Many existing electrocardiogram biometric authentication methods primarily focus on intra-session recognition, which limits their practical applicability. To address temporal variations in cross-session scenarios, we propose a Siamese Autoencoder model that leverages paired inputs of phase-space portraits derived from two heartbeat waveforms. By jointly optimizing reconstruction and contrastive losses, the model learns encoded representations that capture both essential and discriminative features for individual identification. The model was initially trained on the CPSC single-session dataset and subsequently fine-tuned using transfer learning, with independent validation performed on the CYBHi database. Cross-session evaluation results demonstrate that the proposed Siamese Autoencoder outperforms a conventional Siamese Network, achieving a lower equal error rate of 7.79% on the CYBHi database. These findings highlight the effectiveness of incorporating an autoencoder within the Siamese framework to mitigate cross-session variability.