Table 5: Performance comparison (expressed in %)
between A-MAMBA using (Raw Audio) and
A-MAMBA using (augmented audio) for binary and
multi-class classification task.

English						
Model	Dataset	Variant	Binary	Multiclass-Top-1	Multiclass Top-5	Multiclass Top-10
A-MAMBA + SSM	Augmented dataset	Small	53.87	42.54	69.2	76.56
		Medium	**55.79**	**44.86**	**74.77**	**78.55**
		Large	47.2	36.5	66.8	73.3
A-MAMBA + SSM	Original dataset	Small	**52**	**40.59**	**66.87**	**68.97**
		Medium	51.62	37.26	65.23	66.77
		Large	44.62	33.48	60.85	61.89
A-MAMBA + CSM	Augmented dataset	Small	54.11	44.34	70.23	76.67
		Medium	**58.23**	**45.17**	**77.04**	**79.43**
		Large	50.44	37.12	71.43	76.89
A-MAMBA + CSM	Original dataset	Small	**55.56**	**43.26**	**69.11**	**74.77**
		Medium	51.98	41.42	68.44	71.21
		Large	45.86	35.94	63.8	66.48
Bengali						
Model	Dataset	Variant	Binary	Multiclass-Top-1	Multiclass Top-5	Multiclass Top-10
A-MAMBA + SSM	Augmented dataset	Small	52.39	40.66	68.88	75.3
		Medium	**53.37**	**40.79**	**70.47**	**76.11**
		Large	45.65	37.12	66.9	71.08
A-MAMBA + SSM	Original dataset	Small	**51.21**	**38.42**	**68.75**	**74.32**
		Medium	50.34	36.43	65.87	68.76
		Large	44.89	33.37	62.75	66.76
A-MAMBA + CSM	Augmented dataset	Small	54.04	40.23	70.43	75.96
		Medium	**55.6**	**43.41**	**74.68**	77.22
		Large	46.39	36.32	70.9	**77.88**
A-MAMBA + CSM	Original dataset	Small	**52.78**	**39.44**	**70.11**	**75.76**
		Medium	50.79	38.2	65.78	70.87
		Large	45.78	34.61	63.11	66.88

Table 6: Perfromance comparison between baseline
Transformer and A-MAMBA CSM model

English							
# of layers	Model	Params	Variant	Binary	Multiclass Top-1	Multiclass Top-5	Multiclass Top-10
1	Transformer	12.7M	Small	46.43	38.06	55.55	61.24
2		19.4M	Medium	**51.82**	38.66	62.15	66.7
4		48.2M	Large	47.86	**38.79**	**65.77**	**68.87**
1	A-AMAMBA + CSM	11M	Small	**55.56**	**43.26**	**69.11**	**74.77**
2		18.5M	Medium	51.98	41.42	68.44	71.21
4		45.26M	Large	45.86	35.94	63.8	66.48