Integrated Intelligent Research (IIR)
International Journal of Computing Algorithm
Volume: 04 Issue: 02 December 2015 Pages:54-58
ISSN: 2278-2397

Classification and Prediction of Heart Disease from Diabetes Patients using Hybrid Particle Swarm Optimization and Library Support Vector Machine Algorithm

C.Kalaiselvi[1], G.M.Nasira[2]
[1]Research Scholar, Karpagam University, Coimbatore,
Assistant professor and Head, Dept of Computer Applications, Tiruppur Kumaran College for Women, Tirupur, Tamilnadu
[2]Assistant Professor of Department of Computer Applications, Dept of computer science, Chikkanna, Govt Arts College, Tirupur,Tamilnadu
Email: kalaic29@gmail.com, nasiragm99@yahoo.com

Abstract—The multi-factorial chronicle, severe disease among human is diabetes. As a result of abnormal level of glucose in body leads to heart attack, kidney disease, renal failure, Hyperglycemia and also cancer in organs like liver and pancreas. Many studies have been proved that several types of heart diseases are possible in diabetic patients having a high blood sugar. Many approaches were proposed to diagnose both Diabetes and heart diseases. Most of the diabetes people can also have heart diseases called as Diabetic Cardiomyopathy. The earliest manifestation of diabetic cardiomyopathy is needed certain processes. The objective of the study is to examine the association of heart disease and diabetes. The relationship between diabetes and cardiovascular diseases are examined by taking into account of age, sex and associated diabetic and cardiovascular risk factors. The data are collected from patients with diabetes. From these data, features are selected by ant colony optimization and those selected features are given to hybrid PSO-LIBSVM to classify abnormal and normal data. This performance is evaluated using performance metrics and proved this classifiers efficiency for detection of Diabetic Cardiomyopathy.

Keywords-Diabetes Mellitus, Cardio vascular, Cholesterol, Prediction, Blood Pressure

I. INTRODUCTION

When the production of insulin is affected, which controls the glucose level of the body leads to a disease called diabetes mellitus. The diabetes mellitus causes nerve damages, kidney disease, and also cancer. Many researches are going on this area and some of the statistics report shows that most of the people will die because of the high glucose level lead to either cancer or heart diseases. Diabetes mellitus cannot be healed fully but it can be controlled by using medicines such as insulin, food items. Diabetes is a chronic disease that occurs from either when the pancreas does not produce enough insulin or completely stops secretion of insulin. Insulin is a hormone secreted by pancreas that regulates the body's blood sugar. The effects are uncontrolled and over time lead to serious damage to many human body systems, especially for the nerves and blood vessels. There are three types of diabetes [1]. Type 1 diabetes, Type 2 diabetes and Gestational diabetes. Type1 diabetes are mostly happens to children. Type2 diabetes is called adult-onset diabetes [2]. It is common in adults. Gestational diabetes is only in women during pregnancy. Many researches have been conducted to find the relationship between diabetes mellitus and heart diseases recently in diabetic patients. The diabetes is a group of metabolic disorders, and the reinterpretation is required, so, the possibilities of heart diseases in diabetic patients are high. The diabetic patients are affected by metabolic and hormonal disorders. Some of the risk factors of the heart diseases and diabetes are lack of physical activity, alcohol, smoking and obesity. The biological link mechanisms between heart diseases and diabetes are hyper-insulinemia, hyperglycemia, and inflammation. Many new techniques are used for diagnosing both heart diseases and diabetes. The diagnosis using machine learning techniques is one of the existing techniques which have a transparent diagnostic knowledge. Machine learning is classified into two types which are connectionist learning and symbolic learning. The user can easily understand the rules of symbolic learning techniques and are considered as comprehensible techniques.

The best example for the symbolic technique is rule induction which is extensively used for medical diagnosis High Blood sugar damages blood vessels can lead to blockage. People with diabetes have two to four times risk of developing heart disease. Blocked vessel in the legs can cause pain and also impair circulation. Two major types of heart and blood vessel disease also called as cardiovascular disease are Coronary Artery Disease (CAD) and cerebral vascular disease are common in people with diabetes. According to WHO 347 million people worldwide are having diabetes. More than 80% of people with diabetes are low and middle income groups. Diabetic Heart Disease (DHD) may have no signs or symptoms of heart disease. It is called silent heart disease. Two types of heart &blood vessel disease called as cardiovascular disease.
i) Coronary Artery disease.
ii) Cerebral vascular disease.

Some of the symptoms of heart failure are breathing problems, swelling in the ankles, feet, legs, abdomen, and veins in neck. People who have heart failure can live longer and more active lives if the condition is diagnosed earlier and if they follow their treatment plans regularly. People having diabetes can have a chance to develop heart disease and stroke [3-5].Diabetes causes blockage in blood vessels. People having diabetes are at risk of heart disease and stroke. A chance of heart disease is higher in diabetic patients. The risk factors of

International Journal of Computing Algorithm
Volume: 04 Issue: 02 December 2015 Pages: 54-58
ISSN: 2278-2397

heart disease are obesity, abnormal blood cholesterol levels, high blood pressure, smoking etc. High blood glucose levels may cause blockage of blood vessels result into irregularity of heartbeats. It may cause poor blood circulation in legs and feet, heart attack or stroke.The symptoms of heart disease are pain in chest, shoulder, arms, jaw, breath shortness, sweating, giddiness, nausea and light-headedness. These three major problems of coronary heart disease, high blood pressure and diabetes may weaken the heart. Then the body may release proteins and other substances into the blood.The parameters used to predict heart disease in diabetic patients arewaist circumference (40 & above in men)(35 & above in women)

- Elevated levels of triglycerides (>=150 mg/dl)
- Low levels of HDL cholesterol (below 40mg/dl in men)(below 50 mg/dl in women)
- LDL cholesterol: under 100 mg/dl
- Triglycerides: under 150 mg/dl
- DL cholesterol: for men above 40 mg/dl and for women above 50 mg/dl
- Blood Pressure>= 130mmHg for low blood pressure >= 85mmHg for high blood pressure
- HBA1C = below7 % (average blood glucose level for every 3 months once)
- Blood Glucose levels: Brandial and post Brandial blood sugar
- ECG

Data mining derived from the name as the similarities between searching for valuable information in large database. It is also called as knowledge discovery in databases. In computer science it is the process of discovering interesting and useful patterns. Classification is one of the data mining techniques used to predict group membership for data instances. For example, it can be applied in prediction [6]. Popular classification techniques include decision trees and neural networks.

II. RELATED WORK

ShishirMurarka [8] have proposed a new technique RAAS, it is a diabetes mellitus associated and increase the oxidative damage to endothelial cell apoptosis to the diabetic patients. It is a direct signalling via the angiotensin-1 receptors result in increased NADPH oxidise activity and elevation of ROS. The NADPH improves the performance compared with ROS. Loren E. Word [9] said that a Renin-Angiotensin System (RAS) is one of the oldest hormone systems that regulate blood pressure and water balance. It is a strong control system for salt conservation, blood volume and blood pressure. Threats occur from low blood volume and low blood pressure. Renin is an enzyme produced by kidney. The RAS compared with DOCA-salt the performance is lower than the sequence alignment. It is vascular damage by shifting per vascular.Ronald M. Wittele [10] has proposed a new technique on insulin resistant cardiomyopathy. It is increasing the insulin resistance from the etiologic factor in the development of non scheme Heart Failure (HF). It is a fundamental feature of the type II diabetes mellitus. Treat patients with diabetes mellitus also be useful for the Insulin-Resistant Cardiomyopathy (IRCM).

Xiaoyan Zhang [11] said that Tongxinluo for invention in Streptozotication (STZ) is used for diabetic ratings.. In this work can use the Accountable Care Organization (ACO) and then it is downloadable file. It includes information regarding ACO performance rates and Diabetes Mellitus (DM) and one Coronary Artery Disease (CAD). In this dataset is publicly available for the Pioneer Program ACO.Mohammed K. Ali [12] has proposed a new technique on Action to Control Cardiovascular Risk in Diabetes (ACCORD). ACCORD participants had diabetes on average of 10 years. Different levels of risk factors were taken depending on each patient's individual risk profile. George L. Bakris [13] have proposed an analysis of Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) subgroup with diabetes. It is a higher cardiovascular event rate. Then it is a diuretic group they have the greatest worsening of glycemic control. It is designed to validate and elucidate the significance of HF events in ALLHAT. The findings are physician diagnoses were 71%, 80%, and 84% for ALLHAT. It also improves the performance compared with physician diagnoses. M. Harris [14] proposes a new technique on risk assessment tool is designed to identify people at risk and the development of diabetes without the need for laboratory tests. The risk score included age, BMI, waist circumference, history of antihypertensive drug use or blood glucose level, physical activities, consumption of fruit and vegetables. A score of had a sensitivity of 77% and specificity of 66% and PPV of 7% are coherent. RAT is compared with PPV here the performance is 88%. Anderson KM [15] has proposed a new technique on Risk Assessment Algorithm. Cardiovascular risk assessment tools provide a measure of a patient's absolute cardiovascular risk.

Use the algorithm and the charts on the next pages to calculate absolute risk.Barbara [16] said that an Advanced glycation end products (AGEs) form when sugar is non-enzymatically linked with proteins, inducing cross-linking of the glycated proteins. AGE formation and biochemistry cellular receptors for AGE area signal transduction receptor. The AGE performance is compared with RAGE here the performance is 77% lower than RAGE. Alessandra Saldanha de [17] has proposed a new technique on Cardiovascular Autonomic Neuropathy (CAN). CAN is a very important complication of both type 1 and type 2 diabetes could be early diagnosed by some specific test and should be taken into account in all the cases, even the asymptomatic ones. It and during deep breath, resting heart rate, heart rate ratio as blood pressure monitoring and ECG registration and heart rate variability analysis there. Some parameters which are more frequently used: LF and HF (low & high frequency) reflects sympathetic activity and is lower than normal in CAN, HF reflects parasympathetic activity and LF/HF rate which reflects sympathetic.

III. PROPOSED METHODOLOGY

A. Efficient feature selection

Diabetes is one of the disease which affects heart and risk factor arises is called as Diabetic Cardiomyopathy (DCM) which affects myocardium in diabetic patients. The following dataset features are taken from diabetic patients.Age(yrs), Duration of diabetes(yrs), BMI(kg/m^2), Systolic blood

International Journal of Computing Algorithm
Volume: 04 Issue: 02 December 2015 Pages: 54-58
ISSN: 2278-2397

pressure(mmHg), Diastolic blood pressure(mmHg), VO_2max [ml/kg/min], Left ventricular mass(gm/m^2), Triglycerides(mmol/L), HbA1c(%), Insulin sensitivity (QUICKL), PINP(µg/L), PIIINP(µg/L), PICP(ng/mL), E/A RATIO, Mean S_m (cm/s), Mean Strain (%), Mean post – contrast T1 (value) (ms) and Backscatter (dB).These features are given for feature selection using ant colony optimization and classified using particle swarm optimization and library support vector machine.Feature selection is used to select a subset of relevant features for building robust and bestlearning models. These are used in machine learning and provide better understanding of the data by selecting important features within the data. The spectral analysis of heart rate is calculated from 24 hours. The structure and function of the myocardiums of diabetic rats was observed to ascertain the myocardial protective effect of the drug, to provide a theoretical basis for clinical treatment

B. Ant colony optimization

This Ant Colony Optimization is introduced by Marco Dorigo and his co-authors in the early 1990's [4, 5, and 6]. The development of these algorithms is done based on ant colonies. Ants are type of insects. They live in the form of colonies and their behaviour is survived by the goal of colony. The behaviour for ACO is the ant's tries to find shortest paths between food sources and its place. When searching for its food ants initially explore the area surrounding their nest in a random manner. Ants leave a chemical pheromone trail on the ground while moving. They smell pheromone and while choosing their way, they tend to choose, in probability, way is marked by strong pheromone concentrations. After finding the food source by ants, it evaluates the quantity and the quality of the food and carries some it back to the nest. While returning, the capacity of pheromone that an ant leaves on the ground may depend on the capacity and its flavour of the food. The other ants will follow the same pheromone to the food source. The ACO is successfully applied to optimization problems including data mining, telecommunication networks and in vehicle routing, etc. [1-3]. Especially the bioinformatics and biomedical fields show an increasing interest in ACO.

Based on the above figure 1 shows the ACO based feature selection, for each iteration this function updates the pheromone value. If an ant chooses c_i^j pheromone on this, assign more increments and ants should select c_i^j with higher probability in the next iteration. For every iteration, the objective function is updated using

$$\tau_i^j(t+1) = \rho.\tau_i^j(t) + \Delta\tau_i^j(t) + P_i^j(t) \quad [1]$$

Where,

$$\Delta\tau_i^j(t) = \frac{1}{|S_i^j|}\sum_{s\in S!} f(s) \quad [2]$$

$$P_i^j(t) = \begin{cases} Pc_i^j \in S_{best} \\ 0 \quad otherwise \end{cases} \quad [3]$$

The S_i^j is the set of generated solution for the t^{th} iteration when passes through c_i^j. The S_{best} is the best solution obtained and P is a constant.

a) The fitness function

Based on the ant's solution for the selected feature, the solution obtained in terms of classification accuracy is evaluated by classifying the data sets using the selected features. Then the test accuracy is calculated by correctly classified result. In addition, the number of features in the set is also considered in the quality function. The function $f(s)$ of a solution s is defined as follows:

$$f(s) = \frac{N_c}{1 + \lambda N_f}$$

Where, $N_c$is the number of data correctly classified and N_f is the number of features selected in s, λ is the constant to adjust the accuracy and number of features selected.

Initialize the dataset.
Load the features.
Generate a population of Ants.
Initialize the ACO parameters.
FOR each Ant, Generate a subset S.
Evaluate each subset S.
IF the fitness valuebetter than previous global best value.
Set the current subset's accuracy as global best accuracy.
END IF
Update the pheromone values.
Repeat the process until criteria met.
END FOR
Report best feature subset as final Appropriate set.
END

The above algorithm explains the overall flow of the feature selection for this Diabetic Cardiomyopathy data.

C. Classification using Particle Swarm Optimization

The classification of abnormality and normality is done here by using PSO with LSVM technique.Particle Swarm Optimization (PSO) is one of the population-based optimization algorithms modelledusing the simulation of social behaviour of birds in a flock [7]. This PSO algorithm is initialized with particles and it searches optimal value by updating generations. Each particle is flown a group of random through the search space where the position is adjusted based on its distance from its own personal best position and the distance from the best particle of the swarm. The performance of each particle is measured using a fitness function that depends on the optimization problem.Each particle iin an n-dimensional search space, R^n, and maintain the following function:

x_i, the current position of i^{th} particle (x-vector)
p_i, the personal best position of i^{th} particle (p-vector), and
v_i, the current velocity of i^{th} particle (v-vector).

The best position associated with a particle, i, is the best position that the particle has visited so far. If f denotes the fitness function, then the personal best of i at a time step t is updated as:

$$p_i(t+1) = \begin{cases} p_i(t) \text{ if } f(x_i(t+1)) \geq f(p_i(t)) \\ x_i(t+1) \text{ if } f(x_i(t+1)) < f(p_i(t)) \end{cases}$$

If the position of the global best particle is denoted by gbest, then:

$$gbest \in \{p_1(t), p_1(t), \ldots\ldots, p_m(t)\} = \min\{f(p_1(t)), f(p_2(t)), \ldots\ldots f(p_m(t))\}$$

International Journal of Computing Algorithm
Volume: 04 Issue: 02 December 2015 Pages: 54-58
ISSN: 2278-2397

The velocity updates are calculated as a linear combination of position and velocity vectors. Thus, the velocity of particle i is updated and the position of particle i is updated by the following equations.

$v_i(t+1) = w.v_i(t) + c_1 r_1\big(p_i(t) - x_i(t)\big) + c_1 r_1\big(gbest - x_i(t)\big) x_i(t+1) = x_i(t) + v_i(t+1)$

In the formula, w is the inertia weight [8], c_1 and c_2 are the acceleration constants, r_1 and r_2 are random numbers in the range $[0, 1]$ and V_{i1} must be in the range $[-V_{max}, V_{max}]$, where V_{max} is the maximum velocity.

D. Library Support Vector Machine

LIBSVM is a library for Support Vector Machines (SVMs). The goal is to easily apply SVM to their applications. LIBSVM has gained wide popularity in machine learning and many other areas. In this work, we present implementation of LIBSVM Issues such as solving SVM optimization problems theoretical convergence multiclass classification probability estimates and parameter selection.A typical use of LIBSVM involves two steps: first, training a data set to obtain a model and second, using the model to predict information of a testing data set. For SVC and SVR, LIBSVM can also output probability estimates.This is same as SVM technique, where in training SVM the An m by 1 vector of training labels (type must be double) is taken.Then parameters for gamma in LIBSVM are taken from PSO algorithm. And the Cost parameter is set as the parameter C of C-SVC is taken.

E. Kernels

Kernel methods in general have gained increased due to the grown of popularity of the Support Vector Machines. Support Vector Machines are linear classifiers and regressors that, through the Kernel trick, operate in reproducing Kernel Hilbert spaces and are thus able to perform non-linear classification and regression in their input space.

Here Radial Bias Function Kernel is used and it is expressed as

$$RBF = \exp\left(\frac{1}{2\sigma^2 \|x - x_i\|^2}\right)$$

By this the classification of abnormal and normal Diabetic Cardiomyopathy is trained for few dataset and tested for various patients easily and this technique is proved by its performance metrics.

IV. SAMPLING DATA AND SIMULATION RESULTS

The effectiveness of models was tested using different methods. Confusion matrix is one of this. The purpose of this is to determine which model gives the highest percentage of correct predictions for diagnosing diabetic patients with heart disease. This work was designed as a prospective, single centre study. Totally the patients 520 men and 375 women of 40 to70 years old were taken at baseline with 3 years follow-up. The records were split into two classes as 400 records for training set and 120 records for testing dataset. The attribute class value 1 for patients with heart disease and value 0 for patients with no heart disease. If 50% of test dataset is processed, 48% of test dataset is predicted correctly.Finally comparison of results with other methods such asnaïve bayes, k-nearest neighbour can be done by checking the confidence interval with others confidence interval.Of these datasets 30% of men and 20% of women are under risk of having heart disease.

V. CONCLUSION

The Algorithm was developed using categorical variables. Symptoms of heart disease may differ in both men and women. This work was designed as a prospective, single centre study. Totally the patients 520 men and 375 women of 40 to70 years old were taken at baseline with 3 years follow-up. 30% of men and 20% of women are under risk of having heart disease. 1 in 3 women dies because of heart disease each year. 90% of both men and women may have risk factor for developing heart disease when they are having diabetes.

REFERENCES

[1] M. CasacciaBertoluci and A. Pimazoni "Diabetes and cardiovascular diease: from evidence to clinical practice-postion statement 2014 of Brazilian diabetes society", Diabetology and Metabolic syndrome, 6:58, 2-18, 2014.

[2] Duckworth W and Abraira, "VADT Investigators: Glucose Control and VasculaComplications in Veterans with Type 2 Diabetes", N Engl J Med 360(2):129–139, 2009.

[3] Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. JAMA 2002; 287:356 –9.

[4] Lakka HM, Laaksonen DE, Lakka TA, et al. The metabolic syndrome and total and cardiovascular disease mortality in middleaged men. JAMA, 288:2709, 16, 2002.

[5] Lloyd-Jones DM, Larson MG, Leip EP, et al. Lifetime risk for developing congestive heart failure: the Framingham Heart Study. Circulation 106:3068, 72, 2002.

[6] From AM, Leibson CL, Bursi F, et al. Diabetes in heart failure: prevalence and impact on outcome in the population. Am J Med, 119:591–9, 2006.

[7] AllamAppaRao Croft JB, Giles WH, Rashidee A. Hospitalization of patients with heart failure: National Hospital Discharge Survey, 137:352, 60, 2009.

[8] ShishirMurarka and Line iden Berge "Prevalence and characteristics of depressive disorder in type 1 diabetes", BMC Research Notes, 6:543, 2-5, 2013.

[9] Loren E wold "Depressive Disorders among Michigan Adults by Diabetes Status", depression and diabetes among Michigan adult, 6, 1-2, 2012.

[10] Ronald M. Witteles "Insulin resistant cardiomyopathy and heart failure" Journal of American College of cardiomyopathy, vol.2, 94-100, 2008.

[11] Xiaoyan Zhang and Richard E "Heart Failure in SAVOR-TIMI 53: The hindsight of diabetic retinopathy" Keenan research center Journal of diabetes, 2-10, 2014.

[12] Mohammed K. Ali "Diabetes & coronary heart disease: Current perspectives" Indian journal of medical research, 132(5), 584–597, 2010.

[13] George L. Bakris "Preclinical diabetic cardiomyopathy: prevalence, screening and outcome" Internal Medicine University Department, 1-27, 2009.

[14] M.Harris "The role of primary health care in preventing the onset of chronic disease, with a particular focus on the lifestyle risk factors of obesity, tobacco and alcohol" Centre for Primary Health Care and Equity, UNSW, 1-21, 2008.

[15] Anderson KM ". Correlation of regional cardiovascular disease mortality in India with lifestyle and nutritional factors. Int J Cardiol 2006; 108: 291-300.

[16] Barbara "Type 2 diabetes in Asian Indian youth. Pediatr Diabetes 2007; 8 (Suppl 9): 28-34.32.

[17] Alessandra Saldanha de "Impact of Diabetes on Cardiovascular Disease: An Update", International Journal of Hypertension, 653789, 1-15, 2013.

[18] G. Di Caro and M. Dorigo. "AntNet: Distributed stigmergetic control for communications networks". Journal of Artificial Intelligence Research, 9:317–365, 1998.

International Journal of Computing Algorithm
Volume: 04 Issue: 02 December 2015 Pages: 54-58
ISSN: 2278-2397

[19] R.S. Parpinelli; H.S. Lopes; A.A. Freitas, "Data mining with an ant colony optimization algorithm", IEEE Transactions on Evolutionary Computation, 6: 321 - 332 2002.
[20] G. Di Caro and M. Dorigo. "AntNet: Distributed stigmergetic control for communications networks". Journal of Artificial Intelligence Research, 9:317–365, 1998.
[21] Dorigo M, Gambardella LM. Ant colony system: A cooperative learning approach to the traveling salesman problem. IEEE Trans Evolutionary Comput 1997;1(1):53–66.
[22] Dorigo M, Maniezzo V, Colorni A, Positive feedback as a search strategy. Technical Report 91-016, Dipartimento di Elettronica, Politecnico di Milano, Italy, 1991.
[23] Dorigo M, Maniezzo V, Colorni A. Ant System: Optimization by a colony of cooperating agents. IEEE Trans Syst Man Cybernet Part B 1996;26(1):29–41.
[24] J. Kennedy, and R. Eberhart, Swarm Intelligence, San Francisco: Morgan Kaufmann Publishers, 2001.
[25] Y. Shi, and R. Eberhart, "A modified particle swarm optimizer," in the Proceedings of the IEEE International Conference on Evolutionary Computation, Piscataway, NJ, pp. 69-73, 1998.Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin," A Practical Guide to Support Vector Classification", April 15, 2010.

AUTHORS AND AFFILIATIONS

B.Kalaiselvi received her M.C.A., degree from Mother Teresa Womens University and M.Phil degree from Manonmaniam Sundaranar University. She is having 16 years of teaching experience and presently working as Associate professor& Head in the dept. of Computer Applications, Tiruppur Kumaran college for women,Tirupur. She is pursuing her PhD at Karpagam University, Coimbatore. Her area of interest includes Data Mining, Networks and communication security, Software engineering.

Dr. G.M. Nasira is working as Assistant Professor of Computer Application, Department of Computer Science, Chikkanna Government Arts College, Tiruppur, Tamilnadu, India. She has around 18 years of teaching experience. Her Areas of Interest include Artificial Neural Networks, Data mining, Optimization, and Soft Computing. She has 102 publications and organized 21 Seminars / Conferences, Workshop / FDP / Orientation. She has been profiled in various organizations her academic contributions as member of board of studies. She is life member and senior member of ISTE, IACSIT etc. She is the recognized trainer associate for UGC's capacity building programs. She received best faculty award from various institutions.