Loss Convergence Curve - Transformer_Seq2Seq trained with Time Window Length 5 back and 1 forward