International Journal of Swarm Intelligence Research
Volume 13 • Issue 1

Assessment of Risk Factors in Medical Data Using Improved Binary Artificial Fish Swarm Algorithm With Classification Upon Evaluation From F-Test

Sheik Abdullah A., Thiagarajar College of Engineering, India

ABSTRACT

The objective of this research work is to effectively deploy improved binary artificial fish swarm optimization algorithm (BAFSA) with the data classification techniques. The improvement has been made with accordance to the condition of visual scope and the movement of fish to update towards the central position, and chasing behavior towards best point of movement has been modified among the given population. The experimental results show that feature selection by BAFSA and classification by decision trees and Gaussian naïve bayes algorithm provides an improved accuracy of about 89.6% for Pima Indian diabetic dataset, 91.1% for lenses dataset, and 94.4% for heart disease dataset. Statistical analysis has also been made using Fisher's F-Test for two sample variance and the selected risk factors such as glucose, insulin level, blood pressure for diabetics datasets, spectacle prescription, tear production rate for lenses dataset and trestbps, cholesterol level, thalach, chest pain type for heart disease dataset are found to be significant with P value <0.001 respectively.

KEYWORDS

Data Classification, Electronic Medical Record, Feature Selection, Predictive Modelling, Risk Factors, Swarm Intelligence

INTRODUCTION

Data analytics is the science of processing data and changing into information that is employed to enhance productivity and business profit. Data is extracted and differentiated to recognize and analyze behavioral data and patterns, and techniques utilized by organizations vary from one organization to another. Data analytics are employed in several industries to allow firms to form a higher business decisions. There are different types of analytics. They are Predictive Analytics, Descriptive, survival, and Social media analytics. When considering all these analytics, predictive analytics is beneficial to predict future events. Predictive analytics uses many techniques such as statistical learning, machine learning, data mining, and artificial intelligence. The patterns found in the past and transactional data can recognize risks opportunities for the future. Applications of predictive analytics are as follows; Customer Relationship Management(CRM), Fraud Detection in Banking sectors, Risk Management, Direct Marketing, Healthcare, etc.,

Mainly Predictive analytics are used by healthcare professionals to process patient data, forecast the potential for illness, identify high-risk patients, and reduce hospital readmission rates, and so on. Risk stratification enables in ranking clinical progress, reducing system waste, and makes financially

DOI: 10.4018/IJSIR.2022010105

efficient population management. For instance, a calculated risk score will promote decrease system waste by setting progress priorities for patient follow-up inpatient populations. Using this mechanism, care managers are prompted to concentrate on those patients the highest risk and preemptively get involved with medication reconciliation, planning home visits, or follow-up appointments.

The applications in machine learning have been adherently increased in recent years. Various techniques have been proposed to remove and filter the non-important features has become challenging task Youchuan W *et al*. (2016). The feature selection paradigm helps to locate the most essential patterns in the data. It makes to improve the overall prediction performance concerning time and space complexity Girish Chandrashekar G & Sahin F (2014). There are about three methods in feature selection, such as a wrapper, filter, and embedded method (Deepa, N *et al*. 2020). The working principle behind the filter method is the ranking of attributes. In wrapper methods, the classification and feature selection paradigm is wrapped to get the desired output. In the embedded way, the best subset of features is observed with an improvement over accuracy Iman K *et al.* (2015).

Modern medicine generates almost daily, vast amounts of heterogeneous data. For example, medical data may contain SPECT images, signals like ECG, clinical information like temperature, cholesterol levels, etc., as well as the physician's interpretation. Those who deal with such data understand that there is a widening gap between data collection and data comprehension. Extracting useful knowledge and providing scientific decision-making for the diagnosis and treatment of disease from the database is becoming necessary Chen *et al.* (2017).

The health sector needs to be improved by enhancing medical facilities, disease-specific risk factor determination, and by providing health awareness among the people. In addition to the health sector, there lays an individual responsibility and awareness particular to the disease concerned. Enhancing and rendering health-based service also depends upon the likelihood and habits of the people around a particular region. If the risk specific syndromes are detected in advance, cost-effectiveness and treatment expenses can be avoided, and thereby, we can render population-based healthcare service.

Computerized techniques are needed to help humans address this problem. Data mining in medicine can deal with this problem, as given by Stolba and Tjoa (2002). As more and more medical procedures employ imaging as a preferred diagnostic tool, there is a need to develop techniques for efficient mining in databases of images. Stuhlinger *et al.* (2003) discussed the other significant features such as security and confidentiality in handling the medical data by which the reports based on each patient can be managed efficiently. This research focuses on the development of a decision support model using the Binary Artificial Fish Swarm optimization algorithm with data classification techniques upon parameter estimation Chen et al. (2015). The deployment of the model will provide an efficient way to determine the predominant risk factors that contribute towards medical data and its co-morbidities.

The realm of nature-inspired computing lies precisely at the feature selection process and it's confirming parameters. In the existing paradigm, there occurs no such statistical validation and evaluation with feature selection and classification for medical data analysis. This research work entirely focuses on developing a decision support model with a Binary Artificial Fish Swarm optimization algorithm along with data classification techniques for medical data. The model has been explicitly developed to handle medical data with variant features and its significance. The evaluation of the model has been tested with significance corresponding to Fisher's F-Test and its validation process.

LITERATURE REVIEW

Aleksandar K *et al.* (2013) proposed a solution called PE-CMR solution for estimating the cardio-metabolic risk. The group consists of 1281 respondents (692men and 589women) aged from 18 to 67 years. The risk factors are BP, WH ratio, Glycaemia, uric acid. The improvement can be made with the combination and refinement over the development of the neural network model. The evaluation