Part 1

2.

	F	M	NSL F	NSL M	SL F	SL M	Average Recording duration (s)	Total Duration(h:m:s)
0	1991	706	1166	459	825	247	9.07166	24466.3 | 06:47:46
1	967	372	673	178	294	194	8.66955	34990.3 | 09:43:10
2							9.03036	47662.2

The distribution is not very homogeneous, having a lot more female recordings than male recordings. We also have a lot more NSL than SL data. Both of these discrepancies will take an effect in our models.

3.

eGeMAPS has 88 features and IS11 has 4368 features.
eGeMAPS has a lot of functionals such as loudness and some low-level like shimmer.
IS11 has both as well.

Part 2

1.

Accuracy: (TP + TN) / N
Recall: TP / (TP + FN)
Precision: TP / (TP + FP)
F1-score: 2*(Precision * Recall) / (Precision + Recall)

2.

0 as the objective:

- Accuracy: 0.6355489171023152
- Recall: 1.0
- Precision: 0.6355489171023152
- F-measure: 0.7771689497716895

1 as the objective:

- Recall: 0
- Precision: 0
- F-measure: 0

3.

Unweighted average recall is best suited for our dataset since our data is quite unbalanced, the number of training samples of each class is quite different!

5.

The IS11 wasn't able to converge so we ended up using eGemaps, although we expected better results from IS11 since it's what they use in the paper from the competition. We combined this with oversampling from SMOTE to get better data. The results table is in the code.

	Hyperparameters	Train				Development			
		Acc	Precision	Recall	F1	Acc	Precision	Recall	F1
IS11	kernel='linear', C=1, class_weight='balanced'	1	1	1	1	0.512322628	0.502516383	0.502701739	0.497496294
eGeMAPS	kernel='rbf', C=100, class_weight='balanced', gamma=0.01	0.990461538	0.99049293	0.990461538	0.990461385	0.520537714	0.502461269	0.502609032	0.500059320

We had models with accuracy of 66/67 but we thought UAR was more important so we stuck with these that had better recall.

Part 3

2.1.

The model's performance increases (very slightly in our case, barely noticeable) with the additional layers but will start to overfit if we add more layers, which is expected as we don't have much data.

2.2.

The dropout will introduce some randomness in our data and prevent some overfitting. In our models the difference was quite slim.

Batch normalization did as expected, it improved the speed of our NN and some performance, although not quite noticeable.

3.

Our architecture is made of 3 layers, all sequential linear. We went with this option since after a lot of trial and error was the best results we were able to get. Our best feature set was IS11 based on Recall

	Train				Development			
	Acc	Precision	Recall	F1	Acc	Precision	Recall	F1
IS11	0.8089	0.841371697	0.8089230769	0.8042719104	0.587	0.5206951529	0.5160033519	0.5074969554
eGeMAPS	0.5191	0.5466165845	0.5190769231	0.4357400095	0.6049	0.4732114036	0.4912073067	0.4312060906

We also had models with better accuracy but we chose these with better recall.