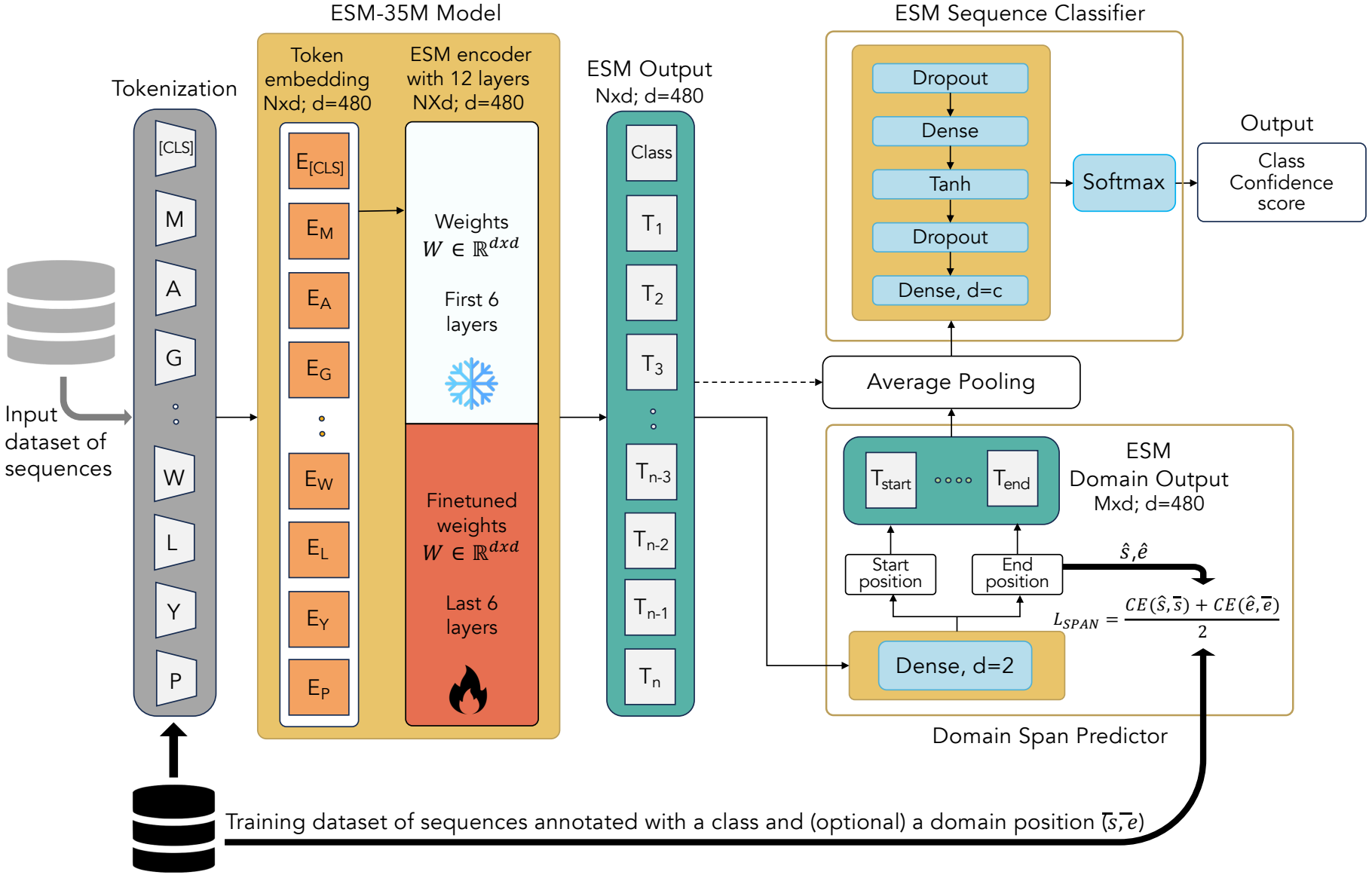
ESM-35M Model
ESM Sequence Classifier
Tokenization
Token embedding Nxd; d=480
ESM encoder with 12 layers NXd; d=480
ESM Output Nxd; d=480
Dropout
Dense
Tanh
Dropout
Dense, d=c
Softmax
Output
Class Confidence score
[CLS]
M
A
G
W
L
Y
P
E[CLS]
EM
EA
EG
EW
EL
EY
EP
Weights $W \in \mathbb{R}^{dxd}$
First 6 layers
Finetuned weights $W \in \mathbb{R}^{dxd}$
Last 6 layers
Class
T1
T2
T3
Tn-3
Tn-2
Tn-1
Tn
Input dataset of sequences
Average Pooling
Tstart
Tend
ESM Domain Output Mxd; d=480
Start position
End position
$\hat{s},\hat{e}$
$L_{SPAN} = \frac{CE(\hat{s},\overline{s}) + CE(\hat{e},\overline{e})}{2}$
Dense, d=2
Domain Span Predictor
Training dataset of sequences annotated with a class and (optional) a domain position $(\overline{s},\overline{e})$