Validation Loss Comparison of Different Number of Attention Heads (d_model=384, dropout=0.1)