State-of-the-Art Study on Search Algorithms

“ our title over here”

Put names and formatting

# Chapter 1

## Introduction To the Project

In the realm of real-world problem-solving, a common challenge emerges: the management of vast datasets with numerous attributes or features. Negotiating this data labyrinth can be an intricate task. Within these datasets lie attributes, some of which may be redundant, or irrelevant, all of which can significantly impede the performance of machine learning models.

Dimensionality too poses issues, potentially leading to overfitting and prolonged learning times. The crux of the matter is to reduce the dimensionality of these datasets while retaining their predictive accuracy. This is where the concept of feature reduction comes into play.

Feature reduction, branches into two key facets: feature construction and feature selection. Feature construction entails the creation of novel features derived from the original dataset, while feature selection involves the process of picking the most pertinent attributes from the dataset. We mainly focus on feature selection.

When dealing with a dataset containing 'n' features, 2^n subsets become potentially relevant. As 'n' grows larger, assessing the performance of the model for each subset becomes impractical. To mitigate this, various methodologies have been introduced. Exhaustive search, greedy search, random search, and more have been deployed to tackle feature selection challenges. However, many of these methods suffer from issues such as premature convergence, unnecessary complexity, and high computational costs. Consequently, the spotlight has shifted towards metaheuristic algorithms as a means of addressing these challenges.

These algorithms offer efficiency and effectiveness, capable of finding optimal feature subsets while preserving accuracy. Pruning features via these algorithms mitigates overfitting, reduces vulnerability to missing data, and facilitates better model interpretation and generalization.

For instance, in image analysis, feature selection helps discern the salient visual components, such as pixels and colours.

Mathematically, the feature selection problem can be framed as :

Given: dataset 'S' containing 'd' features.

Goal: to select the most relevant features.,

Consider dataset S = {f1, f2, …, fd}, the objective is to extract a subset D = {f1, f2, ..., fn},

where n < d, and f1, f2, f3, …, fn represent the attributes of the dataset.

Feature extraction for images encompasses the derivation of higher-level features from raw data, such as edges, corners, or image segments. Some lower-level primary features include size, colour values, intensity.

Yet, the inherent challenge in image datasets lies in their vastness, as they often comprise thousands of features due to the multitude of pixels. The quality of an image is directly correlated with the number of pixels, which, in turn, amplifies computational complexity.

Feature selection for images can be achieved through wrapper methods, filters, and embeddings. Wrapper methods evaluate feature subsets by training models, making them suitable when labelled data is available. Filters, on the other hand, select features based on statistical criteria or domain knowledge, making them computationally efficient.

Embeddings transform high-dimensional image data into lower dimensions, often using techniques like PCA (Principal Component Analysis) or other embeddings. The choice depends on the task, dataset size, and computational resources, with a combination of methods often yielding the best results. Pre-trained networks are commonly used for extracting image embeddings, offering a powerful feature selection approach.

Given the size of the feature space, feature selection can be framed as an optimization problem, thereby making it an ideal candidate for the application of metaheuristic-based optimizers. These algorithms have demonstrated their prowess in solving complex optimization challenges.

A growing area of research pertains to the use of image classification for disease detection, particularly through X-ray and CT-scan images. Over the past few years, metaheuristic algorithms have gained traction for optimizing deep neural network architectures. Their simplicity, flexibility, and problem-agnostic nature render them ideal candidates for tackling intricate optimization problems.

Metaheuristic algorithms constitute a class of optimization methods that excel at seeking optimal or near-optimal solutions for a variety of optimization problems. These algorithms are notable for their lack of reliance on derivatives and their stochastic nature. They initiate their optimization process by generating random solutions, obviating the need for derivative calculations typical of gradient-based techniques. What sets metaheuristic algorithms apart is their simplicity and flexibility, making them open to customization for specific problems. A defining characteristic of these algorithms is their ability to avoid premature convergence, a pitfall often associated with optimization. Their stochastic nature treats the problem as a black box, enabling them to explore the search space thoroughly and evade local optima effectively. Metaheuristic algorithms strike a delicate balance between exploration and exploitation, investigating promising search spaces before delving into local searches in the regions identified during exploration.

To delve into the metaheuristic algorithms applied in feature selection, we delve into the realm of binary vector representations. In this paradigm, each solution vector is represented as a sequence of binary values, such as (10101100...). Here, a '1' indicates the selection of a particular feature, while '0' signifies its exclusion from the subset. This binary representation forms the foundation for metaheuristic algorithms striving to identify optimal feature subsets.

# Chapter 2

## Literature Survey

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SNo | TITLE | METHOD/DESCRIPTION | RESULTS | COMMENTS |
| [1] | An Effective Feature Selection Scheme for Healthcare Data Classification Using Binary Particle Swarm Optimization  Conference Proceedings:  2018 9th International Conference on Information Technology in Medicine and Education (ITME)  Author: Yiyuan Chen  Publisher: IEEE  Date: October 2018 |  |  |  |
| [2] | [Computers in Biology and Medicine](https://www.sciencedirect.com/journal/computers-in-biology-and-medicine) [Volume 144](https://www.sciencedirect.com/journal/computers-in-biology-and-medicine/vol/144/suppl/C), May 2022, 105344 Metaheuristics based COVID-19 detection using medical images: A review Authors:Mamoona Riaz, Maryam Bashir, Irfan Younas |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# Chapter 3

## Conclusion and Plan