

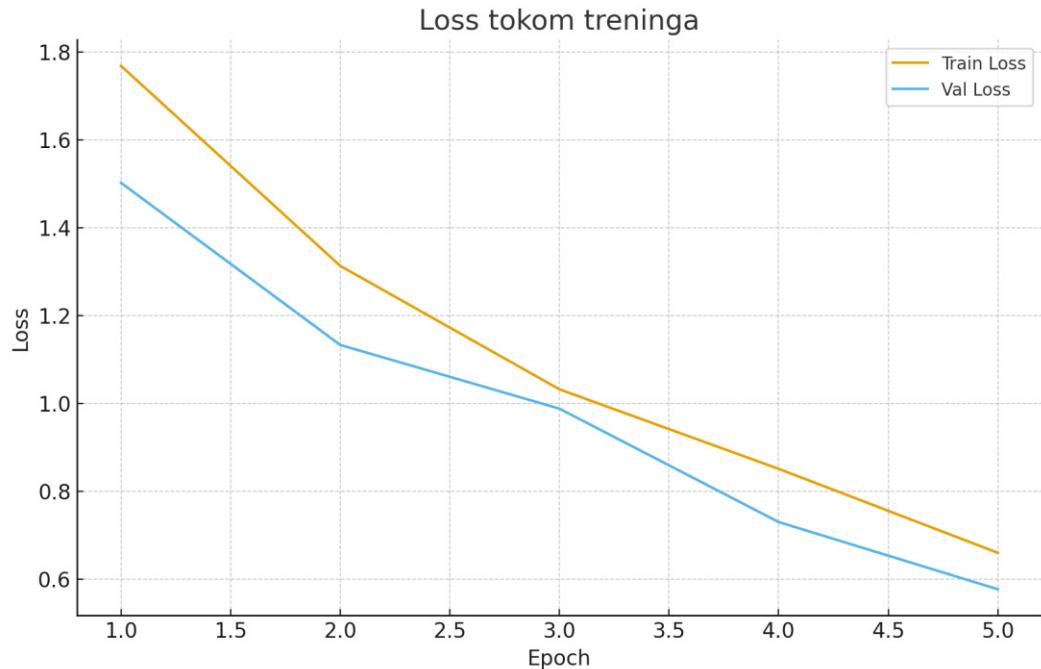
Instrument Classification – Report (Validation Split, ~12k)

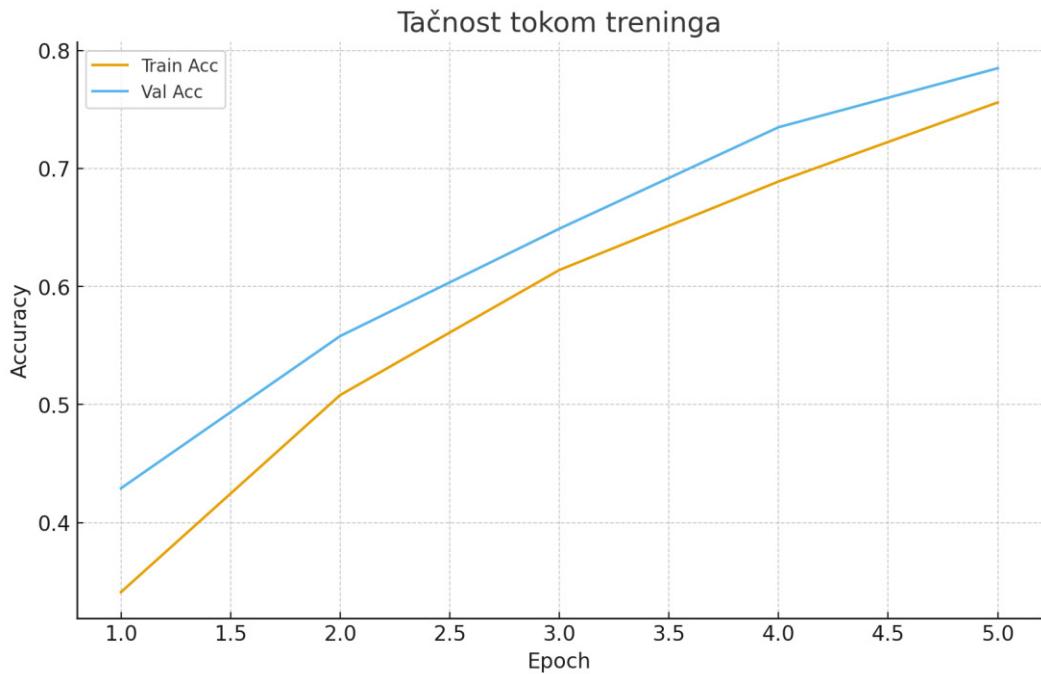
What We Did (Brief Team Summary)

We transitioned from the mini-subset to a larger validation split (~12k samples). We maintained the same pipeline (WAV -> log-mel -> small CNN), trained for 5 epochs again, and recorded the results. Our goal was to confirm that a larger dataset leads to more stable learning and better generalization.

Training Trends

Loss and accuracy per epoch:





Summary (Epochs):

Ep	Train Loss	Train Acc	Val Loss	Val Acc
1	1.768	0.341	1.502	0.429
2	1.313	0.508	1.133	0.558
3	1.032	0.614	0.988	0.649
4	0.851	0.689	0.730	0.735
5	0.660	0.756	0.577	0.785

Performance per Class (Validation)

class	precision	recall	f1	support
bass	0.72	0.60	0.66	160
brass	0.93	0.81	0.86	160
flute	0.95	0.87	0.91	94
guitar	0.80	0.46	0.58	160
keyboard	0.63	0.68	0.66	160

mallet	0.89	0.96	0.92	133
organ	0.93	0.96	0.94	160
reed	0.80	0.96	0.87	144
string	0.62	0.96	0.76	160
vocal	0.78	0.95	0.86	81

Globally: accuracy ≈ 0.78 , macro F1 ≈ 0.79 , weighted F1 ≈ 0.78 . The highest confusion was observed between related families (guitar/keyboard/string), while mallet and organ demonstrated very stable performance.

Conclusion and Next Steps

- The larger dataset raised the validation accuracy to approximately 0.79–0.80 (F1 ~0.79)
- Next step: Implement a more complex network model.