**International Journal of Healthcare Information Systems and Informatics**
Volume 15 • Issue 1 • January-March 2020

# A Novel Hybrid Approach for Chronic Disease Classification

Divya Jain, Computer Science and Engineering, The NorthCap University, Gurugram, India
Vijendra Singh, Computer Science and Engineering, The NorthCap University, Gurugram, India

## ABSTRACT

A two-phase diagnostic framework based on hybrid classification for the diagnosis of chronic disease is proposed. In the first phase, feature selection via ReliefF method and feature extraction via PCA method are incorporated. In the second phase, efficient optimization of SVM parameters via grid search method is performed. The proposed hybrid classification approach is then tested with seven popular chronic disease datasets using a cross-validation method. Experiments are then conducted to evaluate the presented classification method vis-à-vis four other existing classifiers that are applied on the same chronic disease datasets. Results show that the presented approach reduces approximately 40% of the extraneous and surplus features with substantial reduction in the execution time for mining all datasets, achieving the highest classification accuracy of 98.5%. It is concluded that with the presented approach, excellent classification accuracy is achieved for each chronic disease dataset while irrelevant and redundant features may be eliminated, thereby substantially reducing the diagnostic complexity and resulting computational time.

## KEYWORDS

Chronic Disease Diagnosis, Hybrid Approach, Optimized Support Vector Machines, PCA, Relief

## 1. INTRODUCTION

Lately, the devastating effects of chronic diseases on the health and wellbeing of populations all over the globe are becoming worrisome. In 2016, the World Health Organization (WHO) estimates a total of 383,000 deaths in 2015 arising from asthma with 235 million people suffering from this debilitating chronic illness that commonly affects children (WHO, 2016). Also, in 2015, the WHO (2016) statistics reveal that around 13.4 lakh deaths occurred due to viral hepatitis. More recently, in 2017, WHO reports that approximately 205 million women suffer from diabetes worldwide (WHO, 2017). Moreover, these effects are likely to be exacerbated with the rapidly aging populations growing in many countries all over the world. Conceivably, it is essential, in times such as these, that efficient diagnostic mechanisms be developed to render best possible remedy to patients inflicted with various chronic diseases in the least amount of time to sustain a decent level of population health and wellbeing in the developed, developing and under-developing countries of the world.

Accordingly, early and accurate diagnosis of chronic diseases plays a key role in medical informatics applications and clinical care services. With the advent of the digital era, data analytics in the medical field have continued to grow in handling voluminous amount of patient generated data with multidimensional characteristics. Big data mining provides efficient tools and techniques that can aid in health insurance decision-making (Tan et al., 2009) and other clinical decision-making (Han et al., 2011; Larose & Larose, 2014). As medical datasets contain all sorts of data from heterogeneous

DOI: 10.4018/IJHISI.2020010101
This article, originally published under IGI Global's copyright on September 27, 2019 will proceed with publication as an Open Access article starting on January 14, 2021 in the gold Open Access journal, International Journal of Healthcare Information Systems and Informatics (converted to gold Open Access January 1, 2021), and will be distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

sources, there is an utmost need to develop innovative analytical methods and to formulate effective learning algorithms to deal with high-dimensional data.

The rest of this paper is organized as follows. Section 2 discusses the background on existing classifiers for high data dimensionality reduction and argues the fundamental basis for undertaking our proposed approach for data analytical innovation. Section 3 details the dataset materials and methods used in this research. Section 4 depicts the hybrid model being proposed. Following this, Section 5 presents the study results and their comparative evaluation vis-à-vis other analytical methods. Section 6 summarizes the findings, highlighting key effects of the various statistical significance tests. Finally, Section 7 concludes the paper by reviewing the study contributions, provides insights on implications of the findings and potential future work for others interested in this line of research.

## 2. BACKGROUND ON EXISTING CLASSIFIERS

Over the years, numerous machine-learning methods have been applied successfully for effective disease diagnosis (Azzawi et al., 2016; Dora et al., 2017). Owing to its many excellent features and remarkable generalization performance (Cortes & Vapnik, 1995; Critianini & Shawe-Taylor, 2005), support vector machine (SVM) classifier is a widely used learning method in huge demand today. SVM is based on finding the maximum-margin hyper-plane for the separation of two classes as wide as possible. It has gained popularity in a wide range of biological applications (Noble, 2006). Also, with the use of kernel functions, SVM works efficiently with linear as well as non-linear datasets.

With the optimization of parameters used in kernel functions, the SVM classification accuracy increases at a significant rate. Grid search technique, a widely used method for optimizing SVM classifier, helps in finding best parameters to tune the performance of SVM (Lin et al., 2008; Tang et al., 2009). Although SVM has exceptional advantages, its performance will significantly improve if dimensionality reduction techniques are applied before the classification of data (Xie et al., 2010). Notably, researchers use feature selection extensively to reduce high data dimensionality (Dash & Liu, 1997; Guyon & Elisseeff, 2003; Chizi & Maimon, 2009; Kumar & Minz, 2014; Ang et al., 2015; Shardlow, 2016). For disease diagnosis, feature selection eliminates the attributes that are least significant to a particular disease.

The extant literature cumulates studies demonstrating noteworthy results from authors who have conducted research in the field of high data dimensionality reduction with classification. Polat & Gunes (2007), for example, proposed a novel medical system based on principal component analysis (PCA), fuzzy approach and adaptive neuro-fuzzy inference system (ANFIS) classifier for the diagnosis of lymph diseases, giving promising results with a high classification accuracy of 88.83%. Babaoğlu et al. (2010) investigated the effect of PCA-SVM based approach for the diagnosis of coronary artery disease (CAD). Their experimental analysis and results demonstrated that the PCA-SVM system not only diminishes the high data dimensionality and training-test timing, but also increases the accuracy of the SVM classifier.

More recently, Pang et al. (2015) applied ReliefF-SVM based method for the computer-aided diagnosis of breast tumor, yielding positively appealing results with 90.0% accuracy rate, 98.7% sensitivity and 73.8% specificity rate. Shah et al. (2017) employed SVM classifier with probabilistic PCA through parallel analysis for the automatic detection of heart disease. Statistically, their results showed that the presented method reduced approximately 70% of the dimensionality of the dataset with highest classification accuracy of 91.30%. Sangaiah & Kumar (2018) presented a hybrid approach using ReliefF method and entropy-based genetic algorithm for the diagnosis of breast cancer. The results they obtained depicted a superior performance of the approach with reduced dimensionality and significant accuracy rate. Finally, Singh & Pandey (2018) proposed an efficient diagnostic system using a unified two-phase approach based on PCA and K-nearest neighbor classifier (PCA-KNN) for the detection of liver disease. Their prediction model attained remarkable results in terms of various evaluation measures such as positive v. negative predictive value, accuracy, specificity and sensitivity.

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/a-novel-hybrid-approach-for-chronic-disease-classification/239563

## Related Content

Clustering-Based Recommendation System for Preliminary Disease Detection
Gourav Jain, Tripti Mahara, S. C. Sharma, Om Prakash Vermaand Tarun Sharma (2022). *International Journal of E-Health and Medical Communications (pp. 1-14).*
www.irma-international.org/article/clustering-based-recommendation-system-for-preliminary-disease-detection/313191

An In-Depth Analysis of Smart Healthcare Monitoring Systems for Humans Based on IoT
Rubal Jeetand Sandeep Singh Kang (2023). *International Journal of Healthcare Information Systems and Informatics (pp. 1-16).*
www.irma-international.org/article/an-in-depth-analysis-of-smart-healthcare-monitoring-systems-for-humans-based-on-iot/324931

Pediatric Telepsychiatry as Innovation in Healthcare Delivery
Katherine M. Boydell, Tiziana Volpeand Antonio Pignatiello (2010). *Health Information Systems: Concepts, Methodologies, Tools, and Applications (pp. 2044-2056).*
www.irma-international.org/chapter/pediatric-telepsychiatry-innovation-healthcare-delivery/49980

Transformation of a Reluctant Patient to a Proactive Health Advocate
Peg Ford (2013). *International Journal of User-Driven Healthcare (pp. 1-12).*
www.irma-international.org/article/transformation-reluctant-patient-proactive-health/76680

USE IT: The Theoretical Framework Tested on an Electronic Prescription System for General Practitioners
Ton A.M. Spil, Roel W. Schuringand Margreet B. Michel-Verkerke (2006). *E-Health Systems Diffusion and Use: The Innovation, the User and the Use IT Model (pp. 147-176).*
www.irma-international.org/chapter/use-theoretical-framework-tested-electronic/9042