

Informatics Institute of Technology

In Collaboration With

University of Westminster, UK

TAM 

**Shorthand Romanized Tamil to Tamil Transliteration
using Novel Hybrid Approach**

A dissertation by

Miss. Anuja Dilrukshi Herath

w1790023 / 20191180

Supervised by

Mr. Deshan Sumanathilake

May 2022

Submitted in partial fulfilment of the requirements for the BSc (Hons) Computer Science
degree at the University of Westminster.

ABSTRACT

Transliteration from Tamil to the Roman script is an important aspect of communication, education, and digital technology. However, the Tamil script has 12 vowels, while the Roman script has only 5 vowels. This poses a challenge for Romanized Tamil transliteration without vowels. This paper presents a literature review of recent research in Romanized Tamil to Tamil transliteration without vowels. The review includes an analysis of various methods proposed in recent years, such as rule-based, context-based, and machine learning-based approaches. The results of these studies suggest that machine learning-based methods, such as convolutional neural networks, can achieve high accuracy in predicting the missing vowels in Tamil words. This paper highlights the importance of this area of research for improving communication, education, and digital technology for Tamil speakers worldwide.

Keywords: NLP, n-gram, Computer Linguistic, Hybrid Recommendation Systems, Machine Learning, Transliteration System, Rule Based recommendation