Semantic Driven Healthcare Monitoring and Disease Detection Framework from Heterogeneous Sensor Data

Diksha Hooda
Computers Science & Engineering Department,
Thapar Institute of Engineering and Technology,
Patiala, India
dhooda_phd17@thapar.edu

Rinkle Rani
Computers Science & Engineering Department,
Thapar Institute of Engineering and Technology,
Patiala, India
raggarwal@thapar.edu

Abstract—**As the age profile of various societies keeps going up all over the world, in addition to the continuous surge in the populace suffering from long term illnesses, including diabetes, hypertension, cardiovascular problems, obesity, etc. supporting and providing healthcare services, both physically and mentally, is of utmost importance in order to assure and maintain independent living. So, regular and uninterrupted health monitoring is evidently an assured and functional solution that can methodically deliver health-related services to elderly people and patients. The emergence of the sensor based technologies has given rise to immense possibilities in the domain of healthcare. Also, it has been adopted precisely in the construction and deployment of healthcare systems for effective regular health monitoring, illness diagnosis and further treatment. This paper proposes a semantic driven healthcare framework in which an ontology is developed that conceptualises the process of disease detection on the basis of dynamic sensed clinical data. The proposed system provides semantic interoperability by handling the issue of heterogeneity and promoting exchange of consensual knowledge. Our work designs reasoning rules over the semantic knowledge base for classification and disease detection task. Further, the model is validated by deploying semantic querying to retrieve desired results. The results portray the potential and efficiency of the proposed ontology based system to grow into a more sophisticated approach for healthcare monitoring and treatment.**

Keywords—Healthcare, Ontology, Knowledge engineering, Sensor Fusion, Disease detection,

I. INTRODUCTION

Healthcare is the most crucial and consumed service by every age group in the world. It becomes extremely significant for the elderly people. The existing healthcare systems need enhancement and improvement to facilitate continuous monitoring and treatment of illnesses and health issues, trigger alert whenever an urgent health related situation arises and disease prevention[1]. The healthcare industry captures huge amounts of health related data that needs to be mined to extract hidden knowledge for efficient decision making. The world-wide surge in the mortality due to multiple diseases like heart problems and long term exposure to diabetes etc and the availability of large amount of patients' data to extract meaningful information has motivated researchers to apply data mining techniques to assist health care professionals in the diagnosis and treatment [2]. Data mining is the task of analyzing large datasets to retrieve hidden and previously unknown patterns, relationships and information that is tough to capture with conventional statistical techniques [3]. Data mining applications in the field of healthcare can contribute to better health policy-making, no scope of hospital errors, timely detection and early warning, prevention and effective treatment of diseases and hence, preventable patient deaths.

In the past few decades, advanced information technologies like IoT and sensing technologies [4], have been deployed widely in the field of healthcare [5]. So, in order to achieve the objective to offer enhanced health services and hence contribute to enhance life quality of the elderly patients, there is a need to design and deploy healthcare systems using IoT based sensing technologies. These technologies allow interworking of various heterogeneous devices and applications to result in smooth, and seamless services [6].

IoT devices comprise of multiple sensors embedded and deployed to the subject under consideration to collect ambient and dynamic medical data which ensures incessant monitoring of health condition. Henceforth, there arises a need of some standardized format and structures to support interoperability and exchange of sensory database. Semantic based ontological models are appropriate for conceptual modelling and representation. Such models represent the real world entities as concepts and relationships are defined over the set of real world concepts. Ontologies deliver a formal language that represents consensual domain knowledge in a formalized format which facilitates the knowledge exchange among doctors/clinicians/caregivers in an sophisticated and effective manner. A thorough semantic based model in sensory data that provides regular health monitoring and disease detection should comprise of vocabulary and terms correlated to the healthcare sector, sensory equipments, patient profiling in order to generate useful knowledge in the context of all constituents forming the interconnected IoT based system.

In the proposed work, we present a sensing technology driven Healthcare system that works on the basis of a newly developed ontological model. The proposed work presents an ontology-based system that formally represents the captured heterogeneous sensor data and performs rule-based reasoning on the modelled and semantically annotated information to identify health conditions and generate desired results for the doctors, patients and caregivers. Additionally, the proposed work integrates already existing ontologies in the domain of

IoT and healthcare in order to promote meaningful information representation and its sharing.

II. BACKGROUND AND RELATED WORKS

In recent years, multiple studies have been conducted to investigate the usability of ontologies in disease diagnosis and decision support systems. Natural language applications offer limited support and lack potential due to absence of a formal structure and the non-collective implications and interpretations. On the contrary, ontologies [7] can capture the domain knowledge in a meaningful way and specify modelling primitives effectively. Ontologies are defined as a formalised way of knowledge representation through the use of controlled vocabularies designed to facilitate knowledge sharing, computation and reasoning mechanism [8]. In recent years, there has been extensive ongoing research in applicability of ontology along with sensing technology in the domain of healthcare, certain applications have already commercialised and been made available for deployment in the market. Ontologies have gained tremendous relevance in the domain of biomedical, since, they allow researchers to stay abreast of crucial biomedical information and promote the exchange and understanding of such knowledge.

A. Ontologies in the domain of Healthcare

Existing research focuses on assisting patients experiencing problems in independent living, like, elderly people having certain chronic illnesses [9], for example heart illnesses, diabetes, hypertension and Alzheimer's. Varatharajan et al. [10] presented a real time warping driven algorithm for on time detection of Alzheimer while using sensors that can be worn and attached to the body, while Romero et al. [11] proposed a framework to diagnose and monitor symptoms of Parkinson's disease. In [12], the authors developed Hypoglycaemia Prediction Algorithm, abbreviated as HPA, that identifies the crucial stage of hypoglycaemia, which can be fatal for a person. In another work, the authors proposed a universal model for prediction of glucose concentration in the blood [13]. In [14], authors presented a Classification Tree Technique for Heart Disease Prediction. The work analyzed the classification tree method in the field of data mining. In [15], authors used AI and machine learning based techniques to carry out data-driven decision making to prescribe antibiotics with respect to the disease. The authors proposed an ontology based methodology for continuous patient centered chronic disease monitoring and management[16].

In recent years, various ontologies encoding medical concepts in the form of vocabulary have been proposed. Such ontologies represent complete and exhaustive descriptions of various medical concepts and core domain knowledge. They provide reusable vocabulary resources for clinical frameworks and systems and management of structured knowledge and cooperative work among healthcare networks. For instance, Systematized Nomenclature of Medicine - Clinical Terms (abbreviated as SNOMED CT) [17] and International Classification of Diseases (abbreviated as ICD) [18]. SNOMED CT formalized ontology offers a elaborate representation and illustration of terms existing in the medical world. Various decision support systems have applied SNOMED CT as their core component [19-20]. The ICD ontology represents the standard generalised diagnostic procedure for almost all the health issues and diseases. It elucidates diseases and their symptoms experienced in corresponding to the specific health issue.

B. Ontologies in Sensor Technology Domain

The past few years have seen implementation of the semantic web technology for the representation and portrayal of the domain of IoT. Various ontologies are presented that effectively represent sensory devices, their attributes and features and the respective parameters. For instance, the Semantic Sensor Network ontology, mostly referred to as SSN ontology [21], Sensor Observation Sample and Actuator ontology, mostly referred to as SOSA ontology [22], and The IoT-A ontology [23] are some of the ontologies that are considered to be well recognised standards for representing sensing processes.

C. Ontologies for Sensor-based Healthcare applications

A surge in the number of research that has been conducted in the implementation and deployment of semantic web technology and IoT in the domain of bio medical is observed.

In this work [24], authors proposed an ambient intelligent system that provides real-time and continuous monitoring of the concerned patient suffering from congestive heart failure condition. The system facilitates the real time observation of the health status and condition of the patient under consideration, triggering alerts to caretakers and hence, leading to detection of the risk stage by implementing ontology-driven knowledge model and SWRL based inference. In this work [25], authors presented a healthcare monitoring framework based on sensor devices. This system is developed on the basis of a set of rules to trigger and issue up alarms in case the health of the patient in terms of BP, heartbeats and rate and lipid profile etc. is out of the standard consensual range.

Hence, there have been various researches that deploy ontology based model in IoT driven frameworks for the domain of healthcare. However, most of the existing frameworks based on ontological knowledge modeling concentrate on processing and analyzing of the captured measurement. Additionally, the aspect of interoperability is not facilitated between existing systems and captured observation.

Additionally, existing researches focus on on ailment for detection and further treatment options. So this gives the need to develop and implement a general application that fits all to address multiple issues and ailments simultaneously.

III. PROPOSED FRAMEWORK FOR HEALTHCARE MONITORING AND DISEASE DETECTION

The proposed model is an ontology-based model for healthcare monitoring and disease detection. Currently, the ontology supports the detection of heart problem and diabetes. The proposed system constitutes of the following modules: 1. Data Acquisition module 2. Patient Profiling Module 3. Knowledge Discovery module 4. Information Retreival module. Fig. 1 illustrates the architecture of the proposed healthcare monitoring system, whose constituent modules are detailed below:

A. *Data Acquisition*

This module involves the task of sensory data collection from various bodily/wearable sensors that are deployed in order to capture symptoms/readings. The data acquired from various sensors capable of measuring various physiological parameters, contains noisy and missing values. The following preprocessing tasks are performed on the captured data to make it fit for further storage and analysis:

- **Missing values**: There are various records in the sensor data that contain empty cells or have 'nan' value in the cell. These need to be omitted for efficient analysis. Such missing values in the records are replaced by the value captured by the sensor in the preceeding interval. In case, there are a significant number of records with continuous missing values then these records are simple deleted from the sensor data file.
- **Redundant values**: The records that have redundant values are deleted to avoid false emergency conditions related to the health of the patient.

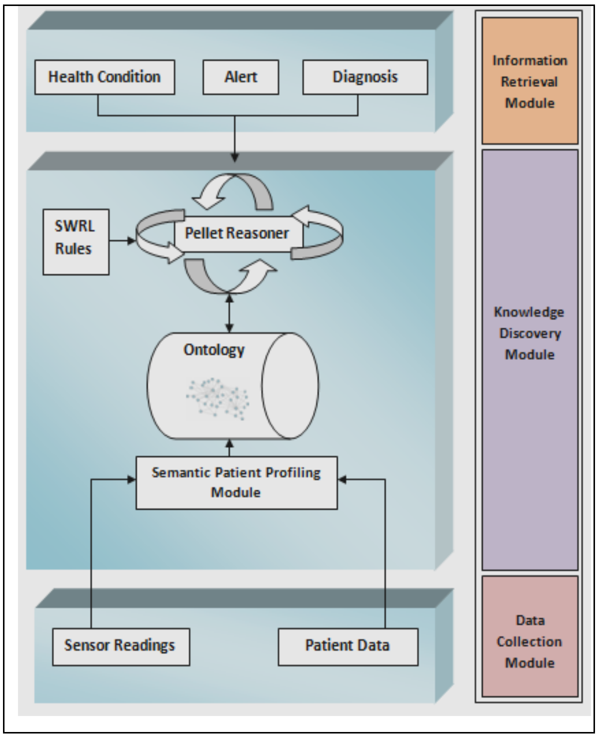

Fig. 1. Architecture of the proposed system

B. *Patient Profiling Semantic Module*

This mode of representation captures the patient profile information similar to the information available as a part of the Patient's Electronic Medical Record. Patient data is usually complex and diverse, exists in heterogeneous formats which makes the task of processing, knowledge mining and analysis of clinical data a tough task, hence, adversely affecting the decision making process. The proposed system deploys a semantic patient profiling module that adds to the patient data an intrinsic semantic meaning. The patient semantic information is stored in the newly developed ontology. So, the patient related features are semantically annotated by classes which model entities like patient name, patient ID, patient history,sex etc as,"hmonitor:pname", "hmonitor:pID", "hmonitor:history","hmonitor:sex",respectively. The well defined classes have properties between them that model the real world relationships in the patient profiling.

C. *Knowledge Discovery Module*

The proposed system is a knowledge driven model that captures consensual knowledge about the health monitoring aspects and disease diagnosis parameters and then adds a formalized structure to it so as to carry out the task of detection of diseases. This module performs two main tasks:

- **Conceptualization:** The aim behind execution of this task is to organise and represent the dynamic sensor data and the acquired domain knowledge in the format of a semi-formalized model that can be understood and interpreted by caretakers, clinicians, domain experts. The acquired knowledge in the form of parameters essential for the scenario of disease detection is represented classes and relationships between them. This constitutes the knowledge base for the proposed system. It consists of a well defined vocabulary over which a set of IF-THEN rules are designed to achieve knowledge reasoning of the developed ontology.
- **Disease Detection:** The proposed model currently supports the detection of cardiac problem and diabetes in patients from the dynamic sensory data captured by sensing devices. The main idea behind the disease classification is for the early prediction of the disease provided a patient's vital clinical data from the sensors is given. Early disease prediction can lead to the treatment of the patients before the condition worsens.
  The detection is performed by firing the inference engine on the semantically annotated data. The well defined rules are taken as input by the reasoner to classify the physiological reading of a particular patient at a specific instant of time into a classifier representing a disease or a normal scenario. Table 1 shows the rules for disease detection task. This table shows rules that have been created in OWL language with the help of built in as well as user defined atoms. After the atoms are created, rules are created and added to the ontology to perform reasoning tasks for disease detection. The proposed model is a multidisciplinary system to promote interoperability and reusage among other existing systems as well.

D. *Information Retrieval Module*

In this module, SPARQL endpoint is used to interface with the data store to allow access to the underlying data. SPARQL is a query language used to query the knowledge base of RDF triples and retrieve results.

IV. IMPLEMENTATION AND EVALUATION

The proposed framework facilitates regular monitoring chronic-care service for elderly to minimize the dependability upon caregivers, and hence diminishing the

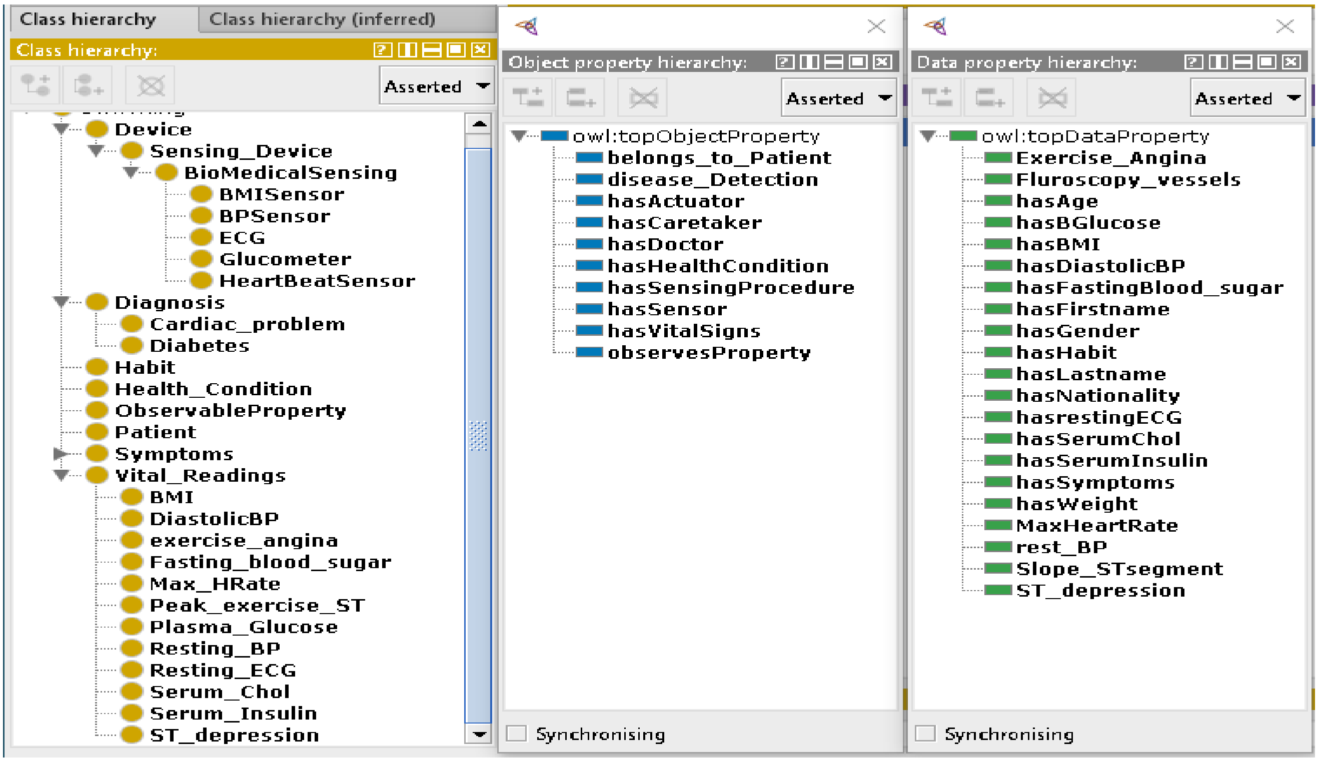

Fig. 2. Taxonomy of the proposed ontology

investment in terms of time, money etc and other resources allocated for elderly health care. The proposed framework is segregated into a four module structure based on the functionalities it supports: Data acquisition module allows the capturing and preprocessing of the sensed observations so as to further add a semantic meaning for data mining tasks to be performed. The semantic patient profiling module is storing patient clinical information in a meaningful manner. The knowledge discovery module is the module which performs interpretation of the captured sensory data and ensures real-time monitoring of the patient by semantically storing the observations and the alert management functionality which is responsible for alerting the caregivers/doctors in case any urgent or emergency situation occurs. In our work, the data acquisition module captures various sensing data such as BP, BMI, Sugar in Blood, Heart Beat, ECG etc, as well as the semantic profile of the subject or patient along with the clinical data, and the knowledge discovery module comprises of conceptualised knowledge in the form of ontological database and is developed implementing the Apache JENA framework. The Protégé tool is used for editing the ontology developed and SWRL language is used to construct SWRL rules that are fed to the reasoner to classify a patient's health status based upon health parameters. It also performs the task of disease detection. The task of inference and reasoning is done by the Pellet inference engine that acts upon the semantically annotated observational data. The SPARQL based query engine implements semantic query language that retrieves data stored in the ontological format. The reasoner used is built upon the Apache JENA API. The results generated as a result of the reasoning process is then stored in the ontological database to be further retrieved by the information retrieval module. The information retrieval module is intended for doctors/clinicians/caregivers and even patients to retrieve classified or regular monitoring data. The proposed system allows doctors/patient to monitor and retrieve their health condition by querying the sensor related data and use it for feedback.

A. *Dataset*

PIMA Dataset from the Indians Diabetes Database captured at the National Institute of Diabetes and Digestive and Kidney Diseases [26] and Cleveland Heart Disease dataset [27] was utilized in this study for the validation of the proposed framework. In case of the PIMA dataset, only five attributes are selected and considered for the classification model and validation. The selection of the attributes is on the basis of similarity of the vital signs data with the sensor devices: 2 hour glucose tolerance value, diastolic blood pressure reading, body mass index, and age. The diastolic blood pressure reading could have been generated from a blood pressure sensitive sensor device, while the 2 hour glucose tolerance value was similar to that captured from the blood glucometer based sensing device.

In the Cleveland heart disease dataset, only 12 sensor based reading attributes are selected for validation and analysis.

V. CONCLUSION

The proposed system supports smoother and a much sophisticated approach for diagnosis and health monitoring between the patient and the caretaker/caretaker by implementing semantic technology framework based on OWL and RDF languages.

Conclusively, the proposed work demonstrates how to operationalize semantic technology for medical diagnosis. The classification and detection of diseases are performed based on the assumptions that all vital sensor based readings

TABLE I. SWRL RULES FOR INFERENCE

Subject	Rule in natural language	SWRL rule
Patient	If the diastolic BP value is greater than 90 (mmHg) then the patient condition is High blood pressure and diagnosis is Hypertension.	Patient(?t) ^ BPSensor(?bpx) ^ Diagnosis(?dg) ^ observedBySensor(?t, ?bpx) ^ DiastolicBP (?bp) ^ observesProperty(?bpx, ?bp) ^ swrlb:greaterThan(?dg, 86) ^ swrlb:lessThanOrEqual(?dg, 90) -> hasHealthCondition(?t, "HighBloodPressure") ^ hasDiagnosis(?t, "Hypertension")
Patient	If the blood sugar value lies in the range 140 and 180 (mg/dL) then diagnosis is Diabetes.	Patient(?t) ^ Glucometer(?gx) ^ observedBySensor(?t, ?gx) ^ Plasma_Glucose(?pg) ^ observesProperty(?gx,?pg) ^ hasBGlucose (?pg, ?gx) ^ swrlb:greaterThan(?pg, 140) ^ swrlb:lessThanOrEqual(?pg,180) -> hasHealthCondition(?t, "HighSugar") ^ hasDiagnosis(?t, "Diabeties")
Patient	If blood sugar value sensed lies in the range 100 to 125, and patient has BMI greater than 27 and age between 40 and 65 , heart beat range 140-160 then the diagnosis is prediabetic	Patient(?t) ^ hasAge(?a) ^ swrlb:greaterThan(?a, 40) ^ swrlb:lessThanOrEqual(?a,65) ^ hasBMI(?bm) ^ swrlb:greaterThan(?bm, 27) ^ hasFastingBlood_Sugar(?bs) ^ swrlb:greaterThan(?bs, 100) ^ hasHeartRate(?h)^ swrlb:lessThanOrEqual(?bs,125) ^ swrlb:greaterThan(?h, 140) ^ swrlb:lessThanOrEqual(?h,160) -> hasHealthCondition(?t, "HighSugar") ^ hasDiagnosis(?t, "PreDiabetic")
Patient	If Diastolic BP value lower than 60mmHg then the patient has Low blood pressure and diagnosis is Hypotension.	Patient(?t) ^ hasAge(?a) ^ swrlb:greaterThan(?a, 40) ^ Diagnosis(?dg) ^ swrlb:lessThanOrEqual(?a, 65)^ observedBySensor(?t, ?x) ^ DiastolicBP (?bp) ^ observesProperty(?x, ?bp) ^ swrlb:lessThanOrEqual(?d, 60) -> hasDiagnosis(?t, "Hypotension") ^ hasHealthCondition(?t, "LowBloodPressure")

have equal weightage in rendering a diagnosis. This is unlikely to be reflective of scenario otherwise.

Having access to patient related health and clinical information is of utmost importance in clinical research. Extracting this information and making it available in a searchable form would be of immense help in clinical research. The proposed work presented a framework to collect patient profile data along with other health-related information and then further process it through rule-driven reasoning in which SWRL language is used for rule development to classify patient's health status and identify adverse situations. The main advantages of the semantic driven approach are as follows: 1) cost-effective maintenance of the Ontology knowledge base due to a clear demarcation among the knowledge base which makes it easier to add new functionality by simply adding new concepts and relationships and SWRL rules. 2) The ontology supports decision making operations that are not implemented in the existing systems and databases technologies. 3) The system also facilitates reuse and exchange of the conceptualized knowledge base.

For future scope, we are aiming to expand the developed system to support various other chronic illnesses and complete other modules that constitute the framework.

REFERENCES

[1] D. G. Korzun, A. V. Borodin, I. V. Paramonov, A. M. Vasilyev,and S. I. Balandin, "Smart spaces enabled mobile healthcare services in internet of things environments," International Journal of Embedded and Real-Time Communication Systems (IJERTCS), vol. 6, no. 1, pp. 1–27, 2015.

[2] S. Palaniappan, and R. Awang," Intelligent heart disease prediction system using data mining techniques", In 2008 IEEE/ACS international conference on computer systems and applications, IEEE, pp. 108-115, March 2008.

[3] S. H. Liao, C. Pei-Hui, and H. Pei-Yuan ,"Data mining techniques and applications–A decade review from 2000 to 2011." Expert systems with applications, vol. 39, no. 12, pp. 11303-11311, 2015.

[4] A. M. Rahmani, T. N. Gia, B. Negash et al.,Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: A fog computing approach," Future Generation Computer Systems,vol. 78, pp. 641–658, 2018.

[5] P.-L. Wu, W. Kang, A. Al-Nayeem, L. Sha, R. B. Berlin Jr., and J. M. Goldman, "A low complexity coordination architecture for networked supervisory medical systems," in Proceedings of the 4th ACM/IEEE International Conference on Cyber-Physical Systems, ICCPS 2013, pp. 89–98, usa, April 2013.

[6] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer networks, vol. 54, no. 15, pp. 2787–2805,2010.

[7] A. Maedche,"Ontology Learning for the Semantic Web", Springer Science & Business Media, New York, 2012.

[8] P. Robinson , S. Bauer, " Introduction to Bio-Ontologies",CRC Press, Boca Raton , 2011.

[9] N.G. Hallfors, M. Alhawari, M. Abi Jaoude, Y. Kifle, H. Saleh, K. Liao, M. Ismail, A .F. Isakovic, "Graphene oxide: Nylon ECG sensors for wearable IoT healthcare—nanomaterial and SOC interface", Analog Integr. Circ. Sig. Process,vol. 1, pp. 1–8 ,2018.

[10] R. Varatharajan, G. Manogaran, M.K. Priyan, R. Sundarasekar, "Wearable sensor devices for early detection of alzheimer disease using dynamic time warping algorithm", Clust. Comput, vol. 6, pp. 1–10, 2017.

[11] L.E. Romero, P. Chatterjee, R.L. Armentano, "An IoT approach for integration of computational intelligence and wearable sensors for Parkinson's disease diagnosis and monitoring", Health Technol. Vol. 6, no. 3, pp. 167–172, 2016.

[12] E. Dassau , F. Cameron, B W Bequette, H. Zisser, L Jovanovic, H. P. Chase, D M Wilson, B A Buckingham and F J Doyle, "Real-time hypoglycemia prediction suite using continuous glucose monitoring", Diabetes Care, vol. 33, no. 6, pp. 1249–1254, 2010.

[13] A. Gani, A V Gribok, W K Ward, R A Vigersky and J Reifman,"Universal glucose models for predicting subcutaneous glucose concentration in humans",IEEE Trans. Inf. Technol. Biomed, vol. 14,no. 1, pp. 157–165, 2010.

[14] S. Palaniappan and R. Awang, "Intelligent heart disease prediction systemusing data mining techniques," pp. 108–115, 2008.

[15] D G Prieto, R I Castilla, E González, and M L Couce, "Automated generation of decision-tree models for the economic assessment of interventions for rare diseases using the RaDiOS ontology", Journal of Biomedical Informatics, pp.103563, 2020..

[16] Y F Zhang, L. Gou, T S Zhou, D N Lin, J L Zheng, and J S Li," An ontology-based approach to patient follow-up assessment for continuous and personalized chronic disease management" Journal of biomedical informatics, vol. 72, pp.45-59, 2017.

[17] Https://bioportal.bioontology.org/ontologies/SNOMEDCT.

[18] "International classification of diseases, version 10,"https://bioportal.bioontology.org/ontologies/ICD10

[19] M. Hussain, A. Khattak, W. Khan, I. Fatima, M. Amin, Z. Pervez,R. Batool, M. Saleem, M. Afzal, M. Faheem et al., "Cloudbased smart cdss for chronic diseases," Health and Technology, vol. 3, no. 2, pp. 153–175, 2013.

[20] M. Marcos, J. A. Maldonado, B. Martínez-Salvador, D. Boscá, and M. Robles, “Interoperability of clinical decision-support systems and electronic health records using archetypes: a case study in clinical trial eligibility,” Journal of biomedical informatics,vol. 46, no. 4, pp. 676–689, 2013.

[21] “Semantic sensor network (SSN) ontology,”https://www.w3.org/2005/Incubator/ssn/ssnx/ssn.

[22] “Sensor, observation, sample, and actuator (sosa) ontology,” https://www.w3.org/ns/sosa/.

[23] W. Wang, S. De, R. Toenjes, E. Reetz, and K. Moessner, “A comprehensive ontology for knowledge representation in the internet of things,” in Trust, Security and Privacy in Computing and Communications (TrustCom), IEEE 11th International Conference on. IEEE, pp. 1793–1798, 2012.

[24] A. Hristoskova, V. Sakkalis, G. Zacharioudakis, M. Tsiknakis,and F. De Turck, “Ontology-driven monitoring of patient’s vital signs enabling personalized medical detection and alert,” Sensors, vol. 14, no. 1, pp. 1598–1628, 2014.

[25] G. Zhang, C. Li, Y. Zhang, C. Xing, and J. Yang, “Semanmedical: A kind of semantic medical monitoring system model based on the iot sensors,” in e-Health Networking, Applications and Services (Healthcom), 2012 IEEE 14th International Conference on. IEEE, pp. 238–243, 2012.

[26] J W Smith, J E Everhart, W C Dickson, W C Knowler, R S Johannes, “Using the ADAP learning algorithm to forecast the onset of diabetes mellitus”, In Proceedings of the Symposium on Computer Applications in Medical Care,Washington, DC, USA, IEEE Computer Society Press: Los Alamitos, CA, USA, pp. 261–265,1988.

[27] D Kibler, W A David, and K A Marc, "Instance□based prediction of real□valued attributes." Computational Intelligence, vol. 5, no. 2,pp. 51-57,1989.