Chapter 24
Predictive Analytics on Female Infertility Using Ensemble Methods

Simi M. S.
Adi Shankara Institute of Engineering and Technology, India

Manish T. I.
https://orcid.org/0000-0002-1347-5510
SCMS School of Engineering and Technology, India

ABSTRACT

With the accessibility of healthcare data for a significant proportion of patients in hospitals, using predictive analytics to detect diseases earlier has become more feasible. Identifying and recording key variables that contribute to a specific medical condition is one of the most difficult challenges for early detection and timely treatment of diseases. Conditions such as infertility that are difficult to detect or diagnose can now be diagnosed with greater accuracy with the help of predictive modeling. Infertility detection, particularly in females, has recently gained attention. In this work, the researchers proposed an intelligent prediction for female infertility (PreFI). The researchers use 26 variables for the early diagnosis and determine a subset of these 26 variables as biomarkers. These biomarkers contribute significantly to a better prediction of the problem. The researchers designed PreFI using ensemble methods with biomarkers and improved the performance of the predictive system.

INTRODUCTION

The amount of data in our medical systems has steadily increased with the advent of electronic medical records and increased computing power (IHTT, 2013). The number of patients and the amount of data stored per patient have both increased, resulting in an increase in data. As a result, in the health-care industry, implementing a solid data analytics platform has become critical (Raghupathi, 2010). The process of generating actionable insights by defining problems and applying statistical models and analysis to

DOI: 10.4018/978-1-7998-8892-5.ch024

existing data is referred to as data analytics (Cooper, 2012). The analysis of this large dataset can be used to generate data that will help doctors diagnose diseases earlier and more accurately (Raghupathi, 2014).

Electronic Health Records (EHR) have been incorporated to provide more coordinated and patient-centered care. The use of an Electronic Health Records (EHR) in the ICU significantly reduces central line-associated bloodstream infections and surgical intensive care unit mortality rates (Flatow, 2015). Electronic Health Records (EHR) provide secure access to patient data, which improves care quality and productivity (Tharmalingam, 2016). Electronic Health Records (EHR) systems have been used to manage chronic diseases such as diabetes, and it has been discovered that if providers participate in health information exchanges, regular use of the Electronic Health Records (EHR) can reduce data fragmentation and increase provider continuity of care (Rinner, 2016). Using patient data, specialized AI systems assist specialists in their clinical workflow by recognizing and diagnosing various diseases (Simi, 2017). In the emergency department (ED), using a decision tree with Electronic Health Records (EHR) improves medical decision making, increases patient quality of life, and is cost-effective (Ben-Assuli, 2016). Another cost-benefit analysis of using Electronic Health Records (EHR) to collect data yielded encouraging results (Beresniak, 2016).

One of the most common diseases affecting humans is infertility. In accordance with World Health Organization (WHO), this issue affects 60 to 80 million people (WHO, 2004), with infertility affecting 17% of females between the ages of 20 and 24. More than 186 million people worldwide are infertile, with the majority living in developing countries (Bittles, 2010). Female infertility can occur for a variety of reasons. In some cases, the disease could be caused by physiological factors. Sometimes there is no obvious cause for the disease. Ovulation disorders, endometriosis, tube damage, uterine disorders, and even lifestyle and environmental elements can all contribute to infertility (Amoako, 2015).

The excessive time it takes to detect the true reason of infertility is one of the most typical trends. A test to confirm a condition can take up to six months, however this delay in diagnosis can alter the likelihood of total cure or the pace with which the disease is cured. Our research focuses on the early detection of unexplained infertility issues. Because clinicians are often unable to diagnose the causes of unexplained infertility, the couple must undergo a battery of costly tests to determine the cause of infertility. Clinicians can easily predict unexplained infertility using our proposed system, and the couple can opt for assisted reproductive technology (ART). As there is no time lag between detection and treatment, the success rate of ART can be significantly improved.

Predictive modelling for infertility diagnosis is still in its initial phases of development. The majority of articles only predict infertility as certain or uncertain (Idowu, 2016). They don't look into the data's causes or conclusions. The majority of this research was done in hospitals with limited population data sets. In this work, the authors classify a broader range of inferences and identify likely, unlikely, and other probable (but not imminent) cases of infertility. For five types of ensemble learners, the authors predict with greater than 90% accuracy. The researchers expanded the number of variables in our work to include twenty-six variables in total, thirteen of which the researchers used for the first time. Our work also made significant contributions to prediction by adapting random forest (RF) (T.K. Ho, 1995) and J48 (Quinlan, 1993).

In this work, the researchers explored various available predictive techniques for early diagnosis of female infertility problems and proposed an intelligent prediction for female infertility (PreFI). The major contributions of our work are (a) the identification of key variables that contribute to female infertility and (b) expanding the prediction system from binary classification to a problem of prediction among 9