Artificial Intelligence-based Novel Techniques for Accelerating Drug Discovery

Thesis submitted in partial fulfillment
for the award of the degree of

Doctor of Philosophy

by

Vishakha Singh

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
(BANARAS HINDU UNIVERSITY)
VARANASI - 221005

Roll No. 20071505 *Year 2024*

CERTIFICATE

It is certified that the work contained in the thesis titled *"Artificial intelligence-based novel techniques for accelerating drug discovery"* by *Vishakha Singh* has been carried out under my supervision, and 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.

Prof. Sanjay Kumar Singh
Professor
Dept. of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi

DECLARATION BY THE CANDIDATE

I, ***Vishakha Singh***, certify that the work embodied in this Ph.D. thesis is my own bonafide work carried out by me under the supervision of ***Prof. Sanjay Kumar Singh*** from ***December 2020*** to ***January 2024*** 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 wilfully 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: 25.07.24

Place: Varanasi

Vishakha Singh

Vishakha Singh

CERTIFICATE BY THE SUPERVISOR

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

Prof. Sanjay Kumar Singh
Professor
Dept. of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi

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

Signature of Head of the Department

आचार्य एवं विभागाध्यक्ष
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)
वाराणसी-221005/Varanasi-221005

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Artificial intelligence-based novel techniques for accelerating drug discovery

Name of the Student: Vishakha Singh

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: 25.07.24

Place: Varanasi

Vishakha Singh

Vishakha Singh

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 beloved parents.

ACKNOWLEDGEMENT

The completion of my doctorate as a Prime Minister Research Fellow has provided me with a wonderful opportunity to express my heartfelt gratitude towards people who were an integral part of this journey. I am grateful to everyone who helped make this dissertation possible by constantly motivating and assisting me to the best of their efforts and resources. First and foremost, I would like to thank my esteemed supervisor, Prof. Sanjay Kumar Singh, for his extraordinary support and supervision during my Ph.D. He guided and motivated me using his knowledge and expertise, from the beginning until the end of my degree, by nudging me in the right direction. I am also grateful to my doctoral committee (Prof. Neeraj Sharma and Dr. Ravi Shankar Singh) for reviewing my research progress and providing invaluable insights as to how I can improve upon it. This acknowledgment cannot be complete without my parents and siblings, to whom I owe this thesis. I want to express my heartfelt gratitude to my mother, Dr. Punam Singh, and father, Mr. Sriniwas Singh, for their unwavering support and understanding and for being my pillars of strength throughout this journey. I am also thankful to my sister, Dr. Aakansha Singh, and brother, Mr. Shashank Shekhar Singh, for motivating and encouraging me. As I complete my doctorate, my heart goes out to my late grandmother, Mrs. Sushila Singh, who was a crucial part of my journey. This note of thanks would be incomplete without mentioning my friends and colleagues (Ms. Deepti Chauhan, Dr. Ritesh Sharma, Dr. Ekta Sharma, Mr. Jayashankara M. and Mr. Rajat Gupta), to whom I am grateful for their encouragement and constructive criticism, which enriched the quality of this thesis. Lastly, I am obliged to all the department's technical and non-technical staff members, Mr. Ravi Kumar Bharti, Mr. Shubham Pandey, and Mr. Prakhar Kumar.

This thesis marks a significant milestone in my career. With heartfelt gratitude, I look forward to apply the lessons learned from this experience to my future endeavors.

Vishakha Singh

(Vishakha Singh)

Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Description
ABPs	Antibacterial Peptides
AI	Artificial Intelligence
ANOVA	Analysis of Variance
AVPs	Antiviral Peptides
AUC	Area under the receiver operating characteristic curve
B3P2s	Blood-Brain Barrier Penetrating Peptides
BERT	Bidirectional Representations from Transformers
BiLSTM	Bidirectional Long Short-Term Memory
ESKAPEE	Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, Enterobacter spp. and Escherichia coli
GSA	Gravitational Search Algorithm
NP	Neurological Peptides
ReLU	Rectified Linear Unit
RFs	Random Forests
SHAP	Shapley Additive explanations
SVM	Support Vector Machine
TCNs	Temporal Convolutional Networks

List of Symbols

Symbol	Description
a_t	attention vector
$\alpha_{t,t'}$	attention weight
$\tilde{c}_t$	candidate state
c_t	cell state
cv_t	context vector
fit	fitness value
f_t	forget gate
h_t	hidden state
i_t	input gate
o_t	output gate
Pop	population of peptides
r_t	reset gate
z_t	update gate