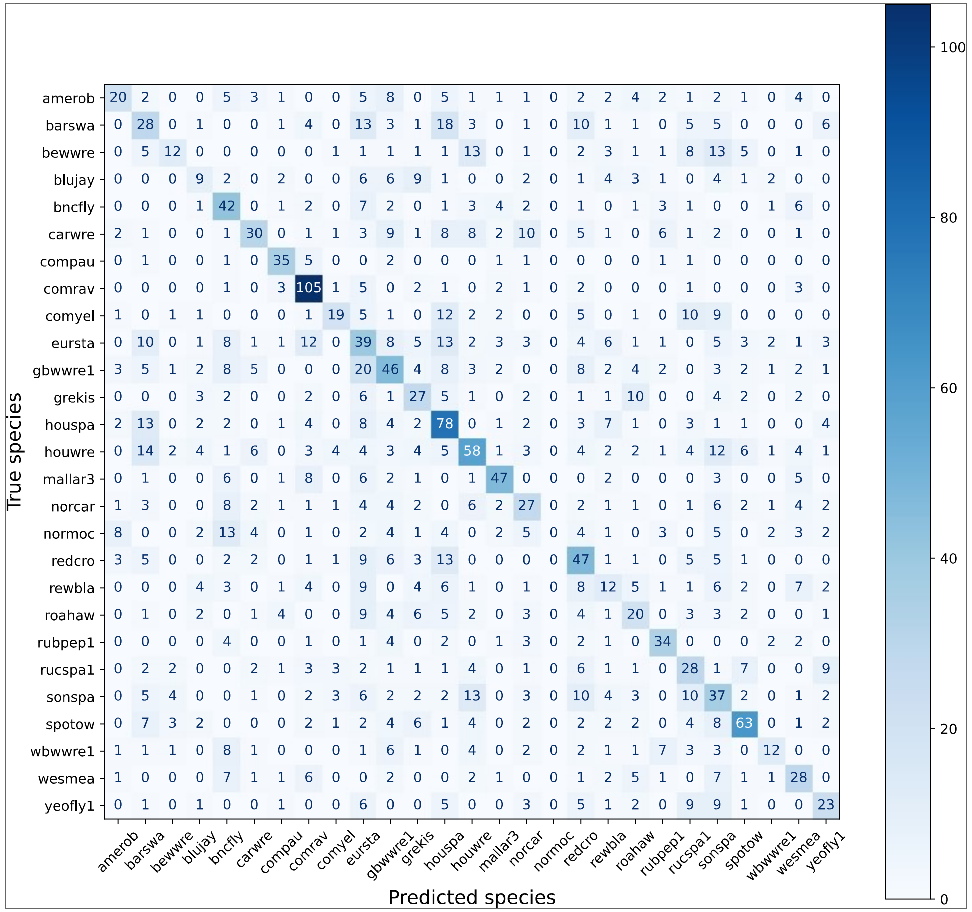
Classification report

	precision	recall	f1-score	support
amerob	0.31	0.23	0.26	70
barswa	0.54	0.19	0.28	101
bewwre	0.24	0.14	0.18	70
blujay	0.18	0.09	0.12	53
bncfly	0.35	0.53	0.42	78
carwre	0.32	0.78	0.46	93
compau	0.54	0.31	0.39	48
comrav	0.43	0.34	0.38	127
comyel	0.30	0.04	0.07	70
eursta	0.43	0.55	0.49	132
gbwwre1	0.36	0.81	0.50	132
grekis	0.44	0.17	0.25	69
houspa	0.43	0.55	0.48	139
houwre	0.27	0.13	0.17	149
mallar3	0.36	0.17	0.23	83
norcar	0.23	0.17	0.19	82
normoc	0.14	0.05	0.07	66
redcro	0.46	0.30	0.37	105
rewbla	0.10	0.05	0.07	77
roahaw	0.22	0.07	0.11	71
rubpep1	0.29	0.30	0.30	57
rucspa1	0.38	0.33	0.35	76
sonspa	0.24	0.24	0.24	112
spotow	0.28	0.74	0.41	118
wbwwre1	0.34	0.27	0.30	55
wesmea	0.38	0.22	0.28	67
yeofly1	0.30	0.46	0.36	67
accuracy			0.34	2367
macro avg	0.33	0.31	0.29	2367
weighted avg	0.34	0.34	0.31	2367

ML Models - LSTM



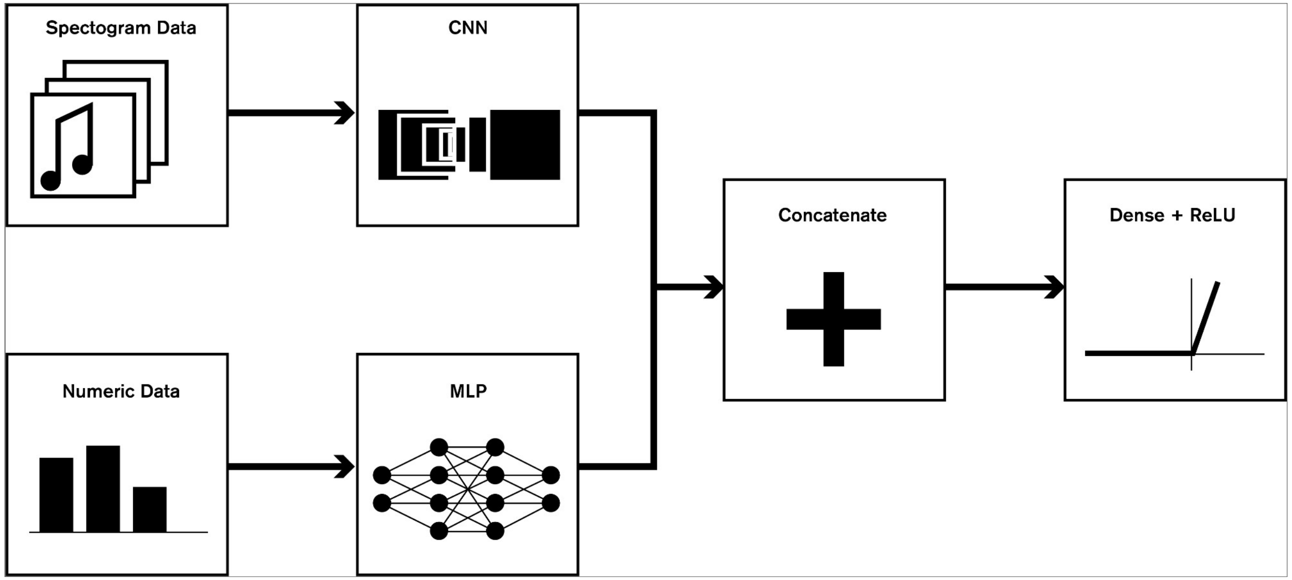
ML Models - LSTM



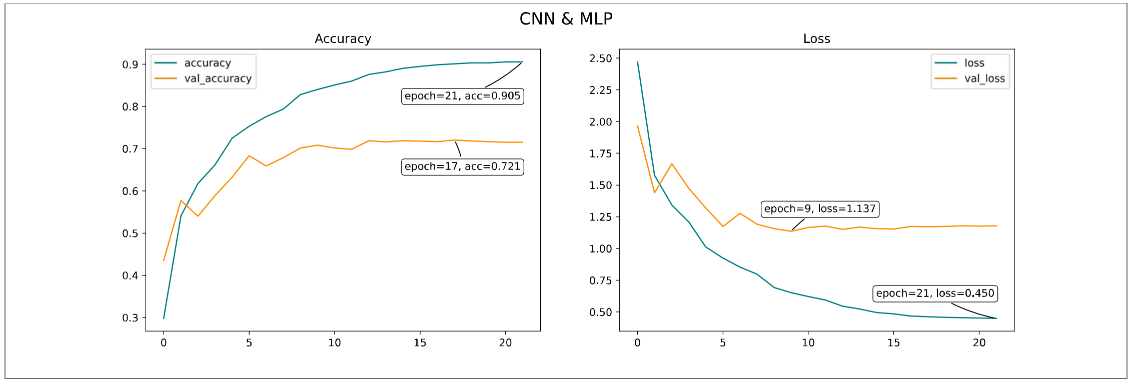
Classification report

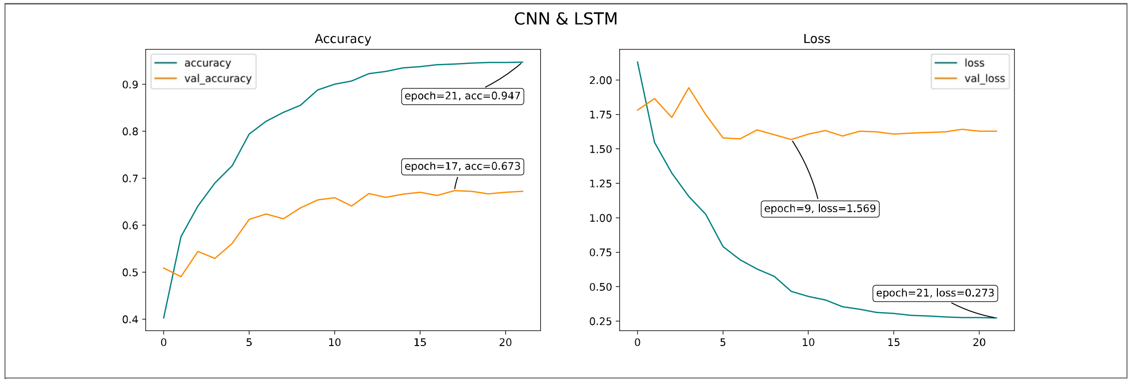
	precision	recall	f1-score	support
amerob	0.48	0.29	0.36	70
barswa	0.27	0.28	0.27	101
bewwre	0.46	0.17	0.25	70
blujay	0.26	0.17	0.20	53
bncfly	0.34	0.54	0.42	78
carwre	0.51	0.32	0.39	93
compau	0.64	0.73	0.68	48
comrav	0.62	0.83	0.71	127
comyel	0.54	0.27	0.36	70
eursta	0.22	0.30	0.25	132
gbwwre1	0.35	0.35	0.35	132
grekis	0.33	0.39	0.36	69
houspa	0.40	0.56	0.47	139
houwre	0.44	0.39	0.41	149
mallar3	0.65	0.57	0.61	83
norcar	0.34	0.33	0.34	82
normoc	0.00	0.00	0.00	66
redcro	0.33	0.45	0.38	105
rewbla	0.20	0.16	0.18	77
roahaw	0.29	0.28	0.28	71
rubpep1	0.53	0.60	0.56	57
rucspa1	0.28	0.37	0.32	76
sonspa	0.24	0.33	0.28	112
spotow	0.62	0.53	0.57	118
wbwwre1	0.48	0.22	0.30	55
wesmea	0.37	0.42	0.39	67
yeofly1	0.40	0.34	0.37	67
accuracy			0.39	2367
macro avg	0.39	0.38	0.37	2367
weighted avg	0.39	0.39	0.38	2367

ML Models - Mixed

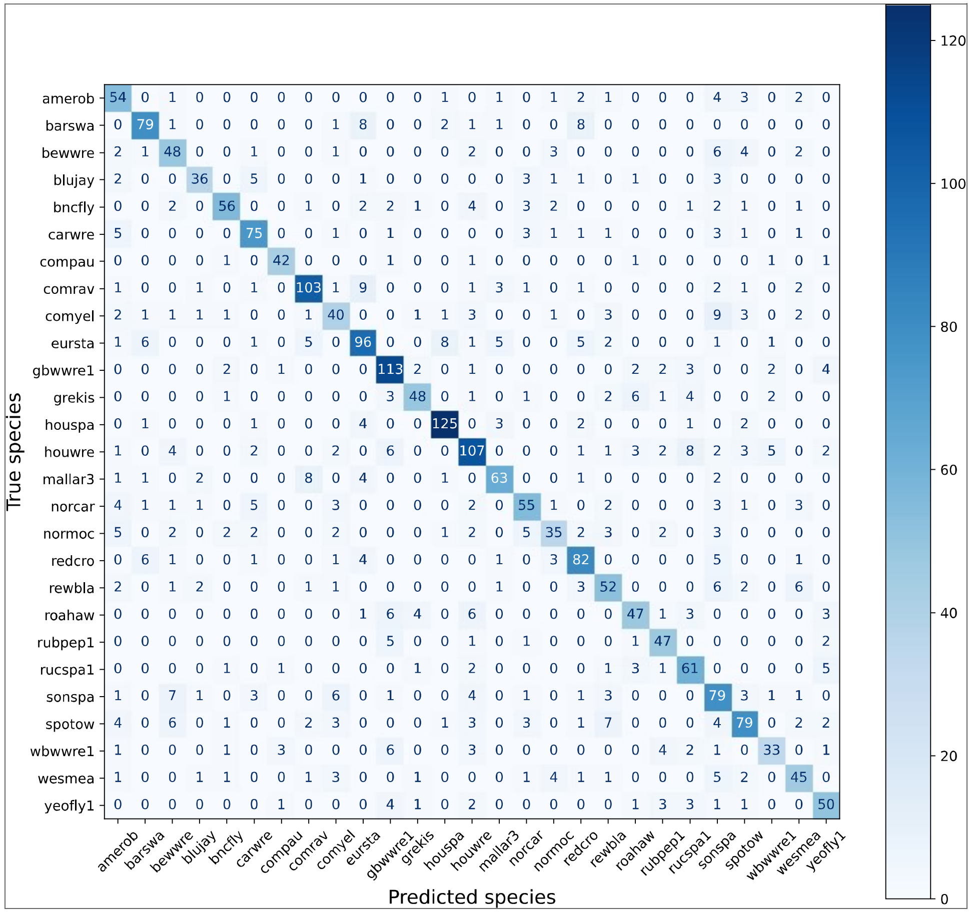


ML Models - Mixed





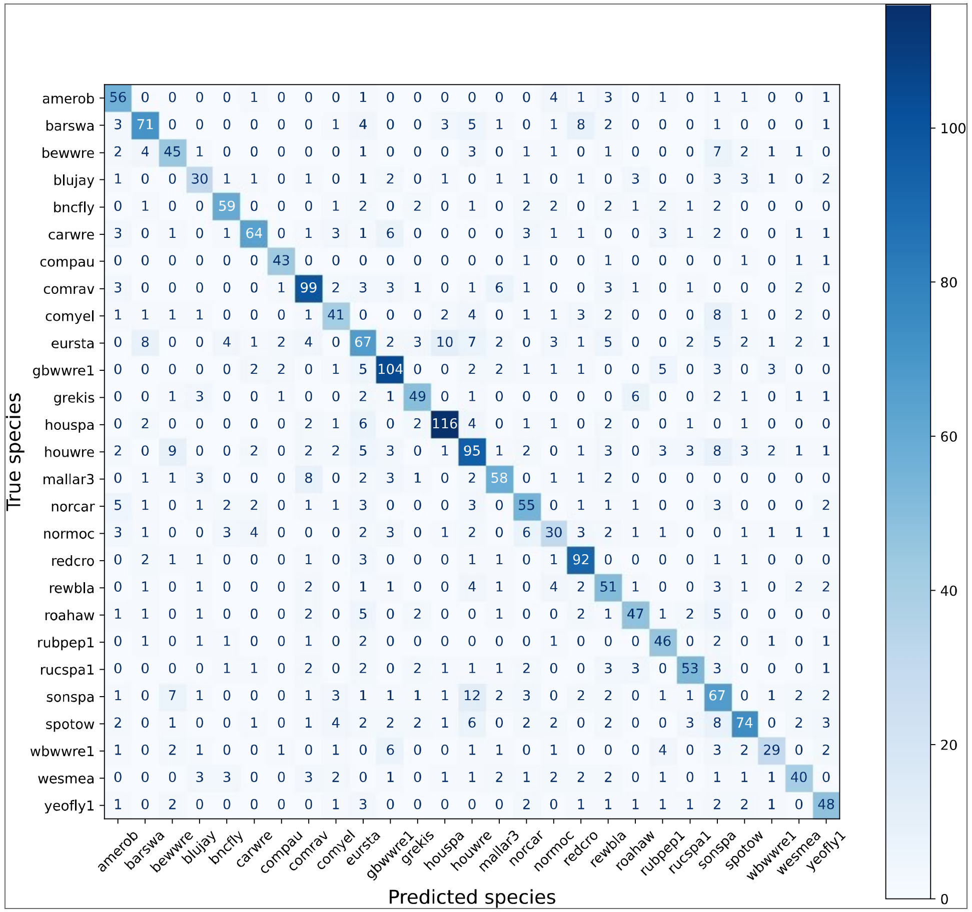
ML Models - CNN & MLP



Classification report

	precision	recall	f1-score	support
amerob	0.62	0.77	0.69	70
barswa	0.82	0.78	0.80	101
bewwre	0.64	0.69	0.66	70
blujay	0.80	0.68	0.73	53
bncfly	0.84	0.72	0.77	78
carwre	0.77	0.81	0.79	93
compau	0.88	0.88	0.88	48
comrav	0.84	0.81	0.83	127
comyel	0.62	0.57	0.59	70
eursta	0.74	0.73	0.74	132
gbwwre1	0.76	0.86	0.81	132
grekis	0.81	0.70	0.75	69
houspa	0.89	0.90	0.90	139
houwre	0.73	0.72	0.72	149
mallar3	0.81	0.76	0.78	83
norcar	0.71	0.67	0.69	82
normoc	0.67	0.53	0.59	66
redcro	0.73	0.78	0.76	105
rewbla	0.66	0.68	0.67	77
roahaw	0.72	0.66	0.69	71
rubpep1	0.75	0.82	0.78	57
rucspa1	0.71	0.80	0.75	76
sonspa	0.56	0.71	0.62	112
spotow	0.75	0.67	0.71	118
wbwwre1	0.73	0.60	0.66	55
wesmea	0.66	0.67	0.67	67
yeofly1	0.71	0.75	0.73	67
accuracy			0.74	2367
macro avg	0.74	0.73	0.73	2367
weighted avg	0.74	0.74	0.74	2367

ML Models - CNN & LSTM

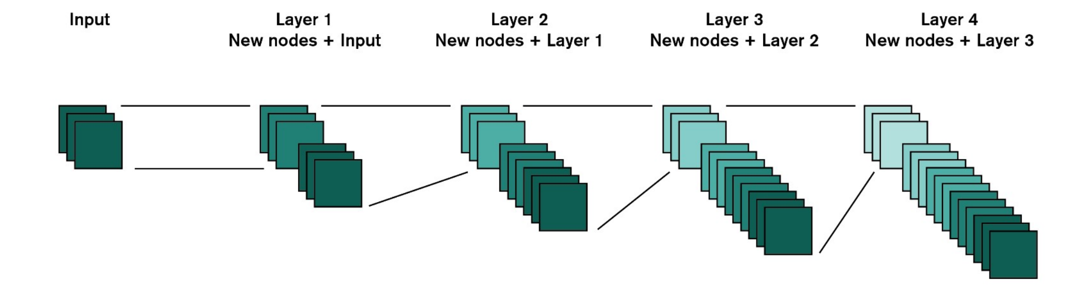


Classification report

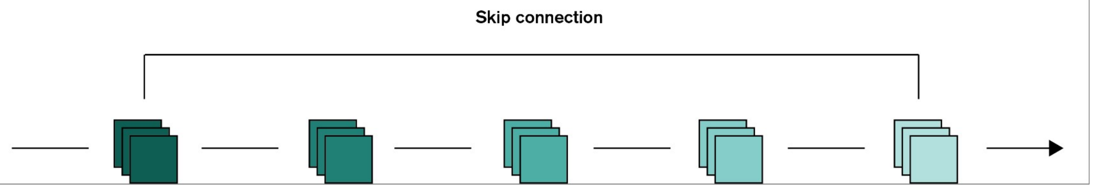
	precision	recall	f1-score	support
amerob	0.66	0.80	0.72	70
barswa	0.75	0.70	0.72	101
bewwre	0.63	0.64	0.64	70
blujay	0.62	0.57	0.59	53
bncfly	0.79	0.76	0.77	78
carwre	0.81	0.69	0.74	93
compau	0.86	0.90	0.88	48
comrav	0.75	0.78	0.76	127
comyel	0.64	0.59	0.61	70
eursta	0.54	0.51	0.52	132
gbwwre1	0.75	0.79	0.77	132
grekis	0.75	0.71	0.73	69
houspa	0.84	0.83	0.84	139
houwre	0.61	0.64	0.62	149
mallar3	0.72	0.70	0.71	83
norcar	0.65	0.67	0.66	82
normoc	0.53	0.45	0.49	66
redcro	0.75	0.88	0.81	105
rewbla	0.56	0.66	0.61	77
roahaw	0.72	0.66	0.69	71
rubpep1	0.67	0.81	0.73	57
rucspa1	0.77	0.70	0.73	76
sonspa	0.48	0.60	0.53	112
spotow	0.76	0.63	0.69	118
wbwwre1	0.69	0.53	0.60	55
wesmea	0.69	0.60	0.64	67
yeofly1	0.68	0.72	0.70	67
accuracy			0.69	2367
macro avg	0.69	0.68	0.69	2367
weighted avg	0.69	0.69	0.69	2367

ML Models - Transfer Learning

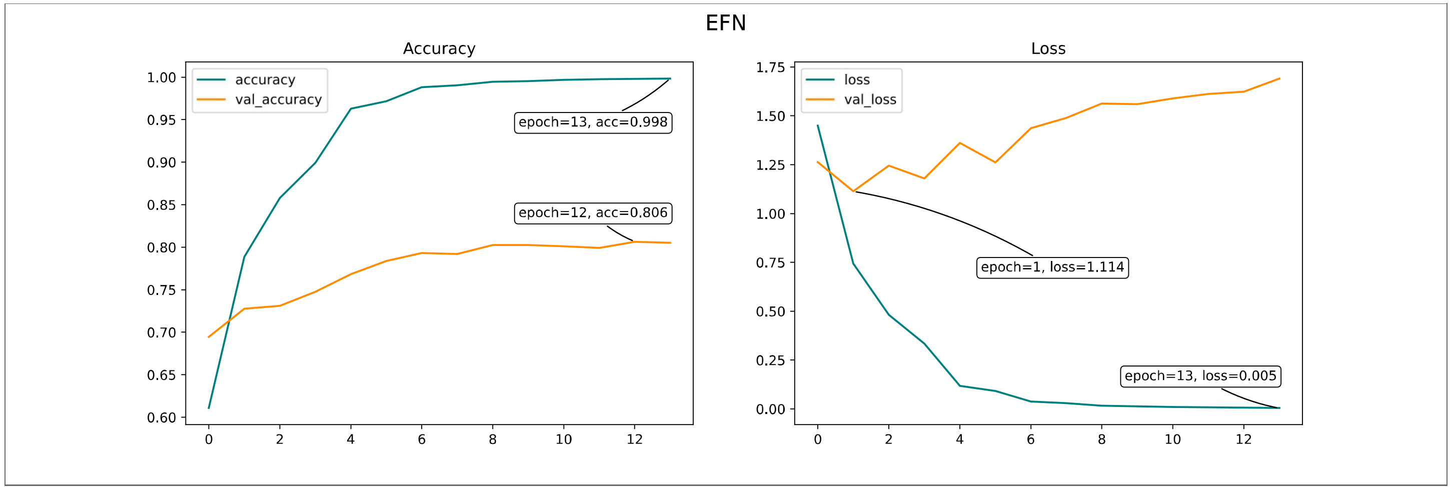
DenseNet



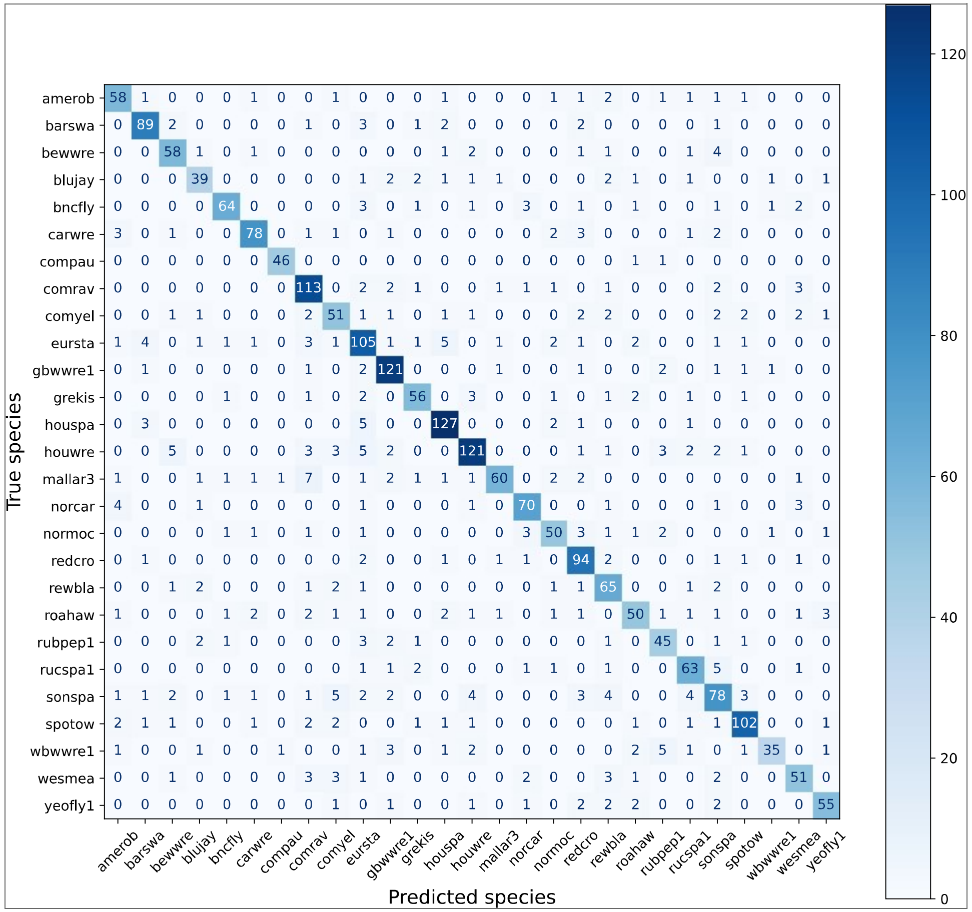
ResNet



ML Models - EfficientNet B1



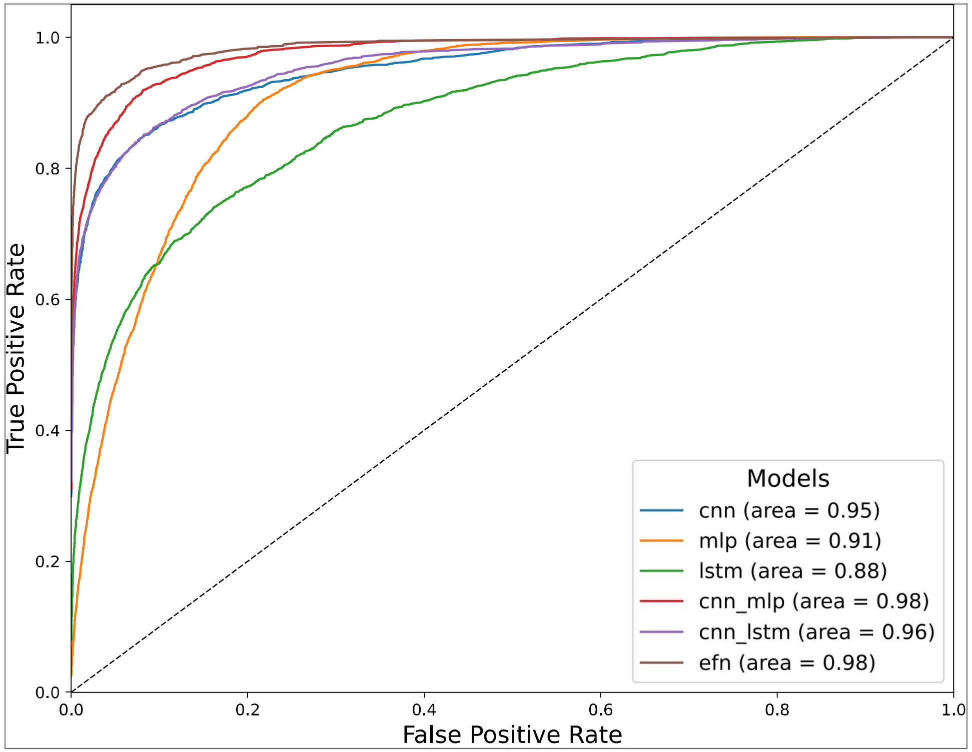
ML Models - EfficientNet B1



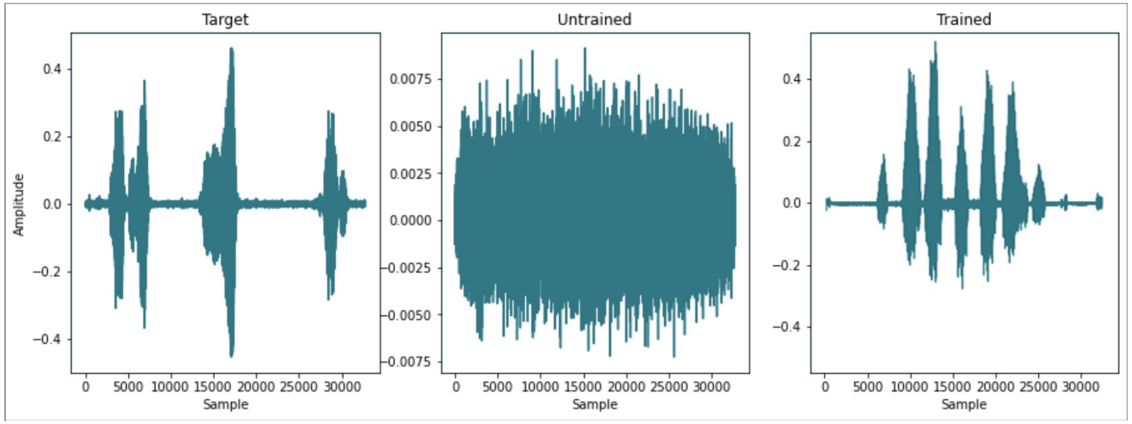
Classification report

	precision	recall	f1-score	support
amerob	0.81	0.83	0.82	70
barswa	0.88	0.88	0.88	101
bewwre	0.81	0.83	0.82	70
blujay	0.80	0.74	0.76	53
bncfly	0.90	0.82	0.86	78
carwre	0.90	0.84	0.87	93
compau	0.96	0.96	0.96	48
comrav	0.80	0.89	0.84	127
comyel	0.72	0.73	0.72	70
eursta	0.73	0.80	0.76	132
gbwwre1	0.86	0.92	0.89	132
grekis	0.84	0.81	0.82	69
houspa	0.88	0.91	0.90	139
houwre	0.86	0.81	0.84	149
mallar3	0.91	0.72	0.81	83
norcar	0.85	0.85	0.85	82
normoc	0.78	0.76	0.77	66
redcro	0.78	0.90	0.84	105
rewbla	0.72	0.84	0.78	77
roahaw	0.78	0.70	0.74	71
rubpep1	0.75	0.79	0.77	57
rucspa1	0.80	0.83	0.81	76
sonspa	0.70	0.70	0.70	112
spotow	0.89	0.86	0.88	118
wbwwre1	0.90	0.64	0.74	55
wesmea	0.78	0.76	0.77	67
yeofly1	0.87	0.82	0.85	67
accuracy			0.82	2367
macro avg	0.82	0.81	0.82	2367
weighted avg	0.82	0.82	0.82	2367

ML Models - Macro-average ROC curve of multi-class predictors



ML Models - GAN



Convolutional GAN

0:00:00 / 24:16:21

WaveGAN

0:00:00 / 37:16:58

Summing up

Further challenges

- Testing on the soundscape dataset

Methods for improvement

- Vary length of audio snippets and resolution parameters
- Use different kinds of spectograms
- Use more numeric data
- Use a tree based algorithm on the numeric data
- Remove audio snippets without birdcalls from training data
- Distinguish between types of birdcalls
- Adjust for non-uniform class distribution
- Source separation

If we had more time...

- Automate hyperparameter optimization
- Try unsupervised clustering
- Try self-supervised wav2vec
- Score by top 3 appearences

References

BirdCLEF 2021 https://www.kaggle.com/c/birdclef-2021/overview
WaveGAN https://arxiv.org/abs/1802.04208v3
ResNet https://github.com/KaimingHe/deep-residual-networks
DenseNet https://arxiv.org/abs/1608.06993v5
EfficientNet https://arxiv.org/abs/1905.11946
Bird Vocalizations https://en.wikipedia.org/wiki/Bird_vocalization
Very good talk about audio classification https://www.youtube.com/watch?v=uCGROOUO_wY
Good talk specifically about this subject https://www.youtube.com/watch?v=pzmdOETnhI0

Appendix

Packages

For audio:

- Librosa (loading audio, generating spectograms)
- noisereduce (reducing noise)
- audiomentations (augmenting audio)
- PIL (handling spectogram images)

Machine Learning:

- tensorflow v2 (For building models)
- scikit-learn (Evaluating models)

Data:

- numpy
- pandas
- scipy

Visualization:

- matplotlib
- seaborn
- plotly

Utility:

- joblib
- tqdm

Optimization

Multi-processing

Loading all the audio files, preprossesing them, and saving the final mel-spectograms takes a lot of time.

In order to speed things up, we used *joblib* as to enable multi-processing for the procedure. We chose this framework because it is also what is used internally in the Librosa package, and as such yielded us the best results.

Using 6 cores as simultaneous workers, we recuded time usage by a factor 3, from 12 hours to 4 hours.

Hardware

In order to run our models more efficiently, we took advantage of Google Colab's GPU ressourcers. This sped things up significantly, by a factor 20 from 300 seconds per epoch to 15.

Additionally, we relied on Colab's large amount of RAM (25 GB) to train the mixed models (and even then it would sometimes crash).

Model summaries

CNN

```
Model: "model_19"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_22 (InputLayer)        [(None, 48, 128, 1)]      0
_________________________________________________________________
conv2d_8 (Conv2D)            (None, 46, 126, 16)       160
_________________________________________________________________
batch_normalization_8 (Batch (None, 46, 126, 16)       64
_________________________________________________________________
max_pooling2d_8 (MaxPooling2 (None, 23, 63, 16)        0
_________________________________________________________________
conv2d_9 (Conv2D)            (None, 21, 61, 32)        4640
_________________________________________________________________
batch_normalization_9 (Batch (None, 21, 61, 32)        128
_________________________________________________________________
max_pooling2d_9 (MaxPooling2 (None, 10, 30, 32)        0
_________________________________________________________________
conv2d_10 (Conv2D)           (None, 8, 28, 64)         18496
_________________________________________________________________
batch_normalization_10 (Batc (None, 8, 28, 64)         256
_________________________________________________________________
max_pooling2d_10 (MaxPooling (None, 4, 14, 64)         0
_________________________________________________________________
conv2d_11 (Conv2D)           (None, 2, 12, 128)        73856
_________________________________________________________________
batch_normalization_11 (Batc (None, 2, 12, 128)        512
_________________________________________________________________
max_pooling2d_11 (MaxPooling (None, 1, 6, 128)         0
```

MLP

```
Model: "model_20"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_23 (InputLayer)        [(None, 15)]              0
_________________________________________________________________
dense_58 (Dense)             (None, 64)                1024
_________________________________________________________________
dense_59 (Dense)             (None, 64)                4160
_________________________________________________________________
dense_60 (Dense)             (None, 64)                4160
_________________________________________________________________
dropout_32 (Dropout)         (None, 64)                0
_________________________________________________________________
dense_61 (Dense)             (None, 27)                1755
=================================================================
Total params: 11,099
Trainable params: 11,099
Non-trainable params: 0
_________________________________________________________________
```

LSTM

```
Model: "model_21"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_24 (InputLayer)        [(None, 48, 128)]         0
_________________________________________________________________
lstm_61 (LSTM)               (None, 48, 36)            23760
_________________________________________________________________
lstm_62 (LSTM)               (None, 48, 32)            8832
_________________________________________________________________
lstm_63 (LSTM)               (None, 48, 28)            6832
_________________________________________________________________
lstm_64 (LSTM)               (None, 48, 24)            5088
_________________________________________________________________
lstm_65 (LSTM)               (None, 48, 20)            3600
_________________________________________________________________
lstm_66 (LSTM)               (None, 16)                2368
_________________________________________________________________
dense_62 (Dense)             (None, 64)                1088
_________________________________________________________________
dropout_33 (Dropout)         (None, 64)                0
_________________________________________________________________
dense_63 (Dense)             (None, 64)                4160
_________________________________________________________________
dropout_34 (Dropout)         (None, 64)                0
_________________________________________________________________
dense_64 (Dense)             (None, 32)                2080
_________________________________________________________________
dropout_35 (Dropout)         (None, 32)                0
```

Merge layers

CNN & MLP

```
Model: "model_27"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
input_27 (InputLayer)           [(None, 48, 128, 1)] 0
__________________________________________________________________________________________________
conv2d_16 (Conv2D)              (None, 46, 126, 16)  160         input_27[0][0]
__________________________________________________________________________________________________
batch_normalization_16 (BatchNo (None, 46, 126, 16)  64          conv2d_16[0][0]
__________________________________________________________________________________________________
max_pooling2d_16 (MaxPooling2D) (None, 23, 63, 16)   0           batch_normalization_16[0][0]
__________________________________________________________________________________________________
conv2d_17 (Conv2D)              (None, 21, 61, 32)   4640        max_pooling2d_16[0][0]
__________________________________________________________________________________________________
batch_normalization_17 (BatchNo (None, 21, 61, 32)   128         conv2d_17[0][0]
__________________________________________________________________________________________________
max_pooling2d_17 (MaxPooling2D) (None, 10, 30, 32)   0           batch_normalization_17[0][0]
__________________________________________________________________________________________________
conv2d_18 (Conv2D)              (None, 8, 28, 64)    18496       max_pooling2d_17[0][0]
__________________________________________________________________________________________________
batch_normalization_18 (BatchNo (None, 8, 28, 64)    256         conv2d_18[0][0]
__________________________________________________________________________________________________
max_pooling2d_18 (MaxPooling2D) (None, 4, 14, 64)    0           batch_normalization_18[0][0]
__________________________________________________________________________________________________
conv2d_19 (Conv2D)              (None, 2, 12, 128)   73856       max_pooling2d_18[0][0]
__________________________________________________________________________________________________
batch_normalization_19 (BatchNo (None, 2, 12, 128)   512         conv2d_19[0][0]
__________________________________________________________________________________________________
max_pooling2d_19 (MaxPooling2D) (None, 1, 6, 128)    0           batch_normalization_19[0][0]
```

CNN & LSTM

```
Model: "model_33"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
input_31 (InputLayer)           [(None, 48, 128, 1)] 0
__________________________________________________________________________________________________
conv2d_24 (Conv2D)              (None, 46, 126, 16)  160         input_31[0][0]
__________________________________________________________________________________________________
batch_normalization_24 (BatchNo (None, 46, 126, 16)  64          conv2d_24[0][0]
__________________________________________________________________________________________________
max_pooling2d_24 (MaxPooling2D) (None, 23, 63, 16)   0           batch_normalization_24[0][0]
__________________________________________________________________________________________________
conv2d_25 (Conv2D)              (None, 21, 61, 32)   4640        max_pooling2d_24[0][0]
__________________________________________________________________________________________________
batch_normalization_25 (BatchNo (None, 21, 61, 32)   128         conv2d_25[0][0]
__________________________________________________________________________________________________
max_pooling2d_25 (MaxPooling2D) (None, 10, 30, 32)   0           batch_normalization_25[0][0]
__________________________________________________________________________________________________
conv2d_26 (Conv2D)              (None, 8, 28, 64)    18496       max_pooling2d_25[0][0]
__________________________________________________________________________________________________
input_32 (InputLayer)           [(None, 48, 128)]    0
__________________________________________________________________________________________________
batch_normalization_26 (BatchNo (None, 8, 28, 64)    256         conv2d_26[0][0]
__________________________________________________________________________________________________
lstm_73 (LSTM)                  (None, 48, 36)       23760       input_32[0][0]
__________________________________________________________________________________________________
max_pooling2d_26 (MaxPooling2D) (None, 4, 14, 64)    0           batch_normalization_26[0][0]
__________________________________________________________________________________________________
lstm_74 (LSTM)                  (None, 48, 32)       8832        lstm_73[0][0]
```

Species legend

Label	Name
amerob	American Robin
barswa	Barn Swallow
bewwre	Bewick's Wren
blujay	Blue Jay
bncfly	Brown-crested Flycatcher
carwre	Carolina Wren
compau	Common Pauraque
comrav	Common Raven
comyel	Common Yellowthroat
eursta	European Starling
gbwwre1	Gray-breasted Wood-Wren
grekis	Great Kiskadee
houspa	House Sparrow
houwre	House Wren
mallar3	Mallard
norcar	Northern Cardinal
normoc	Northern Mockingbird

Label	Name
redcro	Red Crossbill
rewbla	Red-winged Blackbird
roahaw	Roadside Hawk
rubpep1	Rufous-browed Peppershrike
rucspa1	Rufous-collared Sparrow
sonspa	Song Sparrow
spotow	Spotted Towhee
wbwwre1	White-breasted Wood-Wren
wesmea	Western Meadowlark
yeofly1	Yellow-olive Flycatcher