ISSN : 2278-3075

INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND EXPLORING ENGINEERING (IJITEE)

VOLUME NO. 12
ISSUE NO. 3
SEPTEMBER - DECEMBER 2023

ENRICHED PUBLICATIONS PVT. LTD

S-9, IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-47026006

ISSN 2278-3075

International Journal of Innovative Technology and Exploring Engineering (IJITEE)

Aim & Scope

AIM

International Journal of Innovative Technology and Exploring Engineering (IJITEE) is having ISSN 2278-3075 (online), monthly international journal, being published in the months of January, February, March, April, May, June, July, August, September, October November, December by Blue Eyes Intelligence Engineering & Sciences Publication (BEIESP) Bhopal (M.P.), India since year 2012 and processed papers will be forwarded for inclusion in the Scopus database. It is academic, online, open access (abstract), peer reviewed international journal. The aim of the journal is to:

- disseminate original, scientific, theoretical or applied research in the field of Engineering and allied fields.
- dispense a platform for publishing results and research with a strong empirical component.
- aqueduct the significant gap between research and practice by promoting the publication of original, novel, industry-relevant research.
- seek original and unpublished research papers based on theoretical or experimental works for the publication globally.
- publish original, theoretical and practical advances in Computer Science & Engineering, Information Technology, Electrical and Electronics Engineering, Electronics and Telecommunication, Mechanical Engineering, Civil Engineering, Textile Engineering and all interdisciplinary streams of Engineering Sciences.
- impart a platform for publishing results and research with a strong empirical component.
- create a bridge for significant gap between research and practice by promoting the publication of original, novel, industry-relevant research.
- solicit original and unpublished research papers, based on theoretical or experimental works.

SCOPE

International Journal of Innovative Technology and Exploring Engineering (IJITEE) covers all topics of all engineering branches. Some of them are Computer Science & Engineering, Information Technology, Electronics & Communication, Electrical and Electronics, Electronics and Telecommunication, Civil Engineering, Mechanical Engineering, Textile Engineering and all interdisciplinary streams of Engineering Sciences. The main topic includes but not limited to:

1.Smart Computing and Information Processing
  - Signal and Speech Processing
  - Image Processing and Pattern Recognition
  - WSN
  - Artificial Intelligence and machine learning
  - Data mining and warehousing
  - Data Analytics
  - Deep learning
  - Bioinformatics
  - High Performance computing
  - Advanced Computer networking
  - Cloud Computing
  - IoT
  - Parallel Computing on GPU
  - Human Computer Interactions

2.Recent Trends in Microelectronics and VLSI Design
  - Process & Device Technologies
  - Low-power design
  - Nanometer-scale integrated circuits
  - Application specific ICs (ASICs)
  - FPGAs
  - Nanotechnology
  - Nano electronics and Quantum Computing

3.Challenges of Industry and their Solutions, Communications
  - Advanced Manufacturing Technologies
  - Artificial Intelligence
  - Autonomous Robots
  - Augmented Reality
  - Big Data Analytics and Business Intelligence
  - Cyber Physical Systems (CPS)
  - Digital Clone or Simulation
  - Industrial Internet of Things (IIoT)
  - Manufacturing IOT
  - Plant Cyber security
  - Smart Solutions – Wearable Sensors and Smart Glasses
  - System Integration
  - Small Batch Manufacturing
  - Visual Analytics
  - Virtual Reality
  - 3D Printing

4.Internet of Things (IoT)
  - IoT & IoE & Edge Computing
  - Distributed Mobile Applications Utilizing IoT
  - Security, Privacy and Trust in IoT & IoE
  - Standards for IoT Applications
  - Ubiquitous Computing
  - Blockchain-enabled IoT Device and Data Security and Privacy
  - Application of WSN in IoT
  - Cloud Resources Utilization in IoT
  - Wireless Access Technologies for IoT
  - Mobile Applications and Services for IoT
  - Machine/ Deep Learning with IoT & IoE

-
    - Smart Sensors and Internet of Things for Smart City
    - Logic, Functional programming and Microcontrollers for IoT
    - Sensor Networks, Actuators for Internet of Things
    - Data Visualization using IoT
    - IoT Application and Communication Protocol
    - Big Data Analytics for Social Networking using IoT
    - IoT Applications for Smart Cities
    - Emulation and Simulation Methodologies for IoT
    - IoT Applied for Digital Contents
    -

5.Microwaves and Photonics

- Microwave filter
- Microstrip antenna
- Microwave Link design
- Microwave oscillator
- Frequency selective surface
- Microwave Antenna
- Microwave Photonics
- Radio over fiber
- Optical communication
- Optical oscillator
- Optical Link design
- Optical phase lock loop
- Optical devices

6.Computation Intelligence and Analytics

- Soft Computing
- Advance Ubiquitous Computing
- Parallel Computing
- Distributed Computing
- Machine Learning
- Information Retrieval
- Expert Systems
- Data Mining
- Text Mining
- Data Warehousing
- Predictive Analysis
- Data Management
- Big Data Analytics
- Big Data Security

7.Energy Harvesting and Wireless Power Transmission

- Energy harvesting and transfer for wireless sensor networks
- Economics of energy harvesting communications
- Waveform optimization for wireless power transfer
- RF Energy Harvesting
- Wireless Power Transmission
- Microstrip Antenna design and application
- Wearable Textile Antenna
- Luminescence
- Rectenna

8.Advance Concept of Networking and Database

- Computer Network
- Mobile Adhoc Network
- Image Security
- Application Artificial Intelligence and machine learning in the Field of Network and Database
- Data Analytic
- High performance computing
- Pattern Recognition

9.Machine Learning (ML) and Knowledge Mining (KM)

- Regression and prediction
- Problem solving and planning
- Clustering
- Classification
- Neural information processing
- Vision and speech perception
- Heterogeneous and streaming data
- Natural language processing
- Probabilistic Models and Methods
- Reasoning and inference
- Marketing and social sciences
- Data mining
- Knowledge Discovery
- Web mining
- Information retrieval
- Design and diagnosis
- Game playing
- Streaming data
- Music Modelling and Analysis
- Robotics and control
- Multi-agent systems
- Bioinformatics
- Social sciences
- Industrial, financial and scientific applications of all kind

10.Advanced Computer networking Computational Intelligence

- Data Management, Exploration, and Mining
- Robotics
- Artificial Intelligence and Machine Learning
- Computer Architecture and VLSI
- Computer Graphics, Simulation, and Modelling
- Digital System and Logic Design
- Natural Language Processing and Machine Translation
- Parallel and Distributed Algorithms
- Pattern Recognition and Analysis
- Systems and Software Engineering
- Nature Inspired Computing
- Signal and Image Processing
- Reconfigurable Computing
- Cloud, Cluster, Grid and P2P Computing
- Biomedical Computing
- Advanced Bioinformatics
- Green Computing
- Mobile Computing
- Nano Ubiquitous Computing
- Context Awareness and Personalization, Autonomic and Trusted Computing
- Cryptography and Applied Mathematics
- Security, Trust and Privacy
- Digital Rights Management
- Networked-Driven Multicourse Chips
- Internet Computing
- Agricultural Informatics and Communication
- Community Information Systems
- Computational Economics, Digital Photogrammetric
- Remote Sensing, GIS and GPS
- Disaster Management

11.Communications

- Microstrip Antenna
- Microwave
- Radar and Satellite
- Smart Antenna
- MIMO Antenna
- Wireless Communication
- RFID Network and Applications
- 5G Communication
- 6G Communication

12.Algorithms and Complexity

- Sequential, Parallel And Distributed Algorithms And Data Structures
- Approximation And Randomized Algorithms
- Graph Algorithms And Graph Drawing
- On-Line And Streaming Algorithms
- Analysis Of Algorithms And Computational Complexity
- Algorithm Engineering
- Web Algorithms
- Exact And Parameterized Computation
- Algorithmic Game Theory
- Computational Biology
- Foundations Of Communication Networks
- Computational Geometry
- Discrete Optimization

13.Software Engineering and Knowledge Engineering

- Software Engineering Methodologies
- Agent-based software engineering
- Artificial intelligence approaches to software engineering
- Component-based software engineering
- Embedded and ubiquitous software engineering
- Aspect-based software engineering
- Empirical software engineering
- Search-Based Software engineering
- Automated software design and synthesis
- Computer-supported cooperative work
- Automated software specification
- Reverse engineering
- Software Engineering Techniques and Production Perspectives
- Requirements engineering
- Software analysis, design and modelling
- Software maintenance and evolution
- Software engineering tools and environments
- Software engineering decision support
- Software design patterns
- Software product lines
- Process and workflow management
- Reflection and metadata approaches
- Program understanding and system maintenance
- Software domain modelling and analysis
- Software economics
- Multimedia and hypermedia software engineering
- Software engineering case study and experience reports
- Enterprise software, middleware, and tools
- Artificial intelligent methods, models, techniques
- Artificial life and societies
- Swarm intelligence
- Smart Spaces
- Autonomic computing and agent-based systems
- Autonomic computing
- Adaptive Systems
- Agent architectures, ontologies, languages and protocols
- Multi-agent systems
- Agent-based learning and knowledge discovery
- Interface agents
- Agent-based auctions and marketplaces
- Secure mobile and multi-agent systems
- Mobile agents
- SOA and Service-Oriented Systems
- Service-centric software engineering
- Service oriented requirements engineering
- Service oriented architectures
- Middleware for service based systems
- Service discovery and composition
- Service level agreements (drafting, negotiation, monitoring and management)
- Runtime service management
- Semantic web
- Data modelling, mining and data analytics
- Aggregation, integration, and transformation
- Web and text mining
- Data mining methods, techniques, and tools
- Data analytics modelling and algorithms
- Patterns and frameworks
- Data visualization
- Knowledge systems and engineering
- Knowledge acquisition
- Knowledge-based and expert systems
- Knowledge representation and retrievals
- Knowledge engineering tools and techniques
- Time and knowledge management tools
- Knowledge modelling, integration, transformation, and management, Uncertainty knowledge management
- Knowledge visualization
- Mobile computing and mobile system engineering
- Mobile App design and development
- Innovative mobile applications
- Pervasive computing
- Mobile system validation and test automation
- Software & System Quality of Service
- Soft computing
- Quality assurance process, standards, and systems
- Software safety systems
- Software test automation and tools
- Software dependability, reliability, scalability
- Software & System Security
- Software and system security and privacy
- Healthcare Informatics
- Information Ecology and Knowledge Management
- Irrigation Informatics
- Neuro-Informatics
- Open Source: Challenges and opportunities
- Web-Based Learning: Innovation and Challenges
- Soft computing Signal and Speech Processing
- Natural Language Processing

- Mobile app security and privacy
- Encryption methods and tools
- Security service systems
- Cloud, sensor cloud and mobile cloud security
- Mobile system integrity, security, and fault tolerance
- Emergent Topics
- Cloud computing and Innovative cloud-based application systems
- Mobile cloud computing and application services
- Big data analytics application systems
- Innovative sensing cloud and systems
- Large-scale cyber systems
- IOT and smart city application systems
- Big data quality validation
- Big data application quality services
- Smart learning and innovative education systems
- Learning software design engineering
- Smart learning methods and environments
- Online learning systems
- Mobile enabled learning systems and tools

14.Computer Networks and Inventive Communication Technologies

- Addressing & Location Management
- Broadband Access Technologies
- Adhoc and Sensor Networks
- Cross Layer Design and Optimization
- Heterogeneous Wireless Networks
- High Speed Networks
- Internet and Web Applications
- Measurement & Performance Analysis
- Mobile & Broadband Wireless Internet
- Mobile Networks & Wireless LAN
- Network Architectures
- Network Based Applications
- Network Protocols & Wireless Networks
- Network Operations & Management
- Network Security Trust, & Privacy
- Next Generation Internet & Web Architectures
- Peer to Peer and Overlay Networks
- QOS And Resource Management
- Recent Trends & Developments in Computer Networks
- Routing, Switching and Addressing Techniques
- Self-Organizing Networks and Networked Systems
- Ubiquitous Networks
- Virtual & Overlay Networks
- Wireless Communications
- Wireless Mesh Networks and Protocols
- Wireless Multimedia Systems
- Personal and wearable networks
- Wireless ad hoc & sensor networks
- Information-centric networking
- Embedded networks
- Opportunistic networking
- Delay-tolerant networks
- Cognitive radio networks
- Vehicular networks
- Smart Grid communications
- Underwater sensor networks
- Cyber-physical systems
- Social networks
- Mobile and ubiquitous networking
- Green networking
-
- Overlay and peer-to-peer networks
- Local-area and metropolitan-area networks
- Storage-area networks
- Routing and transport protocols
- Big Data Networking
- Cloud computing and networking
- Software Defined Networking and Network Function Virtualization
- Internet of Things
- Link technologies
- Adaptive networking applications
- Authentication, authorization, accounting
- Security and privacy
- Cross-layer optimization
- Mobility and Location-dependent services
- Multimedia and real-time communication
- Machine-to-machine communications for smart environments
- Smart Cities
- Network traffic characterization and measurements
- Network management, reliability and QoS
- Performance evaluation of networks
- Testbeds for network experiments
- Network coding
- Optical and high-speed access networks

15.Resent Engineering and Nano Technology

- Advanced Computing Architectures and New Programming Models
- Bio-science and Bio-technology
- Collaborative applications
- Computational Techniques in Civil Engineering
- Control and Automation
- Concrete Engineering
- CAD/CAM/CIM
- CAE
- Database Theory and Application
- Data Base Management System
- Data Mining
- Decision making
- Digital signal processing theory
- Distributed Sensor Networks
- Distributed Computing
- Mechanical Design
- Evolutionary computing and intelligent systems
- Expert approaches
- Environmental Engineering
- Highways
- Human Computer Interaction (HCI)
- Image analysis and processing
- Information and data security
- Internet Technologies, Infrastructure, Services & Applications
- Nano Technology
- Mobile Computing and Applications
- Multimedia Communications
- Network Modelling and Simulation
- Network Performance; Protocols; Sensors
- Networking theory and technologies

- Open Models and Architectures
  Pattern Recognition
- Perception and semantic interpretation
- Production
- Real-time information systems
- Remote Sensing
- Security Technology and Information Assurance
- Soft Computing
- Software Engineering & Its Applications
- Signal Control System & Processing
- Speech interface; Speech processing
- Communication architectures for pervasive computing
- Computer Vision
- Computer Science and Its applications
- Structural Engineering
- Thermal Engineering
- Ubiquitous Multimedia Computing
- Vision-based applications
- VLSI Algorithms
- Web Technologies
- Textile Engineering

16.Recent on Mechanical Engineering and Technology

- Fluid Mechanics
- Applied Mathematics and Mechanics
- Biomechanics
- Heat Transfer
- Solid Mechanics
- Refrigeration and Air Conditioning
- Renewable Energy Technology
- Materials Engineering
- Composite Materials
- Marine Engineering
- Petroleum and Mineral Resources Engineering
- Textile Engineering
- Industrial Engineering
- Operational Research
- Manufacturing Processes
- Machine Design
- Quality Control
- Mechanical Maintenance
- Tribology
- CAD/CAM
- Aeronautical Engineering
- Production Engineering
- Welding Technology
- Robotics Engineering
- Aerospace Engineering
- Metallurgy
- Rock Mechanics and Mining Sciences
- Solid and Structural Mechanics
- Theoretical and Applied Fracture Mechanics

17.Advance Civil Engineering and Technology

- Assemblage and System
- Behaviour of Structures
- Behaviour of Structures under Seismic Loads
- Building and Environmental Acoustics
- Building Climate Systems
- Building Energy
- Civil and Environmental Engineering
- Coastal Engineering
- Composite Materials
- Concrete Structures
- Construction Economics
- Construction Engineering
- Design and Performance of Green Building
- Design Optimization of Structures
- Earthquake Engineering
- Energy Efficient Building Technology
- Energy Saving Building Materials
- Evaluation of Building Envelope Systems under Structural and Environmental Loads
- Evaluation of Glazing Systems for Energy Performance
- Fire Engineering
- Foundations Dynamics
- Geotechnical Engineering
- Health Monitoring and Life Prediction of Structures
- High Performance Concrete
- Hydraulic Engineering
- Life Cycle Engineering
- Materials and Durability
- Materials Engineering
- Mechanics and Materials Aspects of Advanced Construction Materials
- Municipal or Urban Engineering
- Nondestructive Testing and Evaluation
- Numerical Modelling of Structures
- Optimal Design of Structures
- Properties and Mechanics of Concrete
- Residential, Commercial, Industrial and Public Works
- Seismic Evaluation of Building Nonstructural Components
- Simulation Optimization and Risk Management
- Soil-Structure Interaction
- Structural Engineering
- Structural Evaluation of Panelized and Masonry Wall Systems
- Structural Reliability Analysis
- Surveying
- Sustainable Structures
- Transportation Engineering
- Ventilation and Indoor Air Quality
- Water Supply and Drainage

ISSN 2278-3075

Editor-In-Chief

Dr. Shiv Kumar
Ph.D. (CSE), M.Tech. (IT, Honors), B.Tech. (IT), Senior Member of IEEE, Member of the Elsevier Advisory Panel
Additional Director, Technocrats Institute of Technology and Science, Bhopal (MP), India

Associate Editor-In-Chief Chair

Dr. Hitesh Kumar
Ph.D.(ME), M.E.(ME), B.E. (ME)
Professor and Head, Department of Mechanical Engineering, Technocrats Institute of Technology, Bhopal (MP), India

Dr. Anil Singh Yadav
Ph.D(ME), ME(ME), BE(ME)
Professor, Department of Mechanical Engineering, LNCT Group of Colleges, Bhopal (M.P.), India

Dr. Gamal Abd El-Nasser Ahmed Mohamed Said
Ph.D (CSE), MS(CSE), BSc(EE), Department of Computer and Information Technology, Port Training Institute, Arab Academy for Science, Technology and Maritime Transport, Egypt

Members of Associate Editor-In-Chief Chair

Dr. Hai Shanker Hota
Ph.D. (CSE), MCA, MSc (Mathematics)
Professor & Head, Department of CS, Bilaspur University, Bilaspur (C.G.), India

Dr. Mayank Singh
PDF (Purs), Ph.D(CSE), ME(Software Engineering), BE(CSE), SMACM, MIEEE, LMCSI, SMIACSIT
Department of Electrical, Electronic and Computer Engineering, School of Engineering, Howard College, University of KwaZulu-Natal, Durban, South Africa.

Scientific Editors

Prof. (Dr.) Hamid Saremi
Vice Chancellor of Islamic Azad University of Iran, Quchan Branch, Quchan-Iran

Dr. Moinuddin Sarker
Vice President of Research & Development, Head of Science Team, Natural State Research, Inc., 37 Brown House Road (2nd Floor) Stamford, USA.

Prof. (Dr.) Nishakant Ojha
Principal Advisor (Information &Technology) His Excellency Ambassador Republic of Sudan& Head of Mission in New Delhi, India

Dr. Shanmugha Priya. Pon
Principal, Department of Commerce and Management, St. Joseph College of Management and Finance, Makambako, Tanzania, East Africa, Tanzania

Dr. Veronica Mc Gowan
Associate Professor, Department of Computer and Business Information Systems,Delaware Valley College, Doylestown, PA, Allman, China.

Dr. Fadiya Samson Oluwaseun
Assistant Professor, Girne American University, as a Lecturer & International Admission Officer (African Region) Girne, Northern Cyprus, Turkey.

Dr. Robert Brian Smith
International Development Assistance Consultant, Department of AEC Consultants Pty Ltd, AEC Consultants Pty Ltd, Macquarie Centre, North Ryde, New South Wales, Australia

Dr. Durgesh Mishra
Professor (CSE) and Director, Microsoft Innovation Centre, Sri Aurobindo Institute of Technology, Indore, Madhya Pradesh India

Prof. MPS Chawla
Member of IEEE, Professor-Incharge (head)-Library, Associate Professor in Electrical Engineering, G.S. Institute of Technology & Science Indore, Madhya Pradesh, India, Chairman, IEEE MP Sub-Section, India

Dr. Vinod Kumar Singh
Associate Professor and Head, Department of Electrical Engineering, S.R.Group of Institutions, Jhansi (U.P.), India

Dr. Rachana Dubey
Ph.D.(CSE), MTech(CSE), B.E(CSE), Professor, Department of Computer Science & Engineering, Lakshmi Narain College of Technology Excellence (LNCTE), Bhopal (M.P.), India EMail: adash.research@gmail.com

ISSN 2278-3075

Executive Editor Chair

Dr. Deepak Garg
Professor, Department Of Computer Science And Engineering, Bennett University, Times Group,
Greater Noida (UP), India

Members of Executive Editor Chair

Dr. Vahid Nourani Professor, Faculty of Civil Engineering, University of Tabriz, Iran.	**Dr. Saber Mohamed Abd-Allah** Associate Professor, Department of Biochemistry, Shanghai Institute of Biochemistry and Cell Biology, Shanghai, China.
Dr. Xiaoguang Yue Associate Professor, Department of Computer and Information, Southwest Forestry University, Kunming (Yunnan), China.	**Dr. Labib Francis Gergis Rofaiel** Associate Professor, Department of Digital Communications and Electronics, Misr Academy for Engineering and Technology, Mansoura, Egypt
Dr. Hugo A.F.A. Santos ICES, Institute for Computational Engineering and Sciences, The University of Texas, Austin, USA.	**Dr. Sunandan Bhunia** Associate Professor & Head, Department of Electronics & Communication Engineering, Haldia Institute of Technology, Haldia (Bengal), India.

Dr. Awatif Mohammed Ali Elsiddieg
Assistant Professor, Department of Mathematics, Faculty of Science and Humatarian Studies,
Elnielain University, Khartoum Sudan, Saudi Arabia.

Technical Program Committee Chair

Dr. Mohd. Nazri Ismail
Associate Professor, Department of System and Networking, University of Kuala (UniKL),
Kuala Lumpur, Malaysia.

Members of Technical Program Committee Chair

Dr. Haw Su Cheng Faculty of Information Technology, Multimedia University (MMU), Jalan Multimedia (Cyberjaya), Malaysia.	**Dr. Hasan. A. M Al Dabbas** Chairperson, Vice Dean Faculty of Engineering, Department of Mechanical Engineering, Philadelphia University, Amman, Jordan.
Dr. Gabil Adilov Professor, Department of Mathematics, Akdeniz University, Konyaaltı/Antalya, Turkey.	**Dr. Ch. V. Raghavendran** Professor, Department of Computer Science & Engineering, Ideal College of Arts and Sciences Kakinada (Andhra Pradesh), India.
Dr. Thanhtrung Dang Associate Professor & Vice-Dean, Department of Vehicle and Energy Engineering, HCMC University of Technology and Education, Hochiminh, Vietnam.	**Dr. Wilson Udo Udofia** Associate Professor, Department of Technical Education, State College of Education, Afaha Nsit, Akwa Ibom, Nigeria.
Dr. Ch. Ravi Kumar Dean and Professor, Department of Electronics and Communication Engineering, Prakasam Engineering College, Kandukur (Andhra Pradesh), India.	**Dr. Sanjay Pande MB** FIE Dip. CSE., B.E, CSE., M.Tech.(BMI), Ph.D.,MBA (HR) Professor, Department of Computer Science and Engineering, G M Institute of Technology, Visvesvaraya Technological University Belgaum (Karnataka), India.
Dr. Hany Elazab Assistant Professor and Program Director, Faculty of Engineering, Department of Chemical Engineering, British University, Egypt.	**Dr. M.Varatha Vijayan** Principal, Department of Mechanical Engineering, Mother Terasa College of Engineering and Technology, Pudukkottai (Tamil Nadu) India.

ISSN 2278-3075

International Journal of Innovative Technology and Exploring Engineering (IJITEE)

(Volume No. 12, Issue No. 3, September - December 2023)

Contents

ISSN 2278-3075

Photo Voltaic (PV) Cell Characteristics Design Using M. File in Matlab

Linnet Jaya Savarimuthu[1], **Kirubakaran Victor**[2]
[1]Research Scholar, Centre for Rural Energy, Gandhigram Rural Institute - Deemed to be University, Gandhigram, Tamilnadu, India.
[2]Assistant Professor, Centre for Rural Energy, Gandhigram Rural Institute - Deemed to be University, Gandhigram, Tamilnadu, India

ABSTRACT

Solar energy is an emergent trend suitable for power production in both industrial and household appliances. The distributed renewable resource like solar energy is projected to act as a major responsibility in the forthcoming smart grid applications and technology. For the generation of electricity from solar power, it is essential to analyze the performance characteristics of the solar Photo Voltaic (PV) module, for instance, the power output of a PV panel and the prominent conversion efficiency. The performance of the electrical characterisation of a Photo Voltaic (solar) cells or module delivers the bond among the generated current and voltage on a typical solar PV cell which is termed as a V-I characteristic curve of solar cells. In this paper, a single diode correspondent circuit has been considered to inspect Voltage (V-I) and Power (P-V) characteristics for different insolation levels of a typical 100 W polycrystalline solar PV module. In order to validate the graphical depiction of the solar cell or module operation, M.file in MATLAB software was used. The generated characteristic curves summarise the connection between the current (I) and voltage (V) at the existing state of temperature with different irradiance. The obtained Power-Voltage (P-V) characterisation grant the essential information for building a solar electric power system to drive close up to its maximum peak powerpoint while feasible. The resulted graphs reveal that while considering the single diode model, the level of insolation varies with series resistance and by the generation of photo-current which in turn delivers the rapport of efficiency of solar cells. The proposed system is the initial step to learn a hybrid power system where some other renewable sources can be combined along with a solar power generation system.

Keywords: Characteristic Curves, Efficiency, Irradiance, M.file, Photovoltaic Cell, Power Output, Solar Energy

I. INTRODUCTION

Photovoltaic (PV) modules based renewable energy generation systems, nowadays signify the most appropriate and outstanding solution, for both domestic and industrial power levels, to reduce CO2 emissions and the energy consumption formed by oil and gas. These sources are sustainable and generate clean energy. One of the most essential renewable sources is solar energy. It has been attracting the awareness of many engineers, scientists and researchers. Every day the earth receives an enormous quantity of energy from the sun. The vast amount of solar radiation received by the earth can be employed to produce electrical energy with the help of Photo Voltaic (PV) modules (Photo Electric effect).The unique design has to be considered for modelling a PV system because the solar power system is dynamic in nature which in turn results in unpredictable power output and also the V-I parameters are affected by ambient conditions like the temperature and the intensity of radiation. Hence the efficiency of a solar cell is not stable at all the epoch, and so a simulation should be carried out by means of the required parameter setting to learn, analyze and to develop the PV system model based on the fluctuations of solar radiation values along with the obtained operating temperature [1].

The solar photovoltaic system physically reveals a nonlinear Voltage-Current (V-I) and Voltage-Power (V-P) output plots that diverge with the rise in illumination level, cell temperature, load current, and load potential. Hence on shady location, the efficiency of a solar PV power output characterisation has an exponential attribute related to facilitate the diode ideality factor. When solar energy (photons) is absorbed by a solar chamber which is a semiconducting material, it boosts up the energy of electrons in the valence band causes the thrusting of electrons in the conduction band. This process happens only when the incident photon energy is higher than that of the bandgap energy and it leads to developing a combination of electrons and holes. These free carriers are swept at a distance which is concealed with the ability of the centralized electric fields of the p-n junction (depletion layer) that results in the generation of current which is proportional to the intensity of incident radiation [2].

This paper describes the vital modelling along with simulation for a photovoltaic (PV) panel. For developing the system, module LE12P100 has been considered. This module generates a maximum electrical power output of 100 W. The M.file script in MATLAB software was used to study the Solar Photovoltaic Characterisation which generates the required power (Voltage x Current) for different intensity levels and by keeping the cell temperature as constant. The resulted plots deliver the Photovoltaic (PV) panel or module's capacity for the conversion of sunlight into electric power.

II. PV MODULE MODELLING

A model solar cell can be considered as a current source where the current produced is directly proportional to solar radiation declining on it. The realistic performance of the solar cell is diverged from an ideal model because of its optical and electrical form of losses [3].

The single diode correspondent circuit diagram of a PV cell is represented in fig.1.

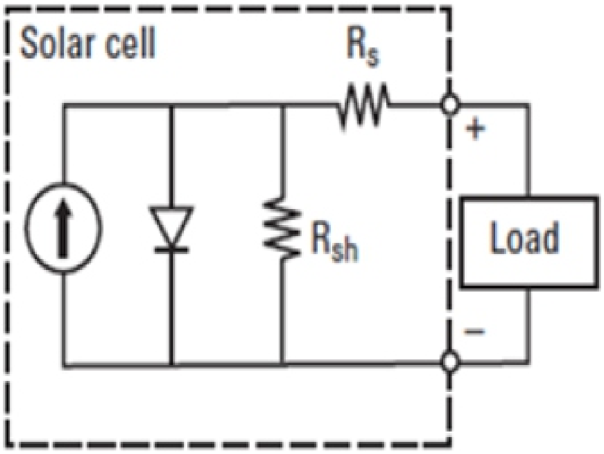


Fig.1 The single diode correspondent circuit diagram of a PV cell

Photovoltaic (solar) cells are made with semiconducting materials that have the capability of converting radiation which incident in the solar PV spectrum into the electric currents. The increase in the intensity of light will proportionally increase the emission of photoelectrons rate generated in the photovoltaic material.

This paper proposes the mathematical modelling with the aid of numerical equations of a PV module and is described through a single diode correspondent circuit of the solar cell. The parameters such as the photocurrent, the diode current, and also the series and shunt resistor have been taken to build an electrical power output curves from the system designed. The governing equation of a photovoltaic (solar) cell is represented in equation (1) [4].

$$I_o = (N_p.I_{ph}) - (N_p.I_{rs}) * (\exp\left(\frac{q}{KTA}\right) * \left(\frac{V_o}{N_s}\right) - 1) \quad (1)$$

Where

$$I_{ph} = I_{scr} + k_i(T - T_r)(\frac{S}{100}) \quad (2)$$

$$I_{rs} = I_{rr} * \left(\frac{T}{T_r}\right)^3 * \exp\left(\frac{q*E_g}{KA}\right) * (\frac{1}{T_r} - \frac{1}{T}) \quad (3)$$

I_{rs} - reverse saturation current of a photovoltaic (solar) cell
T – the cell temperature in Celsius
k - Boltzmann's constant - $1.381 * 10^{-23}$ J/K
q - Elementary charge - $1.602*10^{-19}$ C
K_i - Short circuit current temperature coefficient at Iscr
S - Solar irradiation
I_{scr} - Short circuit current at 25^0 C
I_{ph} - Photon current
E_g - Bandgap energy,
T_r- Reference temperature
R_{sh} –parallel shunt resistance,
R_s - Series resistor

III. RESULT ANALYSIS AND DISCUSSION

A. Reference module

The photovoltaic module used here is LE12P100 which generates a maximum electrical power output of 100 W. The table.1 gives of the key specifications of a photovoltaic module LE12P100 (Polycrystalline) at STC 250C. The given structure is implemented by using the M.file script in MATLAB.

Table-1: Typical rating of Lubi 100 W Photovoltaic Module

CHARACTERISTICS	DETAILS
Model	LE12P100
Maximum electrical power rating (P_{max})	100 W
Rated Operating Voltage (V_{max})	17 V
Rated Operating Current (I_{max})	5.88 A
Open-circuit voltage (V_{oc})	21.5 V
Short-circuit current (I_{sc})	6.2 A

B. Simulation results

Mathematical Modelling of solar PV cells has been done using MATLAB/m.file script. For developing an m.file script, the necessary governing equations of a solar photovoltaic cell has been considered as specified in the equations from (1) - (3) and also taking account of electrical and physical parameters, for instance, solar irradiance (insolation), the cell temperature, and the electrical resistance, etc. PV cells are modelled and the resulted from the V-I plot of an illuminated photovoltaic cell where depends on the spectrum of the incident light and the temperature. The conversion of solar energy into the form of electrical energy gives the power output characteristics of the solar photovoltaic module which can be computed by means of corresponding parameters.

Fig. 2 shows the curves of output current and output voltage under different insolation levels for the Lubi 100 W PV panel.

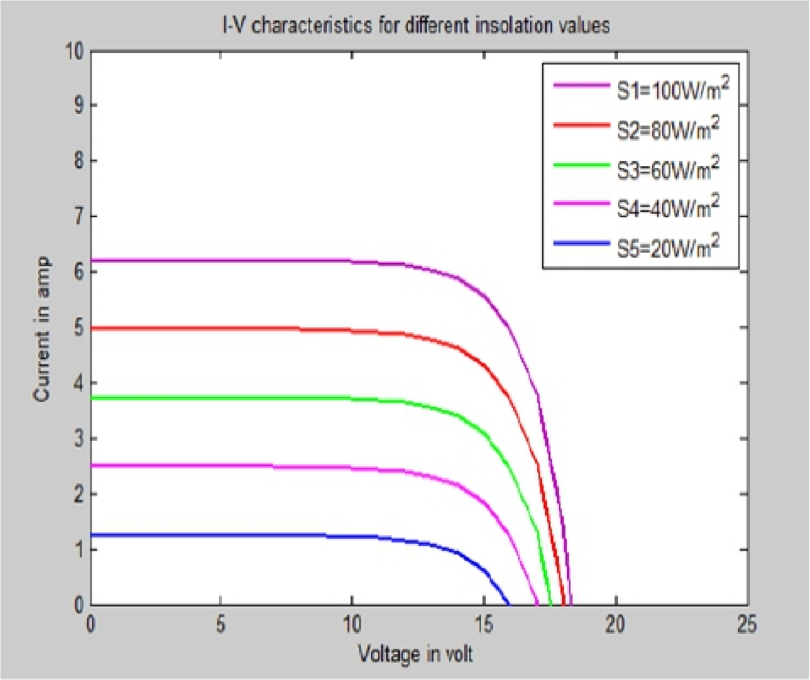


Fig. 2 V-I characterisation for various insolation values

Fig. 3 shows the curves of output power and output voltage under different insolation levels for the Lubi 100 W PV panel.

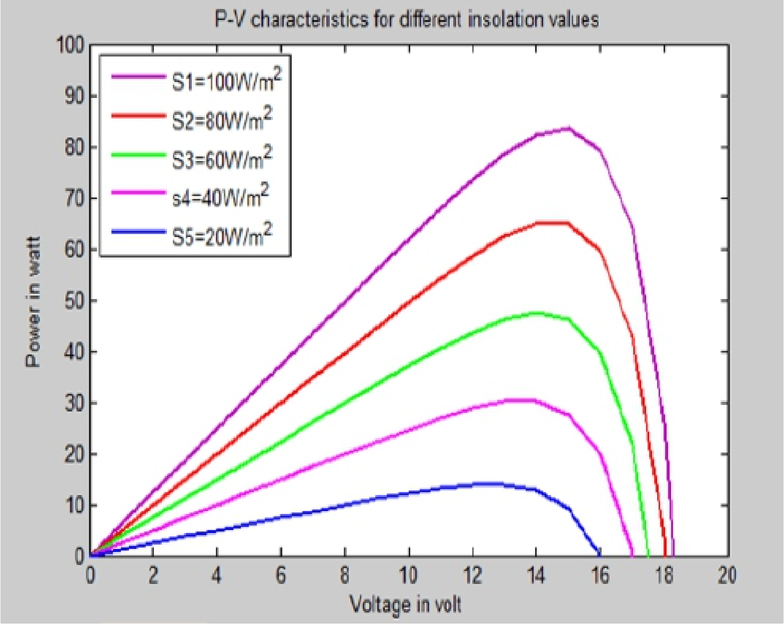


Fig. 3 P-V characterisation for various insolation values

Fig. 4 shows the curves of output power and output current under different insolation levels for Lubi 100 W PV panel.

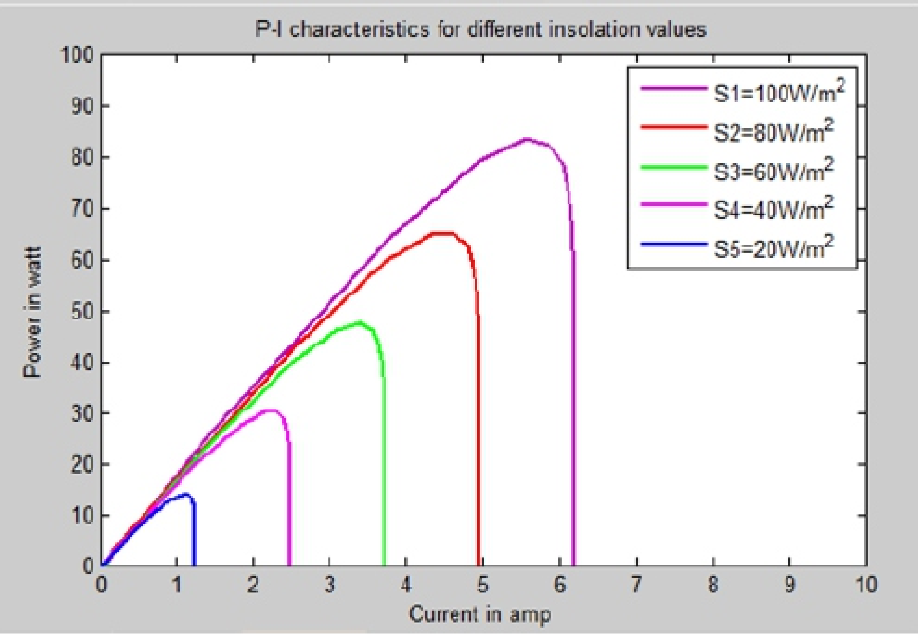


Fig. 4 P-I characterisation for various insolation values

IV. CONCLUSION

Based on the literature assessment and theories, mathematical modelling of the PV module has been considered and developed using the M.file script in MATLAB software by considering the single-diode correspondent model of a photovoltaic (solar) cell. The proposed model referred to the datasheet ideals of a classic 100 W solar PV module along with the various sunlight/insolation. This paper also delivers an apparent and brief perceptive of the bonding between current, power and the potential developed in the solar PV module and the consequences of changes in solar insolation levels on the parameters involved. Moreover, the proposed model serves as a contrivance for the prediction of the responses which includes a photovoltaic (solar) cell, solar panel and a solar array under the variations of the environmental and electrical parameters. The obtained plots also reveal that the voltage-current characterisation persists several steps where the voltage-power characterisation delivers plenty of local peaks with the maximum peak powerpoint which can be termed as the global peak. This proposed model helps to build a hybrid power system in the future using some other renewable sources along with the proposed solar energy system.

REFERENCES

1. *Chetan Singh Solanki, Solar Photovoltaics, PHI publication.*
2. *Farhan A. Salem. 2014. "Modeling and Simulation issues on PhotoVoltaic systems, for Mechatronics design of solar electric applications". IPASJ International Journal of Mechanical Engineering. Volume 2, Issue 8.*
3. *G. Venkateswarlu, P.Sangameswar Raju. 2013. "Simscape Model of Photovoltaic Cell". International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. Vol. 2, Issue 5.*
4. *Habbati Bellia, Ramdani Youcef, Moulay Fatima. 2014. "A detailed modeling of photovoltaic module using MATLAB". NRIAG Journal of Astronomy and Geophysics. Volume 3, 53–61*
5. *Huan - Liang Tsai, Ci - Siang Tu, and Yi -Jie Su. 2008. "Development of Generalized Photovoltaic Model Using MATLAB/SIMULINK". Proceedings of the World Congress on Engineering and Computer Science.*
6. *JA. Ramos-Hernanz, JJ. Campayo, E. Zulueta, O. Barambones, P. Eguía, and I. Zamora. 2013. "Obtaining the characteristics curves of a photocell by different methods". International Conference on Renewable Energies and Power Quality. No.11.*

7. *Jasmina Radosavljević, Amelija Đorđević. 2001. "Defining of the Intensity of Solar Radiation on Horizontal and Oblique Surfaces on Earth", Facta Universitatis Series: Working and Living Environmental Protection. Vol. 2, No 1.*
8. *Jay Patel, Gaurag Sharma. 2013. "Modeling and Simulation of Solar Photovoltaic Module Using MATLAB / SIMULINK". International Journal of Research in Engineering and Technology Volume: 02, Issue: 03.*
9. *Mittal V. Joshi. 2015. " Mathematical Modeling and Simulation of PV System using MATLAB/Simulink". Indian Journal of applied research Volume: 5, Issue: 6.*
10. *Mukesh Kr. Gupta and Rohit Jain. 2013. "Design and Simulation of Photovoltaic Cell using Decrement Resistance Algorithm", Indian Journal of Science and Technology. Vol 6 (5).*
11. *N. Lakshmi Tirupathamma, M. Rajesh, K. Naga Vamsi, R. Lohitha, P. Sravan Kumar. 2014. "Matlab Simulation of Grid Connected PV System Using Hysteresis Current Control Inverter". International Journal of Research Studies in Computer Science and Engineering. Volume. 1, Issue 5.*
12. *Rituraj Tamrakar, Archana Gupta. 2015. "A Review: extraction of solar cell modelling Parameters". International Journal Of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering. Vol. 3, Issue 1.*
13. *Sakshi Shrivastava, Anmol Ratna Saxena, and A. K. Wadhwani. 2015. "MPPT Based Optimization of Photovoltaic System Using DC-DC Converter". International Journal of Hybrid Information Technology. Vol.8, No.10.*
14. *Shankhamitra Sunani , Madhusmita Panda, Radha Nath Patra. 2015. "Mathematical Modelling of Simulink based Photovoltaic Power Generation System". International Journal of Advanced Engineering and Global Technology I. Vol - 03, Issue -12.*
15. *Tarak Salmi, Mounir Bouzguenda, Adel Gastli, Ahmed Masmoudi. 2012. "MATLAB/Simulink Based Modelling of Solar Photovoltaic Cell", International Journal Of Renewable Energy Research, Vol.2, No.2.*

AUTHORS PROFILE

Linnet Jaya Savarimuthu is a Research Scholar at The centre for Rural Energy, Gandhigram Rural Institute - Deemed to be University, Gandhigram, Tamilnadu, India. She received a bachelor's degree in Electronics and Instrumentation Engineering and a Master's degree in Instrumentation and Control Engineering.

Dr. Kirubakaan Victor is an Assistant Professor in The centre for Rural Energy, Gandhigram Rural Institute - Deemed to be University, Gandhigram, Tamilnadu, India. He holds a Ph.D. and Master's degree in Energy Engineering and an expert in the field of Biomass Gasification.

Data Confidentiality in Cloud using Multi-Party Computation

A.Vijaya Kumar[1], **P. Ramya**[2], **R. Poojitha**[3], **M. Lahari**[4]
[1]Associate Professor, CSE Department, KL University, Guntur, Andhra Pradesh, India.
[2]Computer Science and Engineering, KL University, Guntur, AndhraPradesh ,India.
[3]Computer Science and Engineering, KL University, Guntur, Andhra Pradesh ,India.
[4]Computer Science and Engineering, KL University, Guntur, Andhra Pradesh ,India.

ABSTRACT

A new era has approached where we are storing our information in cloud and performing several computations on powerful servers remotely. In cloud, data is not completely secured and sometimes under the control of untrusted Third parties. Some secured protocols are being implemented. The secure multi-party computation protocol, which is existing, demands the inputs to be encrypted using a public key. So, these reasons limit this Secure Multi-party computation to be employed. In the current paper, we put forward a protocol named homomorphic encryption where the input function is being encrypted by different key. This paper also uses Multi-party computation which is one of the most secured technique in cryptography.

Keywords: - Cloud Computing, Confidentiality, Multi-party computation, Homomorphic encryption.

I. INTRODUCTION

Data Confidentiality can be achieved through Secure Multiparty Computation [1] [2] [3] [4] [5] [6] [7] [8] [9] [11] and Homomorphic Encryption. In Secure Multiparty Computation a joint computation function is being performed on the given inputs by taking it from the users where as in this data cannot be completely secured and this can achieved through Homomorphic Encryption [1] [3] [4] [5] [6] [13] [14] [15] by not asking the inputs and computations will be performed on the encrypted data as if they were performing on the original data and for that we used Paillier Algorithm for better security purpose. In section II, the preliminaries i.e., Secure Multi-party Computation and Homomorphic Encryption terms are discussed. In section III, the work related to this paper/Literature survey is mentioned. The methodology proposed is in section IV. The results are mentioned in section V. The applications of the proposed methodology are mentioned in section VI. The conclusion and future work are in section VII and VIII respectively.

II. PRELIMINARIES

(A) Secure Multi-party Computation:

Secure multiparty computation is also named as multiparty computation, privacy preserving and secure computation which is also the field in cryptography that also to compute a joint function on the inputs provided to us.

Secure MPC makes us to work and analyze the information for the welfare of the people without revealing the private information which is very confidential.

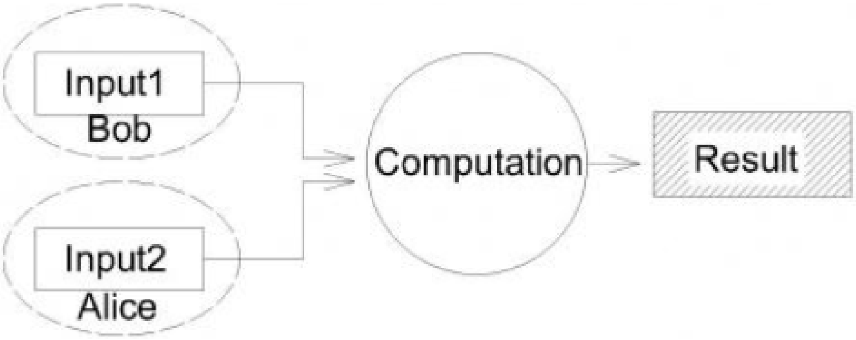


Fig: Multi-party Computation

So as to comprehend what the careful meaning of SMC (secure multiparty calculation) is, we should ready to think about Shamir's Secret Sharing calculation. This algorithm tells us to divide and distribute one value which should be secret to the users. In order to retrieve that available secret value, a smaller number of users must combine their data together. This type of algorithm that is Shamir's algorithm, also used to carry out computations on the secretly shared number. Here, without using a number, we can replace secret as user's personal data. Secure Multi Party Computation also perform the task in the same way as Shamir's algorithm, thus by dividing the user's data into many smaller parts, and the data which is encrypted is sent to a different server. For example, a group of people with their information that is to be stored in cloud needed to be secured before itself, so they jointly compute a function on all their inputs:

- Give a function for X: b1, Y: b2, Z: b3 on which we are going work or compute jointly
- These outputs a 3-tuple: (k1(b1, b2, b3), k2(b1, b2, b3), k3(b1, b2, b3))
- Most of the times all the k1, k2, k3 may not be necessarily same.
- Always people wish to preserve some security properties.
- This security either must be provided either by the adversarial behavior of the participants or by the third party.

(B) Homomorphic Encryption:

This is the process of changing the information or the data into the ciphertext that can be viewed and also we can work on it as if it as a original information which is not actually provided by the user and also the encrypted one. On the encrypted data we can perform the computations without compromising as on the original data. What the homomorphic encryption is changing the form of information or data into another while keeping the same relationships among the data elements in either of the data sets. In Homomorphic encryption confounded numerical tasks are performed on the scrambled information. In distributed computing Homomorphic encryption plays a significant role.

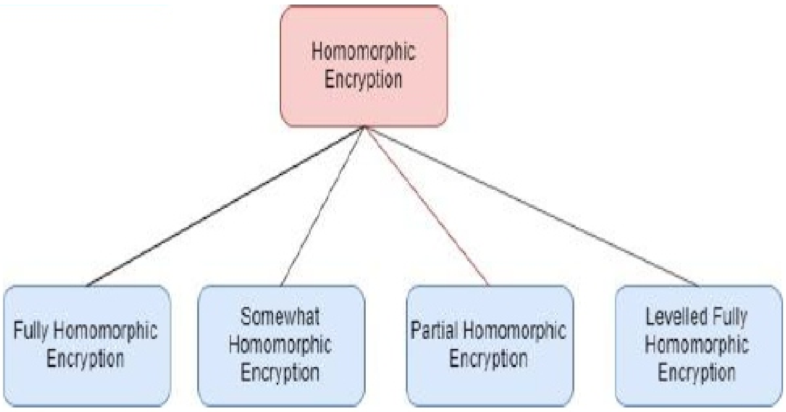


Fig: Types of Homomorphic Encryption

Here we will explain a easy example in what way a homomorphic encryption uses cloud computing in performing operations.

- Let us consider a business XYZ, which is extremely occupied, so it asks cloud supplier to execute the activity. As we realize that lone a cloud supplier will just have the entrance to the encoded set, Finds the distinction of 30-20 and offers the response as 10.

A small example to describe this is:

- An organization named Creta has a very crucial information of data sets which has numbers of 10 and 20. To encrypt this it multiplies each element in it with 3, which creates a new set as 30 and 60.
- Later it sends the encrypted data to cloud. A few months later the other party contacts the Creta and requests for the data.
- As the Creta is busy with their own work it asks the cloud provider to perform the operation which only has the access of encrypted data find the sum as 90 and returns the answer.
- Then the Creta decrypts the data and provides 30 as the output to the other party.
- Hence with this security has been provided to the data.

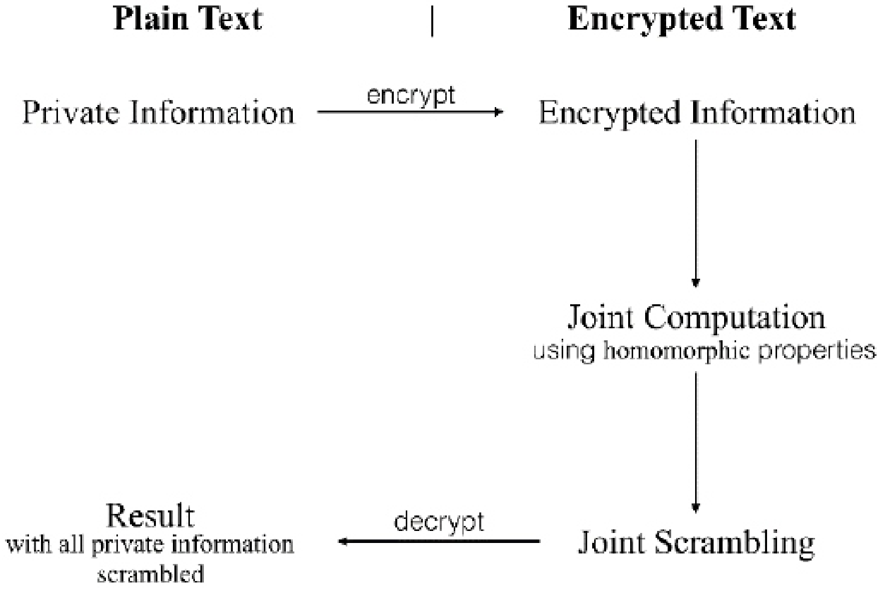


Fig: Homomorphic Encryption

III. RELATED WORK

(A) RELATED WORK OF SEGMENTATION:

Debasis Das et al [1] To store the vast amount data these days a great infrastructure is needed that provides the support for the storage and processing of the data. This is becoming a great challenge for all large MNC's and also for various other sectors to keep their data. Hence Cloud Computing provides the on-demand access and also very much convenient in sharing the resources. So, the MNC's or other organizations tie-up with the cloud and use all the available resources properly. Cloud platform extracts and analyze the useful data from big data. Concerns of cloud computing are confidentiality and privacy. The arrangement is to send the encoded information to the cloud server. Regardless we need to help the valuable estimations on figured information and FHE in supporting manner. Note that mechanisms other than this exists for secure computation, which require diff data providers for exchange of data. The fully homomorphic encryption is better to adapt to the scenario in which we have a lot of data resources. SMPC gives the correct output to everyone for the joint computation not else about anyone info, the users which are performing the calculations may be actively malicious. SMPC can be done by the arbitrary calculations and more number for parties In this way, we can see the SMPC conventions with the goal that the compilers will take input details for mapping, yield is the convention that computes the capacity all the more safely. SMPC gives both privacy and honesty which can be superior to FHE and the confirmed figuring. Security to the information is given by utilizing the cushioning idea called as Optimal Asymmetric Encryption Padding(OAEP) alongside the Hybrid encryption calculation that relies upon RSA to enable the numerous individuals to give the security to their information by utilizing one joint register work on them and it additionally gives the Integrity and Confidentiality to the information. Cryptographic procedures here are Homomorphic encryption and the safe multiparty Computation. OAEP is a Feistel Network. It is a productive cryptographic strategy for cushioning multi party information without figuring it. Weaves information is scrambled by utilizing a plan called cushioning called Optimal Asymmetric Encryption Padding Combine with the Hybrid encryption Algorithm which depends on RSA. Emily Shen et al. [2] Analysts attempt to make secure multiparty calculation—that takes numerous contributions of information from various gatherings and together figures a capacity combinedly and keep the sources of info mystery. "Data sharing ought to be directed in an exceedingly way that ensures the security and common freedoms of individuals, that jelly business classification, that shields the information being shared, which secures the capacity of the Government to watch, explore, avert, and answer to digital dangers." Trying to accomplish greater security to the information use SMC. Cryptographers has been making a decent attempt to build up this kind of innovation, called as secure multiparty calculation for three decades. Secure Multi party calculation adjusts that every one of them get familiar with the accurate yield of the joint calculation however nothing else about different data sources. indeed, even the dynamic execution of calculation by a portion of the clients. Secure Multi party calculation would be accomplished for various calculations and for various gatherings. Along these lines, we can see the safe SPC conventions as like of compilers which accepts contribution as detail of the capacity and yield will figures the capacity safely. So as to perceive how it fills in as of compilers let us take a procedure which is called mystery sharing .A m-of - n mystery sharing plan separates a mystery contribution to k number of pieces which are held by various individuals such that m number of individuals can consolidate their offers to recreate the mystery. For instance, you need to make a 2 of 3 mystery sharing plan utilizing the lines in 2-dimensional space. So as to share the mystery s sum three unique individuals, we will pick an arbitrary line. In this y-blocks equivalent to mystery s. Each individual would share their s, which are various focuses on hold. For any two points that diversely characterize a line, an any 2 individuals join to register the mystery key, conversely a solitary offer doesn't uncover anything about mystery key. This sort of procedure can be summed up to ant=y m utilizing the level of polynomials m-1. On the off chance that any two privileged insights p and r have shared, at that point individuals can compute portions of the total p+r by disentangling including their offer together. Portions of the item can likewise be processed p*r through entangled control of the portions of p and r. In the specific situation, a couple of number of structures are additionally accessible for secure MPC advancement of programming, which require little measure of information on the cryptography. This can be empowered by a designer to give a functionals examine of the significant level wanted calculation. through an exceptional language, alongside the accessible explanations that are accommodating which the given information is allowed so as to uncover. These sort of programming instruments can isolate the errands of programming specialists and cryptographers, streamline their association so as to improve ease of use that likewise diminishes the time being developed. Andriana Lopez Alt et al. [3] In on the fly MPC, each client is included on the underlying phase of transferring his/her encoded information to the cloud server and in conclusive outcome in the decoding stage the outcomes are uncovered. Multifaceted nature of both is free on the capacity that is being processed to the complete number of individuals in the framework. Clients ought to choose which capacity ought to be registered in the further or who they have to figure with. They additionally need the identified with past endorse that in the end pick capacities and the capacities by which information is assessed. As we realize that the cloud offers increasingly number of preferences in both cost just as usefulness, it likewise bring up issues of privacy, since put away information can be helpless against snooping either by more security in applications like face recognition with feasibility.

the cloud supplier or by other cloud suppliers in the cloud. Everybody realizes that the data contains the most touchy information, it is basic for the clients to store the information in the wake of encoding as it were. FHE is the calculation which is reasonable in the setting time where it includes a solitary or significantly more clients, on the grounds that the data sources ought to be encoded under a similar kind of key. All the more ever, there are numerous circumstances where numerous clients, Who have just moved their huge information to store in the cloud in figure structure, and afterward need to choose to fathom joint elements of their information. For instance, They can wish cloud to process the joint factual data all alone information bases, to find the most widely recognized documents in their record assortments, Which runs a calculations part dependent on their post information to arrive at goal or when all is said in done, in the challenge where more clients would need to consolidate their data so as to achieve a most shared objective. Finally, another type of completely homomorphic encryption that they call a multi key FHE which allows a calculation on the data scrambled under the more noteworthy number of disconnected keys As there are increasingly number of clients engaged with any of the calculation or figuring in the arrangement must be limited, as that the all out number of clients in the framework is random.Min Zhao et al. [4] The efficiency of different algorithms are compared based on their security characteristics. These are single homomorphic algorithms. These algorithms differ in many ways based on which homomorphism they perform. Different types of five fully homomorphic encryption algorithms are researched to find out the performance of those algorithms based on security. From this paper we can be able to know which algorithm will provide high security to the data. The GM Algorithm [4] is based on quadratic residue with semantic security and it satisfies only addition homomorphism and it has low efficiency. The GM algorithm is introduced in 1984. In 1985 The ELGAMAL algorithm is introduced to provide the security for data. This calculation depends on open key cryptosystem and elliptic bend cryptosystem and it fulfills any multiplicative homomorphism and can be utilized for encryption. The Hill figure calculation fulfills expansion homomorphism. The RSA calculation depends on number hypothesis and it fulfills any duplication homomorphism encryption. In 1994, a calculation called Improved probabilistic encryption calculation is presented. It encodes n bits of information at once yet can just play out a limited number of expansion homomorphism activity. Paillier encryption algorithm is based on a quadratic residue and it can perform any additive homomorphic operation. At a low possible cost, the confidentiality of data will be guaranteed. The technique used is depends upon encryption scheme called Bulk Copy Scheme. Because of untrusted third parties, this new technique is implemented for outsourcing the secured data and apt to the purpose of Secure multiparty Computation. The data is encoded with many different public keys. There is no interaction of users in this scheme. This scheme provides.

IV. OUR CONTRIBUTION

Paillier Cryptosystem is at first delineated in a very 1999 paper by Pascal Paillier, that outlines associate degree uneven non-deterministic cryptography algorithmic rule with homomorphic additive properties. The cryptosystem additionally provides detached signatures, however not inline signatures as provided by RSA.

Paillier Key cryptography is one of the homomorphic algorithms. The clients who utilize this calculation will have an open key and a private key. The information is first scrambled with the open key and send to the recipient. Then the receiver will decrypt the data with his private key. Paillier key cryptography provides "additive homomorphism".

For example, Bob is the sender and Alice is the recipient. Bob has two messages that are encoded with receiver's public key. Bob add the messages together and will send to receiver. Alice decodes the data with his private key and get the result as the sum of both messages.

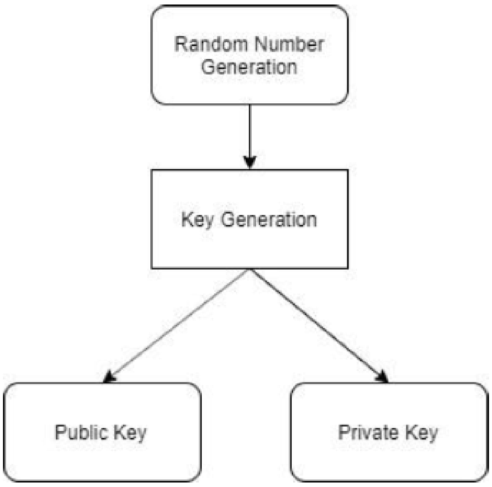


Fig: Paillier Cryptosystem

Example: Casting a ballot System

We need to engrave all polling forms for a political decision board hence outsiders can't check the votes

We need the political race board to have the option to tell who won the political decision

We don't wish the political race board to illuminate us that voter casted a ballot that up-and-comer

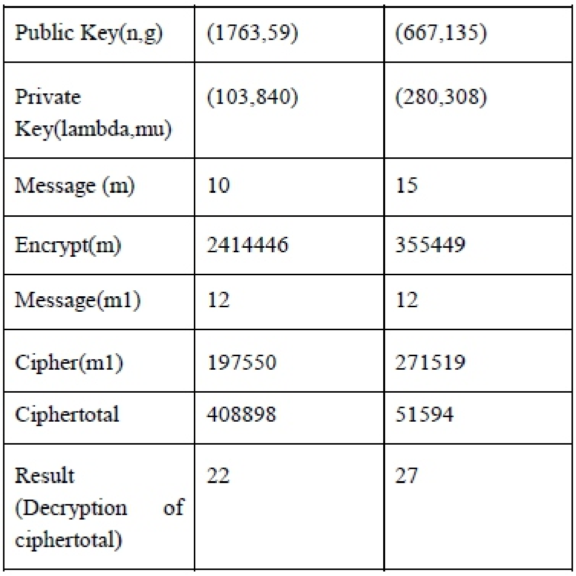
With antiquated cryptography and coding there's no methodology for the political decision board to get a handle on who won the political decision while not decoding each vote.

The voting system can be again designed by using additive homomorphism as pursues:
Voters scramble their polling forms with the political race board's open key

Voters send their scrambled polling forms to an examining server

The exploring server includes every one of the polling forms along and sends them to the political race board The political race board decodes the voting form all out, reports who won the political race while not regularly perusing a voter's voting form.

IV. RESULTS



Public Key(n,g)	(1763,59)	(667,135)
Private Key(lambda,mu)	(103,840)	(280,308)
Message (m)	10	15
Encrypt(m)	2414446	355449
Message(m1)	12	12
Cipher(m1)	197550	271519
Ciphertotal	408898	51594
Result (Decryption of ciphertotal)	22	27

DISCUSSION OF RESULTS:

The data confidentiality is achieved as the receiver get the ciphertotal i.e., the combined result of inputs and decrypts it. The receiver don't know the multiple inputs. The data is secure and cannot be modified.

VI. APPLICATIONS

A. Electronic voting:

In Electronic voting, The people had to cast their vote individually without disclosing to whom they vote to other people. Here, while they cast their votes, they encrypt the vote. So that it is not known to anyone to whom they vote. But, these votes are collected by Election commission to know how many people casted their votes. Then election commission will have the product of the votes casted by the people. They will decrypt that data and acquire results to know how many of the people of the total population casted their votes. People may not be able to know to whom others voted to.

B. Electronic cash:

A new notion is feature of self blinding. The flexibility to alter one coded text into another while not dynamically the content of the decrypting. This application is useful where there is no need of vendors in a shop to pay the amount through our credit card. This application is similar to e-voting. The goal of both the applications is same. We can purchase the items online without disclosing our details and promoting the notion of e-cash.

VII. CONCLUSION

In this paper, Data in cloud will be secured based on confidentiality. In real world, any third party is not trusted to store and secure the data in cloud. There are many homomorphic encryption algorithms which had provided security but not to full extent. There are algorithms with some disadvantages like Hill Cipher i.e., It is symmetric encryption algorithm and data will be easily decrypted. So, paillier algorithm is better to implement to this approach. This method is only implemented for digits but in future it can be extended for strings and images.

REFERENCES

1. *Debasis Das, Department of Computer Science and Information System, BTIS Pilani, Secure Cloud Computing Algorithm Using Homomorphic Encryption and Multi-Party Computation*
2. *Emily Shen, Mayank Varia and Robert K Cunningham, W. Konard Vesey Cryptographically Secure Computation. 0018-9162 / 15/$31 .00 2015 E E E*
3. *Adriana Lopez - Alt,Eran Tromer, Vinod Vaikuntanantham, On-the-Fly Multiparty Computation on the Cloud via Multikey Fully Homomorphic Encryption*
4. *Min Ahao, Yang Geng, Homomorphic Technology for cloud computing*
5. *Yuangang Yao, Jinxia Wel, JianyiLiu, Ru Zhang, efficiently secure multipart computation based on Homomorphic Encryption Proceedings of CCIS201*
6. *Ivan Damgard, Valerio Pastro, Nigel Smart and Sarah Zakarias. Multiparty Computation from Somewhat Homomorphic Encryption . R. Safavi. Naini and R.Canetti (Eds.): CRYPTO 2-12,LNCS 7412,pp.643-662,201 2.International Association for Cryptologic Research 2012*
7. *Valtteri Niemi, Ari Renvall. Secure multiparty computation without computers. Theoretical Computer Science 19.1(1998) 173-183*
8. *Alfredo cuzzocrea, Elissa Bernito. Privacy Preserving OLAP over Distributed XML Data:A Theoretically-Sound Secure Multiparty-computation Approach. Journal of Computer and System Sciences 77(2011)965-987*
9. *Marcm Andrybowicz, Stefan Dziembowski. Daniel Malinowski, Lukasx Mazurek.Secure multiparty Computation on Bitcoin,2014 IEEE Symposium on security and privacy.*
10. *Ronald Cramer, Vanesa Daza Ignacio Gracia, Jorge Jimenez Urroz, Gregor Leader, Jaume Martin-Frame and Carles Pardo. On codes, Matriods and secure Multiparty Computation from Linear Secret -Sharing Schemes, IEEE TRANSACTIONS ON INFORMATION THEOREY, VOL. 54, NO.6, JUNE 2008.*
11. *Andreas Peter, Erik Tews and Stefam Katzenbessier. Efficiently Outourcing Multiparty Computation*
12. *Under Multikey Keys, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL 8, NO. 12, DECEMBER 2013*
13. *Ivan Damgard, Martin Geisler and Mikkel Kreigard Homomorphic encyption and secure compasion.*
14. *Andrea B.Alexandru Student Member,IEEE Konstantinos Gatsis, Member, IEEE, Yasser Sboukry, Member, IEEE, Sanjit A. Seshai, Fellow, IEEE, Paulo Tabula, Fellow, IEEE and George J,Pappas,Fellow, IEEE.Cloud-Based Quadratic Optimization with Partially Homomorphic Encyption.*
15. *Jean Lousis Raissro, Gwangbae Choi Sylvain pradwevand Raphal Colsenet, Natlle Jacquernot, Nicolas Rosat, Vincet Mooser, and Jean-Prierre hulxux Protecting Privacy and Security of Geonomic Data in i2b2 with Homomorphic Encryption and Differential Privacy*

AUTHORS PROFILE

Mr. A. Vijaya Kumar, Associate Professor, CSE Department, KL University. Email: vijay.cse@kl university.in

P. Ramya, CSE Department, K L University, Email: ramyapotru459@gmail.com

R. Poojitha, CSE Department, K L University, Email: pujithaaravipati@gmail.com

M. Lahari, CSE Department, K L University, Email: laharimarella21@gmail.com

ISSN 2278-3075

Inverted Indexing for Information Retrieval from Motifs and Domains of Proteins

Kumud Pant[1], **Bhasker Pant**[2], **Devvret Verma**[3], **Promila Sharma**[4], **Vikas Tripathi**[5]
[1]Department of Biotechnology, Graphic Era Deemed to be University, Dehradun, India.
[2]Department of Computer Science & Engineering Graphic Era Deemed to be University, Dehradun, India.
[3]Department of Biotechnology, Graphic Era Deemed to be University, Dehradun, India.
[4]Department of Biotechnology, Graphic Era Deemed to be University, Dehradun, India.
[5]Department of Computer Science & Engineering Graphic Era Deemed to be University, Dehradun, India.

ABSTRACT

The recent advancement in technologies are generating huge amount of data and extracting information from it is being outpaced by data accumulation. The development of hybrid approaches by combining different algorithms for extraction of required from the stock-pile of data is a demand of the hour. One such algorithm is vector space model for inverted indexing that has been used traditionally for search engine indexing in computers. In bioinformatics also it has been used for assembly of DNA fragments generated after sequencing. But it has not been applied for retrieval of relevant protein sequence to the query, based on presence or absence of motifs and domains in it.

In this paper the concept of inverted indexing has been applied on small motif/domain data of proteins contained in Motivated Proteins database at http://motif.gla.ac.uk/motif/index.html. The index has been built using 17 small hydrogen bonded motifs present in a dataset of 430 proteins. The entire dataset of 430 proteins has been divided into 19 classes. Seven classes' example cyanovirin, antibiotic and concavalin etc. had very few instances (1 or 2), hence have been omitted from further studies. Rest 12 classes with more than 10 proteins were considered further for testing information retrieval (IR) strategy. The document vector of all the proteins belonging to one class was averaged and 12 queries with averaged vector were prepared for testing. The similarity coefficient (SC) was then compared between query and all the proteins of the dataset.

This approach could successfully classify the query as belonging to the class from which it derived. To further validate the importance of document vector as novel attribute for classification, entire dataset of document vector was clustered to ten (10) clusters. Testing was then performed with similarity coefficient (SC) of the query with clusters obtained above. The allocation of cluster to the 12 query sequences followed the same pattern as done with relevant document search using inverted indexing approach. But clustering allocated the queries to only four (4) classes. Maximum number of query proteins (7 proteins or 58%) were found belonging to cluster 5.

Keywords : Information Retrieval, Motif/Domain, Clustering, Inverted Indexing Computing Classification System: I.4

I. INTRODUCTION

In a world laden with data, retrieval of useful data or information has been most demanding, as evident from development of new algorithms, models and software. Concepts from various branches have been incorporated for relevant data (Information) retrieval from biological data warehouse. This becomes especially important since all modern day high throughput technologies are outpacing information generation over data accumulation. The realm of information retrieval here involves picking a document (out of bulk) that is most similar to the input query. There may be many documents sharing great amount of similarity with query, but through similarity coefficient the weight or relevance of ranking can be used to ascertain the one most near to the query [1].

Traditionally IR has been used in digital libraries, media search, search engines and many more. Recently it has started to be used in biomedical literature, the information of which has been compiled on the Biomedical Literature Mining Publication (BLIMP) [2].

From a biologist point of view getting the highest match of a documents (protein or DNA or RNA sequence or macromolecular structure) with the query can be of great value. The Entrez search engine of NCBI devised by NCBI's John Wilbur and several bioinformatics including BLAST and FASTA search performs the same task of match finding [1].

However, specialized algorithms have been formulated as a part of information retrieval strategy. The most notable are Vector Space Model, Probabilistic Retrieval Strategies, Inference Networks, Language Models, Neural Networks, Extended Boolean Retrieval, Latent Sementic Indexing, Genetic Algorithms and Fuzzy Set Retrieval.

The above mentioned approaches use document vectors or vector space model for relevant document search. Document vectors represent frequency of occurrence of a term in the document. The inverted document frequency (idf) is then used to assign weight to each term in the query. The greater number of times a term appears in document the lesser is its weight, and the lesser times it appears the more is its weight.

Proteins, as biological entity, are macromolecules composed of multiple motifs and domains with each motif or domain present multiple times. With same motifs and domains present in both query and database an ideal biological information retrieval strategy can be formed. Although IR strategy for DNA alignment has been developed in the past with nucleotide sequence as attribute but no IR search with domain and motif information has been done till now [3].

To understand the role of frequency of occurrence of motif and domain information in relevant protein search, the vector space model has been implemented on protein small motif dataset obtained from http://motif.gla.ac.uk/. The Motivated Protein database contains a list of 430 proteins with various frequencies of 17 small hydrogen bonded motifs.

As a part of IR strategy an attempt have been made at generating inverted index for motif or domain based information of protein dataset obtained from Motivated Protein database [4]. The frequency of motif/domain information of 430 proteins multiplied with their inverse document frequency produced a document vector (DV). The DV constituted the training dataset. For testing the model, twelve (12) queries were made by averaging the document vector of all proteins belonging to a class in the training dataset itself. Hence it was a case of self-validation. The similarity coefficient (SC) between query and 430 proteins in the training dataset was calculated. The queries were said to belong to the class with which the similarity coefficient was the highest. Ten (10) out of twelve (12) proteins were found to belong to the respective classes from where derived but two queries were not able to locate their respective classes on the basis of similarity coefficient.

To analyze the working of inverted indexing and similarity coefficient based approach for protein classification, the entire dataset of 430 proteins with 17 attributes was considered without class label and was used to train an unsupervised clustering algorithm. On testing with document vector of 12 queries the document vector based classifier allocated the 12 query proteins into 4 classes. Maximum number of query proteins (7 proteins or 58%) were found belonging to cluster 5.

The unsupervised learning based clustering algorithm has been used on the above dataset to highlight the importance of document vector in partitioning data. Although SC based approach has been successful in classifying queries into their respective classes, application of unsupervised learning algorithm further strengthened the power of attribute frequency based methods for the same.

II. MATERIAL AND METHODS

1. The motif data set obtained from http://motif.gla.ac.uk/ was used to retrieve the comprehensive list of the small motifs found in 430 proteins obtained from PDB database.
2. Weka suite of software at http://www.cs.waikato.ac.nz, is a software for performing various data mining activities [5]. Clustering was performed in protein motif dataset for assigning groups to previously unsupervised data.

III. RESULTS

Initially an Inverted Index indicating the presence as well as frequency of all the 17 small hydrogen bonded motifs/domains in 430 proteins was built. The 17 motifs taken for the study are Alpha Beta Loop, Asx motif, Asx Turn, Beta Bulge, Beta Bulge Loop, Beta Bulge Turn, Beta Turn, Bulge Loop Motif, Crown Bridge, Crown Bridge Loop, Nest, Niche, Schellman Loop, ST Loop, ST Motif, ST Staple and ST Turn. The proteins have been taken from various classes and have various disease implications like cancer, schizophrenia and TB to name a few. The construction of inverted index avoids lengthy sequential scan through every document to find terms in the query. It is a look up table for every term occurring in the document.

The inverse document frequency of each term was thereafter calculated with the formula idf = log (d/dfi), where d is number of documents (430) and idf is the occurrence frequency of each motif in the entire document set. The idf for all the17 motifs is shown in Table I.

Table- I: The inverse document frequency for all 17 motifs

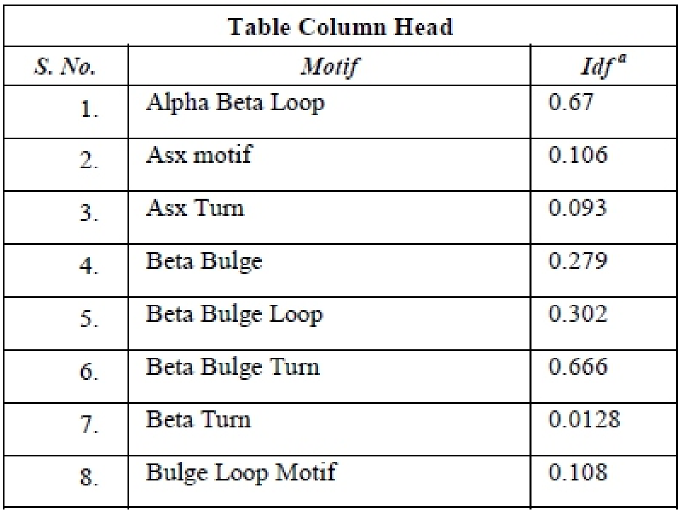


Table Column Head		
S. No.	*Motif*	*Idf*^{a}
1.	Alpha Beta Loop	0.67
2.	Asx motif	0.106
3.	Asx Turn	0.093
4.	Beta Bulge	0.279
5.	Beta Bulge Loop	0.302
6.	Beta Bulge Turn	0.666
7.	Beta Turn	0.0128
8.	Bulge Loop Motif	0.108

9.	Crown Bridge	1.047
10.	Crown Bridge Loop	1.2
11.	Nest	0.019
12.	Niche	0.0128
13.	Schellman Loop	0.171
14.	ST Loop	0.265
15.	ST Motif	0.156
16.	ST Staple	0.106
17.	ST Turn	0.15

a. idf : inverse document frequency

A document vector with seventeen terms for few proteins is shown in Table II. It reflects the importance of each term appearing in the document and is obtained by multiplying the idf with frequency of its occurrence in each document.

Table- II: Document vector for first 23 proteins

Motif → Protein ID ↓	ABL	AM	AT	BB	BBL	BBT	BT	BLM	CB	CBL	NT	N	SL	STL	STM	STS	STT
1BS9	0.4164	0.308197355	0.343001414	0	0	0	0.084950284	0	0	0	0.051222716	0.139824133	0.210810083	0.644420767	0.292308335	0.442972849	0
1BDO	0	0	0	1.138421981	0	1.294775921	0.050970171	0	0	0	0	0.038133855	0	0	0	0	0.0935
2ACT	0.2776	0.205464903	0	0.487895135	0.178199489	0	0.067960228	0	0	0	0.089639753	0.076267709	0.210810083	0.161105192	0.097436112	0.316409178	0.0935
1BTE	0	0.102732452	0.085750353	0.162631712	0	0	0.025485085	0	0	0	0.025611358	0.038133855	0	0	0	0	0
2ACY	0.1388	0	0	0.162631712	0	0	0.016990057	0	0	0	0.064028395	0.02542257	0.105405041	0.161105192	0	0.189845507	0.0935
1MUN	0.7634	0.102732452	0	0	0	0	0.059465199	0	0	0	0.102445432	0.146179776	0.210810083	0	0.292308335	0.379691013	0
1QB7	0.4164	0.102732452	0.085750353	0	0	0	0.042475142	0	0	0	0.102445432	0.152535418	0.105405041	0.483315575	0.487180559	0.126563671	0.187
2DPM	1.1104	0.308197355	0.343001414	0.162631712	0.178199489	0	0.050970171	0.332648823	0	0	0.102445432	0.088978994	0.105405041	0	0.584616671	0.189845507	0.0935
1BX4	0.9716	0.205464903	0.171500707	0	0	0	0.084950284	0	0.420050399	0.445688336	0.115251111	0.120757206	0.421620166	0	0.194872224	0.316409178	0.0935

1QHV	0	0.61639471	0.514502121	0.162631712	0.17819989	0	0.084950284	0.332648823	0	0	0.038417037	0.101690279	0	0.322210383	0	0.126563671	0.187
1ZIN	0.4164	0.205464903	0.085750353	0	0.356398978	0	0.042475142	0	0	0	0.140862469	0.050845139	0.421620166	0	0.389744447	0.189845507	0
1QS1	0.2776	0.719127161	0.171500707	0.97579027	0	0.971081941	0.152910512	0	0	0	0.153668148	0.190669273	0.316215124	0.96663115	0.292308335	0.316409178	0.374
1OUW	0	0	0.171500707	1.138421981	0.17819989	0	0.033980114	0	0	0	0.025611358	0.02542257	0	0	0	0	0.0935
1BD0	0.4164	0.410929806	0.343001414	0.813158558	0	0.647387961	0.135920455	0	0.420050399	0	0.12805679	0.152535418	0.527025207	0.161105192	0.389744447	0.4429728	0.0935
3BTO	0.6246	0.410929806	0.600252474	0.487895135	0.17819989	0.32369398	0.152910512	0	0.420050399	0.445688336	0.243307902	0.184313631	0.421620166	0.483315575	0	0.506254684	0.2805
1DOS	0.9716	0.205464903	0.171500707	0	0	0	0.067960228	0	0	0	0.166473827	0.076267709	0.421620166	0	0.097436112	0.696100191	0.187
1ADS	0.5552	0.719127161	0.25725106	0	0.17819989	0	0.11043537	0	0	0	0.243307902	0.165246703	0.632430249	0.161105192	0.194872224	0.316409178	0.187
1VJS	0.7634	1.130056967	0.943253888	0.487895135	0.356398978	0.32369398	0.144415484	0.665297646	0	0	0.294530618	0.222447485	0.316215124	0.322210383	0.097436112	0.442972849	0.374
1PEN	0.2082	0	0	0	0	0	0.008495028	0	0	0	0	0.012711285	0	0	0	0	0
1QQ4	0.0694	0	0.171500707	1.626317116	0.534598467	1.294775921	0.11043537	0	0	0	0.064028395	0.076267709	0.105405041	0	0	0	0.0935
1TUD	0	0	0.171500707	0.325263423	0	0.32369398	0.016990057	0	0	0	0	0.044489497	0	0	0	0	0
1MRJ	0.5552	0.513662258	0.514502121	0.162631712	0.534598467	0	0.084950284	0.665297646	0.420050399	0.445688336	0.076834074	0.108045921	0	0.322210383	0.097436112	0.759382027	0
1AAC	0	0	0.085750353	0.325263423	0	0.647387961	0.050970171	0	0	0	0.012805679	0.057200782	0	0	0	0	0

ABL: Alpha beta loop, AM: Asx motif, AT: Asx Turn, BB: Beta Bulge, BBL: Beta Bulge Loop, BBT: Beta Bulge Turn, BT: Beta Turn BLM: Bulge Loop Motif, CB: Crown Bridge, CBL: Crown Bridge Loop, NT: Nest, N: Niche, SL: Schellman Loop, STL: ST Loop, STM :ST Motif, STS: ST Staple, STT: ST TURN

To further implement this vector space model a set of twelve (12) hypothetical queries was prepared that are shown in Table III. The proteins in the Motivated Protein database have been obtained from PDB database at www.rcsb.org [6]. The PDB classified the 430 proteins into more than 19 classes. Few classes had one or two proteins only, therefore, in an alternative strategy only 12 protein classes out of 19, from PDB having more than 10 instances, were considered for similarity coefficient (SC) based analysis. The document vectors of all proteins belonging to one class were averaged. The averaged vector comprised a query vector for testing. Following the similar protocol 12 queries from 12 protein classes were prepared. Now the similarity coefficient of 12 queries was calculated.

Table- III: Twelve (12) queries prepared by averaging document vector of all proteins belonging to one class

Small Motif → Query No ↓	Alpha beta loop	asx motif	asx turn	beta bulge	beta bulge loop	beta bulge turn	beta turn	bulge loop motif	Crown bridge	Crown bridge loop	Nest	Niche	Schellman loop	ST Loop	ST Motif	ST Staple	ST Turn	Class
Q1	1.34	0.1855	0.11625	0.279	0.2265	0	0.064	0	0	0	0.09025	0.0928	0.21375	0	0.156	0.0795	0.1875	Hydrolase
Q2	0.67	0.159	0.093	0.279	0.302	0	0.0448	0	0	0	0.133	0.0704	0.2565	0	0.156	0.106	0.225	Electron transport
Q3	4.69	0.689	0.465	0	0	0	0.1344	0	0	0	0.2185	0.16	0.7695	0.3975	0.234	0.583	0.225	Sugar binding protein
Q4	1.34	0.053	0.0465	0.1395	0	0	0.0448	0	0	0	0.0475	0.0768	0.171	0	0.234	0.106	0.225	DNA binding protein
Q5	6.03	0.848	0.372	0.6975	0	0.666	0.1792	0	1.047	0.6	0.266	0.3392	1.1115	0.3975	0.468	0.689	0.15	Isomerase
Q6	7.37	0.689	0.8835	0.1395	0.755	0	0.1984	0.054	0	0	0.285	0.4096	0.7695	0.265	0.78	0.901	0.45	Lyase
Q7	0.67	0.053	0.1395	0	0	0	0.064	0	0	0	0.095	0.0832	0.0855	0.265	0.468	0.159	0.3	Metal binding protein
Q8	3.015	0.53	0.651	0.9765	0.453	0.333	0.1664	0.054	0.5235	0.6	0.228	0.2496	0.4275	0.53	0	0.53	0.225	Oxidoreductase
Q9	5.36	0.689	0.372	0.4185	0.453	0	0.1472	0	0	0	0.1425	0.2816	0.342	0	0.312	0.265	0.3	Transferase
Q10	0.67	0.318	0.279	0.1395	0.151	0	0.064	0.054	0	0	0.0475	0.1088	0.0855	0.265	0	0.265	0.15	Viral protein
Q11	0.67	0.159	0.0465	0	0.151	0	0.0192	0.054	0	0	0.019	0.0064	0.0855	0	0.156	0.053	0	Transcription regulator
Q12	1.675	0.053	0.186	0.1395	0.302	0	0.0576	0.054	0	0	0.076	0.0832	0	0.1325	0.078	0.106	0.15	Toxin

The SC based approach successfully retrieved relevant proteins to the query. The top three similarity coefficient based relevant documents for the query are shown in Table IV. The similarity coefficient (SC) reflecting relevance of the query with each of the 430 sequences in the database was calculated by multiplying document vector of each sequence with query vector shown in table 2. On summing up the similarity coefficient (SC) for every document, the highest SC was given the greatest significance or a particular query was found belonging to the document having the highest SC.

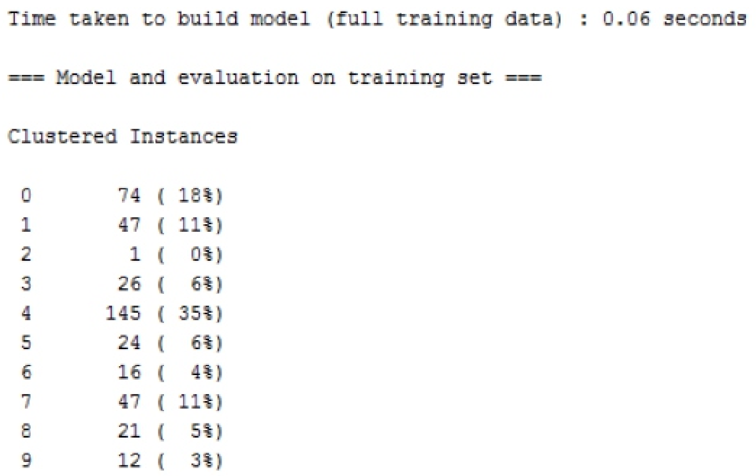
Table- IV: Relevant document allocation on the basis of similarity coefficient

Query (Initial Class)	Similarity coefficient (highest similarity with protein ID) (Class)	Similarity coefficient (second highest similarity)	Similarity coefficient (third highest similarity)
Q1 (Hydrolase)	1SMD (19.42) (Hydrolase)	1VJS (15.30) (Hydrolase)	1CS1 (14.62) (Lyase)
Q2 (Electron transport)	1PLC (19.082) (Electron Transport)	1RIE (15.0058) (Electron Transport)	1JER (14.47) (Electron Transport)
Q3 (Sugar Binding Protein)	1OUW (3.266)(Sugar Binding Protein)	1A7S (3.212) (Ligand Binding Protein)	2IGD (3.198) (Igg Binding Protein)
Q4 (DNA Binding Protein)	1C1K (17.78) (DNA Binding Protein)	3HTS (15.17) (DNA Binding Protein)	1AAC (14.5214) (Electron Transport Protein)

Q5 (Isomerase)	1QRE (19.18) (Lyase)	1HMT (14.97) (Lipid Binding Protein)	1AYL (14.31) (Lipid Binding Protein)
Q6 (Lyase)	4XIS (17.85) (Isomerase)	1QIP (15.02) (Lyase)	3STD (14.50) (Lyase)
Q7 (Metal Binding Protein)	5ICB (19.30) (Metal Binding Protein)	1CYD (15.05) (Oxidoreductase)	1A8E (14.59) (Metal Transport)
Q8 (Oxidoreductase)	1CYD (18.25) (OxRd)	1B4V (15.19) (OxRd)	1ADS (14.57) (OxRd)
Q9 (Transferase)	1QB7 (19.28) (Transferase)	1A8D (15.19) (Toxin)	1BGF (14.57) (Transcription Regulator)
Q10 (Viral Protein)	1EGW (19.19) (Transcription Regulator)	1XWL (15.19) (Tranferase)	1YTB (14.42) (Transcription Rregulator)
Q11 (Transcription regulator)	1MOF (16.2481863) (Viral Protein)	1MFI (15.2345) (Viral Protein)	1BKB (15.05) (Translation Initiation Factor)
Q12 (Toxin)	3SEB (17.234) (Toxin)	1KPT (17.04567) (Toxin)	1PEN (15.0534) (Toxin)

For further validating the importance of document vector as classification attribute the document vector for all 430 proteins were clustered using WEKA 3.8.1. The dataset was considered to be unsupervised, since no class labels were initially provided. The k-means method of clustering was adopted for the study since it produces tighter clusters with large amount of attributes [7]. With document vector for 430 proteins as training dataset, ten (10) clusters were prepared as shown in Figure 1.

The test data set comprised of query vectors for all the 12 queries. The allocation of cluster to the 12 query sequences followed the same pattern as done with relevant document search using inverted indexing approach. But clustering allocated the queries to only four (4) classes. The results for clustering analysis are shown in Figure 2. Maximum number of query proteins (7 proteins or 58%) were found belonging to cluster 5



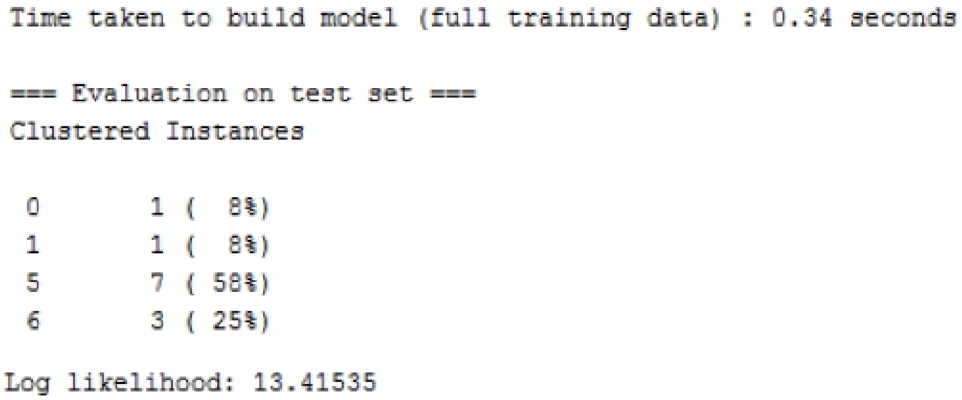
```
Time taken to build model (full training data) : 0.06 seconds

=== Model and evaluation on training set ===

Clustered Instances

 0       74 ( 18%)
 1       47 ( 11%)
 2        1 (  0%)
 3       26 (  6%)
 4      145 ( 35%)
 5       24 (  6%)
 6       16 (  4%)
 7       47 ( 11%)
 8       21 (  5%)
 9       12 (  3%)
```

Fig. 1. Cluster information obtained from Weka 3.8.1 (Eibe Frank et.al.; 2013)



```
Time taken to build model (full training data) : 0.34 seconds

=== Evaluation on test set ===
Clustered Instances

 0       1 (  8%)
 1       1 (  8%)
 5       7 ( 58%)
 6       3 ( 25%)

Log likelihood: 13.41535
```

Fig. 2. Cluster allocation for 12 query proteins

IV. DISCUSSION AND CONCLUSION

Initially the 413 proteins downloaded from http://motif.gla.ac.uk/ on careful observation were found to belong to more than 19 different classes. Few classes had very few instances example class Cyanovirin had only 1 instance. Class antibiotic and Concavalin had 2. Therefore, only those classes were considered for analysis that had more than 10 protein instances. There were twelve (12) classes that fulfilled this criterion, as shown in table 3. The document vector of proteins belonging to one class was averaged. This was repeated for all 12 classes to generate 12 queries. The similarity coefficient of all the queries with all the proteins was generated. 10 protein queries namely were able to locate their respective class of origin. On careful observation it was found that query 5 and query 6 obtained from class Isomerase and Lyase shared document vector at many places. Similarly document vector for query 10 and query 11 of viral protein and transcription regulator shared many values. Hence queries on the basis of similarity coefficient could not be distinguished into two classes.

To analyze the importance of document vector as an attribute for classification, the dataset was made to undergo unsupervised learning in the form of clustering. The training of k-means algorithm for clustering was performed with document vectors as attributes. The classifier clustered the dataset into 10 instances. On testing Weka clustering algorithm with vectors of query, the k-means algorithm allocated the query into 4 groups as shown in figure 2. Although clustering based classifier could not very well distinguish the query into respective groups, it could partly be due to less demarcation between document vectors. Hence document vector can prove to be an effective attribute for protein classification. This approach can also help in classifying a new protein with motif or domain information, as belonging to a particular class.

The implementation of vector space model using inverted index can be very well applied to big datasets. It can avoid lengthy sequential search through every document to find the most relevant one to the query.

ACKNOWLEDGMENT

The authors express the deep sense of gratitude to the Department of Biotechnology, Department of Computer Science & Engineering of Graphic Era Deemed to be University for all the support, assistance, and constant encouragements to carry out this work.

REFERENCES

1. *Nadkarni, P. M., 2002, An introduction to information retrieval: applications in genomics. The Pharmacogenomics Journal. 2(2), 96–102.*
2. *Biomedical Literature Mining Publication (BLIMP) at http://blimp.cs.queensu.ca/.*
3. *Ling, W., Zhao KaiYong., 2013, A new DNA alignment method based on inverted index. At https://arxiv.org/abs/1307.0194v1.*
4. *Motivated Protein database at http://motif.gla.ac.uk/.*
5. *Frank, E., Mark A. Hall., Ian H. Witten., 2016, The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016. The Weka suite of software at http://www.cs.waikato.ac.nz.*
6. *RCSB Protein Data Bank at www.rcsb.org.*
7. *Technology for you at http://playwidtech.blogspot.in/2013/02/k-means-clustering-advantages-and.html.*

AUTHORS PROFILE

Dr. Kumud Pant is working as an Assistant Professor in the Department of Biotechnology, Graphic Era University, Dehradun, Uttarakhand, India. She has completed her Masters in Biotechnology from Jiwaji University, Gwalior and PhD in Bioinformatics from MANIT, Bhopal in 2014. She got air fare Grant from CSIR for attending IC2IT held at Bangkok, Thailand in 2009. She is a supervisor for two (2) PhD Scholar. She has successfully completed a project entitled "Reverse vaccinology approach for detecting major virulent proteins of Encephalitis virus", granted by UCOST, Dehradun. Her area of research is Bioinformatics, Computational Biology and Data Mining. She has more than 25 research publication in reputed journals.

Dr. Bhasker Pant completed his graduate from M.B. Khalsa College, Indore, post graduate from Barkatullaha University, Bhopal, and doctoral studies at Maulana Azad National Institute of Technology, Bhopal, India in 2000, 2003 and 2012 respectively. He has done significant research work on Data Mining, Machine Learning, and Bioinformatics. His research interest includes Internet-of-Things and soft-computing too. He has 7 doctoral students and supervised more than 20 masters students. He has delivered invited talks, guest lectures and presented his research results in various counties like China, Singapore and India. He has more than 40 publications in top quality international conferences and journals. He is holding senior level academic positions the university time to time. He is currently the Dean (Research and Development) in Graphic Era University. He is a recipient of the Eminent Research Award (Computer Science & Engineering Engineering) of the Graphic Era University.

Devvret Verma is a research scholar in the Department of Biotechnology, Graphic Era Deemed to be University Dehradun. He did B.Tech in Bioinformatics from Amity University, Uttar Pradesh and M.Tech. in Biotechnology from Graphic Era Deemed to be University, Dehradun. He has written 2 book chapters and published 5 Research Papers in peer reviewed journals. He has got a Young Scientist Award through UCOST in 11th Science Congress. He had secured second Best Oral Presentation award in conference organized by Uttaranchal University Dehradun. His area of research is Bioinformatics, Drug Designing and Data Mining.

Dr. Promila is working as an Associate Professor in Department of Life Sciences, Graphic Era University, Dehradun, Uttarakhand, India. She did her B.Sc (H) Human Biology from AIIMS, New Delhi. She completed her postgraduation in Biochemistry from Jamia Hamdard, New Delhi and Ph.D from Department of Biotechnology, Jamia Hamdard, New Delhi only. Her area of research is Cell and Molecular Biology. Currently she is working on NO and nitric oxide synthase. She has written three book chapters and published 13 research papers in peer reviewed journals. She is also member of editorial board of International Journal of Plant Biotechnology.

Dr. Vikas Tripathi has done BE in information technology from Technocrats institute of technology, Bhopal, M. Tech in Software engineering from Indian institute of information technology Gwalior and PhD from Uttarakhand technical university, Dehradun. He is actively involved in research related to Software engineering, Computer Vision, Machine learning and Video Analytics. He has published many papers in reputed international conferences and journals. Currently he is working as an associate professor in Graphic era deemed to be university Dehradun, India.

ISSN 2278-3075

Diagnosis of Parkinson's Disorder through Speech Data using Machine Learning Algorithms

Abhishek M. S[1], Chethan C. R[2], Aditya C. R[3], Divitha D[4], Nagaraju T. R[5]
[1]Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysore, India.
[2]Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysore, India.
[3]Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysore, India.
[4]Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysore, India.
[5]Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysore, India.

ABSTRACT

Parkinson's disease is a neurodegenerative disorder that affects millions of people around the globe. Detecting Parkinson's disease at an earlier stage could help to better diagnose the disease. Machine learning provides potentially large opportunities for computer-aided identification and diagnosis that could minimize unavoidable health care errors and inherent clinical uncertainty, provide guidance, and improve decision-making. In this paper, we explore the feature extraction and prediction algorithms used to predict Parkinson's disease and provide a comprehensive comparison of these algorithms.

Keywords: Parkinson's disease, PD, datasets, SVM, KNN, Genetic Algorithm, UPDRS.

I. INTRODUCTION

Parkinson's disease is a neurological dysfunction of the central nervous system that affects the physical movements of the body. The motor system is the series of central and peripheral nervous system structures that support motor, i.e. movement functions. Skeletal muscle and neural tissue connections may include peripheral structures. PD is caused by the loss of nerve cells in the brain, known as substantial nigra, which releases a chemical known as dopamine. Dopamine serves as a messenger to regulate and organize body movements between the brain and nervous system parts. If the nerve cells are destroyed or damaged, dopamine will diminish in the brain, causing slow and abnormal movements .Worldwide, over millions of people affected by PD. The occurrence of Parkinson's disease increases exponentially, but it is estimated that 4% of people with PD have been diagnosed before 50 years of age. In Men, the chance of Parkinson's disease is 1.5 times greater than in women. Usually, the symptoms appear slowly. Screams, slow motion (bradykinesia), muscle rigidity, weakened equilibrium and stability, motor function impairment, voice improvements, changes in writing are some of the main symptoms. Machine learning provides almost enormous potential for computer-aided detection and diagnosis that could minimize unavoidable errors and latent clinical uncertainty in healthcare, provide feedback and speed up decision-making. Voice information [1-3], gait patterns [4-5], force monitoring data [6], single-photon emission computed tomography (SPECT) scan data [7] and odor recognition data have been used to solve the problem of PD identification via machine learning techniques.

A speech dataset can be used accurately to determine whether or not a person suffering from Parkinson's disease [8-10] [16-17]. Factors such as voice pitch, jitter, and frequency can be used to distinguish PD patients from normal persons. Considering different attributes, supervised and unsupervised training methods are used to assess the person on a generic measurable scale. In this paper, we explore different types of machine learning models applied to speech datasets to analyze and predict the symptoms of Parkinson's disease. predict the symptoms of Parkinson's disease.

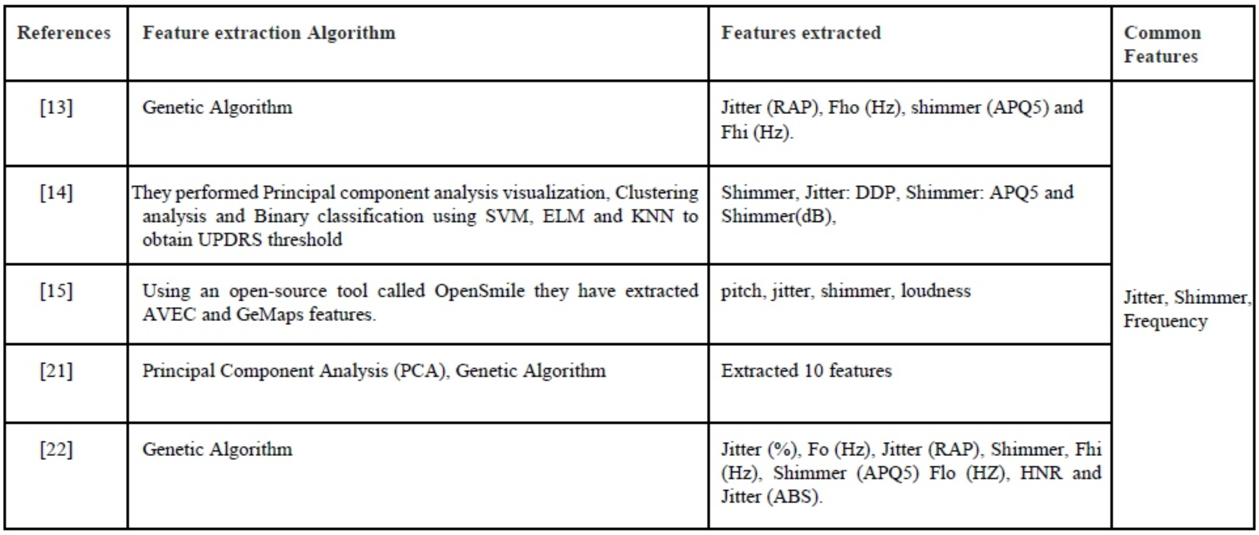
II. METHODOLOGY FOR VOICE FEATURE EXTRACTION

Mohammad Shahbakhi et al. suggested a new algorithm based on voice analysis to diagnose PD [8]. The first step is to select the optimized characteristics of all extracted characters using a genetic algorithm (GA). The genetic algorithm is an efficient heuristic algorithm focused on the choice of human and genetic ideas for evolution. It is one of the most important methods are used in the classification of data to pick optimized features. Mohammad Shahbakhi et al. used a sample of 195 continuous vowel phonations of 31 men and women, of which almost 23 were diagnosed with PD. An average of six phonations of 1 to 36 seconds was reported in each subject. Phonations were recorded at 8 cm from the lips on an IAC sound-processed booth with a head-mounted microphone (AKG C420). The speech signals have been captured with the 16-bit resolution directly to the machine with CSL 4300B (Kay Elemetrics) equipment. Betul Erdogdou Sakar et al. proposed a model [9] with a data set of 42 PD patients tracked at weekly intervals over 6 months. The PD suffered patients who were analyzed within the past five years at the preliminary beginning on the off chance that he/she had any two of the accompanying side effects: stiffness, rest tremor, without proof of different types of Parkinson's. Some simple frequency measurements, different amplitude varying measures, harmonic-to-noise ratios, signal fractal scaling exponent, nonlinear dynamic complexity analysis, and pitch entropy are included in the feature setThe Parkinson's (UPDRS) threshold is first determined in the differential grouping. They use a two-stage approach in their model. Firstly, they determine the optimum UPDRS threshold value, which can be distinguished by the vocal function with the lowest possible error rate. Such functions are provided for Support of Vector Machinery (SVM), Extreme Learning Machines (ELM) and k-classifiers (K-NN) is used for each of the binary classification problems solved by various UPDRS values. They also take into consideration the metric of the Matthew Correlation Coefficient (MCC) to determine the maximum predictable threshold value of UPDRS. Finally, simulation and clustering of the main components are conducted in order to find the desired threshold values. The significance of the human vocal characteristics is later quantified using Mutual Information (MI) with the discrete UPDRS ranking. Therefore, the vocal characteristics that change dramatically concerning the degree of motor system disorders as determined by UPDRS have been defined for PD detection.Timothy J. Wroge et al. [10] suggested a system where the human voice was captured and the raw audio is actively removed with the Voice Activation Detection (VAD) algorithm [11] to isolate and eliminate background noise from the audio until feeding into the feature extraction algorithm. Methods derived from the 2013 Audio-Visual Emotion Recognition Challenge (AVEC) [12] are used for preliminary audio analysis and the Minimum Redundancy Maximum Relevance (mRMR) [13] method is applied to these 2013 AVEC audio features. After that, the cleaned audio is transmitted through two separate extraction algorithms before entering the system models. Then, the raw audio is transferred to the Geneva Minimalistic Acoustic Parameter Set (GeMaps) algorithm [14] for extracting the features using an open smile toolkit [15] before it is sent to machine learning patterns. The GeMaps feature algorithm extracted lower-level characteristics, including loudness, pitch, shimmer, jitter, and harmonic-to-noise ratio.Satyabrata Aich et al. have implemented a model [16] which helps to classify Parkinson affected people from normal people based on voice datasets. The model's database contains speech samples for 31 individuals, 23 of whom have been diagnosed with PD. For data reduction and function collection, they have used two algorithms. First, they have used Primary Component analysis (PCA) which reduces the space dimension in order to visualize the data in a small dimensional space. They also used feature sets based on the genetic algorithm (GA) to pick features.

R. Arefi Shirvan et al [17] proposed a method for detecting PD using the Genetic Algorithm and KNN Classification Method. They collected voice signal samples of healthy people and Parkinson affected people and they extracted various features such as Fo (Hz), Jitter (%), Flo (HZ), Jitter (ABS), Shimmer, Fhi (Hz), Jitter (RAP), Shimmer (APQ5) and Harmonics-to-Noise Ratio (HNR) from the voice signals. Later, using the genetic algorithm, they extracted the optimized features which mainly affect the process of classification of data.

Table 1 shows the features extracted from the voice signal for PD prediction. The common features which are extracted and used for the disease prediction are Jitter, Shimmer, Frequency.

Table- I: Features Extracted from speech signals for Parkinson Disease Prediction



References	Feature extraction Algorithm	Features extracted	Common Features
[13]	Genetic Algorithm	Jitter (RAP), Fho (Hz), shimmer (APQ5) and Fhi (Hz).	Jitter, Shimmer, Frequency
[14]	They performed Principal component analysis visualization, Clustering analysis and Binary classification using SVM, ELM and KNN to obtain UPDRS threshold	Shimmer, Jitter: DDP, Shimmer: APQ5 and Shimmer(dB),	
[15]	Using an open-source tool called OpenSmile they have extracted AVEC and GeMaps features.	pitch, jitter, shimmer, loudness	
[21]	Principal Component Analysis (PCA), Genetic Algorithm	Extracted 10 features	
[22]	Genetic Algorithm	Jitter (%), Fo (Hz), Jitter (RAP), Shimmer, Fhi (Hz), Shimmer (APQ5) Flo (HZ), HNR and Jitter (ABS).	

III. PREDICTION MODELS

In [8] the Support Vector Machine (SVM) for the Gaussian kernel function is used after the optimized features were removed. It is a two-layered neural network using a hidden layer of radial units and a single output neuron. The way to build that network is organized and its parameters are learned. The test samples are classified with 94.50 percent precision per 4 optimized features, as well as 93.66% per seven optimized features and 94.22% per nine optimized features: jitter (RAP), Fhi (Hz), shimmer (APQ5) and Fho (Hz) provide the best grading accuracy. In the case of [9], samples of patients that have a UPDRS score above that threshold are removed, and a new dataset of samples whose UPDRS score is lower than this threshold, consisting of PD patient environments, and of eight healthy subjects is used for analysis. The analysis assesses the efficacy of speech characteristics in discrimination between early PD suffered patients and healthy people with classifiers ELM, k-NN, and SVM, as well as shows the accuracy and Matthew Correlation Coefficient (MCC) of each classifier for various parameter values and types of kernels.In [10], 90 percent practice and 10 percent of research were separated following the implementation of the extraction algorithm. The dataset is presented to various models such as Random Forest, Artificial Neural Network, Decision Tree Classifier, Gradient Boosted Classifier, Extra Tree Classifier, and SVM. Among these Gradient Boosted Decision Tree provided the best classification accuracy of about 86%.In [16], the various classifications have been used: C4.5, Bagging Cart, PART, BoostedC5.0, RPART, Random Wood and SVM. All classifications are non-linear decision-making tree-based classification with the exception of SVM. SVM is the most popular method of linear grading. Throughout grouping, it uses different kernel functions. They used the kernel function radial base in this paper. They compared performance measures such as specificity, precision, positive predictive value (PPV), sensitivity and negative predictive value (npv) following a different classification approach. For GA-based feature sets, SVM classifiers offer maximum reliability of 97.57%.The below flow diagram shows the flow chart of the system for PD prediction.

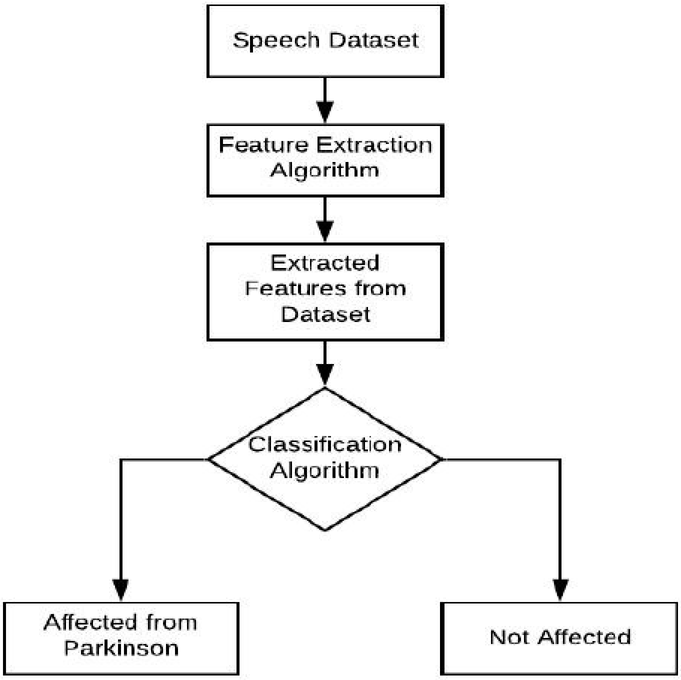


Figure 1 System design for PD prediction

In [17] the information classification is performed using the KNN classification system after the introduction of the genetic algorithm. The highest percentage of precision achieved is 98.2 per 9 optimized features. The 9 features showing the greatest accuracy include Fhi (Hz), Jitter(percent), Flo (HZ), Jitter (ABS), Shimmer, Fo (Hz), Jitter (RAP), Shimmer (APQ5) and HNR. Also, a 93.7 percent accuracy per four optimized features, with a 94.8 percent accuracy per seven optimized features is obtained.

IV. RESULT ANALYSIS

The comparison of different Parkinson's speech prediction classification algorithms is shown in Figure 2. Various algorithms for machine learning have been used in the prediction of Parkinson's disease. In the KNN classifier [17], 9 tailored characteristics are usable at the maximum level of accuracy for Parkinson's disease prediction. The SVM classifier provides 97.5 percent accuracy and has the nearest performance to KNN [17]. Figure 2 shows the comparison of the exactness of the various algorithms.

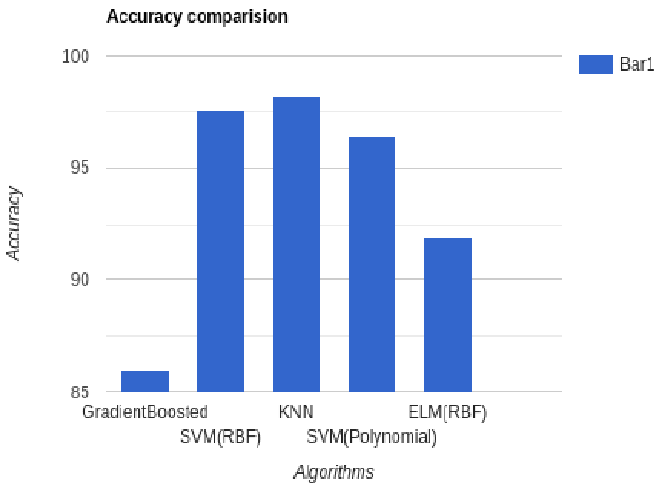


Figure 2 Accuracy Comparison

V. CONCLUSION

A genetic algorithm is the most common method used to extract voice signal characteristics. For analysis, the common features considered are jittering, shimmering and voice signal frequency. Various algorithms for machine learning have been used in the prediction of Parkinson's disease. The range of prediction accuracy from applied algorithms lies between 85% and 95%, which demonstrates that machine learning algorithms can be used to predict Parkinson's disease by analyzing voice signals.

REFERENCES

1. *Alex Frid, Edmond J. Safra, Hananel Hazan, et al. "Computational Diagnosis of Parkinson's Disease Directly from Natural Speech Using Machine Learning Techniques", in IEEE International Conference on Software Science, Technology and Engineering, 2014.*
2. *Made Satria Wibawa, Hanung Adi Nugroho, Noor Akhmad Setiawan, "Performance Evaluation of Combined Feature Selection and Classification Methods in Diagnosing Parkinson Disease Based on Voice Feature", in ICSITech, 2015.*
3. *HakanGunduz, "Deep learning based Parkinson's Disease classification using vocal feature sets", in IEEE.*
4. *Peng Ren, Esin Karahan, Chao Chen, Ruixue Luo, et al. "Gait Influence Diagrams in Parkinson's Disease", in IEEE.*
5. *Juliana P. Felix, Flavio H. T. Vieira, Alisson A. Cardoso, et al. "A Parkinson's Disease Classification Method: An Approach Using Gait Dynamics and Detrended Fluctuation Analysis" in IEEE (CCECE), 2015.*
6. *.R. Brewer, S. Pradhan, G. Carvell, A. Delitto, "Feature selection for classification based on fine motor signs of Parkinson's disease", in Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009.*
7. *I.A. Illan, J.M. Gorrz, J. Ramirez, F. Segovia et al, "Automatic assistance to Parkinson's disease diagnosis in DATASCAN SPECT imaging".*
8. *Mohammad Shahbakhi, Danial Taheri Far, et al, "Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine", February 2014.*
9. *Betul Erdogdu Sakar, Gorkem Serbes, C., et al. "Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease", August 9th, 2017.*
10. *Timothy Wroge, Yasin Serdar Ozkanca, David C Atkins, "Parkinson's Disease Diagnosis Using Machine Learning and Voice", November 2018.*
11. *Xiaoling Yang, Baohua Tan, Jiehua Ding, et al, "Comparative Study on Voice Activity Detection Algorithm", International Conference on Electrical and Control Engineering, 2009.*
12. *M. Valstar, B. Schuller, K. Smith, F. Eyben, et al, "Avec 2013: the continuous audio/visual emotion and depression recognition challenge," in ACM international workshop on Audio/visual emotion challenge. ACM, 2013.*
13. *H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," in IEEE Transactions on pattern analysis and machine intelligence, 2005.*
14. *Florian Eyben, Klaus Scherer, Bjorn Schuller, Johan Sundberg, et al, "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing", in IEEE Transactions on Affective Computing, 2016.*
15. *Florian Eyben, Martin Wollmer, et al, "Opensmile: the Munich versatile and fast open-source audio feature extractor", in 18th ACM international conference on Multimedia, 2010.*
16. *Satyabrata Aich, Hee-Cheol Kim, Ahmed Abdulhakim et al "A Supervised Machine Learning Approach using Different Feature Selection Techniques on Voice Datasets for Prediction of Parkinson's Disease", in ICACT Transactions on Advanced Communications Technology (TACT), May 2018*
17. *R. Arefi Shirvan, E. Tahami, "Voice Analysis for Detecting Parkinson's Disease Using Genetic Algorithm and KNN Classification Method", in 18th Iranian Conference on BioMedical Engineering, December 2011, Iran.*

AUTHORS PROFILE

Abhishek M. S. is currently pursuing a Bachelor of Engineering in Computer Science at Vidyavardhaka College of Engineering, Mysuru. His research interests include Machine Learning, Networks, Blockchain.

Chethan C. R. is currently pursuing a Bachelor of Engineering in Computer Science at Vidyavardhaka College of Engineering, Mysuru. His research interests include Machine Learning, Networks, Blockchain.

Dr. Aditya C. R. is an associate professor of computer science and engineering at Vidyavardhaka College of Engineering. He is a member of AAAI and his areas of research include machine learning and medical informatics.

Divitha D is currently pursuing a Bachelor of Engineering in Computer Science at Vidyavardhaka College of Engineering, Mysuru. Her research interests include Machine Learning, Deep Learning.

Nagaraju T. R. is currently pursuing a Bachelor of Engineering in Computer Science at Vidyavardhaka College of Engineering, Mysuru. His research interests include Machine Learning, Deep Learning.

Real-Time Remote Healthcare and Telemedicine Application Model using Ontology Enabled Clustering of Biomedical & Clinical Documents

R. Sandhiya[1], M. Sundarambal[2]

[1] Assistant Professor, Department of Information Technology, Coimbatore Institute of Technology, Coimbatore -641014, Tamilnadu, India.

[2]Professor, Department of Electrical and Electronics Engineering, Coimbatore Institute of Technology, Coimbatore -641014, Tamilnadu, India.

ABSTRACT

Remote health monitoring has become a hot topic research due to its multi-dimensional benefits to the society. This paper is aimed at developing a novel remote health monitoring model through wireless sensor networks to ensure efficient telemedicine process. The proposed model, Real-time Remote Healthcare and Telemedicine (RRHT) utilizes the concept of model based design to provide low cost and time saving efficiency. First the low power consuming sensor nodes are placed at specified body points with facility to monitor and reduce the power consumption at each stage of the designed model. These nodes collect the patient data and transmit them in wireless medium through the gateway where the data are combined to form documents/notes. Autonomous optimized routing algorithm is employed at this stage for transmission through efficient wireless paths to the processor connected at the hospitals or health centers. At the processor, the transmitted patient data documents are clustered using ontology enabled clustering models using chicken swarm optimization (CSO) and genetic chicken swarm optimization (GCSO). The clustered results are comparatively analyzed with the previous patient database and to determine the change in health readings. Based on these findings, the suitable medication details along with advice on hospital visits are suggested by the decision module and are sent to the physicians or medical experts for approval and further diagnosis. The performance analysis shows that the proposed RRHT system with GCSO clustering is highly reliable and accurate with better speed and lower cost. These results also prove that the RRHT significantly improved the healthcare application through the utilization of better strategies in document clustering of patient data.

Keywords: Telemedicine, Remote health care, wearable sensors, Autonomous optimized routing, ontology enabled clustering, decision support systems.

I. INTRODUCTION

Tele-health and telemedicine concepts are being widely utilized in the fields of medicine where remote health care applications are widely regarded as the future prospects [1], [2]. These two concepts are vital components of remote healthcare system [3]. As per the civil laws, it is compulsory to ensure basic medical facilities to each person in country. But due to various factors of financial crisis and practical difficulties, this basic medical right is just a literal guarantee to the people living in remote locations. Remote health care technologies can be a greater solution to this social issue in most developing and under developed countries [4]. Even the developed countries can utilize such technologies during the periods of natural calamities and emergency scenarios. However the major factors influencing against such systems are varied which increases the research interest in them [5]. The practical feasibility, cost, reliability and adaptability of these systems are major challenges that increase the research society to come up with more liable options. Developing such options can increase the applications of e-health systems in hospitals, clinics, nursing homes, e-health centers, rural health centers, even schools and prisons.

In recent years, wireless sensor network (WSN) has been utilized in various monitoring applications in military, banking and security services [6]. Based on these applications, researchers have utilized WSN for secured transmissions in medical fields [7]. Recently, it has been extensively utilized for the telemedicine and telecare applications due to its simplified operation and efficient throughput in improving quality of health care processes [8], [9]. Even the wireless multimedia sensor networks with efficient multi-path routing [10] can also be used. The affordable price of the personal computers and associated devices along with easy internet access has made it feasible to use WSN and other wireless networks in health care. However there are many limitations and challenges in power consumption and accuracy that need to be resolved to improve the efficiency [11]. This can significantly contribute towards the doctors to reduce their workload in analyzing the medical data of the patients.

In this paper, a novel remote health care and telemedicine model has been developed using the wireless sensor networks for the transmission of patient data. Sensor nodes are placed at different points of the patient body to collect health data readings. These are combined into clinical records and transmitted through the gate-way using Autonomous optimized routing algorithm based on self-optimization concept [12] in wireless sensor networks. Transmitted data are collected and fed to the Processor containing the decision support system and clustering module where the ontology enabled clustering algorithms cluster and analyses the data with the previous patient data to determine the changes in health condition. The clustering algorithms namely ontology based TF-IGM enriched semantic smoothing based clustering [13], chicken swarm optimization (CSO) based model and genetic-chicken swarm optimization (GCSO) based models for biomedical and clinical document clustering are utilized for this purpose. The clinical results analyzed in this stage along with suggestion for medications are forwarded to the e-health centers and/or physicians for approval before transmitting them to the patient care or medical staff attending the patients. The complete operation of this model is illustrated in this article. Section 2 provides brief discussion about the related works. Section 3 presents the methods discussed above while the proposed model is evaluated and the results are presented in section 4. Section 5 makes a conclusion about the work presented in this article.

II. METHODS

The proposed RRHT model of remote health care system consists of three main processes: patient data collection and processing, clustering and analyzing, and the decision approval process. Initially, the sensor nodes are placed at the patient body parts to collect the patient data which are combined into clinical records and transmitted from the gateway using Autonomous optimized routing algorithm of WSN. The clinical documents are collected and fed to the processor where the clustering is performed and decisions are made based on patient health data. The suggestions on medications are forwarded to the doctors at e-health centers for approval. This model of remote health care and telemedicine is provided in Fig. 1.

From the figure, it can be learnt that the decisions are not merely taken from the patients' health data received currently but also from the past data collected during previous medical treatments. The clustering of the documents is performed using the ontology based TF-IGM enriched semantic smoothing based clustering, CSO based model and GCSO based clustering models which were developed previously for efficient biomedical documents and clinical records clustering. Similarly, the routing of the clinical documents is performed efficiently by developing a novel Autonomous optimized routing algorithm based on the self-optimization concept.

A. Sensor nodes and data collection

The wireless medical sensor nodes are primarily utilized for monitoring and collecting various signals for the patients' body and then transmit the collected data to the destination node through the wireless radio frequency (RF). The wireless sensor nodes used in this purpose consists of Microcontroller unit (MCU), RF module, sensor module and the power supply module as described in [14].

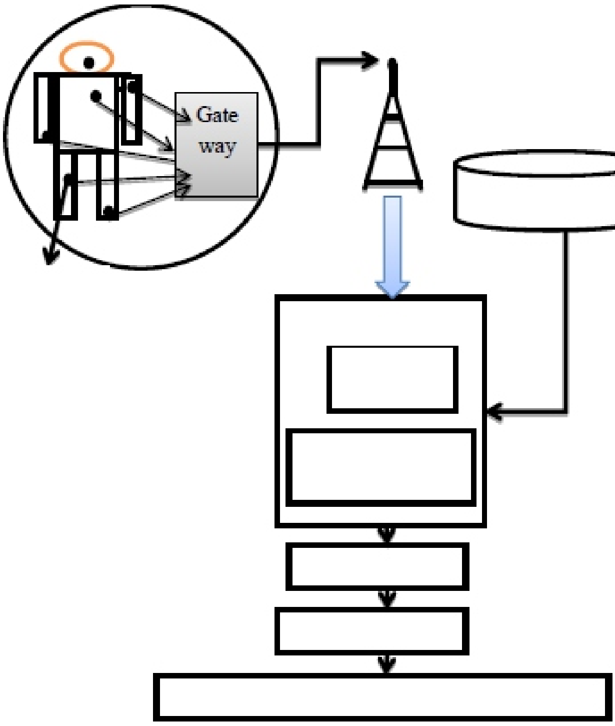


Fig.1. Ontology enabled clustering based remote health care and telemedicine model

With consideration of model cost and resource requirement, the 8bit 80C51MCU is employed as the MCU of the sensor nodes while nRF401 is used as the basic communication module. 80C51MCU has several analog-to-digital interfaces and I/O ports for easy programming and a watchdog timer for resetting the system when fault occurs. The reason for selecting nRF401 is because of its small size and low power consumption. Thus the utilized medical sensors have better performance than other sensor nodes.The sensor data are collected at the health care sink node which deals with the data from individual sensors and fuse them into clinical documents before for-warding to the central server or processing unit. The MCU of the sink node handles the data received from the patient medical sensors and also controls the functioning of the sink node. The RF module receives the data from the sensor nodes and is fused at the gateway by the sink node to form clinical documents. These documents are transmitted to the processor situated at distant location mostly in the distant health care centers in wireless medium using autonomous optimized routing algorithm.

B. Power consumption reduction scheme

The power consumption in a battery-operated sensor node can be reduced by setting the node to wake up once in specified duration to check the health parameter. This can satisfy the needs to consume as little power as possible in order to decrease the need for battery replacement. Many researchers have employed fuzzy approaches [15] which utilize a set of rules to find an effective solution that achieves the target while eluding complex and computationally costly solutions. In our proposed model of RRHT, network lifetime optimization can be accomplished with the reduction of packet transmission power and reduced data processing.

Energy consumption in WSN can be determined by three elements: sensing, processing and transmission of data. Energy consumption for sensing data can be computed by the specified characteristics of the node and determined by the device datasheet. Meanwhile, energy consumption for processing data for sensor node i and processing task k denoted as $E_{i,k}^{proc}$ can be computed which is directly proportional to the complexity of task k and to the average energy consumption per instruction in k of node I denoted as E_i^{ins}. It can be given as Eq. (1).

$$E_{i,k}^{proc} = I_k \times E_i^{ins} \qquad (1)$$

Where I_k is the number of instructions needed to execute task k.

Similarly, the energy consumption for transmission can be computed based on the transmission and reception power consumption. Transmission power consumption and reception power consumption are given by Eq. (2), (3).

$$P_{i,j}^{T} = P_i^{T_0} + P_i^{A}(\delta_{ij}) = P_i^{T_0} + \frac{P_i^{T_x}(\delta_{ij})}{d_{ij}} \qquad (2)$$

$$P_j^{R} = P_j^{R_0} \qquad (3)$$

Where $P_{i,j}^{T}$ and P_j^{R} are power consumption values for transmitting and receiving respectively; $P_i^{A}(\delta_{ij})$ denotes the consumed power of Power amplifier depending on distance δ_{ij} between nodes i to j; $P_i^{T_0}$ and $P_j^{R_0}$ denotes the power consumption components of transmitter and receiver circuitries; $P_i^{T_x}$ represents the transmitting output power depending on δ_{ij} while d_{ij} denotes the drain efficiency of power antenna at node i.

Considering the channel with path loss component, the transmitted power $P_i^{T_x}(\delta_{ij})$ undergoes some modifications.

$$P_i^{T_x}(\delta_{ij}) = P_j^{R_x} \times A_{ij} \times \delta_{ij}^{\alpha PL} = \varphi_{ij} \times \delta_{ij}^{\alpha PL} \qquad (4)$$

Where for reliable communication $\varphi_{ij} = P_j^{R_x} \times A_{ij}$ is the product of minimum reception power at node j with antenna parameter; αPL is the path loss component and A_{ij} is the antenna parameter defining gain and efficiency of antenna.

Substituting these modifications, the transmission energy per bit data at rate R from node i to j at transmitter ($E_{ij}^{T_x}$) and receiver ($E_j^{R_x}$) respectively, can be expressed as Eq. (5) & (6).

$$E_{ij}^{T_x} = \frac{P_{i,j}^{T}}{R} = \frac{1}{R}\left(P_i^{T_0} + \frac{\varphi_{ij} \times \delta_{ij}^{\alpha PL}}{d_{ij}}\right) \qquad (5)$$

$$E_j^{R_x} = \frac{P_j^{R}}{R} = \frac{P_j^{R_0}}{R} \qquad (6)$$

After computing the energy consumption values $E_{i,k}^{proc}$, $E_{ij}^{T_x}$ and $E_j^{R_x}$ for each data bit sensed, processed and transmitted, the fuzzy rules are applied to determine which command to be sent to the nodes based on input variables: throughput to workload and battery level. The main concept is to schedule the sleep of nodes that are on low energy or idle. Three membership functions (Low, Medium, High) are defined the input and output variables. The input variables ranges are determined with throughput to workload (0-100%) and battery level (0-1024 mWh) with smaller value to larger value (S to L). The main commands to the nodes are the sleep and wake-up. The sleep time is determined based on the above membership function and rules. In case of emergency, at the far end of network lifecycle when the number of dead nodes increases, the wake-up message is sent to wake a sleeping node to act as substitute until the dead node is recharged or replaced. This process helps in minimizing the overall power consumption and increases the network lifetime. The fuzzy inference approach requires minimal memory and energy but does not depend on batteries of the sensor nodes. These are powered through dynamic power source which generates power through the patient body heat and thus does not affect the WSN network model.

C. Ontology enabled clinical document clustering algorithms

This module is the vital cog in the healthcare system with implications on the overall output. The patient data are combined into documents or clinical records for transmission convenience which are transmitted with higher security using the autonomous self-optimized routing strategy in WSN. Thus received data documents are normally in the form of patient health readings like heart beat rate/pulse rate, temperature distribution, respiration rate, etc. which is standalone readings that does not provide clear indication of disease or sickness specifically. Hence these data are needed to be clustered in such a manner to utilize for detecting abnormal health condition. The concept of ontology can provide better priority based semantic similarity relationship in clustering these patient data. The ontology enabled clustering algorithms utilized in this work are TF-IGM enriched k-means & hierarchical model, CSO based clustering and GCSO based clustering models.

Ontology based semantic smoothing models have been prevalent in the field that utilizes TF-IDF for term weighting while n-gram techniques are employed for phrase detection in the documents [16]. However, this model has certain drawbacks. The n-gram technique does not perform well when the original document has been modified and also it only considers fixed length phrases and as document size grows, dimensionality increases tremendously. The TF-IDF factor is not fully effective in text document classification and affects the overall accuracy while the semantic smoothing model is effective in handling Sparsity of words but not effective in handling the density of general words. These limitations have created the opportunity to develop efficient strategies.

The first model is the ontology based TF-IGM enriched clustering model [13] which is a hybrid of concept based approach & improved semantic smoothing model. This model utilizes TF-IGM factor and improved background elimination which improves the density handling of general words while also TF-IGM incorporates a new statistical model to precisely measure the class distinguishing power of a term across different classes of text in documents. As n-gram has problems in identifying the phrases, the modified n-gram are used which detects the cases of substitution and deletion in the documents and averts them for better phrase identification. The developed clustering strategy is employed on k-means and hierarchical clustering models for biomedical documents clustering and has found to improve accuracy significantly. The second model developed based on the chicken swarm optimization algorithm also utilizes TF-IGM and modified n-grams technique. Additionally it utilizes Part-of-speech (POS) tagging and concept mapping for determining phrases and mapping of similar concept patterns in the documents. However, the dimension of these document features are considerably higher and hence a dynamic dimension reduction technique is utilized which prune the non-useful text features along with dimension reduction. Finally, the CSO based optimized clustering model clusters the documents into their suitable classes.

From this clustered results, the patient data can be analyzed more effectively. The procedures involved in CSO based biomedical document clustering algorithm are given as follows:

1. Pre-process the input documents
2. Apply semantic annotation, POS tagging and concept mapping
3. Perform modified n-grams duplication removal and TF-IGM
4. Perform dynamic dimension reduction
    a. Compute dynamic document frequency for each documents
    b. When dynamic document frequency ≤ threshold, delete the document features
    c. Form new subset of features
5. Perform CSO clustering
    a. Compute fitness for each document set
    b. Rank the documents based on fitness
    c. Establish hierarchical order and cluster them into groups
    d. Determine relationships between documents based on similarity
    e. Update each cluster of documents
6. Return clustered document sets.

The third model is based on the ontology enabled genetic-chicken swarm optimization (GCSO) based clustering in which the semantic annotation, concept mapping and modified n-grams are employed as in CSO based model. This model additionally utilizes phrase-based detection method for detecting and avoiding redundant copies documents. The term weighting is performed using Length Feature Weight (LFW) scheme as it can improve the weights based on the new factors at the level of the document. Adaptive and dynamic dimension reduction is employed in the form of Multi-level Bhattacharyya distance based Dimensionality Reduction (MBDR). MBDR discovers elliptical clusters for more effective dimensionality reduction by using only the low-dimensional subspaces. MBDR is highly scalable in terms of data size and dimension and it is also dynamic and adaptive to insertions after which the optimized clustering is performed. By theoretical comparison, it was found that this model of biomedical document cluster is highly accurate than the former two models. The major reason for this development is the use of LFW and MBDR. The steps involved in GCSO based biomedical document clustering algorithm are given as follows:

1. Pre-process the input documents
2. Apply semantic annotation, POS tagging and concept mapping
3. Modified n-gram and phrase based duplication removal
4. Term weighting using LFW
5. Perform MBDR
    a. Model documents as data stream
    b. Compute Bhattacharyya distance and error rate for each data stream
    c. Analyze and sort sub-space features based on minimum Bhattacharyya distance
    d. Set threshold for Bhattacharyya based projection distance and error rate
    e. Reduce dimension for high dimension features
    f. Form new subset of features

```
6. Perform GCSO clustering
   a. Compute fitness for each document set
   b. Rank the documents based on fitness
   c. Select 2N best fitness documents and apply two-parent crossover and 20-50% mutation
   d. Re-rank documents based on fitness
   e. Establish hierarchical order and cluster them into groups
   f. Determine relationships between documents based on similarity
   g. Update each cluster of documents
7. Return clustered document sets
```

D. Decision support module and final approval

Once the clinical documents are clustered into respective classes, it will pave way for analyzing the measured reading for specified sickness. At this stage, the patient's past data collected during the previous trial is utilized along with an ideal measurement threshold data. The present data are compared with these data to determine the change in measurements of heart beat, temperature, etc. If the data readings are greater or lesser than past data but within the ideal threshold then some precautionary medications are suggested. When the patient readings compared are beyond the ideal threshold levels then the medications are suggested for curing and also the possible hospital visits are prescribed. In worst cases, the emergency treatment are also initiated with an alert to all respective teams of health centers, ambulances and also the family members of the patient. These suggestions are subject to approval from the physicians as fully automatic health care medication systems are still not completely acceptable in the medical fields. In this research, the blood pressure is considered for evaluation and hence the data are clustered and the BP readings are compared with that of patients' past data to evaluate the changes. Table 1 shows the treatment process [17], [18] suggested by the proposed RRHT decision support system for the BP patients connected in the remote health care system.

Based on this Table-I, the decision module prescribes the treatment process and respective medications. The medications suggested for the stage-1 and stage-2 hypertension with recommended doses for selected antihypertensive agents for outpatient management of hypertension in children and adolescents [17]-[21]. These prescriptions are solely provided only based on 2004 report by American Academy of Pediatrics [22] and are verified the doctors' prescriptions of medicine and not based on the brands or companies. The medications suggested are provided only through extensive research with the help of medical professionals and hence there is high reliability in the decision support system. Thus the proposed RRHT model reduces the travel expenses and trouble faced by the unhealthy patients in reaching the hospitals without even compromising the quality of treatment.

III. PERFORMANCE EVALUATION

The proposed RRHT application is evaluated in a constrained simulation environment in MATLAB to verify the theoretical expertise. The simulation settings are utilized as in [13]. The proposed RRHT model utilized biomedical document clustering algorithms in the form of CSO based clustering and GCSO clustering. The two models are compared with existing clustering models namely MSVM [23] and ELM [24] based models. The comparisons are made in terms of accuracy, precision and recall. The experimental validation process is performed for remote patients living 10–100 km away from the hospital server in a test environment.

Table-I: Decision module Treatment process for BP patients

Systole pressure mmHg	Diastole pressure mmHg	Level	Recheck	Prescription
120	80	Normal BP	1year	Nil
120-129	<80	Elevated BP	6months	Lifestyle changes
130-139	80-89	Stage 1 hypertension	2months	Starting dose
>140	>90	Stage 2 hypertension	1-4 weeks	Maximum dose
>180	>120	Hypertensive Crisis	Emergency	Direct to hospital

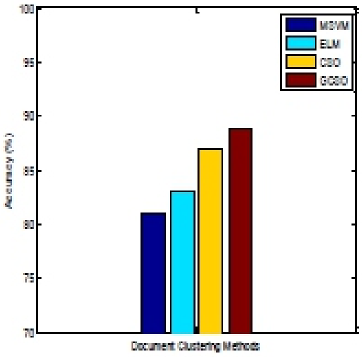


Fig.2. Clustering accuracy

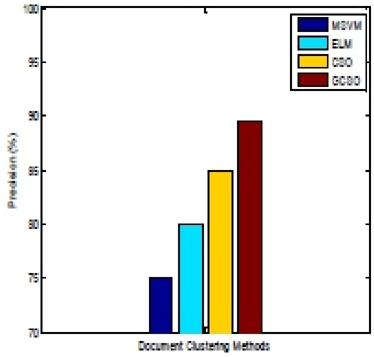


Fig.3. Clustering precision

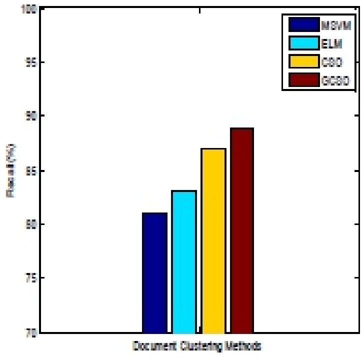


Fig.4. Clustering recall

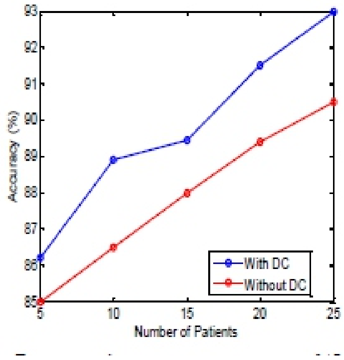


Fig.5. Accuracy with Document Clustering

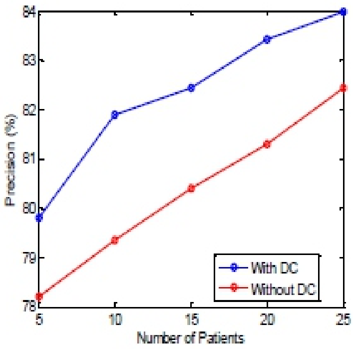


Fig.5. Accuracy with Document Clustering

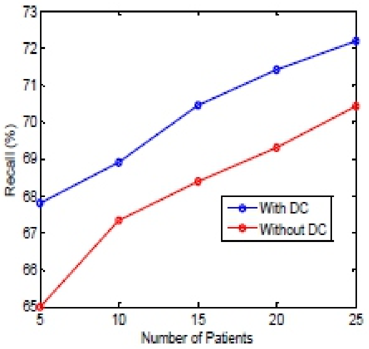


Fig.7. Recall with document Clustering

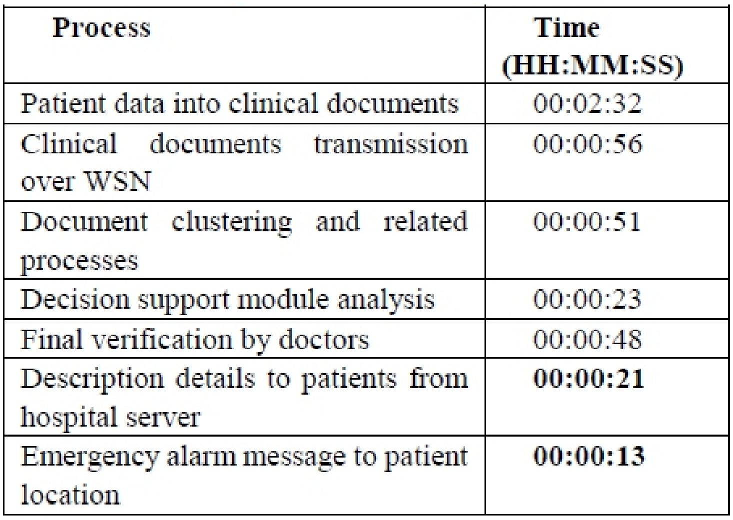
Fig. 2 shows the clustering accuracy while Fig. 3 and Fig. 4 show the precision and recall of the clustering algorithms, respectively. It can be seen that the GCSO model has better performance with higher values of accuracy, precision and recall. It has around 2%, 4% and 2% higher values of accuracy, precision and recall, respectively than the CSO based model while the MSVM and ELM based clustering models have much lesser performance values.

Currently, there are many versions of remote health monitoring and telecare applications and the major difference in the proposed RRHT model is the utilization of document clustering concept. The influence of document clustering concept can be illustrated by comparing the RRHT model with and without using the document clustering algorithm in terms of accuracy, precision and recall.

Fig. 5, Fig. 6 and Fig. 7 respectively shows the accuracy, precision and recall of the RRHT with and without document clustering model. It can be seen that the use of document clustering improves the accuracy, precision and recall of the proposed RRHT application system. RRHT with document clustering achieves 3%, 2% and 2% greater accuracy, precision and recall than that without document clustering. In addition to the validation parameters, the RRHT has been evaluated in terms of time consumed for sending the collected patient data to the hospital and the time for sending and receiving alarm messages over WSN. Table 2 shows the average time taken for different processes in RRHT. Time for each process in the table is the average of 20 iterations.

The evaluation results prove that the RRHT can significantly improve remote healthcare application with the use of document clustering while the GCSO based document clustering has better performance among the clustering algorithms. From the experiments, it has also been found that one physician can observe 15–20 patients if exclusively appointed on proposed system; however, this number reduces to 5–12 patients if a physician is requested to monitor patients through recommended scheme along with regular patient treatment in hospital. Moreover, it is important to state that the number of patients depends on the condition of the patients in hospital as well as patients monitored by the recommended RRHT scheme.

Table-II: Average time for RRHT processes over WSN



Process	**Time (HH:MM:SS)**
Patient data into clinical documents	00:02:32
Clinical documents transmission over WSN	00:00:56
Document clustering and related processes	00:00:51
Decision support module analysis	00:00:23
Final verification by doctors	00:00:48
Description details to patients from hospital server	**00:00:21**
Emergency alarm message to patient location	**00:00:13**

IV. CONCLUSION

This paper proposed the development of RRHT for the telecare and telemedicine application for remote patients. This model employed the wearable sensors to collect data while WSN for transmission through autonomous optimized routing algorithm. The model employed ontology enabled document clustering algorithms for improving the detection of patient health data through efficient clustering of clinical records and subsequent comparison with past data. The decision module detects the consistencies in the patient health data readings and suggested the suitable treatment process and/or medications. This model has been evaluated to determine its efficiency and the valid performance results proved that this model is highly efficient than most existing models in the lines of remote health monitoring and telemedicine. In future, this model is intended to include more disease detecting capabilities through high intelligent systems. It is also planned to build this model by tackling practical difficulties and increasing the accuracy of diagnosis.

REFERENCES

1. P. Y. Chau, and P. J. H. Hu, "Investigating healthcare professionals' decisions to accept telemedicine technology: an empirical test of competing theories," Information & management, vol. 39, no. 4, pp. 297-311, 2002.
2. S. Koch, "Home telehealth—current state and future trends," International journal of medical informatics, vol. 75, no. 8, pp. 565-576, 2006.
3. K. Maharatna, E. B. Mazomenos, J. Morgan, and S. Bonfiglio, "Towards the development of next-generation remote healthcare system: Some practical considerations," In 2012 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, pp. 1-4, 2012.
4. S. Youm, G. Lee, S. Park, and W. Zhu, "Development of remote healthcare system for measuring and promoting healthy lifestyle," Expert Systems with Applications, vol. 38, no. 3, pp. 2828-2834, 2011.
5. F. Touati, and R. Tabish, "U-healthcare system: State-of-the-art review and challenges," Journal of medical systems, vol. 37, no. 3, p. 9949, 2013.
6. L. Gavrilovska, S. Krco, V. Milutinović, I. Stojmenovic, and R. Trobec, (Eds), "Application and multidisciplinary aspects of wireless sensor networks: concepts, integration, and case studies," Springer Science & Business Media, 2010.
7. J. A. Stankovic, Q. Cao, T. Doan, L. Fang, Z. He, R. Kiran, and A. Wood, "Wireless sensor networks for in-home healthcare: Potential and challenges," In High confidence medical device software and systems (HCMDSS) workshop, vol. 2005, 2005.
8. R. Dobrescu, D. Popescu, M. Dobrescu, and M. Nicolae, "Integration of WSN-based platform in a homecare monitoring system," Latest Trends on Communications and Information Technology, pp. 165-170, 2008.
9. A. Milenković, C. Otto, and E. Jovanov, "Wireless sensor networks for personal health monitoring: Issues and an implementation," Computer communications, vol. 29, no. 13-14, pp. 2521-2533, 2006.
10. H. D. E. Al-Ariki, and M. S. Swamy, "A survey and analysis of multipath routing protocols in wireless multimedia sensor networks," Wireless Networks, vol. 23, no. 6, pp. 1823-1835, 2017.
11. S. S. Lin, M. H. Hung, C. L. Tsai, and L. P. Chou, "Development of an ease-of-use remote healthcare system architecture using RFID and networking technologies," Journal of medical systems, vol. 36, no. 6, pp. 3605-3619, 2012.
12. K. Saleem, N. Fisal, S. Hafizah, S. Kamilah, and R. A. Rashid, "A self-optimized multipath routing protocol for wireless sensor networks," International Journal of Recent Trends in Engineering, vol. 2, no. 1, pp. 93-98, 2009.
13. R. Sandhiya, and M. Sundarambal, "Clustering of biomedical documents using ontology-based TF-IGM enriched semantic smoothing model for telemedicine applications," Cluster Computing, vol. 22, no. 2, pp. 3213-3230, 2019.
14. H. Huo, Y. Xu, H. Zhang, Y. H. Chuang, and T. C. Wu, "Wireless-Sensor-Networks-based healthcare system: a survey on the view of communication paradigms," International Journal of Ad Hoc and Ubiquitous Computing, vol. 8, no. 3, pp. 135-154, 2011.
15. M. Collotta, G. Pau, and G. Scata, "A fuzzy system to reduce power consumption in Wireless Sensor Networks: A comparison between WirelessHART and IEEE 802.15. 4," In 2014 IEEE International Energy Conference (ENERGYCON), IEEE, pp. 766-771, 2014.
16. S. Logeswari, and K. Premalatha, "Ontology-based semantic smoothing model for biomedical document clustering," International Journal of Telemedicine and Clinical Practices, vol. 1, no. 1, pp. 94-110, 2015.
17. B. Falkner, S. R. Daniels, J. T. Flynn, S. Gidding, L. A. Green, J. R. Ingelfinger, and A. P. Rocchini, "The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents," Pediatrics, vol. 114, no. 2 III, pp. 555-576, 2014.
18. M. B. Lande, and J. T. Flynn, "Treatment of hypertension in children and adolescents," Pediatric Nephrology, vol. 24, no. 10, pp. 1939-1949, 2009.
19. M. Riley, and B. Bluhm, "High blood pressure in children and adolescents," American family physician, vol. 85, no. 7, 2012.
20. E. Lurbe, R. Cifkova, J. K. Cruickshank, M. J. Dillon, I. Ferreira, C. Invitti, and W. Rascher, "Management of high blood pressure in children and adolescents: recommendations of the European Society of Hypertension," Journal of hypertension, vol. 27, no. 9, pp. 1719-1742, 2009.
21. J. T. Flynn, and B. E. Falkner, "New clinical practice guideline for the management of high blood pressure in children and adolescents," Hypertension, vol. 70, no. 4, pp. 683-686, 2017.
22. American Academy of Pediatrics, "National high blood pressure education program working group on high blood pressure in children and adolescents," Pediatrics, vol. 114(Supplement 2), pp. iv-iv, 2004.
23. A. Temko, and C. Nadeu, "Classification of acoustic events using SVM-based clustering schemes," Pattern Recognition, vol. 39, no. 4, 682-694, 2006.

24. *C. K. L. Lekamalage, T. Liu, Y. Yang, Z. Lin, and G. B. Huang, "Extreme learning machine for clustering," In Proceedings of ELM-2014, Springer, Cham, vol. 1, pp. 435-444, 2015.*

AUTHORS PROFILE

Ms. R. Sandhiya has obtained her Bachelors in Information Technology at the CSI College of Engineering from Anna University in 2011. She obtained her Master's in Information Technology at Anna University Regional Centre, Coimbatore from Anna University in 2014. She is currently working as Professor in the Department of Information Technology at Coimbatore Institute of Technology, Coimbatore, India.

Dr .M. Sundarambal has obtained her Bachelors in Electrical and Electronics Engineering at the Government College of Technology from Madras University in 1981. She obtained her Master's in Applied Electronics at PSG College of Technology from Bharathiar University in 1984. Dr.M.Sundarambal is currently working as Professor in the Department of Electrical and Electronics Engineering at Coimbatore Institute of Technology, Coimbatore, India. Her specializations include MANET, Wireless Networks, Wireless Sensor Networks, Robotics; Agent based Intelligent System and High Tech Prosthetics. She has produced many Ph.D. Scholars and has more than 100 publications in international journals and national conferences. She is a member of professional bodies, Indian Society of Technical Education, System Society of India, and Institution of Engineers.

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).
2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);
3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);
4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

Address of the Editorial Office:

Enriched Publications Pvt. Ltd.
S-9,IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR -5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-45525005