

CERTIFICATE

This is to certify that the thesis entitled “**A Hybrid Machine Learning Approach for Detecting Intentional and Unintentional Insider Threats with Mitigation Through Behavioral Biometrics and User Profiling Mechanism**” submitted to the Avinashilingam Institute for Home Science and Higher Education for Woman, Coimbatore, for the award of the degree of **Doctor of Philosophy in Computer Science**, is a record of original research work done by **S. Asha (20PHCSF005)**, during the period of her study in the Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, under my supervision and guidance and the thesis has not formed the basis for the award of any Degree / Diploma / Associateship / Fellowship or Seminar title to any candidate of any University.



Signature of the

Head of the Department

B. Kalpana, M.Sc., M.Phil., Ph.D.
Professor and Head
Department of Computer Science
Avinashilingam Institute for Home Science
and Higher Education for Women
(University), Coimbatore - 641 043



Signature of the Supervisor



Signature of the Dean

Dr. (Mrs) V. RADHA
Dean, School of Physical Sciences and
Computational Sciences
Professor, Department of Computer Science
Coordinator, B.Voc (AI & ML)
Avinashilingam Institute for Home Science
and Higher Education for Women
Coimbatore - 641 043

DECLARATION

I hereby declare that the thesis titled “**A Hybrid Machine Learning Approach for Detecting Intentional and Unintentional Insider Threats with Mitigation Through Behavioral Biometrics and User Profiling Mechanism**” is the result of investigations carried out by me in the Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, under the supervision and guidance of **Dr. D.Shanmugapriya**, Assistant Professor, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Woman, Coimbatore, and that it has not been submitted for the award of any Degree / Diploma / Associateship / Fellowship or similar title of any University or Institute.



Signature of the Supervisor



Signature of the Research Scholar

ACKNOWLEDGEMENT

At the outset, I would like to express my sincere gratitude to the God Almighty for his/her constant love, blessing and grace showered on me making my meaningful and worthwhile to the society.

I express my reverential gratitude to **Late Sri. T.S. Avinashilingam Ayya Avl.**, Founder and the First Chancellor, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for providing this temple of learning for women.

My reverential gratitude to **Late Dr. (Tmt) Rajammal P. Devadas Amma Avl.**, M.A., M.Sc., Ph.D. (OHIO State), Hon. D.Sc., Hon D.Sc., Former Chancellor, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, *for her heavenly blessings.*

I express my gratitude to the **Late Padma Shri. Dr. P. R. Krishnakumar**, Former Chancellor, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for providing me with an opportunity, infrastructure and all the amenities to carry out the research.

I express my heartfelt thanks to Former **Chancellor Dr. S. P. Thyagarajan**, D.Sc., M.D., Ph.D., Avinashilingam Institute for Home Science and Higher Education for women, for providing me the moral support for the conduct of research work.

I express heartfelt thanks to **Chancellor Sri. T.S.K. Meenakshisundaram**, M.A., M.Phil., Ph.D., Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for providing the facilities to conduct the study.

I express my immense gratitude to **Dr. (Mrs.) Premavathy Vijayan**, M.Sc., M.Ed., Dip. Spl. Edn. (U.K.), M.Phil., Ph.D., Former Vice Chancellor, Avinashilingam Institute for home science and Higher Education for women for her constant encouragement throughout the research work.

I express my sincere gratitude to **Dr. (Mrs) V. Bharathi Harishankar** Ph.D., FRSA, Vice Chancellor, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for providing the necessary platform and support for research work.

My sincere thanks to **Dr. (Mrs) S. Kowsalya**, M.Sc., M.Phil., Ph.D., Former Registrar, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for providing the opportunity to do this research.

My sincere thanks to **Dr. (Mrs) H. Indu**, M.Sc (Phy)., M.Ed., SLET (Edn)., Dip in Multimedia., M.Phil (Education)., Ph.D (Education)., M.B.A (Education Management), Registrar, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for providing the opportunity to do this research.

I sincerely thank to **Dr. (Mrs) K. Manimozhi**, M.Sc., B.Ed. M.Phil., Ph.D., Former Controller of Examination, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for giving suggestions and support to me at all times of need.

I owe deep gratitude to **Dr. K. Sambath Rani** M.R.Sc, M.Phil., M.Ed, (MR+VI), Ph.D., NET, Controller of Examination, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for her constant encouragement, support and wishes.

I record my gratefulness to **Dr.(Mrs.) P.Lalitha**, M.Sc., M.Phil., Ph.D., Dean, R&D, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore for her timely help, support and encouragement in carrying out the research work.

I am also thankful to **Dr. (Mrs). G. Padmavathi**, M.Sc., M.Phil., Ph.D., Former Dean, School of Physical Sciences and Computational Sciences, for granting the facilities required. I am obliged for their extended support throughout my work.

I sincerely thank **Dr. (Mrs). V. Radha**, M.Sc., B.Ed., PGDOR., PGDCA., M.Phil., Ph.D., Dean, School of Physical Sciences and Computational Sciences, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for providing her constant encouragement, motivation, and suggestions for my research work.

My thanks are also due to **Dr. (Mrs.) S.N. Geethalakshmi**, MCA., M.Phil., Ph.D., Former Professor and Head,, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for her blessings, support, timely help, encouragement and cooperation rendered towards the completion of this research.

My thanks are also due to **Dr. (Mrs.) B. Kalpana**, M.Sc., M.Phil, Ph.D., Professor and Head, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for her blessings, support, timely help, encouragement and cooperation rendered towards the completion of this research.

My heartfelt and deep sense of gratitude to my beloved supervisor, **Dr. (Mrs.) D. Shanmugapriya**, M.Sc., M.Phil., Ph.D., SET, Assistant Professor, Department of Information Technology., who enlivened and gave a golden opportunity to carry out my research on this topic, which helped me a lot to learn many new things. I am very grateful for her continuous guidance with all the useful discussions and brainstorming sessions, especially during the difficult conceptual development stage of my research. She has always been my thriving force and inspiration. Her unwavering enthusiasm for research kept me constantly engaged and fortified my passion for research.

I thank the Doctoral Committee Member **Dr. G. R. Karpagam**, B.E, M.E, Ph.D., Professor, Department of Computer Science and Engineering, PSG College of Technology, Coimbatore for helping me to fine-tune my research work through her valuable discussions, comments and suggestions.

I am thankful to the Librarians of Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore for utilizing various e-resources available in our university library. I also thank the technical staff members, who are the helping hands technically to carry out my research work safely and successfully.

I owe a deep sense of gratitude to reviewers, editors, scientists, professors, research forums, publishers, peer-reviewed journals, reputed conferences, workshops and seminars, the research community, resource persons, professionals, research scholars, invited speakers and colleagues for contributing their scientific ideas, knowledge, resources & materials, a precious suggestion which significantly supported me to learn, gain knowledge, and explore my research work to a great extent.

I am so thankful to all research scholars, staff members, Department of Computer Science and members of Computer Center, Department of Information technology, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore for their support, cooperation, suggestions and wishes.

I extend my heartfelt thanks to *all my family members* for their unwavering love, prayers, support, blessings and encouragement throughout this endeavor. I am also grateful to my friends and everyone else who has offered their support and well-wishes along the way. Besides this, I would like to extend my gratitude to one, and all who have knowingly or unknowingly, directly or indirectly help me in the successful completion of this work.

I am also grateful to the Almighty for the blessings throughout the journey which walks me in the correct direction for successful completion of this work.

Asha S

LIST OF FIGURES

S. No.	FIGURE No.	FIGURE TITLE	PAGE NO.
1	1.1	Classification of Insider Threat	4
2	1.2	Categories of Intentional Insider Threat	5
3	1.3	Categories of Unintentional Insider Threat	7
4	1.4	Categorization of the Potential Consequence of Intentional and Unintentional Insiders	11
5	1.5	Significant Contributions	17
6	2.1	Categories of Biometric Techniques	53
7	3.1	Methodology Overview	73
8	3.2	Techniques Proposed and Outcome Achieved	77
9	4.1	Methodology Overview of P&ID phase	81
10	4.2	Overview of Preprocessing Layer	87
11	4.3	Data Integration from Multiple Sources	89
12	4.4	Classification of Sampling Approaches	91
13	4.5	Limitations Observed in NM-2 Technique	99
14	4.6	Overview of B-SVM Technique	104
15	4.7	Results of P&ID Phase After Applying B-SVM using CERT Insider and CIC Darknet Datasets	120
16	5.1	Methodology Overview of UIM Phase	125
17	5.2	Overview of Feature Engineering Layer	131
18	5.3	Three Stages of CKPCA Technique for Feature Engineering	132
19	5.4	Overview of the DBN Technique	143
20	5.5	Results of Population Subset Selection	147

S. No.	FIGURE No.	FIGURE TITLE	PAGE NO.
21	5.6	Results of Feature Mapping	148
22	5.7	Results of Dimensionality Reduction	149
23	5.8	Results of UIM Phase After Applying CKPCA-DBN using CERT Insider and CIC Darknet Datasets	159
24	6.1	Methodology Overview of IIM Phase	164
25	6.2	Overview of Data Preprocessing Layer	169
26	6.3	Overview of Decision Tree Technique for Model Training and Evaluation	172
27	6.4	Performance of IIM phase	181

LIST OF TABLES

S. No.	Table No.	Table Title	Page No.
1	1.1	Analysis on Taxonomy and Typology of Intentional Insider Threats	6
2	1.2	Analysis on Taxonomy and Typology of Unintentional Insider Threats	8
3	1.3	Analysis of Different Types of Insider Threats	10
4	1.4	Significant Consequences of Intentional Insider Threats	12
5	1.5	Significant Consequences of Unintentional Insider Threats	14
6	1.6	Recent Statistics on the Consequence of Insider Threat	15
7	2.1	Review of Significant Research in Intentional Insider Threat Detection	30
8	2.2	Review of Significant Research in Unintentional Insider Threat Detection	38
9	2.3	Review of Significant Research in Class Imbalance Problem	45
10	2.4	Review of Recent Findings in Keystroke Dynamics	63
11	2.5	Analysis of Research Gap Identified and Proposed Contribution	69
12	4.1	Dataset Versions Associated with Five Scenarios	82
13	4.2	Dataset Description of CERT Insider Dataset	83
14	4.3	Dataset Description of the CIC Darknet Dataset	84
15	4.4	CIC Darknet Application Classes	86
16	4.5	Working Procedure of P&ID Phase	87
17	4.6	Detailed Description for CERT Log Files	89

S. No.	Table No.	Table Title	Page No.
18	4.7	Feature Description of Integrated Data Before Data Encoding	90
19	4.8	Feature Description of Integrated Data After Data Encoding	91
20	4.9	Working Criteria of Sampling Approaches	94
21	4.10	Summary of Sampling Approaches	94
22	4.11	Performance Metrics for Evaluating Sampling Approaches	95
23	4.12	Simulation Parameters for SVM	96
24	4.13	Results of Various Sampling Approaches	97
25	4.14	Before and After Undersampling Approaches	98
26	4.15	Sampling Strategy of NM-2	100
27	4.16	Results of Tuned Nearmiss2 using Four Versions of the CERT Datasets	101
28	4.17	Results of Data sampling using Tuned Nearmiss2	102
29	4.18	Analysis of Decision using Proposed B-SVM Algorithm	106
30	4.19	Performance Metrics to Evaluate B-SVM for Insider Threat Detection	110
31	4.20	Result of Data Integration using Simple Feature Concatenation Technique in CERT Insider Dataset	112
32	4.21	Result of Data Encoding using Categorical Encoding Technique in CERT Insider Dataset	113
33	4.22	Result of Data Encoding using Categorical Encoding Technique in CIC Darknet Dataset	114
34	4.23	Result of Data Sampling using Tuned Nearmiss2 Approach in CERT Insider Dataset	115

S. No.	Table No.	Table Title	Page No.
35	4.24	Result of Data Sampling using Tuned Nearmiss2 Approach in CIC Darkent Dataset	116
36	4.25	Result of Proposed B-SVM in the P&ID Phase	117
37	4.26	Mean Performance of Proposed B-SVM in the P&ID Phase	118
38	4.27	Performance Evaluation of B-SVM in the P&ID Phase vs Other State-of-art-methods	119
39	5.1	Temporal Events of the CMU Keystroke Dataset	126
40	5.2	Dataset Description of CMU Keystroke Dataset	127
41	5.3	Temporal Events of the Collected Keystroke Dataset	129
42	5.4	Dataset Description of Collected Keystroke Dataset	129
43	5.5	Working Procedure of the UIM Phase	131
44	5.6	Hyperparameter Used for Population Subset Selection using CSA	134
45	5.7	Best Hyperparameter Values for Population Subset Selection using CSA	134
46	5.8	Parameter Used for Feature Mapping using KME	136
47	5.9	Hyperparameter Used for Feature Mapping using KME	136
48	5.10	Performance Evaluation of KME and RF	136
49	5.11	Parameter Used for Dimensionality Reduction using PCA	138
50	5.12	Hyperparameter Used for Dimensionality Reduction using PCA	138
51	5.13	Performance of RF for Dimensionality Reduction using PCA	138
52	5.14	Parameter Specification of Proposed CKPCA	139

S. No.	Table No.	Table Title	Page No.
53	5.15	Advantage of CKPCA over Traditional PCA Approach	141
54	5.16	Analysis of Decision using DBN	145
55	5.17	Performance Metrics to Evaluate DBN for Core Behavior Identification	146
56	5.18	Result of Population Subset Selection using Clonal Selection Algorithm in CMU Keystroke Dataset	150
57	5.19	Result of Population Subset Selection using Clonal Selection Algorithm in Collected Keystroke Dataset	152
58	5.20	Performance of Feature Mapping using CMU Keystroke Dataset	153
59	5.21	Performance of Feature Mapping using Collected Keystroke Dataset	153
60	5.22	Performance of Dimensionality Reduction using CMU Keystroke Dataset	154
61	5.23	Performance of dimensionality reduction using collected keystroke dataset	154
62	5.24	Results of DBN in the UIM Phase using CMU Keystroke Dataset	155
63	5.25	Results of DBN in the UIM Phase using Collected Keystroke Dataset	156
64	5.26	Mean Performance of CKPCA-DBN in the UIM Phase	157
65	5.27	Performance Evaluation of CKPCA-DBN in the UIM Phase vs Other State-of-art-methods	158
66	6.1	Sample of Combined CERT Insider Dataset	165
67	6.2	Dataset Description of the Combined CERT Insider Dataset	166

S. No.	Table No.	Table Title	Page No.
68	6.3	Dataset Description of the Combined CIC Darknet Dataset	167
69	6.4	Working Procedure of the IIM Phase	169
70	6.5	Feature Description of the Combined Data Before Data Encoding	170
71	6.6	Feature Description of the Combined Data After Data Encoding	171
72	6.7	Performance Metrics to Evaluate a Decision Tree to Predict User Risk	176
73	6.8	Result of Data Encoding in the CERT Insider Dataset	177
74	6.9	Result of Data Encoding in the CIC Darknet Dataset	177
75	6.10	Results of Model Training and Evaluation using CERT Insider and CIC Darknet datasets	180
76	7.1	Significant Contribution in Proposed Methodology	186
