

# Chapter 12

## Decision Support System for Diabetes Classification Using Data Mining Techniques: Classification Using Data Mining Techniques

**Ahmad M. Al-Khasawneh**  
*Hashemite University, Jordan*

### **ABSTRACT**

*The use of data mining algorithms in health information systems has played a significant role in developing applications that help to diagnose different diseases. The type of the disease determines the selection of the algorithm, parameters to be used, and dataset pre-processing steps, etc. In this chapter, diagnosing diabetes mellitus is the target since it has gained significant attention in the last few decades due to the increased severity of the disease. Four predictive data mining approaches are being used in diagnosing diabetes. Four models were implemented to diagnose diabetes from PIMA dataset: k-nearest neighbor, support vector machine, multilayer perceptron neural network, and naive Bayesian network. Giving the highest classification accuracy, support vector machine technique outperformed the others with a value of 78.83%.*

### **INTRODUCTION**

The number of diabetes patients had increased remarkably over the last few decades. Diabetes is known as an escalate percentage of sugar in the blood exceeding the normal levels that can't be absorbed properly by the body cells. There are two types of Diabetes; (type1) happens due to inability of the pancreas to produce enough insulin or (type2) inability of the body cells to respond to the produced insulin (IDF Diabetes Atlas, 2013).

DOI: 10.4018/978-1-5225-5460-8.ch012

It is known that medical diagnosis is a classification problem (Saidi, Chikh, & Settouti, 2011). In order to diagnose diabetes, many researchers developed clinical decision support systems to help clinicians to professionally diagnose the syndrome in its early stages. Most of the researchers applied data mining to classify information required for decision making within Clinical Decision Support Systems (Kazemzadeh et al., 2010). Clinical decision support systems (CDSSs), or diagnosing decision support systems (DDSSs) help specialist to decide whether a patient suffers from a specific disorder depends on the medical signs, symptoms, and tests (Kazemzadeh et al., 2010) (Al-Khasawneh & Hijazi 2014)

One of the best computer assistance methods that can be helpful for healthcare decision-making is data mining (Hajiheydari et al., 2013). The use of data mining in health care applications is usually known as clinical data mining (CDM). CDM involves the conceptualization, extraction, analysis, and interpretation of the available clinical data for practical knowledge-building, clinical decision making, and partition reflection (Jacob & Ramani, 2012).

CDSS usually employ predictive data mining to diagnose a disease (Al-Khasawneh & Hijazi 2014). The availability of previous data is a key factor in Predictive data mining, predictive data mining build models from historical data to predict the outcome for new patients. Data mining models are objective models since it is driven by the available data. In the predictive data mining, the data set consists of instances, each instance is characterized by attributes or features and another special attribute represents the outcome variable or the class (Bellazzi & Zupanb, 2008).

Predictive data mining is a supervised model building algorithm (Williams, 2011) which tries to predict trends and future behaviours depending on historical variables (Omari, 2013) and values wherein the probable values of the outcome are specified previously.

In this chapter, Four models had been implemented to diagnose diabetes; k-nearest neighbour, support vector machine, multilayer perceptron neural network, and naive bayesian network. All of the models were implemented from the Pima Indian diabetes dataset and validated using 10-cross validation techniques.

## **BACKGROUND**

Many researchers applied data mining techniques in solving clinical problems, predictive models are utilized to help in disease diagnosis process, specifically, the diabetes illness. A survey was conducted by (Tomar and Agarwal, 2013) on data mining approaches (classification, regression, and clustering approaches) used in healthcare. Classification models predict the class of a new observation among predefined categories of the target variable, while in regression modelling the output is a numeric value (Williams, 2011).

Predictive data mining builds both classification and regression modelling using several algorithms such as decision trees, random forests, boosting, support vector machines, linear regression, neural networks, naive Bayesian classifier, Bayesian networks, and the k-nearest neighbours' models (Bellazzi & Zupanb, 2008; Williams, 2011; Al-Khasawneh & Hijazi 2014). Pradhan et al. (2011) introduced general guidelines to design a predictive model for diagnosing diabetes using hybrid of soft computing and data mining techniques

A classifier to diagnose coronary heart disease using SVM algorithm with a radial basis function (RBF) was proposed by Kumari and Chitra (2013). They compared the results they obtained with another linear discriminate analysis (LDA) classifier. They found that SVM classifier surpass the LDA classifier in the cross-validated prediction accuracy. They also used SVM with radial basis function to classify

21 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/chapter/decision-support-system-for-diabetes-classification-using-data-mining-techniques/205129](http://www.igi-global.com/chapter/decision-support-system-for-diabetes-classification-using-data-mining-techniques/205129)

## Related Content

---

### The Impact of Word of Mouth on Brand Utility in the Healthcare Service Sector

Nail Burak Öztanand Gül Banu Dayanç Kyat (2023). *Handbook of Research on Quality and Competitiveness in the Healthcare Services Sector* (pp. 379-403).

[www.irma-international.org/chapter/the-impact-of-word-of-mouth-on-brand-utility-in-the-healthcare-service-sector/320861](http://www.irma-international.org/chapter/the-impact-of-word-of-mouth-on-brand-utility-in-the-healthcare-service-sector/320861)

### Resilient Vaccine Supply Chain: The Adoption of Cloud-Based Blockchain

Yui-yip Lau, Andrew Yang Wu, Yuk Ming Tangand Jessica Wong Lok Man (2023). *Revolutionizing Digital Healthcare Through Blockchain Technology Applications* (pp. 207-226).

[www.irma-international.org/chapter/resilient-vaccine-supply-chain/320973](http://www.irma-international.org/chapter/resilient-vaccine-supply-chain/320973)

### Colour Theory in Healthcare Corporate Identity

Inês Veiga Pereira, José Duarte Santosand Inês Nunes de Carvalho (2021). *Management and Marketing for Improved Competitiveness and Performance in the Healthcare Sector* (pp. 168-187).

[www.irma-international.org/chapter/colour-theory-in-healthcare-corporate-identity/285806](http://www.irma-international.org/chapter/colour-theory-in-healthcare-corporate-identity/285806)

### Sensorized Garments Developed for Remote Postural and Motor Rehabilitation

Giovanni Saggio, Valentina Sabatoand Roberto Mugavero (2015). *Healthcare Administration: Concepts, Methodologies, Tools, and Applications* (pp. 511-536).

[www.irma-international.org/chapter/sensorized-garments-developed-for-remote-postural-and-motor-rehabilitation/116232](http://www.irma-international.org/chapter/sensorized-garments-developed-for-remote-postural-and-motor-rehabilitation/116232)

### Mobile Device Application in Healthcare

Sandeep Lakkarajuand Santhosh Lakkaraju (2015). *Healthcare Administration: Concepts, Methodologies, Tools, and Applications* (pp. 640-663).

[www.irma-international.org/chapter/mobile-device-application-in-healthcare/116238](http://www.irma-international.org/chapter/mobile-device-application-in-healthcare/116238)