

FOUNDATIONS OF DEEP LEARNING

Università degli studi di Milano-Bicocca

2024-2025

Pedestrian and bicyclist classification



Elisa Princic 886476
Sasha Risoluti 870667
Andrea Torre 886051

INTRODUCTION

	SOURCE	<u>MATLAB dataset</u> of radar signatures
	DIMENSION	(25000 ,144 , 400)
	TARGET	5 balanced classes
	GOAL	Building a deep learning network to classify micro-doppler signals
	CHALLENGES	Traffic noise and double objects classes

DIMENSIONALITY REDUCTION

WHY?

- Heavy size
- RAM issues
- Training speed

HOW MUCH?

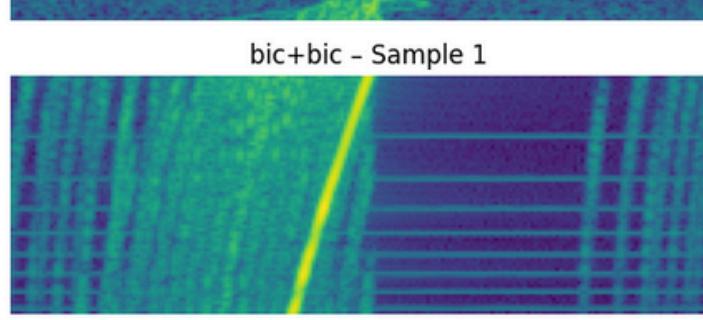
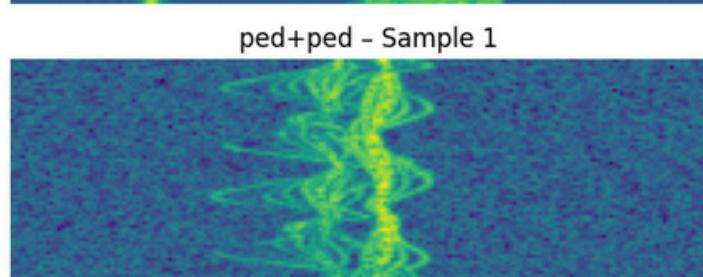
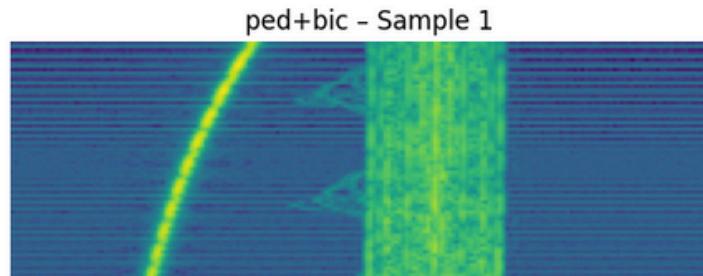
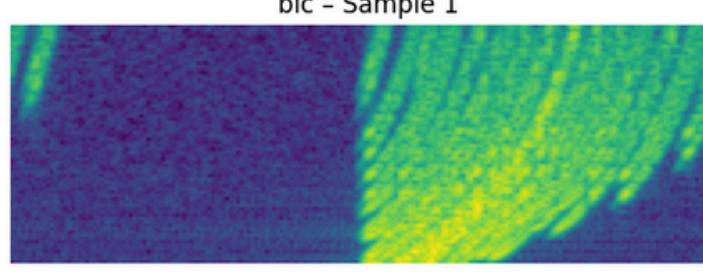
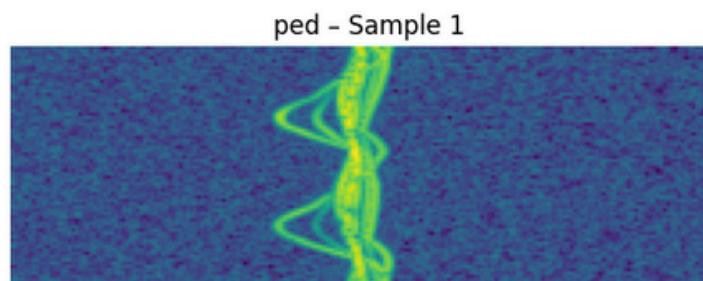
from 144x400
to 72x200

HOW?

tf.image.resize
with method **area**
→ average of
neighboring pixel

DATA EXPLORATION

SPECTROGRAM



MEAN

0.4786

0.4964

0.4987

0.4726

0.5205

STANDARD DEVIATION

0.1156

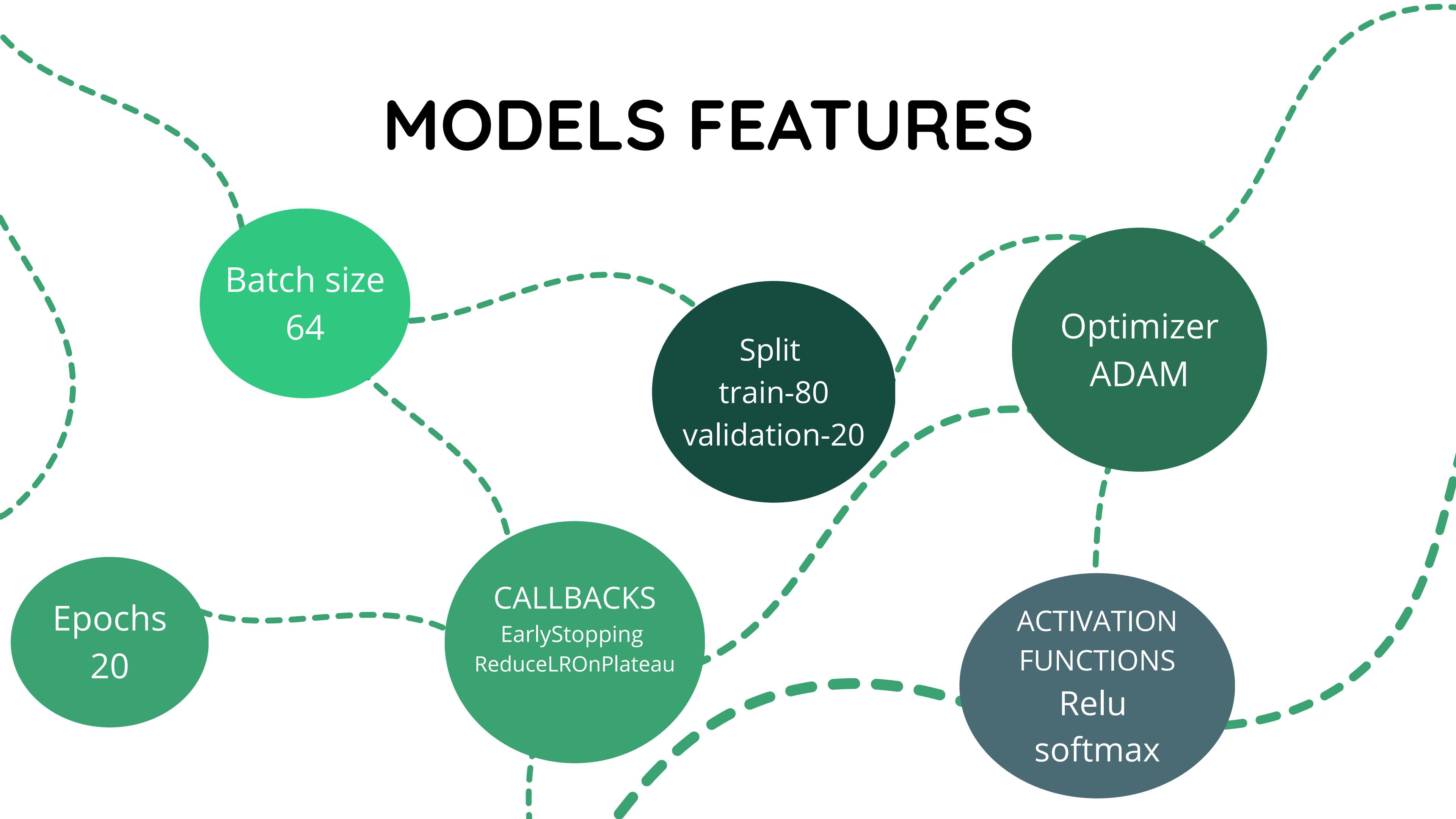
0.1486

0.1513

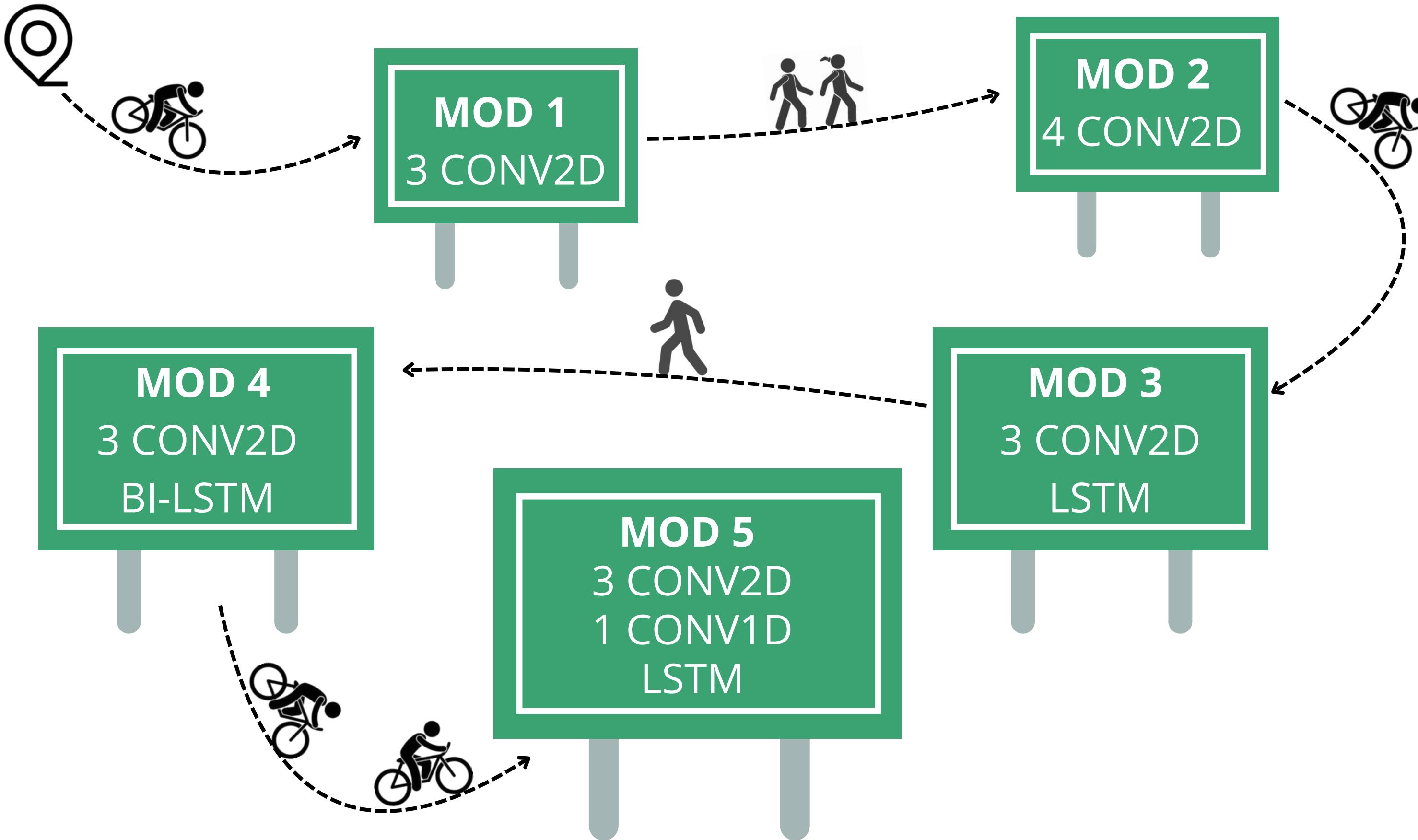
0.1231

0.1619

MODELS FEATURES



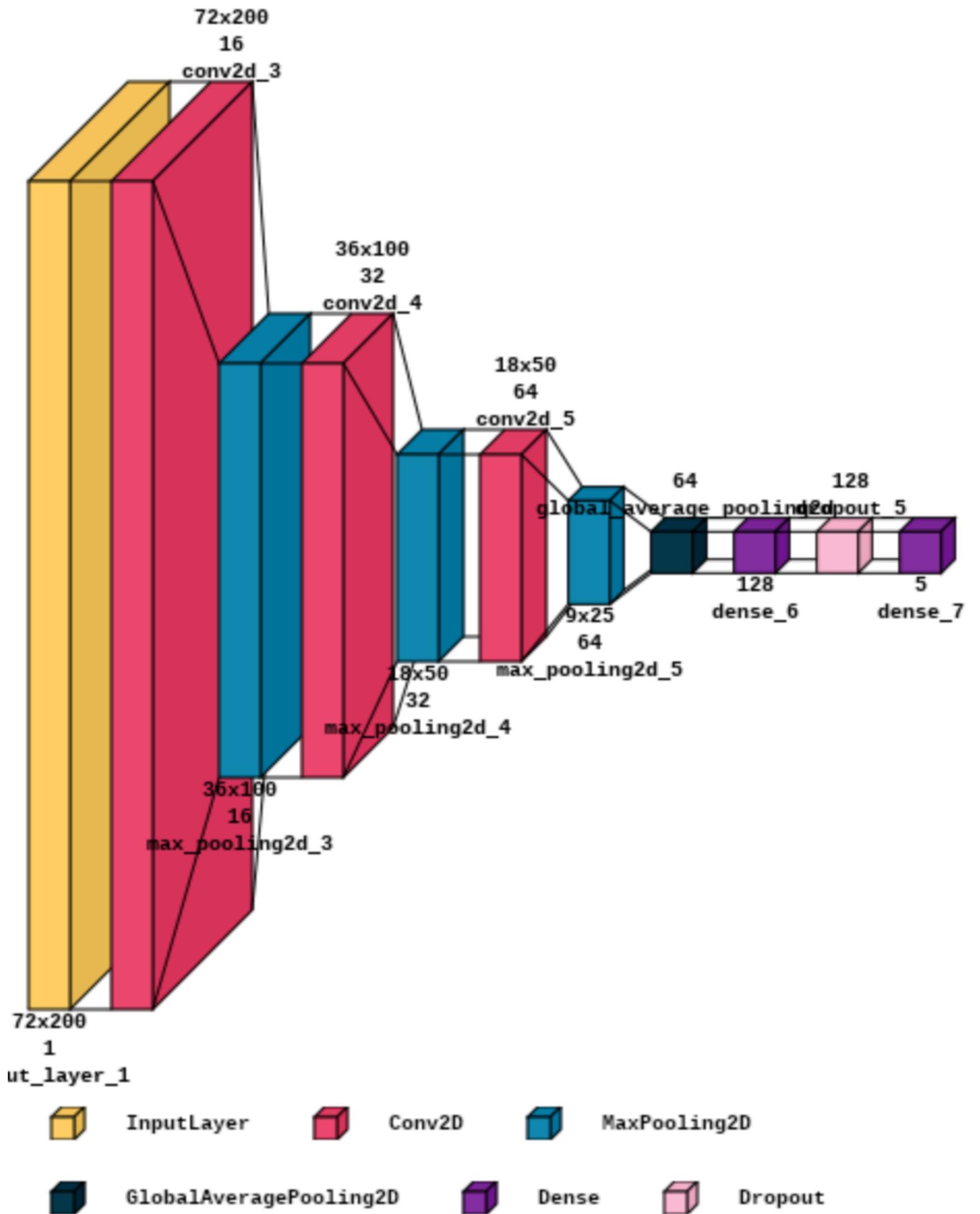
ARCHITECTURES



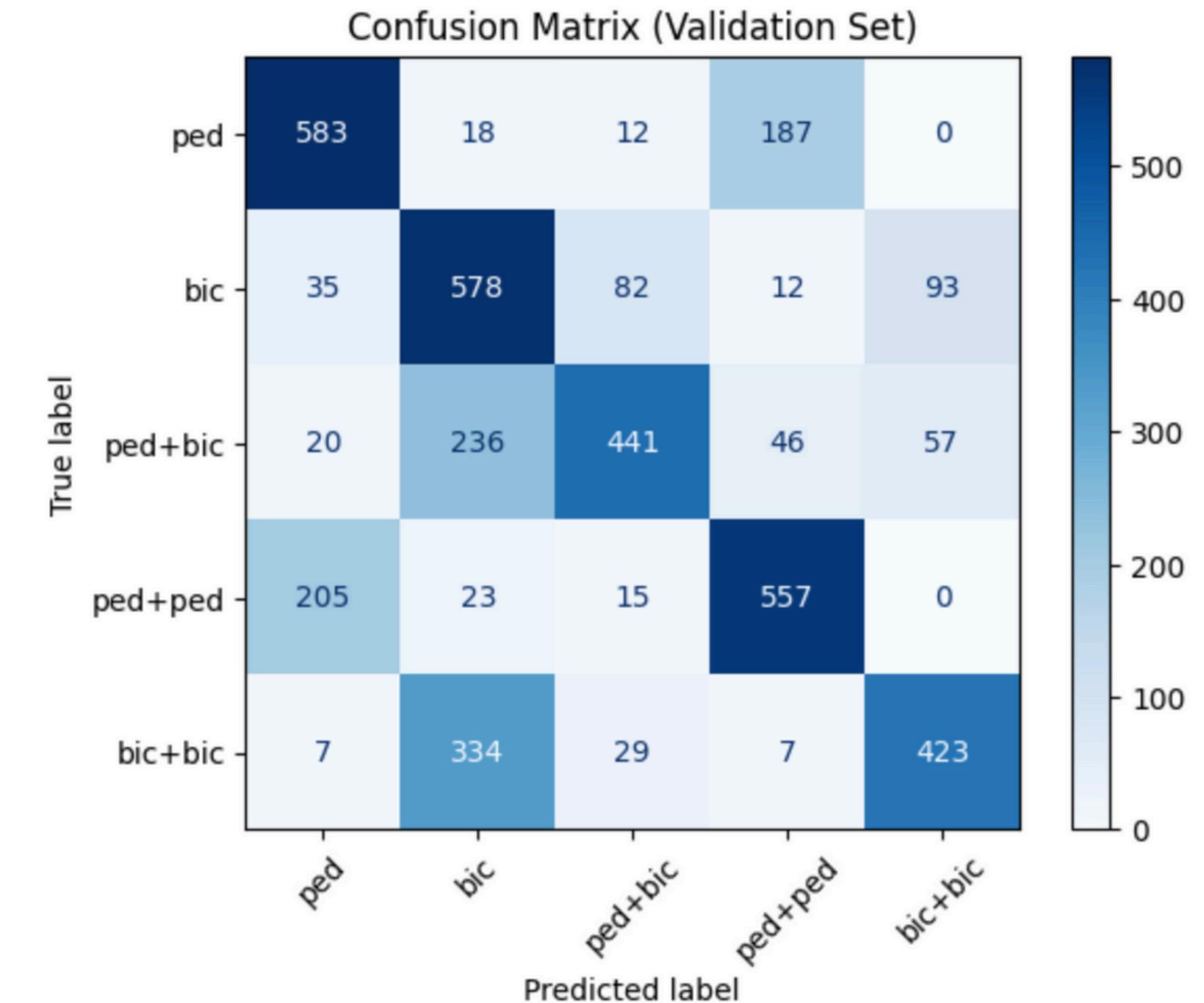
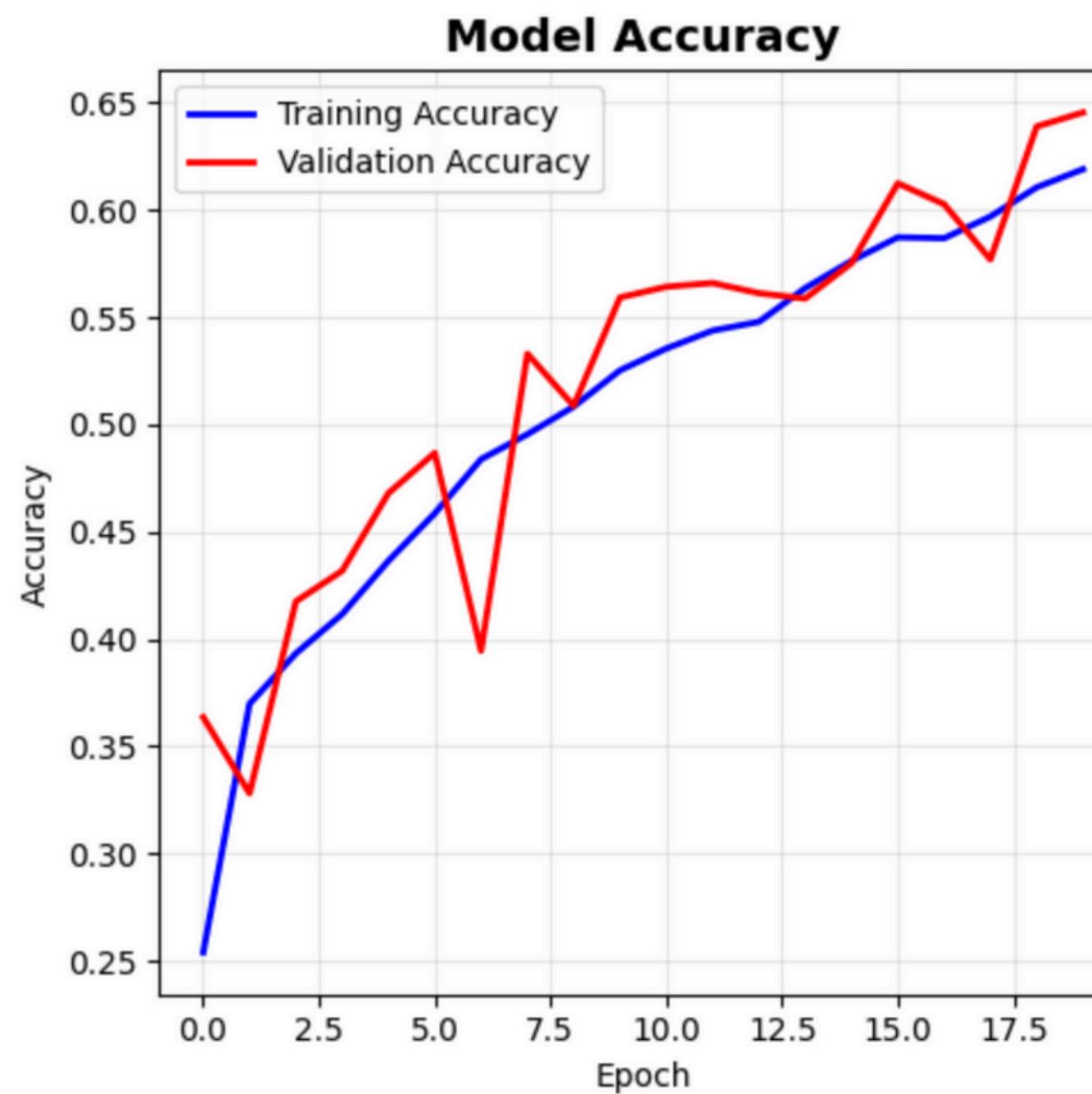
MODEL 1

Parameters: 32,261

The model has three Conv2D with 3x3 kernel size. The number of filters doubles at each layer. Each one is followed by a 2D max pooling (2x2). Lastly, we apply Global Average Pooling2D, a dense layer and dropout layer with rate of 0.3.



MODEL 1



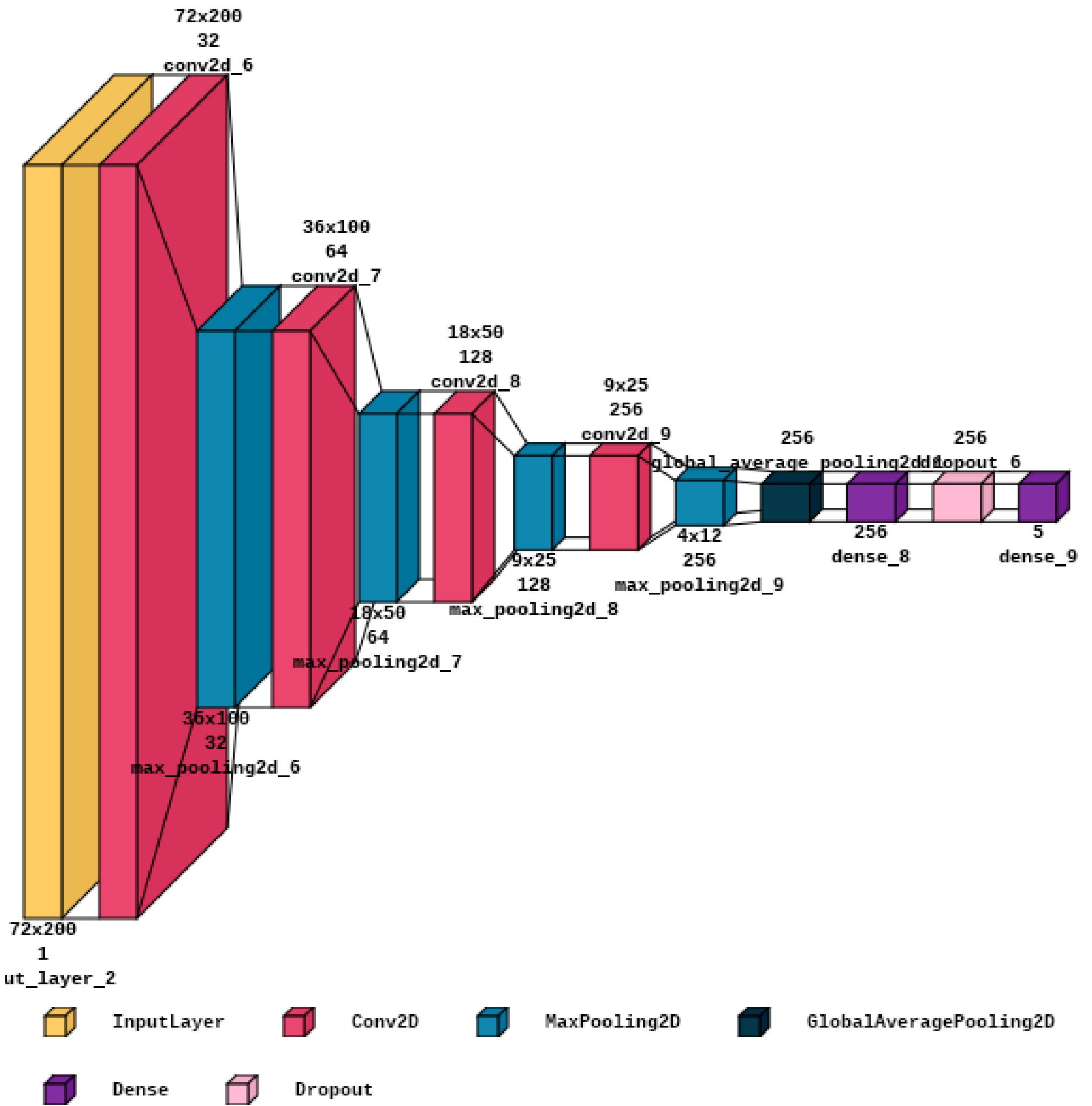
Train Accuracy: 0.622

Validation Accuracy: 0.6455

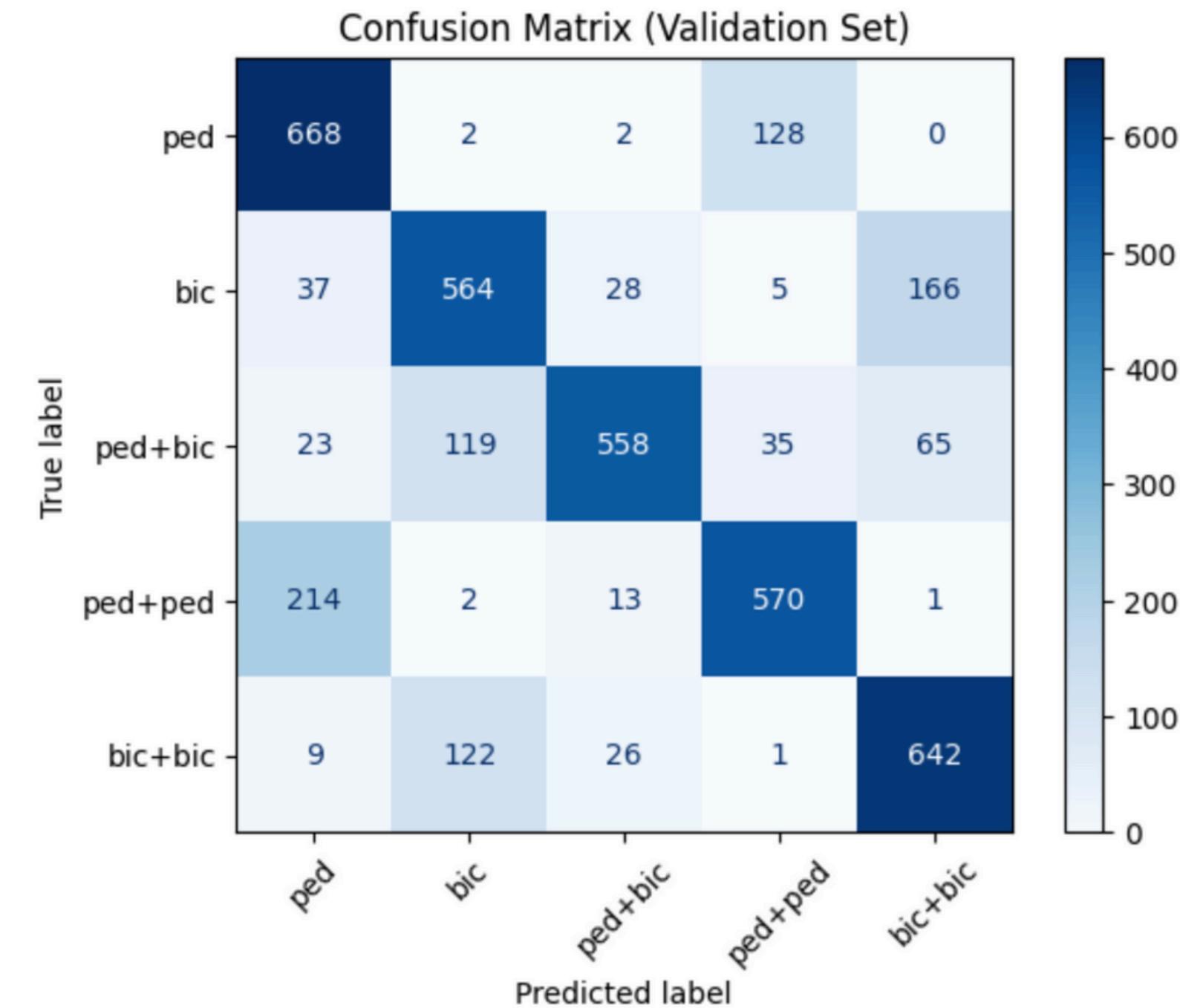
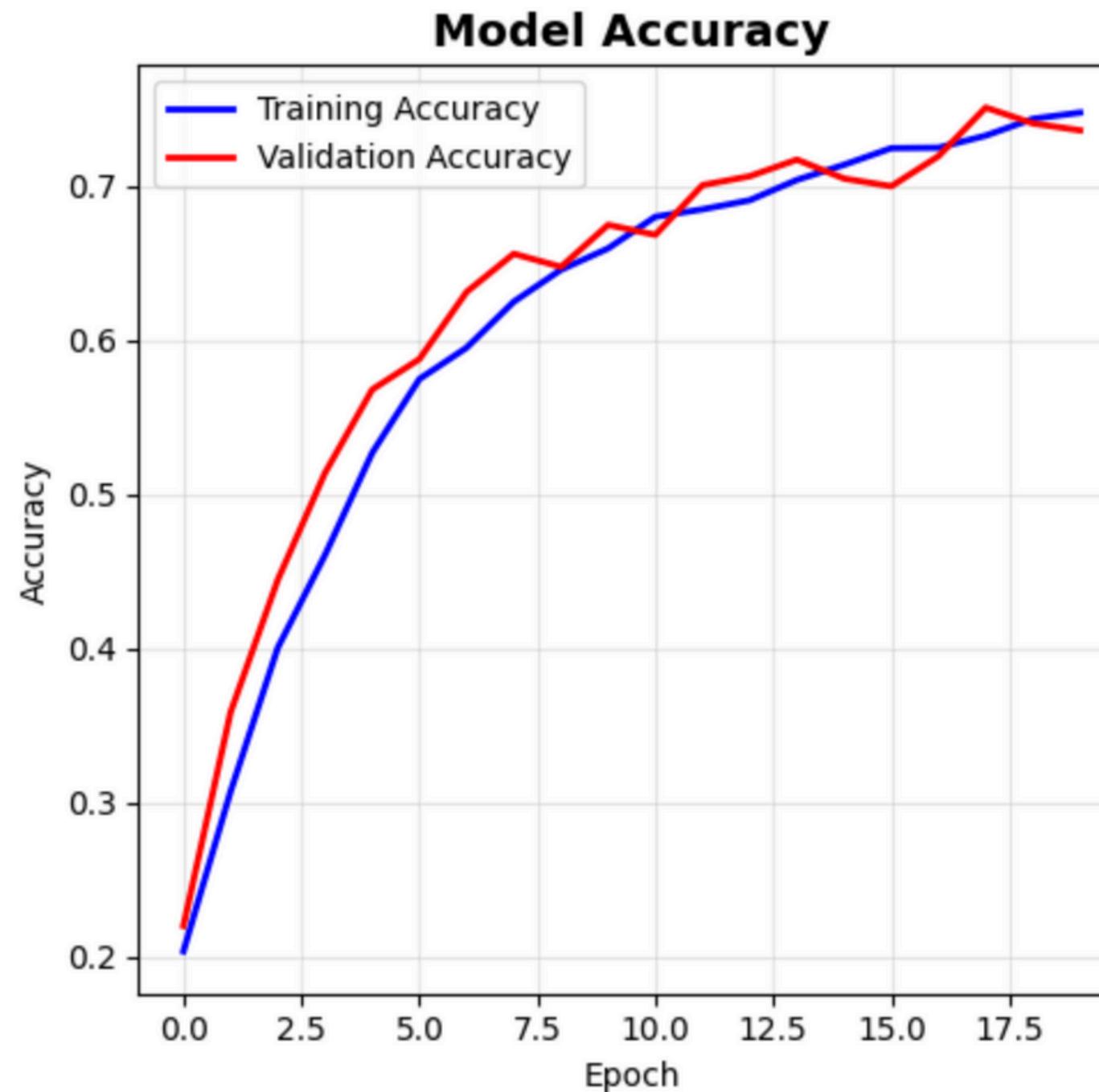
MODEL 2

Parameters: 454,917

We add complexity with a fourth 2D - convolutional layer with size 128. The dense layer now has 256 units. We increase dropout rate to 0.5. Everything else stays the same.



MODEL 2



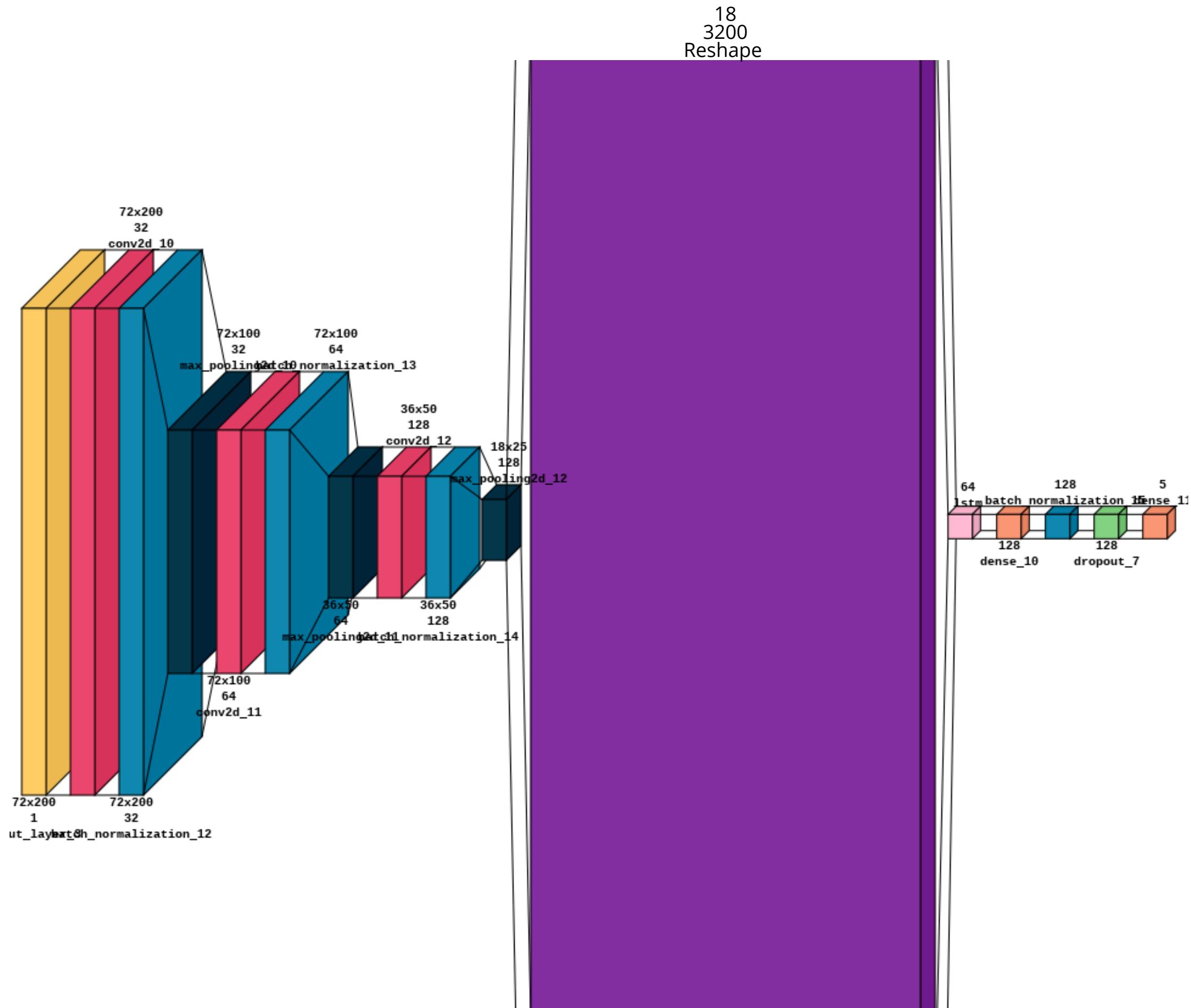
Train Accuracy: 0.7706
Validation Accuracy: 0.7505

MODEL 3

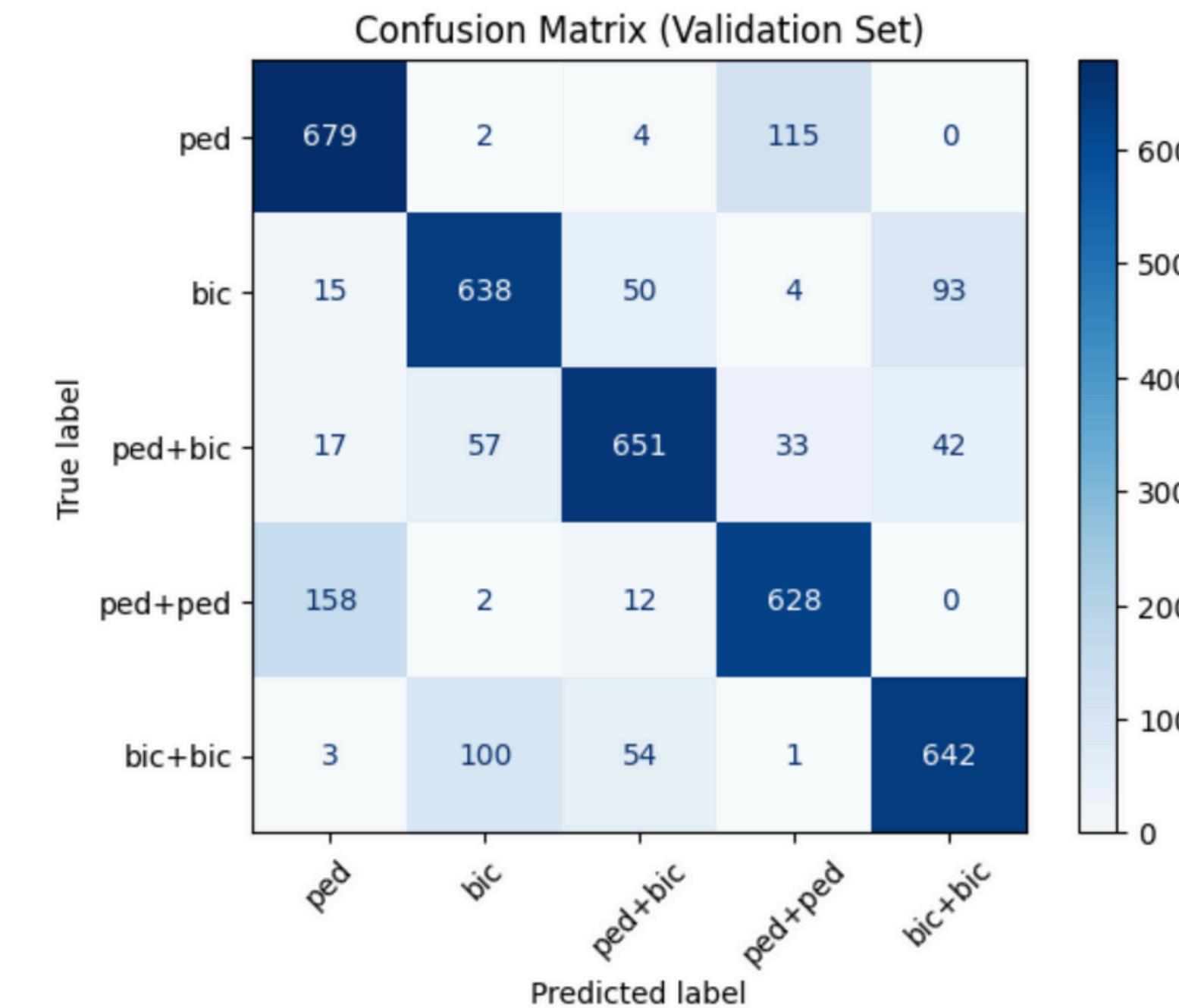
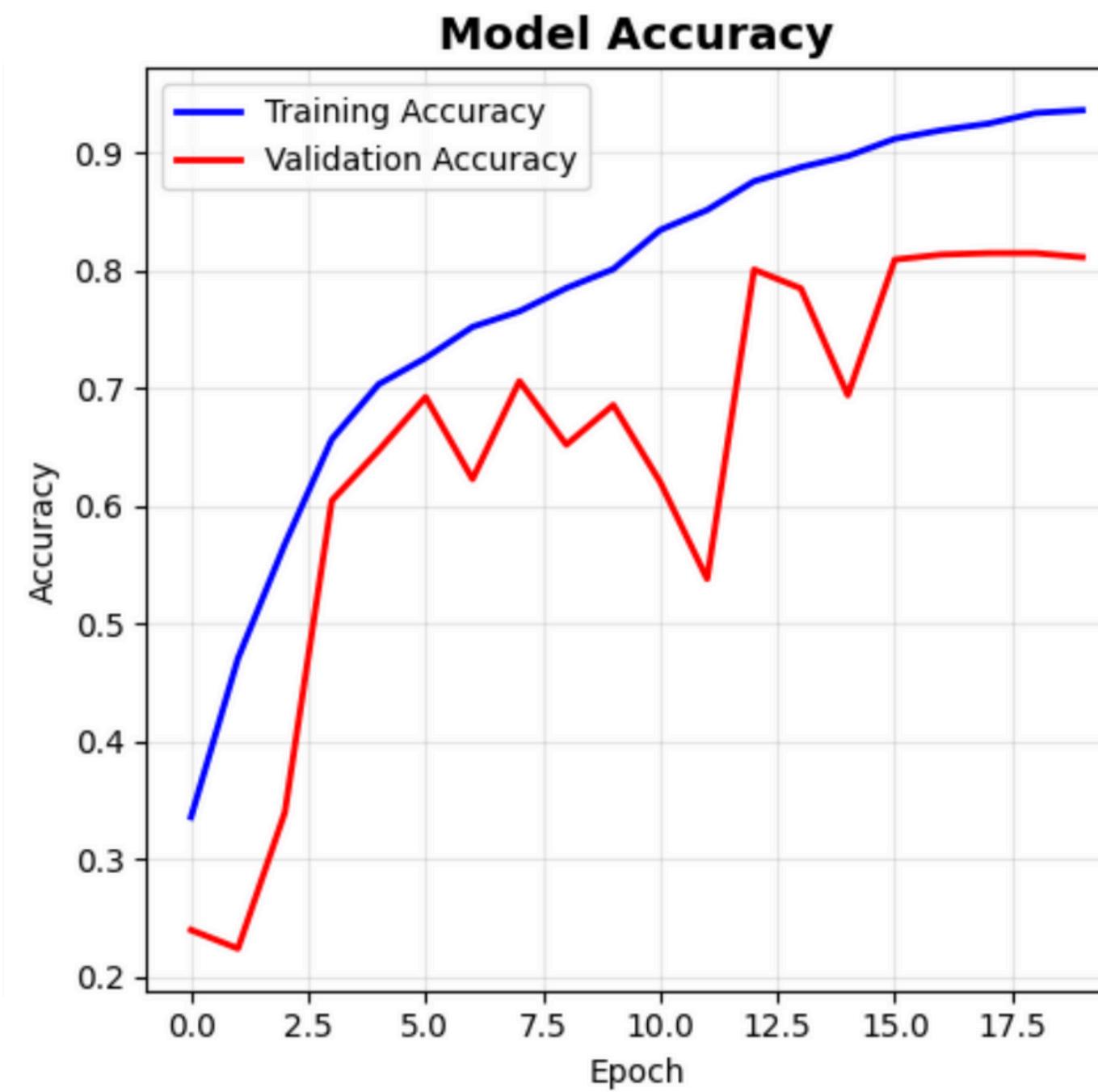
Parameters: 1,111,429

The spectrograms can be considered as time series, so an LSTM layer has been added.

The peculiar kernels have been chosen to catch the differences between the signals of pedestrians and bicycles.



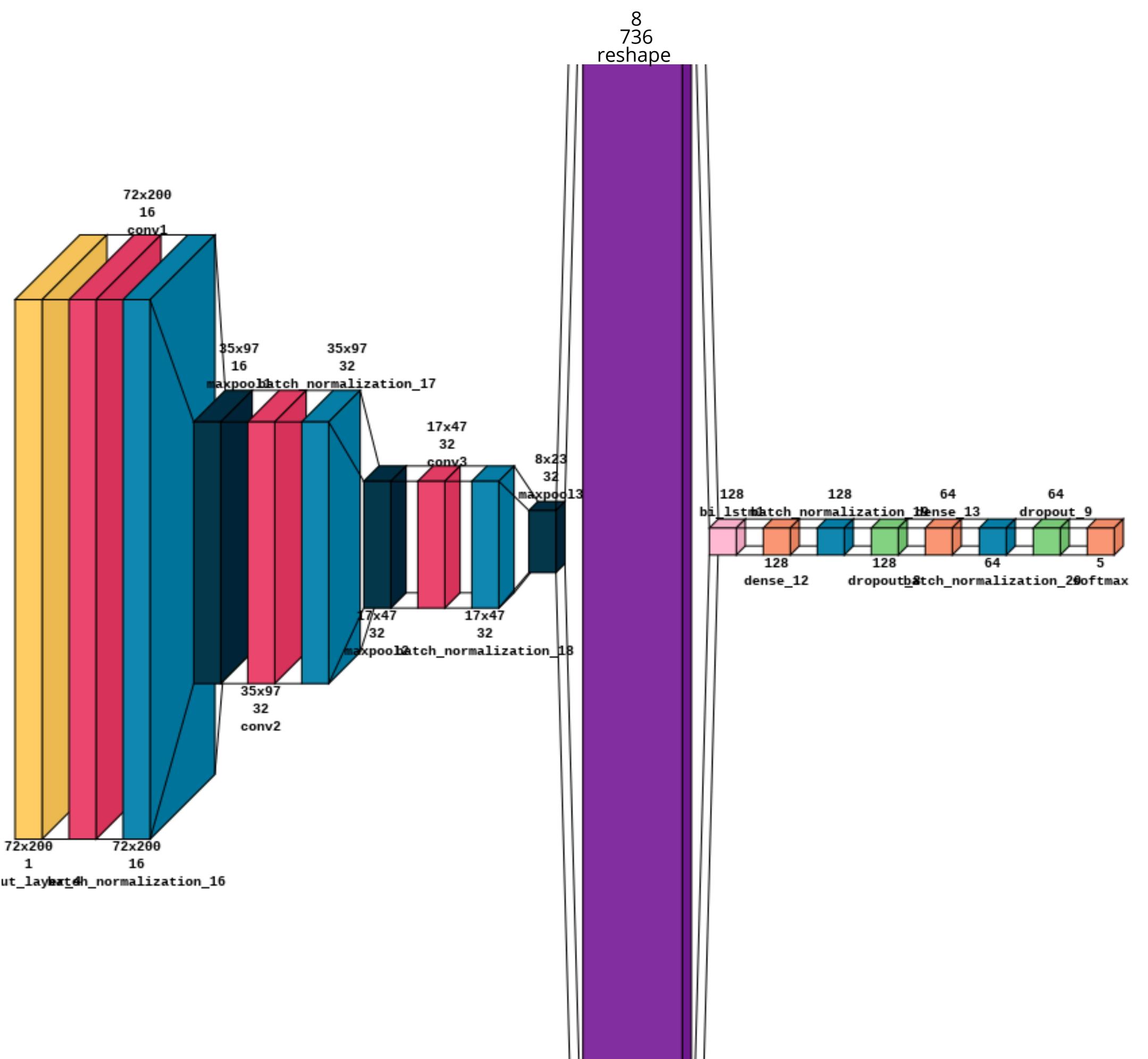
MODEL 3



Train Accuracy: 0.9238

Validation Accuracy: 0.8095

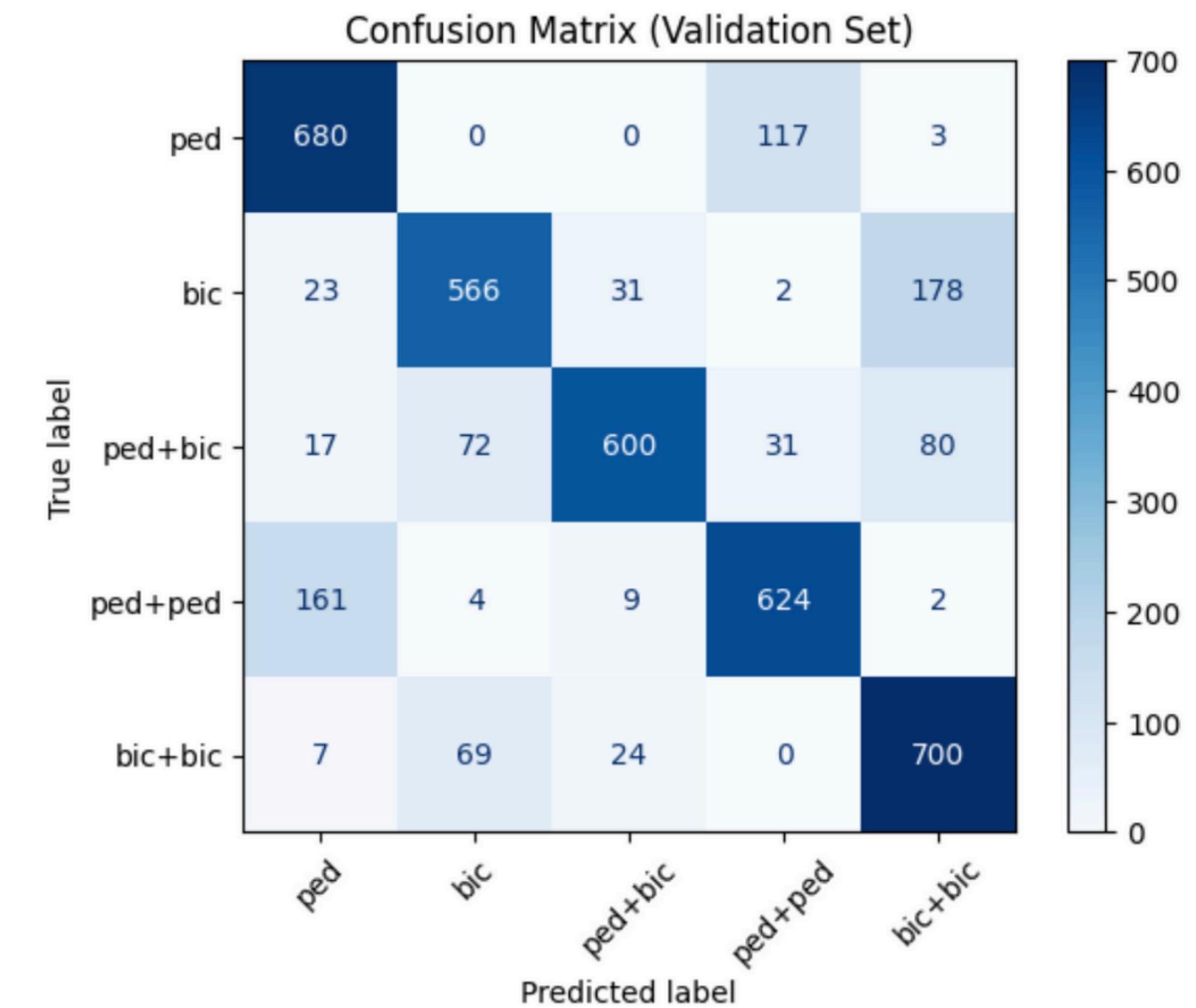
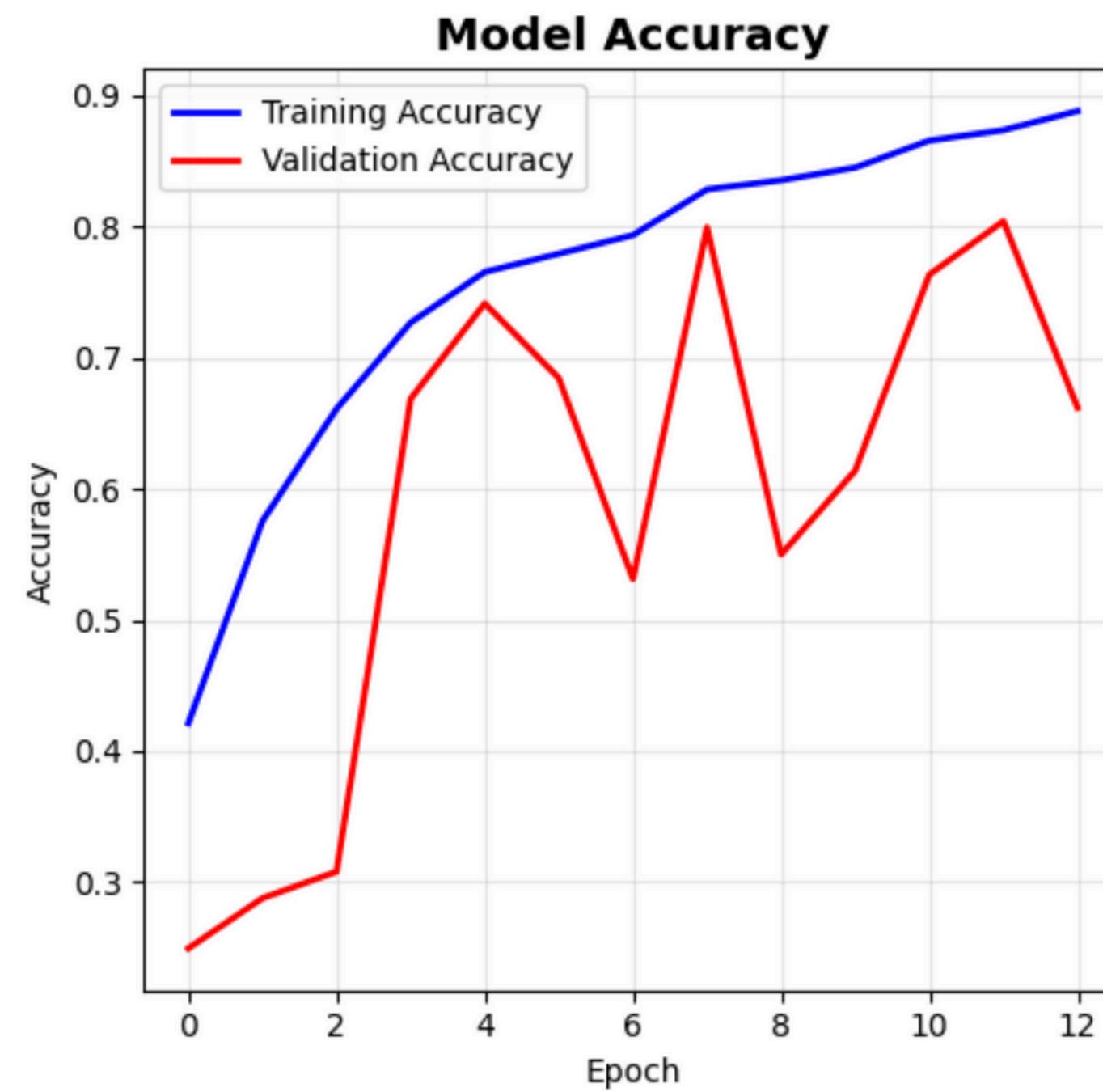
MODEL 4



Parameters: 476,373

We substitute LSTM with BI-LSTM to catch overlapping features as it analyses movements in both directions. We modify the dimension of the filters, increasing also the maxpooling2D kernel size. We also add another dense layer.

MODEL 4

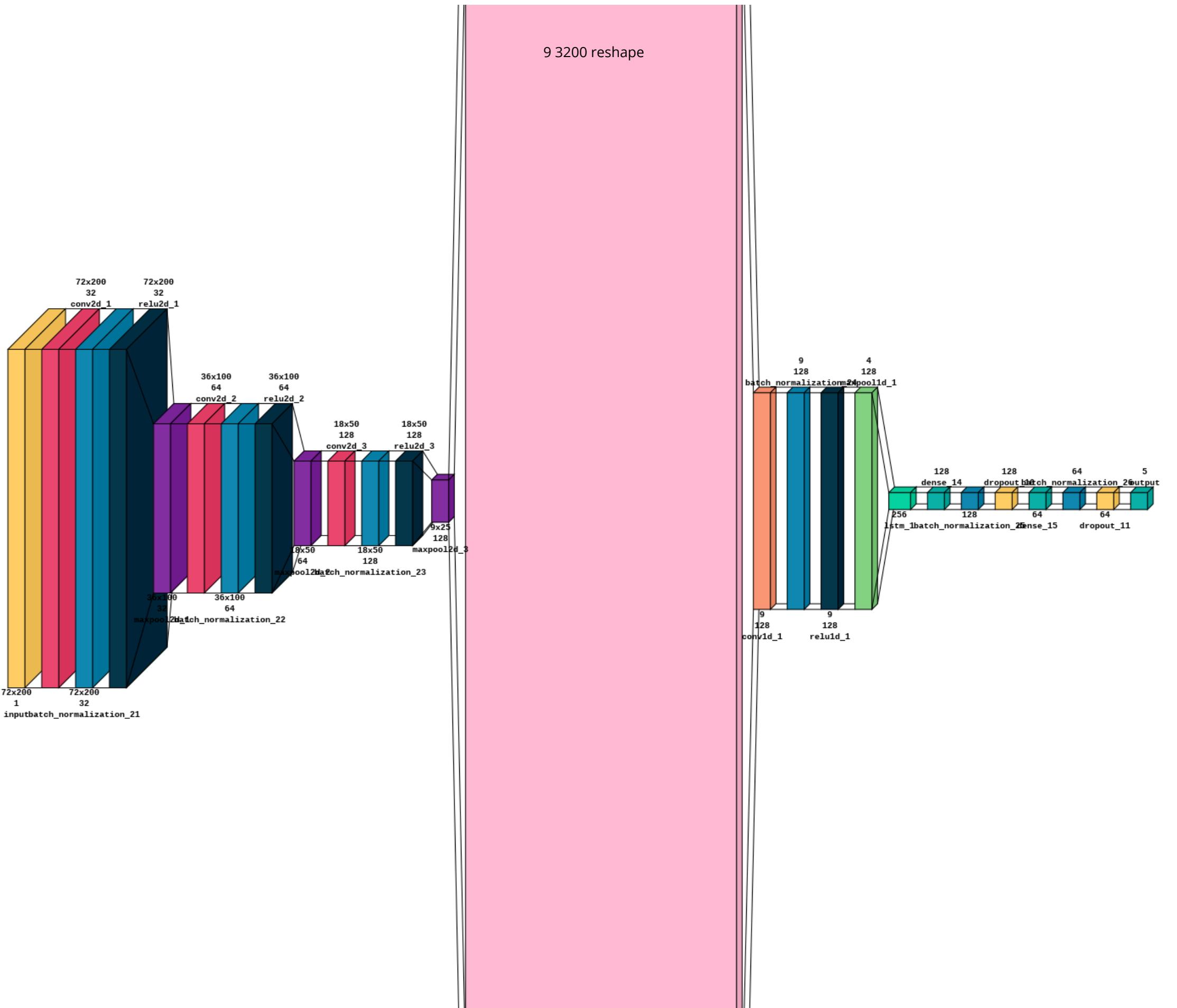


Train Accuracy: 0.8322

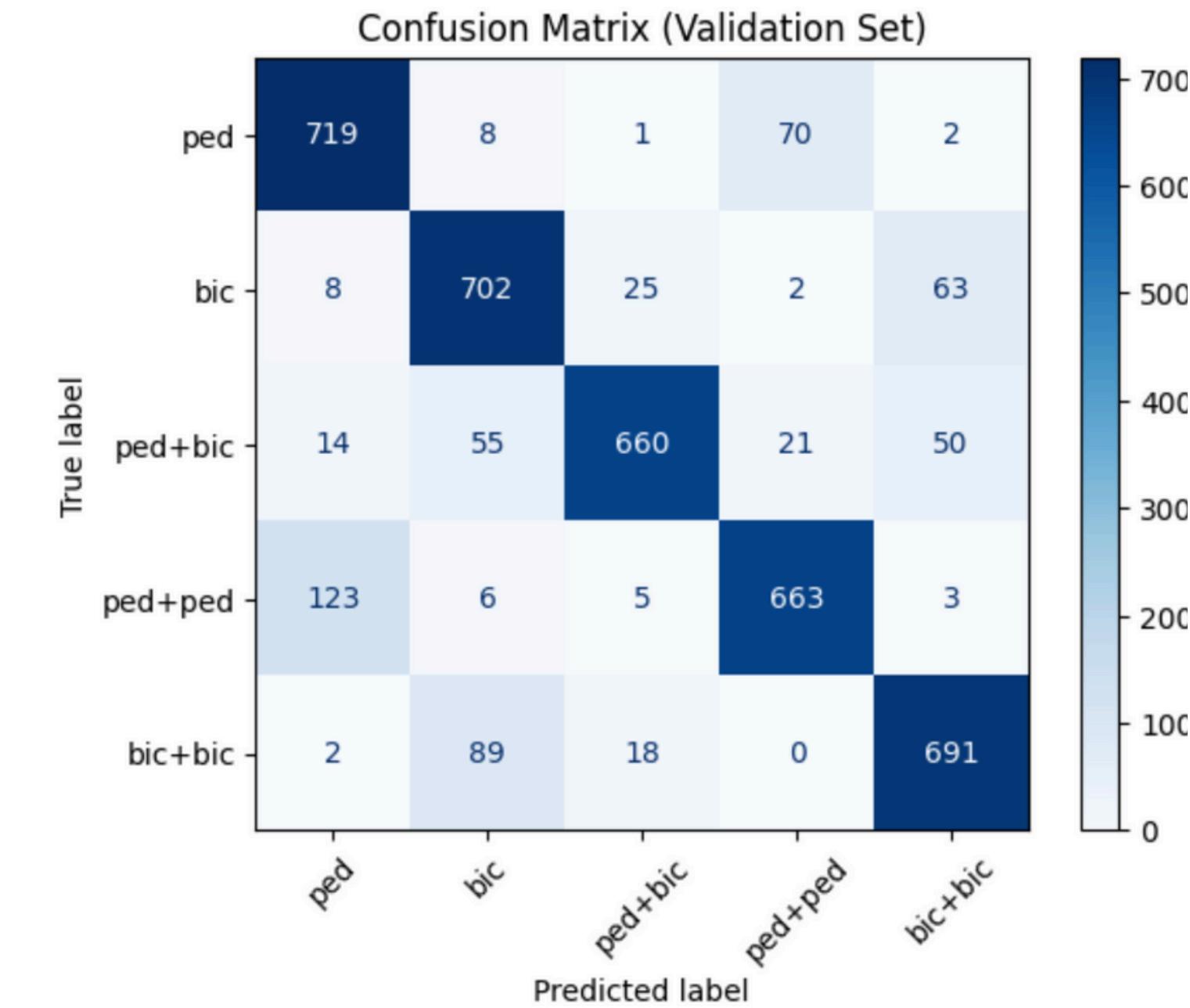
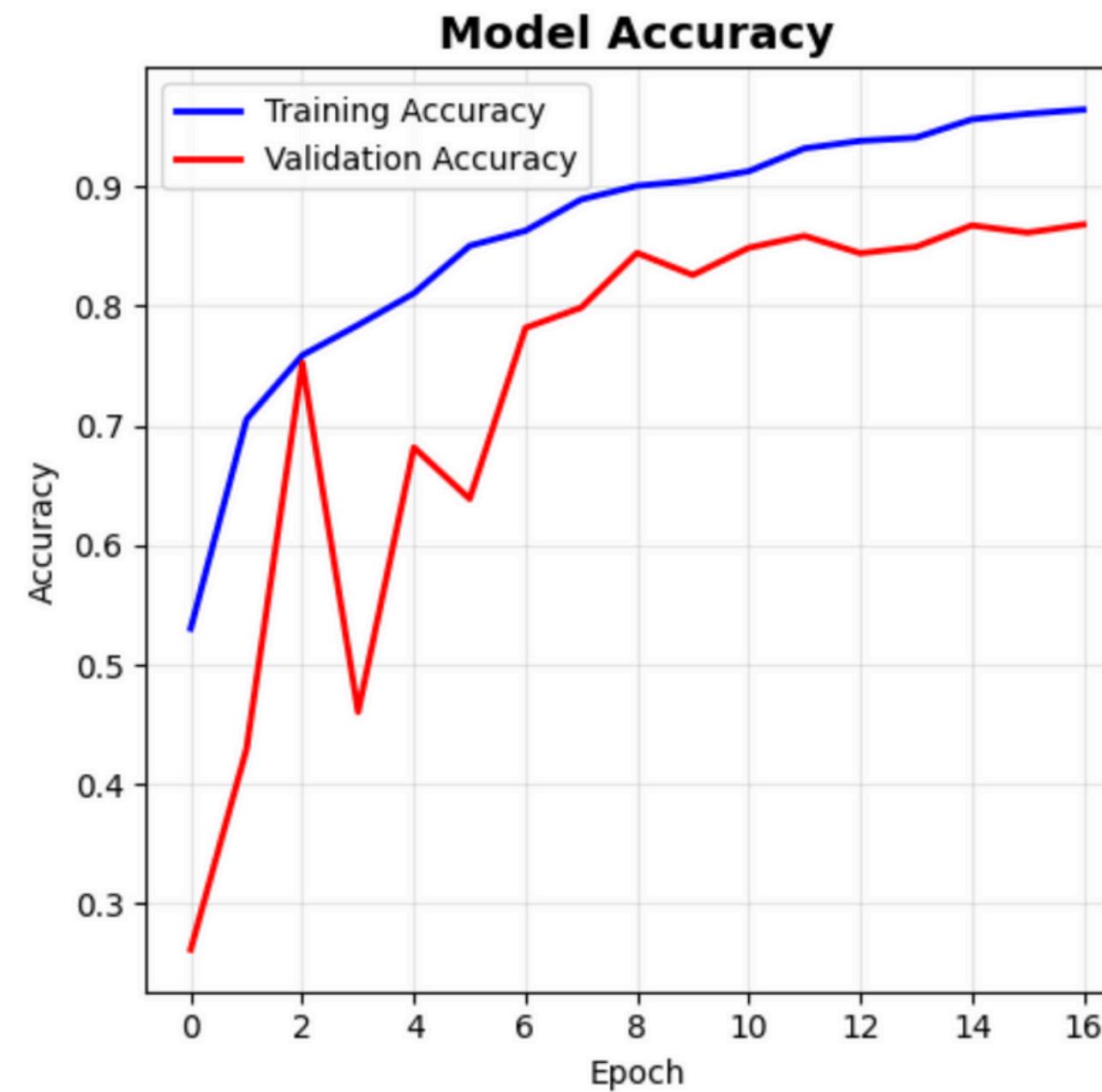
Validation Accuracy: 0.7995

MODEL 5

Parameters: 1,793,541
Since the 2D convolutional layer are treating spectrograms as static images, we add a 1D convolutional layer (128, kernel= 3) to detect how patterns evolve. Additionally, it functions as a filter for LSTM, which has now 256 units.

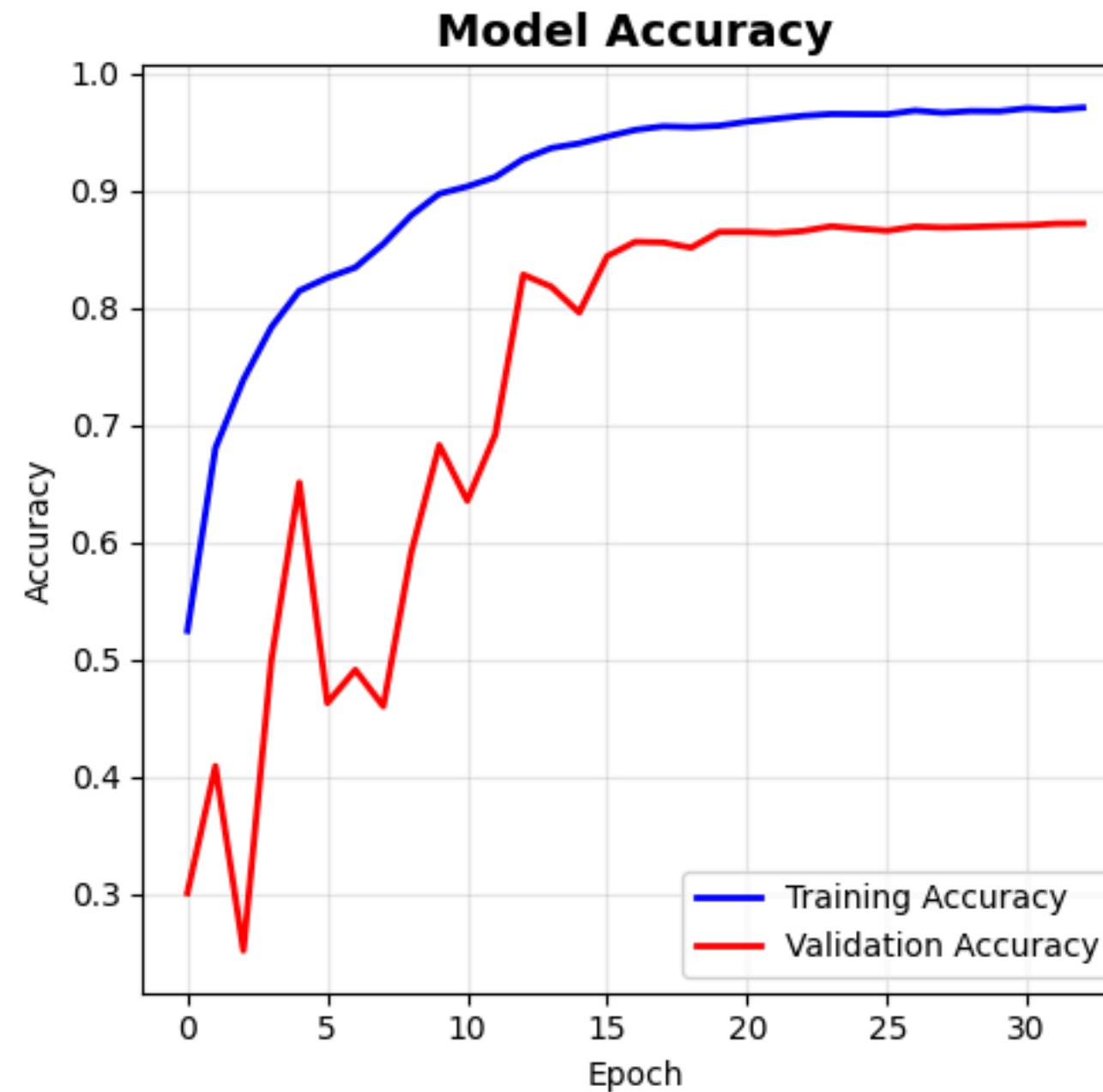


MODEL 5

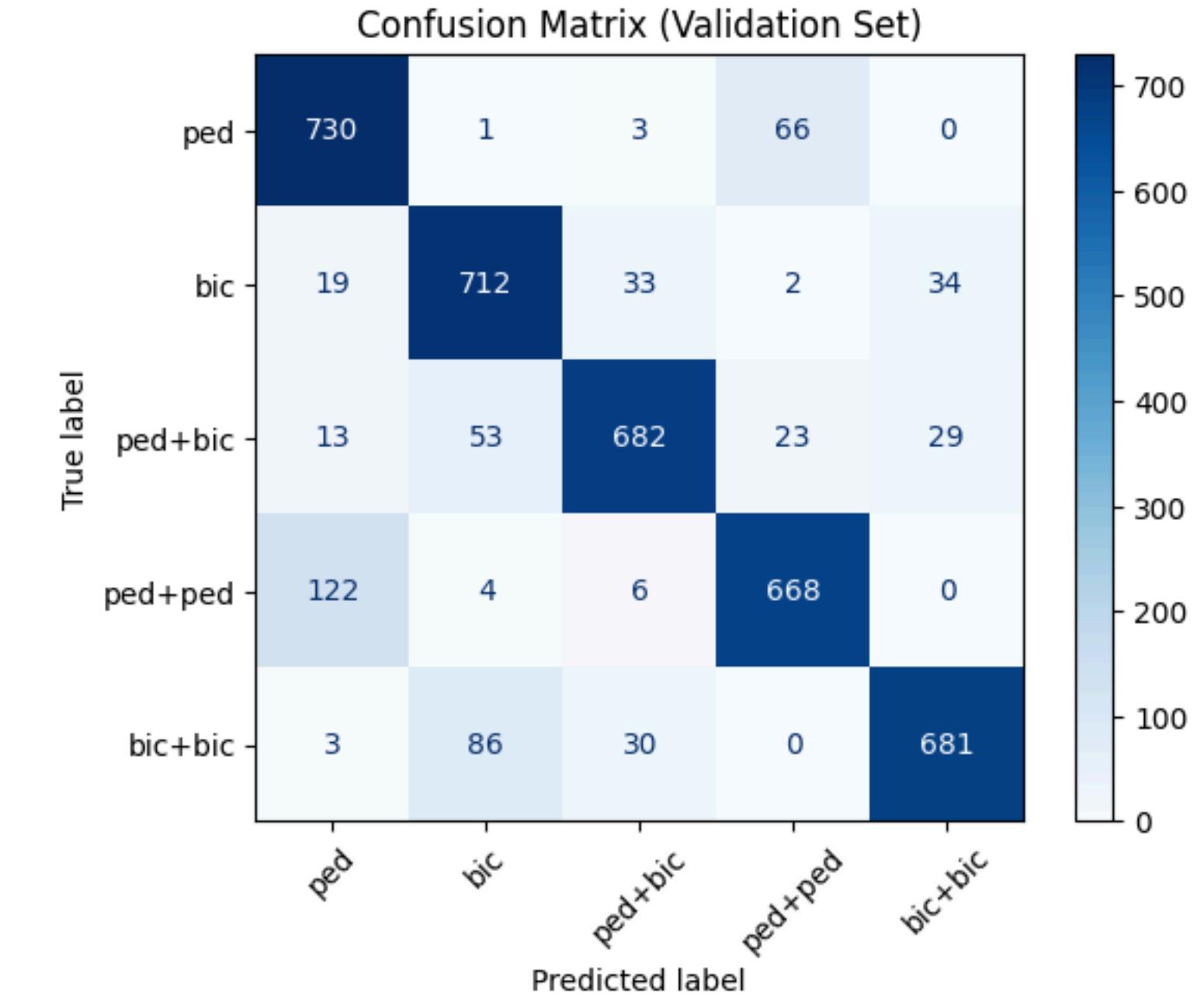


Train Accuracy: 0.9399
Validation Accuracy: 0.8587

MODEL 5- AUGMENTED



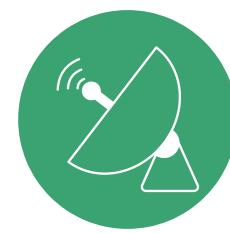
Train Accuracy: 0.9774
Validation Accuracy: 0.8683



Application of time-warping on PED+PED and BIC+BIC, deforming horizontal axis. We generated 500 sample per class.

	ACCURACY	LOSS	PARAMETERS
MODEL 1	0.6455	0.802	32,261
MODEL 2	0.7505	0.5712	454,917
MODEL 3	0.8095	0.4895	1,111,429
MODEL 4	0.7995	0.4968	476,373
MODEL 5	0.8587	0.4009	1,793,541
MODEL 5-A	0.8683	0.4112	1,793,541

TEST SET



We choose the fifth augmented model
as the best one



Test set: 5000 observations



Dimensional reduction and preprocessing



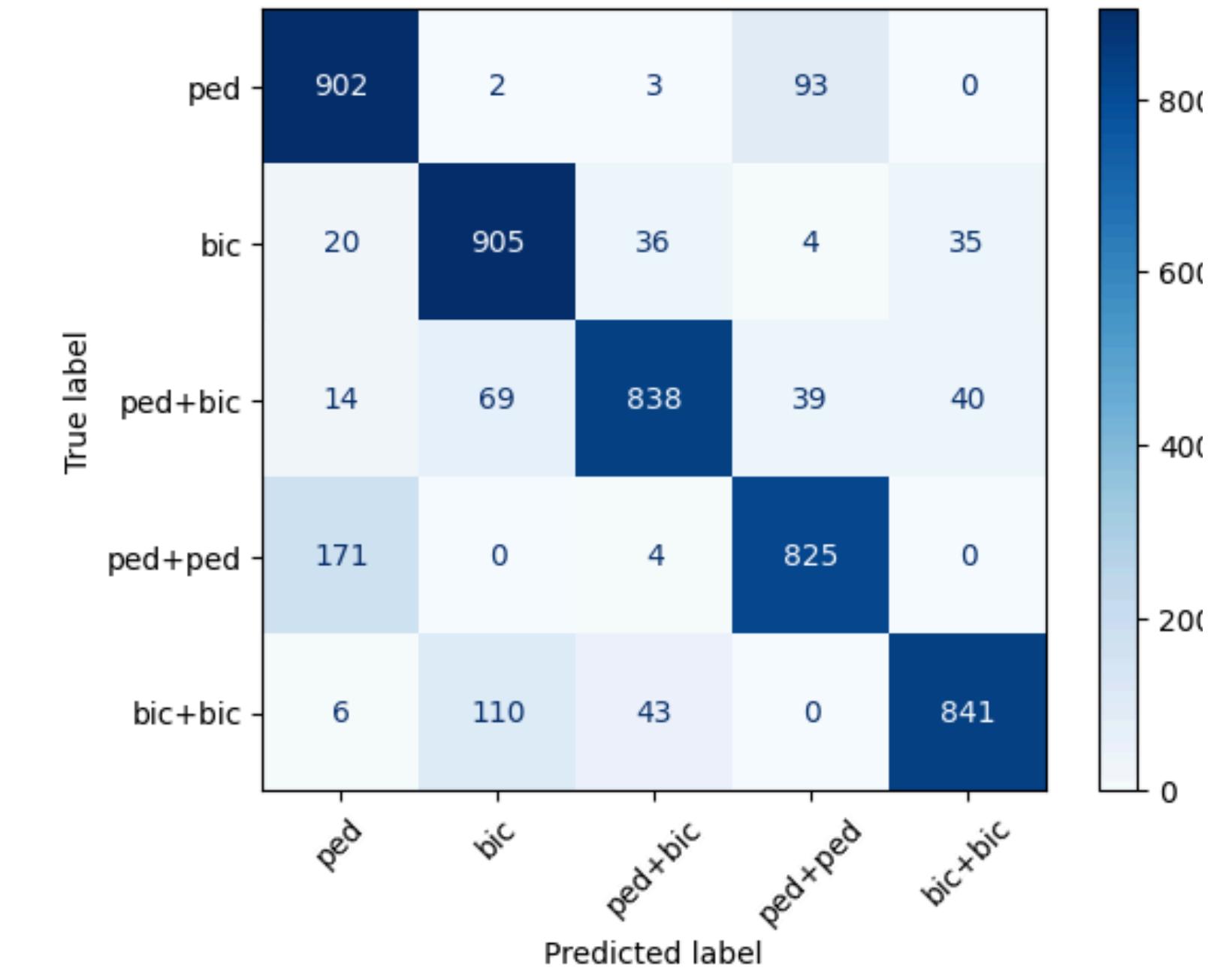
Evaluation of the best model



Final accuracy: 86.22%

METRICS AND CONFUSION MATRIX

	PRECISION	RECALL	F1
PED	0.810	0.902	0.854
BIC	0.833	0.905	0.868
PED+BIC	0.907	0.838	0.871
PED+PED	0.858	0.825	0.841
BIC+BIC	0.918	0.841	0.878



CONCLUSIONS

- The final model is a deep architecture with 3Conv2d, 1Conv1D, a LSTM layer trained on augmented data .
- The final result is acceptable (86.22%), but the problems about the two classes 'Ped+Ped' and 'Bic+Bic' still remains.
- Possible future improvements include parameter optimization and the design of architecture to better deal with the noise.