A Novel Method for Identification of Cardio Vascular Disease using KELM Optimized by Grey Wolf Algorithm

S. Sharmila, M.P. Indra Gandhi

Abstract--Timely discovery of the presence of cardiovascular disease can be the difference between life and death. There has been great importance in the construction of processing tools for prognosis and diagnosis of cardiac disease and, especially, cardio vascular events. Classifying data is a customary duty of machine learning. Data mining in health care is a forthcoming arena that attains huge importance for delivering prognosis and a profound realization of medical data. The usage of SVM dependent methodologies in identification the cardio vascular diseases has some restrictions. The important drawbacks of SVM is the severe absence of transparency of outcomes. The ELM learning algorithm is a simple process and it provides accurate result when compared to other traditional algorithms. As the proposal of this research, to enhance the generalization capacity of ELM, KELM is utilized. To improve the classification accuracy of KELM, in this paper a nature inspired swarm intelligence Grey Wolf Algorithm is utilized. Grey Wolf Algorithm is utilized in optimizing the parameter of KELM. By performing classification, accuracy is improved along with high precision and low error rate. Experimental results clearly indicate that the proposed GWO – KELM classifier performs better on comparison with some classifiers that are currently used for the identification of the Cardio Vascular Disease.

Keywords: Cardio Vascular Disease, Machine Learning, Data mining, Kernel Extreme Learning Machine, Grey Wolf Optimizer.

I. INTRODUCTION

KNOWLEDGEunearthing in databases is well-explained procedure comprising of various discrete steps. Data mining is the fundamental phase, which leads in the identification of hidden but fruitful information from enormous databases. A proper explanation of Knowledge identification in databases is given as follows: “Data mining is the non-trivial abstraction of implied before hand unidentified and possibly valuable material about data” [1]. Data mining strategy offers a customer-based strategy to unique and hidden patterns in the data. The latest development in the data mining strategies has offered a stage to abundant applications in healthcare segment. It has developed an active research field because of its huge scale capability in the area of Disease identifying and determining its obtaining strategies. It also put up the investigators in the arena of healthcare in expansion of efficient guidelines, and diverse framework to protect diverse kinds of disease. The data needed concerning such systems can be the particulars of the patients, hospitals, diseases and their treatments. Data mining is very much supportive for the investigation of different parameters, which are accountable for diseases distribution around, like working atmosphere, existing circumstances, food quality, and obtain ability of clean water, health facilities and many others [2].

Precisenesstakes a crucialpart in the medical arena of Cardiology as itsapprehensions with the life of aspecific human being. From this time, it’s very significant and critical while taking decisions, which embraces both the past knowledge and the present circumstances. Data mining in the medical fieldfunctions on the past knowledge (data acquired) and investigate them to determine the over-alltendencies and possibleexplanations to the contemporarycircumstances. Hereafterpossessing this concept in mind, the crucial nature of the arenaisselected for investigating the several mining methodologies and lastlya conclusion is arrived that considering into a real-lifesituation one can’t envisage the preciseexplanation or the action to be followed for the identification of the patient on the foundation of machine createdforecasted values from the previousoutcomes. Since the whole thingfunctions on the likelihood, should even reflect the mined consequences to be likely and not the final. [3]

Artificial Intelligence (AI) has modified fundamentalcharacteristics of human life. Machine Learning (ML), which is a subdivision of AI wherein machines individuallygather information by mining patterns from huge databases, has been progressively utilized within the medical community,and particularly within the arena of cardiovascular diseases. Disease discovery and identification of sicknesses is at the leading stage of Machine Learning research in medicine. Machine learning and processingclevernessmethodologies have large amount ofcapability to modify healthcare by offering objective conclusionhelp tools to contribute medical professionals in identification and treatment of patient circumstances.Machine Learning (ML), an elongation of the century-long quest for Artificial Intelligence (AI), has modifiedgatheredcommencement of information and its apparentlyunlimitedcapacity for managerial change. Machine learning is widely defined as the capacity of a system to individually gain knowledge by mining patterns from huge data sets. Prepared with innovative ML frameworks, swelling in theprocessing power and the obtainability of big data, the ML community is now focusing its efforts directly at complicated tasks in the healthcare sector. More recently, in clinical

Revised Manuscript Received on July 10, 2019.

S. Sharmila: Research scholar, Department of Computer Science, Mother Teresa Womens University, Kodaikanal, Tamilnadu, India.

M.P. Indra Gandhi: Assistant Professor, Department of Computer Science, Mother Teresa Womens University, Kodaikanal, Tamilnadu, India

cardiology, it has been demonstrated that ML is more skillful in the forecastingof either cardiovascular or all-cause death than clinical or imaging modalities utilized separately. [4]

Cardio Vascular disease requiredutmost support of machine learning to identify Cardio Vascular disease in initial stage, because it cannot be cured and also provides great complexity to our health system. The precise diagnosis of diseases with high occurrence rate such as CVD, Cancer etc. is one of the most significant biomedical challenges whose management is authoritative. Soas to identify positive indications of CVD, medical specialists recommendingseveral tests such as angiography, nuclear scan, and C-reactive protein test which are moderatelycostly and needs technical specialists therefore, investigators are looking forattention to construct a less expensive and an efficient substitute to the costly prescribed test.

For the purpose of minimizing the involvement by clinical along with specialist's effort is along with conserving the time resulting in patient's lives and cost, smart systems which are involved in prediction of Cardio Vascular disease was constructed utilizing machine learning techniques along with data mining strategies was established.[5]. Suitably, many ML techniques are adequatelyadaptable to achieve both types of analysis with only smaller adjustments, although limitations such as interpretability, processing cost, and type of obtainable data required to be measured in tailoring the selection of technique. Datasets utilized in ML projects are characteristicallysegregated into learning, authentication, and test subsets: the training set, which comprised the bulk of all obtainable data, is utilized for the principal growth of the model; the authentication set is utilized to assess complete model functionality or to fine-tune its hyper parameters. It's clearly evident that machine learning puts another arrow in the tremble of clinical conclusion making. Still, machine learning advances itself to some procedures better than others. Also, those with large image dataset, such as radiology, cardiology, and pathology, are robust contenders[6].

With an investigational platform and machine learning executing in the contextual, the human algorithm the extra layer of a back-up physician wouldn't be necessary. The analytics engine would have substantially more data than any one person could ever compute. It would have treatment choices available with identifications of how long they would be efficient, mortality rates, side effects, and cost. Irrespective of all the exertion by a human caregiver, an investigational platform could put in substantially more work behind the sections and provideconclusive information to the doctor in real time.

The remainder of this paper is organized as succeeding: Section II Presents the Survey of various researches done by the various scholars in the area of Automatic prediction of cardio vascular disease using classification Section III establishes the statement of the Problem Section IV explains the proposed strategy, Evaluation Results of the proposed technique are described in Section V. Lastly, Conclusion of the paper will be provided in Section VI.

II. LITERATURE REVIEW

Healthcare systems delivertailored services in extensive spread domains to assist patients in fitting themselves into their regular activities of life.Cardio Vascular Disease is one of the most significantthoughtfulencounters in the medical field,since most of the time is spent in diagnosing the disease. Classification is one of the most significantpolicymakingprocedures in many real-world difficulties.The main motivation is to categorize the data as Heart Patient or non-Heart Patient and enhance the classification accuracy. Machine learning for identification of Cardio Vascular Disease, is about learning structures from the Cardio Vascular Disease dataset which is delivered. Machine learning in current years have been the developing, dependable and supportive tool in medical domain. This research is attentive on the prediction of CVD types of patients reliant on their personal and clinical information utilizingmachine-learning classifiers. In this chapter, summarization of the works projected by various researches in the past decade is done. It is useful to identify the limitations in the works which are proposed in the area of Cardio Vascular disease diagnosis-based machine learning classifiers. The problem in diagnosis of Cardio Vascular is a dynamic area in the research field.

Ajad*et al (2017)* [7]pronounced about a model utilizing data mining techniques mainly Naïve Bayes and WAC (Weighted Associated Classifier). The Naive Bayes model is an outdatedstrategy for classification and predictor choice that is enjoying a renaissance because of its straightforwardness and steadiness. WAC is also utilized to diagnosis the heart disease. Weighted Associative Classifier (WAC) is a new notion that utilized Weighted Association Rule for classification. Weighted ARM utilized Weighted Support and Confidence structure to segment Association rule from data repository. The WAC has been projected as a novelProcedure to obtain the significant rule instead of flooded with irrelevant relation. The Mainstages in WAC method was projectedevidently. The projectedeffort can be additionallyimproved and prolonged for the automation of Heart disease identification. Real data from Health care organizations and agencies requires to be gathered and all the obtainableprocedures will be associated for the optimum correctness. They established an Automotive and efficient heart disease prediction techniques utilizing data mining. Medical diagnosis is measured as animportant yet complicated task that requires to be performed outaccurately and proficiently. To identify the heart disease Naïve Bayes, Decision Tree, Regression algorithm were the efficient one and deliversoutcomeproficiently.

For protecting the heart from failure and providing the timely treatment, identification of disease in heart without any error at the primitive stage itself is the most obligatory task for increasing the survival chance of patients.By utilizing past history, detection of cardio vascular disease was not involved in consideration of various facts which general physician recommend for tests by adopting conservative methods.Strategies that relies with machine learning techniques falls under the

category of noninvasive approacheswill be considered as competent strategy for identification of cardio vascular diseases in healthy people.Construction of machine learning dependent identification with the help of dataset that contains the information of heart disease was suggested by Haqet al., (2018), [8]For the purpose of providing the assurance of validity utilization of famous machine learning approaches, strategies involved in choice of features,procedures involved in cross -authentication, along with functional assessment pertaining to classifiers by utilizing measures like classification exactness, specificity, sensitivity, Matthews' correlation coefficient, along with execution time. Identification of heart diseasein normal persons utilizing the machine learning technique will be involved in less complicated procedure.Computation with respect to area under for under the curves pertaining to optimization characteristic of receiver was performed.Classifiers, feature selection strategies, preprocessingtechnique, authentication approach, and metric for functionalityassessment of classifiers were utilized.Functionality belongs to plannedstructurewas authenticatedwith comprehensive features along with the functionality belongs to abridgedgroup of features was related. The features minimizationmight have an influencewith respect to classifiersoperation in terms of accuracy and implementation time of classifiers.

Progressive data mining techniques can be utilized to identify hidden pattern in data. These procedures will be useful for medical practitioners to consider efficient decision. In the paper projected by Shylaja and Muralitharan, (2018) [9] data mining classification techniques RIPPER classifier, Decision tree, Artificial Neural Network (ANN), Naive Bayes, Support Vector Machine (SVM), are examined on heart disease dataset and its competence is investigated. Performance of these practices is associated through sensitivity, specificity, Accuracy, true positive Rate and False Positive Rate.

In the paperprojected by Yahyaoui, and Yumusak, (2018), [10]they utilized the Support Vector Machines (SVM) process to identify chest diseases and for the first time, theyscrutinize the functionality of the Adaptive Support Vector Machine (ASVM) technique for chest disease diagnosis. This comprisesenhancing the SVM by determining its most suitability Bias term value. These methods are appraised utilizing an investigational dataset from Diyarbakir chest sicknesses hospital and associating them with the Neural Network method utilized in preceding analysis. The investigationaloutcomesexhibited the competence of these approaches, particularly ASVM, which could accomplishtalentedoutcomes and long-established that it, can be effectivelyutilized in chest diseases.

For identification of outcomes pertaining to Stroke with the help of utilizing Least-Squares Support Vector Machines (LSSVMs) was established by Sarihan, M. E., & Hanbay, D. (2017) [11]. Medical dataset associatedwith stroke was accomplishedwith clinical database that belongs to department of emergency in medicine.Providing the authentication with respect to vague data might be performed by 28 forecasters. For the purpose of minimizing the dimension, correlations among input and target (stroke) variables were considered for evaluation. Diverse LS-SVMs prototypes were investigated with Radial Basis Function (RBF), proportional along with polynomial kernels. Cross authentication strategy was utilized in comprisingphases to accomplish the best model by means of all of the data. Parameters selected for conducting the investigation of functionality will be correctness along with Area with respect to Receiver Operating Curve. Initially, opting the feature was accomplished. Subsequently, determination of 14 input variables was performed. LS-SVMs prototypemight be involved in identification of consequence of stroke. Outcomes demonstrated that LS-SVMs with proportional kernel containedadditionalmeticulousness along with values for forecasting stroke.

However, ECG feature abstraction is obligatory and this may minimize diagnosis exactness in conservative shallow learning models, whereasBackward Propagation (BP) algorithm utilized by the outdated deep learning models has the shortcomings of local minimization and slow convergence rate. To determine solutions these complications, a new deep learning algorithm termedDeep Kernel Extreme Learning Machine (DKELM) is projected byLi (2018) [12]merging the extreme learning machine auto-encoder (ELM-AE) and kernel ELM (KELM). In the new-fangled DKELM framework with hidden layers, ELM-AEs are utilized by the visible M hidden layers for feature extraction in the unsupervised learning process, which can effectively extract abstract features from the original ECG signal. To alleviate the \dimension disaster" problem, the kernel function is established into ELM to act as classifier by the hidden layer in the supervised learning process. Though the accuracy of Convolutional Neural Network (CNN) is nearly the similar as DKELM, the processing time of CNN is much longer than DKELM.

The motivation of theanalysis byLiu et al., (2017) [13]is to support the identification of heart disease utilizing a hybrid classification system depending on the ReliefF and Rough Set (RFRS) technique. The projectedarrangementcomprises two subsystems: the RFRS feature choosing system and a classification system with an ensemble classifier. The first organizationcomprises three phases: (i) data discretization, (ii) feature extraction using the ReliefF algorithm, and (iii) feature minimizationutilizing the heuristic Rough Set minimization algorithm that theyconstructed. In the second scheme, an ensemble classifier is projecteddepending on the C4.5 classifiers. The Statlog (Heart) dataset, attained from the UCI database, was utilized for investigations. A maximum classification accuracy of 92.59% was accomplished as per to a jackknife cross-authenticationarrangement. The outcomesestablish that the functionality of the projectedarrangement is greater to the performances of beforehanddescribed classification procedures.

In the research work by Kalaivani and Uma (2017) [14] the imbalanced data categorizationissue could be addressed in a methodicalapproach by enforcingimproved stacking with cuckoo optimization which supports to quickens the functionality of learning algorithms. Also, in this investigational work

theyprojected anidentificationarrangement to classify heart disease utilizing cuckoo search optimization to minimize feature with altered stacking also fuzzy logic system is utilized to identify the heart disease. They demonstrate that ensemble classifier is more accurate than any single constituent classifiers. Originally the data composed are imbalanced. Theystabilize the data and make it better forecasting. To identify Ischemic Heart Disease in more precisethey project a novel ensemble assorted classifier buildingprocedures with improved stacking to engender domain-specific configurations. Theydeliberate both the field ensemble learning and disseminated data mining to generate an effective and improvedvariety of the customary stacking ensemble learning procedures by utilizing fuzzy logic and cuckoo optimization producing the Meta classifier.

For the purpose of identification of heart disease utilization of Extreme Learning Machine optimized by Dragon Fly approach was established by Salam et al., [2016] [15]. (ELM) EXTREME Learning Machine approachcontains the training strategy that will be extremely rapid.On the other hand, requirement of huge quantity of computing element with respect to concealed layer was observed in Extreme Learning Machines. Improvement in processing duration time was noticed by utilization of huge quantity of processing elements.As a consequence of utilization of utilization huge quantity of processing elements, the strategy involved might be devoid of providing the assurance of weights along with biases settings with respect to concealed layer. Latest fascinating strategy involved in copying the navigating conduct belongs to moths was evolved and termed as Dragon fly algorithm. For the purpose of reducingcomputing elementsutilized in concealed layer belongs to extreme learning machine utilization of Dragon Fly approach which might be involved in choosing least quantity of processing elements with respect to concealed layer for speeding up classification operation belongs to Extreme Learning Machine.Dragon fly strategy might be utilizedfor choosingbest possible weights along with biases associated with concealed layer.Group of valuationpointers utilizedin assessingsuggested along withassociatedprocedures over data sets that might be related with the procedure of regression that was obtained from UCI repository. Results provide the assurance of capability pertaining to suggestedHybrid of Dragon fly -Extreme Learning Machineprototype in probing for selecting the best feature combinations that might be involved in improving the simplification capability belongs to Extreme Learning Machine along withidentification accuracy.

PROBLEM STATEMENT

Cardio Vascular Disease will beforemostreason of non-communicabledisease and death will beestimated to be the principalreason of death worldwide by 2030 [WHO -I]. Majority of prevailinginvestigationexertionswas intendedin the direction of the identification, alteration and treatment of individual-level risk factors. In spite ofprominentprogressions happened with medical industry, gross inequalities endure to persevere over space along withduration. Thoughcumulatingof diseases were observed at diverse rates universally, the strength of swellforoccurrence of numerous cardiovascular dangerinfluenceswasfocused in altering investigationexertionsforinvestigating reasonsinvolved in producing the disease, which comprise the community determinants pertaining to health of an individual. Requirement of classifier that might be involved in automatically identification of Cardio Vascular diseases in a precise manner will be considered as most obligatory. Accurate identification along withexplanationof information will provide the considerable quantity of challenge in performing the process of categorization.Degree of disease along with identification during primitive stage will be considered as significant in developing fast harmlessperiodsin the direction ofcircumventingrisk. Classification of dataset will provide the considerable amount of complexity especially in the field of medicine.With respect to majority of circumstances functionalitybelongs to technique falls back in terms of rapidity or correctness or both. To enhance the correctness in identification along with training with data in a rapid manner, numerous procedures were utilized. Byexaminingentireapproaches,a competent hybrid classification procedure utilizing Extreme Learning Machine Learning optimized with the help of Grey Wolf Algorithm is constructedin identification of Cardio Vascular Disease.

III. PROPOSED METHODOLOGY

Main influenceprovided in suggestedresearch workcanbesummarizedassucceeds: Fresh nature-inspired technique, that operates with the principle of meta-heuristics termed as Grey Wolf Optimizerwill be successfully utilizedinfindingbest possible parameters for utilization of Kernel Extreme Learning Machine in classification. Potential prototype, that combines both Grey Wolf Optimization along with Kernel Extreme Learning Machine, will be successfully enforced in identificationofCardio vascular disease with the view of providing the capablepreliminarycautionary tool for cardiovascular disease infield of diagnosis.Suggested GWO-KELM will be formulated with the aim of accomplishingimproved classification, along withproduceadditionalsteadiness in addition with robust outcomes in differentiating the cardiovascular disease from healthy ones when compared withvariousadditionalapproaches.

A famous machine learning strategy which finds its application in various areas is Extreme learning machine [16]. Establishment of Arbitrary generation of input weight in addition with bias terms corresponding to concealed layeraccording to quantity of neurons associated with input in addition with processing elements belongs to concealed layer along with calculation weight matrix pertaining to output with respect to Moore-Penrose generalized inverse criteria will be performed. Although extreme learning machine approachcontains numerousadvantages when compared with conservative neural networks, strategy of extreme learning machine resulted in instability due to the utilization of arbitrary input weight along with bias terms related with concealed layer.With the motivation of accomplishing higher preciseness, theworksuggestsstrategy which will be involved in providing the optimization toExtreme Learning Machine by means

of optimization strategy termed asGrey Wolf Optimization [17].

A. Grey Wolf Algorithm

Grey Wolf Algorithm established by Mirjalili et al. (2014) [18] falls under the category of fresh swarmintelligence-based approach that will be involved in determining the best solutions for the provided challenges.Imaginative approach will be modelling the community ranking along with along withtarget findingconduct ofgreywolvesinwildlife.Fordemonstratingthesocialhierarchy behaviorofgreywolf,thegroupisseparatedintofourparts: alpha (α), beta(β), delta(δ), andomega(ω) as shown in Fig. 1. α is measured tobethebest appropriate solution succeeded by β and δ, correspondingly,along withrestof answers will be containing the place to ω.Initialsuitableanimalwhich will be nearby totargetare α, β along with δ who leads ω inexploration of targetwith respect to hopeful space for exploration.While performing the process of surrounding the target, mentioned animals will be involved in improvingthe location surrounding α, β, or δ. The strategy is established in expressions provided below.

$$\vec{D} = |\vec{C}\vec{X}_p(t) - \vec{X}(t)| \quad (1)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A}\vec{D} \quad (2)$$

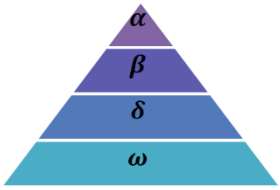

Fig 1Arragnement of Grey Wolves

while t will bepresent step$\vec{X}_p(t)$ marks the present location of target along with$\vec{X}(t)$ refers the present location belongs to mentioned animal. $\vec{D}$ will be distance amongmentioned animals along withtarget, in addition with coefficient vectors $\vec{A}$ *and* $\vec{C}$will be mathematically articulated as provided in the following expression

AUTHORS PROFILE

Dr. Indra Gandhi M P completed her Master's degree in Computer Application from Mother Teresa Women's University, Kodaikanal and Ph.D. in computer science from Mother Teresa Women's University, Kodaikanal. Shee joined the Computer Science Department at MotherTeresa Women's University University in 1999 and is currently an Assisstant professor in the department. Her research interests include Image Processing, Data Mining, Image Registration, Software Engineering, Biometric, Computer Networks. She has published over 30 papers in refereed journals and conference proceedings. She is a member of the Advisory Committee Sri Changara College, Kodaikanal and Doctorial Committee at Gandhigram Rural Institute, Gandhigram. She is also a lifemember of CSI.

Shamila. S has completed her graduation in Computer Science from Mother Teresa Women's University, Kodaikanal. and master's degree in Computer Applications Mother Teresa Women's University, Kodaikanal. She joined as Assitant Professor in the department of Computer Science at Govt Arts College Ramanathapuram, She is also a research scholar at Mother Teresa Women's University, Kodaikanal. Her research interests include Digital Image Processing and Machine Vision.