CERTIFICATE

It is certified that the work contained in the thesis titled ***"Low Resource Similar Language Machine Translation in Common Phonetic Space with Multilingual, Adversarial and Reinforced Deep Learning"*** by ***Amit Kumar*** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all requirements of Comprehensive Examination, Candidacy, and SOTA for the award of Ph.D. Degree.

Supervisor
Dr. Anil Kumar Singh
Associate Professor,
Dept. of Computer Science and Engineering,
Indian Institute of Technology (BHU),
Varanasi, India 221005.

पर्यवेक्षक/Supervisor
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(काशी हिन्दू विश्वविद्यालय)
Banaras Hindu University)
Varanasi-221005

Co-Supervisor
Dr. Ajay Pratap
Assistant Professor,
Dept. of Computer Science and Engineering
Indian Institute of Technology (BHU),
Varanasi, India 221005.

सहायक आचार्य/Assistant Professor
संगणक विज्ञान एवं अभियांत्रिकी विभाग/Department of Computer Sc. &
भारतीय प्रौद्योगिकी संस्थान/Indian Institute of Technolo
काशम हिन्दू यूनिवर्सिटी)/(Banaras Hindu University,
वाराणसी /Varanasi-221005

DECLARATION BY THE CANDIDATE

I, ***Amit Kumar***, certify that the work embodied in this Ph.D. thesis is my own bonafide work carried out by me under the supervision of ***Dr. Anil Kumar Singh*** and co-supervision of ***Dr. Ajay Pratap*** from ***July 2017*** to ***December 2022*** at ***Department of Computer Science and Engineering***, Indian Institute of Technology (BHU) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, *etc.* reported in journals, books, magazines, reports, dissertations, theses, *etc.*, or available at websites and have not included them in this thesis and have not cited as my own work.

Date: 06/12/2022

Place: Varanasi

Amit Kumar

(Amit Kumar)

CERTIFICATE BY THE SUPERVISOR

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Supervisor)
Dr. Anil Kumar Singh
Associate Professor,
Dept. of Computer Science and Engineering,
Indian Institute of Technology (BHU),
Varanasi, India

पर्यवेक्षक/Supervisor
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(काशी हिन्दू विश्वविद्यालय)
(Banaras Hindu University)
वाराणसी/Varanasi-221005

(Co-Supervisor)
Dr. Ajay Pratap
Assistant Professor,
Dept. of Computer Science and Engineering
Indian Institute of Technology (BHU),
Varanasi, India

सहायक आचार्य/Assistant Professor
संगणक विज्ञान एवं अभियांत्रिकी विभाग/Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान/Indian Institute of Technology
(बनारस हिन्दू यूनिवर्सिटी)/(Banaras Hindu University)
वाराणसी/Varanasi-221005

Signature of Head of Department

Professor & Head
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(बनारस हिन्दू यूनिवर्सिटी)
(Banaras Hindu University)

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Low Resource Similar Language Machine Translation in Common Phonetic Space with Multilingual, Adversarial and Reinforced Deep Learning

Name of the Student: Amit Kumar

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University) Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the *Doctor of Philosophy.*

Date: 06/12/2022

Place: Varanasi

Amit Kumar

(Amit Kumar)

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

Dedicated to my parents,

Shri. Panna Lal

and

Smt. Bhagirathi Devi

ACKNOWLEDGEMENT

First and foremost, I would like to thank my supervisor, Dr. Anil Kumar Singh, and my co-supervisor, Dr. Ajay Pratap for their invaluable support and assistance. I feel immense pleasure in expressing my profound sense of gratitude and sincere regard for their constant feedback and expertise during all these years. I am eternally grateful to have the opportunity to work on my thesis under their supervision.

My cordial thanks to all the members of the Department of Computer Science and Engineering for creating an excellent working atmosphere. I would also like to thank the other members of my Doctoral committee, Dr. Sukhda, Department of Humanities Studies, and Dr. Sukomal Pal, Department of Computer Science and Engineering, for their help and support throughout the tenure of my studies. Special thanks to Dr. Sriparna Saha, Department of Computer Science and Engineering, IIT Patna, for her valuable suggestions. I would also like to convey my sincere gratitude to Prof. Sanjay Kumar Singh, Head of the CSE Department, and all the RPEC and DPGC members for their suggestions and endorsement of this work.

I am grateful to my colleagues and friends for the long discussions and their brilliant insights that have helped me to overcome the challenges I have faced in the development of this work. Finally, I express my heartfelt gratitude to my parents Smt. Bhagirathi Devi and Shri. Panna Lal for their constant support, love, encouragement, and sacrifices. Their affectionate love and care cannot be expressed in words.

With limitless humility, I would like to praise and thank the **" Satguru Kabir Sahab Ji"**. The almighty, the Merciful compassionate who bestowed me with all the favourable circumstances to achieve the desired goal of life through this crucial juncture.

(Amit Kumar)

List of Figures

List of Tables

List of Symbols

Symbol	Description
X	Source sequence
Y	Target sequence
x_i	Source word embedding at i^{th} position in X
y_j	Word at position j in target sequence
q_t	Hidden state at time t
α_{ti}	Attention weight between source word x_i and target word y_t
$output\_length$	Length of predicted sentences
$reference\_length$	Length of reference sentences
$attn$	Attention in Transformer
Q	Query
K	Key
V	Value
d_k	Dimension of key
$\mathcal{L}_{CE}$	Cross-entropy loss
$\mathcal{L}_{LM}$	Cross-entropy loss of language model
PP	Perplexity
w_i	i^{th} word embedding
W_i	initial weight of each sentence pair S_i (X_i,Y_i)
c	State
a	Action
θ	Models parameter in policy network
$\pi_\theta(c, a)$	Policy
$\mathcal{N}$	Gaussian distribution
μ	Mean
σ	Standard deviation
R	Reward

Symbol	**Description**
sw_i	Subword embeddings at i^{th} position
so_i	Sub-phonetics embeddings at i^{th} position

Abbreviations

Abbreviation	Description
NLP	Natural Language Processing
MT	Machine Translation
NMT	Neural Machine Translation
SMT	Statistical Machine Translation
CNN	Convolutional Neural Network
LSTM	Long-Short Term Memory
GAN	Generative Adversarial Network
NN	Neural Network
RNN	Recurrent Neural Network
GRU	Gated Recurrent Unit
ReLU	Rectified Linear Unit
IPA	International Phonetic Alphabet
HRLs	High Resource Languages
LRLs	Low Resource Languages
ZRP	Zero-Resource Problem
ZST	Zero-Shot Translation
TER	Translation Edit Rate
TL	Transfer Learning
DA	Domain Adaptation