**SPRINGER BRIEFS IN APPLIED SCIENCES AND**

**TECHNOLOGY ** FORENSIC AND MEDICAL BIOINFORMATICS

P. Venkata Krishna Sasikumar Gurumoorthy Mohammad S. Obaida**t**

**Internet of Things and Personalized Healthcare Systems**

Contents

1. [Sensitivity Analysis of Micro Mass Optical MEMS Sensor](#_bookmark0)

[for Biomedical IoT Devices](#_bookmark0) [1](#_bookmark0)

Mala Serene, Rajasekhara Babu and Zachariah C. Alex

* 1. [Introduction](#_bookmark0) [1](#_bookmark0)
  2. [Modeling and Simulation](#_bookmark1) [2](#_bookmark1)
  3. [Different Shapes of Cantilever](#_bookmark2) [3](#_bookmark2)
  4. [Rectangular-Shaped Micro Mass Optical MEMS Sensor](#_bookmark4) [4](#_bookmark4)
  5. [Trapezoidal/Triangular-Shaped Micro Mass Optical MEMS](#_bookmark6)

[Sensor](#_bookmark6) [5](#_bookmark6)

* 1. [Step Pro*ﬁ*le-Shaped Micro Mass Optical MEMS Sensor](#_bookmark10) [7](#_bookmark10)
  2. [Results and Discussion](#_bookmark13) [8](#_bookmark13)
  3. [Conclusion](#_bookmark18) [11](#_bookmark18)

[References](#_bookmark18) [11](#_bookmark18)

1. [Enhancing the Performance of Decision Tree Using NSUM](#_bookmark35)

[Technique for Diabetes Patients](#_bookmark35) [13](#_bookmark35)

Nithya Settu and M. Rajasekhara Babu

* 1. [Introduction](#_bookmark35) [13](#_bookmark35)
  2. [Related Work](#_bookmark37) [16](#_bookmark37)
  3. [Mutual Information](#_bookmark38) [17](#_bookmark38)
     1. [Symmetric Uncertainty](#_bookmark38) [17](#_bookmark38)
     2. [Proposed Algorithm](#_bookmark41) [18](#_bookmark41)
  4. [Experimental Result and Discussion](#_bookmark41) [18](#_bookmark41)
  5. [Conclusion and Future Scope](#_bookmark43) [19](#_bookmark43)

[References](#_bookmark43) [19](#_bookmark43)

1. [A Novel Framework for Healthcare Monitoring System](#_bookmark71)

[Through Cyber-Physical System](#_bookmark71) [21](#_bookmark71)

[K. Monisha and M. Rajasekhara Babu](#_TOC_250009)

* 1. [Introduction](#_bookmark72) [22](#_bookmark72)
  2. [Related Work](#_bookmark73) [23](#_bookmark73)
     1. [Wireless Body Area Network (WBAN)](#_bookmark73)

[in Healthcare System](#_bookmark73) [23](#_bookmark73)

* + 1. [Electronic Health Record (EHR) Assisted](#_bookmark74)

[by Cloud](#_bookmark74) [24](#_bookmark74)

* + 1. [Data Security in Healthcare Application](#_bookmark75) [26](#_bookmark75)
  1. [Framework for Healthcare Application Through CPS](#_bookmark76) [27](#_bookmark76)
  2. [Internet of Medical Things (IoMT)](#_bookmark78) [29](#_bookmark78)
  3. [Proposed Method](#_bookmark79) [30](#_bookmark79)
  4. [Result and Discussion](#_bookmark83) [32](#_bookmark83)
  5. [Conclusion](#_bookmark85) [33](#_bookmark85)

[References](#_bookmark87) [34](#_bookmark87)

1. [An IoT Model to Improve Cognitive Skills of Student](#_bookmark133)

[Learning Experience Using Neurosensors](#_bookmark133) [37](#_bookmark133)

[Abhishek Padhi, M. Rajasekhara Babu, Bhasker Jha and Shrutisha Joshi](#_TOC_250008)

* 1. [Introduction](#_bookmark133) [37](#_bookmark133)
     1. [Needs or Requirements](#_bookmark133) [37](#_bookmark133)
     2. [Why This Work?](#_bookmark134) [38](#_bookmark134)
     3. [ThinkGear Measurements (MindSet Pro/TGEM)](#_bookmark136) [40](#_bookmark136)
  2. [Existing Methods](#_bookmark137) [41](#_bookmark137)
  3. [Proposed Method](#_bookmark139) [45](#_bookmark139)
  4. [Result and Discussion](#_bookmark141) [47](#_bookmark141)
  5. [Conclusion](#_bookmark144) [48](#_bookmark144)

[References](#_bookmark146) [49](#_bookmark146)

1. [AdaBoost with Feature Selection Using IoT to Bring](#_bookmark175)

[the Paths for Somatic Mutations Evaluation in Cancer](#_bookmark175) [51](#_bookmark175)

[Anuradha Chokka and K. Sandhya Rani](#_TOC_250007)

* 1. [Introduction](#_bookmark175) [51](#_bookmark175)
     1. [AdaBoost Technique](#_bookmark176) [52](#_bookmark176)
     2. [Feature Selection Techniques](#_bookmark176) [52](#_bookmark176)
     3. [Internet of Things (IoT)](#_bookmark177) [53](#_bookmark177)
     4. [Challenges in Sequencing](#_bookmark177) [53](#_bookmark177)
  2. [Existing Models](#_bookmark178) [54](#_bookmark178)
  3. [Methodology](#_bookmark179) [55](#_bookmark179)
     1. [Redundancy and Relevancy Analysis Approach](#_bookmark179) [55](#_bookmark179)
     2. [Feature Redundancy and Feature Relevancy](#_bookmark180) [56](#_bookmark180)
     3. [De*ﬁ*ning a Framework of AdaBoost Technique](#_bookmark180)

[with Feature Selection](#_bookmark180) [56](#_bookmark180)

[Schematic Representation for the Proposed](#_bookmark181)

[Algorithm](#_bookmark181) [57](#_bookmark181)

* + 1. [Algorithm and Analysis](#_bookmark181) [57](#_bookmark181)
    2. [IoT Wearables to Detect Cancer](#_bookmark187) [58](#_bookmark187)
  1. [Conclusions](#_bookmark191) [61](#_bookmark191)

[References](#_bookmark195) [62](#_bookmark195)

1. [A Computational Approach to Predict Diabetic Retinopathy](#_bookmark325)

[Through Data Analytics](#_bookmark325) [105](#_bookmark325)

[Ashraf Ali Shaik, Ch Prathima and Naresh Babu Muppalaneni](#_TOC_250002)

* 1. [Introduction](#_bookmark325) [105](#_bookmark325)
     1. [Steps in Algorithm](#_bookmark326) [107](#_bookmark326)
  2. [Methodology](#_bookmark326) [107](#_bookmark326)
     1. [Description of Dataset](#_bookmark326) [107](#_bookmark326)
     2. [Attribute Information](#_bookmark328) [108](#_bookmark328)
     3. [Cross-Validation](#_bookmark328) [108](#_bookmark328)
     4. [Classi](#_bookmark328)*ﬁ*[cation Matrix](#_bookmark328) [108](#_bookmark328)
     5. [Bagging and Boosting](#_bookmark329) [109](#_bookmark329)
  3. [Performance Measures](#_bookmark329) [109](#_bookmark329)
     1. [Accuracy](#_bookmark329) [109](#_bookmark329)
     2. [Sensitivity](#_bookmark329) [109](#_bookmark329)
     3. [Speci*ﬁ*city](#_bookmark331) [110](#_bookmark331)
     4. [Classi*ﬁ*cation Matrix](#_bookmark331) [110](#_bookmark331)
  4. [Tools Used and Results Discussion](#_bookmark331) [110](#_bookmark331)
  5. [Conclusion](#_bookmark333) [111](#_bookmark333)

[References](#_bookmark337) [112](#_bookmark337)

**Chapter 1**

# Sensitivity Analysis of Micro-Mass Optical MEMS Sensor for Biomedical IoT Devices

##### Mala Serene, Rajasekhara Babu and Zachariah C. Alex

**Abstract** Micro-electromechanical systems (MEMS) have tremendous applications in the field of biomedical and chemical sensors. There are different readout tech- niques like piezo-resistive and piezo-electric which are used to measure the stimuli absorbed by the cantilever into electrical signals. In this paper, we used the open- source Ptolemy software to model MOEMS sensor with novel optical read out. To enhance the deflection and sensitivity, four micro-mass optical MEMS sensor mod- els were developed using four different shapes of the cantilever. The detectable

mass range measured by the triangular cantilever using parylene as material is 50.97 μg–23.996 mg.

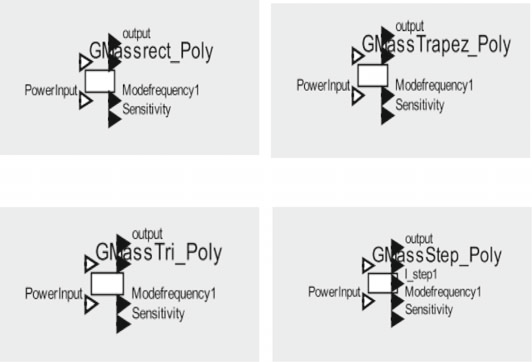
* 1. **Introduction**

Microcantilever sensors offer a highly promising area to sense various physical stim- uli, chemical vapors, and measure very small masses. The different shapes of the cantilever are used to detect different diseases, chemical, and micro-masses. Many readout methods like piezo-resistive and piezo-electric are used to convert the stimuli present on the cantilever. The optical lever method uses the atomic force microscopy which can accurately measure the deflection of the cantilever when compared to the above-mentioned techniques [[1](#_bookmark19)].

### Modeling and Simulation

*A.* Modeling and simulation play a key role in the silicon electronic industry. Through the modeling process, the system can developed and its behavior can be studied in a particular environment [[13](#_bookmark31)]. As it is already mentioned that the system-level software is not available for optical MEMS, Ptolemy II is an open-source framework chosen

* 1. Modeling and Simulation 3
     1. **(b)**



**(c)**

**(d)**

**Fig. 1.1** micro-mass actor for **a** rectangular, **b** trapezoidal, **c** triangular, **d** step profile rectangular using Ptolemy

for modeling, simulation, and design of concurrent systems [[14](#_bookmark32)]. In this chapter, to enhance the sensitivity of the sensor, four different shapes of the cantilever are used in the micro-mass optical MEMS sensor. The software codes for laser actor, pho- todetector actor, and four micro-mass actors using four different shapes of cantilever are developed using Ptolemy software [[15](#_bookmark33)]. Each mass actor consists of two optical fibers and one of the four shapes of cantilever like rectangular, trapezoidal, triangu- lar, and step profile cantilever to make micro-mass actor. The developed micro-mass actors for different shapes of cantilever like rectangular, trapezoidal, triangular, and step profile are shown in Fig. [1.1](#_bookmark3).

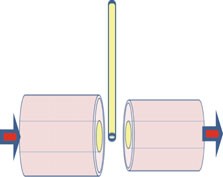
### Different Shapes of Cantilever

To enhance the sensitivity of the mass sensor, the study is carried out using two different polymers and four shapes of microcantilever. The polymers are chosen as microcantilever materials, because of its low Young’s modulus, which gives more deflection than silicon. The polymer cantilevers can be easily fabricated. The fabri- cation cost also is lesser than silicon. But the polymers are temperature sensitive, so the cantilever should kept in a protective environment. The mathematical equation needed to develop the sensor and the micro-mass optical MEMS sensor for four different shapes are discussed below.

4 1 Sensitivity Analysis of Micro-Mass Optical MEMS Sensor …

The fundamental resonant frequency (f0) of the rectangular cantilever is given by

**(a)**



Mass

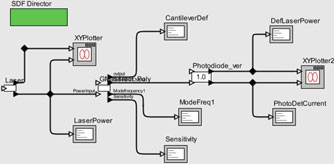
Op- tical fiber

cal fiber

Cantile- ver

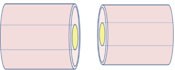
Pho- to De- tec- tor

La ser So ur- ce

**(b)**

**Fig. 1.2 a** Micro-mass optical MEMS sensor using rectangular cantilever. **b** Micro-mass optical MEMS sensor using rectangular-shaped cantilever using Ptolemy (both Polyimide and Parylene)

**(a)**



Canti

Op ti- cal

Op-

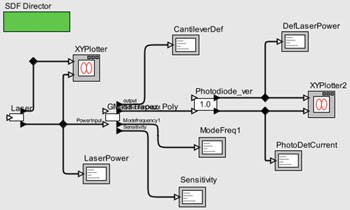
ti-

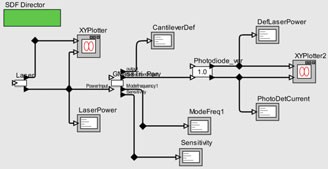
L

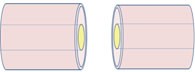
a s e

Ph oto De tec tor

**(b)**

**Fig. 1.3** Micro-mass optical MEMS sensor using trapezoidal cantilever. **a** Block diagram. **b** Model using Ptolemy (both polyimide and parylene)

**(b)**



**(a)**

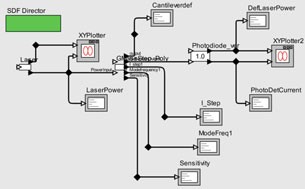
La- ser Sou rce

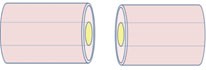
Optical fi- ber

Optical fi- ber

Pho- to De- tec- tor

**Fig. 1.4** Micro-mass optical MEMS sensor using triangular cantilever. **a** Block diagram. **b** Model using Ptolemy (both polyimide and parylene)

**(b)**



Step Pro- file

Opti- cal fiber

Op- tical fiber

Pho- to De- tec- tor

La- ser Sou rce

**(a)**

**Fig. 1.5** Micro-mass optical MEMS sensor using step profile cantilever. **a** Block diagram. **b** Model using Ptolemy (both polyimide and parylene)

### Results and Discussion

The range of the mass measured by the four micro-mass optical sensor models using four shapes are given in Table [1.1](#_bookmark15). The lowest mass is measured by triangular-shaped cantilever when compared to other shapes. Step profile rectangular can measure 70.1441% less mass when compared to rectangular cantilever. Triangular cantilever can sense 76.5217% less mass than trapezoidal cantilever is shown in Fig. [1.6](#_bookmark16).

The fundamental frequency of the rectangular cantilever is lowest, and the trian- gular is highest shown in Fig. [1.6](#_bookmark16). Sensitivity of the triangular shape is highest when compared to other shapes. Sensitivity of the step profile cantilever is 14.445% more than rectangular cantilever when Polyimide is used as cantilever material shown in Fig. [1.6](#_bookmark16). Parylene is used as cantilever material in the micro-mass optical MEMS sensor. The lowest mass is measured by triangular-shaped cantilever when com- pared to other shapes. Step profile rectangular can measure 70.1299% less mass when compared to rectangular cantilever. Triangular cantilever can sense 76.6067% less mass than trapezoidal cantilever which is shown in Fig. [1.7](#_bookmark17). The fundamental frequency of the rectangular cantilever is lowest, and the triangular is highest shown in Fig. [1.7](#_bookmark17). Sensitivity of the triangular shape is highest when compared to other shapes. Sensitivity of the step profile cantilever is 21.0181% more than rectangular cantilever when parylene is used as cantilever material shown in Fig. [1.7](#_bookmark17). Parylene

**Table 1.1** Results of different shapes of cantilever Micro-mass optical MEMS sensor

|  |  |  |
| --- | --- | --- |
| Shape of the cantilever | Polyimide | Parylene |
| Mass range | Mass range |
| Rectangular | 101.9 μg–98.37 mg | 101.9 μg–83.2 mg |
| Trapezoidal | 101.9 μg–63.61 mg | 101.9 μg–53.81 mg |
| Triangular | 81.55 μg–28.39 mg | 50.97 μg–23.996 mg |
| Step profile cantilever | 203.9 μg–47.299 mg | 50.97 μg–39.96 mg |

1.7 Results and Discussion

9

Results and Discussion 11

cantilever can sense approximately 15.4% less mass than Polyimide cantilever in all shapes. Parylene cantilever fundamental frequency is approximately 6.4% less than Polyimide. Parylene cantilever is approximately 6.2% less sensitivity than polyimide cantilever. The different shapes of the cantilever are modeled and simulated using the open-source framework Ptolemy. The deflection, fundamental frequency, and sensitivity of each shape were measured.

Parylene cantilever can sense approximately 15.4% less mass than polyimide cantilever in all shapes. Parylene cantilever fundamental frequency is approximately 6.4% less than polyimide. Parylene cantilever is approximately 6.2% less sensitivity than polyimide cantilever. The different shapes of the cantilever are modeled and simulated using the open-source framework Ptolemy. The deflection, fundamental frequency, and sensitivity of each shape were measured.

The triangular cantilever able to measure lowest mass among the four shapes, and fundamental frequency and sensitivity of the triangular is higher than other shapes. So far the system-level model has developed using Ptolemy framework.

### Conclusion

The four micro-mass optical MEMS sensors are modeled using four different shapes of the cantilever in open-source framework Ptolemy. The deflection, fundamental fre- quency, and sensitivity of each shape are measured. The micro-mass optical MEMS sensor using triangular cantilever able to measure lowest mass among the four shapes and fundamental frequency and sensitivity of that sensor is high than other shapes. Two polymers are used as the sensor material for the cantilever. The detectable mass range measured by the triangular cantilever using polyimide as material is

81.55 μg–28.39 mg. The detectable mass range measured by the triangular can- tilever using parylene as material is 50.97 μg–23.996 mg.

### References

1. P.G. Waggoner, H.G. Craighead, Micro- and nanomechanical sensors for environmental, chem- ical, and biological detection. Lab Chip **7**(10), 1238–1255 (2007)
2. A. Gupta, D. Akin, R. Bashir, Single virus particle mass detection using microresonators with nanoscale thickness. Appl. Phys. Lett. **84**(11), 1976–1978 (2004)
3. M.Z. Ansari, C. Cho, J. Kim, B. Bang, Comparison between deflection and vibration charac- teristics of rectangular and trapezoidal profile microcantilevers. Sensors **9**, 2706–2718 (2009)
4. H.F. Hawari, Y. Wahab, M.T. Azmi, A.Y. Shakaff, U. Hashim, S. Johari, Design and analysis of various microcantilever shapes for MEMS based sensing. J. Phys. Conf. Ser. **495**, 1–9 (2014)
5. M. Abdel-Jaber, A. Al-Qaisia, M.S. Abdel-Jaber, Nonlinear natural frequencies of a tapered cantilever beam. Advanc. Steel Construct. **5**, 259–272 (2009)
6. S. Morshed, B.C. Prorok, Tailoring beam mechanics towards enhancing detection of hazardous biological species. Experiment. Mech. **47**, 405–415 (2007)

**Chapter 2**

# Enhancing the Performance of Decision Tree Using NSUM Technique

**for Diabetes Patients**

##### Nithya Settu and M. Rajasekhara Babu

**Abstract** Diabetes is a common disease among children to adult in this era. To prevent the diseases is very important because it saves the human lives. Data mining technique helps to solve the problem of predicting diabetes. It has steps of processes to predict the illness. Feature selection is an important phase in data mining pro- cess. In feature selection when dimension of the data increases, the quantity of data required to deliver a dependable analysis raises exponentially. Numerous different feature selection and feature extraction techniques are present, and they are widely used filter-based feature selection method is proposed which takes advantage of the wrapper, Embedded, hybrid methods by evaluating with a lower cost and improves the performance of a classification algorithm like a decision tree, support vector machine, logistic regression and so on. To predict whether the patient has diabetes or not, we introduce a novel filter method ranking technique called Novel Symmetrical Uncertainty Measure (NSUM). NSUM technique experimentally shows that com- pared to the other algorithms in filter method, wrapper method, embedded method and hybrid method it proves more efficient in terms of Performance, Accuracy, Less computational complexity. The existing technique of symmetric uncertainty measure shows less computational power and high performance, but it lacks in accuracy. The aim of the NSUM method is to overcome the drawback of the filter method, i.e., less accuracy compared to other methods. NSUM technique results show high perfor- mance, improved accuracy, and less computational complexity. NSUM method runs in 0.03 s with 89.12% as accuracy by using Weka tool.

**2.1 Introduction**

Diabetes is the disorder that outcomes from lack of insulin in a human being blood. There are one more types of diabetes called is diabetes insipidus. When the patient mentions “diabetes,” they mean diabetes mellitus (DM) [[1](#_bookmark46)]. A human with dia- betes mellitus is called “diabetics.” Diabetes symptoms include frequent urination, increased hunger, and thirst. If this is untreated, it will lead to serious complica- tions. This complication includes kidney failure, stroke, damage to the eyes, heart disease, and foot ulcers. If there is a decrease in the sugar level in the blood, it will be called as a pre-diabetes [[2](#_bookmark47)]. Diabetes is causes when the pancreas does not secrete enough insulin to the body. Diabetes mellitus is of three types, namely type I dia- betes mellitus, type II diabetes mellitus, and gestational diabetes. This is explained in detail. Type I diabetes is caused when the pancreas fails to yield enough insulin. It is referred as IDDM which is “insulin-dependent diabetes mellitus” alias “ju- venile diabetes.” The root cause is unknown. It will affect the people from below 20 years of age. It will continue throughout their life. They should follow strict diet and exercise. Type II diabetes starts when the insulin stops working in the human body. When the disease increases, the insulin level will be reduced. This is called as NIDDM non-insulin-dependent diabetes mellitus. The reason for this diabetes is obesity and lack of exercise. Type III diabetes is called as gestational diabetes. This will occur during the pregnancy. In the history of the patient, it will develop high sugar level.

These features should be helpful to accurately predict the outcome of the improved performance. It has been classified into four types, namely filter, wrapper, embedded, and hybrid.

* 1. Introduction 15

###### Filter Method

Filter method is independent of the building model during the time of execution. It has a very low computational cost that is an advantage but the accuracy of the model will be produced is not be promised [[10](#_bookmark55)].

###### Wrapper Method

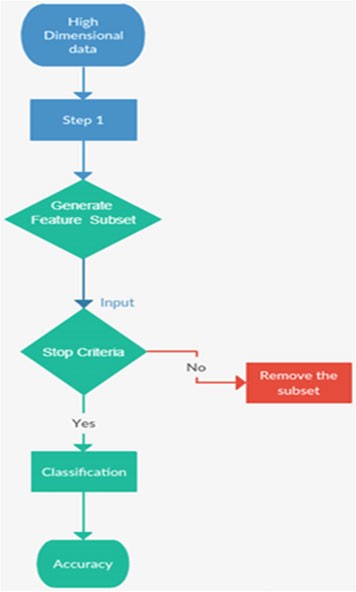
Wrapper method is dependent on building a model during classification or clus- tering based on the selected attributes [[11](#_bookmark56)]. It produces the high accuracy, but the computational complexity is too high [[12](#_bookmark57)].

###### Embedded Method

Embedded technique uses the FS as a portion of the training process. This method comparatively produces less accuracy than the wrapper method [[13](#_bookmark58)].

###### Hybrid Method

Hybrid technique is a mixture of wrapper and filter methods to achieve more accuracy, low computational time, and less cost [[14](#_bookmark59)] (Fig. [2.1](#_bookmark36)).

**Fig. 2.1** Feature subset selection process

16 2 Enhancing the Performance of Decision Tree …

The aim of this paper is to improve the performance of the filter algorithm by using symmetrical uncertainty measure (SUM). We proposed a novel algorithm for SUM technique which is called as NSUM.

18 2 Enhancing the Performance of Decision Tree …

#### Proposed Algorithm

Symmetric uncertainty ranking-based feature selection Input dataset—(f1, f2, f3 … fn, C), threshold ň.

Output dataset—an optimal subset of features.

* + - 1. Begin the algorithm.
      2. Calculate the SUi for each feature fi.
      3. Check SUi is greater than the ň value.
      4. Store fi into D’ array variable.
      5. Find the length of the D’ array.
      6. Calculate the middle index (MID). For D’.
      7. Select the first value in the pointer. Travel (first point to the midpoint).
      8. Select the last value in the pointer. Travel (last value point to the midpoint).
      9. Using the temporary variable, swap the data according to the ranking. First part.
      10. Using the temporary variable, swap the data according to the ranking. Second part.
      11. /\*\*\* {Select the First feature! Midpoint} Sort the data according to the feature.

=

* + - 1. Variable array 1 Store the sorted data.

=

* + - 1. /\*\* {Select the last feature! Midpoint} Sort the data according to the feature.

=

* + - 1. Variable array 2 Store the sorted data.

=

* + - 1. Add the variable array 1 and variable array 2.
      2. OUTPUT is optimized features.

### Experimental Result and Discussion

In our new work, we evaluated the efficiency of the recommended technique. The aim of our plan is to assess the method in terms of speed, no of selected attributes, predictive accuracy for a J48 classifier selected feature. The algorithm matched in contradiction of some previously existing techniques: SOM, chi-square, relief, and FCBF on the diabetes high-dimensional datasets. NSUM approach outcome is less number of features as compared to SOM, FCBC, and Relief, grades in the reduction of time for the resultant mining algorithm. A list of datasets used in our approach is from the UCI repository [[26](#_bookmark70)]. A brief summary of datasets is described in Table [2.1](#_bookmark42).

**Table 2.1** Feature technique run time

|  |  |  |  |
| --- | --- | --- | --- |
| Technique | Time (ms) | Correctly identified instances | Incorrectly identified instances |
| SUM | 0.06 | 79.08 | 20.92 |
| NSUM | 0.03 | 87.12 | 12.88 |

### Conclusion and Future Scope

Decision tree data mining technique is used to help healthcare specialists in the diag- nosis of diabetes millitus disease. Applying health mining is helpful to healthcare, disease diagnosis, and treatment. The future scope will be using a hybrid model increase the accuracy and performance optimization (Figs. [2.2](#_bookmark44) and [2.3](#_bookmark45)).

### References

* + 1. S. Siddiqui, Depression in type 2 diabetes mellitus—a brief review. Diabetes Metab. Synd. Clin. Res. Rev. **8**(1), 62–65 (2014)
    2. K. Rajesh, V. Sangeetha, Application of data mining methods and techniques for diabetes diagnosis. Int. J. Eng. Innov. Technol. (IJEIT) **2**(3) (2012)
    3. S. Sarma Kattamuri, Predictive modeling with SAS enterprise miner: practical solutions for business applications (SAS Institute, 2013)
    4. W. Gregg Edward et al., Association of an intensive lifestyle intervention with remission of type 2 diabetes. JAMA **308**(23), 2489–2496 (2012)
    5. A.R. Mire-Sluis, R.G. Das, A. Lernmark, American diabetes association. Diabetes/Metab. Res. Rev. **15**(1), 78–79 (1999), [http://www.diabetes.org](http://www.diabetes.org/)
    6. I. Yoo et al., Data mining in healthcare and biomedicine: a survey of the literature. J. Med. Syst. **36**(4), 2431–2448 (2012)
    7. R. Li et al., Cost-effectiveness of interventions to prevent and control diabetes mellitus: a systematic review. Diabetes Care **33**(8), 1872–1894 (2010)
    8. J.-H. Lin, P.J. Haug, Data preparation framework for preprocessing clinical data in data mining, in *AMIA Annual Symposium Proceedings* (American Medical Informatics Association, 2006)
    9. M. Luboschik et al., Supporting an early detection of diabetic neuropathy by visual analytics, in *Proceedings of the EuroVis Workshop on Visual Analytics (EuroVA)* (2014)

**Chapter 3**

# A Novel Framework for Healthcare Monitoring System Through

**Cyber-Physical System**

##### K. Monisha and M. Rajasekhara Babu

**Abstract** In recent years, the major concern with people is healthcare. Humans are susceptible to various chronic diseases such as diabetes insipidus, kidney diseases, and eating disorders. The patient suffering from the mentioned diseases should be monitored and treated regularly to avoid any serious conditions. Thus, an embedded technology is developed to transfer the patient’s health information through sensor to network and then to the cloud storage. The existing technologies usually monitor the patient’s clinical data and share the sensor data to cloud. But, the system does not perform any data analysis or actuation process for efficient remedial treatment. In critical situations, the patient also requires the doctors and clinical assistants to be alongside to provide treatment immediately. Therefore, it requires a smart improve- ment in the current technology. In our methodology, we implement cyber-physical system (CPS) technique for healthcare system. CPS technology classifies the imple- mentation into three parts, namely communication, computation, and actuation or control. CPS continuously monitors the patient’s health parameters such as blood glucose (BG) level, blood pressure (BP) level, body temperature (BT) level, and heart beat (HB) rate. When the health parameter value reaches their critical bound, then through actuators the patients are treated inevitably as a remedial measure. The proposed system benefits the patients, doctors, and clinical assistants in reducing the overhead of assisting all the patients during the inconvenience period. Due to increased physical connectivity constraints, embedded systems and networks have more security exposures. Especially in healthcare systems, the lack of importance on device security has headed to numerous cyber-security gaps. Therefore, a proper investigation is needed on the CPS security issues to make sure that systems are working safe. Furthermore, security resilience and robustness are discussed. Finally, some healthcare data security challenges are elevated for the future study. The pro- posed CPS model decreases the overhead of medical representatives. This approach also decreases the time and cost complexity compared to the previous works.

**3.1 Introduction**

Cyber-physical system (CPS) is a recent research topic that has received widespread attention across different domains, including smart grids, smart hospitals, smart house, and energy [[1](#_bookmark88)]. Generally, CPS is a technology comprised of 3Cs namely communication, computation, and control, respectively. CPS forms closed loop by connecting machine to machine where the physical components (e.g., sensors and Wi-Fi boards) actively interact with the cyber-space network (e.g., Internet) for trans- mission of data to cloud storage and response back to the physical space using actuator machine. CPS mutates the process of interaction occurring in the corporal world, as distinct device needs different form of safety levels built on robustness of control system and the sensitivity of the data that is exchanged. CPS has challenges in pre- serving the security and privacy in each application. In healthcare system [[2](#_bookmark89)], it is important to facilitate the security, privacy, reliability, and assurance for effective health device communication. Thus, implementing an efficient healthcare system using CPS requires a secure pervasive network model. A model, namely Perva- sive Social Network (PSN) for healthcare-based system, is promoted for sharing the patient’s health data collected from the medical sensor over the secure network [[3](#_bookmark90), [4](#_bookmark91)]. The data used in e-healthcare application is represented as Electronic Healthcare Record (EHR) which has a vital scope in improving the healthcare usability, human experience, and data intelligence. It is observed that EHR could eventually store huge volume of data allowing effective retrieval of clinical records [[5](#_bookmark92)].

In healthcare system, sharing of patient’s medical record should help in making the user experience smarter, better endorsement for both doctors and patients, under- standing the data patterns and diseases to provide better healthcare quality service. Due to the significant feature of CPS, its application is used everywhere with testified results. The Health CPS also has a prompting growth due to the evolution of hardware techniques having the standard bandwidth by integrating the intellectual radio-based networks to disclose the utilization range of frequency band. Machine-to-machine communication in CPS, wireless sensor network (WSN), and cloud computing has become a fundamental part for any Internet-based applications [[6](#_bookmark93)]. As many parts included in the IoT applications, it also required to have a concern on security prob- lems relating to WSN, CPS, and cloud computing. The security concerned problems can appear in the background of IoT or CPS having Internet protocol standard for connectivity. Consequently, in recent surveys many effects have been taken to han- dle the security issues in the CPS model. Different security approaches are followed namely, providing security to only particular layer and in some cases providing end- to-end security to the CPS applications [[7](#_bookmark94), [8](#_bookmark95)]. Relating to the security issues in the CPS network layer, it is observed that more than many thousand health consumer devices were in consent to distribute spam mails, brute force attack, and other out- breaks.

### Related Work

#### Wireless Body Area Network (WBAN) in Healthcare System

In [[16](#_bookmark103)] methodology, a wireless body area network (WBAN) is designed for imple- menting healthcare application. WBAN uses the clinical band to transfer the patient’s physiological parameters from the sensor node through microcontroller using wire- less communication system. To increase the synchronization aspect between the sensor node devices and other network node devices, clinical bands are introduced to reduce the interventions at different health centers [[17](#_bookmark104), [18](#_bookmark105)].

The proposed system employs the multi-hoping method to transfer the collected sensor data from one location to another isolated location using wireless gateway board. The exchange of information happens by connecting the sensor node to the Wi-Fi node or local area network (LAN). The proposed WBAN for medical applications ensures in facilitating the health centers, doctors, and clinical assistants to access the patient’s physiological parameters at anywhere through both offline and online [[19](#_bookmark106), [20](#_bookmark107)]. The defined methodology also reduces the medical cost, human faults, and periodical checkup for patients attended by medical professionals. In [[21](#_bookmark108)], WBAN security and privacy aspects are discussed. In smart technologies, it is important to provide a high-level security and privacy which is a vital scope for healthcare monitoring applications. Healthcare monitoring system is responsible for observing and transferring the patient health data over the network to the cloud for storage purposes. Hence, it is essential to protect the health data parameters from the intruder’s exploitation. Therefore, the proposed system works in deploying the WBAN based on the privacy and security aspects [[22](#_bookmark109), [23](#_bookmark110)].

#### Electronic Health Record (EHR) Assisted by Cloud

In [[28](#_bookmark115)], Clinical Document Architecture (CDA) is generated and integrated with health records for secure exchange of information using cloud computing. It is noted that electronic health record is used for storing the patient’s physiological param- eters. Hence, prerequisite of interoperability is required for deploying EHR for an improvised patient healthcare and security. CDA is a fundamental document stan- dard developed by HL7 for interoperability concept between heterogeneous domains.

3.2 Related Work 25

The proposed Health CPS model consists of three layers, namely (a) data collec- tion layer—It consists of all e-health standards integrated together as a collection;

(b) data service-oriented layer—This layer is responsible for providing all Health CPS-related services; (c) data management layer—The management layer controls parallel computing and distributed data storage in healthcare system. Finally, it is observed that a smart healthcare system is implemented using cloud and big data technologies. In [[34](#_bookmark121)] proposed system, the electronic patient centric records are han- dled and stored in cloud using a secure role-based technique. In recent observations, cloud technology encounters a rapid growth applied in different applications. Even- tually, a cloud server has to adapt a larger data storage with increasing popularity in smart technologies such as smart hospital, smart grid, smart city, and smart energy. Hence, many hospitals started to store patients’ record in an electronic form rather than manual data. These electronic health records are stored through cloud-based mechanism for better retrieval of data and quality of service [[35](#_bookmark122)]. However, despite cloud having the advantage in storage, it also has the issues in security aspect involv- ing to unauthorized users. A cryptographic technique, namely role-based encryption model, is implemented to frame a secure and flexible cloud-based system to store electronic health records. The role-based encryption system ensures in framing the policies in the cloud system by avoiding the unauthorized user access [[36](#_bookmark123), [37](#_bookmark124)]. The proposed role-based encryption (RBE) system also establishes the security and reliability consent with Personally Controlled Electronic Health Record (PCEHR) system developed by the research center. Thus, the implemented system has the capability to deploy its role-based accessible secure method in any healthcare related appli- cations. The methodology also observes experimental access procedures based on the roles and delivers secure storage access in cloud server imposing these access specific strategies.

#### Data Security in Healthcare Application

In [[38](#_bookmark125)] survey, big data technology has become a driving factor for many applications such as healthcare research, information technology, and educational institutions [[39](#_bookmark126), [40](#_bookmark127)]. Big data technology has many advantages such as time and cost reduction, and advanced product development. However, big data technology also encounters many challenges and impediments in providing security, privacy, and proficient talents in software development. One among those applications in big data is e-healthcare system where the health records are most susceptible to the attackers. Those attackers can easily find out the sensible data and spread them across the network which eventually leads to data breach [[41](#_bookmark128), [42](#_bookmark129)]. Hence, authentication is an important aspect in the healthcare system to protect those sensitive data from breaching by using various techniques such as.

##### Data Encryption

Encryption allows protecting the ownership of the data by avoiding any unauthorized user access to database. Encryption algorithms such as RSA, DES, RC4, AES are used as an encryption scheme for any efficient data privacy management.

##### Authentication

It involves authenticating the users to access the e-healthcare records by applying cryptographic protocols such as secure socket layer (SSL) and transport layer proto- col (TLP).

##### Access Control

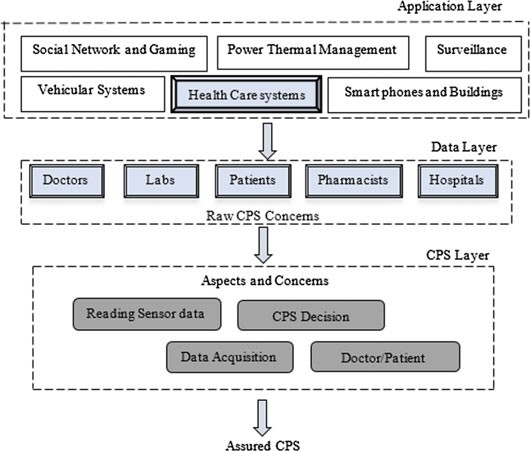
When an authenticated user accesses the e-health system database, they are regulated by the access control policy though the user is authenticated. Here the user gets their rights and privilege only when they are authorized as patients. Some of the techniques used for access control are sequence access control (SAC) and role-based control (RBC).

### Framework for Healthcare Application Through CPS

Healthcare system requires a constant improvisation in its organization resources and structure. Accordingly, many health research organizations manage in improving the efficiency and reliability of Electronic Health Records (EHR). The medical institu- tions improved their proficiency through unification adapters and health monitoring devices over the network module. These organizations also make an operable func- tion over the influenceable variables cached in their healthcare server [[43](#_bookmark130)]. However, the operations defined in the server defect in their vital extensions as the structure of the healthcare system is more complex than the predicted one. The modifica- tions that happen frequently or rarely in the server frameworks can affect the service delivered by the wellness program. The changes can affect the service standards by performing in an unusual behavior. For example, a doctor or medical assistants will be unable to provide proper treatment to patients in given time due to irregular update along with unexpected costs. Hence, a smart system is required by integrating the service-oriented cloud with other smart solutions to monitor the patients regu- larly. The patient’s heath parameters are observed by sensors, microcontrollers, and other smart devices such as computers and mobiles. The interconnected solutions are accessible to clinical data which is presented through some algorithms and frame- works. The patterns are recognized through the algorithms for each patient with the responses stored in the data servers.

Thus, to provide the best solution to the healthcare care organizations, smart systems are employed. The smart system with effective machine-to-machine com- munication is provided through cyber-physical system (CPS). CPS framework is deployed for effectual healthcare monitoring system. CPS is a mechanism developed using problem-solving algorithms connected to the Internet users through network adapters. CPS is a technique built upon logically by merging the optimized algo- rithms with the networks and smart physical devices. CPS is employed in the plat- form whenever a smart implementation is required in an environmental application. In Fig. [3.1](#_bookmark77), a framework is designed for healthcare monitoring system by applying CPS notions. In the design, the framework is divided into three layers, namely (a) application layer—It consists of the applications defined for CPS technologies; (b) data layer—It includes entities or the members who analyze data for further concern in the system; (c) CPS layer—This layer consists of actual CPS implementation for smart hospital.

28 3 A Novel Framework for Healthcare Monitoring System …



**Fig. 3.1** Health CPS framework

Each layer in the defined framework makes vital scope for effective healthcare monitoring system through assured CPS. The objective is to provide the well-defined framework in coordination with common architectural standards in the scope of deploying the smart hospital. Application layer—It consists of the domains for the smart system, namely smart grid, smart hospital, smart energy, smart city, smart vehicle, and smart house. In the proposed system, smart hospital is implemented using CPS framework. Data layer—In this layer, members or entities to analyze the medical data are represented. The entities are patients, laboratories, doctors, pharmacists, and hospitals. The doctors and clinical assistants’ analyze the data stored in cloud for providing the treatment to patients. This layer receives the assured and measured patient’s health record.

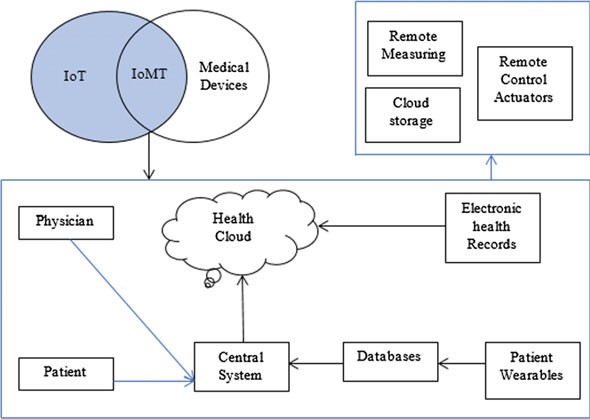
CPS layer—This layer includes aspects and concern of smart hospital. The actual implementation is resided in this layer. The sensors are placed over the patient’s body making each sensor area as a node. The sensor sends the physiological values to the microcontroller, thereby sending to the cloud storage. In cloud, decisions are made whether to provide treatment to patient or not based on the physiological parameters which are termed as CPS decision. Data acquisition happens when the doctor or any clinical assistants access the patient’s data from cloud. After accessing, the doctors or nurses decide the kind of treatment to give to the observed patient. Thus, CPS enables an active interaction between the doctors and patients by enabling a proficient communication and computation model over the network. Hence, CPS provides an assured mechanism or algorithmic concept for implementing smart hospital.

* 1. Internet of Medical Things (IoMT) 29

### Internet of Medical Things (IoMT)

Internet of Medical Things (IoMT) is a technology of connecting the IoT devices with Medicare application in the IT system through embedded networks [[44](#_bookmark131)]. IoMT applies the concept of machine-to-machine communication using the Wi-Fi-enabled devices. In IoMT, embedded devices transfer the health records over the computer networks and store the data in cloud for future analysis. IoMT includes remote monitoring of patient suffering from long-term or chronic diseases such as heart ailment, stroke, and diabetes. IoMT also tracks the patient’s health conditions or orders, patient movement in the hospital or home, and patient’s wearable e-health devices. IoMT collects the medical records and sends to the cloud for the caretakers to analyze the data.

The microcontroller or Wi-Fi-enabled device is connected to the data analytical dashboard and to the sensors equipped with patient’s bed. These sensors and dash- boards which observe the physiological parameters can be deployed as IoMT tech- nology. IoMT comprises both software and hardware architecture which is used as a foundation for future low-power and wireless communication of wearable devices. These wearable devices are placed on the patient’s body and communicate non- invasively through body tissues. IoMT allows the following features in healthcare system, namely (a) monitoring the patient remotely and storing the physiological parameters in cloud observed by wearable sensors; (b) controlling the actuators remotely deployed in the patient’s body; (c) machine-to-machine communication enabling the system to function as closed-loop application. In Fig. [3.2](#_bookmark80), the basic concept of IoMT architecture is represented along with the components included to design the IoMT model. IoMT is described when the medical devices are compro- mised or connected to the IoT technology by framing as Internet of Medical Things. As standard IoT, IoMT also contains physical space consists of hardware boards and sensors, where the observed sensor data is transferred to the central database in the form of electronic health record. The EHR format of patient’s physiological parameter allows for efficient monitoring of any remote sensing model. In the above IoMT representational diagram, it is explained that the data from the patient’s wear- able devices is exchanged to the database through Wi-Fi-enabled microcontrollers. The records are then stored in the central storage system such as cloud server. The data in the cloud server is stored in the form of Electronic Health Records hav- ing the standard health parameter values. The patient’s health value stored in the cloud server is termed as Health Cloud allowing for remote monitoring and sensing of patient’s health condition at anywhere. Remote monitoring also includes remote control actuators, remote measuring of physiological parameters, and cloud storage having electronic medical records.



**Fig. 3.2** IoMT basic architecture

### Result and Discussion

To evaluate the use of OCDA approach, we need to evaluate the communication cost using the data collected through various sensors such as pulse rate sensor, tempera- ture and humidity sensor, ultrasonic sensor, and blood pressure sensor. Intel Edison Arduino boards are used to measure the data transmission cost of virtual resources. The boards are connected to the same Wi-Fi network, and each board executes one resource at a time. At a time, each board communicates with the other by sending 1000 sequential requests. Once the request is sent, the resource must wait for the acknowledgment. Likewise, the resources wait and proceed with the future requests. Figure [3.6](#_bookmark86)a, b displays the time taken to send the data of various sizes using OCDA approach. Before experimenting with the real-time data, let us test our approach with the available benchmark dataset. For our experimentation, we have obtained diabetes dataset from [[45](#_bookmark132)]. There are almost 20 data fields comprised of insulin dose, blood glucose measurement, hypoglycemic symptoms, and so on. Each data value is organized in such a way that one resource is communicated at a time. To avoid over- loading, the resources are scheduled in a proper way and communicated

### Conclusion

The main objective of implementing CPS is to monitor the patient suffering from chronic diseases effectively to overcome the severity in patient’s health condition. The modeling also involves active observation of patient’s physiological parameters such as body oxygen (BO) level, heartbeat (HB) rate, blood glucose (BG) level, and blood pressure (BP) level. The observed values are then uploaded to the cloud server and analyzed using some defined framework for determining the patient’s body condition. CPS categorizes the implementation into three spaces, namely phys- ical space—It consists of hardware components such sensors and microcontroller; cyber-space—It includes the actual computation where the sensor data that is trans- ferred over the network is store in cloud for computing purposes; social interaction

34 3 A Novel Framework for Healthcare Monitoring System …

space—In this space, the actual interaction between machine and machine occurs, and it also involves the interaction between the patient and doctors or clinical assis- tants. In critical situations, the data analyzed is the cloud and fixes a status that if value is higher than the threshold value, then a notification is sent to the doctors or clinical assistants’ mobile devices to ensure the patient’s condition. Hence, it is observed that the patient’s health condition data is very sensitive and important to handle while transferred over the network module. In this paper, a CPS framework is developed for remote monitoring of patient along with some security or safety measures that are also implemented to protect the electronic health records from cyber-attacks. In the proposed method, a novel OCDA approach is used for preserving the data by dividing the data into chunks or blocks and sending it to different cloud servers. The proposed architecture uses the concept of blockchain methodology for distributed storage system. The approach uses the concept of distributed data storage with the perception of single data server as similar to interleaved memory. The objective is to split the sensor data key value and store the key values into chunks or blocks each holding different parameter value. These block values are then transferred to different cloud servers to avoid data breaching or any other cyber-attacks. The storage of key values in different cloud server allows efficient data storage system. The experimen- tation shows us a clear difference on response time when the data size is increased. Moreover, the medical resource data can be well organized and stored properly in the defined method. Further researches are on progress for well-organized medical data storage and speed data access.

### References

1. C. Konstantinou, M. Maniatakos, F. Saqib, S. Hu, J. Plusquellic, Y. Jin, Cyber-physical systems: a security perspective, in 2015 20th IEEE European Test Symposium (ETS), pp. 1–8 (2015)
2. V. Buzduga, D.M. Witters, J.P. Casamento, W. Kainz, Testing the immunity of active implantable medical devices to CW magnetic fields up to 1 MHz by an immersion method. IEEE Trans. Biomed. Eng. **54**(9), 1679–1686 (2007)
3. J. Zhang, N. Xue, X. Huang, A secure system for pervasive social network-based healthcare. IEEE Access **4**, 9239–9250 (2016)
4. Y.M. Huang, M.Y. Hsieh, H.C. Chao, S.H. Hung, J.H. Park, Pervasive, secure access to a hier- archical sensor-based healthcare monitoring architecture in wireless heterogeneous networks. IEEE J. Sel. Areas Commun. **27**(4), 400–411 (2009)
5. M.A. Khan, K. Salah, IoT security: review, blockchain solutions, and open challenges. Future Gener. Comput. Syst. (2017)
6. X. Yue, H. Wang, D. Jin, M. Li, W. Jiang, Healthcare data gateways: found healthcare intelli- gence on blockchain with novel privacy risk control. J. Med. Syst. **40**(10) (2016)
7. Z. Zheng, S. Xie, H. Dai, X. Chen, H. Wang, An overview of blockchain technology: architec- ture, consensus, and future trends, in *Proceedings—2017 IEEE 6th International Congress on Big Data, BigData Congress* (2017), pp. 557–564
8. M. Samaniego, R. Deters, blockchain as a service for IoT, in *Proceedings*—*2016 IEEE Inter- national Conference on Internet of Things; IEEE Green Computing and Communications; IEEE Cyber, Physical, and Social Computing; IEEE Smart Data, iThings-GreenCom-CPSCom- Smart Data 2016*, (2017), pp. 433–436

**Chapter 4**

# An IoT Model to Improve Cognitive Skills of Student Learning Experience Using Neurosensors

##### Abhishek Padhi, M. Rajasekhara Babu, Bhasker Jha and Shrutisha Joshi

**Abstract** In a classroom, during the teaching period, there is a need of analyzing the basic level of understanding in a student in order to improve the teaching method for better teaching experience in a class. This model is required so that the concentration level of students can be monitored in a systematic manner, and after analyzing the concentration level, proper steps can be taken to improve it accordingly. This model presents designing an apparatus to record EEG waveform and then compare it to pre- recorded reading of different mind states using Arduino Brain Library and processing IDE to obtain the result as the emotion of the student. In the proposed method, EEG waveforms are obtained, which are the mathematical representation of the emotions; on analyzing those emotions, we can understand the level of concentration of the student in an efficient manner. It does not use any guesswork, and hence, the results obtained are reliable, and required actions can be taken on basis of that.

**Keywords** EEG Waveform Arduino Brain Library Processing IDE Electrodes · Neurosensor · IoT · Cognitive library

· · ·

* 1. **Introduction**

#### Needs or Requirements

The education of students plays a vital role in the development of society, and so, student learning experience is a big area of interest. Proper learning or capturing of subjects taught in a class is majorly based on the states of brain and how students understand the concept taught. There are already a variety of methods which include interpreting the facial expression of the students, examining their hand–eye coordina- tion and asking them question regarding class activities and studies. These methods have several drawbacks. They are inefficient as they involve very basic ways such as interpreting the facial expression/hand–eye coordination of the students and then changing teaching methods accordingly, but this method is not very reliable as they are very ambiguous and may lead one to take inappropriate actions to reach the goal [[1](#_bookmark148)].

This model concentrates on the brain activity which is the real factual data to analyze the state of mind to give appropriate result. The current state of brain or brain activity can be analyzed and studied with the help of brain waves since brain emits electrical waves which are called brain waves. These brain waves can be cap- tured using EEG neurosensors, and they are referenced as EEG waveforms. The EEG signals are complex, multi-component periodic curves that are composed of high amplitudes which range between 1 and 50 Hz waves. These amplitude ranges are hence divided in eight parts, namely delta (1–3 Hz), theta (4–7 Hz), low alpha (8–9 Hz), high alpha (10–12 Hz), low beta (13–17 Hz), high beta (18–30 Hz), low gamma (31–40 Hz), high gamma (41–50 Hz). These states define the current state of mind such as relaxed, attentive, sleeping. These states can be recorded using EEG sensors, and hence, in this model, Mind Flex headset is used which contains Neu- roSky chip which captures the brain waves and then converts these waves according to the frequency and amplitude to these eight parameters [[2](#_bookmark149)].

In the proposed model, Mind Flex headset is used with Arduino UNO as hard- ware to collect input from brain and get the data in the form of these parameters onto Arduino IDE. In Arduino IDE, this input will be analyzed and results will be extracted using the Arduino Brain Library which is a library, especially for brain waves analysis. Hence, these parameters will be analyzed and then processed with the help of processing IDE which will give the final result of how much a student has understood a particular topic.

#### Why This Work?

The ability to concentrate in class despite distraction, lack of interest or fatigue is an art that requires a lot of self-discipline and hard work. It is very difficult for one to focus on a specific task when there are multiple things going around, mind anyhow wanders away. Although the concentration time of a person and the factors that distract the person will vary from one person to another, hence, we can say that the actions that are required to improve the concentration level of a person will also vary and methodology will change for every other individual.

Talking about other feelings such as saying truth in difficult situations or express- ing emotions, some people find themselves in a difficult situation where they deviate from the truth or they prefer to hide their emotions because it makes them nervous or they are just not capable of expressing their emotions even if they want to. These above-mentioned situations are required to be handled in an exclusive manner for every individual [[3](#_bookmark150)] (Fig. [4.1](#_bookmark135)).

A. About the NeuroSky chip:

* 1. Introduction



**Fig. 4.1** Overview of NeuroSky headset

A few points of interest from the data given by the organization are as follows:

* + 1. It is obtained from the product family, ThinkGear AM, where A corresponds to ASIC and M corresponds to module.
    2. Next, demonstrate the number of the chip which is TGAM1, Revision Number 2.3.
    3. So, the dimensions of the module are round about 29.9 mm 15.2 mm 2.5 mm (1.1 in. 0.60 in. 0.10 in.).

× ×

× ×

* + 1. The module weight is 130 mg (0.0045 oz).
    2. The working voltage of this module is about 2.97–3.63 V.
    3. The maximum input noise which the module can possibly filter is 10 mV from peak to peak. We will then measure our noise and will ensure that the noise is in the module range for ideal outcomes.
    4. Maximum power consumption of the module is 15 mA @3.3 V. We will check these quantities with a multimeter and will measure every parameter. It will be enjoyable to check these values by our own.
    5. ESD protection of the gadget is 4 kV for the contact discharge and 8 kV for the air discharge. It is critical to take note of the fact that electrostatic discharge is the flow of electricity between two charged items caused by contact, dielectric breakdown, or electric shock. It is principally caused by a static charge of two bodies. The friction-based electricity can be built by induction or tribocharging (certain materials turn out to be electrically charged after they came into contact with various material) [[4](#_bookmark151)].
    6. The gadget can communicate serially with 9600, 1200, 57,600 bps baud rate. There are arrangement pins by the assistance of which, we can change the baud rate.

How does headset work? What does the ear movements signify? The working involves the following mentioned steps:

Step 1: The electrical impulses are sensed by the EEG sensor placed on forehead because of the neurons which are bombarded in the brain giving of the waves.

Step 2: The headset captures brainwave data, filtering out the environmental dis- turbances in the form of electrical noise, and interprets it with NeuroSky’s attention and meditation algorithms.

Step 3: This mental state is then presented in the form of ear movements and shared.

From these ear movements, headset senses the attention and presents it in the form of ears shooting straight up. In relaxed phase, the ears droop down. Also, during highly focused and relaxed mode, the ears wiggle up and down [[5](#_bookmark152)].

In P3,

1. 1 is GND “ ”

−

1. 2 is VCC “+”
2. 3 is RXD “R”
3. 4 is TXD “T” In P4,
4. 1 is VCC “+”
5. 2 is GND “−” In P1,

In this paper titled “A brain eeg classification system for the mild cognitive impairment analysis,” A. Nancy, Dr. M. Balamurugan, and Vijaykumar S. observes that electroencephalogram (EEG) signals is a demanding and challenging task, and hence, some of the classification techniques which includes discrete wavelet trans- form (DWT), discrete cosine transform (DCT), and fast Fourier transform (FFT) are frequently used in the existing works across the world [[29](#_bookmark173)]. Yet, it had a few drawbacks; for example, the previously mentioned strategies speak to the structure estimations of input EEG signal in light of separated component of eye flickering estimation dataset [[30](#_bookmark174)]. To conquer this issue, this work proposed another framework: integrated pattern mining (IPM)—support vector machine (SVM) for the EEG signal order. In this work the EEG signals as input are pre proposed by using multiband spectral filtering and hence the specifications of the filtered signals are obtained. From that point onward, the ordinary or unusual mind states are grouped from the given flag utilizing SVM arrangement method. From the obtained output, the execu- tion of the proposed IPM-SVM technique is assessed and compare in terminologies like False Rejection Rate (FRR), False Acceptance Rate (FAR), Genuine Acceptance Rate (GAR), exactness, review, affectability, specificity and precision. The principle favorable position of this proposed framework is that it precisely characterizes the anomalous classification of intellectual weakness by enhancing the characterization execution of the signal classification framework [[15](#_bookmark159)].

In this paper titled “Discriminating different color from EEG signals using Interval-Type 2 fuzzy space classifier (a neuromarketing study on the effect of color to cognitive state),” Arnab Rakshit and Rimita Lahiri analyze that color perception is one of most important cognitive features in human brain and hence different cog- nitive activity is led by different color. Since color plays an important role, hence in this paper color-based recognition is shown using EEG sensors. Neuromarketing research based on color stimuli is a considerable tool for marketing research. It con- sidered to consider first the color detection in mind in order to get different colors from EEG sensors. EEG sensors are hence used as a market based research tool in which the focus remains to detect various colors using the EEG sensors and thus the mentioned stimulus were obtained. This paper includes an interval type II fuzzy space classifier to differentiate between different stimuli which are considered for the ongoing experiment. Research says that red color has maximum classification rate and minimum is yellow. In this paper, red, yellow, blue, and green are the four colors to be considered for judgment

### Proposed Method

This project involves the calculation of EEG wave on various aspects such as low alpha, high alpha, low beta, high beta, delta, theta, low gamma, high gamma, and then categorizing them according to their digital value. This digital value of the various aspects is then matched with the experimentally calculated value, and then, the level of emotions of the person is analyzed. By this way, accurate results can be known such as concentration level, level of distraction, attention level, and hence, proper steps can be taken to improve the cognitive learning of the student [[17](#_bookmark161), [18](#_bookmark162)].

Conclusion that can be included from the value of all these waves [[19](#_bookmark163)] is as follows:

##### Gamma Waves

If high: anxiety, stress, high arousal

If low: depression, ADHD, learning disabilities

Optimal: cognition, information processing, binding senses, learning, perception, REM sleep.

##### Beta Waves

If high: anxiety, high arousal, inability to relax, stress, adrenaline If low: daydreaming, ADHD, depression, poor cognition Optimal: memory, conscious focus, problem solving [[20](#_bookmark164)].

##### Alpha Waves

If high: inability to focus, daydreaming, too relaxed If low: high stress, anxiety, insomnia, OCD [[21](#_bookmark165)] Optimal: relaxation [[20](#_bookmark164)].

##### Theta Waves

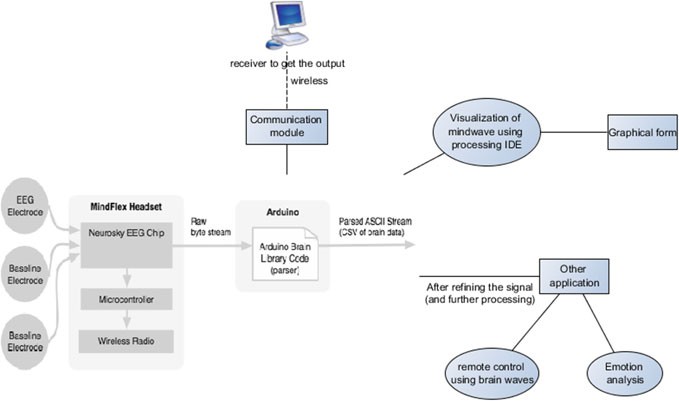
If high: depression, ADHD, hyperactivity, impulsivity, inattentiveness If low: poor emotional awareness, anxiety, stress

Optimal: creativity, emotional connection, intuition, relaxation.

##### Delta Waves

If high: learning problems, brain injuries, inability to think, severe ADHD

If low: inability to rejuvenate body, poor sleep, inability to revitalize the brain Optimal: immune system, natural healing, restorative/deep sleep [[19](#_bookmark163)].

46 4 An IoT Model to Improve Cognitive Skills …

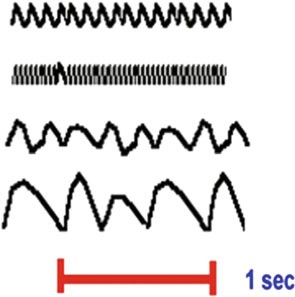
**Fig. 4.3** Flow diagram representation of our model

The procedures followed while carrying out the work involve the various stages mentioned in the Fig. [4.3](#_bookmark140).

1. It starts with recording EEG waves using the NeuroSky chip. The analysis of the EEG waves is made using Arduino Brain Library. Next, various aspects of EEG waves are measured such as theta, delta, low alpha, high alpha, low beta, high beta, high gamma, low gamma.
2. Now, the visualization of the EEG waves is made using the processing IDE.
3. Next, the analyzed and recorded values are sent to other computers or databases.
4. After further analysis of various values, mind waves can be used to predict other emotions of an individual; it can also be used to control remote controls, etc. [[22](#_bookmark166)].

The implementation also involves the use of the Wi-Fi and Bluetooth modules to increase or widen the range/area and increase the reach of the model to far-off distances, including the places of the needy, so that it can help to uplift and develop them.

In Fig. [4.4](#_bookmark142), the four waveforms are of alpha, beta, theta, and delta, respectively. It is a sample which shows how the waveforms of various aspects of EEG waves look in a time interval of 1 s, when they are generated in a computer or any hardware.

**Fig. 4.4** Sample waveform representation of various

aspects of EEG

### Result and Discussion

There is a fact related to Table [4.1](#_bookmark143) that higher value of low beta shows that the person has a higher level of concentration and attention [[23](#_bookmark167)]. So, from Table [4.1](#_bookmark143), we can conclude that concentration level is generally low in the morning (because of various reasons such as say feeling sleepy) which then gradually increases, and after a few hours, it decreases again (might be due to exhaustion) (Fig. [4.5](#_bookmark145)).

A. Output Dataset

In Fig. [4.6](#_bookmark147), the representations are A1-signal strength, A2-attention, A3-meditation, A4-delta, A5-theta, A6-low alpha, A7-high alpha, A8-low beta, A9-high beta, A10- low gamma, A11-high gamma [[24](#_bookmark168), [25](#_bookmark169)].

### Conclusion

On feasibility background, currently, the headset is the most expensive part of our model because the headset has not officially been launched in India, so to buy the headset, one has to import it from America. But if it is made available in India, the expenses to make the model can reduce to its quarter price, which would make it very economical in comparison to heavy EEG machines available in hospitals. We found out that this model provides very accurate reading, and based on these readings, one can differentiate between different emotions of an individual. Its use can be expanded to various fields and occupations such as psychiatric, neurological, neurotherapy, medical, education [[26](#_bookmark170)]. So, with the inclusion of Brain Library in Arduino [[27](#_bookmark171)], the Mindflex can be used for various applications such as giving instruction and controlling hardware devices such as prosthetic arm [[28](#_bookmark172)], wheelchair. It does not need any prior knowledge/experience or any specialization to operate this device; hence, it can be used even by a layman. So, using the communication modules, live status of the mind of an individual can be analyzed from distance; hence, it can be used in long-distance learning. Using IoT with this module, its application can be expanded into vast areas and the data will not be bounded by the physical distance.

### References

* + 1. K.V. Thomas, A.P. Vinod, A study on the impact of neurofeedback in EEG based attention- driven game (2016)
    2. J. Katona, I. Farkas, T. Ujbanyi, P. Dukan, A. Kovari, Evaluation of the NeuroSky MindFlex EEG headset brain waves data (2014)
    3. P. Sri Sai Chaitanya, S. Agnadi, A Review on improving technologies in wireless communica- tions (2013)
    4. M. Chaumon, D.V. Bishop, N.A. Busch, A practical guide to the selection of independent components of the electroencephalogram for artifact correction. J. Neurosci. Methods **250**, 47–63 (2015)
    5. <https://www.engineersgarage.com/articles/understanding-neurosky-eeg-chip-detail-part-213>
    6. P.S. Yalagi, T.S. Indi, M.A. Nirgude, Enhancing the cognitive level of novice learners using effective program writing skills (2016)

**Chapter 5**

# AdaBoost with Feature Selection Using IoT to Bring the Paths for Somatic Mutations Evaluation in Cancer

##### Anuradha Chokka and K. Sandhya Rani

**Abstract** Nowadays, the research in bioinformatics helps in finding out numerous ways in storing, managing organic information, and developing and analyzing the computational tools for better understanding. So far, much of the research has been carried out to overcome the difficulties in experimental methods while storing vast amounts of the data in different sequencing projects. In this process, many of the computational methods and clustering algorithms were brought to light in the past to diminish blocks between newly sequenced gene and genotypes by applying identi- fied jobs. The latest specific applications invented in bioinformatics are paving way for more advancement by adding developments in machine learning and data min- ing fields. Because of a large quantity of applications acquired by various feature encoding methods, the existing classification results remained inadequate. Hence, the present study is intended to create awareness among the readers on the various possibilities available in finding somatic mutations by using machine learning algo- rithm, AdaBoost with feature selection, a classification in various feature selection techniques with their applications, and detailed explanation on the distinct types of advanced bioinformatics applications. This study presents the statistical metric- based AdaBoost feature selection in detail and how it helps in decreasing the size of the selected feature vector, and it explains how the improvement can be attributed through some measurements using performance metrics: correctness, understanding, specificity, paths of mutations, etc. The present study suggests some IOT devices for early detection of breast cancer.

**Keywords** Bioinformatics Somatic mutations Machine learning AdaBoost Feature selection · IoT

· · ·

* 1. **Introduction**

It is found in previous investigations that tumor samples in cancer patients display several types of genetic defects which have been infected to the mankind during somatic mutation developments from a normal cell condition. Somatic mutations are accumulated in every cell continuously where the effect of one gene is dependent on the presence of one or more modifier genes. This phenomenon is known as epistasis which plays a vital role in molecular evaluation and while limiting the continuous flow of mutation built up. The size of epistasis connections relies on the fitness function of the space in all genotypes. Therefore, it can be stated that the genotypes noticed in the growth samples square measure the results of a varied set of alteration methods envisioning a posh fitness landscape. The somatic mutation forms helps to understand the progressive ways of the developments in cancer. This Significant information specifies how somatic mutations are influenced by the epistatic gene interactions among them. In this scenario, it is highly difficult to pull out how cancer is developing in the unknown fitness landscapes conditions, and examining such a huge data with hundreds of intermittently genes is one of the highly demanding areas in the research of bioinformatics. Since the existing computational methods are unable to overcome setbacks in the path of success, there is an urgent need to develop appropriate methods for the advancement in medicine. The AdaBoost algorithm is now a well-known and deeply studied method to build ensembles of classifiers with very good performance with high accuracy. This study also focuses on IoT a latest technology which is being adopted in the healthcare systems to detect and diagnose the cancers earlier to save the lives of the people as well as the money of the victims.

#### 

#### AdaBoost Technique

AdaBoost (adaptive boost) is a machine learning classification technique, which builds ensembles of classifiers in order to give good result [[1](#_bookmark196)]. This algorithm creates group of weak classifiers which form sequentially to give final classifier. Weight will be given to each set of training data, and the weight from weak classifier will be updated to the next classifier. This process will be continued until the last training data tests to get the final strong classifier. The weight for the weak classifier will be zeros likewise the accuracy [[1](#_bookmark196)]. When the weight increases, the accuracy for that classifier will also increase. Each training instance will be reweighted according to its misclassification by the previous classifiers.

#### Feature Selection Techniques

Here, we tend to report organic process progression methods for neoplasm samples from body part, brain tumor, respiratory organ, and female internal reproductive organ cancer problem persons (patients). EPPs area unit is derived for using a machine learning machine technique to reconstruct ancestral genotypes from observed growth genotypes, referred to as feature selection techniques (FSTs) [[2](#_bookmark197)]. The main purpose of the feature selections is manifold, and the very first crucial point is: (a) It is used to avoid overfitting, and it also improves the model performance in a great way,

* 2. Introduction 53

that is prediction of performance at intervals. The study of each case is supervised classifications and to have better even good cluster search or detection at intervals. The second point is (b) to have faster even cheaper models. The third point is (c) to be grateful for deeper underlying and neat processes which generates the data. As AdaBoost algorithm is having advantages over existing techniques, In this paper this algorithm is considered to develop a classification model. However, the benefits of FSTs are worth full. FS techniques dissent from or to each other at intervals of the approach, and this search is incorporated at intervals of the feature subsets. The FS techniques are broadly classified into three important categories; those models are filter ways, wrapper ways, and an embedded way. The filtering principles are used to assess how relevant the selected data at the intrinsic properties. In majority of cases, FS score is calculated and small scoring choices unit of measurement are removed. Afterward, this type of choices is taken as input to the classification formula for assessment. The Feature selection is a onetime process, and these can be used for development and analysis of different classifiers.[[3](#_bookmark198)]. That is, each feature is taken into consideration on an individual basis.Therefore, to address the ignoring the feature dependencies among the variables, filter principles were introduced. By rendering this method, the analysis of a specific set of choices is obtained by testing a specific classification model to a specific classification formula. The second method, wrapper ways utilises various searching algorithms to extract significant features. The third method of Feature selection techniques is termed as embedded technique.

#### Internet of Things (IoT)

Presently, the world is at the Internet of Things World Forum, we’ve been hearing a great deal about the transformational estimation of the Internet of things (IoT) crosswise over numerous enterprises—producing, transportation, horticulture, bril- liant urban communities, retail, back, and medicinal services. Such a large number of new arrangements are in plain view that helps associations either spare or profit. In any case, in medicinal services, IoT can really accomplish more than that; it can possibly spare lives.

#### Challenges in Sequencing

Single cell sequencing (SCS) has so many recent and advanced methodologies which have come into picture to expose the growth of a tumor unsimilarity and well- endowed resolution at very high level. Even though there are multiple benefits in SCS, it has many of its own problems. The foremost problem is noise which is iden- tified in different genotypes [[4](#_bookmark199)]. It is also observed in several instances that these genotypes include false +ve and false –ve mutations with missing values. Because of this persistent problem of noise, the clustering methods were unable to recognize the subpopulations in the sequenced cell and even a simpler task like mapping cells to clones has become a difficult issue to resolve. The second issue occurs in unnoticed subpopulations. Because of partiality in sampling, under sampling, or in the disap- pearances of these subdomains, the exemplificated cells are used to correspond to the division of the subdomains which emerges in the lumps total life history. Hence, approaches are required to understand the unnoticed ancestral subpopulations to find out the development of a tumor exactly.

### Existing Models

Navodit Misra expressed that BML is a predicated model on a probabilistic biological process path from traditional genotype to other neoplasm genotype that incorporates a nonzero chance. The model BML initially estimates the chance that the selected combinations of mutations that reach extreme degree in each one cell population that is been evolved from a standard cell gene and can in the long run attain a neoplasm cell gene [[5](#_bookmark200)]. Here, these users can talk over with it called evolutionary genes G. The probability of these genes G, i.e., P(G) is the process of genes which makes equals the total of path chances for each mutation source from which the tradi- tional genotype that it passes through the tip as a neoplasm genotype.

### Methodology

Many feature selection strategies are there in literature in order to perform dimen- sionality reduction for terribly huge data. Feature choice strategies provide North American country the simplest way of reducing computation time, up prediction performance, and a far better understanding of the information in machine learning or pattern recognition applications. In this paper, we offer a summary of a number of strategies gift in the literature. The target is to produce a generic introduction to variable elimination which might be applied to a good array of machine learning issues. We tend to concentrate on filter, wrapper, and embedded strategies. We tend to conjointly apply a number of the feature choice techniques on commonplace datasets to demonstrate the pertinence of feature selection techniques.

#### Redundancy and Relevancy Analysis Approach

Despite the spectacular achievements within the current field of feature choice, we have a tendency to observe nice challenges arising from domains admire genomic microarray analysis and text categorization wherever knowledge might contain tens of thousands of options. Initial of all, the character of high spatiality of knowledge will cause the questionable downside of curse of spatiality. Secondly, high-dimensional knowledge usually contains several redundant options. Each theoretical analysis and empirical proof show that besides impertinent options, redundant options addition- ally have an effect on the accuracy [[7](#_bookmark202)], speed, and vibrant of machine learning algorithms and sought to eliminate yet. Existing feature choice ways principally exploit two approaches: individual analysis and set analysis. In individual analysis rank options in keeping with their importance in differentiating instances of vari- ous categories and might solely take away impertinent options as redundant options doubtlessly have similar rankings. Ways of set analysis look for a minimum set of options that satisfies some goodness live and might take away impertinent options yet as redundant ones. However, among existing heuristic search methods for set analy- sis, even greedy sequent search that reduces the search house from O(2N) to O(N2) will become terribly inefficient for high-dimensional knowledge [[3](#_bookmark198)]. The restrictions of existing analysis clearly counsel that we should always pursue a special framework of feature choice that permits economical analysis of each feature connectedness and redundancy for high-dimensional knowledge.

#### Feature Redundancy and Feature Relevancy

In normal, feature selection has concentrated so far in studying the relevant features. Even though latest study has focused on the presence of feature redundancy along with its results, there is some work to be accomplished in the explicit treatment of feature redundancy [[7](#_bookmark202)]. With a view to achieve the target, this study presents a tradi- tional method of feature relevance and also explains the reason why it is impossible to feature redundancy to deal with alone and also introduces a suitable formal definition for feature redundancy that leads to the removal of redundant features effectively. On the base of the definitions given by John, Kohavi, and Pfleger, the feature redundan- cies are divided into three categories. They are strong relevant features, weak relevant features, and irrelevant features. Let F be a full set of features, Fi a feature, and Sai Fa {Fai}. These three categories could be regularized in the following manner. Generally, these categories are in relation to feature correlation. It has been agreed that two features are redundant when their values are correlated fully (e.g., features F2 and F3). In practical situations, it is very difficult to fix feature redundancy where a feature is related to other sets [[3](#_bookmark198)]. Hence, we propose a feature redundancy to formulate a method to explicitly recognize and remove redundant features.

= −

#### Defining a Framework of AdaBoost Technique with Feature Selection

To classify the given datasets accurately, we use the advance machine learning tech- nique called AdaBoost (adaptive boost). It is machine learning’s boosting technique which helps us to combine multiple weak classifiers into a final strong classifier [[1](#_bookmark196)]. To remove redundant features, the modern feature selection techniques should depend upon the method for the subset assessment that completely deals the feature redundancy with the support of feature relevance [[2](#_bookmark197)]. These modern techniques are able to show improvement in the results when we apply both combinations. However, the main drawback lies in this technique; unbearable computational cost in the search of subset made them weak while handling a huge amount of dimensional data. In view of finding out a suitable method for this issue, the study presents a new approach in AdaBoost with feature selection that completely overcomes the drawbacks in the previous methods by introducing an explicitly handling feature redundancy process. The main goal of present study is to find out somatic mutations and bringing dif- ferences between strong relevance and irrelevant redundancy. Identification of these differences can be achieved when the definition of relevance is completely under- stood and by the achievement of the following two steps [[3](#_bookmark198)]. First, we find out cancer mutations using AdaBoost technique by classifying given datasets. Second, by removing redundant features and subsets by considering relevant features of the relevance analysis the advantage of the modern process is dividing the redundancy and relevance in the analysis process.

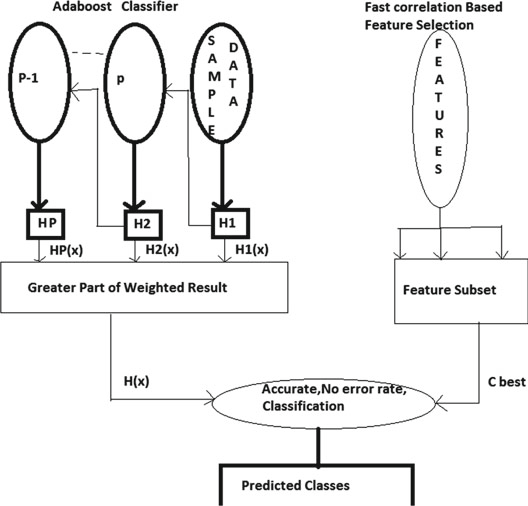
#### Schematic Representation for the Proposed Algorithm

See Fig. [5.1](#_bookmark188).

#### Algorithm and Analysis

The function used in fourth line of the Algorithm 1 h → {0, 1} is described as h(a) =

1, when a ≥ 0, and h(a) = 0 when a < 0. The classifier Hp(xi) and yi represent the



**Fig. 5.1** Schematic representation for the algorithm AdaBoost with feature selection

values { 1, +1}, and the term errrate p is the weighted error rate. The final classifier is the summation of all the weak classifiers with sign [[1](#_bookmark196)]. Thereby, it finally classifies the result with great accuracy as mentioned in ([1](#_bookmark182)). The approximation methodology for connectedness associated redundancy analysis conferred before is completed by using an algorithmic specified by the authors in [[2](#_bookmark197)] ([2](#_bookmark183)) choosing predominant options from relevant ones. Using ([1](#_bookmark182))–([4](#_bookmark185)) for a knowledge set, it calculates the uncertainty of symmetrical (US) feature price for every feature.

−

#### IoT Wearables to Detect Cancer

The innovations in IOT related to wearable’s, remote checking, execution helps to enhance well - being and well ness of the people and also to detect bosom dis- ease. With inserted temperature sensors, this new sort of wearable innovation tracks changes in temperature in bosom tissue after some time. It utilizes machine learning and prescient investigation to recognize and group unusual examples that could show beginning period bosom growth.

##### AdaBoost Classification Algorithm.

Input: dataset M = {M1, M2, …, M*N* } with *Mi* = (*xi* , *yi* )

* 1. Methodology 59

where *xi* k and *yi* { 1, +1} P, the highest no. of classifiers

∈ ∈ −

Result: A classifiers H: K → {−1, +1}

* + 1. Initialize the weights *W* (1) = 1 , i ∈{1, …, N}, and set p = 1; learner on **M**

using

* + 1. While p ≤ P do;

*i N*

( *p*)

* + - 1. Run weak weights *Wi* yielding classifier *Hp*:K → {−1, +1}
      2. Compute *errratep* = .*N W* (*p*)*h*(−*yi* ; *Hp*(*xi* ));

*i* − *i* =1

*i*

2

*errrate*

* + - 1. Compute *bp* = 1 log(

*errratep* ) /\* Weak learner **weight \*/**

*p*

* + - 1. For every sample i = 1, …, L, update the weight

*V* (*p*) = *w*(*p*) exp(−*by Hp*(*xi* ))

*i*

*i*

* + - 1. Renormalize the weight: Calculate *Sp* = .*N Vj* and for i = 1, …, N; *W p*+1 =

*i*

*V* (*p*)/*Sp*;

* + - 1. Increase the iteration counter: p p++

←

* + - 1. End of while

*j* =1 *i*

* + - 1. H*(*X*)* = sign(.*p bc Hc*(*X* ))

*c*=1

Algorithm 1. AdaBoost Structure Learning.

1. **Fast Correlation Feature Selection Algorithm Input**: *C* ( *f*1*, f*2*, ..., f N,d* ) /**\*A** training Data Set \* /

*α*, /\* predefined Threshold \*/

Output = *Cbest* /\* Final Best Subset \*/

1. Begin
2. For i 1 to N do begin

=

1. Calculate U S*i,d* for F*i*
2. If(US*i,d* ≥ *α*) 1
3. Append *fi* to *Clist*
4. End;
5. Order *C* 1 in descending US*i,d* value

*list*

1. F*v* = get First Element *C* 1 ;

*list*

1. Do begin
2. *fw* = get Next Element (*C* 1

*list*

, f*v*)

1. if( *fw* <> NULL)
2. do begin
3. *f* 1 = *fw*

*w*

1. If (US*v,w* ≥ US*w,d*) 1
2. Remove *fw* from *Clist*
3. *fw* = get Next element (*C* 1

*list*

*, f* 1)

1. else *fw* = get Next Element (*C* 1

*w*

*list*

*, fw*)

1. End until ( *fw* == NULL)
2. *fw* = get next element (*C* 1

*list*

1. end until ( *fv* == NULL);

, *fv*);

60 5 AdaBoost with Feature Selection Using IoT …

**Table 5.1** Feature set considered for fast correlation-based feature selection

|  |  |
| --- | --- |
| Sl. no. | Features |
| 1 | Married status |
| 2 | Basis of diagnosis |
| 3 | Age |
| 4 | Occupation |
| 5 | Topography |
| 6 | Received surgery |
| 7 | Morphology |
| 8 | Received radiation |
| 9 | Stage |
| 10 | Survivability (classes) |

### 5.4 Conclusions

This paper discusses classification of cancer mutations using the fundamental details of AdaBoost algorithm and fast correlation feature selection technique. As a pop- ular machine learning method, adaptive boost technique uses are in many different kinds of real-time applications, and feature selection’s fast correlation method is helpful for finding the subset of features to reduce the redundancy; thereby combin-

62 5 AdaBoost with Feature Selection Using IoT …

ing both AdaBoosts with fast correlation technique, we get accurate classification of finding the somatic mutations in human beings. The requirements for the specific redundancy analysis are identified for tumor analysis in cancer and formulated a feasible definition to feature redundancy and selected features lineage, occupation, and stage of cancer to find mutations fast. By this, we were able to present a new type of approach which is suitable for the analysis of genes and related approach that practices the redundancy analysis. This feature rule is applied and examined in detailed experiments. The results of this AdaBoost with feature choice are confirmed by machine learning algorithms. Modeling the organic process events resulting in cancer and robustness scenery of cancer prisons guarantees the advanced applica- tions in the analysis of cancer. On such item is the IoT inner wears. The gadget is a bra or other suit with implanted sensors that recognize little temperature changes in bosom tissue after some time. Enthused with the conceivable outcomes, Cisco supported a film in view of the innovation, called Detected.

### References

1. A. Ferreira, Boosting algorithms: a review of methods, theory, and applications, in *Instituto de Telecomunicacoes* (Portugal, Chap. 3, April 2012), pp. 7–9
2. J. Thongkam, AdaBoost algorithm with random forests for predicting breast cancer survivability. IEEE Explore, 978-1-4244-1821-3/08 (2008)
3. L. Yu, Feature selection for high-dimensional data: a fast correlation-based filter solution, in *Proceedings of the Twentieth International Conference on Machine Learning* (Washington DC, 2003)
4. Y. Wang, X. Fan, A comparative study of improvements pre-filter methods bring on feature selection using microarray data. Health Inf. Sci. Syst. **2**, 7 (2014)
5. E.M. Ross, F. Markowetz, OncoNEM inferring tumor evolution from single-cell sequencing data. Ross and Markowetz Genome Biol. **17**(69) (2016)
6. N. Misra, E. Szczurek, Inferring the paths of somatic evolution in cancer, vol. 30 (Oxford University Press, May 2014), pp. 2456–2469
7. C.S. Attolini et al., A mathematical framework to determine the temporal sequence of somatic genetic events in cancer. Proc. Natl. Acad. Sci. U.S.A. **107**, 17604–17609 (2010)
8. Yu. Lei, Efficient feature selection via analysis of relevance and redundancy. J. Mach. Learn. Res. **5**, 1205–1224 (2004)
9. J. Thongkam, Breast cancer survivability via adaboost algorith ms: Australian computer society, in *Conferences in Research and Practice in Information Technology*, vol. 80 (2008)

**Chapter 10**

# A Computational Approach to Predict Diabetic Retinopathy Through Data Analytics

##### Ashraf Ali Shaik, Ch Prathima and Naresh Babu Muppalaneni

**Abstract** Making use of estimating methods in the field of medicine has been the powerful research recently. Diabetic retinopathy is a retinal disease which causes huge blindness. Recurrent screening for prior disease detection has been a highly labor force—and resource—powerful process. So computerized diagnosis of these diseases through estimating methods would be a great remedy. Through this paper, a novel estimation strategy for computerized disease prognosis is suggested, which uti- lizes retinal image analysis and mining methods to accurately differentiate between the retinal images as normal and affected. Eighteen feature relevance and three vari- ations algorithms were analyzed and used to identify the contributing features that provided better conjecture results.

**Keywords** Diabetic retinopathy Classification CTree SVMC Bagging and boosting

· · ·

**10.1 Introduction**

Exploration in neuro-scientific medicine shows that excessive pressure and glucose levels are a significant reason behind several critical health problems. Many of this astigmatism can lead to other problems in a variety of body parts. This paper gives awareness of Diabetic Retinopathy, being a classic disorder in the retina of the atten- tion triggered due mainly to Diabetes [[1](#_bookmark338)] creating in decrease of vision and the second option being associated with a boost of pressure in the interest in the end creates harm to the optic nerve system [[2](#_bookmark339)]. Diabetic retinopathy is asymptomatic in your beginning stage and results say that treatment may be useful only once clinically diagnosed in the beginning. Standard screening process having risky of the condition might help discover the condition at the beginning stage. Finding retinal disorders in the charac- teristics made by screening process program is a moment task. Systematic diagnosis of the problem from the retinal data is an important part of recurring research [[3](#_bookmark340)–[6](#_bookmark341)]. In this research, we place attention to automated analysis medical prognosis of eyesight malocclusions (diabetic retinopathy) wherein the data is generally cleaned and statistical, GLCM organized, and bins setup measurements are computed. The classifier categorizes the choroid image to the problem category to which it belongs. This kind of research is geared toward robotic diagnosis of diabetic retinopathy through data and feature distinction using data gold exploration techniques.

Information mining maintains great prospective for the healthcare industry to allow health systems to systematically use information and statistics to identify lack- ing and best methods that improve care and reduce costs [[7](#_bookmark342), [8](#_bookmark343)]. Descriptive analytics describe what has happened. Predictive analytics [[9](#_bookmark344)] predicts what will happen. The disease estimations play a natural part in data gold mining. Info mining tools have recently been developed for effective examination of medical information, to be able to help clinicians for making better diagnosis to be treated purposes.

The two traditional algorithms, gold mining techniques proposed are CTree, sup- port vector machine, also to classify the person with and without diabetes predicting the results Baye’s Theorem is proposed.

1. CTree (Conditional Inference Tree): It is the decision tree known as CTree, it is a non-variable class of r-Trees in tree-structured regression models into a well-learned theory of conditional procedures.
2. SVMC (Support Vector Machine Classifier): SVMCs also support vector clas- sifiers [[1](#_bookmark338)] are supervised learning models with associated learning codes that analyze data used for classification and regression evaluation. Given a collec- tion of training examples, each proclaimed as owned by one or the other of two categories, an SVMC training algorithm builds an unit that assigns new illustrations to one category or the other, rendering it a non-probabilistic binary linear class (although methods such as plot scaling exist to work with SVMC in a probabilistic classification setting). SVMC model is a representation of the examples as points in space, mapped so types of the independent categories are divided with a clear space that can be as extensive as it can be. New examples are then mapped into that same space and expected to participate in a category based on which aspect of the space they fall.

Furthermore, to performing linear category, SVMCs can proficiently do a nonlin- ear classification using precisely what is called the kernel trick, with ought a shadow of doubt maps their inputs into high-dimensional feature spaces.

Data is not tagged, checked learning is difficult, and an unsupervised learning approach is essential, which makes an effort to find natural clustering of the infor- mation to types and then map new data to these created groups. The clustering standards which gives an improvement to the support vector devices is referred to as support vector clustering [[2](#_bookmark339)] and is often [citation needed] used in professional applications either when data is not tagged or when only some data is defined as a preprocessing for a category pass.

Bayes Theorem: Bayes divisers are a family group of simple probabilistic clas- sifiers based with strong (naive) independence presumptions of the features Bayes classifiers are highly ductile, requiring a quantity of parameters linear in the num- ber of value predictors in a learning problem. Maximum-likelihood training can be

* 1. Introduction 107

achieved by accessing a closed-form expression, [[1](#_bookmark338)] which takes step-wise time, so expensive iterative approximation is used for most various divisors.

#### Steps in Algorithm

1. Each single data test is viewed by an n size vector, Y sama dengan (Y1, Y2, …, Yn), describing sizes made on quality from n values, correspondingly A1, A2, An.
2. Assume that there are n classes, C1, C2, …, Cn. Given an anonymous data test,

Y. No classes designate, the classifier will predict that X is one of the classes getting the best posterior possibility, trained and only when

P*(*Ca|Y*) >* P*(*Cb|Y*)* for all those a *<*= b *<*= n and b! = a

Thus, we increase P(Ca|Y). The category Ca that P(Ca|Y) is strengthened is known as the utmost posteriori hypothesis. Simply B theorem

P*(*Ca|Y*)* = (P(Y|Ca)P(Ca))*/*P*(*Y*)*

1. P(Y) is regular for all those classes; only P(Y|Ca)P(Ca) must be maximized. When the probabilities are not known, then it is often guessed that the similarly is likely, i.e., P(C1) P(C2) sama dengan… P(Cn), and we would increase P(Y|Ca). Normally, P(Y|Ca) · P(Ca) is increased. Do not forget that the course probabilities may be approximated by P(Ca) sa/s, where sa is the amount of training {samples of category Ca}, and s is the overall range of data training samples.

=

= =

The manuscript is designed in this way. Section [10.2](#_bookmark327) targets the materials and methods. Section [10.3](#_bookmark330) shows the functionality measures while Sect. [10.4](#_bookmark332) presents tools used and experiment results. Section [10.5](#_bookmark336) concludes the research work.

### Methodology

#### Description of Dataset

[The dataset prevailed at https://archive.ics.uci.edu/ml/machine-learning-databases/ 00329/messidor\_features.arff. This dataset contains 1151 records with nineteen fea-](https://archive.ics.uci.edu/ml/machine-learning-databases/00329/messidor_features.arff) tures acquired from the Messidor image set to predict whether an image as indications of diabetic retinopathy or not.

108 10 A Computational Approach to Predict Diabetic Retinopathy …

#### Attribute Information

1. Binary output of quality assessment is 0 bad class 1 sufficient class.

= =

1. Binary response to pre-screening, where one indicates severe retinal furor and 0 its shortage.

(2–7) Results of Messidor attribute detection. Each quality value is known for the number of Messidor attributes available at the confidence levels first 0.5, …, 1, respectively.

=

(8–15) contain the same information as (2–7) for issues. However, as issues are represented by a group of points rather than the number of cells constructing the abrasion, these features are normalized by dividing the number of abrasion with the size of the ROI to compensate different image sizes.

1. The Euclidean distance of the center of the macula and the middle of the optic disk to provide important information about the patient’s condition. This feature is also normalized with the size of the ROI.
2. Diameter of the optic disc.
3. Binary reaction to the AM/FM-based classification.
4. Class ingredients label: 1 contains symptoms of DR (accumulative ingredients label for the Messidor classes 1, 2, 3), zero no indications of DR.

=

=

Because Bayes algorithm will not grant regular data type, all the ideals in the dataset are cared for as categorical.

#### Cross-Validation

Cross-validation (CV) is the typical data retrieving way of examining performance of classification strategy. Mainly it is used to judge the challenge rate of an learning strategy. In CV, a dataset is portioned in n folds up, where each can be used for screening and the rest can provide for training. The process of screening and training is repeated n times so that all rupture of collapse can be used once for verification.

#### Classification Matrix

Classification matrix is a creation tool which is often used to provide the precision of the divisors in classification. It can be used showing the connections between final results and expected classes.

The entries in classification matrix have next meanings in framework of our research:

* p is the amount of accurate estimations value is negative,
* q is the amount of inaccurate estimations value is positive,
* r is the amount of inaccurate estimations value is negative,
  1. Methodology 109
* s is the amount of accurate estimations value is positive.

#### Bagging and Boosting

Bootstrap aggregating, also called bagging, is a machine learning ensemble meta- algorithm designed to increase the stableness and accuracy of machine learning algorithms used in statistical classification and regression. Additionally, it reduces variance and helps to avoid over-fitting.

Boosting is a machine learning ensemble meta-algorithm for mostly reducing tendency, and also variance in supervised learning, and a family of machine learning algorithms which convert fragile learners to strong ones. Algorithms that achieve speculation boosting quickly became simply known as “boosting.”

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model by means of an ensemble of weak prediction models, typically decision trees.

### Performance Measures

The concerned algorithm’s ability to produce exact results was determined in this paper by the consumption of four characteristics: accuracy, sensitivity, specificity, and classification matrix.

#### Accuracy

In statistical scrutiny of binary classification, the F-measure is a way of computing a test’s correctness. This considers both accuracy a and the recall r test to compute the report. a is number of correct excellent results divided by the amount of all excellent results, and r is amount of accurate excellent results divided by the amount of good results which should have recently been delivered.

#### Sensitivity

This is regarded as the opportunity that the relevant survey can be retrieved by the worried query. Also known as recall or true positive rate, it could be found in binary or nominal datasets but cannot be 100% depended after as a way of measuring developed.

#### Specificity

Additionally, it is called true negative rate and actions the percentage of negatives that are correctly thought as a result. Specificity belongs to the test’s ability to effectively identify patients without a condition.

#### Classification Matrix

A classification matrix is a table that is often used to describe the performance of a classification model (or “classifier”) on a group of test data for which the actual classes are known.

### Tools Used and Results Discussion

The diabetic retinopathy dataset includes features extracted from the Messidor image set to predict whether an image contains signs of diabetic retinopathy or not. It includes 1151 instances with twenty attributes. In this dataset, all the attributes are numeric except the category label. The attributed used here are about quality assess- ment, Euclidean distance, size of the optic disk, and so on.

To obtain and calculate the test results of Bayes classifiers, the dataset instances are classified, i.e., size: 1151 is divided into training set (75%) 863 and test set (25%) 288.

=

=

To anticipate the occurrences of medical issues of individuals, it is extremely much necessary to examine the preceding data, utilizing data mining methods, especially for classification goal. As informed above, the classification methods that are taken into account are Bayes classifier, support vector machine classifier, decision tree, bagging, and boosting. In this paper, the research has been completed utilizing the wide open source data mining tool R.

The repository considered mostly includes 1151 instance of all the datasets and applied all the preferred classification algorithms as shown in Table [10.1](#_bookmark334). Out of this table, it could be recognized that the retinopathy expose higher accuracy of the diabetic dataset, it was examined in three specific ways: considering 1151 is divided into training set (75) 863 and test set (25%) 288 and last but not least with the help of nominal features. The results of the methods are shown in Table [10.2](#_bookmark335).

= =

Finally, if we convert all the characteristics to nominal, the information becomes lossy which is not strongly recommended in the medical field as shown in Table [10.2](#_bookmark335). Because of this, it is extremely suggested maximum change of information into nominal should not exceed 50% of the complete data of the concerned attributes and this modification should be under the advice of doctors.

* 1. Conclusion 111

**Table 10.1** Classification matrix

|  |  |  |
| --- | --- | --- |
| Actual | Predicted | |
|  | −ve | +ve |
| −ve | p | q |
| +ve | r | s |

**Table 10.2** Comparison of results on the classification algorithms for diabetic retinopathy datasets over different instances and their data types. R was the tool used

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Algorithm | Accuracy | Sensitivity | Specificity | Classification matrix | |  |
| Naïve Bayes | 0.6146 | 0.5642 | 0.7714 | *Prediction* | Reference | |
| 123 | 16 |
| 95 | 54 |
| Decision tree | 0.6042 | 0.8633 | 0.3624 | *Prediction* | Reference | |
| 120 | 95 |
| 19 | 54 |
| Support vector machines | 0.5868 | 0.4029 | 0.7584 | *Prediction* | Reference | |
| 56 | 36 |
| 83 | 113 |
| Bagging | 0.6597 | 0.6835 | 0.6376 | *Prediction* | Reference | |
| 95 | 54 |
| 44 | 95 |
| Boosting | 0.5729 | 0.3669 | 0.7651 | *Prediction* | Reference | |
| 51 | 35 |
| 88 | 114 |

### 10.5 Conclusion

This kind of study evidently demonstrates the results are used for the information mining techniques of problem in medical directories.

In this paper, decision support machine classifier system was well suited for diabetic retinopathy. The machine can offer as training tool for medical students. Also, it can be heading be big hands for doctors. The machine can be further increased and extended; it can assimilate other medical properties besides in the Table [10.2](#_bookmark335) particular, and yes it can be integrated other gold mining techniques. Constant data can be utilized rather than just nominal data.

112 10 A Computational Approach to Predict Diabetic Retinopathy …

### References

1. J. Han, M. Kamber, *Data Mining Concepts and Techniques* (Morgan Kaufman Publishers, 2006)
2. S.-C. Liao, I.-N. Lee, Appropriate medical data categorization for data mining techniques. Med. Inform. **27**(1), 59–67 (2002)
3. V. Balakrishnan, M.R. Shakouri, H. Hoodeh, L. Hakso-Soo, Predictions *Using Data Mining and Case-based Reasoning: A Case Study for Retinopathy*, vol. 63 (World Academy of Science and Technology, 2012)
4. R. Klein, B.E.K. Klein, S.E. Moss, T.Y. Wong, L. Hubbard, K.J. Cruickshanks, M. Palta (2004)
5. The relation of retinal vessel caliber to the incidence and progression of diabetic retinopathy: XIX: The wisconsin epidemiologic study of diabetic retinopathy, archives of ophthalmology, vol. 122, pp. 76–83
6. Ch-L. Chan, Y.Ch. Liu, Sh-H. Luo, Investigation of diabetic micro vascular complications using data mining techniques, in *International Joint Conference on Neural Networks (IJCNN 2008)* (2008)
7. X. Fang, Are you becoming a diabetic? A data mining approach, in *Sixth International Confer- ence on Fuzzy Systems and Knowledge Discovery* (2009)
8. M. Shouman, T. Turner, R. Stocker, Using data mining techniques in heart disease diagnosis and treatment, in *Japan-Egypt Conference on Electronics, Communication and Computers* (2012)
9. M. Salehi, N.M. Parandeh, A. Soltain Sarvestani, A.A. Savafi, Predictioned breast cancer surviv- ability using data mining techniques, in *2nd International Conference on Software Technology and Engineering (ICSTE)* (2010)