

# Heart Disease Prediction Using Bio Inspired Algorithms

<sup>1</sup>Mr.S.Asalam Shareef

<sup>2</sup> MULA RAVALI (183T1A0568)

<sup>3</sup>MATTA TEJASWINI (183T1A0564)

<sup>4</sup>SHAIK SHAMMILA (183T1A0597)

<sup>5</sup>SYEDA HIBA SHAHID (183T1A05A2)

<sup>1</sup>Guide Assisstant professor <sup>2345</sup>UG SCHOLAR  
<sup>2345</sup>Ravindra College of Engineering for Women

## ABSTRACT

Heart related diseases or Cardiovascular Diseases (CVDs) are the main reason for a huge number of deaths in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases. This project presents a survey of various models based on such algorithms and techniques and analyzes their performance. The main objective of this project is to provide efficient algorithm to predict heart disease. So, at the end we compare our algorithm (Genetic algorithm) with BAT and BEE algorithms and we prove that the produced algorithm is efficient one among all.

## I. INTRODUCTION

As indicated by a report by McKinsey, half of Americans have at least one persistent sickness, and 80% of American clinical consideration expense is spent on persistent illness treatment. With the improvement of expectations for everyday comforts, the rate of persistent sickness is expanding. The US has spent a normal of 2.7 trillion

USD every year on persistent infection treatment. This sum includes 18% of the whole yearly Gross domestic product of the US. The medical services issue of ongoing illnesses is likewise vital in numerous different nations. In China, persistent sicknesses are the primary driver of death, as indicated by a Chinese report on sustenance and persistent illnesses in 2015, 86.6% of

passings are brought about by constant infections. Thusly, performing risk evaluations for constant diseases is fundamental. With the development in clinical information [2], gathering electronic wellbeing records (EHR) is progressively advantageous. Furthermore, first introduced a bio-enlivened elite execution heterogeneous vehicular tele-matics worldview, with the end goal that the assortment of versatile clients' wellbeing related continuous large information can be accomplished with the organization of cutting edge heterogeneous vehicular organizations. Patients' factual data, test results furthermore, illness history are kept in the EHR, empowering us to distinguish potential information driven answers for decrease the expenses of clinical contextual investigations. Proposed an ideal large information sharing calculation to deal with the entangle informational collection in telehealth with cloud strategies. One of the applications is to identify high-risk

patients which can be used to diminish clinical expense since high-risk patients frequently require costly medical services. Also, in the principal paper proposing medical services digital actual framework, it creatively presented the idea of expectation based medical care applications, counting wellbeing risk evaluation. Expectation utilizing customary sickness risk models ordinarily implies an AI calculation (e.g., strategic relapse and relapse investigation, and so forth), and particularly a directed learning calculation by the utilization of preparing information with names to prepare the model. In the test set, patients can be ordered into gatherings of either high-chance or generally safe. These models are important in clinical circumstances and are generally considered. Not with standing, these plans have the accompanying qualities and deformities.

The informational collection is ordinarily little, for patients and

infections with explicit circumstances, the attributes are chosen through experience. In any case, these pre-chosen qualities perhaps not fulfill the progressions in the illness and its affecting elements. With the improvement of enormous information examination innovation, more consideration has been paid to illness forecast from the viewpoint of huge information investigation, different explores have been led by choosing the qualities naturally from countless information to work on the exactness of hazard order instead of the recently chosen qualities. Nonetheless, that current work for the most part thought of organized information. For unstructured information, for instance, utilizing convolutional brain organization (CNN) to separate text attributes consequently has previously drawn in wide consideration and additionally accomplished awesome outcomes. Be that as it may, to the best of our insight, none of past work handle Chinese clinical text information by CNN. Thigh-risk

patients which can be used to diminish clinical expense since high-risk patients frequently require costly medical services. Also, in the principal paper proposing medical services digital actual framework, it creatively presented the idea of expectation based medical care applications, counting wellbeing risk evaluation. Expectation utilizing customary sickness risk models ordinarily implies an AI calculation (e.g., strategic relapse and relapse investigation, and so forth), and particularly a directed learning calculation by the utilization of preparing information with names to prepare the model.

In the test set, patients can be ordered into gatherings of either high-chance or generally safe. These models are important in clinical circumstances and are generally considered. Notwithstanding, these plans have the accompanying qualities and deformities. The informational collection is ordinarily little, for patients and infections with

explicit circumstances, the attributes are chosen through experience. In any case, these pre-chosen qualities perhaps not fulfill the progressions in the illness and its affecting elements. With the improvement of enormous information examination innovation, more consideration has been paid to illness forecast from the viewpoint of huge information investigation, different explores have been led by choosing the qualities naturally from countless information to work on the exactness of hazard order, instead of the recently chosen qualities. Nonetheless, that current work for the most part thought of organized information. For unstructured information, for instance, utilizing convolutional brain organization (CNN) to separate text attributes consequently has previously drawn in wide consideration and additionally accomplished awesome outcomes. Be that as it may, to the best of our insight, none of past work handles Chinese clinical text information by CNN.

## II. EXISTING SYSTEM

### SVM ALGORITHM

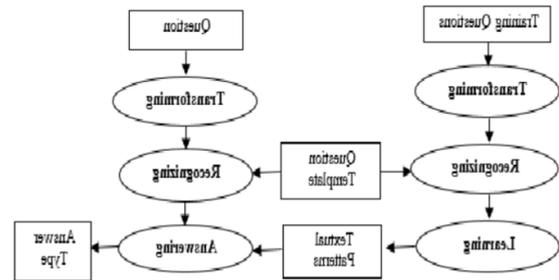
In machine learning, support-vector machines (SVMs, also support-vector networks[1]) are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. Developed at AT&T Bell Laboratories by Vladimir Vapnik with colleagues (Boser et al., 1992, Guyon et al., 1993, Cortes and Vapnik, 1995,[2] Vapnik et al., 1997[citation needed]) SVMs are one of the most robust prediction methods, being based on statistical learning frameworks or VC theory proposed by Vapnik (1982, 1995) and Chervonenkis (1974). Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as

Platt scaling exist to use SVM in a probabilistic classification setting). SVM maps training examples to points in space so as to maximise the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

When data are unlabelled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The support-vector clustering[3] algorithm, created by Hava Siegelmann and Vladimir Vapnik, applies the statistics of support vectors, developed in the support vector

machines algorithm, to categorize unlabeled data.



Classifying data is a common task in machine learning. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new data point will be in. In the case of support-vector machines, a data point is viewed as a  $\{p\}$ -dimensional vector (a list of  $\{p\}$  numbers), and we want to know whether we can separate such points with a  $\{(p-1)\}$ -dimensional hyperplane. This is called a linear classifier. There are many hyperplanes that might classify the data. One reasonable choice as the best hyperplane is the one that represents the largest separation, or

margin, between the two classes. So we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized. If such a hyperplane exists, it is known as the maximum-margin hyperplane and the linear classifier it defines is known as a maximum-margin classifier; or equivalently, the perceptron of optimal stability.[citation needed]

More formally, a support-vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection.[4] Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin, the lower the generalization error of the classifier.[5]

Whereas the original problem may be stated in a finite-dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed[6] that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products of pairs of input data vectors may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function  $\{ \displaystyle k(x,y) \} \{ \displaystyle k(x,y) \}$  selected to suit the problem.[7] The hyperplanes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is constant, where such a set of vectors is an orthogonal (and thus minimal) set of vectors that defines a hyperplane. The vectors defining the hyperplanes can be chosen to be linear

combinations with parameters  $\alpha_i$  of images of feature vectors  $x_i$  that occur in the data base. With this choice of a hyperplane, the points  $x$  in the feature space that are mapped into the hyperplane are defined by the relation  $\sum \alpha_i k(x_i, x) = \text{constant}$ . Note that if  $k(x, y)$  becomes small as  $y$  grows further away from  $x$ , each term in the sum measures the degree of closeness of the test point  $x$  to the corresponding data base point  $x_i$ . In this way, the sum of kernels above can be used to measure the relative nearness of each test point to the data points originating in one or the other of the sets to be discriminated. Note the fact that the set of points

$x$  mapped into any hyperplane can be quite convoluted as a result, allowing much more complex discrimination between sets that are not convex at all in the original space.

Heart diseases have emerged as one of the most prominent cause of death all around the world. According to World Health Organization, heart related diseases are responsible for the taking 17.7 million lives every year, 31% of all global deaths. In India too, heart related diseases have become the leading cause of mortality. Heart diseases have killed 1.7 million Indians in 2016, according to the 2016 Global Burden of Disease Report, released on September 15, 2017. Heart related diseases increase the spending on health care and also reduce the productivity of an individual. Estimates made by the World Health Organisation (WHO), suggest that India have lost up to \$237 billion, from 2005-2015, due to heart related or cardiovascular diseases. Thus, feasible and accurate prediction of

heart related diseases is very important.

Medical organizations, all around the world, collect data on various health related issues. These data can be exploited using various machine learning techniques to gain useful insights. But the data collected is very massive and, many a times, this data can be very noisy. These datasets, which are too overwhelming for human minds to comprehend, can be easily explored using various machine learning techniques. Thus, these algorithms have become very useful, in recent times, to predict the presence or absence of heart related diseases accurately.

### **III. PROPOSED SYSTEM**

Dimensionality Reduction involves selecting mathematical representation such that one can

relate the majority of, but not all, the variance within the given data, thereby including only most significant information. The data considered for a task or a problem, may consists of a lot of attributes or dimensions, but not all of these attributes may equally influence the output. A large number of attributes, or features, may affect the computational complexity and may even lead to over fitting which leads to poor results. Thus, Dimensionality Reduction is a very important step considered while building any model. Dimensionality Reduction is generally achieved by two methods - Feature Extraction and Feature Selection.

To implement this algorithms I am using Heart disease dataset which contains 14 attributes and 4 class labels where 0 refers to No heart Disease and 1 refers to stage1 disease and 2 and 3 refers stage 3 and 4 disease.

Below are some values from dataset to train algorithms

age,sex,cp,trestbps,chol,fbs,restecg,t  
halach,exang,oldpeak,slope,ca,thal,c  
lass

63.0,1.0,1.0,145.0,233.0,1.0,2.0,150.  
0,0.0,2.3,3.0,0.0,6.0,0

67.0,1.0,4.0,160.0,286.0,0.0,2.0,108.  
0,1.0,1.5,2.0,3.0,3.0,2

67.0,1.0,4.0,120.0,229.0,0.0,2.0,129.  
0,1.0,2.6,2.0,2.0,7.0,1

37.0,1.0,3.0,130.0,250.0,0.0,0.0,187.  
0,0.0,3.5,3.0,0.0,3.0,0

First records contain dataset column names and remaining records are the values of dataset. In last column we have class values as 0, 2, 1 and 3 as disease stage.

Test dataset also contains record values but it will not have class labels and application will apply that test values on train dataset to predict it class labels. Some values from test dataset.

age,sex,cp,trestbps,chol,fbs,restecg,t  
halach,exang,oldpeak,slope,ca,thal

63.0,1.0,1.0,145.0,233.0,1.0,2.0,150.  
0,0.0,2.3,3.0,0.0,6.0

67.0,1.0,4.0,160.0,286.0,0.0,2.0,108.  
0,1.0,1.5,2.0,3.0,3.0

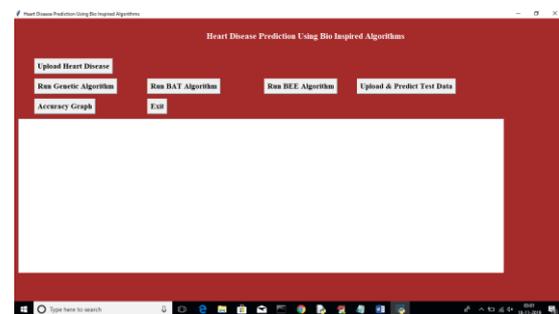
67.0,1.0,4.0,120.0,229.0,0.0,2.0,129.  
0,1.0,2.6,2.0,2.0,7.0

In above test dataset we can see there is no class name and application will predict it. All this files are available inside ‘heart dataset’ folder.

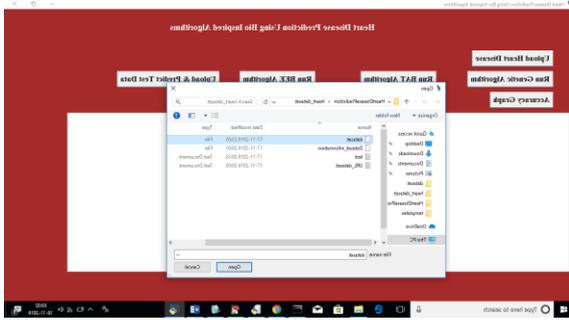
In that ‘heart dataset’ folder I kept dataset URL and information of dataset for references.

Screen shots

To run this project double click on ‘run.bat’ file to get below screen



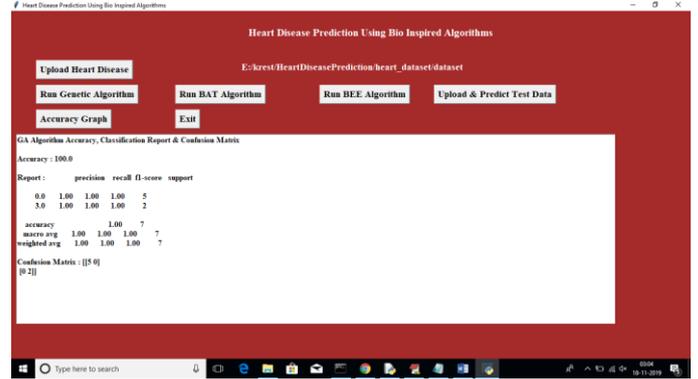
In above screen click on ‘Upload Heart Disease’ button and upload heart disease dataset. See below screen



In above screen uploading dataset file, after uploading will get below screen



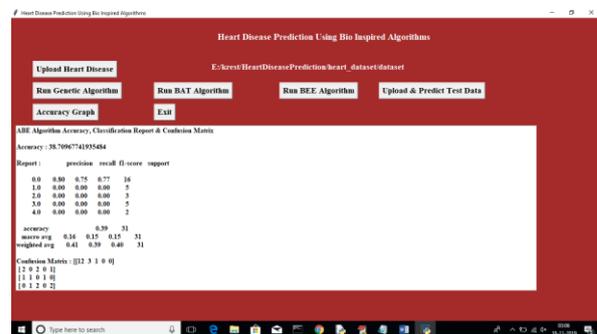
Now click on 'Run Genetic Algorithm' button to run genetic algorithm on dataset and to get its accuracy details. While running this algorithm u can see black console to see feature selection process, while running it will open empty windows, u just close all those empty windows except current window



In above screen for GA accuracy, precision and recall we got 100% result. Now click on 'Run Bat' algorithm button to get its accuracy

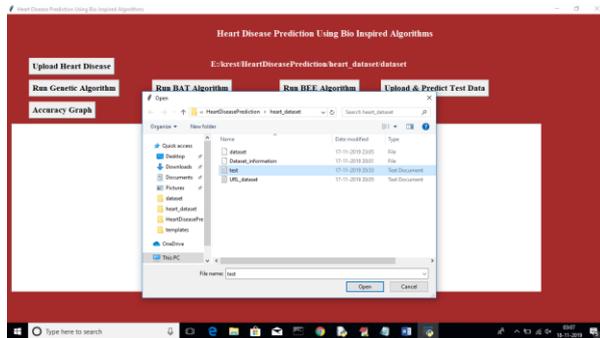


In above screen for BAT we got 45% accuracy, now click on 'Run BEE Algorithm' button to get BEE accuracy

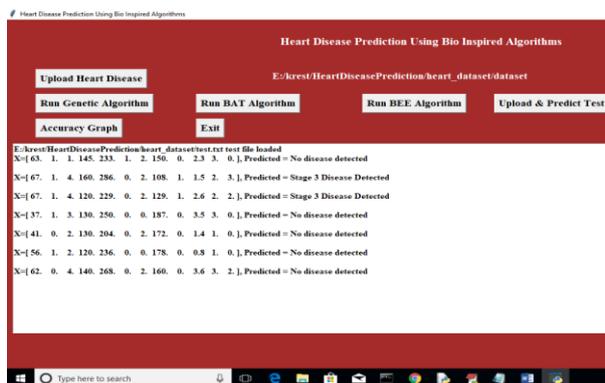


In above screen for BEE we got 38% accuracy, now click on 'Upload &

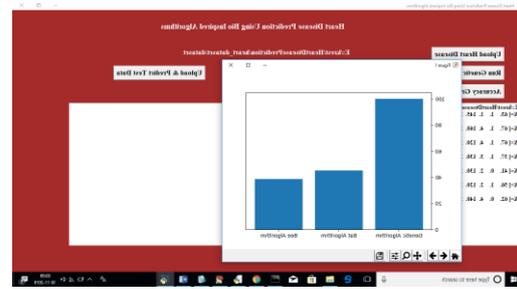
Predict Test Data' button to upload test data and to predict it class



In above screen I am uploading test file which contains test data without class label, after uploading test data will get below screen



In above screen application has predicted disease stages. Now click on 'Accuracy Graph' button to view accuracy of all algorithms in graph format



In above graph x-axis represents Algorithm Name and y-axis represents accuracy of those algorithms

#### IV. ALGORITHMS

##### PROPOSED ALGORITHM

Heart Disease Prediction Using Bio Inspired Algorithms

In this project student want to detect heart disease from dataset using Bio Inspired 4 features optimizing algorithms such as Genetic Algorithm, Bat, Bee and ACO. Here ACO algorithm is design in python to solve Travelling Salesman Problem to find shortest path and it cannot be implemented with heart disease dataset, so I am implementing 3 algorithms called Genetic, Bat and Bee.

Bio inspired algorithms design to optimized features used in dataset for training classification algorithms to increase prediction accuracy, sometime some datasets may have irrelevant values inside dataset and those irrelevant attributes or values may degrade classification accuracy so using optimize algorithms we can reduce features (attribute values) from dataset. This optimize algorithms will be applied on dataset to check whether all values are related to dataset or not, if any attribute found unrelated then it will removed from dataset.

### **What is SDLC?**

SDLC stands for Software Development Life Cycle. A Software Development Life Cycle is essentially a series of steps, or phases, that provide a model for the development and lifecycle management of an application or piece of software.

SDLC is the process consisting of a series of planned activities to develop or alter the software products.

### **Benefits of the SDLC Process**

The intent of a SDLC process it to help produce a product that is cost-efficient, effective, and of high quality. Once an application is created, the SDLC maps the proper deployment and decommissioning of the software once it becomes a legacy. The SDLC methodology usually contains the following stages: Analysis (requirements and design), construction, testing, release, and maintenance (response). Veracode makes it possible to integrate automated security testing into the SDLC process through use of its cloud based platform.

#### **1. Requirements Gathering:**

In this phase we gather all the requirements from the client, i.e. what are the client expected input, output.....

#### **2. Analysis:**

In this phase based upon the client requirements we prepare one documentation is called “High Level Design Document”. It contains Abstract, Functional Requirements, Non Functional Requirements, Existing System, Proposed System, SRS,.....

### 3. Design:

It is difficult to understand the High Level Design Document for all the members, so to understand easily we use “Low Level Design Document”. To design this document we use UML (Unified Modeling Language). In this we have Use case, Sequence, Collaboration.....

### 4. Coding:

In this phase we develop the coding module by

module. After developing all the modules we integrate them.

### 5. Testing:

After developing we have to check weather client requirements are satisfied or not. If not we are again going to develop.

### 6. Implementation:

In testing phase if client requirements are satisfied, we go for implementation. i.e. we need to deploy the application in some server.

### 7. Maintenance:

After deployment, if at all any problems come from the client side; we are providing maintenance for that application.

## V. SYSTEM DESIGN

### UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

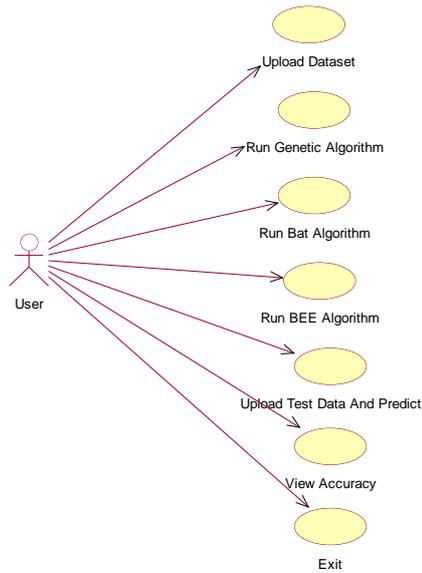
The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

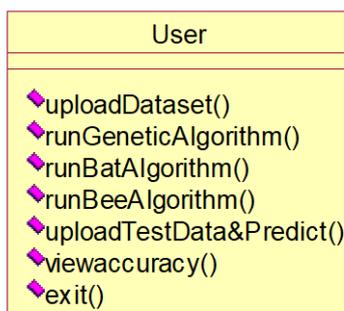
### USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



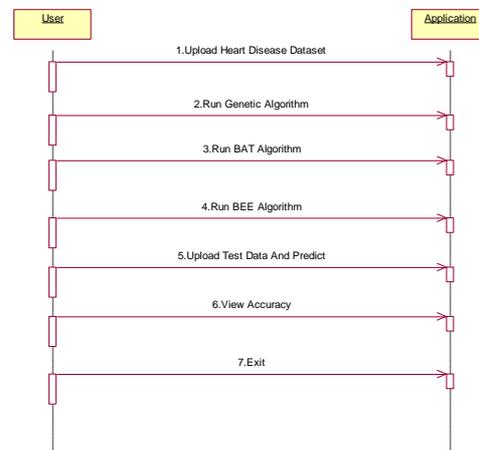
**CLASS DIAGRAM:**

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

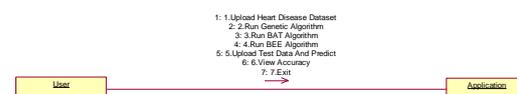


**FLOW CHART**

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.



**Collaboration Diagram:**



Reform Business Technology Office,  
2016.

## I. CONCLUSION

In this paper, we propose a new convolutional neural network based multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured data from hospital. To the best of our knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Compared to several typical prediction algorithms, the prediction accuracy of our proposed algorithm reaches 94.8% with a convergence speed which is faster than that of the CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm.

## II. REFERENCES

[1] P. Groves, B. Kayyali, D. Knott, and S. van Kuiken, The 'Big Data' Revolution in Healthcare: Accelerating Value and Innovation. USA: Center for US Health System

[2] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171–209, Apr. 2014.

[3] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: Towards better research applications and clinical care," *Nature Rev. Genet.*, vol. 13, no. 6, pp. 395–405, 2012.

[4] D. Tian, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, "A dynamic and self-adaptive network selection method for multimode communications in heterogeneous vehicular telematics," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3033–3049, Dec. 2015.

[5] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, "Wearable 2.0: Enable human-cloud integration in next generation healthcare system," *IEEE Commun.*, vol. 55, no. 1, pp. 54–61, Jan. 2017.

- [6] M. Chen, Y. Ma, J. Song, C. Lai, and B. Hu, “Smart clothing: Connecting human with clouds and big data for sustainable health monitoring,” *ACM/Springer Mobile Netw. Appl.*, vol. 21, no. 5, pp. 825–845, 2016.
- [7] M. Chen, P. Zhou, and G. Fortino, “Emotion communication system,” *IEEE Access*, vol. 5, pp. 326–337, 2017, doi: 10.1109/ACCESS.2016.2641480.
- [8] M. Qiu and E. H.-M. Sha, “Cost minimization while satisfying hard/soft timing constraints for heterogeneous embedded systems,” *ACM Trans. Design Autom. Electron. Syst.*, vol. 14, no. 2, p. 25, 2009.
- [9] J. Wang, M. Qiu, and B. Guo, “Enabling real-time information service on telehealth system over cloud-based big data platform,” *J. Syst. Archit.*, vol. 72, pp. 69–79, Jan. 2017.
- [10] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, “Big data in health care: Using analytics to identify and manage high-risk and high-cost patients,” *Health Affairs*, vol. 33, no. 7, pp. 1123–1131, 2014.
- [11] L. Qiu, K. Gai, and M. Qiu, “Optimal big data sharing approach for telehealth in cloud computing,” in *Proc. IEEE Int. Conf. Smart Cloud (SmartCloud)*, Nov. 2016, pp. 184–189.
- [12] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, “HealthCPS: Healthcare cyber-physical system assisted by cloud and big data,” *IEEE Syst. J.*, vol. 11, no. 1, pp. 88–95, Mar. 2017.
- [13] K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, “Localization based on social big data analysis in the vehicular networks,” *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2016.2641467.
- [14] K. Lin, M. Chen, J. Deng, M. M. Hassan, and G. Fortino, “Enhanced fingerprinting and trajectory

prediction for iot localization in smart buildings,” IEEE Trans. Autom. Sci. Eng., vol. 13, no. 3, pp. 1294–1307, Jul. 2016.

[15] D. Oliver, F. Daly, F. C. Martin, and M. E. McMurdo, “Risk factors and risk assessment tools for falls in hospital in-patients: A systematic review,” Age Ageing, vol. 33, no. 2, pp. 122–130, 2004.

[16] S. Marcoon, A. M. Chang, B. Lee, R. Salhi, and J. E. Hollander, “Heart score to further risk stratify patients with low TIMI scores,” Critical Pathways Cardiol., vol. 12, no. 1, pp. 1–5, 2013.

[17] S. Bandyopadhyay et al., “Data mining for censored time-to-event data: A Bayesian network model for predicting cardiovascular risk from electronic health record data,” Data Mining Knowl. Discovery, vol. 29, no. 4, pp. 1033–1069, 2015.

[18] B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, “A relative similarity based method for interactive patient risk prediction,”

Data Mining Knowl. Discovery, vol. 29, no. 4, pp. 1070–1093, 2015.

[19] A. Singh, G. Nadkarni, O. Gottesman, S. B. Ellis, E. P. Bottinger, and J. V. Guttag, “Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration,” J. Biomed. Inform., vol. 53, pp. 220–228, Feb. 2015.

[20] J. Wan et al., “A manufacturing big data solution for active preventive maintenance,” IEEE Trans. Ind. Informat., to be published, doi: 10.1109/TII.2017.2670505

[21] W. Yin and H. Schutze, “Convolutional neural network for paraphrase identification,” in Proc. HLT-NAACL, 2015, pp. 901–911.

[22] N. Nori, H. Kashima, K. Yamashita, H. Ikai, and Y. Imanaka, “Simultaneous modeling of multiple diseases for mortality prediction in acute hospital care,” in Proc. 21th ACM SIGKDD Int. Conf. Knowl.

Discovery Data Mining, 2015, pp. 855–864.

[23] S. Zhai, K.-H. Chang, R. Zhang, and Z. M. Zhang, “Deepintent: Learning attentions for online advertising with recurrent neural networks,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 1295–1304.

[24] K. Hwang and M. Chen, Big Data Analytics for Cloud/IoT and Cognitive Computing. Hoboken, NJ, USA: Wiley, 2017.

[25] H. Chen, R. H. Chiang, and V. C. Storey, “Business intelligence and analytics: From big data to big impact,” MIS Quart., vol. 36, no. 4, pp. 1165–1188, 2012