1 Hyperparameters

Table 1: Hyperparameter search space explored with Optuna. Ranges marked *log* are sampled log-uniformly. Derived quantities are not sampled: they are determined by other choices, which keeps the trunk output width equal to d_{model} and lets the feed-forward budget scale with model width. Head- and backbone-specific blocks are mutually exclusive; a single study searches the shared block plus exactly one head and one backbone.

Hyperparameter	Search space	Sampling
Shared		
Learning rate η	$[10^{-5},\ 3 \times 10^{-3}]$	continuous, log
Weight decay λ	$[10^{-6},\ 10^{-2}]$	continuous, log
Dropout p	$[0.0,\ 0.4]$	continuous
Model width d_{model}	{16, 32, 64, 128}	categorical
Distribution head		
Label width $\sigma_{\min}$	$[2.0,\ 25.0]$	continuous
Refinement order w	$\{1, \ldots, 7\}$	categorical
Regression head		
Regression loss	$\{\ell_1,\ \ell_2\}$	categorical
TCN backbone		
Depth L	$\{1, 2, 3, 4\}$	integer
Channels per layer	d_{model}	derived
Kernel size	{2, 3, 5, 7}	categorical
Transformer backbone		
Attention heads h	{2, 4, 8}	categorical
Encoder layers	$\{1, 2, 3, 4\}$	integer
FF multiplier r	{1, 2, 4}	categorical
FF dimension	$r\, d_{\text{model}}$	derived

Table 2: Patch length search space, in samples, as a function of the input sampling rate. Values are chosen so that a patch always spans the same set of wall-clock durations across rates; at 5 min resolution only the two longest durations are representable.

Input rate	Patch length (samples)	Span
10 Hz	{600, 1800, 3000, 6000}	1–10 min
1 Hz	{60, 180, 300, 600}	1–10 min
1 min	{1, 3, 5, 10}	1–10 min
5 min	{1, 2}	5–10 min