MULTI-DOMAIN NEURAL MACHINE TRANSLATION WITH KNOWLEDGE DISTILLATION FOR LOW RESOURCE LANGUAGES

Velayuthan Menan

238043D

Master of Science (Major Compoment Research)

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

March 2025

MULTI-DOMAIN NEURAL MACHINE TRANSLATION WITH KNOWLEDGE DISTILLATION FOR LOW RESOURCE LANGUAGES

Velayuthan Menan

238043D

Dissertation submitted in partial fulfillment of the requirements for the degree
Master of Science (Major Compoment Research)

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

March 2025

DECLARATION

I declare that this is my own work and this Dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 27/05/2025

The supervisors should certify the Dissertation with the following declaration.

The above candidate has carried out research for the Master of Science (Major Compoment Research) Dissertation under our supervision. We confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr. Nisansa de Silva

Signature of the Supervisor: Date: 27/05/2025

Name of Supervisor: Dr. Surangika Ranathunga

Signature of the Supervisor: Date:

DEDICATION

To my dearest parents, Velayuthan (Appa) and Kanageshwary (Amma),
and to my fiancée, Sambavi,

Thank you for your unconditional love, unwavering support, and endless belief in me.
This journey would not have been possible without you.
This work is dedicated to you.

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude to my supervisors, Dr. Nisansa de Silva and Dr. Surangika Ranathunga, for their continuous support and guidance throughout the course of this research. Without your insights and tremendous mentorship, I would not have been able to achieve this level of success. Thank you for always stepping in to help whenever I needed, despite your busy schedules.

I also wish to extend my heartfelt appreciation to the Research Coordinator of the Department of Computer Science & Engineering, Dr. Kutila Gunasekera, for his unwavering support in ensuring this was a smooth and enriching experience. My sincere thanks also go to the academic and non-academic staff of the Department of Computer Science & Engineering for their assistance and for providing the resources necessary to carry out my research.

This work was funded by the Google Award for Inclusion Research (AIR) 2022, received by Dr. Surangika Ranathunga and Dr. Nisansa de Silva. I would like to thank the National Languages Processing (NLP) Centre at the University of Moratuwa for providing the GPUs used to execute the experiments related to this research.

ABSTRACT

Multi-domain adaptation in Neural Machine Translation (NMT) is crucial for ensuring high-quality translations across diverse domains. Traditional fine-tuning approaches, while effective, become impractical as the number of domains increases, leading to high computational costs, space complexity, and catastrophic forgetting. Knowledge Distillation (KD) offers a scalable alternative by training a compact model using distilled data from a larger teacher model. However, we hypothesize that sequence-level KD primarily distills the decoder while neglecting encoder knowledge transfer, resulting in suboptimal adaptation and generalization, particularly in low-resource language settings where both data and computational resources are constrained.

To address this, we propose an improved sequence-level distillation framework enhanced with encoder alignment using cosine similarity-based loss. Our approach ensures that the student model captures both encoder and decoder knowledge, mitigating the limitations of conventional KD. We evaluate our method on multi-domain German–English translation under simulated low-resource conditions and further extend the evaluation to a bona fide low-resource language, demonstrating the method's robustness across diverse data conditions.

Results demonstrate that our proposed encoder-aligned student model can even outperform its larger teacher models, achieving strong generalization across domains. Additionally, our method enables efficient domain adaptation when fine-tuned on new domains, surpassing existing KD-based approaches. These findings establish encoder alignment as a crucial component for effective knowledge transfer in multi-domain NMT, with significant implications for scalable and resource-efficient domain adaptation.

Keywords: Neural Machine Translation, Knowledge Distillation, Multi-Domain Adaptation, Low-Resource Languages, Encoder Alignment, Sequence-Level Distillation

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
EWC	Elastic Weight Consolidation
GRU	Gated Recurrent Unit
KD	Knowledge Distillation
LRL	Low Resource Languages
LSTM	Long Short-Term Memory
M	Million
MT	Machine Translation
NLP	Natural Language Processing
NMT	Neural Machine Translation
RNMT	Recurrent Networks for Neural Machine Translation
RNN	Recurrent Neural Networks
SMT	Statistical Machine Translation