The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board
Chairman
Nikhil R. Pal, Indian Statistical Institute, Kolkata, India
e-mail: nikhil@isical.ac.in

Members
Rafael Bello Perez, Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba
e-mail: rbellop@uclv.edu.cu

Emilio S. Corchado, University of Salamanca, Salamanca, Spain
e-mail: escorchado@usal.es

Hani Hagras, University of Essex, Colchester, UK
e-mail: hani@essex.ac.uk

László T. Kóczy, Széchenyi István University, Győr, Hungary
e-mail: koczy@sze.hu

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA
e-mail: vladik@utep.edu

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ctlin@mail.nctu.edu.tw

Jie Lu, University of Technology, Sydney, Australia
e-mail: Jie.Lu@uts.edu.au

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico
e-mail: epmelin@hafsamx.org

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: nadia@eng.uerj.br

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland
e-mail: Ngoc-Thanh.Nguyen@pwr.edu.pl

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: jwang@mae.cuhk.edu.hk

More information about this series at http://www.springer.com/series/11156
Jugal Kalita · Valentina Emilia Balas
Samarjeet Borah · Ratika Pradhan
Editors

Recent Developments in Machine Learning and Data Analytics

IC3 2018

Springer
Recent Developments in Machine Learning and Data Analytics is a collection of research findings of the Second International Conference on Computing and Communication. The conference is centered upon the theme of machine learning and data analytics.

The works incorporated in this volume can roughly be divided into three parts, namely data analytics, natural language processing, and soft computing. The following section contains a brief information about the various contributions to this volume.

In the first paper, Aski et al. provide an architectural overview of IoT-enabled ubiquitous healthcare data acquisition and monitoring system for personal and medical usage powered by cloud application. The next one is also an IoT-based paper where Ishita Chakraborty, Anannya Chakraborty, and Prodipto Das discuss sensor selection and data fusion approach for IoT applications. An overview of Hadoop MapReduce, Spark, and scalable graph processing architecture is provided by Talan et al. in their paper. On the other hand, Hore et al. discuss a machine intelligence-based approach to analyze social trend toward girl child in India. Analyzing class-imbalanced data is found to be a difficult task always. In the next paper, an improvement in boosting method for class-imbalanced datasets is discussed by Kumar et al. In their paper, Ambika Choudhury and Deepak Gupta provide a survey on medical diagnosis of diabetes using machine learning techniques. Another classification-related issue on diabetes data is presented by Santosh Kumar Majhi in his research work. Findings on the usefulness of big data technologies for statistical homicide dataset are discussed by Askew et al. The next research work discusses a journal recommendation system through content-based filtering approach. Another content-based filtering and collaborative filtering technique for movie recommendation is provided by Bharti et al in their research work.

The next 13 contributions are from the domain of natural language processing. The first work of such kind discusses a word-sense disambiguation for Assamese language. In addition to this, other two works are found for Assamese language where Choudhury et al. present a context-sensitive spelling checker for Assamese
language and Mirzanur Rahman and Shikhar Kumar Sarma discuss a hybrid approach to analyze the morphology of an Assamese word. The next work presents an aptitude question paper generator and answer verification system. Ghosh et al. discuss affinity maturation of homophones in a word-level speech recognition in their work. This follows a discussion on feature map reduction in CNN for handwritten digit recognition. Multilingual text localization from camera-captured images is presented by Dutta et al. The technique is based on foreground homogeneity analysis. Jajoo et al. propose script identification from camera-captured multiscr ipt scene text components in their research work. Again, Khan et al. present a distance transform-based stroke feature descriptor for text–non-text classification.

The volume also includes a contribution on emotion mining. This follows two works on Nepali language, where Thapa et al. discuss a finger spelling recognition for Nepali sign language and Yajnik et al. present a work on parsing in Nepali language.

A number of contributions are found which can roughly be categorized under the domain of soft computing. Mishra et al. discuss a BFS-NB hybrid model in intrusion detection system, whereas Saikia et al. propose an effective alert correlation method in their research work. An application of ensemble random forest classifier for detecting financial statement manipulation of listed Indian companies is discussed by Hiral Patel and co-authors. Another security-related paper is discussed on dynamic shifting genetic non-adjacent form elliptic curve Diffie–Hellman key exchange procedure for IoT heterogeneous network. This follows few classification-related works, where Vijaya et al. discuss fuzzy clustering with ensemble classification techniques to improve the customer churn prediction in telecommunication sector and Ahmed et al. propose a technique to remove the bottleneck of FP tree. Additional works include elephant herding algorithm, improved K-NN algorithm through class discernibility and cohesiveness, a reduction-level fusion of PCA and random projection for high-dimensional data, a stable clustering algorithm for mobile ad hoc networks (MANETs), and interval-valued complex fuzzy concept lattice and its granular decomposition.

Umesh Gupta and Deepak Gupta discuss their findings on twin-bounded support vector machine based on L2-norm. A work to perform natural scene labeling using neural networks is presented by Das et al. Kalvapalli et al. present their findings on a city-scale transportation system using XGBoost. A selfish controlled scheme in an opportunistic mobile communication network is presented by Moirangthem Tiken Singh and Surajit Borkotokey. Few quality works on image processing techniques are also included in this volume. The first part of these papers discusses a fusion-based underwater image enhancement using weight map techniques. The next work proposes an algorithm for automatic segmentation of pancreas histological images for glucose intolerance identification. An edge detection technique using ACO with PSO for the noisy image is discussed by Aditya Gautam and Mantosh Biswas in their research work. Another work discusses improved convolutional neural networks for hyperspectral image classification. Mohanraj et al. present a neural network-based approach for face recognition. A method on automated vision inspection system for cylindrical metallic components is proposed by
Govindaraj and co-authors in their research work. The volume also includes a research work on gene selection of microarray datasets. A case study on geo-statistical modeling of remote sensing data for forest carbon estimation is presented by Kumar et al. Finally, it includes a study of DC–DC converters with MPPT for stand-alone solar water pumping.

IC3 2018 represents a global forum for research on computational approaches to learning. It includes mostly the current works and research findings from various research laboratories, universities, and institutions and may lead to the development of market-demanded products. The works report substantive results on a wide range of learning methods applied to a variety of learning problems. It provides solid support via empirical studies, theoretical analysis, or comparison to psychological phenomena. The volume includes works to show how to apply learning methods to solve important application problems as well as how machine learning research is conducted.

The volume editors are very thankful to all the authors, contributors, reviewers, and the publisher for making this effort a successful one.

Colorado Springs, USA
Arad, Romania
Sikkim, India

Jugal Kalita
Valentina Emilia Balas
Samarjeet Borah
Ratika Pradhan
Contents

IoT Enabled Ubiquitous Healthcare Data Acquisition and Monitoring System for Personal and Medical Usage Powered by Cloud Application: An Architectural Overview .......................... 1
Vidhaydhar J. Aski, Shubham Sanjay Sonawane and Ujjwal Soni

Sensor Selection and Data Fusion Approach for IoT Applications ...... 17
Ishita Chakraborty, Anannya Chakraborty and Prodipto Das

An Overview of Hadoop MapReduce, Spark, and Scalable Graph Processing Architecture ................................. 35
Pooja P. Talan, Kartik U. Sharma, Pratiksha P. Nawade and Karishma P. Talan

Analyzing Social Trend Towards Girl Child in India: A Machine Intelligence-Based Approach ................................. 43
Sirshendu Hore and Tanmay Bhattacharya

Improvement in Boosting Method by Using RUSTBoost Technique for Class Imbalanced Data ............................... 51
Ashutosh Kumar, Roshan Bharti, Deepak Gupta and Anish Kumar Saha

A Survey on Medical Diagnosis of Diabetes Using Machine Learning Techniques ................................................. 67
Ambika Choudhury and Deepak Gupta

How Effective Is the Moth-Flame Optimization in Diabetes Data Classification ....................................................... 79
Santosh Kumar Majhi

Evaluating Big Data Technologies for Statistical Homicide Dataset ...... 89
Roland Askew, Sreenivas Sremath Tirumala and G. Anjan Babu

Journal Recommendation System Using Content-Based Filtering ...... 99
Sonal Jain, Harshita Khangarot and Shivank Singh
# Contents

Recommending Top N Movies Using Content-Based Filtering and Collaborative Filtering with Hadoop and Hive Framework ........ 109
Roshan Bharti and Deepak Gupta

WSD for Assamese Language .............................................. 119
Pranjal Protim Borah, Gitimoni Talukdar and Arup Baruah

Aptitude Question Paper Generator and Answer Verification System ................................................................. 129
Meghna Saikia, Saini Chakraborty, Suranjan Barman and Sarat Kr. Chettri

Affinity Maturation of Homophones in Word-Level Speech Recognition .............................................................. 137
P. Ghosh, T. S. Chingtham and M. K. Ghose

Feature Map Reduction in CNN for Handwritten Digit Recognition ................................................................. 143
Sinjan Chakraborty, Sayantan Paul, Ram Sarkar and Mita Nasipuri

Multi-lingual Text Localization from Camera Captured Images Based on Foreground Homogenity Analysis ......................... 149
Indra Narayan Dutta, Neelotpal Chakraborty, Ayatullah Faruk Mollah, Subhadip Basu and Ram Sarkar

Script Identification from Camera-Captured Multi-script Scene Text Components ..................................................... 159
Madhuram Jajoo, Neelotpal Chakraborty, Ayatullah Faruk Mollah, Subhadip Basu and Ram Sarkar

Implementation of BFS-NB Hybrid Model in Intrusion Detection System .............................................................. 167
Sushruta Mishra, Chandrakanta Mahanty, Shreela Dash and Brojo Kishore Mishra

Context-Sensitive Spelling Checker for Assamese Language .............. 177
Ranjan Choudhury, Nabamita Deb and Kishore Kashyap

Distance Transform-Based Stroke Feature Descriptor for Text Non-text Classification .................................................. 189
Tauseef Khan and Ayatullah Faruk Mollah

A Hybrid Approach to Analyze the Morphology of an Assamese Word ................................................................. 201
Mirzanur Rahman and Shikhar Kumar Sarma

Netizen’s Perspective on a Recent Scam in India—An Emotion Mining Approach .................................................... 211
Aviroop Mukherjee, Agnivo Ghosh and Tanmay Bhattacharya
Finger Spelling Recognition for Nepali Sign Language ............... 219
Vivek Thapa, Jhuma Sunuwar and Ratika Pradhan

Parsing in Nepali Language Using Linear Programming Problem ..... 229
Archit Yajnik, Furkim Bhutia and Samarjeet Borah

MaNaDAC: An Effective Alert Correlation Method .................. 249
Manaswita Saikia, Nazrul Hoque and Dhruba Kumar Bhattacharyya

Fuzzy Clustering with Ensemble Classification Techniques to Improve
the Customer Churn Prediction in Telecommunication Sector .......... 261
J. Vijaya, E. Sivasankar and S. Gayathri

Interval-Valued Complex Fuzzy Concept Lattice and Its
Granular Decomposition .................................................. 275
Prem Kumar Singh, Ganeshsree Selvachandran and Ch. Aswani Kumar

SbFP-Growth: A Step to Remove the Bottleneck of FP-Tree .......... 285
Shaful Alom Ahmed, Bhabesh Nath and Abhijeet Talukdar

Combining Multilevel Contexts of Superpixel Using Convolutional
Neural Networks to Perform Natural Scene Labeling ................ 297
Aritra Das, Swarnendu Ghosh, Ritesh Sarkhel, Sandipan Choudhuri,
Nibaran Das and Mita Nasipuri

A Fast Algorithm for Automatic Segmentation of Pancreas
Histological Images for Glucose Intolerance Identification ........... 307
Tathagata Bandyopadhyay, Shyamali Mitra, Sreetama Mitra, Nibaran Das,
Luis Miguel Rato and Mrinal Kanti Naskar

Elephant Herding Algorithm for Clustering ............................ 317
Karanjekar Pranay Jaiprakash and Satyasai Jagannath Nanda

Fusion-Based Underwater Image Enhancement by Weight Map
Techniques ................................................................. 327
Pooja Honnutagi, V. D. Mytri and Y. S. Lalitha

Analysis and Prediction of City-Scale Transportation System Using
XGBOOST Technique .......................................................... 341
Sai Prabanjan Kumar Kalvapalli and Mala Chelliah

An Application of Ensemble Random Forest Classifier for Detecting
Financial Statement Manipulation of Indian Listed Companies ........ 349
Hiral Patel, Satyen Parikh, Amit Patel and Abhishek Parikh

Selfish Controlled Scheme in Opportunistic Mobile
Communication Network ................................................... 361
Moirangthem Tiken Singh and Surajit Borkotokey
A Study of DC–DC Converters with MPPT for Standalone Solar Water-Pumping System ............................................. 373
Baishyanar Jana, Shriraj Dhandhukiya, Ramji Tiwari and N. Ramesh Babu

Edge Detection Technique Using ACO with PSO for Noisy Image ................................. 383
Aditya Gautam and Mantosh Biswas

Improved Convolutional Neural Networks for Hyperspectral Image Classification ............................................. 397
Shashanka Kalita and Mantosh Biswas

Automated Vision Inspection System for Cylindrical Metallic Components ............................................. 411
Krithika Govindaraj, Bhargavi Vaidya, Akash Sharma and T. Shreekanth

A Hybrid Model for Optimum Gene Selection of Microarray Datasets ............................................. 423
Shemim Begum, Ashraf Ali Ansari, Sadaf Sultan and Rakhee Dam

Lagrangian Twin-Bounded Support Vector Machine Based on L2-Norm ............................................. 431
Umesh Gupta and Deepak Gupta

An Improved K-NN Algorithm Through Class Discernibility and Cohesiveness ............................................. 445
Rajesh Prasad Sarkar and Ananjan Maiti

Modified Energy-Efficient Stable Clustering Algorithm for Mobile Ad Hoc Networks (MANET) ............................................. 455
S. R. Drishya and Vaidehi Vijayakumar

Ensemble of Convolutional Neural Networks for Face Recognition ............................................. 467
V. Mohanraj, S. Sibi Chakkaravarthy and V. Vaidehi

Clustering High-Dimensional Data: A Reduction-Level Fusion of PCA and Random Projection ............................................. 479
Raghunadh Pasunuri, Vadlamudi China Venkaiah and Amit Srivastava

Dynamic Shifting Genetic Non-adjacent Form Elliptic Curve Diffie–Hellman Key Exchange Procedure for IoT Heterogeneous Network ............................................. 489
M. Durairaj and K. Muthuramalingam

Geo-Statistical Modelling of Remote Sensing Data for Forest Carbon Estimation—A Case Study on Sikkim, Himalayas ............................................. 511
Pradeep Kumar, Ratika Pradhan and Mrinal Kanti Ghose

Author Index ................................................ 529
About the Editors

Jugal Kalita is Professor in the Department of Computer Science at the University of Colorado, Colorado Springs, CO, USA. He completed MS and Ph.D. from the University of Pennsylvania, USA; M.Sc. from the University of Saskatchewan, Canada; and B.Tech. from the Indian Institute of Technology Kharagpur, India. His areas of research are artificial intelligence, bioinformatics, information retrieval, and computer security. He has over 185 journal/conference papers, 2 technical reports, 5 edited volumes, and 4 books to his credit. He is Editor-in-Chief of Posoowa, non-resident Assamese monthly magazine and edited 75+ issues. He is also Editor of Asomi, yearly magazine of Assam Society of America. He is also involved in several non-profit organizations, mostly geared toward the southeastern Himalayan foothills region.

Valentina Emilia Balas is currently Full Professor in the Department of Automatics and Applied Software at the Faculty of Engineering, “Aurel Vlaicu” University of Arad, Romania. She holds a Ph.D. in applied electronics and telecommunications from Polytechnic University of Timisoara. She is the author of more than 250 research papers in refereed journals and international conferences. Her areas of research interests are intelligent systems, fuzzy control, soft computing, smart sensors, information fusion, modeling, and simulation. She is Editor-in-Chief of International Journal of Advanced Intelligence Paradigms (IJAIP) and International Journal of Computational Systems Engineering (IJCSysE), Editorial board member of several national and international journals, and an evaluator expert for national and international projects. She is Director of Intelligent Systems Research Centre in Aurel Vlaicu University of Arad. She served as General Chair of the International Workshop Soft Computing and Applications (SOFA) in seven editions 2005–2016 held in Romania and Hungary. She participated in many international conferences as Organizer, Honorary Chair, Session Chair and Member in Steering, Advisory or International Program Committees. She is Member of EUSFLAT and SIAM and Senior Member of IEEE, Member in TC—Fuzzy Systems (IEEE CIS), Member in TC—Emergent Technologies (IEEE CIS), and Member in TC—Soft Computing (IEEE SMCS). She was Vice-President
Samarjeet Borah is currently working as Professor in the Department of Computer Applications, Sikkim Manipal University (SMU), Sikkim, India. Dr. Borah handles various academics, research and administrative activities. He is also involved in curriculum development activities, board of studies, doctoral research committee, IT infrastructure management etc. along with various administrative activities under SMU. Dr. Borah is involved with three funded projects in the capacity of Principal Investigator/Co-principal Investigator. The projects are sponsored by, AICTE (Govt. of India), DST-CSRI (Govt. of India) and Dr. TMA Pai Endowment Fund. Out of which one is completed and two are undergoing. He is associated with ACM (CSTA), IAENG and IACSIT. Dr. Borah organized various national and international conferences in SMU. Some of these events include ISRO Sponsored Training Programme on Remote Sensing & GIS, NCWBCB 2014, NER-WNLP 2014, IC3-2016, ICACCp 2017, IC3-2018 etc. Dr. Borah is involved with various journals of repute in the capacity of Editor/Guest Editor such as IJSE, IJHISI, IJGHPC, IJIM, IJVCSN, JISys, IJIPT, IJDS etc.

Dr. Ratika Pradhan is working as Professor in the Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology since July 1999. She received Ph.D. from Sikkim Manipal University (SMU) in 2011 and M.E. (CSE) from Jadavpur University, Jadavpur, in 2004. Her areas of research interest are digital image processing, remote sensing, and GIS. She has published 25 journal papers, and 5 conference papers.
IoT Enabled Ubiquitous Healthcare Data Acquisition and Monitoring System for Personal and Medical Usage Powered by Cloud Application: An Architectural Overview

Vidhaydhar J. Aski, Shubham Sanjay Sonawane and Ujjwal Soni

Abstract Modern lifestyle, swift adoption of fast-food diet and various environmental changes causing chronic life-threatening diseases stimulates a real necessity of advancements in built-in Internet technology for remote healthcare. Proposed work discourses an IoT paradigm comprising of Wireless Health Sensors (WHS) that allows us to observe essentially continuous monitorable and/or should be monitored biometric parameters like pulse rate, pulmonary functional quality, blood pressure, body temperature, electro cardio activity, etc., which in turn helps us for self-evaluation and control future severity by predictive analysis via smart healthcare systems in a long run diagnostic procedure by a medical practitioner. The work addresses the development of an Arduino-based all-in-one cost-effective, miniaturized Wireless Intelligent Embedded Healthcare Device (WIEHD) that can provide home-based health services to the patient. The sensors are connected to Arduino to track the status which is interfaced to a display as well as Wi-Fi connection in order to transmit collected data and receive user requests. The parameters can be monitored from a smartphone application. This system is based on a cloud platform and keeps track of the device data on daily basis. This data is shared with doctors through a website where the doctor can analyze the condition of the patient and provide further details online and intimate patient about future severity well in time.

Keywords WHS (wireless health sensors) · Arduino WIEHD (wireless intelligent embedded healthcare device) · NodeMCU IoT · Cloud · Database · Wi-Fi

V. J. Aski (✉) · S. S. Sonawane · U. Soni
Manipal University Jaipur, Jaipur, Rajasthan, India
e-mail: vidyadharjinnappa.aski@jaipur.manipal.edu
S. S. Sonawane
e-mail: shubhamsanjaysonawane@muj.manipal.edu
U. Soni
e-mail: ujjwalsoni@muj.manipal.edu

© Springer Nature Singapore Pte Ltd. 2019
J. Kalita et al. (eds.), Recent Developments in Machine Learning and Data Analytics, Advances in Intelligent Systems and Computing 740, https://doi.org/10.1007/978-981-13-1280-9_1
1 Introduction

Domestic and self-remote healthcare plays a vital role when it comes to the frequent ups and downs in the health of elderly and patients having the chronic illness which needs recurrent health examination. The application spectrum shall be extended to a wide variety of people having numerous medical backgrounds like differently abled people, diabetic, asthmatic, cancerous people, etc. need their daily health status for further precautions. This opens up huge opportunities for advancements in healthcare and medical devices. The system we propose is purely meant for home sitting applications but regardless of that, the same technology can be used in medical industry as well. Although similar technologies have emerged, there is still space for better, open-source, and low-cost solution.

The system is designed in such a way that it is capable of measuring various human body factors. Various sensors are used to determine the health structure of the patient. Basic sensors like DS18B20 [1] digital temperature sensor, MLX90614 [2], noncontact infrared temperature sensor, pulse rate sensor, ECG module, spirometry sensor, and more can be used to collect primary body factors. Complex sensors like those used in body composition meters can also become part of the system to provide specific data like fat, water, bone, and muscle percentage. More accurate sensors can even determine weight, body age, BMI, and resting metabolism. These sensors are to be connected to the Arduino Uno or nanomicrocontroller. Arduino is to be programmed in such a way that it is not to be limited to the number of sensors. The Arduino will mainly act as a sensor manager and help the system to collect the data for further processing. The collected data is further sent to NodeMCU [3]. NodeMCU is another microcontroller with inbuilt ESP8266 [3] Wi-Fi module and MCU-based Arduino architecture. NodeMCU will receive the data from Arduino via serial communication established between them. NodeMCU will act in master mode and will then be used to command Arduino for the sensor data that has been requested by the user. The input to the system or NodeMCU will be over intranet where MCU will act as a server to provide a response to the user requests.

This system is based on a cloud platform and hence the data is to be uploaded on to the cloud. When performing this task, MCU goes to client mode and requests the cloud for upload. The system keeps track of the daily data collected in the past. This data is then shared with doctors through a website where the doctor can analyze the condition of the patient and provide further details online. Analytical platforms like Watson and Matlab can be used to perform data analysis on to the information collected by the system over the cloud. Patients and their respective doctor can communicate with one another on this platform and can go back and view the past data any time they desire. If the system detects any abrupt abnormalities in patient’s any of the health condition like heart rate or body temperature, the system automatically notifies the user about the patient’s status over their Android phones or web portal, and also shows details of the patient in real time over the Internet. The system uses ubiquitous data accessing method in IoT-based information system [3] for management of huge data collected on a daily basis. This will make the system
more efficient with big data management allowing the user to share information with medical experts on a large scale.

Patient’s health parameter data stored in the cloud makes it even more beneficial than maintaining the records in physical format or digital memory storage devices. There are chances that these devices can get corrupt and data might be lost, whereas the cloud storage is more reliable and does have minimal chances of data loss. Thus, this cloud and IoT-based health tracking system effectively uses the Internet to monitor patient’s health status and save lives on time.

2 Methodology

Following methods are used in creating this overall research plan.

2.1 Sensors

The system will consist of mainly three health monitoring sensors: temperature sensor, pulse rate, and ECG sensor. But the design is to be made in such a way that the system will not be limited to few sensors. Depending on the patient’s need, the system will provide options for the body status they are looking for. Other important factors that can be added to the system are the body composition sensors which will provide extensive insight of body and visceral fat, skeletal muscle level, BMI, resting metabolism, and more.

Sensors like DS18B20 [4] digital temperature sensor, SEN-11574 pulse rate sensor, AD8232 heart ECG monitoring sensor, and spirometry sensor are to be used in the system. DS18B20 contains a unique 64-bit code stored in ROM [4]. The least significant 8 bits of the ROM code contains the DS18B20’s 1-Wire family code: 28 h. The next 48 bits contain a unique serial number. The most significant 8 bits contain a cyclic redundancy check (CRC) byte that is calculated from the first 56 bits of the ROM code [4].

SEN-11574 is easy to use pulse rate sensor [5]. The power of SEN-11574 ranges from 3v to 5v [5]. It consists of optical heartbeat sensor circuit, an amplification circuit, and noise cancellation circuit [5] all embedded on the single chip.

An analog sensor requires signal conditioning circuit to interface with MCU for better precision, likewise for ECG sensor AD2832 acts as an internal signal conditioning circuit and many other potential bio-parameters [6]. This AD8232 can also act as high-pass filter that removes low-frequency motion objects. This filter has got large gain as it is tightly coupled with amplifier.
2.2 Microcontroller

The sensors will collect the analog and digital input for the microcontroller. The system will use Arduino nano or Uno for processing the data collected by the sensors. Arduino will be programmed in such a way that it should be capable of taking as many body factors as possible into consideration. This will make the system not limited to few sensors. Arduino will mainly be programmed for handling various sensors. As there is no limitation to the sensors we are using, one microcontroller is fully assigned to perform the hand over the task of data collection.

The data will then be sent to NodeMCU, another microcontroller with Wi-Fi connectivity for further processing via hardware serial communication established between both controllers as shown in Fig. 1. NodeMCU has a built-in ESP8266 module. The module supports standard IEEE 802.11 agreement and complete TCP/IP protocol stack [3]. The module can be added to an existing networking device or built-in a separate network controller as NodeMCU [3]. ESP8266 is high integration wireless SOCs, designed for space- and power-constrained mobile platform designers. It provides unsurpassed ability to embed Wi-Fi capabilities within other systems or to function as a standalone application, with the lowest cost, and minimal space requirement [3]. NodeMCU will be our primary microcontroller. It will communicate with Arduino to fetch the data over a hardware serial communication established between them. Arduino will be in slave mode, while NodeMCU will be in master mode. NodeMCU will also be interfaced with a display which will provide output in real time.

Fig. 1 System architecture
NodeMCU is a low-power device capable of operating on 3.3v input, while Arduino takes up to 5v of power. This makes the system more power efficient.

### 2.3 Internet Connectivity

Internet connectivity plays an important role when it comes to IoT platforms. We will use the ESP8266 Wi-Fi module inbuilt in our NodeMCU. It will deal with all the Internet requests sent by the client. The node will be connected to the home Wi-Fi network. Since the system can be controlled by Android or Web application, the request has to be sent to the system. Same goes for requests that system sends to cloud as in Fig. 1.

When receiving requests, NodeMCU will go into a server mode, providing a response to the requests. The response is nothing but an acknowledgement for the user to use requested sensor and storing collected information in the cloud. NodeMCU will command Arduino for the provision of the data that is to be uploaded to the cloud. When uploading data to the cloud, NodeMCU will be in client mode and request the cloud server to upload collected data to its database. Most of the request from a user to the system will be over the intranet, and hence connectivity will only be limited to local area network when the node is in server mode. As for client mode, MCU will switch over to the Internet to access the cloud.

### 2.4 Cloud

Information collected by the system will be stored in a database located on cloud virtual machine. The database will contain all the biodata of the patient. It will also have information of the doctors. The web application and the online server that drives the system will also be part of the cloud virtual machine as in Fig. 1. Cloud platforms like IBM Bluemix [7] can be ideal for this task. IBM Bluemix is open cloud platform that offers mobile and web developer access to IBM and other software for integration, security, transactions, and other key functions [7]. The IBM IoT service lets our apps communicate with and retrieve data collected by the system [7]. Bluemix’s recipes make it easy to get devices connected to the Internet of Things cloud [7]. The web and Android application can then use real-time and REST APIs to communicate with devices and use the data that sensors had been collecting [7].

### 2.5 Application Development

Depending on the requirement, an Android or web application is to be developed which shall contain both patients and doctors portal. The patient can select their
respective doctor(s) so that they can share their daily health status. The website will contain all the past information of the patient, so that doctor or patient can check and analyze the data any time they want.

Various tools like IBM Watson or Matlab can be used in the portal to provide analysis and visualization of the data that is collected by our system time to time. This will provide better user experience and will also help doctors determine current health condition of the patient.

3 Cloud Platform and Database Management

With the usage of the medical devices being high, collecting a huge amount of sensor data from n number of users and a load of data on the cloud platform demands technologies depending upon big data, hence resulting in a requirement for data management on multitenant basis rather than conventional distributed data acquiring modules.

Multilayer data storing architecture module is used for increased efficiency directly in terms of data fetching and data storing speed. The bottom data layer is a cause of tenant data storage layer, which allows us to store collective tenant databases. Next layer is used as controlled data access layer. This constitutes different mechanisms for controlling resources in order to form effective distributed healthcare model. Among all the layers, top layer being the business layer. Data interpretations, controlling operations, and business operations sharing resources between various data formats will occur in the business layer that coordinates shared data as well as interpreted data. Two separate data interoperable protocols are used for creating isolated nature between shared data and interpreted data. In healthcare applications, patient data is used as an extremely big dataset in hospitals and MQTT protocols provide relatively good encryption to datasets for transporting data in remote places. Because various hospitals provide different cloud infrastructures for patient’s database, it is always better to have a protocol that deals with extracting data from different databases by maintaining isolation between each infrastructure [8].

4 Implementation

The sensors which are to be used must be connected to the Arduino Uno depending on the analog and digital pins. Temperature sensor is like DS18B20 digital temperature sensor. The DQ pin of like DS18B20 is connected to any of the Arduino digital pins. It is also connected to 4.5k register which is then interfaced with VCC of the sensor input and 5v Arduino output as shown in Fig. 2. The GND of the sensor goes to the GND of Arduino. Similarly, analog output of the SEN-11574 pulse rate sensor goes to any one of the Arduino analog pins with an input voltage of 5v.
AD8232 [9] bioelectrical signal acquisition development module and the sensor are most important of all as it will be used for electrocardiography, i.e., ECG visualization. It has got the efficient structure to excerpt, strengthening signal by means of amplification and removal of less biopotential signals. In the case of noisy signal condition occurrence like those designed by distant electrode placement [9], this also enables to have extremely low-power ADC embedded in the microcontroller to obtain output signal power more easily [9]. AD8232 is low-power 3.3v device. It will get its input from Arduino 3.3v pin. An analog output of AD8232 will be connected to an analog pin of Arduino, whereas LO— and LO+ will be going to digital pins of Arduino as shown in Fig. 2 (Fig. 3).

The 6 analog pins and 13 digital pins of Arduino Uno being more than enough for above implementation, many more sensor in future can be added to the system using 4051 Multiplexer. The 4051 is an 8-channel analog multiplexer/DE multiplexer with one common and 8 I/O pins. The common pin is connected to any of the analog pins, whereas sensors are connected to 8 channels. 4051 also have three digital pins—A, B, and C; these pins together will determine which channel is to be used by providing

---

**Fig. 2** Sensor interfaced circuit diagram
Fig. 3  Arduino sensor manager flowchart

Start

Initialize Pin 8 to DS18B20
Initialize Pin 9, 10 to AD8232 A, B
Initialize A0 to pulse sensor
initialize and A1 to AD8232 analog

Read Serial Data

If Serial Available

YES

case 1

DS18B20 Temp Sensor Function

case 2

Pulse Rate Sensor Function

case 3

AD8232 ECG module Function

case 4

Other Sensor Function

Delay 500ms

End

NO
Fig. 4 4051 multiplexer with Arduino Uno

A three-digit binary number from 0 to 7 as of respective channel pin of 4051. Pins A, B, and C will be connected to digital pins of Arduino as shown in Fig. 4.

The Arduino is internally programmed to handle all the sensors at the same time. A programmable switch is written in such a way that depending upon what is the request for sensor available on the hardware serial, nothing but a request from NodeMCU in form of an integer. This commands Arduino to call for a function specific to the requested sensor. Each function of sensors will have its own program depending upon its characteristics. The flowchart of the sensor is shown in Fig. 3.

This is only to assure that the user has a home or personal Wi-Fi and is supposed to provide Wi-Fi name and password to the system. Once the system is provided with the details, it will then store this info on to its memory and start accessing the Internet. Once the connection is established, the users are free to use the system to its full potential. This task only has to be performed once. The Wi-Fi details are to be stored in the memory of NodeMCU. NodeMCU has the memory of 128 KB and flash storage of around 4 MB which is more than enough for this task. This memory can also be used to temporarily store the data fetched when the system is in offline mode or has experienced any problems with the connectivity (Fig. 5).

Once an Internet connection is available, the system will upload the data on to the cloud. The flowchart of NodeMCU is shown in Fig. 6.
An application programming interface will be designed in order to provide information to the available doctors and hospital. Since we are using three-layer database management architecture [8], both shared and isolated databases have to be managed. The shared database will contain all the patient’s daily health and body status. Doctor- and hospital-related information also are to be stored in the shared database. Patients and doctors personal information will be stored in an isolated database. The web application will be developed using JSP technology and for storage purpose, we will be using MongoDB document-oriented database which is ideal for our three-layer architecture.

The analytical tool is a vital feature of the system that can be used for predictive measures. Specifically, IBM Watson analytics is used as an analytical tool. This is an essential cloud service platform which enables us to create various regressive models and design supervised training techniques for the purpose of data analytics [9]. Watson analytical tool allows working with extremely simple and common milieu schemes such as .csv, .xsl, .xlsx, etc. As a use case of this platform, it provides an environment and asserts an offset value of data quality automatically after uploading datasets. Along with this, many other factors are also taken into consideration such as it requires less number of records, filled completeness, and various other qualitative parameters. We created a data acquisition system for collecting biomedical parameters such as ECG, body temperature, pulse, and ECG, among which ECG data is very crucial and is extremely necessary to feed Watson with collected ECG data on daily basis for prediction of abnormalities in the functioning of heart.
Fig. 6 NodeMCU request handle flowchart
**Table 1** Results obtained by using DS18b20 digital temperature sensor

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>Sensor result (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>27.4</td>
</tr>
<tr>
<td>5</td>
<td>28.8</td>
</tr>
<tr>
<td>10</td>
<td>29.5</td>
</tr>
<tr>
<td>15</td>
<td>29.8</td>
</tr>
<tr>
<td>20</td>
<td>30.6</td>
</tr>
</tbody>
</table>

The electrocardiogram is a graphical representation of activities that occur after every contraction and relaxation. This can be detailed and observed using electrodes of ECG sensors, and data processing unit (microcontroller) will take care of mapping voltage variations in accordance with the various records observed through ECG sensors placed on the body in response to the heart function. Basically, cardiac activity can be measured through the observations of voltage differences in some predefined placement positions of electrodes on the skin [9]. ECG is a very basic, less expensive, and vital source of diagnostic information in abnormalities of cardiac function since decades and is used to observe electrical activity of cardiovascular stem. Reading and understanding ECG signal was a real challenge for an engineer and can refer various medical journals to study on ECG waves. Varieties of ECG variations cause and intuit different health statuses of the heart. It acquires signature waves whose variations may indicate different problems of the heart. These changes are referred to as heart rate vulnerability (HRV); this generally refers to time length variations between two heartbeats. HRV dataset [9] can be used in the Watson.

## 5 Result

After developing the system using above implementation, we can test the system for the outputs of the sensors and connectivity.

A test run of the system is performed using DS18B20 one wire digital temperature sensor. The DS18B20 is very precise sensors even capable of measuring temperature ranging from $-55$ to $+125$ °C [1].

We tested the temperature sensor on a person with normal body temperature. The sensor data obtained in every 5 s is shown in Table 1.

The data clearly shows that the temperature gradually increases till the certain point and then it becomes stable. In this case, the data is stable at 31.9 °C. Since the sensor is connected to the palm of the patient, the temperature shown in Table 1 is normal palm temperature. The normal skin temperature of the person is said to be 32–33 °C, while normal external body temperature is 34 °C, i.e., 2 °C more than that of skin. Therefore, in this case, the normal body temperature is $31.9 + 2 = 33.9$ °C or 93.02 °F. Normal internal body temperature most commonly referred by doctors is 37 °C or 100 °F which is 3 °C more than that of internal body temperature in this case, that is, $33.9 + 3 = 36.9$ °C or 98.42 °F, which is close enough to normal
body temperature. The output provided by the system will be in degree Fahrenheit. Internal body temperature will be the factor taken into the consideration but external and skin temperature will also be used for better analysis as shown in Fig. 7.

It can be observed that the sensor provides the required data accurately after at least 30 s. Hence, Arduino is programmed for a delay of 30 s after the sensor is requested by the user, to provide more precise body temperature. The resolution of DS18B20 is also set to max for better results. Similarly, results are to be found in data obtained from pulse sensor and ECG sensor.

6 Future Scope

Currently, the system is developed by keeping domestic applications in mind but it can be expanded to the medical industry as well. The system can be integrated into the medical devices in hospitals and data can be fetched from there and then to the cloud. This will allow us to perform analytics on better scale to get more accurate and realistic results. Same results can also be shared with the hospital database using given three-layer model. IBM Watson and its IoT features can be utilized to full extent as it develops itself, making our Internet of Things, Internet that thinks. As per domestic applications of the system are concerned, modern day home automation features can be added to our system as well. The system can be configured with popular AI like Google Home’s Google Assistant [10] and Amazon Echo’s Alexa [11] to provide more immersive and living experience. As for sensor, with growing sensor technology and fields of application of sensors and sensor networks [12], more medical sensors like blood pressure, insulin and sugar monitor, RBC, WBC, and platelet monitors can also become part of the system as well.
7 Conclusion

In this paper, we proposed an architecture of a healthcare system capable of performing basic medical operations for analysis and visualization of body and health factors. The architecture was designed with two microcontrollers, Arduino Uno and NodeMCU. Arduino Uno was set up to manage sensor network of the model, while NodeMCU was responsible for Internet connectivity and user request handling. We also developed flowcharts of the programs that are to be installed in these microcontrollers. We studied the three-layer database management for ubiquitous data access. Visualization of the Android and web application was also taken into consideration in this study. The cloud structure and various analytical tools like Watson were understood. Observation of results of the sensors in real time was done for better understanding of the data collection mechanism.

Acknowledgements Ethical approval and informed consent All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors, and informed consent was obtained from all individual participants included in the study.

References

4. DS18B20 Programmable Resolution 1-Wire Digital Thermometer, Data Sheet, Maxim Integrated Products, Inc. datasheets.maximintegrated.com


Abstract  The wireless sensors which are used in IoT applications are battery powered, which imposes extreme energy constraints on their operations. Therefore, a fundamental challenge lies in designing a sensor network which can maximize its lifetime. To address this issue in WSNs, data aggregation techniques have come up as a novel approach that minimizes the number of transmissions in the network, which in turn can optimize the network and increase its lifetime. In this paper, we have discussed a simple hierarchical data aggregation method for energy optimized data transmission in wireless sensor networks.

Keywords  Wireless sensor network · Data fusion · IoT · Sensor selection · Path optimization

1 Introduction

In the current years of development of wireless technology and micro-electromechanical systems, development of small, cost-efficient sensors with the ability of sensing and wireless communication is paving a huge path in the domain of wireless communication. This amalgam of wireless technology with sensors has given rise to the wireless sensor network communication. WSNs are deployed in various domains like military surveillance, environmental monitoring, civil applications, etc. Wireless sensor networks have been heavily used for acquiring environmental data such as humidity, temperature, sound, pollution, and so on which are later used