### **Abstract:**

**Abstract:**

The Transformer-based model presented here leverages the power of self-attention mechanisms to process sequence data. With an embedding layer that maps inputs to a 256-dimensional space, followed by a two-layer Transformer encoder with four attention heads, this model is designed for classification tasks. The output of the Transformer encoder is passed through a linear classifier to predict class labels. Achieving perfect accuracy on both training and validation sets, this model demonstrates its effectiveness on simpler tasks or well-structured data. However, its performance may degrade with more complex or imbalanced datasets, and it may face computational challenges on larger-scale tasks. The model’s tendency to overfit on smaller datasets highlights the importance of regularization when scaling to diverse real-world applications.