Chapter 11

Two Enhancement Levels for Male Fertility Rate Categorization Using Whale Optimization and Pegasos Algorithms

Abeer S. Desuky
https://orcid.org/0000-0003-1661-9134
Al-Azhar University, Egypt

ABSTRACT

Recently, diseases and health problems that were common only in the elderly became common also among the youth. Some of these medical problems causes include behavioral, environmental, and lifestyle factors. The decrease in fertility rates especially among the male population is one of those problems. Now, machine learning and artificial intelligence algorithms are emerging methodologies as computer-aided decision systems in medical diagnosis and health problems. In this chapter, the incorporation of the bio-inspired whale optimization algorithm (WOA) and Pegasos algorithm are used to enhance the male fertility rate categorization in two levels. Results show that implementing WOA as the second level of enhancement gives better accuracy than the first level of enhancement in Pegasos algorithm with a prediction accuracy value of 90%. Using two machine learning algorithms to categorize the male fertility rate helped in the overall improvement of the proposed system performance to give results that exceeded all recent research results for fertility data.

INTRODUCTION

There has been a remarkable decrease in fertility over the last two decades and the global health community has recently been concerned with the decline in male fertility. Studies showed that thirty percent of infertility is related to male factor. The evaluation of male fertility potential was based mainly on se-

DOI: 10.4018/978-1-6684-5092-5.ch011

men analysis which is the core of evaluating men for fertility or infertility, but late studies have proved that the test is not the only available way for predicting the fertility status of men reliably. Infertility has many causes including hormonal imbalances, psychological, environmental and behavioral factors (World Health Organization [WHO], 2010) (Barrat et al., 1998). To obtain a quick assessment of the seminal profile of the patients based on these factors, machine learning, and artificial intelligence techniques have been applied as computer aided decision systems in the medical diagnosis field. Such systems can be used for the prior assessment of semen quality before more exhaustive, uncomfortable, and costly tests are conducted on patients (Wang et al., 2014).

Computer aided decision (CAD) systems that rely on a wide range of classifiers have been investigated and implemented for the diagnosis of various diseases. Constructing a classifier requires a set of data samples representing previous experience (Mazurowski et al., 2008). Generally, the larger and more representative the set of samples, the better the categorization of new query cases can be obtained. However, in the medical domain, collecting data faces various challenges and practical limitations. One of these is the time consumed in collecting data from patients. Another challenge, gathering large numbers of patients infected with diseases that have low prevalence (imbalance). Moreover, the clinical presentation of the same disease' patients vary dramatically. According to this variability, computer aided systems are requested to handle large numbers of features, often many of which are correlated and) or (have no significant diagnostic value. These described issues (limited sample size, imbalanced data, and large numbers of potentially insignificant/correlated features) can detrimentally affect the performance and development of CAD classifiers (Barrat et al., 1998; Mazurowski et al., 2008; Raudys & Jain, 1991).

As is prevailing in medical data diagnosis (Mazurowski et al., 2008), most available male fertility rating data also have the class imbalance problem, where the number of normal (negative) samples far exceeds the number of altered (positive) samples. Owing to this distorted distribution of data, many learning algorithms often perform inaccurately on the minority class (Wang et al., 2014; Jose et al., 2013). Some researchers have handled classification in the presence of these challenges. David et al. (2012), proposed and compared three machine learning classifiers: multilayer perceptron (MLP), decision trees (DT) artificial and support vector machines (SVM) and the highest accuracy was 86% for SVM and MLP. In another work (Macmillan et al., 2015), Artificial Neural Network (ANN), and Naïve Bayes (NB) classifiers were presented for the characterization of seminal quality, based on environmental factors and lifestyle habits comparisons between the two classifiers showed that their accuracy rate was identical (80%). Particle Swarm Optimization (PSO) was also proposed in (Puneet et al., 2015) compared to another three learning algorithms. PSO showed the highest performance accuracy (88%) whereas SVM and MLP showed less performance accuracy (85%) and DT had much less accuracy (84%).

The objective of this chapter is to enhance the categorization of the male fertility rate by considering some environmental and behavioral factors. Due to the complexity of the problem itself, the data imbalance, and its impact on categorization accuracy, two levels of enhancement will be applied using the Pegasos algorithm as the first enhancement level and the bio-inspired whale optimization algorithm as the second enhancement level.

BACKGROUND

Optimization techniques were utilized for different real-world applications, Negi et al., (2021) proposed a hybrid approach with Particle Swarm Optimization (PSO) and Gray Wolf Optimization (GWO) (Mirjalili