SPECIMEN FORMAT FOR THESES OF MONTH

Faculty	**:**	**School of Physical Sciences and Computational Sciences**
Department	**:**	**Computer Science**
Branch/ Area:	**:**	**Machine Learning**
Sub Subject Heading:	**:**	
Candidate's Name	**:**	**Anuradha.P**
Candidate's Address with email	**:**	118, 2nd Cross, 8th Main, Sir M.V Nagar II block, Ramamurthy Nagar, Bangalore-16 17phcsp007@avinuty.ac.in
Title of the thesis	**:**	PREDICTION OF HEART DISEASES RISK USING NOVEL MACHINE LEARNING TECHNIQUES
(i) In Roman Script (ii) In roman Script		- -
Nomenclature of Degree:	**:**	Doctor of Philosophy
Month & Year of Enrolment:	**:**	July 2017
Month & Year of Registration:	**:**	July 2017
Month &Year of Submission:	**:**	November 2024
Month &Year of Award	**:**	August 2025
Name of Supervisor	**:**	Dr.Vasantha Kalyani David
Designation of Supervisor	**:**	Professor
Centre/department/school in which research was conducted	:	Department of Computer Science, School of Physical Sciences and computational Sciences
University's Name & Address	**:**	Avinashilingam Institute for HomeScience and Higher Education for Women, Deemed to be University, Coimbatore-43, Tamil Nadu, India

Abstract within 300 words:

Heart diseases are a major health concern globally and early detection of the disease plays a crucial role in reducing mortality rates and improving patient outcomes. With the advancements in machine learning, there is an increasing focus on utilizing these techniques to enhance the prediction and diagnosis of heart diseases, even before symptoms manifest. Machine learning (ML) research in this field aims to develop accurate and efficient models that can assist in clinical decision-making. There are a variety of ML methods making use of stand-alone classifiers and hybrids. However, the results from these models vary considerably between various cardiac datasets and/or they are not modelled using both low and high dimensional data. This research proposes novel and improved feature selection and classification techniques towards enhancing heart disease prediction performance.

In the first stage of work, feature selection using Feature Importance (FI) ranking of Gradient Boosting algorithms is done and a significant reduction in the search space of feature subsets is identified. Next, a novel feature selection algorithm called ModifiedBoostARoota (MBAR) is proposed, which identifies the risk parameters that strongly contributes to the prediction of heart disease. This algorithm incorporates CatBoost as the base model and utilizes a novel feature elimination process.

In the second phase of the work, a novel Super Learner Ensemble Model (SLEM) is proposed, which is an integration of diverse ML base models selected by repeated stratified k-fold cross validation. A meta learner logistic regression is employed to learn from the predictions of the base classifiers. By backward elimination method, an optimal combination of classifiers in SLEM was identified.

In the third phase of the work, to further improve the classification in the SLEM model, a new Optimized Super Learner Ensemble Model (OSLEM) is proposed. OSLEM utilizes the Whale Optimization Algorithm and pairwise disagreement accuracy diversity measure to select an optimal set of base learners in OSLEM. The final proposed model comprising of MBAR and OSLEM presented high performance when compared with existing models across the heart datasets.

i) Major objectives :

- To devise novel feature selection techniques to select relevant features for identifying the appropriate risk parameters for the prediction of heart disease with less computational cost
- To develop an ensemble classifier with high prediction accuracy by selecting an appropriate combination of diverse base learners
- To improve the ensemble model by optimizing the selection of classifiers in the ensemble

ii) Hypothesis:

Feature selection by ModifiedBoostARoota algorithm and classification by Optimized Super Learner Ensemble Model (OSLEM) will outperform other models in heart disease classification.

iii) Methodology :

- A wrapper feature selection method, ModifiedBoostARoota, is developed, which uses CatBoost as the base model and a novel feature elimination approach to efficiently identify the risk parameters of heart disease.
- Also, Feature selection by Feature Importance Scores of Gradient Boosting Algorithms is proposed with a significant reduction in the search space of the sets of features
- A novel Super-Learner Ensemble Model (SLEM) with an appropriate combination of diverse classifiers is proposed and by backward elimination method, an optimal combination of classifiers in SLEM was identified.
- An Optimized Super Learner Ensemble Model (OSLEM), using the Whale Optimization Algorithm and pairwise divergence measure to select an optimal set of base learners, is devised to improve the efficiency of the classification model.

iv) Findings:

- Feature Selection improved the classifier performance
- Optimizing the selection of base learners in the ensemble improved the performance of the classifier

Examiners

Internal Examiner :

Dr.Veena Goswami
Professor and Director General
School of Computer Applications
Kalinga Institute of Industrial Technology
Bhubaneswar.

External Examiner :

Dr.Elpiniki I Papageorgion
Professor
Department of AI Energy Systems
University of Thessaly
Larissa, Greece