**INTRODUCTION**

A. MOTIVATION AND OBJECTIVE

THE Short Message Service (SMS) has been widely used as a communication tool over the past few decades as the popularity of mobile phone and mobile network grows. However, SMS users are also suffering from SMS spam. The SMS spam, also known as drunk message, refers to any irrelevant messages delivered using mobile networks [1]. There are several reasons that lead to the popularity of spam messages. Firstly, there is a large number of users who use mobile phones in the world, making the potential victims of the spam messages attack also high. Secondly, the cost of sending out spam messages is low, which could be good news to the spam attacker. Last but not least, the capability of the spam classifer on most mobile phones is relatively weak due to the shortage of computational resources, which limits them from identifying the spam message correctly and efficiently.

Machine learning is one of the most popular topics in the last few decades, and there are a great number of machine learning based classification applications in multiple research areas. Specifically, spam detection is a relatively mature research topic with several established methods. However, most of the machine learning based classifers were dependent on the handcrafted features extracted from the training data [2].

As a class of machine learning techniques, deep learning has been developing rapidly recently thanks to the surprising

growth of coputational resources in the last few decades. Nowadays, deep learning based applications play a significant part in our society, making our lives much easier in many aspects. As one of the most effective and widely used deep learning architectures, Recurrent Neural Network (RNN), as well as its variants such as Long Short-Term Memory (LSTM), were applied to spam detection and proved to be extremely effective during the last few years.

The Transformer [3] is an attention-based sequence-to sequence model that was originally designated for translation task, and it achieved great success in English-German and English French translation. Moreover, there are multiple improved Transformer-based models such as GPT-3 [4] and BERT [5] proposed recently to address different Natural Language Process (NLP) problems. The accomplishments of the Transformer and I s successors have proved how powerful and promising they are. In this paper, we aim to explore whether it is possible to adapt the Transformer model to the SMS spam detection problem. Therefore, we propose a modi\_ed model based on the vanilla Transformer to identify SMS spam messages. Additionally, we analyze and compare the performance of SMS spam detection between traditional machine learning classifers, an LSTM deep learning solution, and our proposed spam Transformer model.