

## Application of Support Vector Machine for Diagnosis of Diabetes: A Systematic review

Ms. Suruchi Pimple  
Sadabai Raisonni Women's College, India

---

**Abstract:** The support vector machine is a training algorithm for learning classification and regression rules from data. The SVM is based on statistical learning theory and is a classification tool used with great accuracy and efficiency. The objective of this study is to conduct an empirical study on application of SVM in diagnosis of diabetes. This paper includes various research papers which have used SVM or Enhanced SVM for diagnosis of diabetes. The goal of the study was to identify research goals, diabetes type, data sets, software and technologies used and the outcomes.

**Keywords:** Data mining, Diabetes mellitus, Support Vector Machine.

---

### I. Introduction

#### 1.1 Diabetes

Diabetes mellitus, or simply diabetes, is a set of related diseases in which the body cannot regulate the amount of sugar in the blood [1]. It is a group of metabolic diseases in which a person has high blood sugar, either because the body does not produce enough insulin, or because cells do not respond to the insulin that is produced. This high blood sugar produces the classical symptoms of polyuria, polydipsia and polyphagia [2]. There are three main types of diabetes mellitus (DM). Type 1 DM results from the body's failure to produce insulin, and presently requires the person to inject insulin or wear an insulin pump. Type 2 DM results from insulin resistance, a condition in which cells fail to use insulin properly, sometimes combined with an absolute insulin deficiency. The third main form, gestational diabetes occurs when pregnant women without a previous diagnosis of diabetes develop a high blood glucose level. It may precede development of type 2 DM. As of 2000 it was estimated that 171 million people globally suffered from diabetes or 2.8% of the population. Type-2 diabetes is the most common type worldwide [3]. Figures for the year 2013 show that the 5 countries with the largest amount of people diagnosed with diabetes were China (98.4 million) India (65.1 million), US (24.4million), Brazil(11.0 million and Russia (10.9 million) [3].

Medical diagnosis is a field where large volume of data is generated continuously. Data mining tools can be used to analyze this data for extracting knowledge. Support Vector machine is an emerging tool which can be used successfully to classify the data representing diabetic and non-diabetic patients.

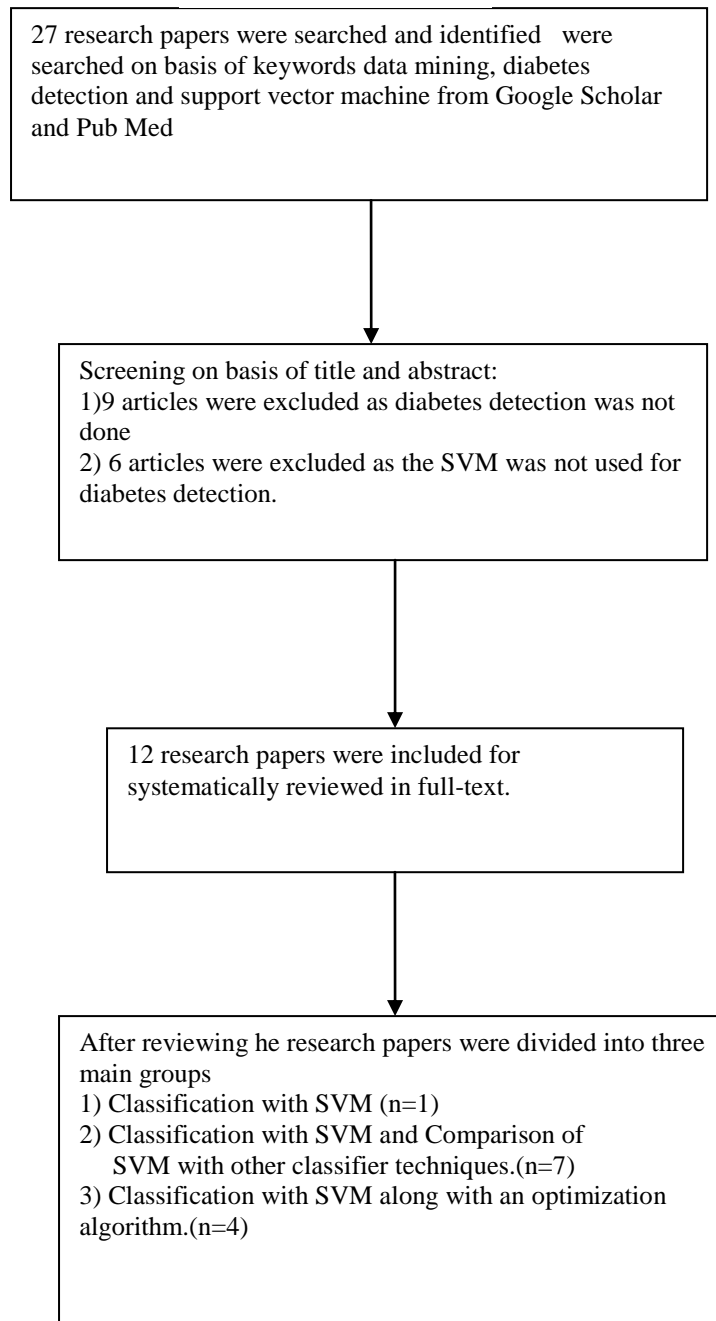
### II. Study selection and data extraction

The research papers were screened on the basis of following factors

- 1)Articles referring to SVM
- 2)Articles for Diabetes detection
- 3) Application of SVM for detection of diabetes.

Potential eligible papers were selected and reviewed. Information extracted from the article was extracted and represented in a tabular form. The table includes 1) research goals, 2) diabetes type,3) data sets,4) software and technologies used and 5) the outcomes.

**Study Flow:**



**Table**

Sr No.	Author name (year)	Research Topic/Goals	Method used	Data sets	Software and technologies used	The outcomes
1	Dr.V.Karthikeyani et.al (2013)	Comparison a Performance of Data Mining Algorithms (CPDMA) in Prediction	C4.5, SVM, K-NN, PNN, BLR, MLR, CRT, CS-CRT, PLS-DA and PLS-LDA	Pima Indian Diabetes Dataset	Tanagra tool	BootstrapValidationErrorrate= 0.2929 Accuracy= 70.6667
2	V. Anuja Kumari, R.Chitra (2013)	Classification Of Diabetes Disease Using Support Vector Machine	SVM(RBF kernel)	Pima Indian Diabetes Dataset	Matlab tool	Accuracy=78%
3	Longfei Han, Senlin Luo, Jianmin Yu, Limin Pan, and Songjing Chen (2015)	Rule Extraction From Support Vector Machines Using Ensemble Learning Approach: An Application for Diagnosis of Diabetes	SVM, RF , C4.5, NBTree, BP NN, SVM +C4.5, SVM +RF	Open dataset China Health and Nutrition Survey (CHNS)	Weka tool	SVM + RF has higher precision
4	Mustafa Serter Uzer, , Nihat Yilmaz, and Onur Inan (2013)	Feature Selection Method Based on Artificial Bee Colony Algorithm and Support Vector Machines for Medical Datasets Classification	SVM+ABC Algorithm	Pima Indian Diabetes Dataset	--	Classification accuracy 79.29% for diabetes dataset
5	Davar Giveki, Hamid Salimi, GholamReza Bahmanyar, Younes Khademian (2012)	Automatic Detection of Diabetes Diagnosis using Feature Weighted Support Vector Machines based on Mutual Information and Modified Cuckoo Search	MI-MCS-FWSVM	Pima Indian Diabetes Dataset	--	93.58% accuracy
6.	K. Rajesh, V. Sangeetha	Application of Data Mining Methods and	CS-RT, C4.5, ID3, K-	Pima Indian Diabetes Dataset	ROSETTA	Error rate for SVM =0.2253

		Techniques for Diabetes Diagnosis	NN, LDA, NAÏVEBAYES, PLS-DA, SVM			
7	Ankita Parashar Kavita Burse Kavita Rawat (2014)	Comparative Approach for Pima Indians Diabetes Diagnosis using LDA-Support Vector Machine and Feed Forward Neural Network	LDA-SVM FF-NN	Pima Indian Diabetes Dataset	WEKA	LDA-SVM Accuracy=75.65
8	Dr. D. Ashok Kumar#1 and R. Govindasamy (2015)	Performance and Evaluation of Classification Data Mining Techniques in Diabetes	SVM REGRESSION BAYES NET NAÏVE BAYES DECISION TABLE	Pima Indian Diabetes Dataset	WEKA	DECISION TABLE +Feature selection Accuracy = 79.81%.
9	Nahla Barakat1, Andrew P. Bradley1 and M. Nabil Barakat (2010)	Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus	SQRex_SVM <i>Eclectic</i>	National Survey of Diabetes data collected in the Sultanate of Oman	--	SVM Rule set accuracy 0.89  SQRex-SVM Rule set accuracy 0.94  Eclectic Rule set accuracy 0.93
10	Chi Yue1, Liu Xin2, Xia Kewen1, Su Chang1	An Intelligent Diagnosis to Type 2 Diabetes Based on QPSO Algorithm and WLS-SVM	QPSO Algorithm and WLS-SVM	Pima Indian Diabetes Dataset	--	WLS-SVM with RBF Kernel Diagnostic rate= 98%
11	Thirumalaimuthu Thirumalaiappan Ramanathan, Dharmendra Sharma (2015)	An SVM-Fuzzy Expert System Design For Diabetes Risk Classification	SVM-Fuzzy	Pima Indian Diabetes Dataset	MATLAB	Classification accuracy of 96%.
12	Omar S.Soliman, Eman AboElhand (2014)	Classification of Diabetes Mellitus using Modified Particle Swarm Optimization and Least Squares Support Vector Machine	PSO LS-SVM	Pima Indian Diabetes Dataset	--	<i>Classification accuracy of 97.833%.</i>

### **III. Conclusion**

Using data mining to deal with the avalanche of clinical data collected from patients and generated from the research and management of diabetes is a valuable asset that can help researchers and clinicians provide better health care for the patients affected by this modern-society disease. Modified Support Vector Machine along with various optimization techniques has shown promising results. This study will be helpful for constructing medical decision support systems in future

### **References**

- [1]. [http://www.emedicinehealth.com/diabetes\\_mellitus\\_type\\_1\\_and\\_type\\_2](http://www.emedicinehealth.com/diabetes_mellitus_type_1_and_type_2)
- [2]. [http://en.wikipedia.org/wiki/Diabetes\\_mellitus](http://en.wikipedia.org/wiki/Diabetes_mellitus).
- [3]. <https://www.idf.org>