An Ontology-based Framework for Querying Heterogeneous Sensor Data

A Thesis

submitted in fulfillment of the requirement

for the award of degree of

Doctor of Philosophy

Submitted By

Diksha Hooda

(Roll No.: 901703004)

Under the guidance of

Dr. Rinkle Rani

(Associate Professor, Computer Science and Engineering Department)

Thapar Institute of Engineering & Technology, Patiala

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR INSTITUTE OF ENGINEERING & TECHNOLOGY

PATIALA - 147004, Punjab

July 2022

Contents

List of Figures

List of Tables

List of Algorithms

Certificate

I hereby certify that the work which is presented in this thesis entitled **"An Ontology-based Framework for Querying Heterogeneous Sensor Data"**, in fulfillment of the requirement for the award of degree of **"Doctor of Philosophy"** submitted in Computer Science and Engineering Department of Thapar Institute of Engineering & Technology, Patiala, is an authentic record of my own work carried out under the supervision of **Dr. Rinkle Rani** and refers other research works which are duly listed in the reference section.

The matter presented in this thesis has not been submitted for the award of any other degree of this or any other university.

Diksha

(Diksha Hooda)
Regn. No. 901703004

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Rinkle

(Dr. Rinkle Rani)
Associate Professor
Computer Science & Engineering Department
Thapar Institute of Engineering and Technology, Patiala, India

Acknowledgement

First and foremost, I would like to express my warmest gratitude to my supervisor, **Dr. Rinkle Rani**, for her excellent guidance, immense support and motivation throughout my research work. Her positive mindset, strength, and words of encouragement have always pushed me to work harder for the entire duration of the research work. The optimistic lessons that I have learned from her will stay with me forever for good.

My sincere thanks to **Dr. Shalini Batra**, Professor and Head, CSED, Thapar Institute of Engineering and Technology, Patiala, for providing me the necessary administrative assistance and infrastructure that helped me in the completion of my research work. I am thankful to my doctoral committee members: **Dr. Ashutosh Mishra**, Assistant Professor, CSED, **Dr. Deepshikha Tiwari**, Assistant Professor, CSED, and **Dr. Kulbir Singh**, Professor, ECED, Thapar Institute of Engineering and Technology, Patiala for their insightful suggestions and regular monitoring of the research work. I am thankful to all the faculty, staff members and research scholars of CSED for their support.

I would like to offer my heartfelt thanks to my father, **Mr. Sujender Singh Hooda**, for his life lessons and kind words that gave me the sense of encouragement and confidence throughout this journey. I am thankful to my mother **Mrs. Reena Hooda**, my confidante, whose love and care have kept me going. Also, I express my thanks to my brother, **Sahil** and sister-in-law, **Paarkhi** for their constant support, suggestions and confidence in me. I also acknowledge the cooperation and encouragement by my friends all this while.

Last but not the least, I bow my head in front of my God for showering His grace on me, without which any of this, would not have come to life.

Diksha

Patiala (Diksha Hooda)

July, 2022

Abstract

Advancement in the Sensing Technology has resulted in the wide spectrum of modality and heterogeneity in data captured by heterogeneous sensors. This randomized temporal data has immense potential to generate information that can drive the process of gaining insightful domain knowledge to achieve efficient decision making. Due to this existing heterogeneity, there is a lack of interoperability of various sensor devices giving rise to complexities in the process of sensor data fusion and aggregation. Hence, there is a need for managing sensor data dynamics in a way that facilitates efficient aggregation, searching, analyzing, reuse and exchange of sensed data. To achieve all this, the aggregation needs to be performed at concept/entity level in meaningful ways to present information in an interpretable format that generates extensive knowledge and value. Semantic approaches - Ontologies - play an important role in solving these issues. In order to achieve the task of gaining knowledge through intelligent analysis of sensed data, Semantic Web technologies are implemented to achieve interoperability of sensing devices and systems. Ontologies are defined as "well-founded mechanism for the representation and exchange of structured information". These technologies can aid in the management, storage, fusion of sensor devices and measurement data, so as to facilitate efficient information retrieval and mining from the Ontology-Based Knowledge Systems.

This thesis presents research work undertaken to perform efficient heterogeneous sensor data analysis with the development of Ontological Knowledge models and fetching information through SPARQL-based query firing. Two ontological knowledge models are proposed for activity recognition. The first one targets heterogeneous sensors generating homogeneous data and the other one is a probabilistic ontological model for targeting and fusing multimodal data from heterogeneous sensors for efficient activity recognition. Another ontological model takes in input as digital heterogeneous data from ECG-based sensors to classify

heartbeat data into arrhythmia classes. It solves the issue of interoperability and heterogeneity in the input sensor data. An interval type-2 fuzzy-based ontological model is developed to target the uncertainty and heterogeneity in the heterogeneous water sensors data. The proposed model ensures effective handling of ambiguous and uncertain sensor observations, a shortcoming of elementary/classical knowledge-driven models.

For all the proposed ontological knowledge models, SPARQL-based querying service is deployed to draw insightful knowledge in the from of inferences. The resultant frameworks solve the issue of sensor heterogeneity and provides solutions that promote reusability, interoperability and exchange of domain knowledge.

Keywords: Heterogeneous Sensor Data Aggregation, Ontology-based Knowledge Systems, Semantic Technology, Sensor Fusion, Knowledge Engineering, SPARQL Querying, SWRL-based Reasoning.

Chapter 1

INTRODUCTION

> This chapter discusses the fundamentals of Semantic Technology and its deployment in the field of Knowledge Engineering from heterogeneous sensor data. Broadly, the discussion revolves around the integration of Ontologies with Knowledge-based Systems developed on multi-modal sensed information from heterogeneous sensors.

In 1993, Gruber[1] defined the concept of Ontology as "an explicit specification of a conceptualization". In 1997, Borst[2] redefined Ontology to be as a "formal specification of a shared conceptualization". This conceptualization was defined to be a shared view rather than an individual view of the specification. In 1998, Studer et al.[3] merged the two definitions and hence, stating, "An Ontology is a formal, explicit specification of a shared conceptualization". The term "Conceptualization" can be defined as an abstract and simplified view of the real world that needs to be represented. Every Knowledge-based System or Agent is bound to some implicit or explicit conceptualization. Due to lack of the shared view of the ontology from various stakeholders, the benefits that can be achieved from Ontology-based knowledge system are restricted. The ontology can eventually turn out to be of no use if it defies the shared aspect of the ontological commitment. Thus, it is considered to be of great importance that the ontologies constructed to support large-scale interoperability are well founded in the sense that the primitives are efficiently chosen to be well understood and distinctive in the general sense.

Ontology is a machine-readable form of concepts, functions, features, attribute, axioms and facts as consensual knowledge and a simplified view of some complex phenomenon through abstract modelling. They are the fundamental terms and facts that define the vocabulary and jargon of a particular domain. It comprises of relations to provide extensions to the already defined facts and axioms. Information extraction usually revolves around the retrieval of information in a specific domain, so explicitly specifying concepts of the domain in the form of formal representation through ontology can prove to be very beneficial for this process. A Knowledge Base comprises of well defined classes along with the instances of those classes. Classes act as the fundamental constituent of Ontology creation. Classes in ontology define the concepts of a domain in concern.

The word "ontology" is applicable in different contexts in respect to different communities. The two major areas where ontology has its applicability are so radically different, one being the philosophical sense and the other being the computational sense.The philosophical context revolves around the nature and structure of "reality". The computational discipline of Ontology has recently emerged in the Knowledge Engineering domain. For AI systems, something that exists can be represented. The computational ontologies formally model the structure of a system representing the entities and their relationships explicitly.

1.1 Knowledge-based Systems

Knowledge-based Systems were first developed by Artificial Intelligent scientists which were primarily rule-based expert systems. These systems were based on facts that were represented as simple statements and declarations in regular repository, driving addition of the rules-based reasoning results into the repository. Increase in complexity of the Knowledge-based systems promoted more sophisticated mechanisms for representation of the Knowledge Base. These systems are not completely dependent on intelligent reasoning and effective search techniques but the strength of problem solving lies in the domain specific real-world knowledge. Ontology building is an efficient way for modelling for declarative knowledge. Ontologies can be created to facilitate sharing and reuse of knowledge and can be merged for additional information[4]. They are based on static domain knowledge which can be dynamically used in the reasoning process for solving real-world problems.

1.2 Ontologies in Knowledge-Based Systems

With the increase in the amount of knowledge being published, a need for information systems that facilitate management, storage and search arises[5]. Ontology editors, e.g. Protege[6], are systems that provide a means for the creation, editing and investigating arbitrary ontologies without forming any assumptions about the ontology domain. But, in order to make an information system readily available and usable by a non-expert in ontological engineering, domain-dependent business logic (e.g. complex computations) and domain-specific user interface (e.g. specific visualization) is required. The design and maintenance of these domain-specific, concept-based information systems that implement semantic ontologies as being their data supply source is non-trivial as compared to other database-based information systems[7]. Firstly, databases are based upon the concept of Closed World Assumption (CWA), which imposes different constraints on the data easily. In Closed World Assumption, if a database authorizes the last name of a person should be known, and if in case, it is absent, as per the closed world assumption, it is inferred that no such name is given and constraint violation is encountered. These integrity constraints ensure data quality but they also end up making the data structure very rigid and limits the flexibility. Even incorporating a small alteration i.e. attribute addition or deletion in the data model, significant work by the application designer like adjusting the logic accordingly, updating the application model, and recompiling the application is required. On the contrary, Ontologies follow Open World Assumption (OWA), which is considered to be appropriate for inferring fresh and new knowledge in the domain[8]. For example, an ontology that authorizes that each and every person should have a last name, can be consistent, regardless of the fact that the last name of any person is unknown or not - if the last name is not present, its existence (although not necessarily its particular value) can be implied and inferred. Though, the differences between (open-world) ontological and (closed-world) database descriptions are of great value but, are usually ignored by knowledge system designers who make assumptions about the underlying ontology structure similar to the database schema. As a result, the ontology evolution[9] that takes place during the runtime of application violates the information system's assumptions that are consistency of the ontology and the dependent application like incomplete ontological data with respect to assumed data structures, at some point. It becomes difficult to assure the desired and correct functionality of the logic in case of the evolution of ontology. Hence, ontology changes might cause unexpected behaviour of the application. This application can cause damage to the ontology by generating ontological data that result in overall inconsistency. Secondly, in comparison to the traditional databases, OWL ontologies are demonstrative enough to depict taxonomies for example, "each woman is a person", complex relationships among objects and entities like "being someone's uncle means being a brother of his/her parent", cardinality constraints like "each person has exactly two arms", negation relations like "every person not being a man must be a woman". This extends the already represented knowledge which can be reused in many other applications. The need for complex query engines that can handle and exploit the expressive knowledge in applications is surfaced. Such query engines need to be efficient enough in order to exploit not only the domain-dependent data but also data about relationships between domain-specific concepts and individuals. With the growing and evolving domain knowledge, and intricate relationship, the implementation of such knowledge model with relational database as the underlying technology, becomes difficult and hard-to-maintain. This was said to be the motivation for representing the domain knowledge by means of semantic ontologies.

1.3 Sensor Data and Ontology-based Knowledge Systems

With the advent and adoption of the Internet of Things (IoT) technology in different domains like health, transportation, and manufacturing, the IoT devices in the world are expected to rise to a number approximately equal to 50 billion by the year 2020[10]. These IoT devices capture huge amount of data through the sensors embedded within them[11,12]. In figure 1.1, as per a Cisco report, the amount of annual global data traffic is expected to touch 10.4 ZB (Zettabytes) by the end of 2019 and 2.5 quintillion bytes daily. This enormous data can be multi-modal in nature consisting of distinct formats like video streams, images, and textual data. The management of such large-scale sensor data and subjecting to processing tasks acts like a key factor in the building smart applications. Semantic approaches- ontologies – play an important role in solving these issues. Ontologies are defined as a "well-founded mechanism for the representation and exchange of structured information".

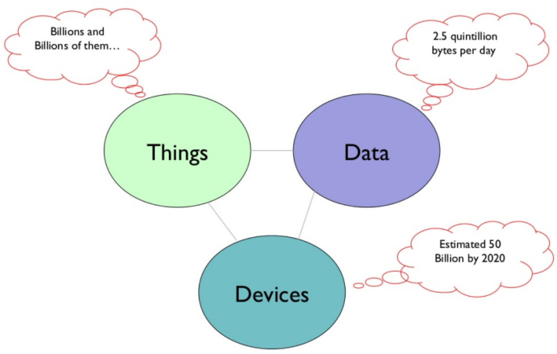


Figure 1.1: Scale of IoT Problem

1.3.1 Sensors

Sensor consists of two main elements: First, a sensing device and its specification and second the measurements and observations, both of these clubbed together are known as sensor data. The sensor data is collected by sensory equipments [13]

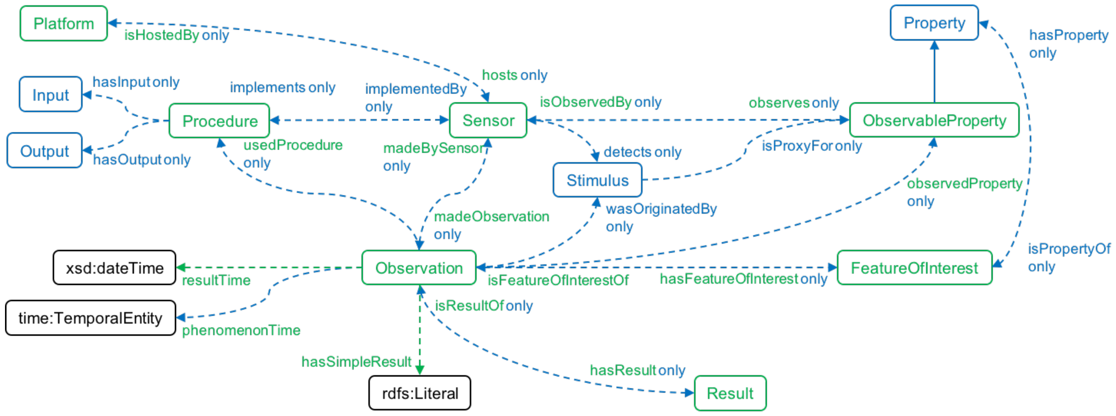


Figure 1.2: Semantic Sensor Network Ontology

and even through crowd sensing sources. It is a function of time and location, since it is captured with respect to a particular time duration or instant. It can be noisy and the quality of the sensed data can vary depending upon various external or device-related factors. Also, the captured data is often continuous and streaming from multiple heterogeneous sources in varying formats and representations. This data from distinct sources when integrated often generates value and creates an ecosystem of systems. The large availability of sensor data throws light on the need of efficient data analysis performed by computer-aided systems over the obtained static data or incoming dynamic streaming data. So, due to the growing prevalence and usage of sensor instruments and systems, management of the sensing devices and the generated volume of data has become a crucial task. The information systems and services that are developed on sensor networks face the challenge of bridging a gap between captured sensor data and domain-specific technology for performing efficient knowledge mining and data analysis. In order to achieve the task of gaining knowledge through intelligent analysis of sensed data, Semantic Web technologies are proposed to achieve interoperability of sensing devices and systems. These technologies can aid in the management, storage, fusion of sensor devices and measurement data, so as to facilitate efficient information retrieval and mining from the Ontology-Based Knowledge Systems. The Semantic Sensor Networks Ontology in the domain of IOT in Figure 1.2 describes sensor devices and their metadata, the observations and the related concepts. It does not describe other domain concepts, time, locations etc. These other concepts can be included and integrated from other ontologies through the OWL imports.

1.3.2 Heterogeneity in Sensor Data

The sensor data is collected by sensory equipments and even through crowd sensing sources. It is a function of time and location, since it is captured with respect to a particular time duration or instant.

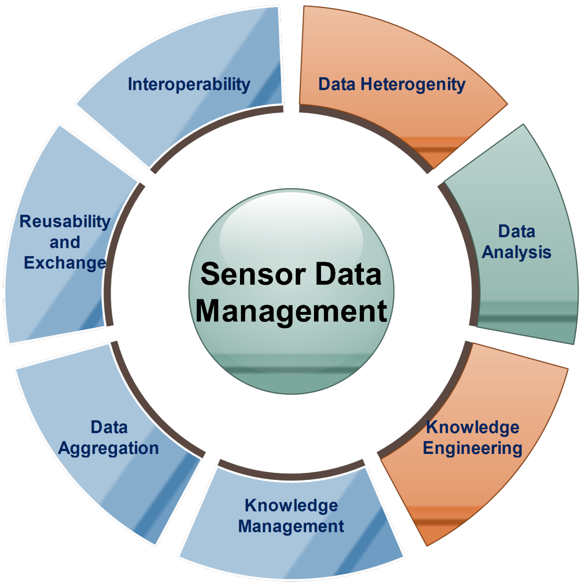


Figure 1.3: Sensor Data Management Aspects

It can be noisy and the quality of the sensed data can vary depending upon various external or device-related factors. This data can be multi-modal and heterogeneous in nature consisting of distinct formats like video streams, images, and textual data. It is often dynamic coming from multiple heterogeneous sources. This data from distinct sources when integrated often generates value and creates an ecosystem of systems.

The large availability of sensor data throws light on the need of efficient data analysis performed by computer-aided systems on static data or incoming dynamic streaming data. So, due to the growing prevalence and usage of sensor instruments and systems, management of the sensing devices and the generated volume of data has become a crucial task. The information systems and services that are developed on sensor networks face the challenge of bridging a gap between captured sensor data and domain-specific technology for performing efficient knowledge mining and data analysis. Figure 1.3 shows all the aspects of Sensor Data management.

In order to achieve the task of gaining knowledge through intelligent analysis of sensed data, Semantic Web technologies are proposed to achieve interoperability of sensing devices and systems. The underlying mechanisms of an Ontology-based Knowledge System like data storage, querying, semantic annotation and retrieval are required to be efficiently devised in order to assist in the effective and desired sensor analytics process. Information or Knowledge Systems need functions exploiting special technologies in order to reduce intricacies and complexities existing within the data and processes. A powerful and rich Knowledge System can be developed when backed up with the Semantic Technology for the crucial data operations.

1.3.3 Semantic Technologies

Semantic Technology utilises formal semantics to annotate raw disparate data scattered around us. This technology provides techniques to connect concepts of the real world and build relations between data in varying sources and formats. Hence, as a result context is built in the data and links are created out of the constructed relationships. It defines and establishes bridges between the data through formal languages that are efficient enough to express self-explanatory and rich relationships of data in a form that is readable and can be processed by machines. Hence, the machines are not only able to process the data but also become capable to store, manage and retrieve information on the basis of annotated and logically defined relationships. The main difference that exists between the semantic technology and the traditional data technology like relational database is that the former revolves around the context and meaning of data rather than its structure and schema. Semantic Technology can be exploited to bridge a gap between different domains in order to gain deeper insights about the data sensed from the sensors represented in the Sensor Ontology. The sensor ontology can be integrated with Weather Ontology (provides information about different weather conditions), Temporal Ontology (introduces temporal i.e. time aspect with the sensed data), Geospatial Ontology (provides concepts to associate sensor data with location of sensor devices when under movement) etc.

Semantic Technology is based on a set of universal standards defined by the World Wide Web Consortium. This technology provides a universal medium that facilitates the exchange and reuse of annotated data by allowing a global common sharing platform of all categories of data like personal, scientific, cultural and commercial data. The main standards that this technology is developed upon are: Resource Description Framework (RDF), SPARQL (SPARQL Protocol and RDF Query Language) and OWL (Ontology Web Language).

- **Resource Description Framework (RDF):** It is a standardized framework for describing resources. It is a model that can be read, understood and processed by the machine. RDF was designed to assist in proving a common platform to describe data and information so as to make it readable and understood by the computer. RDF has functionalities that allow merging of data which leads to enhanced knowledge capturing from the knowledge model. RDF represents data in the form of triples i.e. three positional statements. For example, the fact, "Barack Obama is married to Michelle Obama" can be represented as RDF triple: <Barack Obama> hasSpouse<Michelle Obama> with Barack Obama as being the subject, hasSpouse as the predicate and Michelle Obama as the object of the statement or fact. The subject and the predicate are said to be resources which are uniquely identified by URI and the object can either be a resource or a literal value. The Uniform Resource Identifier (URI) uniquely identifies each resource and serves as the building block of the data structures that can be identified and understood by the machine.

Why RDF? It is considered to be a flexible data model due to its power of linking multiple disparate data sources irrespective of their schemas and provides a standardized platform to achieve interlinking of all that is known and asserted through facts, records, information etc.

- **Ontology Web Language (OWL):** It is a logic-based computational language that provides semantic annotation to the data represented as RDF triples. It represents copious and complex knowledge about the relationships and hierarchies between entities represented as concepts. It complements RDF and formalizes the representation of ontology in a particular domain. The OWL documents which are actually ontologies, when published in the World Wide Web can be accessed by other OWL ontologies for knowledge enhancement and gain deeper insights.

- **SPARQL (SPARQL Protocol):** It is a semantic query language that is used to query and retrieve information stored in RDF stores. SPARQL has functionality to capture patterns in the RDF graphs or triple store and return the input query's conjunctions and disjunctions. The SPARQL query results can be in the form of result sets or even RDF graphs.

1.4 Thesis Organization

The thesis is organized as per the following chapters:

Chapter 1: Introduction

This chapter discusses the fundamentals of Semantic Technology and its deployment in the field of Knowledge Engineering. Broadly, the discussion revolves around the integration of Ontologies with Knowledge-based Systems developed on sensed information.

Chapter 2: Background and Related works

This chapter comprises of a comprehensive survey of Ontology-based application areas, Knowledge-based systems built on ontologies and various Ontology mapping techniques for performance improvement.

Chapter 3: Problem Formulation

This chapter includes the research gaps identified based upon the literature review performed in the initial phase of the study. The gaps are identified in terms of dealing with data heterogeneity, efficient handling and interoperability in Knowledge-based Systems and shortcoming in data analysis using existing Knowledge-based Systems.

Chapter 4: Ontological model for heterogeneous sensor data

In this chapter, semantic reasoning is performed on Ontological Knowledge Base developed on observations captured by heterogeneous sensors and domain knowledge to perform human activity recognition. Further, axioms are designed to achieve efficient data analysis and knowledge entailment.

Chapter 5: Probabilistic Ontological model for heterogeneous data of heterogeneous sensors

Here, Probabilistic Ontological model is developed to efficiently analyze multi-modal/heterogeneous sensor data from heterogeneous sensors. The resultant model is deployed to perform daily living activity recognition through the implemented Ontological Knowledge Base enriched with Probabilistic reasoning. The developed deals with the issue of heterogeneity in data and interoperability for knowledge reuse and exchange. The results showcase the proposed model outperforms existing traditional data-driven approaches in terms of precision and accuracy.

Chapter 6: Ontological Knowledge model for heterogeneous digital sensors data

In this chapter, heterogeneous digitized sensor data is semantically annotated to built a modular Ontological Knowledge model to achieve cardiac arrhythmia detection. The domain knowledge is captured and transformed into novel and newly developed ontologies for SWRL rule-based reasoning and entailment. The proposed framework is compared with old as well as new AI-based approaches to exhibit greater performance and efficiency in the recognition process. Also, knowledge retrieval is performed through querying service deployment over it.

Chapter 7: Interval Type-2 Fuzzy Logic driven Ontological model for heterogeneous sensors data

This chapter discusses in detail the design, development and implementation of Type-2 Fuzzy Logic based Ontological framework for heterogeneous sensors data. This developed knowledge model is instantiated with a real world data from water quality sensors to perform water quality detection. Additionally, the verification and validation process showcases the robustness, flexibility and consistency of the system. The comparison with other existing approaches shows performance enhancement in contrast to other techniques.

Chapter 8: Conclusion and Future Directions

This chapter concludes the thesis by providing a brief overview of proposed approaches and providing insight into the future scope of work. Summarization of future directions in the fields of Ontological Knowledge model implementation and its deployment in various application scenarios. It highlights the contributions and enlists the points that need to be worked in near future.

Chapter 2

BACKGROUND AND RELATED WORK

This chapter comprises of a comprehensive survey of Ontology-based application areas, Knowledge-based systems built on ontologies and various Ontology mapping techniques for performance improvement. Section 2.1 contains a bibliographic study of ontology-based application areas. Section 2.2 contains a comparison of the existing Ontology based IoT Techniques. Section 2.3 contains study of existing heterogeneous sensor data fusion techniques.Section 2.4 comprises of analysis of various ontology mapping and merging approaches for optimized performance.

Semantic Technologies largely hinge on Ontologies for making the cooperation between machines and human beings possible.In the field of Computer Science, in computational context Ontology provides a logical structure for knowledge modeling and facilitating inference tasks crucial for implementation of intelligent systems [14]. This chapter contains a detailed review on various application areas of Ontologies, along with various Ontology mapping techniques that are a part and parcel of designing intelligent systems.

The development of ontologies is driven by the modularization aspect and reuse of the developed ontologies. These ontologies are used together in complex applications. Autonomously developed ontologies emerge quite naturally in different domains (health, tourism, learning, quality of services, etc.). These ontologies, each one built for different purpose, are used together in complex applications. However, how they are combined is usually hidden in the application code, without explicitly expressing the way how ontologies are combined for a specific purpose.

2.1 Ontology-Based Application Areas

Medical Domain: E. Manzini et al. proposed a translation tool for a human phenotype ontology that builds the gap between the clinician and the care receiver. It takes the concepts in the ontology as input space and performs neural model driven translation to give layman understandable output space[15].

Database Enrichment: C. Blankenberg et al. mapped the Ontology as a graph database that effectively aligns ERP objects with conceptual schema and allows for loading of the data objects in real-time efficiently. The ontological data in the form of business objects are semantically queried and visualized for the user[16].

Search Engines: J. Ingram and P. Gaskell proposed a methodology to design and construct an user-centric ontological database that supports development of a search engine for agricultural practitioners. In the field of agriculture, it is important to be able to have access to data intelligently to achieve mobilization of resources for innovation induction. The proposed work aims to achieve the same by deploying the developed ontology across Europe. This ontological approach has advantages over the traditional search engines by incorporating a kind of interactive exchange of meaning to achieve replication of the communication process in its true sense[17].

Semantic Reasoning in IoT: G. Chen et al. proposed a semantic reasoning model based on ontologies for IoT in order to target the needs of the IoT driven health cabin needs. The model developed is based on dynamic ontology modeling and reasoning to perform real time analysis in IoT systems. The developed system deploys rule based reasoner to generate entailments for future decision making. The ultimate application of this work lies in the autnomous classification and attribute based enhancement of behaviour and resources of IoT system. It has largely improved the automatic control of the IoT network and its resources.[18]

Table 2.1: Ontology Application Areas

Author(s)	Application Area	Summary	Outcome
Manzini E. et al.[15]	Medical Domain	Proposed a translation tool for a human phenotype ontology that builds gap between clinicians and patients.	Model driven translation to give layman understandable output space
J. Ingram and P. Gaskell[17]	Semanttics Search Engine	proposed a methodology to design and construct an user-centric ontological database that supports development of a search engine.	This has advantages over the traditional search engines by incorporating a kind of interactive meaningful exchange
Hippolyte J.L. et al.[19]	Ontology web service in energy domain	OWL ontology developed to indentify upper level ontologies along with other specific ontologies.	Evaluation of the developed model by subjecting it to datasets and querying in SPARQL to validate results.
Chen M. et al.[18]	IOT & Semantic reasoning	Semantic reasoning model based on ontologies for IoT in order to target the needs of the IoT driven health cabin needs.	It has largely improved the automatic control of the IoT network and its resources
J. Warner[20]	Disaster Mitigation	work aiming to achieve ontological security in times of disasters.	Efficient emergency management capability to implement ontology-based emergency plans.
Muljarto A. R. et al.[21]	Agriculture	Agri-Food Experiment Ontology to moel viticulture activities and wine making operations.	Resultant system provides linked data to prospective researcher for detailed analysis of various wine making techniques.
Bagherifarda K. et al.[22]	Recommender Systems	Ontologies to represent Content based Filtering (CBF) and Collaborative Filtering (CF) tasks of the Hybrid Recommender System.	Resultant system uses a hybrid approach integrating ontology and clustering to get accurate results.
O.N. Oyelade et al.[23]	Medical Science - Cancer	To propose an approach for ontology-based analysis of Tumor Suppressor Genes (TSG) and non-TSG's.	In future, aim to devise new cancer prevention treatments based on the analysis.

Disaster Management: J. Warner presented the work aiming to achieve ontological security in times of disasters such as the one that occurred in Netherlands in 2000. This plays an important role in maintaining the trust of the community and disaster affected populace in the regulatory organization deploying ontological security[20].

Data Integration: T. Sobral et al. developed an ontology-driven system that supports integration and visualization of the data from Intelligent Transportation and Mobilization Systems. The work includes spatio-temporal aspect of the data to add a large scale semantic to the knowledge represented. As a part of the project, transport related data from Portugal was utilised for demonstration of practical applications of ontological knowledge represenation for each facet[24].

Hybrid Recommender System: K. Bagherifarda et al. have extended recommender system by using a hybrid approach which employs an ontology and enhanced clustering to get improved and accurate results. Ontologies are used in the Content Based Filtering (CBF) and Collaborative Filtering (CF) portions of the Hybrid Recommender System to enhance accuracy of the results obtained. The Hybrid approach is based on ontological inference to promote effective user profiling and achieve an automated recommender system[22].

Discovery: B.J. Cortez proposed a method that generates descriptions in natural language of cartoons and then extracting entities from texts, followed by finding the extracted entities in the existing ontological systems to promote the reuse of data represented. It promoted reusability of existing benchmarked ontologies such as the infamous Gene Ontology, Sequence Ontology, Relations Ontology and ChEBI ontology[25].

Target based Sentiment Analysis: D. Meskele and F. Frasincar presented an ontological hybrid model for aspect based sentiment analysis. In this work the domain knowledge is represented in the form of ontological knowledge base.

The sentence based neural model driven analysis is performed through a lexical domain ontology. The manually developed ontology is integrated to exploit field centric data and knowledge. The ontological aspect of knowledge is integrated with neural network based classification to enhance performance [26].

2.2 Ontology based IoT Techniques

The Internet of Things (IoT) has now become one of the most promising and prospective technologies. IoT-based systems generate huge amounts of data that require meaningful analysis to perform effective decision making. This integration of IoT analysis and Ontological reasoning has booming application in the real-world scenario.

OBDAIR: G. Santipantakis et al. presented an ontology-based distributed framework that performs data retrieval, its integration and reasoning with data extracted from disparate static or dynamically updated heterogeneous sources. The proposed OBDAIR framework supports scalable data-driven domain dependent applications by accomplishing processing of large amounts of data in close vicinity to data sources and hence, supports knowledge mining in a decentralized manner. The authors implemented modular ontology approach for representation of disparate data sources to achieve a distributed framework [27].

Semantic interoperability in IoT infrastructure: F. Ullah et al. proposed an Interoperability Model for Big-data in IoT (SIMB-IoT) to achieve semantic interoperability among different heterogeneous IoT devices in the domain of health care. The proposed model recommends medicines with their respective side effects for different symptoms extracted from heterogeneous IoT bodily sensors. The presented approach takes input two big datasets for the analysis: one dataset contains diseases with prescribed drug details and the second dataset contains medicines with their respective side effects caused [28].

FIESTA-IoT: M. Serrano et al. preented the FIESTA-IoT and its experimental infrastructure that provides to IoT based European experimenters rendering the capability of accessing, sharing and reusage of IoT devices generated semantically annotated dataset in a testbed way. The project objectives enable experiment execution and handling across multiple testbeds for IoT, based on single API for experiment submission and single unique set of credentials for researcher. It allows portability of the IoT based experiments across distinct testbed and provision of allowing for interoperable standards for IoT/cloud technology. This project allows the amalgamation of IoT system's resources, infrastructure of testbeds and the associated applications of the same.[29].

LOV4IoT: A. Gyriad et al. proposed a Machine-to-Machine Measurement (M3) system which is deployed online and it aims at allowing researchers to rapidly design and construct semantic-based IoT driven applications. It promotes reusing of existing ontologies as much as possible in the form of the domain knowledge represented as ontologies, datasets and constrainrs/rules on them. Semantic Web of Things is an emerging field of research integrating Semantic Web and IoT technologies including smart objects and application scenarios connected to Web.[30].

IoT-based Human Acitivity Recognition: L. Buoncompagni et al. proposed a framework to integrate networks of comparatively smaller ontologies to perform knowledge modeling and hence enabling smart devices driven homes to achieve human activity recognition. In the IoT network, nodes are considered as ontologies which allow for data contextualisation, and the edges are computational functionalities /procedures that elaborate data. The proposed system, Arianna+, is a system with flexible interface that runs smoothly between the inputs and outputs, which are known to be as atomic representations of ontological knowledge[31].

Table 2.2: IoT driven Semantic Techniques

Author(s)	Input Data	Proposed Work	Summary
Santipantakis G. et al.[27]	Static and Dynamic Vessel Trajectory data and Vessels/Ports details	Detection of complex events like collision in maritime area.	Deals with RDB data only and lack of spatio and temporal reasoning.
Ullah F. et al.[28]	IOT sensor disease symptoms data, Drug details data and Diseases-Drugs data	System provides medicines to patients with side effects.	Syntactic Interoperability and security issues not addressed.
N. Sharma et al.[32]	Biomedical Signal Data	Ontology based reasoning on sensor driven signal data .	Coozo simulator and OWl are used for implementation
T. Elsaleh et al.[33]	IoT streaming data	Semantic annotations in the pipeline of IoT and performing sensory data analytics	Ontological database of benchmarked ontologies is used for ontology alignment
M. Serrano et al.[29]	IoT testbed data	FIESTA-IoT allows accessing and sharing of IoT based semantically annotated data in a testbed-agnostic format.	Ontology and IoT techiques are integrated for sensor data analytics
A. Gyrard et al.[30]	Iot based streaming data	Machine-to-Machine Measurement (M3) system to design and construct semantic-based IoT applications	Linked Open Vocabularies are deployed for reuse
L. Buoncompagni et al.[31]	Sensor based multimodal data	Modular ontological approach for semantic reasoning and activity recognition	Logic based and data driven approach is used.

2.3 Heterogeneous Sensor Data Fusion Techniques

In order to develop sensor driven intelligent information/knowledge based systems, it is crucial to integrate functionalities to accommodate multiple sensors data, heterogeneous data [34], multi-functions, and man-machine interaction based decision making. However, existing heterogeneous sensor data management and fusion architectures are not comprehensive enough to give efficient performance. Most knowledge systems or frameworks focus only on simplex sensor data fusion, i.e., integration without considering the aspect of decision-making.

Table 2.3: Existing Heterogeneous Sensor Data Fusion Techniques

Author(s)	Fusion Technique	Approach	Limitations
Troyanskaya et al. [35]	Bayesian-based framework	Uses Bayesian reasoning to integrate varying format biological data .	input data is just matrix based, multi-modal data is not targeted.
Khalil et al. [36]	XML Schema-based fusion	XML schema format to denote sensory data	Modality of sensory data is simple and in the same XML Schema format.
Dai et al. [37]	Neural Concept Liking approach	A data integration problem for health-care data through concept linking is proposed	Sensory Data is not multi-modal and heterogeneous
Hachour et al. [38]	Generalized Bayes' Theorem based approach	Proposed a ProbAbilistic matching algorithm for sensory outputs	Data is mono-sensor and synthetic, hence, questionable performance
Poekaew et al. [39]	Adaptive Principal Component Analysis (A-PCA) technique	Dimensionality-reduction technique for heterogeneous sensory data aggregation	Exhibits decreased performance when the sensor data are not related.

2.4 Techniques for ontology mapping

The Overlap among various ontologies can occur due to myriad of reasons. One of the reasons is that a single ontology can not provide sufficient coverage for all the possibilities for the concerned application. This need for representation of sufficient information about the application aspect of the domain gives rise to development of more application based ontologies. But, somehow these newly developed ontologies need to be aligned with the existing benchmarked ontologies to provide an abstract view of knowledge reuse and implemenation.

The Mapping of ontologies has applicability in various large domains like biomedical, healthcare, laboratory investigation etc. This task of mapping of knowledge from different fields of research is driven by the fact that the discovery of the evidence of a relationship/property applying techniques like data/text mining

Making Sense of Sensor Data Using Ontology: M. Stocker et al. proposed application of automated and formalized representation of knowledge captured from the sensor network data. The authors aimed bridge the gap between raw sensor data and measurements and abstract domain terminology or vocabulary. The authors presented how sense can be made out of high-volume numerical and temporal sensor data which is very crucial for real world application base scenarios. The future decision making can be driven by the quality of analysis done in the near past[40].

A Method for Mapping Sensor Data to SSN Ontology: X. Zhang et al. proposed a mechanism to achieve mapping sensor data to SSN (Semantic Sensor Network) ontology automatically based on a predefined XML-based mapping file. The authors developed a mapping language which defines a foundational schema to achieve annotation of sensors and data sources to produce a XML document to achieve mapping. This logical underlying schema becomes the blueprint for the process of ontological mapping and alignment[41].

Connecting IoT Sensors to Knowledge-based Systems: X. Su et al. presented an approach for interlinking SenML-enabled IoT sensors to semantic knowledge-based systems. A real use case study from fish industry is presented and a comparison of SenML with other formats is done and concluded that there is resource-efficiency associated with the usage of SenML. The proposed approach connects a huge number of resource dependent IoT sensors through SenML[42].

Table 2.4: Frameworks/Techniques for Ontology Mapping

Author(s)	Optimization Technique	Approach	Limitations
Stocker M. et al. [40]	Mapping Sensor Data to Ontologies	Sensor data is represented as ontology detecting some events	Not for use case with low frequency sensing
Zhang X. et al. [41]	Sensor Data Mapping	Algorithm based mapping of sensor data to SSN through mapping file	Shows good performance for limited cases.
Su X. et al. [41]	Common format for sensor data	Conversion of SenML data to RDF through mapping	Resource Usage not monitored
Zander S. et al. [43]	Ontological Semantics and Reasoning	Detailed Sensor RDF data representation to achieve semantic reasoning	It can be counter-productive, preventing deducing of extra information
Santipantakis G. et al. [44]	Modular Ontologies	Distributed approach to integrate, store, reason with sensor data	Not efficiently scalable to large network of peers
Myłkar A. et al. [45]	Heterogeneous Sensor data integration	X2R,system to convert different sensor formats into ontology	No parallelization of tasks support.
Ouksili H. et al. [46]	Pattern Oriented RDF querying based on SPARQL	Algorithm based on theme exploration and keyword search.	Lack of optimized clustering process
Nikolaou C. et al. [47]	Dealing with incomplete information in RDF data	Working on relational D/B to deal with unknown data	Lack of handling of incomplete information in subject and object place
Ouksili H. et al. [48]	Efficient Querying on RDF data	Clustering based algorithm for theme identification for querying results	Precision of the clustering and lack of semantic cluster annotation
Angles R. et al. [49]	Graph database for RDF data	Incorporating graph D/B functionality with RDF	Lack of effective query benchmarks for graph-like model

Enhancing the Utilization of IoT Devices Using Ontological Semantics: S. Zander et al. presented an approach that substantiates the assumption that a more detailed description of a device's functionalities and its capabilities to accommodate the device with different system contexts and paves the way for interoperability among disparate IoT devices in sensor networks. The author demonstrated the basic conceptual and technical information formalized as RDF data can be used to automatically classify and identify IoT devices. The authors presented an approach for adding the descriptions and definitions of IoT devices with extra rich information inferred from the formal theoretic semantics of domain –specific ontologies to enhance the utilization and interoperability among sensing devices [43].

Reasoning with data from data sources using modular ontologies: G. Santipantakis et al. proposed a distributed or decentralized framework for retrieving, merging and reasoning with data coming from heterogeneous and distinct data sources. The proposed solution merges the E-SHIQ modular ontology framework with the Ontop ontology-based data access (OBDA) technology. The discussed approach allows the autonomous treatment of data from distinct data sources, accommodating the parallelization of operations, and achieves an efficient reasoning with the data. The presented work achieves processing of data, before data reaches the reasoner, at the time of accessing it near the source of the data [44].

Integration of heterogeneous data in an ontology-based knowledge system: A. Myłka et al. proposed a system, X2R, to integrate heterogeneous data sources in an ontology-based knowledge system. The authors devised a system that presents an unified view of information stored in relational, XML and LDAP data format within an organization in RDF model through a common ontology and following a set of integrity constraints. They also leveraged in a novel approach to achieve semantic optimization of SPARQL queries [45].

Pattern oriented RDF graphs exploration: H. Ouksili et al. introduced two complementary approaches to explore RDF(S)/OWL data: theme-based exploration and keyword search method. These two approaches are based upon the description of patterns to model the requirements of the users while the exploration process.The authors presented, PATEX, a system devised to deal with RDF(S)/OWL datasets using the two proposed search strategies and allowing the user to switch between them interactively. The proposed approach combines theme-based and keyword exploration in the RDF datasets and is proved to reduce the user's effort to browse and explore RDF(S)/OWL dataset without any additional knowledge about the datasets [46].

Querying incomplete information in RDF with SPARQL: C. Nikolaou and M. Koubarakis proposed for enhancing RDF with the functionality for representation of property values that exist but are unknown or partially known through constraints. The authors proposed RDFi, a framework that supplements RDF with the functionality to deal with incomplete information using a first-order constraint representation language. This can prove to be crucial to be able to make effective, meaningful and efficient decisions in the presence of complete information. The incompleteness in the information at the disposal can very well reflect in the decisions ,hence affecting performance [47].

Theme Identification in RDF Graphs: H. Ouksili et al. presented an approach that provides a theme-based view of a given RDF dataset and makes it easier to capture the relevant and desired resources and properties. The author's work combines existing clustering techniques along with some semantic requirements set by the user to identify the desired themes. The proposed approach composes of three steps: capturing user requirements, apply a clustering algorithm to capture and identify the required theme and extract labels or nodes describing each of them [48].

Querying RDF Data from a Graph Database Perspective: R. Angles and C. Gutierrez demonstrated the RDF model from a perspective of a database and compared with other database models, in particular with graph database.The RDF technology from the perspective of graph database modelling is explored by the authors. The authors concluded that the RDF query language should incorporate and integrate with graph database query language primitives so as to achieve better and efficient results. Also, onec the ontology is mapped onto a graphical database, it opens a whole lot of possibilities to be applied onto the represented knowledge in the form of graphical techniques like searching, clustering etc.[49].

Chapter 3

PROBLEM FORMULATION

This chapter includes the research gaps identified based upon the literature review performed in the initial phase of the study. The gaps are identified in terms of dealing with data heterogeneity, efficient handling & interoperability in Knowledge-based Systems and shortcomings in sensor data analysis using existing systems. Considering these gaps four research objectives are identified for developing Ontological Knowledge models for heterogeneous sensor data. Further, querying service is deployed on these models for knowledge engineering and resultant decision making.

3.1 Research Gaps

Knowledge-Based Systems have existed and are being used for a long time now. But, with growing data and knowledge arises a need to enhance the already existing Knowledge Bases in order to provide more efficient data storage, management and data retrieval mechanisms. Ontology-based knowledge systems can also be subjected to enhancement procedures so as to ensure and experience efficiency in data management and knowledge mining techniques like never before. Keeping in mind the scalable aspect of data and future application based requirements, many research gaps exist, which are as follows:

- There is a need for efficient handling of heterogeneous sensor data coming from varying data sources. Heterogeneous data from sensor networks has high potential which can help in capturing useful information.

- There is a need for efficient handling of streaming sensor data which has temporal aspects associated with it in order to exploit the dynamic nature of data to gain deeper insights.

- There are issues in interoperability in data from different sources due to different syntactic attributes and formats.

- Efficient pattern matching is an important task in order to extract extra information from the RDF Graph to gain knowledge and provide to the user while querying the RDF repository. Efficient pattern matching algorithm need to be devised for desired results.

- It is a critical to achieve workload balancing among peers in case of distributed knowledge-based systems on ontologies.

- There is a need to efficiently transform sensed data into RDF form in order to associate it with domain ontology (e.g. SSN ontology) through a mapping language.

- Merging of multiple ontologies into a single one can be very informative, since it can lead to integrated knowledge of different domains and concepts merged together to form a single entity. The merging process needs to be free of resulting consistencies and the integrity of the domain data and relationships needs to be maintained.

- There is a need to devise efficient techniques to achieve query optimization while querying the RDF data so as to get better results with much less complexity with respect to time. Also, to get much refined results in case of ambiguous user queries.

- There is scope of devising efficient result ranking of user queries fired at RDF-based database stores for effective and impressive knowledge sharing with the end users.

- Ontologies are often represented as a graph where nodes represent objects within given ontology and edges represent relations between those objects. Such graphs can be very complex even for medium sized ontologies so a clean and complete graphical representation is needed. Existing solutions do not include all elements or produce unreadable results for big ontologies.

3.2 Research Objectives

The Objectives of the research work are as under:

- To study various techniques and tools available for integration of heterogeneous sensor data and querying knowledge model.

- To design ontology(s) based knowledge model for integration of heterogeneous sensor data.

- To design and deploy querying service over the semantic repository to draw useful inferences.

- Testing and validation of the proposed framework using real-world sensor data against other existing approaches.

3.3 Research Methodology

The following methodology is followed to achieve the above stated objectives. Figure 3.1 shows the flow of research methodology adopted.

Research Methodology for Objective 1: Extensive survey of existing approaches of sensor data analytics is done to gain deep insights and better understanding of the domain. The bibliographic study is carried out considering the following aspects for sensor data analysis:

- Gained understanding of the basic concept of ontologies, usage, inference mechanism and their implementation. Also, understanding and working of various tools for the same is achieved.

- Gathered knowledge about the current sensor data capturing sensor networks, the existing techniques for sensor data cleaning, filtering and processing and studying the sensor data.

- Reviewed the existing Knowledge-based systems and their real-world application to get a deeper understanding of the research trends.

- A Comprehensive review about designing of ontology-based knowledge model involving feature extraction of the heterogeneous sensor data as one of the tasks, is performed to gain better insights.

Research Methodology for Objective 2: Development of Ontological models for heterogeneous sensor data is performed to gain deeper insights into the domain knowledge. It is achieved as:

- Developed an ontological model for homogeneous data from heterogeneous sensors that models the task of Activity Recognition. The ontological knowledge model is implemented through OWL, RDF, RDFS and RDF4j framework and used Protege 5.2 as the ontology editor.

- Designed and developed a probabilistic ontological knowledge model targeting heterogeneous data from heterogeneous sensors to perform activity recognition. It is based on a combination of statistical (probability) and knowledge-driven (ontology) integrated approach. The data preprocessing is performed using pandas 1.4.0 python library and OID_v4_Toolkit. For ontology development, Java is used in OWL framework. For adding the probabilistic enrichment, pellet-based ontological reasoner, Pronto, is deployed.

- Proposed a modular ontology-based reasoning approach that targets fusion of heterogeneous digital ECG recordings from sensors to detect and classify heartbeat data into either of the cardiac arrhythmia classes. A SWRL based ontology classifier performs the task of final classification.

- Proposed an interval type-2 fuzzy logic and fuzzy ontology–based knowledge model that takes in input the heterogeneous sensor data to detect water quality. The fuzzy partitioning, fuzzy sets, fuzzy rules and fuzzy inference are implemented in an integrated process with the ontological approach and methodology.

Research Methodology for Objective 3:

- Semantic SPARQL-based query service is deployed over each of the proposed ontological knowledge-based systems. Identifying patterns in the ontology-based knowledge model in the form of RDF data (triples) stored in RDF repositories.

- The deployed querying service invoked pattern oriented ontological knowledge base exploration. The reasoner/inference engine is activated to work on the axioms/constraints/rules to draw useful knowledge.

Research Methodology for Objective 4:

- Validation of the proposed approaches is done by instantiation with real-life heterogeneous sensor data.

- The performance of the proposed models is evaluated by firing queries to fetch semantic knowledge and then evaluation in terms of true positives (TP), false positives (FP), false negatives (FN) and false positives (FP), precision, accuracy and F1 scores.

- The performance of the proposed systems is compared with conventional existing approaches in terms of performance metrics.

- The scalability, consistency and flexibility is established by instantiation of the proposed systems with newly added knowledge and carry out updations frequently. Also, varying size of data is used for the instantiation process.

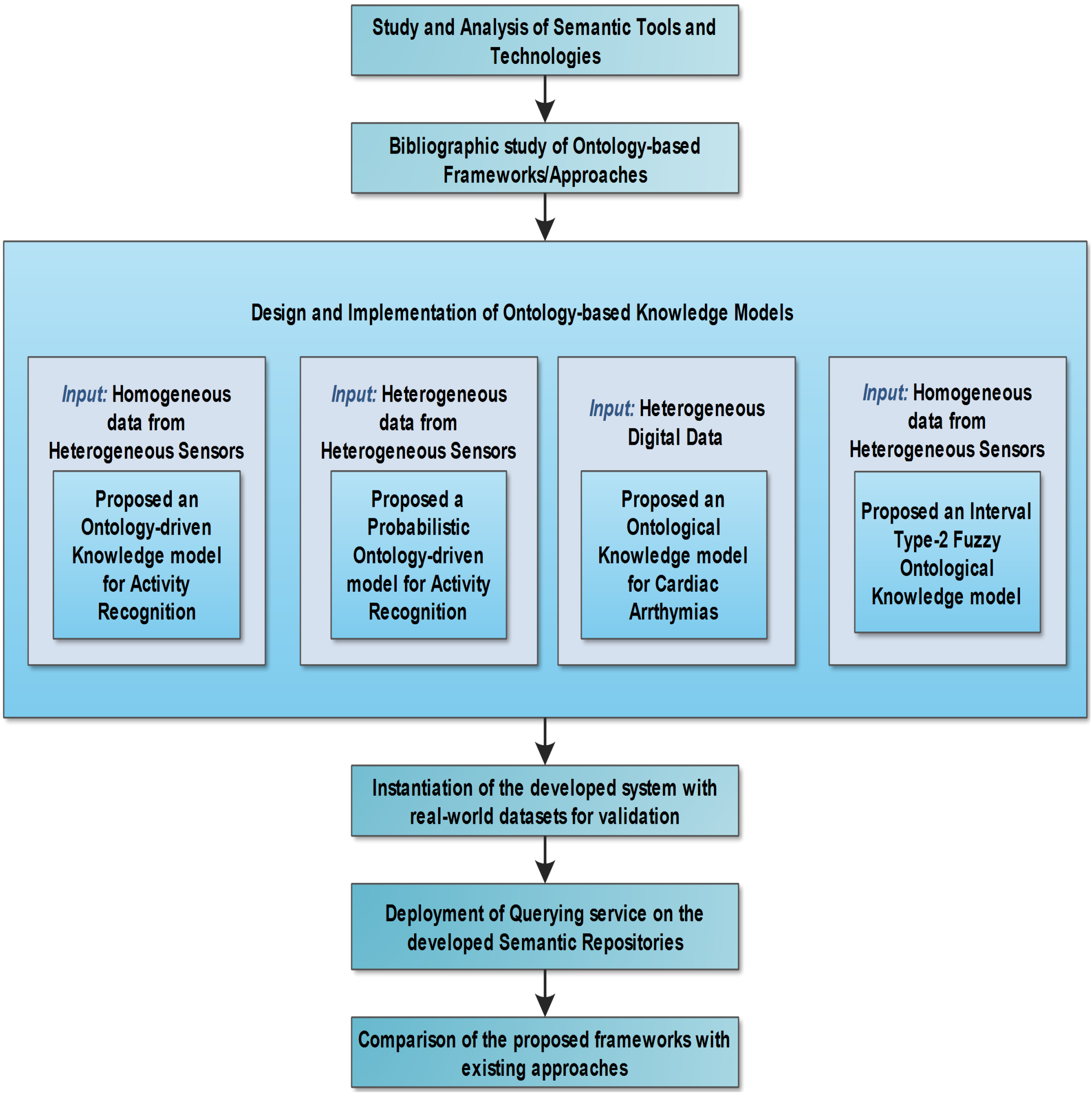


Figure 3.1: Research Methodology

Chapter 4

ONTOLOGICAL MODEL FOR HETEROGENEOUS SENSOR DATA

The objective of this chapter is to perform semantic reasoning on heterogeneous sensor data through the development of Ontological Knowledge Base over timestamped observations. Based upon the domain knowledge represented, the proposed model is instantiated with a real world dataset captured by sensors while daily living activities are being performed. The developed Ontological system deploys querying service and performs inference mechanism through a set of axioms defined to achieve human activity recognition. The resultant system performs efficient sensor data analysis and knowledge entailment.

4.1 Heterogeneity in Sensor Data

The devices with sensors integrated in them have been creating profusing data incessantly. Making this data available to the world to derive useful information from this data will engender new capabilities, unprecedented opportunities in various real-world applications and scenarios. This availability of sensor data in large amounts highlights the need to perform intelligent analysis on the gathered heterogeneous sensor data . Detecting relevant patterns in the sensor data is the key to deduce the knowledge within the data. But, this sensor data is abundant in formats that make it difficult to perform knowledge extraction through traditional data analytics. The multimodal and heterogeneous nature of the enormous sensor data highlights the need to handle it prior to subjecting it to analysis. The existing approaches are not able to perform intelligent analysis on sensors heterogeneous data efficiently.

4.2 Heterogeneous Sensor Data for Activity Recognition

In recent years, sensor-based activity recognition has integrated the field of sensor networks with data mining techniques to model a broad range of human activities and behaviour [50]. Huge amounts of data coming from smart gadgets such as smartphones and smartwatches opens up the possibility of probing and extracting knowledge from the data in the direction of monitoring and health care [51]. Though available smartwatch and smartphone context-aware application are not well equipped to manage raw sensor data and measurements. Even minor changes in conditions become difficult to accommodate and lead to entire application redesign. Further, to develop realistic functional applications, developers need actionable knowledge and information, which cannot be extracted from raw sensor observations [52,53]. Thus, management, representation, processing and analysis of sensory data collected from sensor based systems such as smartphones and smartwatches pose a big challenge. Therefore, exploiting state-of-the-art technologies for semantic representation and interpretation of sensors and observational data can be very helpful in context-aware and human activity recognition applications. The issue of heterogeneity in sensors generating voluminous data through varying measuring procedures and in different formats, makes it hard to retrieve information for users searching for relevant information. The inherent design, features of sensors and absence of adaptability in different conditions restrict the accuracy of the sensed data. The noisy and asynchronous samples collected by heterogeneous sensors result in missing data that restricts applications for performing efficient analysis. Further, improper and incomplete sensors specifications and annotations can hinder the inference mechanism on sensor data. All the above problems can be resolved by efficient integration of multi format heterogeneous sensor data. The sensor data integration can enable retrieval of knowledge that is difficult to infer through individual sensors. The potential residing in the booming sensor technology can be best exploited when there is availability of a well-developed common platform for expressing various aspects of sensors and sensing procedure [54]. Some standards have already been established for sensor devices and measurements for facilitating interoperability and thus, providing a common format for heterogeneity in sensor data repositories and applications [55]. However, these types of standards lack the aspect of semantic description to perform logical and reasoning tasks [56]. Semantic Web technologies can be used to add an additional semantic layer to achieve semantic compatibility and interoperability in terms of sensor devices and the observation data. In this respect, ontologies provide semantic annotations to sensory data through temporal, spatial, and thematic metadata and thus, boost semantic interoperability, understanding, and mapping of links and associations among different mismatching terms [7]. Semantic Annotation of sensory data paves way for the enhanced integration and interoperability of heterogeneous sensory data. It can be reused as contextual information for achieving situational awareness. The ontology creation can revolutionize smartphone and smartwatch context-aware applications through a broader data representation model with the potential to add fresh content and relationships. The ontology can segregate the application part of the knowledge from the operational part, hence, enabling applications to achieve easier knowledge management [57].

This chapter presents an ontology-based knowledge model developed for fusing heterogeneous sensor data. It is instantiated with sensed data while various activities are being performed to achieve human activity recognition. The knowledge model is based on two newly developed ontologies: Sensor

Measurements Ontology to model the heterogeneous sensor data and Activity Recognition Ontology to conceptualise the activity recognition process by capturing the relationships between the low level acts (simple activities) and high level (inferred activity). The resultant knowledge-based system solves the issue of sensor heterogeneity and provides a model that promotes reusability, interoperability and exchange.

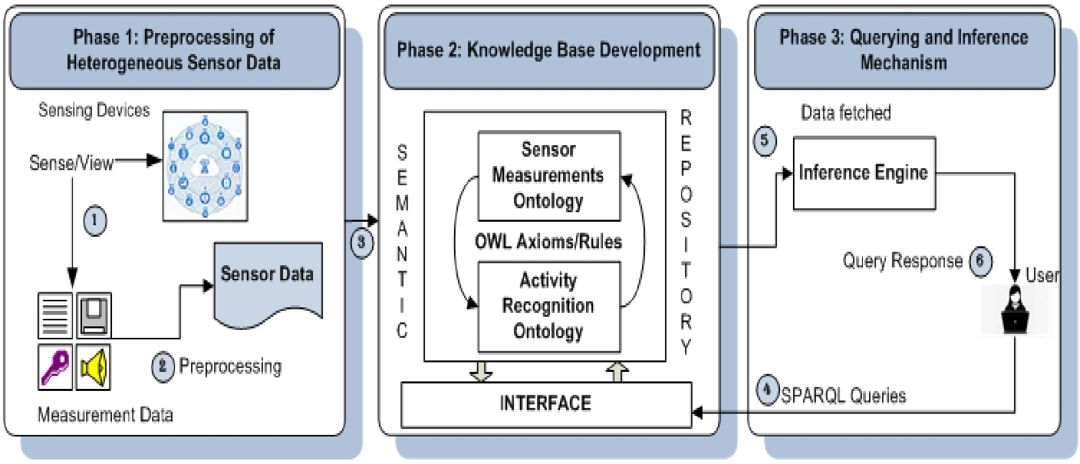


Figure 4.1: Architecture of Proposed System

4.3 Proposed Ontological Knowledge model

An ontology-based model for activity recognition is proposed which consists of three main components: Heterogeneous Sensor Observational Data, Knowledge Base Development and Data Querying & Inference for analysis and knowledge gathering. Figure 4.1 shows the architecture of the proposed model and Figure 4.2 depicts the overall working. The proposed model takes in processed heterogeneous sensor data as input, stores it and performs ontology-based analysis whenever a query is fired. It returns the required knowledge to the client or user (whoever fires the query) based upon the inference mechanism carried out. The result of the query service deployed is the type of activity that is being performed during a particular time interval or at that instant by the person under consideration. The three phases that constitute the activities performed in the proposed ontology-based recognition model are:

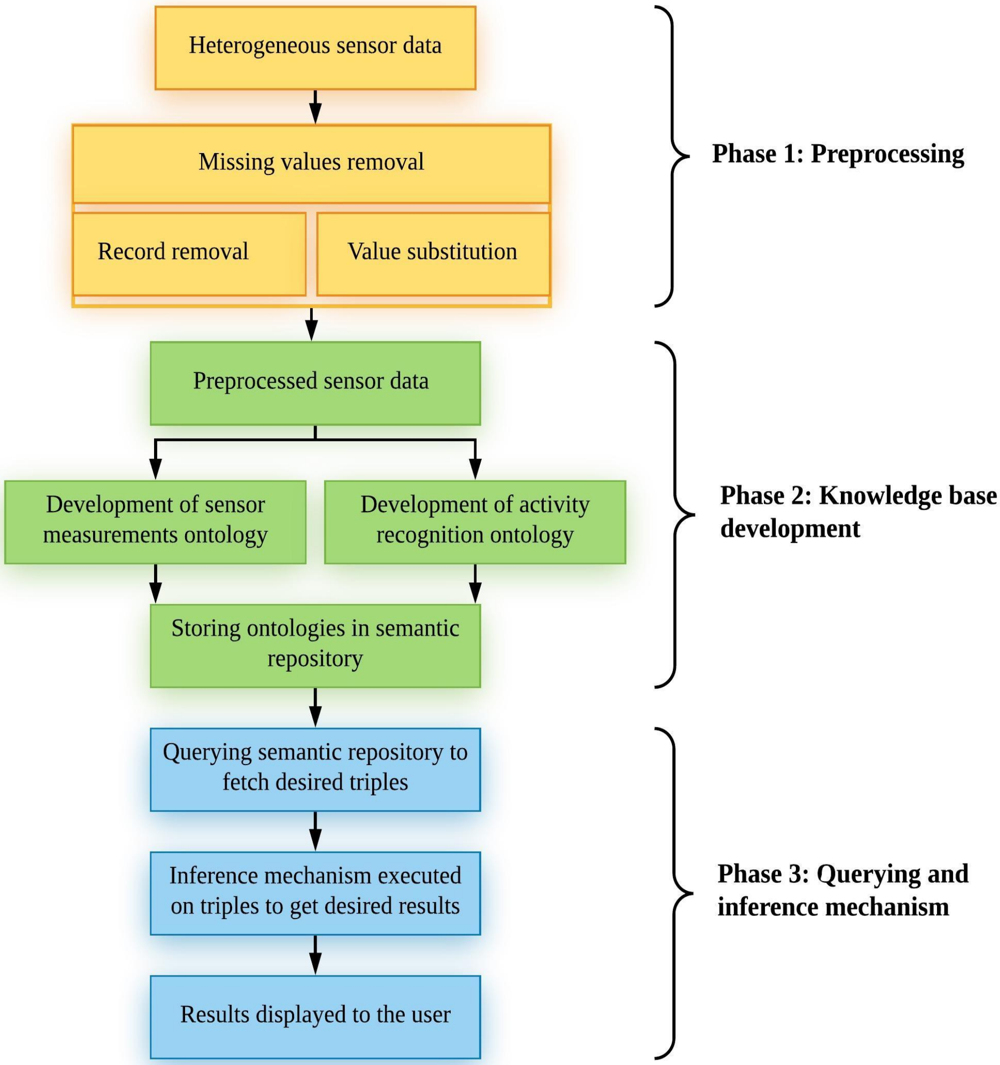


Figure 4.2: Work Flow of Proposed Model

4.3.1 Phase 1: Preprocessing of Heterogeneous Sensor Observational Data

During this phase, preprocessing of the heterogeneous sensor data is performed. The main issue addressed here is the problem of missing values in the data, which is dealt in two ways:

- **Record Removal:** The set of continuously captured sensor observations which contained missing values due to some sensor fault are removed so as to prevent the issue of memory overhead and unnecessary triples in the triple store which increases the data fetch time during inference tasks.

- **Substitution:** The observations which have missing values are replaced with values of the same sensor from the previous epoch or interval, based on the assumption that not much has changed in the next epoch, since the interval length is fixed and same for all the sensors.

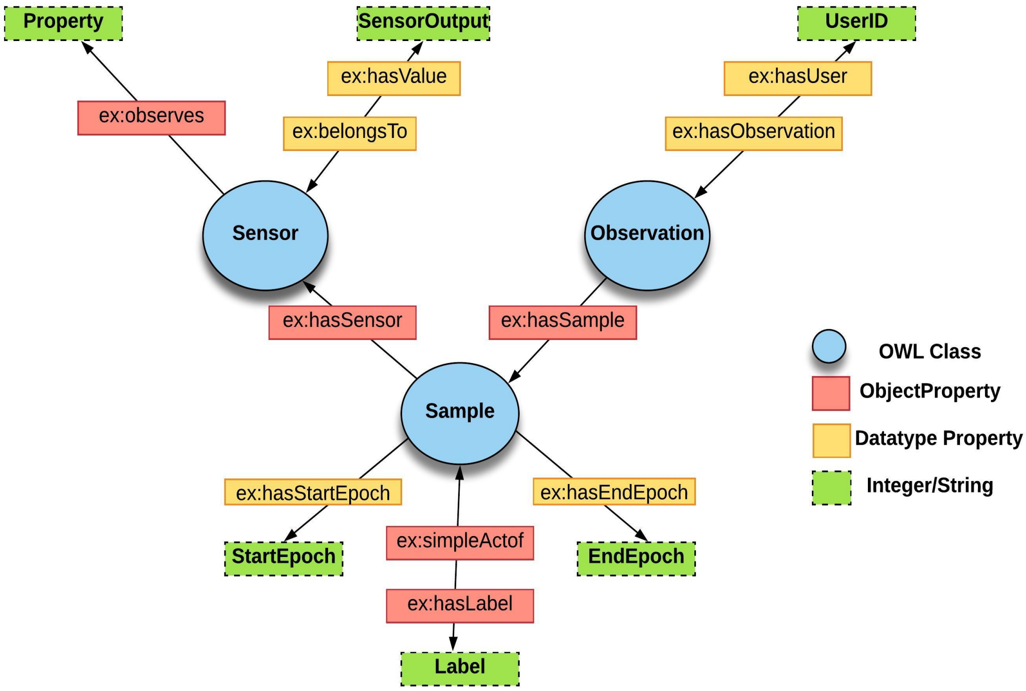


Figure 4.3: Sensor Measurements Ontology

4.3.2 Phase 2: Knowledge Base Development

The knowledge base development phase is subdivided into two main steps. The first step of this phase consists of ontology modelling of the sensor data obtained from the previous phase.

It consists of development of two ontologies which form the knowledge base for our model: Sensor Measurements Ontology and Activity Recognition Ontology. The second step deals with the storage of developed ontologies in a triple store.

Ontology Construction: In the proposed model, ontology development is

Table 4.1: Classes of Sensor Measurements Ontology

Classes	Description
ex:Observation	The class ex:Observation represents the sensor observations with respect to each person. This class can contain thousands of samples taken in pre defined intervals by different sensors.
ex:Sample	The entity ex:Sample models the class which has instances uniquely associated with a timestamp and have values captured by heterogeneous sensors. Each instance of ex:Sample has values captured by different sensors.
ex:Sensor	The entity ex:Sensor models a concept that senses a change in stimuli and provides observational values and data.
ex:SensorOutput	The entity ex:SensorOutput models the processed sensor outputs with respect to a particular instance of ex:Sensor class.
ex:Property	The entity ex:Property models the properties targeted by the class ex:Sensor. For example, Accelerometer can be a member of class ex:Sensor which observes property Acceleration. So, Acceleration is member of the class ex:Property.
ex:UserID	The entity ex:UserId denotes the class of persons or subjects associated with his/her ex:Observation, uniquely identified by an ID.

performed on the basis of the design criteria defined by[58]: clarity in conceptualisation of defined terms, coherence in the axioms over the defined conceptualisation, extendibility in terms of anticipating the shared usage of the developed vocabulary, minimal bias in encoding to ensure independency on any particular symbol-level encoding and minimal ontological commitment to support knowledge sharing. The knowledge base is developed with the timestamped heterogeneous sensor data in the sensor measurements ontology. The activity recognition ontology consists of semantically annotated domain knowledge along with a well defined set of axioms over the entities and their relationships.

Table 4.2: Properties in Sensor Measurements Ontology

Properties	Description
ex:hasSample	Links members of class ex:Observation to members of class ex:Sample
ex:hasUser	Links members of class ex:Observation to members of class:UserID
ex:hasObservation	Links Observations to the class ex:UserID
ex:hasSensor	Links members of ex:Sample to members of class ex:Sensor
ex:observes	Links members of class ex:Sensor to members of class ex:Property
ex:hasStartEpoch	Relationship between a literal timestamp value and class ex:Sample
ex:hasEndEpoch	Relationship between a literal timestamp value and class ex:Sample
ex:hasValue	Links the sensor with the sensor output value
ex:belongsTo	It is the reverse relationship of ex:hasValue
ex:simpleActof	Links the labels associated to ex:Sample
ex:hasLabel	Relationship between Labels or Simple acts and ex:Sample

The Step 1 is accomplished as follows:

Development of Sensor Measurements Ontology

Sensor Measurements Ontology is a temporal ontology consisting of statistically processed heterogeneous sensor observational data. It adds semantic enrichment and significance to the data by including metadata of the sensor data. The RDF4j framework is used for the implementation of this ontology.

The sensor measurements ontology constitute of a set of concepts represented as classes and relationships between them through links in the form of two types of properties: owl:ObjectProperty and owl:DatatypeProperty. It contains 6 core classes namely:

- ex:Observation
- ex:Sample,

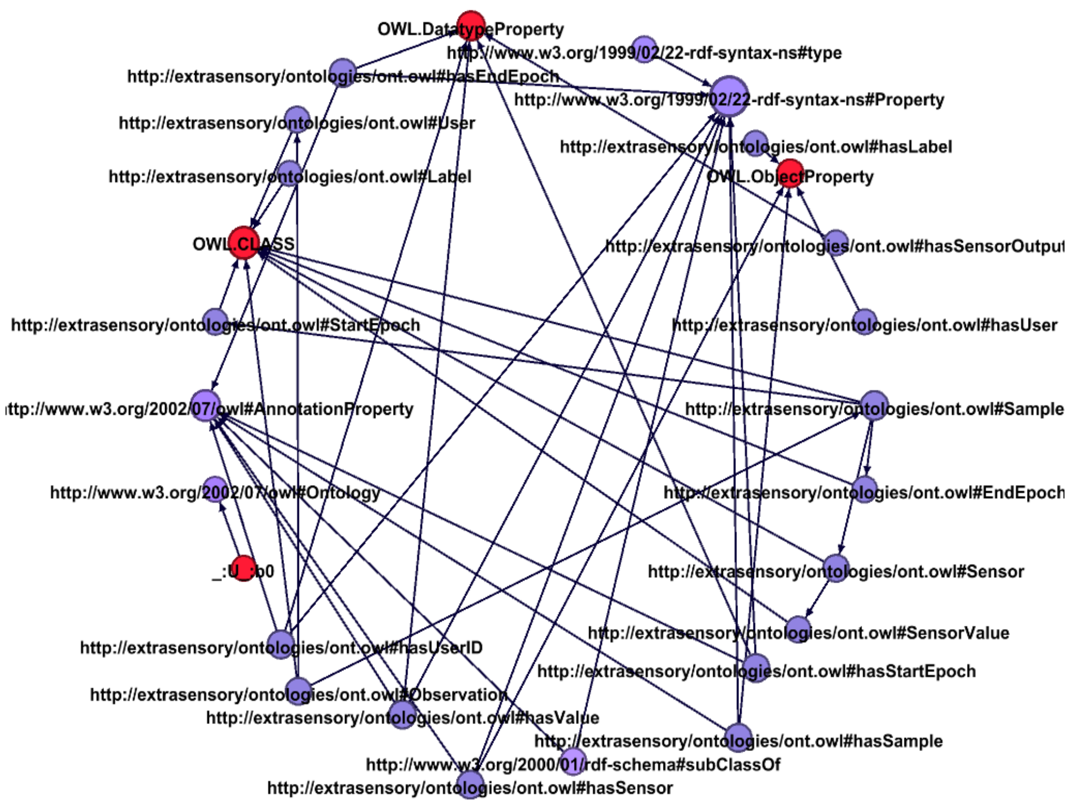


Figure 4.4: Structure of Sensor Measurements Ontology

- ex:Sensor,
- ex:SensorOutput,
- ex:Property and
- ex:UserID.

Figure 4.3 shows the graphical representation of sensor measurements ontology. Figure 4.4 represents the underlying skeleton of the above described sensor measurements ontology. The red colored nodes represent the core OWL classes or entities and the blue ones are the subclasses of the core entities. Algorithm 1 depicts the process of creating a basic skeleton and structure of sensor measurements ontology. Table 4.1 describes different entities or classes of sensor measurements ontology. Table 4.2 depicts relationships between the entities or classes of sensor measurements ontology. Algorithm 2 depicts the instantiation process of the Sensor Measurements Ontology. The W3C incubator group developed SSN (Semantic Sensor Network Ontology) which can subsume the concepts and properties of the developed Sensor Measurements Ontology. The ontology alignment can be defined using the Description Logics in the following manner:

- ex:Sensor $\equiv$ ssn:Sensor
- ex:SensorOutput $\equiv$ ssn:SensorOutput
- ex:Property $\equiv$ ssn:FeatureOfInterest
- ex:observes $\equiv$ ssn:hasObservableProperty

The symbol $\equiv$ denotes equivalence relation between the two entities and $\sqsubseteq$ symbol represents concept inclusion concept. If class A$\sqsubseteq$B, then the concept A is included in the concept B.

Algorithm 1 Creating skeleton of Sensor Measurements Ontology

Input: Set of m entities (E), n properties (P) and r relationship triples (R).
Output: Sensor measurements ontology file in turtle format.
Give format of output file with filename i.e. sensorMO.ttl
Instantiate ModelBuilder class
Give namespace for the ontology
for i $\leftarrow$ 1 to m **do**
    Add triple (E[i], rdfs.subclass, owl.class)
end for
for j $\leftarrow$ 1 to n **do**
    if object property **then**
        Add triple (P[j], rdfs.subclass, owl.objectproperty)
    end if
    if datatype property **then**
        Add triple (P[j], rdfs.subclass, owl.datatypeproperty)
    end if
end for
for k $\leftarrow$ 1 to r **do**
    Add a triple for each relationship b/w classes
end for

Algorithm 2 Instantiation of Sensor Measurements Ontology

```
Input: Set of 'f' heterogeneous sensor data files (F).
Output: Instantiated Sensor measurements ontology file with turtle format.
Create a structure of Sensor Measurements ontology // using Algorithm (1)
Enter path to sensor data files
Instantiate BufferedReader class with path+file name
for i ← 1 to f  do // Read all files one by one
    Add triple (F[i], rdf.type, Observation)
    Add triple (F[i], ex:hasUser, UserID)
    for j ← 1 to total samples  do
        // each observation/file has different set 'Tj'of samples
        Add triple ( T_j, rdf.type, Sample)
        for k ← 1 to total sensors  do
            // each sample has different set 'S'_k of samples
            Add triple ( S[k], rdf.type, Sensor)
            for l ← 1 to all observed outputs  do
                // each sensor has different observed o/p 'O'
                Add triple ( S[k], ex:observes, O_l)
                Add triple (O_l, ex:hasValue, V) // V is sensor output
            end for
        end for
    end for
end for
```

Development of Activity Recognition Ontology

The aim of the activity recognition ontology is to model the domain concepts and perform the reasoning mechanism. This ontology describes the concept of "Action". Action is associated with a person and the restriction that is imposed on the concept "Action" is that it must have a start time while finalization of the Action is not necessary. This is because an Action that is being performed started at a particular instant and is being performed at the current time. The ontology defines the concept Action to be of 2 types (subclasses): Simple Act and Inferred Act. The class "simple act" comprises of simple activities which can imply occurrence of an inferred act during a particular interval. The class "inferred act" is defined to have

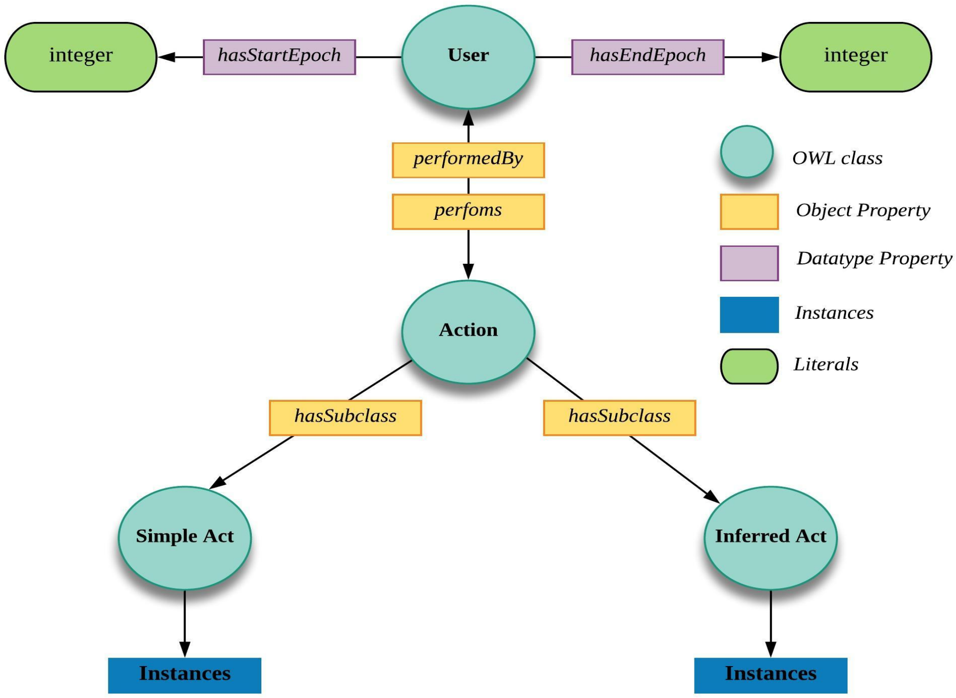


Figure 4.5: Activity Recognition Ontology

6 subclasses: Activity, Inactivity, Exercising, Home Activity, Outside Home and Indoor Activities. It can be comprised of more than these subclasses. Figure 4.5 represents the structure of Activity Recognition Ontology. Situations in the form of labels like lying down, sitting, running, cleaning, computer work etc are members of the class simple act. An inferred act can be determined from the presence of one or more simple acts. In some cases, even a single simple act is enough to infer the type of inferred activity. For example, sitting, a type of simple activity, classifies the action to be a member of the class Inactivity, a type of inferred act. Such axioms are defined in the ontology to classify simple actions (associated with sensor measurements) into high level acts or Inferred Act.

Ontology Storage

In this step, the constructed ontologies are stored in Triple Store for efficient storage and retrieval of data with maintaining data consistency and no loss. Triple Store is a database for persistence storage of the Sensor Measurements Ontology and Activity Recognition Ontology. The ontologies are stored as concepts, the properties and their instances in the form of triplets in Ontotext GraphDB. GraphDB fulfils the requirement of the database to be devised in the form of a triple store. The triple store database allows the reading and editing of the stored ontologies. Further, it facilitates storage of class instances of the ontology. It allows for adding of new instances and retrieving of old instances from the triple store repository.

4.3.3 Phase 3: Designing of Querying and Inference Mechanism

Once the ontologies are stored in the GraphDB, SPARQL[59] queries can be launched on the triple store to fetch results for further inference tasks. In order to connect to the repository, a connection needs to be established through the RDF4j framework. Multiple queries can be generated depending upon the input triples required by the inference engine to get the desired results in a particular use case scenario. Figure 4.6 illustrates the flowchart of querying service and inference mechanism. The query fired fetches the triples that match a pattern specified in the query. The result of the query acts as input data to the hermiT reasoned to perform inference. For example, a query is fired to the repository in the triple store, to get the startEpoch and endEpoch times of the user that performs an activity, in order to classify it into one of the subclasses of the class “inferred act”. The query also fetches all the simple activities that are being performed at that time interval. The aim is to assign the unclassified instance to a high level or inferred activity. This is achieved by activating the reasoner to act upon the defined axioms and rules over the classes and the relationships. The query can also be customised to check if a particular unclassified instance is being continued in the next epoch with the same set of simple acts.

4.4 Experimental Analysis

4.4.1 Dataset

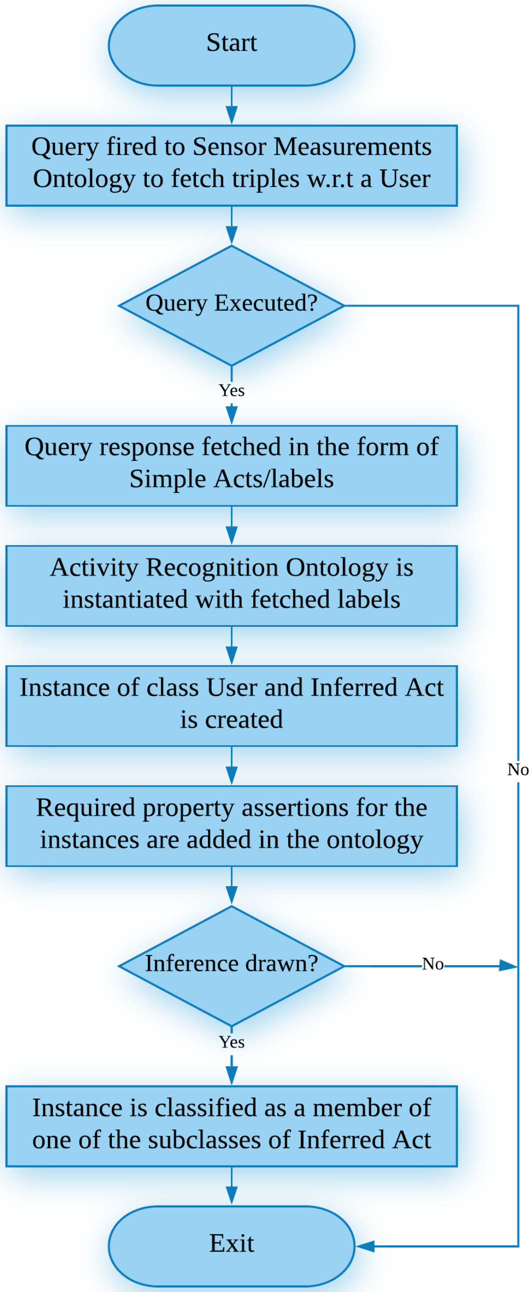


Figure 4.6: Flowchart of Querying and Inference Mechanism

The extrasensory dataset contains observations of multimodal sensors embedded in smartphones and smartwatches[60,61]. The extrasensory dataset contains sensory data extracted from 60 users who were engaged in their day to day activities and exhibiting their natural behaviour.

Table 4.3: Contextual labels and samples in the dataset

Simple Acts/Labels	No. of Samples
OR_indoors	184692
LOC_home	152892
SITTING	136356
PHONE_ON_TABLE	115037
LYING_DOWN	104210
SLEEPING	83055
AT_SCHOOL	42331

Each user is uniquely identified with an identifier value called as UUID (Universally Unique Identifier). Each user has large number of measurements taken by each sensor and extracted every 20 seconds. Each sensor measurement contains low level labels self reported by the user, except some missing values. The age range of the user under consideration is 18-42 with approximately 24 years as the mean age with 36 females and 24 males.

Table 4.3 shows few contextual labels and the corresponding users involved in those contexts. Also, it shows the number of examples or observations associate with a particular contextual label.

4.4.2 Instantiation with input Heterogeneous Sensor data

The ontology-based analysis and inference classifies the sensor measurements into Activities like Home Activities, Exercising, Outside Home, Indoor activities with respect to the labels or simple acts assigned to a particular sensor measurement. The implementation required construction of two ontology: one for observational data and another for conceptual knowledge. The ontology is implemented through OWL (Ontology Web Language), RDF, RDFS and RDF4j framework and

Table 4.4: Tools/Technologies used

Proposed Component	Tools/Technologies used
Knowledge Base	RDF4j, OWL, RDF, RDFS, Protege
Ontology Storage	GraphDB
Inference Mechanism	HermiT Reasoner
Querying	SPARQL

used Protege 5.2[62] as the ontology editor along with the required plugins activated. The hermiT reasoner is used for the inference mechanism and Ontotext GraphDB for storage of constructed ontology in the form of triples. A local repository in the GraphDB is dedicated to storage of the constructed ontology. Table 4.4 shows the tools and technologies used to implement the proposed model.

```
PREFIX ex: <http://extrasensory/ontologies/ont.owl#>
SELECT ?simpleAct
WHERE {
    ?a ex:haslabel ?simpleAct .
    ?a ex:hasStartEpoch "1444146916" .
    ?a ex:hasUser "00EABED2-271D-49D8-B599-1D4A09240601" .
} LIMIT 100
```

Figure 4.7: SPARQL Query

Lets's say, we wish to find out the inferred activity with respect to a certain set of simple activities performed by a particular user uniquely identified by the UIID "00EABED2-271D-49D8-B599-1D4A09240601" at a particular instant or epoch with value "1444146916". The following is the sequence of steps involved to draw the required inferences from the knowledge base:

- **Querying service deployment and fetching data:** The SPARQL query needs to match the required simple activities in the form of labels with the given

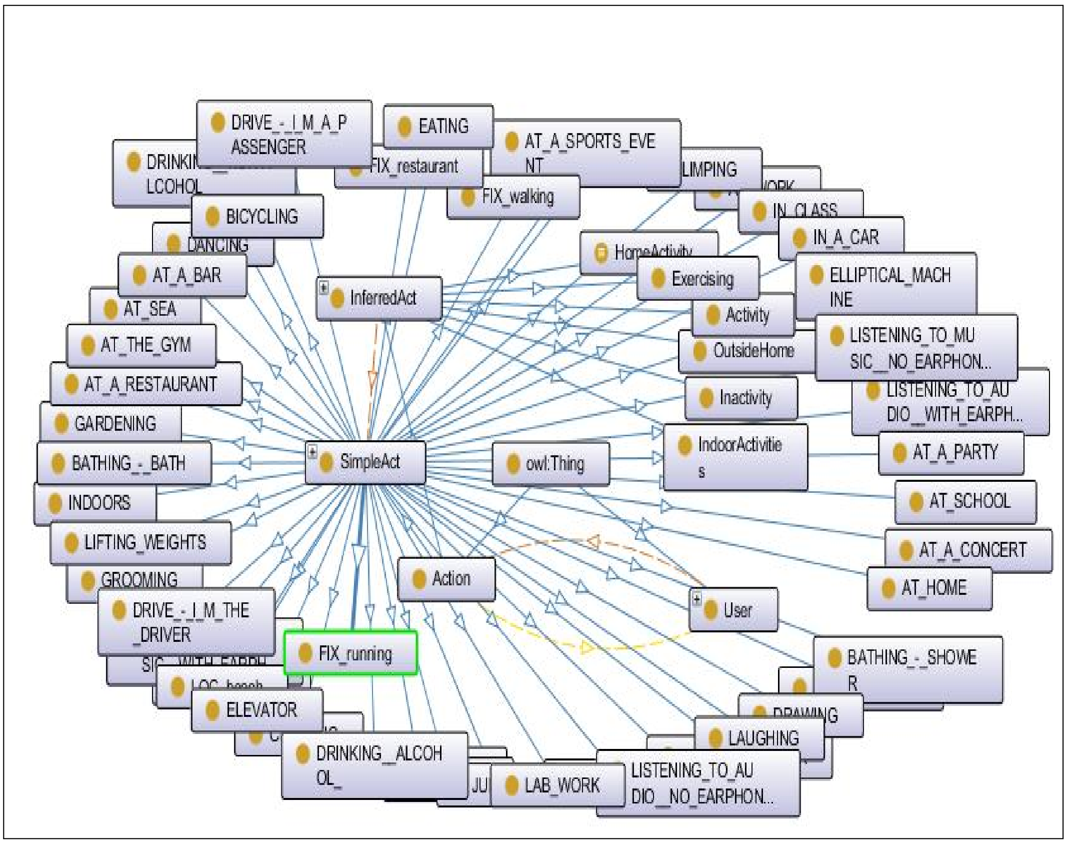


Figure 4.8: Activity Recognition Ontology Instantiation

timestamp/epoch and the user performing the activities. The SPARQL query to fetch this information is given in Figure 4.7. The specified SPARQL query requests the local repository created in the triple store to open a connection and fire the query to retrieve results corresponding to User ID 00EABED2-271D-49D8-B599-1D4A09240601 and timestamp 1444146916. The query performs pattern search in the set of triples and fetches results in the form of labels or simple activities. The result of the query depicts that the person is labelled with Sitting, IN_CLASS, PHONE_ON_TABLE at that specific epoch. So, the obtained result of the query represents the simple acts carried out by the user. After result of the query is obtained, the connection to the repository is closed to assure no data leakage or loss.

- **Instantiate the Ontology:** After, the query results are fetched from the Sensor Measurements Ontology and instantiation of the Activity Recognition Ontology is performed as shown in Figure 4.8. Firstly, the

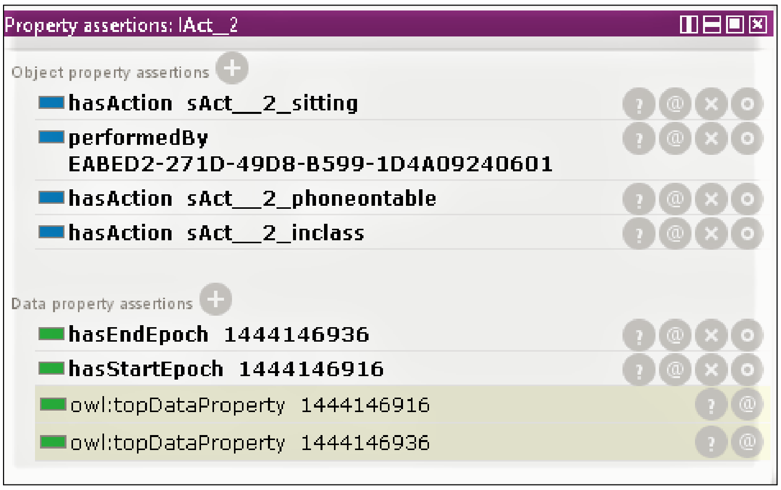


Figure 4.9: Property Assertions on Inferred Act instance

class simple act is instantiated corresponding to the labels returned from the SPAQRL query. The instances created are: sAct__2_sitting, sAct__2_inclass, sAct__2_phoneontable. The instance of the class User is created as 00EABED2-271D-49D8-B599-1D4A09240601. Also, inferred act is instantiated with its member IAct__2 The assertions that we made

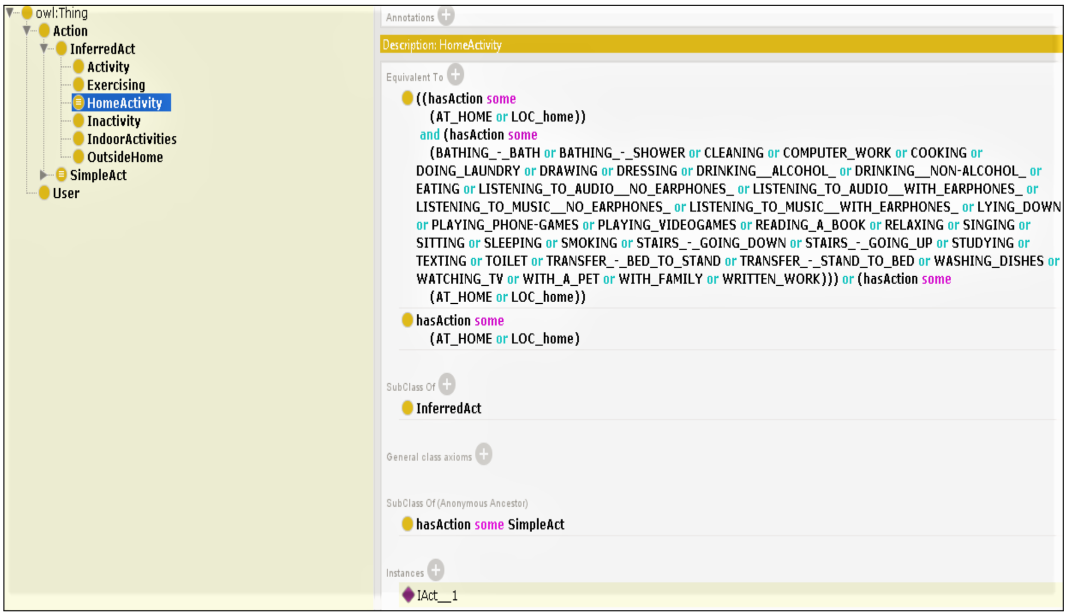


Figure 4.10: Inference Mechanism

are: performedBy(00EABED2-271D-49D8-B599-1D4A09240601, IAct__2), hasAction(IAct__2,sAct__2_sitting), hasAction(IAct__2,sAct__2_ inclass), hasAction(IAct__2,sAct__2_phoneontable), hasStartEpoch(IAct__2,144146916) and hasEndEpoch(IAct__2, 144146936).

Figure 4.9 shows the property assertions made in protégé tool.

- **Inference Mechanism:** After the ontology is instantiated and the required property assertions are made as per the universal and existential restrictions defined on the ontology, inference engine is activated to draw useful inferences. The reasoner is activated giving the set of assertions as input to generate the inference as shown in figure 4.10. The reasoner upon performing the reasoning task classifies the instance IAct__2 as the member of the class Outside Home. Figure 4.11 shows the inference result of the reasoner.

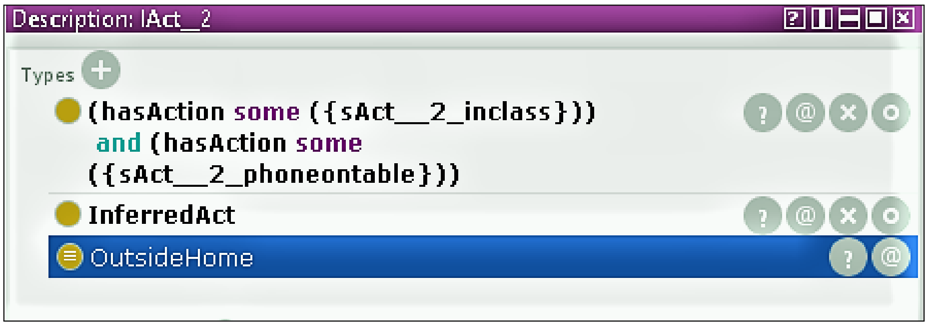


Figure 4.11: Classification of IAct__2 as outside home

4.4.3 Experimental Setup

The parameters setting of the proposed approach is tabulated in table 4.1. The parameter values of other competing algorithms are set as they are recommended in their original papers. The experimentation has been done on Matlab R2014a (8.3.0.532) version in the environment of Microsoft Windows 8.1 using 64 bit Core i-5 processor with 2.40 GHz and 4 GB main memory. The performance of proposed framework is evaluated in terms of classification accuracy. The proposed approach utilizes 30 independent runs in which each run employs 1000 iterations.

Chapter 5

PROBABILISTIC ONTOLOGICAL MODEL FOR HETEROGENEOUS DATA OF HETEROGENEOUS SENSORS

Here, Probabilistic Ontological model is developed to efficiently analyze multi-modal/heterogeneous sensor data from heterogeneous sensors. This kind of sensed data coming from disparate sources have varying output formats. Heterogeneous data from physiological, video/camera and plug sensors is semantically integrated and annotated for knowledge extraction. In this chapter, the proposed model aims to combine the ontological probabilistic knowledge along with the knowledge produced by a multi-class classifier to detect particular objects in image data captured by video sensors. Sensed data from other sensors is fused with this to result in Ontological Knowledge Base of heterogeneous sensed data. This integration not only provides and identifies existing as well as new relationships among the different segments, but also improves the classification precision and accuracy. This work boosts the performance for representing knowledge of the real world, expressed as probabilities of a set of relations among the objects and various other concepts. The developed model deals with the issue of heterogeneity in data and interoperability for knowledge reuse and exchange. The results fetched by various queries designed and launched showcase the proposed model outperforms existing traditional data-driven approaches in terms of precision and accuracy.

5.1 Schematic Heterogeneity in sensor data from different sensors

In recent years, the sensor technology has received more and more attention from the research community due to its underlying potential. It has been deployed in various kinds of application scenarios and environments for distinct purposes but valuable outcomes. Therefore, huge volumes of sensor data is being captured continuously by devices being part of different sensor networks all over the world. However, this sensed data lacks any semantics associated with it and can be heterogeneous at syntax, schema and semantic level. Schematic heterogeneity is the kind of issue that arises as a result of varying sensors capturing data in varying formats. Hence, making the process of sharing and integrating this sensed data a lot more tough. In order to overcome the absence of semantics in the sensor network and to be able to effectively manage this wide spectrum of heterogeneous data, semantic web technology is inculcated to achieve the following tasks:

- Meaningful description of sensor objects, information exchange, sensed data integration and knowledge driven reasoning.
- Semantic annotation of heterogeneous sensor information to improve interoperability.
- Addition of contextual information in terms of temporal and spatial relationships in the sensed data.
- Integration of multi sensor heterogeneous data to further support intelligent analysis.
- Issue of Schematic heterogeneity is resolved to represent sensed data uniformly in a machine as well as human readable format.
- Description, reuse, exchange and sharing capability in context of sensor data

is greatly improved.

5.2 Activity Recognition from Schematically Heterogeneous sensor data

The World Health Organization has estimated that there will be 2 billion people of age 60 and above by 2050[63]. Elders who are dependent and vulnerable in different aspects due to cognitive and physical impairment require assistance in their daily activities, which is mostly provided by caregivers. However, caregivers' assistance is becoming scarce and caregivers become overburdened with continuous monitoring responsibilities with the increase of elderly population. Also, most elderly people prefer to remain independent rather than to be moved in to a home care[64]. Therefore, it is imperative that technologies and services which allow the elders to live independently and happily be developed. One crucial feature of elderly healthcare system is to monitor the activities of daily living, especially when elderly are not accompanied by caregivers[65]. These activities include housekeeping, leisure, etc. Activity recognition involves the process of fusing and classifying contextual information acquired from sensors attached to inhabitant and environment. Numerous fusion and classification techniques have been proposed for activity recognition which can be classified into learning-based and specification-based techniques. Learning-based techniques create activity models from existing activity datasets that may contain noise and determine their correlation with the activity classes[66]. However, human activities are performed in various ways, and even one activity could be carried out in different sequences with difference durations. Therefore, statically trained activity model could not be applied from one person to another. Furthermore, the learning process is time consuming and error-prone because data annotation must be done manually. In the field of healthcare, sensor-based monitoring devices are being actively and widely adopted in modern healthcare settings to support administrative tasks, management of patient-related healthcare data and monitoring [67,68,69]. Today, caregivers are in direct contact with these technologies, increasing the complexity of their day-to-day activities [70]. The caregiver becomes obligated to use such devices to manually refer to, feed in and combine data, even while performing a simple single task. This becomes a very time-consuming and cumbersome process. Due to this improper integration of the technology and the huge amount of patient-related data being generated by the sensor devices, it is quite common for crucial and impactful events to be missed, e.g., prior indication of worsening of the condition of a patient etc. To deal with this issue and complications as a result of this, activity recognition need to be deployed to exploit the health related information made available by the sensor embedded devices to enhance the quality of continuous care of the patient and provide overall health care in a personalized way [71].

5.3 Background and Preliminaries

The feeding in of contextual data about the current situation in an ontology, intelligent algorithms can be developed and deployed that can take advantage of the semantically annotated information, optimizing and personalizing the context-aware applications and their performance.

5.3.1 E-Health ontologies

The Web Ontology Language (OWL) [72] is a standardized language used widely for encoding of ontologies. Since OWL can be mapped into Description Logics (DLs) [73], a family of logics representing decidable fragments of first order logic, the modelled knowledge and data can be logically derived and proved. Additionally, inference of new knowledge as a result of the correlation within the represented data. In order to achieve this reasoning process, reasoners are deployed. Thus, ontologies are able to effectively segregate the domain knowledge, which is agreed upon by researchers all over and can be re-used across various disciplines and application areas, from the logic, which can be stacked as rules, constraints and axioms on top of the developed ontology. The applicability of ontologies in the medical domain is an open research topic, since it has been acknowledged that ontology-based knowledge systems can improve the efficiency and management of complex healthcare monitoring systems [74]. The already developed ontologies revolve around biomedical research to focus on defining medical terminology vividly [75], e.g., Galen Common Reference Model [76], the Foundational Model of Anatomy Ontology (FMA) [77] or the Gene Ontology [78]. But, there is still scope in the development of high-level ontologies, which can model dynamic processes and contextual knowledge that can be utilized across different continuous care applications and settings. However, for some specific subdomains aiming at continuous health care of the patient, ontologies have been developed and deployed e.g., ontologies for representing knowledge in homecare assistance [79], modelling context of the activity performed by the user and dealing with chronic illness management in a homecare settings [80].

5.3.2 Probabilistic prediction model

When the conditions affecting the process are not properly known beforehand, it is not suited best to deploy deterministic prediction model. This paves the way for applying the probabilistic model to deal with uncertainties in the process.

- **Probabilistic inference:** Probabilistic inference is divided into two types: Frequentist and Bayesian. Frequentist inference is a type of statistical methodology which produces results from sampled data based upon the frequency or proportion of data. So, frequentist defines probability as the frequency in which the event occurs from a long-term perspective. Bayesian inference defines probability as the degree of belief for the events. This belief is then verified and modified through the results of the subsequent attempts or data. This inference establishes a relationship between prior and posterior beliefs as new information is obtained.

5.4 Proposed Methodology

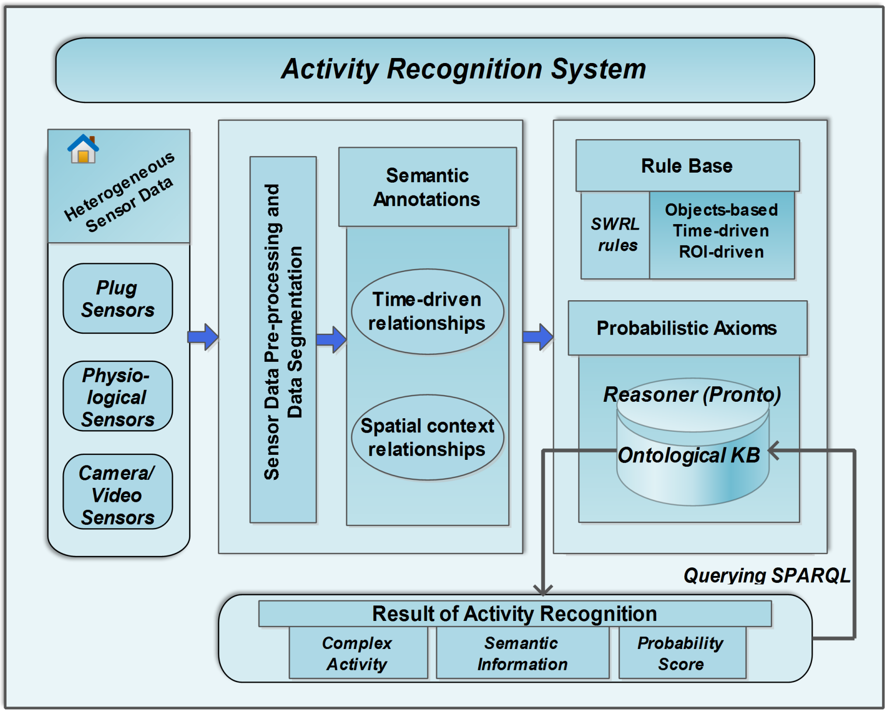


Figure 5.1: Architectural diagram of Proposed System

The proposed system is based on a probabilistic ontological approach to perform activity recognition from heterogeneous sensor-based data. It models the activity recognition process through a combination of data-driven (probability) and knowledge-driven (ontology) hybrid approach. The reasoning over the developed knowledge base formalizing the activity recognition task is based on probabilistic ontological entailment. Figure 5.1 represents the architecture of the proposed system. The domain knowledge that models the task of activity recognition is utilized to develop the knowledge base comprising of the terminology Box (TBox) as well as Assertion Box (ABox) statements in the ontology. The adopted approach involves processing of raw sensor generated heterogeneous data and perform probabilistic reasoning on the simple sub-activities for the detection of a complex activity being performed. It is based on a modular ontology design involving a spatial depiction associated with all the activities to be detected, and a temporal depiction annotated for each of the activities.

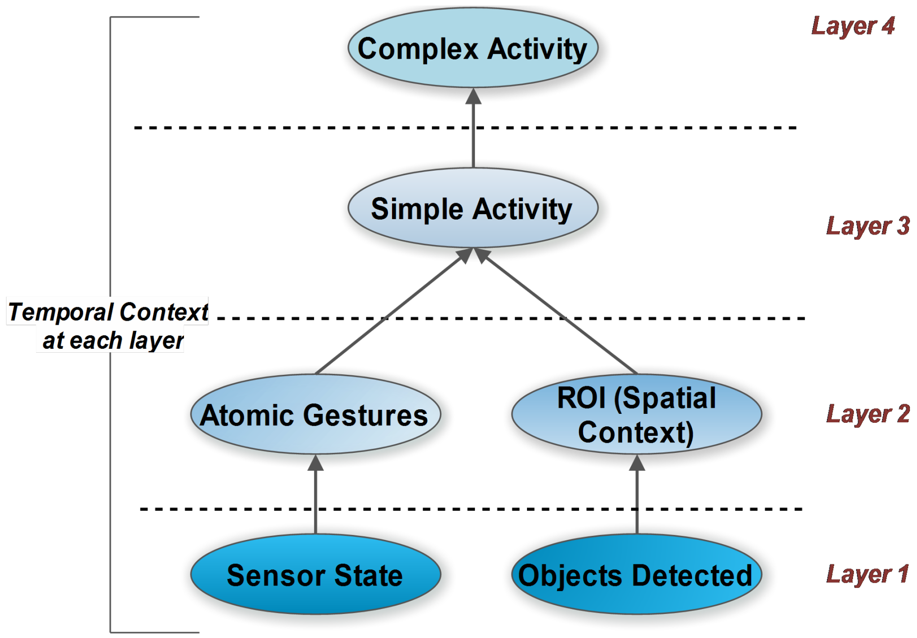


Figure 5.2: Layered structure of concepts of Activity Recognition Ontology

The role of adding probabilistic enrichment to the activity recognition model lies in the task of recognition of activities that requires probabilistic scores describing the regions of interest. These probabilities come from the object classifier that works on the association of each class of objects with the respective region of interest. The probabilistic identification of the regions of interest in the input image along with the spatial relations and temporal aspect further improves the recognition of complex activities. In a more formalised format, in every input image frame, a set S of regions of interest are identified based upon the set of 'C' of detected classes of objects. The Bayesian classifier assigns a probability distribution to each and every possible region of interest identified based upon domain ontological knowledge. This probability distribution is derived from all the possible object classes 'C' detected by the object classifier. Further, the identified region of interest with the class with the highest probability is selected for a particular image frame and time interval. This spatial and temporal relationship represents the baseline of the recognition model. Further, this information along with the domain knowledge is used to identify simple/complex activities being performed in the said time interval. The proposed system is implemented in a hierarchical structure composed of 3 logical modules. In the first module, a data-driven model for object recognition from sensor data is deployed. Secondly, identification of atomic gestures or activities is performed in the activity ontology including the temporal and spatial context of the gesture. Third and the final step is a rule-based probabilistic ontological inference carried out on the set of atomic activities recognised in the previous step.

5.4.1 Object Recognition from Sensor Data

The detection of objects that are crucial for the recognition of the activity being performed in a particular interval or instant provides more robustness to the model, in the case when the entire emphasis is on the activities of daily living. In this step, a multiclass object classifier is learned for a set of detection classes. The goal of this module is to determine a set of regions of interest corresponding to the objects recognised by the classifier. For each identified object category, the image classifier provides a numeric score, calculated in terms of probability. This score depicts the plausibility of the presence of that object in the image frame. Since, it is the data-driven step, the detection classes are carefully selected based upon the insightful analysis of the frames captured from video sequences. The preprocessing of the dataset is performed to make the data fit for training of the model. The Faster RCNN model is trained and applied for the process of object recognition. Algorithm 3 depicts the process of data preparation for Faster RCNN to be applied. The steps involved in the task of data preparation is performed using OIDV4_tootlkit library. The result of this transformation is the form of an annotation file containing the following columns:

- **imageID:** contains the label of the image

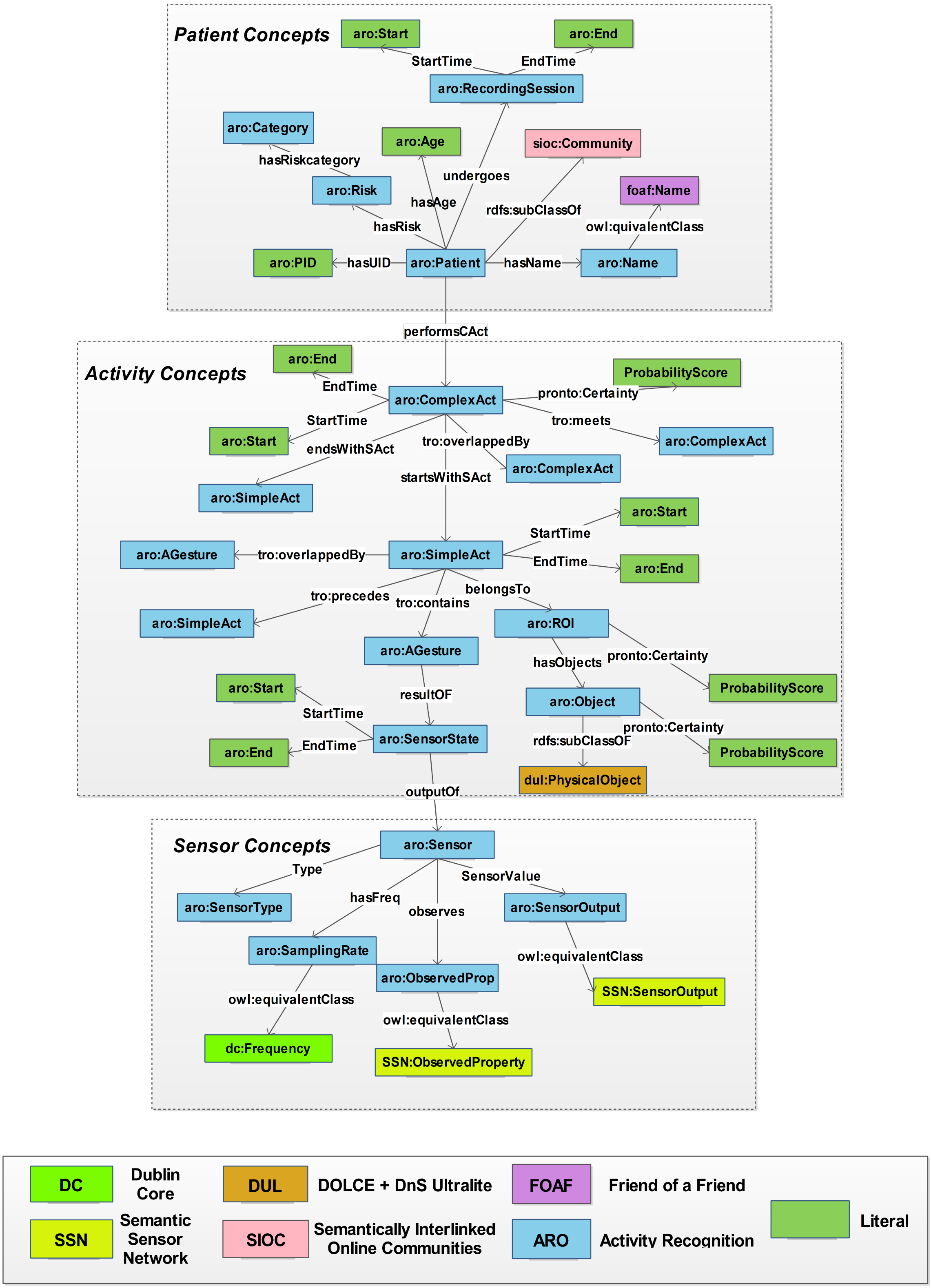


Figure 5.3: Graphical representation of Activity Recognition Ontology

Algorithm 3 Creating skeleton of the Activity Recognition Ontology.

Require: Set E of entities, P of properties and S of statements (S).

Ensure: Ontology ($Onto_{ARO}$) output file in manchester OWL format i.e., ARO.owl

Instantiate OWLOntologymanager class
Create and map IRI to ($Onto_{ARO}$)
Instantiate ManchesterOWLSyntaxOntologyFormat class and assign output format
for each $class_i \in E$ **do**
    Instantiate OWLClass with $class_i$
    Assign IRI 'arr'
    Add declaration axiom to '$Onto_{ARO}$'
    Add subclass/superclass axioms with x $\in$ E : x $\neq$ m
end for
for each $prop_i \in P$ **do**
    Assign IRI
    if $prop_i isObjectProperty$ **then**
        Instantiate OWLObjectProperty class
    end if
    if $prop_i isDataProperty$ **then**
        Instantiate OWLDataProperty class
        Define $datatype_{prop}$ expression
    end if
    Create axioms : $prop_i \sqsubseteq \rho$; $prop_i \equiv \gamma$; $prop_i \equiv \delta$, $\exists\ \rho$, $\exists\ \gamma$, $\exists\ \delta$, $\exists$P
    Add axiom rdfs:range($prop_i$, arr:x), $\exists$ x $\in$ E
    Add axiom rdfs:domain($prop_i$, arr:y), $\exists$ y $\in$ E
end for
for each $assertion_i \in S$ **do**
    Define DataType/ObjectType assertion i.e. ($prop_i$, $class_i$, $class_j$); $class_i \neq class_j$
    Add assertion axiom to '$Onto'_{ARO}$
end for
call saveOntology with args

- **Xmin:** x-coordinate of the bottom left part of the image
- **Xmax:** x-coordinate of the top right part of the image
- **Ymin:** y-coordinate of the bottom left part of the image
- **Ymax:** y-coordinate of the top right part of the image
- **className:** name of the object category in the image

Once the model is trained, it is run on the dataset to provide results as objects recognised in terms of probabilistic values.

5.4.2 Simple Activity Recognition

The core knowledge base of the proposed system is a newly developed ontology which models the task of activity recognition.The contextualised depictions made in the activity ontological knowledge base are based on the prior domain knowledge and the heterogeneous sensory contextual data. The developed activity ontology is organized into 4 different layers of context as depicted in figure 5.2. Each layer of the developed ontology is explicitly represented in the form of concepts upon which a set of axioms/rules are defined.

Figure 5.3 illustrates the taxonomy of the Activity Recognition Ontology segregated into 3 set of contextual concepts: Patient, Activity and Sensor . Algorithm 4 depicts the process of constructing the ontology. Algorithm 5 represents the process of instantiation of the developed ontology with sensor recordings of patients.
The atomic gestures, regions of interest and temporal context are the primitive concepts that describe the simple activity context, which in turn based upon domain and pattern-based knowledge gathered are used to classify a particular interval or instant into a complex activity concept. In order to achieve this, data segmentation is necessary.

Algorithm 4 Function to prepare data for Faster RCNN

Require: Open Image V4 data, class_Descriptions_boxable.csv, annotation_bbox.csv and set of ‘d’ Detection_classes (D) .

Ensure: Data in required format for Faster RCNN.

function Data_Prepare(Open_Image_Data, D)
Open class_Descriptions_boxable.csv in ‘r’ mode
Create $Img_{annotations}$ file
for each $DetectClass_i \in D$ **do**
    Extract ‘I’ image_labels and ‘C’ className from class_Descriptions.csv
    for each $classname_j \in$ ‘C' **do**
        Extract $record_k$
        Append $record_k$ to $Img_{annotations}$ file
    end for
end for
Add column ‘class_Name’to $Img_{annotations}$ file
for each $DetectClass_i \in D$ **do**
    Add $classname_i$ w.r.t image_labels to the column
end for
Extract imageID, Xmin, Xmax, Ymin, Ymax, ClassName from $Img_{annotations}$ file
for each $record_j \in Img_{annotations} file$ **do**
    Add image extension to imageID.
end for Save $Img_{annotations}$ file
end function

The sensor recordings are continuous and need to be segregated into data patches to enable complex activity recognition. The proposed system deploys an algorithm represented as Algorithm 6 that is based on both fixed interval and location driven segmentation. It segregates the data based upon fixed count of image frames and then updating it with the spatial(location-based) touch to generate patches of data allowing for efficient timestamped activity recognition. Also, thresholding is done in the sense that image frame count is kept under check. The output of this algorithm is further given to the proposed system to identify activities.

Algorithm 5 Activity Recognition Ontology Instantiation

Require: Set S of Sensor recordings of patients performing activities.

Ensure: Instantiated ontology.

Load ontology '$Onto_{ARO}$'with OWLOntologyManager //using Algorithm (4)
Instantiate BufferedReader with path+patient's recording
Instantiate DataFactory class
for each $patient_i \in P$ **do**
  Add classtype assertion (Patient, $[patient_i$)
  for each $Act_{Complex_j} \in C$ **do**
    Add axiom performsCAct($patient_i, Act_{Complex_j}$)
    Add annotation Certainty($Patient_i$, $Act_{Complex_j}$)
    Add axioms for start and end time.
    for each $Temporal_{Rel_k} \in T$ **do**
      Add $Temporal_{Rel_k}$ with $Act_{Complex_{j,..,n}}$
      for each $Act_{Simple_l} \in SA$ **do**
        Add start/end time axioms for Act_{Simple_l}
        Add axiom belongsTo(Act_{Simple_l}, ROI)
        Add startsWithSAct($Act_{Complex_j}, Act_{Simple_l}$)
        Add endsWithSAct($Act_{Complex_j}, Act_{Simple_l}$)
        Add time relations with $Act_{Simple_{1,2..,m}}$
        for each $AGesture_o \in AGest$ **do**
          Add axioms for start/end of $AGesture_o$
          Add time relations with $Act_{Simple_{1,2..,m}}$
          Add resultOF($AGesture_o$, Sensor_State)
          Add outputOF($Sensor\_State, Sensor_y$)
        end for
      end for
    end for
  end for
  call saveOntology
end for

A complex activity can be a combination of multiple simple activities or gestures associated with temporal as well a spatial context. A simple activity can be a combination of primitive and manipulative gestures. The rule-based inference mechanism is the process of activity recognition where an upper layer complex activity context is interpreted semantically by the simple activity layer context. The lowest layer depicts the sensor layer which models sensor-based atomic gestures and objects recognised in the visual frames.

Algorithm 6 Algorithm for Segmentation of Input Sensor Data

Require: Sensor based Image data (F), Set of Regions of Interest (ROI), Time duration for all ROI_i ,Image count Threshold value ($Icount_{thresh}$), Set of objects for each Obj_{ROI_i}.

Ensure: Segmented data patches for activity, $Data_k$.

Initialize D_k to null
Initialize T_{D_k} to null
Declare and initialize ($Icount_{thresh}$)
for each $Frame_i \in F$ **do**
  for each i=1 to $Icount_{thresh}$ **do**
    Append $Frame_i$ to $Data_k$
    Update $T_{D_k} = T_{D_k} + TimeStamp_{image_i}$
    Increment i by 1
    while i < $Icount_{thresh}$ **do**
      if $ROI_{i+1} == ROI_{i+2}$ **then**
        Append $Frame_{i+1}$ and $Frame_{i+2}$ to $Data_k$
        $T_{D_k} \leftarrow T_{D_k} + T_{image_{i+1}} + T_{image_{i+2}}$
        Increment i by 2
      end if
    end while
  end for
end for
return D_k with T_{D_k}

Algorithm 7 Function for Rules Generation for Activity Recognition Model

Require: Set of rules R.

Ensure: Ontology file ARO.owl with SWRL rules.

function CreateRule(OWLOntologyManager $\langle object \rangle$) //from Algo (4)
Begin
for each $Arule_i \in R$ **do**
  Initialize Set$\langle SWRLAtom \rangle antecedent_{rule}$
  Initialize Set$\langle SWRLAtom \rangle$ $consequent_{rule}$
  Initialize SWRL variables
  for each $class_k \in Arule_i$ **do**
    Create SWRLClass Atom
  end for
  for each $prop_p \in Arule_i$ **do**
    if $prop_p isObjectProperty$ **then**
      Create SWRLObjectProperty Atom
    end if
    if $prop_p isDataProperty$ **then**
      Create SWRLDataProperty Atom
    end if
  end for
  for each $atom_m \in Arule_{antecedent}$ **do**
    Add atom to antecedent $Arule_{antecedent}$
  end for
  for each $atom_m \in Arule_{consequent}$ **do**
    Add atom to consequent $Arule_{consequent}$
  end for
  for each $builtin_{swrl} \in Arule_{antecedent}$ **do**
    for each $args_n \in Argsbuiltin_{swrl}$ **do**
      List.add(argsn) //Add builtin rule atoms to a list
    end for
    Call Access_BuiltinAtom()
    Generate SWRLBuiltin Atom with IRI and args
    Add Atom to antecedent
  end for
  Initialize SWRLRule class
  Add rule $Arule_i$
  Call applyChange function
end for
Call saveOntology '$Onto_{ARO}$'
End

As, we go up the ontology structure, the recognised objects contribute to detection of a single or various regions of interest, which combined with some temporal context provides information on the person's simple physical activity. From this layer, the recognition process further goes to the complex layer, where combination of one or more simple activities, object-based region of interests and temporal conditions are propagated upwards to infer a complex activity being performed at that particular instant/interval. This process is represented in the form of three steps in figure 5.4.

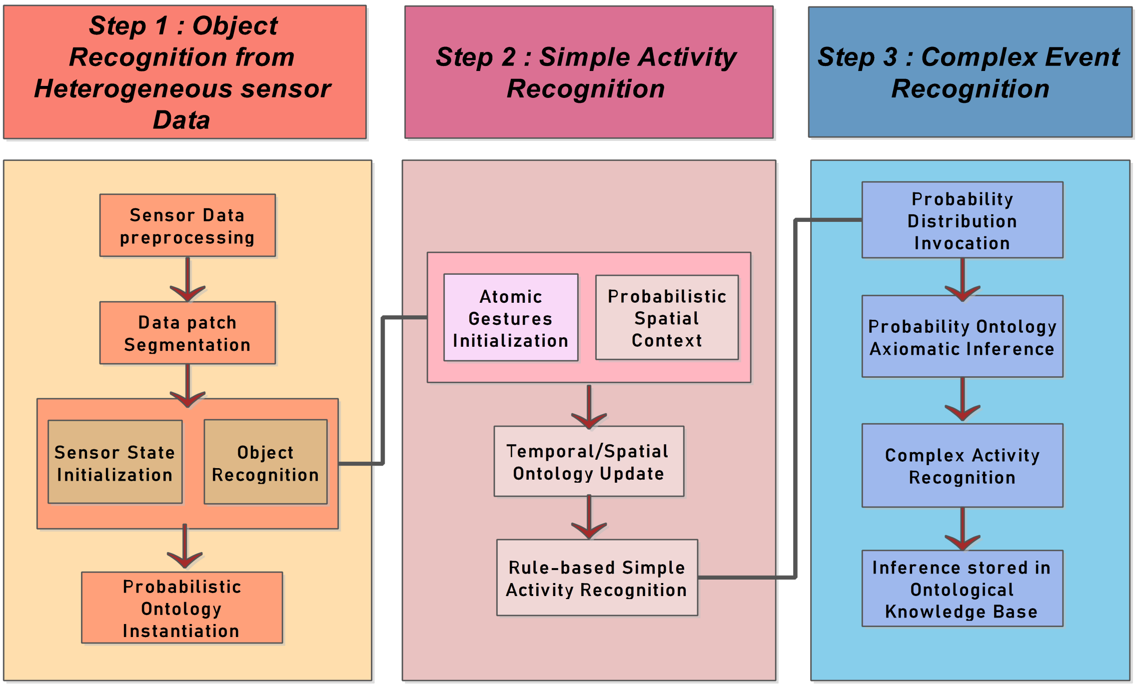


Figure 5.4: Flow of Activity Recognition process

For example, contact switches being installed on a electric kettle lid and Passive Infrared (PIR) sensors deployed in the region of interest i.e., kitchen can give contextual information on electric kettle being used due to its lid being opened or closed and the person under consideration is in the kitchen region. These simple gestures/activities are integrated based upon probabilistic rules/axiom to infer that the person using the appliance is in the kitchen area (simple activity). A complex activity like preparing tea is implied based upon a sequence of simple activities like taking bowl, opening electric kettle and closing electric kettle along with being in the region of interest i.e., kitchen for instance. The developed Activity Recognition Ontology comprises of six main contextual concepts: Sensors, Atomic_Gestures, Temporal, Region_of_Interest, Simple_Activity and Complex_Activity. The various layers of the ontological knowledge base are described in detail as follows:

- **Modelling Sensor-based gestures:** The sensor concept depicts the state of the sensor in the ontology. A sensor concept models a gesture based upon the its state. The sensor associated with an appliance like microwave can represent two gestures: MicrowaveDoor_ON and MicrowaveDoor_OFF.

- **REGIONS OF INTEREST (SPATIAL CONTEXT):** In order to provide spatial context to the Activity Recognition Ontology, spatial ontological relationships are added into it. In the ontological knowledge base, as in the previous modules, objects (e.g. bed, chair,table, cup, book, etc.) and rooms (e.g. bedroom, kitchen, living room, bathroom, etc.) are added as concepts. These concepts are further semantically annotated with their relationships. With these resources, Further, instances of these concepts i.e., objects belonging to a certain category (e.g. rdf:type(chair_12,chair)) and relationships assigning them to various other categories (e.g. belongs(chair_12,bedroom)) or (e.g. hasScore(chair_12, 0.78)). The object recognition module recognises a set of objects in an image frame. A score is assigned to each of the objects in that particular image frame, which implies the probability of that object being in the detected class. The identified set of objects, along with the assigned probability scores are used to identify regions of interest based upon Bayesian Inference[81]. In order to perform this region of interest identification, prior consensual domain as well as pattern-based knowledge is required, which assigns a set of objects to a particular region. The proposed system involves a probabilistic module based on a Bayesian framework[82] to perform the categorization of regions of interest (rooms) based upon on prior knowledge which is: the object classes detected from the previous module with confidence/probability values about their presence), and the domain knowledge of associating objects to their corresponding regions of interests. To exemplify, the proposed approach carries out analysis like: a microwave and a stove are identified and that too with a high value of confidence, so it implies that the region of interest is a kitchen area. The proposed distribution that gives out confidence/probability values from the possible set of regions of interest (room) classes (bathroom, bedroom, kitchen, livingroom, etc.) revolves around the following definitions:

(i) $|\theta|$ denotes the number of detected objects

(ii) *rc* is the number of different types of rooms

(iii) *oc* denotes the different types of object categories

(iv) K_i represents number of observations of $object_i$

To begin with, when there is no prior knowledge about the presence of set of objects, probability of an image frame classified into a region of interest (room) category is:

$$Pr(Room_{category} = X) = \frac{1}{rc} \tag{5.1}$$

In absence of any object detection, the above mentioned probability computation holds true. When any one or more unique objects are detected from the previous object classifier module, the probability distribution is transformed into probability mixture distribution based on Bayes Theorem [?] into the following equation:

$$Pr(Room_{category}|K_i) = \sum_{\theta}^{i=1} Pr(Room_{category}|obj_i) X P(obj_i) \tag{5.2}$$

$Pr(Room_{category}|obj_i)$ is the conditional probability denoting the input room belonging to a category based upon the object detections. The other probability $P(obj_i)$ is the confidence score of that object belonging to a particular detection class, given the object observations. Further, the equation below is based upon a mixture distribution and is simplified/defined by the domain knowledge encoded in the activity recognition ontology. This prior domain knowledge setting includes what classes of objects can appear in which of the regions of interest (room) categories. Hence, the probability distribution is given the following representation:

$$Pr(Room_{category}|obj_i) = \begin{cases} 0.95/|Room_c ategory_{obj_i}| & \exists obj_i \in rc \\ 0.05, & \text{otherwise} \end{cases} \tag{5.3}$$

This identification of each region of interest based upon the classifier scores of each of the classes of objects considered, provides spatial context and relationships between the other regions of interest as well as the simple activity that is being performed in that particular instant.

- **TEMPORAL CONTEXT:** For the task of activity recognition, the input sensor data is broken down into time windows or slots known as time intervals. This division is done based upon the time interval for which the person under consideration keeps interacting with his/her surrounding as depicted in Algorithm 6. In this context, a particular set of objects and the region of interest forms the surrounding environment for the concerned person. An time interval is defined as a concept of Activity Recognition Ontology. The concept "Interval" is further enhanced and semantically annotated by relations defined by Allen's temporal logic[83]. Figure 5.5 shows the pictorial representation of the relations in the ontology, inspired by Allen's Temporal Alzebra. The temporal relations defined in the activity recognition ontology add temporal enrichment and are represented as SWRL rules. These SWRL rules are activated to be a part of composite activity process. The temporal relations mentioned above are modelled for activity recognition as follows:

'Preceds'temporal relation:

$Interval(?a), Interval(?b), tro : starts(?a, ?sx), tro : starts(?b, ?sy),$

$tro : ends(?a, ?ex), tro : ends(?b, ?ey), swrlb : greaterThan(?sy,$

$?sx) \Rightarrow precedes(?a, ?b)$

'Overlaps'temporal relation:

$Interval(?a), Interval(?b), tro : starts(?a, ?sx), tro : starts(?b, ?sy),$

$tro : ends(?a, ?ex), tro : ends(?b, ?ey), swrlb : greaterThan(?sy,$

$?sx), swrlb : greaterThan(?ex, ?sy), swrlb : greaterThan(?ey,$

$?ex) \Rightarrow overlaps(?a, ?b)$

These relations modelled in the Activity Recognition Ontology, provide temporal enrichment to various gestures and activities being performed at different instances of time.

In order to perform the reasoning on the temporal relations, the ontology has a set SWRL rules which are based on the Allen algebra. The operations pictorially represented in the above figure allows for the relative temporal positioning of pair of time intervals. The time intervals are compared by their respective starting and ending time instants. Subsequently, various atomic gestures enriched with spatial and temporal contexts are combined based upon rules to form a conclusion about what simple activity is being performed by the person under consideration. Further, these temporal relation play a key role in the recognition of various types of complex activities.

- **PROBABILITY ONTOLOGY MODELLING**

The probabilistic-driven knowledge classifier combines the result of the data-driven object classifier with the domain knowledge represented in the probabilistic ontology that expresses spatial as well as temporal relationships between instances of objects and regions of interest in terms of probability scores.

The object probabilities are computed by the classifier which was trained with the set of pre-labelled images with object classes, and the spatial relations were constructed based upon the probabilistic object classification. The Bayesian classifier estimated the probability taking into prior domain knowledge and the frequency of occurrence of a set of objects with respect to a particular region of interest in the dataset under consideration. The region of interest detected, associated with some atomic gestures and temporal aspect can be comprehended as the basic description of the input image frame. The knowledge modelling in the proposed system includes integration of probabilistic reasoning with the domain ontology for activity recognition.

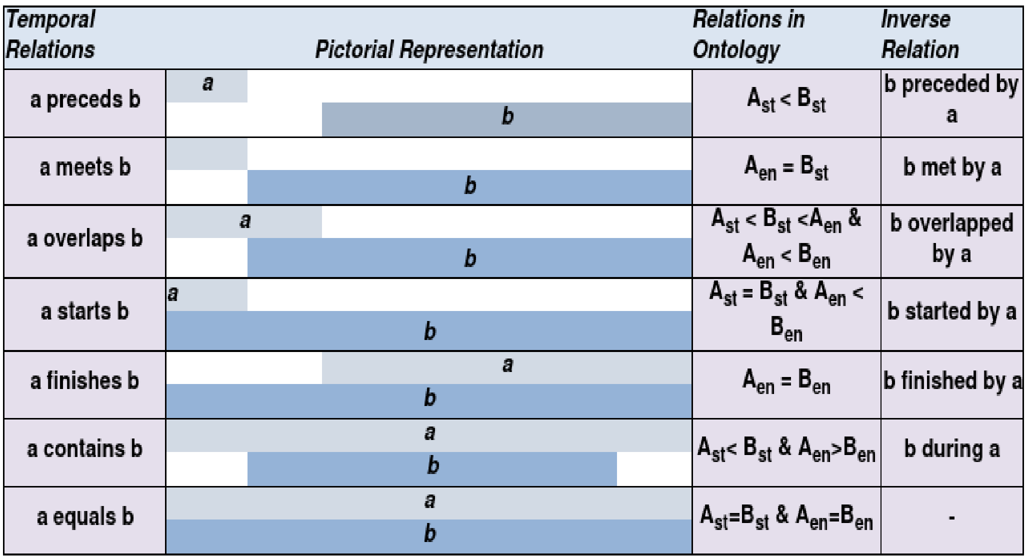


Figure 5.5: Pictorial representation of Temporal Relations in the Activity Recognition Ontology

The OWL is used to develop and formalize the schema of the Activity Recognition Ontology. The Protégé[6] tool is used to edit and visualize the structure of the Activity Recognition ontology. Also, Pronto[84]is used to add probabilistic axioms as well as a reasoner for probabilistic inference. Pronto library facilitates the probability enrichment through the element tag denoted as pronto:certainty. It is added to the ontology to represent a discrete value or range of probability. An example below shows the usage of pronto:certainty tag as part of the Activity Recognition Ontology:

⟨owl11:Axiom ⟩
⟨rdf:subject rdf:resource=‘URI#x’/ ⟩
⟨rdf:predicate rdf:resource=‘&rdfs:subClassOf’/⟩
⟨rdf:object rdf:resource=URI#y /⟩
⟨ pronto:certainty ⟩ 0.6905 ⟨ /pronto:certainty ⟩
⟨ /owl11:Axiom⟩

5.4.3 Complex Event Recognition

Complex Activities are performed in complex patterns and can be denoted as a combination of various simple activities along with temporal and spatial contexts. Composite and Complex activities can be categorized into 3 main types based upon their temporal nature: sequential, concurrent and interleaved. These complex activities are formally represented in the activity recognition model as concepts and are further semantically annotated as axioms/rules along for classification and detection. The different types of composite activities are represented in the activity model as follows: Sequential Complex activities are represented using temporal relations ‘before/after’. The temporal relations ‘before/after’indicate that a gap is present in between the time intervals where the two activities are being performed. On the other hand, the temporal relation ‘meets/metby’indicates absence of any time gap between the intervals of the two composite activities.

Thus, the Interleaved and concurrent complex activities are modelled by simple activities that are represented by completely or partially overlapping and started by time intervals. These time interval relations defined in the temporal context module are extended and included in axioms/rules to achieve Complex Activity recognition.

The mapping between the complex activity recognition process and observations with the the ontological representation is done as follows: Gestures detected from sensor states are depicted as instances of Atomic Gestures concept. Objects identified from object classifier are linked to the instances of physical object concept. As the name suggests, this concept class includes objects that can produce atomic gestures (e.g., kettle,microwave, pillbox, etc). Simple Actions recognised from the recognition module are represented instances of Simple Activity concepts. Finally, an SWRL rule/axiom-based implication from the simple activities can lead to complex activity identification. Algorithm 7 represents the pseudo code for inference rule generation for the inference process. The task of recognition of a complex activity is exemplified where, a simple activity instance 'cleaning'is followed by another simple activity 'prepare breakfast'. These two activities are performed one after the other modelled by the temporal relation 'After'. Hence, the activity 'clean'and 'prepare breakfast', together are implied to be a sequential complex activity. This way the ontological recognition model identified both simple, complex activities and a combination of both.

Additionally, the proposed ontology-based activity model incorporates probabilistic reasoning with the ontological inference mechanism to handle the uncertainty in the process. In the proposed module, the probability of a complex activity (C_{act}) is formalized as a function of probability of the simple activities (S_{act})/gestures (S_{ges}) along with the factors that can influence them . The probability-based Mixture Distribution is used to mix and combine the probabilities of simple activities along with the factors. The factors that intervene are defined as : Aptness (Apt) of the simple act to the complex activity and the Trust factor given the classifier (C_γ) from the previous modules. Equation 6 denotes the mixture probability distribution of the complex activity, given the observed simple acts. Trust (Tr) handles the correctness of the simple activity recognition. It is represented mathematically by Equation 4. It estimates the accuracy of recognition of the simple activities that are part of the complex event. Equation 5 represents, Aptness (Apt), which mathematically maps the contribution of a simple activity in a complex event detection to a value. It implies the simple acts that are important for a particular complex act.

$$Tr = Pr(S\_Act_{\alpha,\beta}|C_\gamma) = \frac{S_{\_True,\beta\gamma}}{S_{\_False,\beta\gamma} + S_{\_True,\beta\gamma}} \tag{5.4}$$

In the above equation, $S\_Act_{True,\beta\gamma}$ denotes the frequency of correct detection of Simple Activity contributing to the complex act by the classifier. Also, $S\_Act_{False,\beta\gamma}$ denotes frequency of detection of simple activity in absence of the complex event under consideration.

$$Apt = \frac{C\_Act_{S\_Act,\beta,\gamma}}{S\_Act_{\beta,\gamma}} \tag{5.5}$$

In the above equation, $C\_Act_{S\_Act,\beta,\gamma}$ denotes the frequency of presence of a Simple Activity for an instance of a complex activity. $S\_Act_{\beta,\gamma}$ denotes frequency of a simple activity is detected.

$$Pr(C\_Act_\beta) = \frac{\Sigma_{S\_Act_{\alpha,\beta,\gamma}} Sig(S\_Act_{\alpha,\beta,\gamma}) X Pr(S\_Act_{\alpha,\beta,\gamma})}{\Sigma_{S\_Act_{\alpha,\beta,\gamma}} Sig(S\_Act_{\alpha,\beta,\gamma})} \tag{5.6}$$

In the above equation, $Sig(S\_Act_{\alpha,\beta,\gamma})$ is computed as follows:

$$Sig(S\_Act_{\alpha,\beta,\gamma}) = Apt + Tr \tag{5.7}$$

Equation 7 shows the sum of Trust factor and Aptness values. Once the probabilities associated with a complex activity are computed, then the probabilistic inference handles the ambiguity among various complex activities.

The inference step takes these probability scores as input to retrieve the complex activity thats most likely to happen event based upon thresholding the probability value. It is implemented over the most likely activities to come to a conclusion if the computed probability value implies to a real-life event or not.

The task of thresholding of the computed probability provides a way that identifies a probability level that shows that the proposed mathematical probability distribution model has sufficient evidence to accurately recognise a real-world activity or event. Thus the recognized simple and composite activity can be further used for response activation. The activated response can be in the form of a simple alert, message or call to the concerned person or caretaker.

5.5 Experimental Evaluation and Results

The entire experimental setup of the proposed system is designed and implemented on Intel Core processor i5 processor with 16 GB RAM and GTX Nvidia card on windows 10 operating system. The sensor data preprocessing in terms of data resampling in performed using pandas 1.4.0 python library. The image data preprocessing and annotation is done with OID_v4_Toolkit for faster RCNN model run. The ontology development and implementation is done using Java programming language in OWL framework. OWL API 3.0 toolkit is used for the development and editing of the ontology. The rule base is coded and implemented through the SWRL framework. For adding the probabilistic enrichment, pellet-based ontological reasoner, Pronto, is deployed which facilitates addition of probabilistic axioms as well as inference on the knowledge base through OWL API. SPARQL endpoint is deployed and executed over the probabilistic ontological knowledge base to perform querying and fetching of inferred results and knowledge using OWL API. Eclipse Java version Photon is chosen as the IDE to create the development environment for probabilistic ontology construction with java development version JDK 8.

For creating environment for faster RCNN, anaconda tool is used. The created environment is launched through spyder 5.2.0. Lastly, Protege 5.2.0 is installed to achieve visualization of the developed ontological knowledge base and the validation of proposed system.

5.5.1 Heterogeneous Dataset

In the considered Dem@care[85] dataset, participants of age approximately 65 years and beyond were chosen by a Greek Institute under a an European project. The chosen patients were suffering from Alzheimer's disease. It comprises sensor-based heterogeneous observations and recordings of the various patients performing several daily routine physical chores/tasks. Various plug, physiological and wearable sensors were deployed to capture activity related observation. Sensor-based experimental video recordings were captured by color-depth sensors at frequency approximately 10 frames/second).

5.5.2 Querying service deployment and System Evaluation

It is pivotal to test the developed and implemented system for it's accuracy, precision, coherency of the domain and the consistency for evaluation. Therefore, the process of evaluating the proposed system is achieved into three different stages. Each stage has a different criteria for evaluation: ontology validation, evaluating by comparing the proposed system with existing approaches, and thirdly evaluating the overall accuracy and efficiency of the proposed system by launching various sets of SPARQL queries designed.

Query-based Validation

The evaluation of the developed probabilistic ontological knowledge base is performed here to track the overall accuracy of the proposed system. The verification and validation of the ontology is very important, since overall performance of the proposed system greatly depends upon its consistency and coherency. Since, the ontology alters frequently at each update or change,

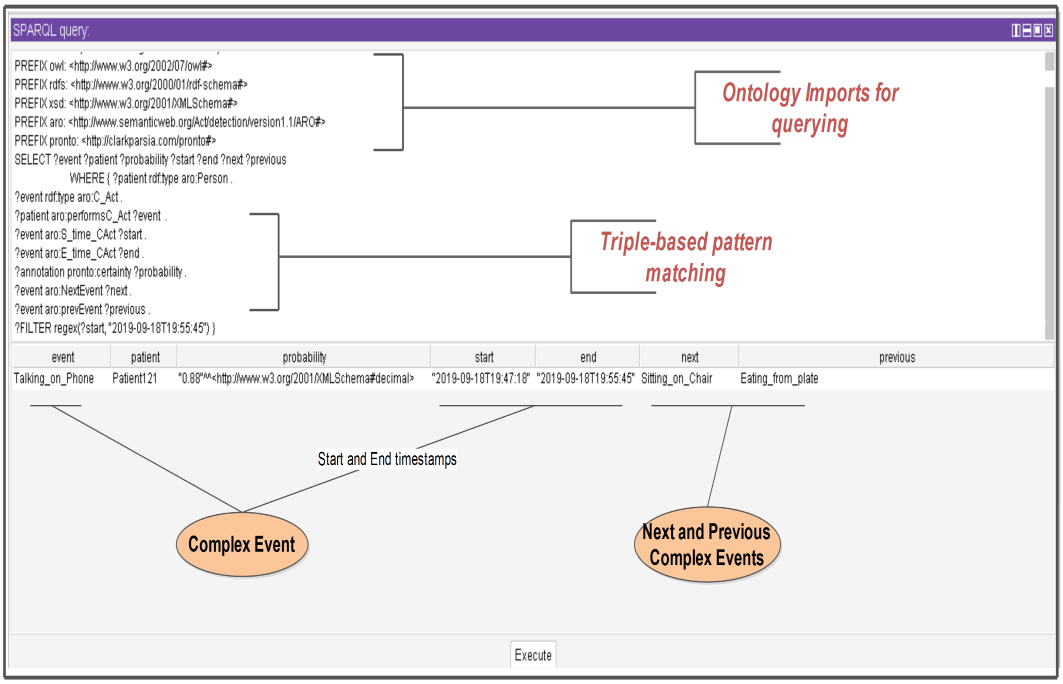


Figure 5.6: Query that shows probability and metadata of a Complex Activity being performed

consistency checking becomes all the more crucial. In order to ensure and validate the probabilistic ontological knowledge base, a Q/A (Question/Answer) method is adopted to test result accuracy, and perform overall evaluation [86]. Protege editor is used to launch SPARQL-based queries and the results are checked for consistency. Various SPARQL/DL [87] queries are fired in OWL language that fetch required information based upon pattern matching through the usage of variables and triples. Some of the queries designed and fired at the constructed ontology to test the overall precision, performance and consistency of the system are as follows:

SPARQL Query 1: Query to retrieve the metadata about a particular complex activity (Washing Dishes) being performed in a particular interval along with the computed probability.

Query:

PREFIX rdf: $\langle http://www.w3.org/1999/02/22rdf syntaxns\#\rangle$

PREFIX owl: ⟨http://www.w3.org/2002/07/owl#⟩

PREFIX rdfs: ⟨http://www.w3.org/2000/01/rdfschema#⟩

PREFIX xsd: ⟨http://www.w3.org/2001/XMLSchema#⟩

PREFIX aro: ⟨http://www.semanticweb.org/Act/detect/ARO#⟩

PREFIX pronto: ⟨http://clarkparsia.com/pronto#⟩

SELECT ?event ?patient ?probability

?start ?end ?next ?previous

WHERE { ?patient rdf:type aro:Person .

?event rdf:type aro:C_Act .

?patient aro:performsC_Act ?event .

?event aro:S_time_CAct ?start .

?event aro:E_time_CAct ?end .

?annotation pronto:certainty ?probability .

?event aro:NextEvent ?next .

?event aro:prevEvent ?previous .

?FILTER regex(?start, "2019-09-18T19:55:45")

}

SPARQL Query 2: Query that retrieves all the simple activities and atomic gestures that are part of the complex activity.

Query:

SELECT ?event ?patient ?probability ?Sagest

?Eagest ?SimpleAct ?TapOnTime ?TapOffTime

WHERE { ?patient rdf:type aro:Person .

?event rdf:type aro:C_Act .

?patient aro:performsC_Act ?event .

?annotation pronto:certainty ?probability .

?event aro:startedBy_Agest ?Sagest .

?Sagest aro:time_Agest ?TapOnTime .

?event aro:endedBy_Agest ?Eagest .

?Eagest aro:time_Agest ?TapOffTime .

?event aro:hasS_Act ?SimpleAct .

}

accuracy and efficiency of the proposed system.

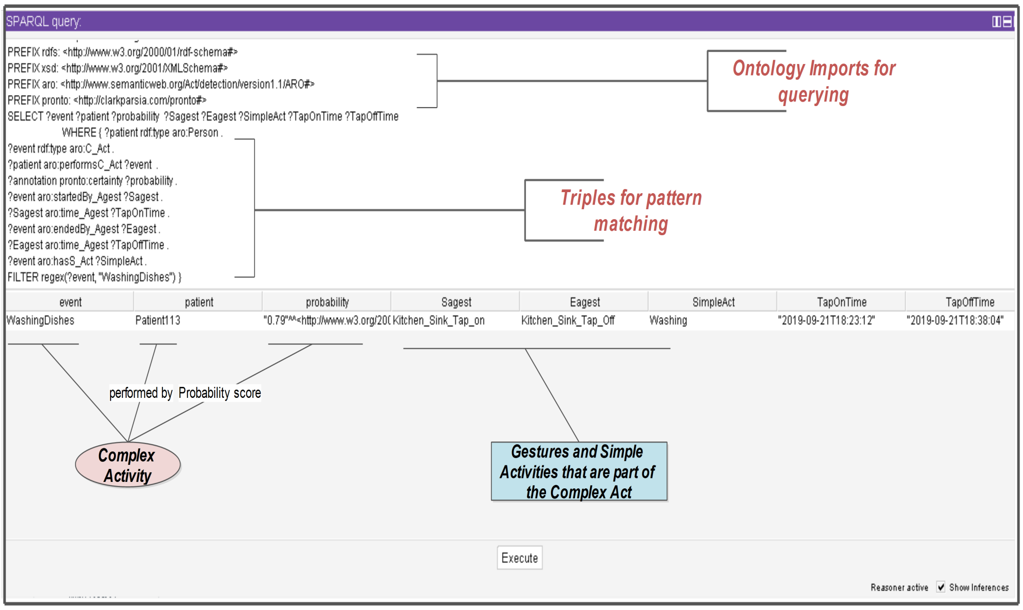


Figure 5.7: Query that shows all the Simple Acts and Gestures that imply the Complex Event being performed

Overall System's Performance and Efficiency

For the evaluation of overall efficiency and precision of the proposed approach, heterogeneous sensor-based activity related data belonging to different subjects or patients with varying conditions drives the way to launch inquiries in the form of queries to the proposed ontological knowledge system. In order to perform this evaluation, the proposed probabilistic ontological system was instantiated with the entire set of sensor samples from the Dem@care dataset that fall under certain observation windows and timeframe, corresponding to a patient/person under consideration, uniquely recognised by patient ID. The S1th record of the dataset is chosen, since it has a wide spectrum of activity/event types, making it suitable for showcasing the overall performance of the system. Additionally, to achieve result verification, this particular chosen of dataset patch has been vividly annotated by the dataset owners and experts. After the fixing of the imports closure of the

Table 5.1: Performance of queries on proposed Activity Recognition System

	StartTime	EndTime	Simple Activity /Atomic Gestures	Complex Activity	Precision
Act1	2019-09-18 T19:31:09	2019-09-18 T19:45:01	Beverage bottle picked,Cup Picked	Drinking	100%
Act2	2019-09-18 T19:47:19	2019-09-18 T19:55:19	Phone_Receiver _Activated	Talking on Phone	100%
Act3	2019-09-18 T19:56:11	2019-09-18 T20:04:17	Eating_from _Plate	Eating	100%
Act4	2019-09-18 T20:06:13	2019-09-18 T20:20:15	Chair_sensor _Activated	Sitting_on _Chair	100%
Act5	2019-09-18 T20:23:11	2019-09-18 T20:30:01	Book picked up, Chairsensor_Active	Reading	100%

ontological knowledge base, the probabilistic reasoner is launched and activated, driven by the developed rule base, the inference mechanism is carried out. The results in the form of inferred triples are fetched and saved, and the the knowledge base is updated. For verification and evaluation, a set of queries are constructed and fired w.r.t each person/patient to fetch semantic knowledge provided by the ontological classifier.

Table 5.1 shows a subset of the results of complex activity recognition from simple acts/atomic gestures from 5 different instances under the s1th record of DS1 containing image frames and DS12 dataset having data from physiological sensors, belonging to a particular person/patient, as a result of the queries launched over the ontological knowledge base of the system. The S1th record of the DS1 dataset consist of a total of 2470 image frames, out of which 1235 image frames are discarded while pre-processing the data, since they are blackened images with no object detection achieved and DS12th patch contains sensor samples whenever any movement is detected by plug/physiological sensors.

Table 5.2: Description of some Complex Activities

Activity	Definition
Washing Dishes 15-20 mins	aro:startswithSAct(Washing)⊓tro:overlappedby(Standing) ⊓tro:startedBy(Kitchen_Tap_on)⊓finishedBy(KitchenTap_Off) ⊓(Kitchen_door_Opened)⊔Near_Burner) ⊓Inside_kitchen
Having Meal 10 mins	startswithSAct(Taking_plate)⊓Near_DiningArea⊓ (Sitting_DiningChair ⊔ Standing _nearDiningTable)
Taking Medicine 45-60 sec	startswithSAct(Taking_PillBox)⊓ tro:startedBy(Kitchen_Tap_on)⊓ Inside_kitchen
Reading Book 30-40 mins	startswithSAct(Reading)⊓tro:startedBy(room_Lamp_on) ⊔tro:finishedBy(room_Lamp_off)⊓(Sitting_on_sofa⊔ Sitting_on_Chair)
Showering 10-20 mins	tro:startedBy(Shower_On) ⊓ tro:finishedBy(Shower_Off) ⊓ Inside_Washroom
Toileting 2-10mins	startswithSAct(Sitting)⊓ tro:startedBy(Toilet_Door_ Open)⊓ (tro:finishedBy(Toilet_Door_close⊔ tro:finishedBy(Flush_pressed))
Watching TV 15-30 mins	(startswithSAct(Sitting_On_Sofa)⊔ startswithSAct(Sit_on_Chair)) ⊓tro:startedBy(TV_On)⊓(tro:finishedBy(TV_Off)⊔ Inside_Living_Room
Watering Plants 5 mins	tro:startedBy(Hose_On)⊓(tro:finishedBy(Hose_Off)⊓ (In_the_Garden ⊔Outdoors)
Prepare Beverage 2-5 mins	tro:startedBy(electric_kettle_On)⊓(tro:finishedBy (electric_kettle_Off) ⊓(Inside_Kitchen⊔ Near_Microwave)

In order to classify a particular instance into a type of activity, firstly segmentation of input sensor data is performed on the basis of fixed time frames i.e. 10

minutes. Beyond 10 minutes, if same set of objects are detected with same Region of Interest even in the next time frame, then more image frames are added into the data patch under consideration for activity recognition. Now, since the timeslot has been updated and fixed for further analysis, the object detection step is carried out by first preparing the image data. The set of objects detected, along with the data from the physiological/plug sensors about the movements of the patient/person under observation in the same time frame, are used to detect simple activities and atomic gestures based upon the ontological SWRL rule base in the knowledge base. The imports closure is now updated after the activity instance is classified into a simple activity type. Also, the region of interest is computed based upon the Bayesian Inference probabilistic distribution model proposed in the system. Then, along with the probability score, the ontological imports closure is again updated which means the activity instance under consideration, is further semantically annotated with the identified spatial context information about the activity being performed. At this point in the recognition process, all the required ontological instantiation has been achieved, so based upon the SWRL rule base, the complex activity being performed as a combination of simple activities along with spatial and temporal context, is recognised. The subset of the domain knowledge required for this recognition is shown in Table 5.2. Additionally, along with it, the probability score is computed from the probabilistic frequency distribution is also added into the ontology, associating this semantic information with the activity instance under consideration. Now, the recognition process has been performed, hence, various queries can be fired at the knowledge base with the aim to retrieve the activity being performed in a specific time interval or instant along with the semantic knowledge associated with it.

Additionally, another way to perform evaluation of the performance of the proposed system in terms of overall efficiency is through thoughtfully constructing and launching all the existing categories of SPARQL query forms on the implemented ontological knowledge base. The knowledge base of the proposed system is instantiated with heteregeneous sensor data patches from the Dem@care dataset. The sensor-based data records that were used were DS12 and DS1. The DS1th record has further 32 data groupings, out of which 6 are selected which have wide spectrum of activities being carried out by the patient/person along with clear activity annotations given for verification purpose. Finally, all the query types are formulated and fired, i.e., Describe, Construct,Select, Ask, Graph and Filter. Additionally, the query execution times of every query form is taken account of for comparison. Table 5.3 showcases the execution times of the queries in miliseconds (ms) associated with variables and triples that are part of the query designed. The length of these execution times imply that the probabilistic ontology is designed and constructed in such a way that promotes consistency. It also shows the well defined structuring of the ontology in terms of its design and implementation. Another implication is that the average execution time of considered query types is approximately within a second. Only those query forms that fetch the results in the form of graph have longer duration than the others fetching output in the form of triples or literals or tables.

Table 5.3: Execution time of different types of queries on the proposed system

Query	Pattern	Execution time(ms) w.r.t. Patient's data						Avg
	Vars,trip.	S2	S7	S8	S18	S21	S30	
DESCRIBE	10,5	0.195	0.189	0.192	0.156	0.172	0.122	0.171
CONSTRUCT	10,5	0.469	0.654	0.549	0.508	0.399	0.584	0.527
SELECT	2,7	0.655	0.423	0.648	0.578	0.464	0.646	0.569
ASK	11,8	0.345	0.299	0.398	0.298	0.400	0.287	0.337
Graph	10,6	0.723	0.546	0.376	0.613	0.695	0.687	0.606
FILTER	12,9	0.709	0.612	0.845	0.765	0.687	0.599	0.702

Comparison of the proposed system with existing systems

Precision, Accuracy, Recall and F-score are the metrics that are used to evaluate performance of the proposed ontological system. In order to perform the comparison of performance of the proposed ontological system with existing systems/approaches, the instantiated knowledge base is put under scrutiny. To

Table 5.4: Comparison of the proposed approach with other existing auto AI-based diagnosis approaches

Author(s) and Year	Methodology	#Classes	Performance
Negin et al.[88] 2016	Supervised Hybrid Framework	7	Acc:95% Sen:74.57% PPV:91.29%
Negin et al.[89] 2015	Unsupervised Activity learning approach	5	Sen:100% PPV:70%
Negin et al.[90] 2019	Unsupervised approach with hierarchical models	7	F-score:0.7 Prec:96.47% Sen:80.94%
Avgerinakis et al.[91] 2016	Activity recognition through SBDD Technique	7	Sen:86.22% PPV:49.33%
Elloumi et al.[92] 2014	Unsupervised using Global motion	7	Sen:77.71% PPV:90.29%
Proposed Approach	Probabilistic Ontology-based SWRL classification	12	Acc:98.45% Sen:94% Spec:95.09% PPV:98.65%

achieve this, the data patches S1 and S12 of the Dem@care dataset are use to instantiate the ontology. Both of the selected data patches contain a random distribution of different activities with clear sensor outputs. Firstly, the the sensor observational data was semantically annotated as per the proposed recognition model. After the recognition process is performed,various set of SPARQL queries relevant to activity detection and classification were launched at the system. Then the value of the above said performance metrics was computed based upon the fact that fetched results are relevant and accurate or not. Table 5.4 shows the comparison of the proposed approach with existing approaches for activity detection. For this, records of the dataset were segregated into 5 folds, and a set of 15 queries were designed and fired at the ontology. The sensor-based samples of activity instances are classified into an complex event among the set of 10 predefined activity classes. The evaluation computation performed has resulted into the following specificity(Sp), sensitivity(Se), accuracy(Acc) and positive predictive values(PPV) to be 95.09%, 94%, 98.45% and 98.65% respectively.

Chapter 6

ONTOLOGICAL KNOWLEDGE MODEL FOR HETEROGENEOUS SENSORS DIGITAL DATA

This chapter presents an ontology-driven knowledge model to semantically fuse digital ECG-based heterogeneous sensor data. The proposed Ontological system is integrated with the arrhythmia domain knowledge. Conceptual relationships relevant to classification of heartbeat into corresponding cardiac arrhythmias are represented in the developed modular ontology for sensor data analysis and knowledge gathering. The newly developed arrhythmia ontology consists of three different modules, each semantically annotating a different aspect of the arrhythmia detection process. A SWRL(Semantic Web Rule Language) based ontology classifier performs classification of patients ECG data into corresponding cardiac arrhythmia types. Further, SPARQL-based querying is performed to fetch the results from the Knowledge Base.

6.1 Digital Heterogeneous Sensor Data

There exists certain sensors that read electrical activity and convert the sensed data into digital signals to be analyzed later on. This syntactic dissimilarity in the sensor output opens up possibilities to integrate another form of heterogeneity associated for semantic reasoning. There are converters deployed along with such sensors that convert electrical activity data into digital data. This digital sensed data has no scope of human interpretability unless translated and fused into a semantic format to perform digital sensed data analysis. Additionally, it cannot be uniformly shared and exchanged due to the presence of syntactic heterogeneity

6.2 Integration of ECG Data from Heterogeneous Electrode Sensors

Advancement in the field of sensors-based smart materials along with information processing and communication technology drives their applicability in the domains of medical and healthcare. The main requirements of any healthcare setup range from mere healthcare services to intelligent and accurate healthcare systems, efficient health issue detection and timely alerts providing the status of health meant for users like doctors, caretakers etc. in interaction[93,94]. This kind of integration aims at improving the quality of life and living. Therefore, it can be convincingly said that sensor technology when implemented in the bio-medical domain can be adopted as a solution to make healthcare more effective and attainable for everyone. Cardiovascular disease is considered to be one of the most detrimental diseases, ironically widespread, contributing to approximately 30% of the total mortality globally. A huge number of cardiovascular deaths are associated with arrhythmogenesis[95]. Arrhythmia is an abnormality that arises in the heart rhythm[96]. The heart deflects from the normal range i.e. it can lower as well as increase the heart rate drastically. It can lead to most critical form of arrhythmias like ventricular fibrillation which are fatal[97]. Majorly, cardiac arrhythmias are categorized into two kinds: life-threatening or fatal and non fatal. The arrhythmias that fall in the life-threatening category can result in cardiac arrest and death all of a sudden like ventricular fibrillation.While atrial fibrillation arrhythmia is also considered a serious diagnosis. This condition is not considered fatal in itself, but can cause potentially life threatening complications like heart stroke and failure. Such circumstances require urgent attention and emergency treatment. While those arrhythmias that are non-life threatening cannot cause heart failure, timely appropriate treatment is still required to avoid any deterioration of the functioning of heart.

In order to capture such arrhythmia-related episodes, a Holter monitoring device is deployed which monitors patients in intensive scenarios or otherwise. Such devices generate long term ECG signal data. The electrocardiogram is a well-established tool which is non-invasive inexpensive for clinicians and patients to work with. ECG consists of vital physiological signals, widely used to process the heart beat data and analyze the functioning of heart. An ECG signal consists of recorded electrical activity of the heart, containing a sequence of P-wave, U wave, T wave and QRS complex[98] . The variations in the functioning and the electrical activity of heart are studied and analyzed to detect any sign indicating presence of any heart disease. Although, this task of analyzing the lengthy ECG records through manual inspection by a cardiologist is cumbersome and time consuming. Heart beats which are irregular and have features that vary from the normal beats usually indicate arrhythmias and such events can imply underlying heart problems. ECG signals consists of various waves and the characteristics of these waves like area, morphology, amplitude, length, and time duration between the major segments and waves are used to differentiate between normal and arrhythmic beats. Such characteristics are referred to as features.

An early diagnosis is considered to be an important aspect in providing a functional treatment for arrhythmias. The detection of various transient, in frequent and short-term arrhythmias at an early stage requires a stable monitoring lasting for longer duration. Intelligent healthcare monitoring systems can make patients spend lesser time obsessing about their health condition, hence more time enjoying and living quality life, while responsibly handling the health issues and care required. In order to turn the analysis into a manageable and easier process for the cardiologists, automated system based analysis is an active topic of research[99,100,101]. Hence, preferably clinicians usually implement computer-driven methods to interpret ECG signals.

6.3 Proposed methodology for Arrhythmia detection from Digital Sensor Data

The proposed system is an ontology-based reasoning system that detects and classifies heartbeats into one of the arrhythmia classes. The approach is influenced by how clinicians and doctors perform illness diagnosis through their expertise and clinical experiences,along with patients' individualistic characteristics.

The ontology-driven classifier processes the long-term ECG observations and performs arrhythmia diagnosis based upon the clinical data of the patient along with conceptual and experiential knowledge mapped into an ontology-based modular design. Figure 6.1 shows the architecture of the proposed system for arrhythmia classification. Algorithm 8 represents the algorithm for the creation of an ontological skeleton from a well-defined conceptual model comprising of a set of entities and relationships between them. The functioning of the proposed system is categorized into three main phases: preprocessing phase, semantic phase and analysis and reasoning phase; detailed below:

6.3.1 Preprocessing Phase

The ECG sensor recording data retrieved from MIT-BIH database is not received from a patient directly. Therefore, it can be effectively assumed that it does not contain disruptive noise in the ECG recordings. But, in order to make the ECG data suitable for use in the proposed system, following preprocessing is performed:

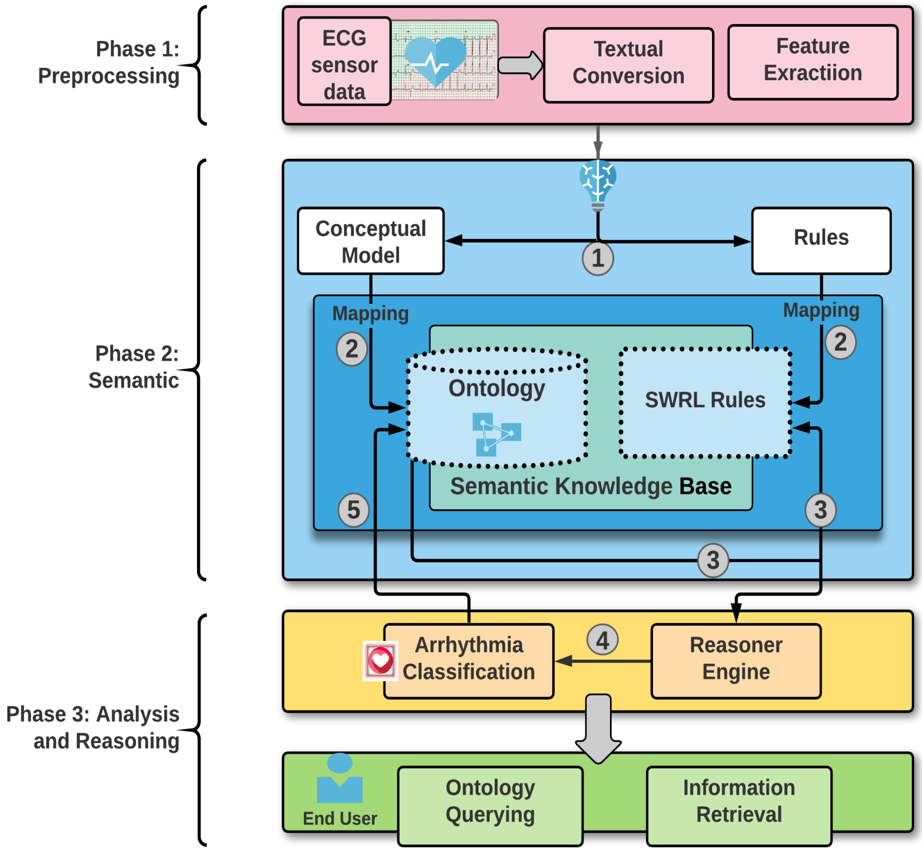


Figure 6.1: Architectural diagram of Proposed System

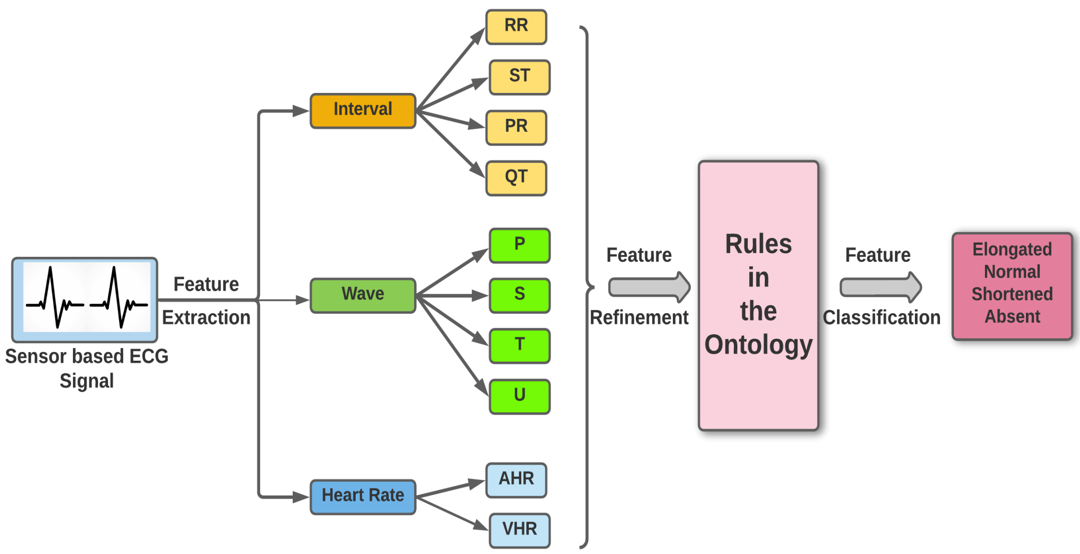


Figure 6.2: Feature Extraction from ECG Signal Data

- **Data Conversion and Integration:** The sensor-based ECG signal data is converted into textual form. The digitized form of the ECG data needs to be transformed into a format where the heartbeat segmentation and feature extraction can be achieved.

- **Feature Extraction:** Further, the heartbeat segmentation is carried out where essential features in the form of time duration is extracted corresponding to elementary forms like waves and intervals in the ECG recording. The digital recordings are transformed into heat beat annotations associated with time instant. The features extracted from ECG data include duration of interval and waves. Figure 6.2 shows the process of extracting features. The features are extracted in the form of interval, waves, and heart rate using the WFDB (Waveform Database) library. Tools and functions like ecgpuwave, rdann, sqrs, ihr, rdamp, etc. are used to detect crucial points in the ECG signal. These crucial points like onset/offset of P-wave, QRS complex, T-wave, PR-interval, RR-interval etc. are further categorized into either of the four feature labels, namely, normal; elongated; shortened and absent based upon the rules defined for feature classification in the ontology.

6.3.2 Semantic Phase

This layer forms the core of the knowledge base for the effective functioning of the proposed system. The knowledge base consists of a flexible ontology-based modular structure,comprising of three modules namely: a) Patient Profile module, b) Observation Diagnosis module, c) Arrhythmia module. The namespace or IRI (Internationalized Resource Identifier) of the ontology database is represented by the prefix "arr" and is added to each of the components of the ontology, i.e., classes and properties. Figure 6.3 shows the diagrammatical representation of various modules of the proposed ontology. The modules are further explained in detail below:

1. **Patient Profile module:** Patient-related data is usually complex and exists scattered in heterogeneous formats, which makes it difficult to perform

knowledge mining to carry out some relevant decision making. This module is a formal representation of clinical data that contains information about the signs and symptoms of the patient. It is developed to personalize the medical knowledge about the patient under consideration, who is given a unique identification ID. The semantic profile and clinical data in this module can be used in the ontological analysis and inference by proceeding layer of the model. It can also be forwarded and given access to clinicians as a part of Electronic Health Record of the patient.

2. **Arrhythmia module:** In the Arrhythmia module, the consensual and conceptual knowledge crucial for the arrhythmia detection is represented in an ontological format. The arrhythmia-related knowledge is given a semantic meaning by mapping the core concepts into a set of classes and meaningful relationships between them. This module is a very crucial module for classification of heartbeats into either of the pre-defined classes in the arrhythmia module.

3. **Observation module:** This module has a dynamic flavor associated with it since it represents all the observational values extracted from the ECG features. All the values extracted from the preprocessed ECG data are represented here in this module and are uniquely identified by the time instant i.e., RR-interval and PR-interval values between two consecutive cardiac cycles, P-wave and R-wave. It semantically annotates all the ECG indicators and their features that are used for the classification based upon the rules in the reasoning layer. It represents the taxonomy of the ECG indicators, i.e., Wave, Intervals and segments, which are further categorized into the respective types.

Algorithm 9 represents the instantiation process of the developed ontological Knowledge Base. It depicts the process of associating domain knowledge with observational sensor-based ECG data, without which no dynamic and situation-driven decision can be made with respect to patients.

Ontology Development Assumptions:

The crucial part of an ontology-based classification system is the development of the knowledge base. In order to develop the ontology, a state-of-the-art ontological engineering approach is deployed which is based on two ontological artefacts [102]:

- Any subject domain that is represented ontologically, should be strongly axiomatized that constraints the original intended meaning of the theory.

- Computability for reasoning and information extraction should be preserved.

So, based upon the above mentioned assumptions, the generic approach for ontology development can be classified into three main steps: specification (acquiring core conceptual knowledge of the domain), conceptualization (organizing and structuring the knowledge in the form of key classes and relationships among them) and implementation (verifying and validating the developed ontology).

Existing Ontology Interoperability:

It is always beneficial to develop a biomedical ontology that is based on top of an existing ontological foundation. The task of ontology alignment is defined as establishing correspondence among ontologies [103]. A correspondence relationship can be in the form of equivalence ($\equiv$), generality ($\sqsubseteq$), and disjointness ($\perp$) relations between the classes of each of the ontologies involved in the alignment.

Table 6.1 shows the alignment relationships with some of the important and foundational or upper-level ontologies. The proposed ontology is semantically aligned with the top-level ontologies.

Algorithm 8 Creating skeleton of the Ontological Knowledge Base (KB).

Require: Set E of entities, P of properties and S of statements (S).

Ensure: Ontology ($Onto_{Out}$) output file in manchester OWL format i.e., Arrhythmia.owl

Instantiate OWLOntologymanager class
Create and map IRI to ($Onto_{Out}$)
Instantiate ManchesterOWLSyntaxOntologyFormat class and assign output format
for each $class_i \in E$ **do**
    Instantiate OWLClass with $class_i$
    Assign IRI 'arr'
    Add declaration axiom to '$Onto_{Out}$'
    Add subclass/superclass axioms with x $\in$ E : x $\neq$ m
end for
for each $prop_i \in P$ **do**
    Assign IRI
    if $prop_i isObjectProperty$ **then**
        Instantiate OWLObjectProperty class
    end if
    if $prop_i isDataProperty$ **then**
        Instantiate OWLDataProperty class
        Define $datatype_{prop}$ expression
    end if
    Create axioms:$prop_i \sqsubseteq \rho$;$prop_i \equiv \gamma$; $prop_i \equiv \delta$,$\exists\rho$,$\exists\gamma$, $\exists\delta$, $\exists$P
    Add axiom rdfs:range($prop_i$, arr:x) , $\exists$ x $\in$ E
    Add axiom rdfs:domain($prop_i$, arr:y) , $\exists$ y $\in$E
end for
for each $assertion_i \in S$ **do**
    Define DataType/ObjectType assertion i.e. ($prop_i$, $class_i$, $class_j$); $class_i \neq class_j$
    Add assertion axiom to '$Onto_{Out}$'
end for
call saveOntology with args

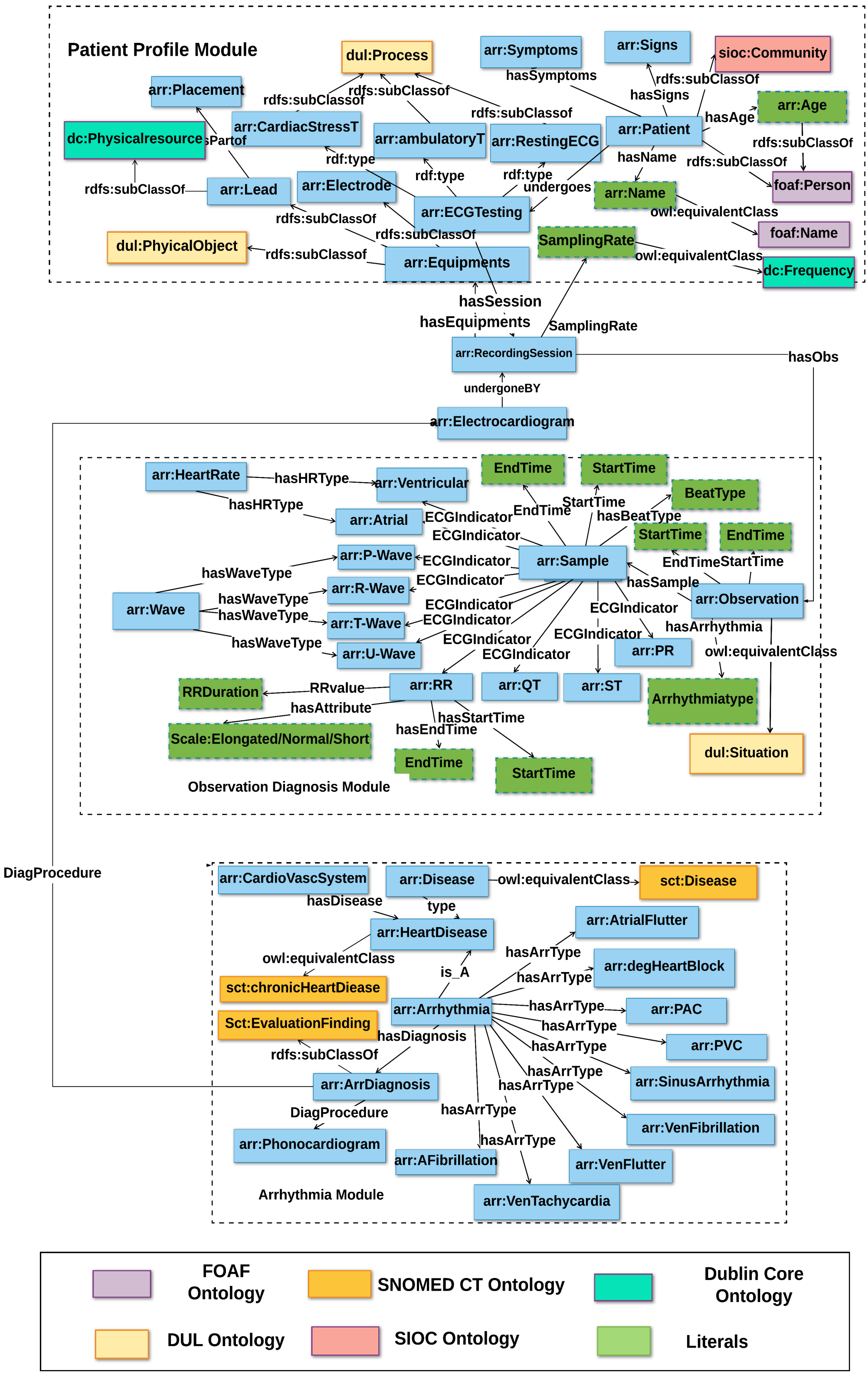


Figure 6.3: Taxonomy of Arrhythmia Ontology

Algorithm 9 Ontological Knowledge Base Instantiation

Require: Set R of ECG recording sessions in textual format.
Ensure: Instantiated ontology with observational ECG data.

Load ontology '$Onto_{Out}$'with OWLOntologyManager object //using Algo (8)
Instantiate BufferedReader class with path+recording session name
Initialize global variables
Instantiate DataFactory class
for each $record_i \in R$ **do**
  Add classtype assertion (RecordingSession, $record_i$)
  Add axiom hasSession($Patient_i$, $record_i$)
  Add datatype axioms for start and end time.
  for each $obs_i \in OB$ **do** // each session has a set of Observations 'O_j'
    Add classtype assertion (Observation, obs_j)
    Add axiom hasObs($record_i$, obs_j)
    Add datatype axioms for start and end time of obs_j
    for each $beat_k \in S$ **do** // each obs has a set of beats i.e., Samples 'S_k'
      Add classtype assertion (Sample, $beat_k$)
      Add start/end axioms for $beat_k$) //each beat has duration
      for each $ECGindicator_m \in EcgI$ **do** // Interval, Waves,H r
        if $ECGindicator_m isInterval$ **then**
          for each interval type **do** //types: RR, PR, QT, ST
            Add axiom for start/end time of $ECGindicator_m$.
          end for
        end if
        if $ECGindicator_m isWave$ **then**
          for each wave type **do** //types: P, R, T,QRS
            Add axioms for start/end time of $ECGindicator_m$
          end for
        end if
      end for
    end for
  end for
  call saveOntology with args
end for

6.3.3 Analysis and Reasoning Phase

This phase comprises of rules-based reasoning task carried out over the ontological database. Based upon the domain knowledge about the ECG analysis and its interpretation, key ECG indicators are fundamental components which become criteria for the classification of different types of arrhythmias.

Table 6.1: Ontology alignment of proposed ontology to existing top-level ontologies

Found -ational /Upper Ontology	Alignment Relations with the proposed ontology	Ontology description
DUL/ DOLCE [104]	(arr:ambulatoryECG$\sqcup$arr:RestingECG $\sqcup$arr:CardiacStressT)$\sqsubseteq$dul:Process; arr:ECGIndicator$\equiv$dul:InformationObject; (arr:Wave$\sqcup$arr:Interval$\sqcup$arr:HeartRate)$\sqsubseteq$ dul:Quality; arr:Observation$\equiv$dul:Situation; arr:Equipment $\sqsubseteq$ dul:PhysicalObject	Upper-level ontology for capturing foundational language and human common sense to widen the scope across multiple domains.
FOAF [105]	arr:Patient$\sqsubseteq$foaf:Person; arr:holdsAccount$\sqsubseteq$foaf:hasAccess; arr:Age$\equiv$foaf:Age; arr:Name$\equiv$foaf:Name	Ontology for describing people,their activities and relations to other individuals and objects.
SNOMED CT [106]	Sct:ClinicalFinding$\sqsupseteq$arr:Disease Sct:Disease$\equiv$arr:Disease Sct:Bradycardia$\sqcap$sct:Tachycardia $\sqsupseteq$arr:Arrhythmia Sct:chronicHeartDisease$\equiv$arr:HrtDisease Sct:EvaluationFinding$\sqsupseteq$arr:ArrDiagnosis	Ontology that represents collection of scientific labels, synonyms and definitions of medical related terms used in the clinical documentation, communication and reporting.
DC [107]	dc:Frequency$\equiv$arr:SamplingRate; dc:PhysicalResource$\sqsupseteq$arr:Lead$\sqcup$ arr:Electrode; dc:Service $\sqsupseteq$ arr:ArrDiagnosis;	Ontology for describing digital as well as physical resources.
SIOC [108]	Sioc:Community$\sqsupseteq$arr:Patient$\sqcup$ arr:Doctors; Sioc:User$\sqsubseteq$foaf:OnlineAccount	Ontology describing information from online communities.

Table 6.2 represents the core knowledge for arrhythmia classification. Based

Algorithm 10 Inference Rules Generation Function

Require: Set R of rules.
Ensure: Ontology ‘O’with SWRL rules.

function SWRLruleCreate(OWLOntologyManager $\langle object \rangle$) //using Algo 8
for each $vrules_i \in R$ **do**
  Initialize Set$\langle SWRLAtom \rangle antecedent_{rule}$
  Initialize Set$\langle SWRLAtom \rangle$ $consequent_{rule}$
  Initialize SWRL variables
  for each $class_i \in vrules_i$ **do**
    Create SWRLClass Atom with retrieved object
  end for
  for each $prop_i \in vrules_i$ **do**
    if $prop_i isObjectProperty$ **then**
      Create SWRLObjectProperty Atom with object/SWRL variable
    end if
    if $prop_i isDataProperty$ **then**
      Create SWRLDataProperty Atom with object/SWRL variable
    end if
  end for
  for each $atom_k \in vrules_{antecedent}$ **do**
    Add atom to antecedent $vrules_{antecedent}$
  end for
  for each $atom_k \in vrules_{consequent}$ **do**
    Add atom to consequent $vrules_{consequent}$
  end for
  for each $builtin_{swrl} \in vrules_{antecedent}$ **do**
    for each $args_i \in Argsbuiltin_{swrl}$ **do**
      List.add(argsi) //Add all builtin atoms in each rule to a list
    end for
    Call Access_BuiltinAtom() function //Retrieve IRI using Algo 11
    Generate SWRLBuiltin Atom with $IRI_{builtinswrl}$ and List of args
    Add Atom to antecedent to $vrules_{antecedent}$
  end for
  Initialize SWRLRule class
  Add rule $vrule_i$
  Call applyChange function to save axiom
end for
Call saveOntology ‘$Onto_{out}$’
End

Table 6.2: Subset of domain knowledge for Arrhythmia Classification

Heartbeat type	Interval						Wave								Rate	
	RR			PR			P				QRS				A	V
	S	N	E	S	N	E	A	S	N	E	S	N	E		Value	Value
Atrial Flutter	√										√				300	150
Atrial Fibrillation	•		•	√			√				√				350- 650	100- 170
PAC					•	•	•	•		•	•	•			-	60- 100
Ventricular Tachycardia													√		60- 100	100- 250
Ventricular Flutter	√														60- 100	150- 300
Ventricular Fibrillation	√						√								-	300- 600
PVC							√						√		-	
Pause												√			-	⟨ 60
Sinus Bradycardia			√		√								√		-	⟨ 60
Sinus Tachycardia	√				√				√				√		100- 180	100- 180
1 deg Heart Block						√						√			-	60- 100

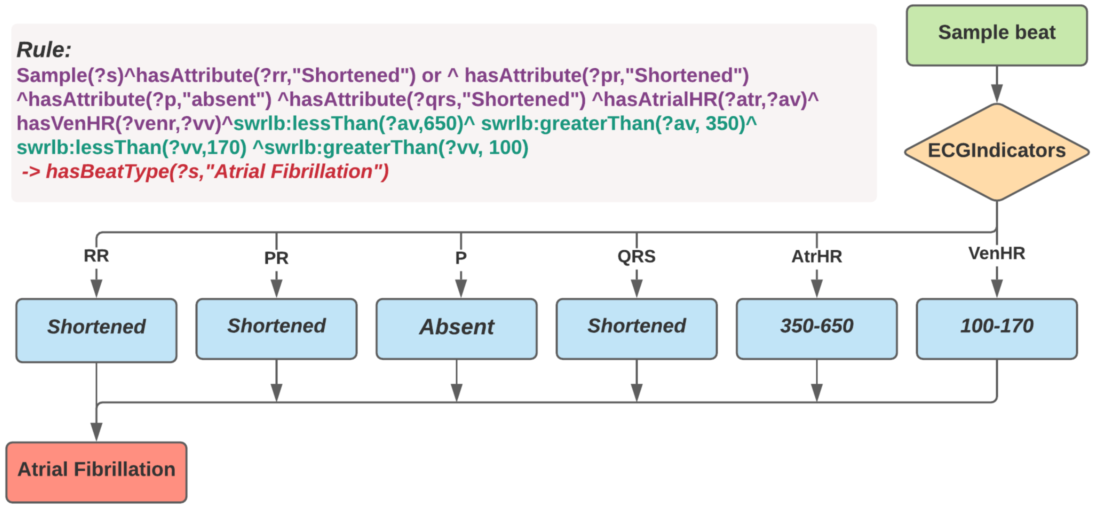


Figure 6.4: SWRL Rule description

upon the domain knowledge gathered, this table defines necessary and sufficient conditions for classification of a segment of an ECG signal (heartbeat sample) into the corresponding arrhythmia type. It lists a subset of features that have been used in the proposed methodology. In this table, '$\surd$' represents intersection (conjunction) of conditions and '•' represents union (disjunction) of conditions necessary for heartbeat classification. For example, (1) disjunction of shortened RR and elongated RR (2) conjunction of shortened PR, absent P-wave, shortened QRS, atrial rate in between 350 and 650 bpm, and ventricular rate in between 100

Algorithm 11 Function to Access IRI of SWRLBuiltin Atoms

Require: Set of 'b'SWRLbuiltin Atoms (B).

Ensure: Hashmap of Atoms and IRI.

1: function Access_BuiltinAtom(Set $\langle String \rangle$ $builtin_{swrl}$)
2: **for each** $atom_i \in B$ **do**
3: Initialize Hashmap$\langle String, String \rangle$
4: Initialize IRI class
5: Call getIRI function with $atom_i$ as argument
6: Add $atom_i$ and IRI_atom to Hashmap
7: **end for**
8: end function

and 170 provides a subset of necessary conditions to classify this ECG segment into Atrial Fibrillation arrhythmia. The domain knowledge represented in this table is mapped into axioms/rules, which are fed to the reasoning engine to perform the task of arrhythmia classification. Algorithm 10 presents the process of creating SWRL rules from the axiomatic domain knowledge captured. A SWRL[109] rule consists of two parts: antecedent and consequent, separated by an implication operator ($\rightarrow$). Further, each of these components comprise of a set of atoms. Atoms can be user-defined as well as can be chosen from the built in library. Algorithm 10 represents the pseudo-code for fetching built in atoms to be added into the rule from the built in library.

The rule generation process involves adding a set of atoms to either side of the rule. Once all the atoms are added, the rule is generated and finally added to the ontological database by the process as mentioned in Algorithm 11. A SWRL rule to classify a Sample beat into an arrhythmia type is shown in figure 6.4.

After the generation of rules, classification is performed by the reasoning engine. The reasoner is activated by giving semantically annotated sample beat as input, which acts upon these well-defined and formalized rules/axioms to classify the beats in one of the arrhythmia classes. Algorithm 12 presents the reasoning task in a phased manner. Firstly, to activate the reasoning engine, it is initialized with the instantiated ontology in the form of imports closure. The imports closure of the proposed ontology consists of all classes, properties, and set of axioms/rules defined at a particular instant of time. The reasoner is activated in buffering mode, which makes it possible to accommodate all the updates made to the ontology. Hence, the reasoner loads a fresh imports closure whenever changes and updation is carried out over the proposed ontology. After the reasoner completes the import of all the changes, consistency of the ontology is checked. If the ontological database turns out to be consistent, i.e., follows all the rules needed to perform further analysis, the rule-based classification is performed. Hence, a particular sample beat is classified into a class that is dedicated to a particular arrhythmia type. Further, all the inferred axioms and triples are further added into the input ontology so that the result is retrieved and knowledge is extracted whenever the ontology is queried.

Algorithm 12 Reasoning Mechanism (Arrhythmia Classification)

Require: Ontology '$Onto_{Out}$'.

Ensure: Ontology '$Onto_{Out}$'with Inferred Axioms.

1: Instantiate OWLReasonerFactory with PelletReasonerFactory())
2: Call createReasoner function with ontology manager of '$Onto_{Out}$'.
3: **for each** applyChange() called **do**
4: Call getmportsClosure() with '$Onto_{Out}$'.
5: Load imports closure to Pellet reasoner instance
6: **end for**
7: Check Consistency of '$Onto_{Out}$'.
8: **if** $Onto_{out} is consistent$ **then**
9: Call Classify function //inbuilt function by Pellet
10: Instantiate InferredOntologyGenerator to save inferred axioms
11: **end if**

6.4 Query-driven Experimental Evaluation

The entire setup of the proposed system is implemented on Intel Core processor with 8 GB RAM and windows 10 operating system. The preprocessing of the dataset is performed using Cygwin DLL 3.1.6 with implementing and deploying WFDB software package for the feature extraction process. The ontology implementation is based mainly on Java programming language using OWL framework. OWL API 3.0 is used for ontology development and editing. The process of rule base development is performed by using SWRL framework. Java-based Pellet reasoner is deployed through OWL API to execute reasoning tasks. SPARQL endpoint is implemented over the ontological knowledge base to perform querying through OWL API. Eclipse IDE Java version Photon is used as the development environment to provide an integrated environment for code execution with JDK 8. Protege 5.2.0 is used for visualization of the knowledge base and validation of the proposed system.

6.4.1 Digital Heterogeneous Dataset

For the purpose of experimental evaluation of the proposed system, the ECG observational data is obtained from the MIT-BIH Arrhythmia database[110] of the physionet library[111]. The dataset contains ECG recordings from 47 subjects under observation at the MIT-BIH Arrhythmia Lab. The MIT-BIH dataset consists of a total of 48 recording belonging to a mixed population of inpatients and outpatients in the hospital. Out of these 47 recordings, 9 records were selected to perform the evaluation process of the proposed work. These recordings contain samples that represent not just common but significant arrhythmia types clinically, along with normal heartbeat samples. Each of these selected total records is approximately 30 min long in duration and contain values of 2 leads, i.e., two different signals recorded by placing sensor-based electrodes at distinct locations on the chest of the patients.The 9 records selected for evaluation and validation are identified uniquely as 100, 103, 105, 111, 113, 200, 202, 210, 212. These records in total consist of 20,390 heartbeat samples, out of which 17,501 beats are normal and 2140 beats are arrhythmic or abnormal beats. All these records are selected out of the whole lot-based upon unbiased distribution of the normal and abnormal beats. Additionally, the annotation given by the human experts corresponding to these recording is vivid and clear for validation and performance metrics computation.

6.4.2 Experimental Evaluation

Once the system is implemented, it is crucial to test and compare the system's accuracy, consistency, and domain coherency for evaluation. Hence, the evaluation process is segregated into three distinct phases, each with a different evaluation criteria: ontology evaluation and validation, a comparison between the proposed system and other existing approaches, and evaluation of the proposed system's overall efficiency.

- **Querying for Validation**

  The performance evaluation of the proposed ontology after implementation is carried out in order to measure the improvement in the overall accuracy of the system. The task of verifying and validating an ontology holds utmost importance since any update or altercations done to the ontology can lead to the declaration of duplicate or redundant instances that disturb the consistency of the ontology. A Q/A (Question/Answer) method is deployed to perform result accuracy, and evaluation[112]. Protege tool allows firing DL queries and SPARQL queries on the developed ontology and provides flexibility along with a user-friendly environment for validation. Additionally, it constructs valid instances or individual members as a result of the input query in the required time. DL/SPARQL[87]queries are constructed in OWL syntax and combine variables and relationships to retrieve the required information. Certain queries that are designed and executed over the proposed ontology to perform evaluation of the overall accuracy and consistency are as follows: **SPARQL Query 1:** Query for listing signs and symptoms corresponding to a patient uniquely identified by PID *12119*

  Query:

  Syntax: PREFIX rdf: $\langle http://www.w3.org/1999/02/22-rdf-syntax-ns\#\rangle$

  PREFIX owl: $\langle http://www.w3.org/2002/07/owl\#\rangle$

  PREFIX rdfs: $\langle http://www.w3.org/2000/01/rdf-schema\#\rangle$

  PREFIX xsd: $\langle http://www.w3.org/2001/XMLSchema\#\rangle$

  PREFIX arr: $\langle http://www.semanticweb.org/arrhythmia/arr\#\rangle$

  SELECT ?patient ?signs ?symp

  WHERE { ?patient rdf:type arr:Patient .

  ?patient arr:P_ID 12119 .

  ?patient arr:hasSigns ?signs .

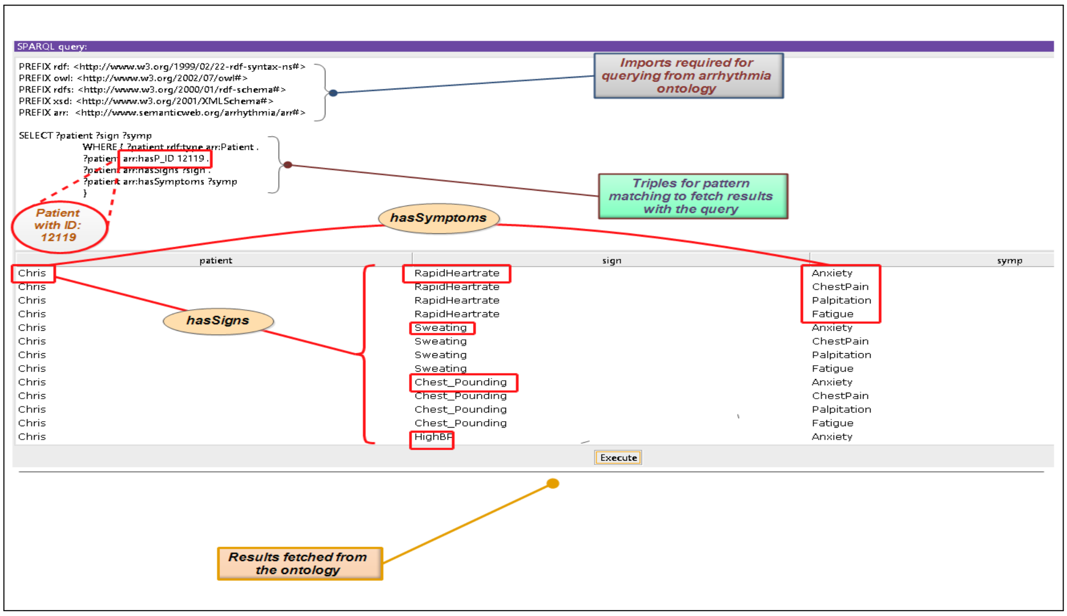


Figure 6.5: Query that displays a Patient's Symptoms and Signs

?patient arr:hasSymptoms ?symp

}

Explanation: While firing the above mentioned query, ontology analyst needs to find out all the signs and symptoms that a patient exhibits. The required information is extracted from the ontology with the input query. The output of the query is dependent on the patient ID that the analyst supplies in the query. Figure 6.5 shows the output of the query that generates a table listing signs and symptoms corresponding to a particular patient, uniquely identified by his/her patient ID.

SPARQL Query 2: Query to retrieve from ontology all Sample Heartbeats of a particular patient that are not normal beats, i.e., exhibiting irregularities or an arrhythmia state.

Query: *SELECT ?s ?b ?p ?o*

WHERE { ?p rdf:type arr:Patient .

?p arr: P_ID 12119 .

?r rdf:type arr:RecordingSession .

?o rdf:type arr:Observation .

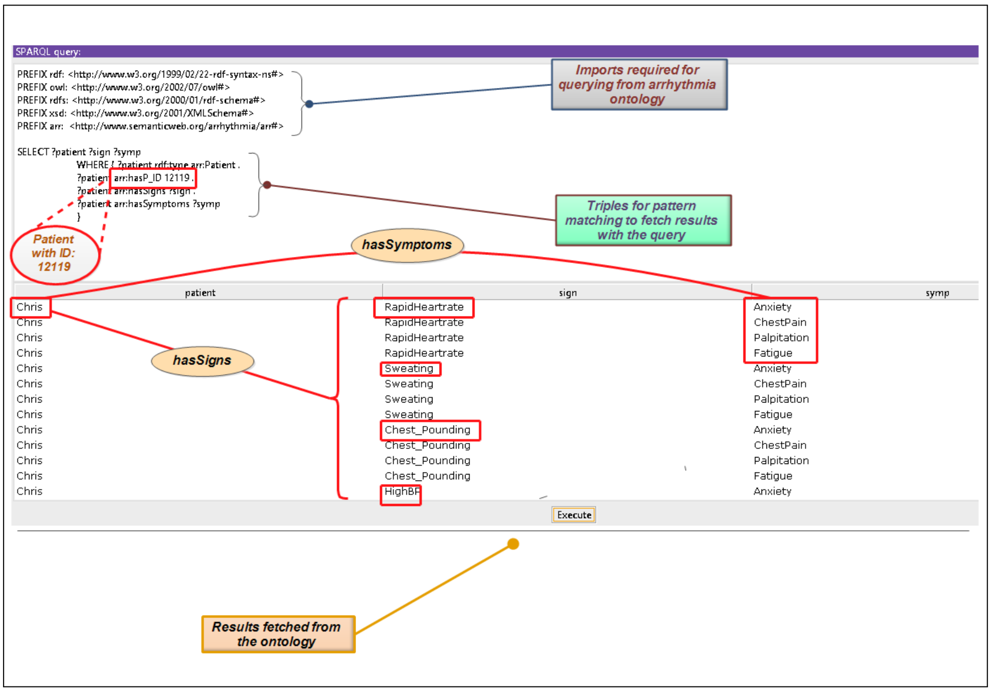


Figure 6.6: Query that shows Heartbeats with Arrhythmia Conditions

?p arr:hasSession ?r .

?r arr:hasObs ?o .

?o arr:hasSample ?s .

?s arr:hasBeatType ?b

FILTER NOT EXISTS {

?s arr:hasBeatType ?b

FILTER(regex(str(?b),"Normal"))

}

}

Explanation: The ontology analysts wishes to retrieve all the sample beats with respect to a patient with a given patient ID, that are classified by the ontology-based classifier to any of the arrhythmia type of a particular observation. Figure 6.6 shows the snapshot of the query result consisting of all the sample beats of a particular duration that are not normal beats.

DL Query 3: Query to list all the heartbeats that are classified into either of the two arrhythmia types:

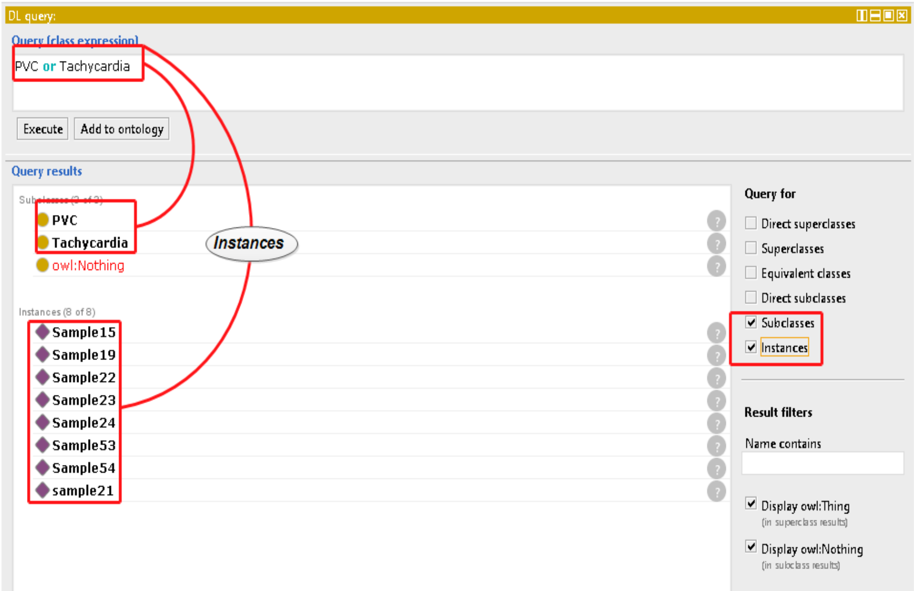


Figure 6.7: Query to list all instances of PVC or Tachycardia Arrhythmia

PVC or Tachycardia.

PVC or Tachycardia

Explanation: The ontology analyst, by firing this query, can retrieve all the instances of Sample Beats that have been classified to be of arrhythmia type PVC or Tachycardia. Figure 6.7 shows the result of the query consisting of all the instances of arrhythmic conditions: PVC or Tachycardia.

- **Overall System's Performance and Efficiency**

  In order to evaluate the overall efficiency of the proposed system, sensor-based ECG data corresponding to various patients/persons with varying conditions is used to make different types of inquiries to the proposed system. For this evaluation, the proposed ontology was instantiated with the entire set of heartbeat samples of 200^{th} record of the MIT BIH database that fall under a particular observation window and timeframe, belonging to a patient uniquely identified by the patient ID. The 200^{th} record is selected because it has wide spectrum of beat types which makes it suitable to showcase the performance of the system.

Additionally, the annotations by the experts are clearly defined for this record for validation of the classified arrhythmic type.

After the imports closure is fixed, the reasoner is activated, and based upon the rule base the inference is performed. The inferred triples are saved, and the ontology is updated. Additionally, a set of queries are formulated for each person/patient to retrieve semantic information attained by ontology classifier. Table 6.3 shows a subset of the results of arrhythmia classification performed on 12 distinct heartbeat samples belonging to 200^{th} record, ECG signal data of a particular patient, based upon the queries fired over the developed knowledge base. The 200^{th} record consists of approximately 2800 heartbeat samples, out of which approximately 1800 heartbeat samples are classified as normal beats and rest as arrhythmic beats.

In order to classify a heartbeat sample, its time elapsed from the beginning of the ECG recording is noted. The preprocessing phase is applied in which ecgpuwave, one of the tools of WFDB, is run, which segments the digital ECG signal into distinct heartbeats with time elapsed of each beat being associated with it. After the heartbeats are segmented, values of the selected features are extracted in time duration corresponding to each of the heartbeats in the recording. Then the feature values are further refined and assigned a feature label. Further, based upon the arrhythmic domain knowledge mapped into the SWRL rules, the heartbeat sample is classified into one of the 12 beat labels/arrhythmic classes. For example, when heartbeat sample 318^{th} is taken into consideration, ecgpuwave tool segments the entire 200^{th} record into well-defined heartbeat samples each having a unique timestamp. It also

Table 6.3: Performance of queries for Ontology-based Arrhythmia Classification

Beat No.	PR	QT	ST	RR	P	QRS	T	AHR	VHR	Type	Precision
Sample 55	0.119	0.361	0.217	0.786	0.100	0.144	0.197	500	160	A.Fib	100%
Sample 56	0.128	Ab.	Ab.	0.483	0.058	0.113	Ab.	120	100	Normal	100%
Sample 57	0.153	0.430	0.253	0.911	0.111	0.177	0.228	100	90	Normal	100%
Sample 58	0.189	0.370	0.267	0.539	0.134	0.103	0.103	90	65	Normal	100%
Sample 59	0.115	0.288	0.255	0.555	0.223	0.130	0.120	100	80	Normal	100%
Sample 60	0.097	0.400	0.357	1.5	0.010	0.05	0.146	400	140	A.Fib	100%
Sample 168	0.142	0.258	0.122	0.458	0.119	0.136	0.12	90	150	Ven. Tachycardia	100%
Sample 318	0.130	0.392	0.161	0.597	0.500	0.120	0.136	80	98	PAC	100%
Sample 367	0.162	0.411	0.261	0.892	0.62	0.150	0.140	85	71	PAC	100%
Sample 991	0.236	0.397	0.241	0.497	0.147	0.156	0.080	90	140	Ven. Tachycardia	100%
Sample 1171	Ab.	0.461	0.338	0.892	Ab.	0.123	0.305	100	80	PVC	100%
Sample 1173	Ab.	0.447	0.303	0.797	Ab.	0.144	0.285	70	65	PVC	100%
Sample 1175	Ab.	0.487	0.312	0.719	Ab.	0.175	0.264	100	60	PVC	100%
Sample 1181	Ab.	0.461	0.311	0.889	Ab.	0.15	0.291	95	55	PVC	100%
Sample 1703	0.136	0.605	0.541	2.147	0.116	0.064	0.111	100	200	Ven. Tachycardia	100%
Sample 2337	0.145	0.377	0.233	0.517	0.122	0.144	0.202	100	160	Ven. Tachycardia	100%
Avg.	0.109	0.384	0.260	0.823	0.141	0.129	0.170	138	100	-	100%
Std. Deviation	0.072	0.129	0.115	0.439	0.177	0.036	0.086	123	43.7	-	100%

Table 6.4: Execution times of different types of SPARQL Queries on proposed system

Type	Execution time(ms) w.r.t. Patient(Record)								Avg Time
	100	103	105	113	200	202	210	212	
DESCRIBE	0.210	0.199	0.198	0.192	0.187	0.200	0.190	0.180	0.192
SELECT	0.555	0.523	0.498	0.544	0.564	0.655	0.487	0.534	0.560
CONS.	0.476	0.456	0.455	0.423	0.454	0.443	0.420	0.400	0.446
GRAPH	0.623	0.599	0.344	0.599	0.578	0.610	0.621	0.585	0.574
ASK	0.324	0.312	0.298	0.364	0.299	0.376	0.255	0.322	0.321
FILTER	0.714	0.665	0.955	0.695	0.654	0.687	0.666	0.715	0.721

gives the value of onset and offset of the features. The time duration of wave-based features (P,T, QRS complex), interval-based features (PR, RR, ST,QT), and heartrate (AHR and VHR) is computed as shown in Table 6.3. Then the features are refined and assigned classes as: P-wave is elongated; PR-interval is normal; QRS complex is normal; VHR has value 98, which falls in between 60 and 100. Further, based upon the knowledge-driven SWRL rules, the sample 318, under consideration is classified into arrhythmic type PAC by the reasoner. The results in Table 6.3 show that the proposed model exhibits a high degree of precision level with respect to the queries fired at the knowledge base. For each of the heartbeat samples, the proposed model classifies into the correct class of arrhythmia. The verification and validation of this classification is done by comparing the annotations added by human experts in the physionet library. The comparison shows the exact match and 100% precision of our model has been achieved. Also, it can be noticed that the model gives correct results in case of some of the missing values of the parameters as well.

Additionally, another way to evaluate the overall efficiency of the proposed system is by designing and firing all the known SPARQL query forms on the developed ontological system, instantiated with a group of selected recordings of the MIT-BIH dataset identified as:

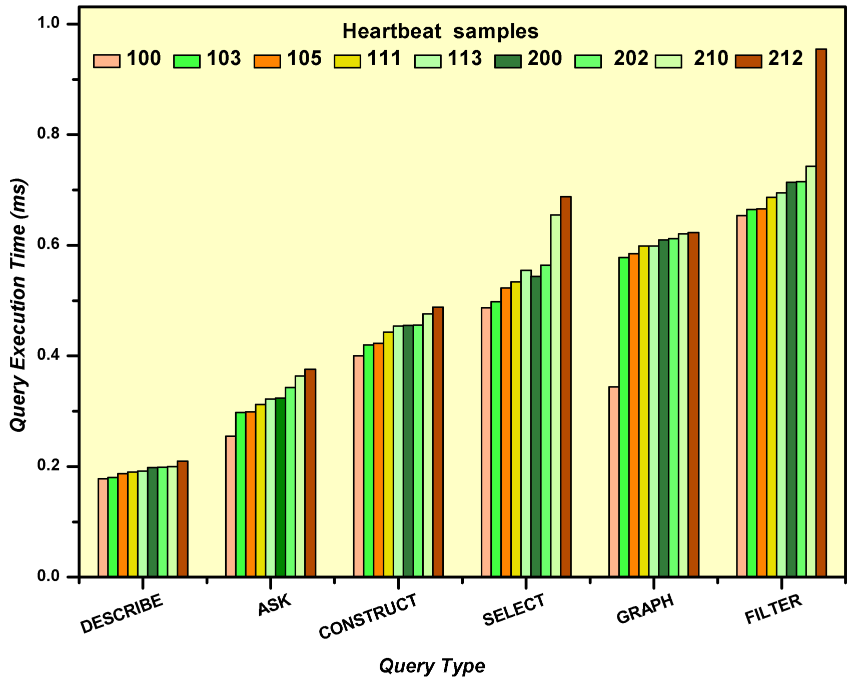
100^{th},103^{th},105^{th},111^{th},200^{th},202^{th},210^{th} and 212^{th}. Thus, the developed knowledge base is queried through all the query forms, i.e., Describe, Select, Construct, Graph, Ask and Filter and the respected query execution time are noted. Table 6.4 shows the query execution times along with the number of variables and triples involved in the designed query. The query execution times show that the knowledge base is developed in a way that is consistent and well structured in the design and implementation since the average execution time corresponding to all query forms is within a second. The query forms that fetch results in the form of a graph take longer than the ones which simply return tables. Figure 6.8a depicts the execution times in the form of bar graph corresponding to various query forms and different dataset records belonging to disinct patients.

Table 6.5 shows the beat-by-beat confusion matrix obtained by evaluating the proposed ontological approach, when the knowledge base is instantiated with a set of 9 recording of the MIT BIH dataset. The performance metrics are denoted by TP (True Positive), FP (False Positive), FN (False Negative) and TN (True Negative). The Positive label denotes arrhythmic classes and the negative label implies normal beats. For example, TP is the metric that depicts number of arrhythmic heartbeat samples, which are classified as arrhythmic beats by the proposed system as well.

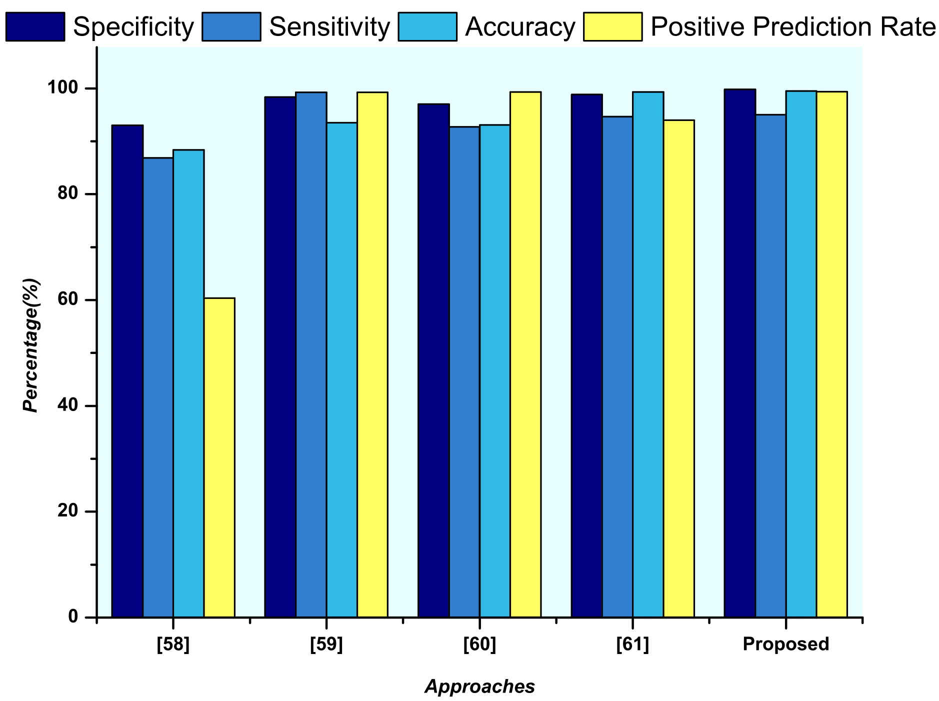
- **Comparison of proposed system with existing systems**

  The performance metrics such as precision, accuracy, recall and function measure are used to evaluate the performance of the proposed classification system[113]. Precision determines degree of closeness of two or more conclusions to one another, and accuracy shows the measure of closeness of a predicted value to a known value.

  In order to compare the performance of the proposed system with other existing approaches, the developed ontological knowledge base is



(a)



(b)

Figure 6.8: (a) Query Execution time w.r.t query type corresponding to different Heartbeat Samples (b) Comparison of existing approaches with proposed model based upon Performance Parameters

Table 6.5: Beat by beat Confusion matrix for Ontology-based Arrhythmia Classification

Record	TP	FP	FN	TN	No. of beats	Normal beats	Arrhythmic beats
100	789	3	4	2065	2861	2059	802
103	330	3	9	1640	1982	1650	332
105	460	88	50	2300	2480	2380	468
111	0	3	3	2855	2861	2859	2
200	940	2	13	1850	2805	1855	950
202	18	0	1	2019	2038	2020	18
210	178	8	1	2315	2502	2317	185
212	2	1	1	2857	2861	2859	2

instantiated. For this evaluation, the records 100 and 200 containing a distribution of both abnormal and normal beats were used. At first, the proposed ontology-based system was instantiated and the observational data was semantically annotated in accordance to the underlying model. Multiple queries that were relevant to arrhythmia detection and classification were fired at the system and based upon the information fetched, values of various performance metrics were calculated. Table 6.6 shows the comparison of the proposed approach with other existing AI-based diagnosis algorithms on the basis of the above said performance metrics. Table 6.7 shows the comparison results of the proposed approach with existing non-AI based engines based upon the accuracy values. Both the comparison results show, promising results with improved values of the performance metrics. For Table 6.8, records of the dataset were segregated into 10 folds, and a set of 20 queries were fired at the proposed system. The sample beats are classified among the set of 12 target classes. Table 6.8 shows the results of the metrics computed compared to other deep learning based existing approaches. Figure 6.8b depicts the computed performance metrics graphically in the form of a bar chart. The computed performance metrics are averaged to exhibit specificity, sensitivity, accuracy and positive predictive values to be 99.79%, 95%,

Table 6.6: Comparison of proposed approach with other existing Auto AI-based Diagnosis approaches

Author(s) and Year	Methodology	Class/Labels	Performance
Dokur et al.[114] 2001	Hybrid neural network	10	Acc: 95% Se: 96% PPV: 94%
Inan et al.[115] 2006	Wavelet transform and timing interval-based neural network	3	Acc: 95%
Sayadi et al.[116] 2010	Wave-based EKF bayesian filtering framework	3	Acc: 98.8% Sen: 98% Spec:97%
Martis et al.[117] 2012	PCA-based SVM with RBF kernel	3	Acc: 97.9% Sen: 98.9% Spec:98.8%
Martis et al.[118] 2013	Higher order statistics based SVM	3	Acc: 92.9% Sen: 99.27% Spec:98%
Li and Zhou[119] 2016	Wavelet packet entropy and random forest	5	Acc:94.1%
Mondjar-Guerra et al.[120] 2018	Ensemble of classifiers and various amplitude RF	5	Acc: 94% Sen: 66% Spec:70%
Proposed Approach	Automated Ontology-based SWRL classification	12	Acc: 99.49% Sen: 95% Spec:99.79% PPV: 99.35%

Table 6.7: Comparison of proposed approach with existing non-AI based Diagnosis approaches

Approach	Year	Labels	Accuracy
Mixture of Experts(MOE) [121]	1997	4	94%
Discrete Wavelet Transform (DWT) [122]	1999	10	95%
Discrete Fourier Transform (DFT) [123]	1999	10	89%
Unsupervised Soft Competitive Learning (USCL) [123]	2001	5	97.8%
Proposed approach	2021	12	99.49%

Table 6.8: Comparison of proposed system with existing Deep Learning-based Frameworks

Approach	SP(%)	Se (%)	Acc (%)	Pp (%)
Z. Zhang et al. [124]	93	86.82	88.34	60.36
R.J. Martis et al. [118]	98.31	99.27	93.48	99.25
M.A. Rahhal et al. [125]	99	92.7	93.1	99.3
P. Plawiak et al. [126]	99.5	94.62	99.37	94
Proposed Approach	99.79	95	99.49	99.35

99.49% and 99.3% respectively.

Chapter 7

INTERVAL TYPE-2 FUZZY LOGIC DRIVEN ONTOLOGICAL MODEL FOR HETEROGENEOUS SENSORS DATA

This chapter discusses in detail the design, development and implementation of Interval Type-2 Fuzzy Logic-based Ontological model with SPARQL-based querying for efficient fusion and analysis of heterogeneous sensors data. The resultant framework is instantiated with a real world data from heterogeneous water sensors to perform water quality detection. Additionally, the verification and validation process showcases the robustness, flexibility and consistency of the system. The comparison with other existing approaches shows performance enhancement in contrast to other techniques.

7.1 Need for Heterogeneous Sensor Data Fusion

The Internet of Things (IoT) has brought an immense revolution how "things" interact; hence generate huge amounts of data. The collection of sensor-based data manually and transporting them to a laboratory for further analysis is considered to be an old-fashioned and conventional approach that proves to be long-running in terms of time, expensive, vulnerable to mishandling of data causing fluctuations in the concentrations of the sampled data and continuous data collection, monitoring and analysis doesn't seem possible[127,128] . With the advancements in the sensor and network capabilities for data distribution over longer geographically distributed ranges and efficient storage provides a capable platform to utilize low cost, high performance and real-time monitoring Wireless Sensor Networks (WSN)[129] for modern business and industrial applications like water quality monitoring and management. The data collected and stored at various geographically distributed locations through the Wireless Sensor Network sensors needs to be processed and intelligently represented before any knowledge extraction technique or algorithm is applied over it. The task of processing of water sensors observations involves processing archival and historical data saved on durable databases, as well as real-time temporal incoming data. Hence, a flexible and efficient knowledge base model is needed to represent the consensual and derived domain knowledge. The researchers all over the world have integrated and worked upon different representational systems and schemes. Since, analysis of this large amount of data that is being produced by smart devices is a cumbersome and tedious process. This analysis can drive efficient decision making for business oriented applications. The approaches to efficiently manage and infer some knowledge out of this sensory data may be classified as data-driven[130,131] and knowledge-based[132] approaches. The methodology of the former approach is aimed at developing robust models by using statistical and machine learning techniques. The best strengths of this approach are their ability in effectively handling noise or incomplete sensor data[133]. Such models have shown good precision and accuracy in domains where semantics is not the key. However, there exists limitations in terms of need for training of data and the overhead of computational time and resources in the setup of dynamic environments and situations. Additionally, data-driven algorithms don't provide abstract reasoning mechanisms to infer the meaning of instances according to semantics. On the other hand, knowledge-based approaches improve interoperability and facilitate adaptation to various dynamic situations. The modern world applications are usually situational, have a dynamic flavour, continuously changing affected by the environment, not always known and taken into account in prior. Hence, these knowledge-driven methods prove to be of more advantage in comparison to data-driven models. Furthermore, features of knowledge-driven techniques are interesting in the applications which deploy a prediction system or recommender system along with the semantics to aid in the detection task, model the consensual knowledge (e.g. machine understandable/readable and easier to comprehend) and perform decision making through logical reasoning on the represented knowledge about future requirements, situations and actions. Some of the knowledge-based techniques are based on logic-driven approaches[134], rule-driven systems[135], and ontology models[136]. Though contemporary approaches are primarily ontology-based[137]. The ontological capability of promoting reuse and sharing of knowledge is the prime reason behind why ontologies are best suited knowledge system for modelling systems like water quality monitoring frameworks[138]. The state-of-the-art Wireless Sensor Networks have implemented Semantic technologies to not just achieve automated real-time monitoring of water bodies, but also provide enrichment to the modelling with semantics. Different intelligent real-time Water Quality Management systems are implemented and currently working, be it centrally managed[139] or distributed nature on sensor nodes[140]. At the decision level, most recommendation or prediction systems deploy classical ontology for knowledge representation and analysis. However, this classical ontology addresses only the certain and crisp data but does not provide profitable results from blurred and uncertain data. To address this critical issue, researchers have integrated fuzzy logic system[141] with crisp ontologies to propose precise solutions to get better results. The combination of both technologies has been proposed in recent years to better answer the user's queries[142]. However, Type-1 fuzzy ontology-based systems are inadequate and can extract relevant knowledge only to a limited extent. Additionally, the dynamic environment is so fuzzy and quality prediction is so swarmed with uncertainties and blurredness that these systems cannot create membership functions that can model in a precise manner. Therefore, the implementation of type-2 fuzzy logic (T2FL) with classical crisp ontology is the preferable option for water quality prediction.

7.2 Background and Preliminaries

7.2.1 Type-2 Fuzzy Logic

Various decision making applications apply fuzzy logic as their core component to accommodate uncertainty for effective decision making. Lofti Zadeh developed and introduced the concept of fuzzy logic in 1965 to extract vague and blurred concepts [143]. It was first developed and deployed as a controller for a steam engine [144] . The type-1 fuzzy set (T1FS) was extended to type-2 fuzzy set (T2FS) by Zadeh ten years later in 1975 [145] . In a type-2 fuzzy set, the membership function gives a fuzzy set for every element of the domain, so it can be seen as a fuzzy-fuzzy set [146]. The Type-2 Fuzzy System is bounded by two membership functions: a lower membership function (LMF) and an upper membership function (UMF) denoted by $\underline{\mu_{\tilde{A}}}(x)|\forall x \in X$ and $\overline{\mu_{\tilde{A}}}(x)|\forall x \in X$ respectively. The bounded region is called footprint of uncertainty which is the union of all the primary memberships i.e.

$$FOU(\tilde{A}) = \bigcup_{x \in X} J_x = (x, \mu) : \mu \in J_x \subseteq [0, 1] \tag{7.1}$$

The lower and upper membership functions are denoted as:

$$\underline{\mu_{\tilde{A}}}(x) = \underline{FOU(\tilde{A})}|\forall x \in X \tag{7.2}$$

and

$$\overline{\mu_{\tilde{A}}}(x) = \overline{FOU(\tilde{A})}|\forall x \in X \tag{7.3}$$

In case of interval type 2 fuzzy sets (IT2 FS),

$$J_x = [\underline{\mu_{\tilde{A}}}(x), \overline{\mu_{\tilde{A}}}(x)] \forall x \in X \tag{7.4}$$

7.2.2 Fuzzy Ontology

In the world of Semantic technology, ontologies form the core technology and component of the data, which can be shared and reused by researchers all over the world due to their consensual knowledge representation and formalism. An ontology models the relationship between the concepts in a given domain. An ontology is defined as "An ontology is an explicit specification of a conceptualization"[57]. In various application areas, the knowledge that can play a significant role in the decision making process is available in the form of an ontology. However, classical (crisp) ontologies are not sufficient and appropriate to represent uncertain and vague knowledge[147] . In order to overcome this issue, the concept of fuzzification of ontology has been introduced. In simpler terms, a fuzzy ontology can be defined as "an ontology which uses fuzzy logic to allow representation of imprecise and vague information and knowledge to ease reasoning over the formalised data"[148]. Recently, combinations of fuzzy sets and ontologies have been implemented by researchers. Also, it has been asserted that type-2 fuzzy logic-based ontological systems are more efficient in dealing with hazy and uncertain information than type-1 fuzzy ontologies[149,150]. In Type-1 Fuzzy System, fuzziness is translated as a crisp membership value $\in$ [0,1]. But, Type-1 Fuzzy System cannot completely represent the fuzziness of these fuzzy variables. Therefore, this drives the motivation to use Type-2 Fuzzy System instead of Type-1 Fuzzy System. An Interval Type-2 Fuzzy System is characterized by another fuzzy set i.e. membership value of each fuzzy set is a fuzzy set in [0,1], instead of being a crisp value. Each fuzzy set is bounded by an interval created by two membership functions: lower membership function

(LMF) and upper membership function (UMF) .They denote the maximum and minimum value of each μ for each x.The footprint of unceratinity with respect to the parameter is the area of uncertainity. It is the aggregation or union of all the primary memberships. This shows uncertainity associated with each value of x belonging to the universe of discourse. The proposed system uses triangular membership function and shoulder function based on 3 points (a,b,c) as shown in Equations (7.5) – (7.7).

$$Triangular(x, a, b, c) = max(0, min(y_1, y_2, 1)) \tag{7.5}$$

where, $y_1 = \frac{x-a}{b-a}$ and $y_2 = \frac{c-x}{c-b}$

$$LeftShoulder(x, a, b) = max(0, min(z_1, 1)) \tag{7.6}$$

where, $z_1 = \frac{x}{a-b} - b$

$$RightShoulder(x, a, b) = max(0, min(z_2, 1)) \tag{7.7}$$

where, $z_2 = \frac{x}{b-a} + a$

7.3 Design architecture of the Proposed System

An ontology driven semantically intelligent system based on interval type-2 fuzzy reasoning for water quality prediction from observations captured by heterogeneous sensors is proposed. Figure 7.1 shows a functional diagram of the proposed system. The data coming from multiple water sensors is captured and refined to get rid of noisy and irrelevant data records for efficient analysis. The preprocessed temporal data is semantically annotated with Water Quality Data ontology. The sensor measurements are intelligently classified by interval type-2 fuzzy logic-based inference mechanism to determine the water quality keeping in mind the fuzziness and uncertainty associated with the sensor data. The classified instances determining the quality of water sample are stored in the ontology, hence the ontology can be queried for information retrieval for effective decision making6. Figure 7.2 shows the scheduling of activities in the proposed system. The sequences of steps that constitute the functionality of the proposed system are as follows:

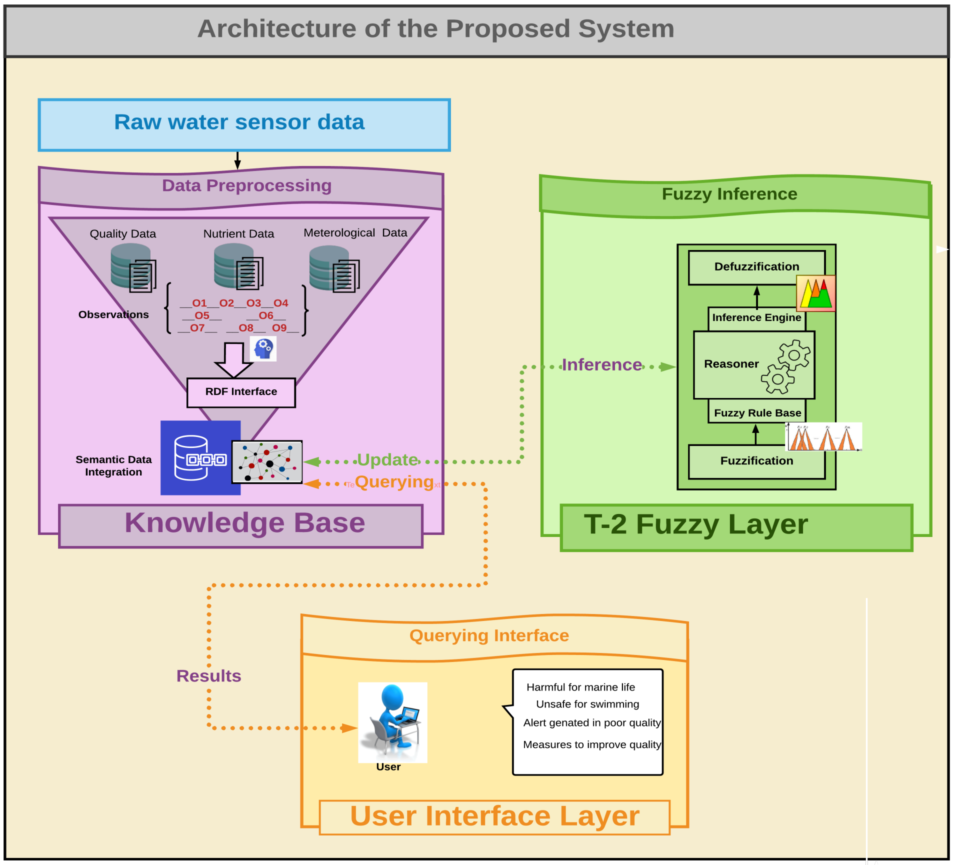


Figure 7.1: Architecture of Proposed System

7.3.1 Water Sensor Data Preprocessing

In this first step, the temporal data from heterogeneous water sensors is improved in terms of its quality. The data is refined by performing preprocessing tasks like: removing outliers and noisy data, dealing with missing and redundant values, encoding of categorical and nominal attributes. The records with missing sensor values are substituted with values in the previous measurement interval. The records with missing timestamp are removed to get rid of noisy data. In order to uniquely identify every sensor sample in the dataset, each row is given an identifier. The station code is appended with the sample number.

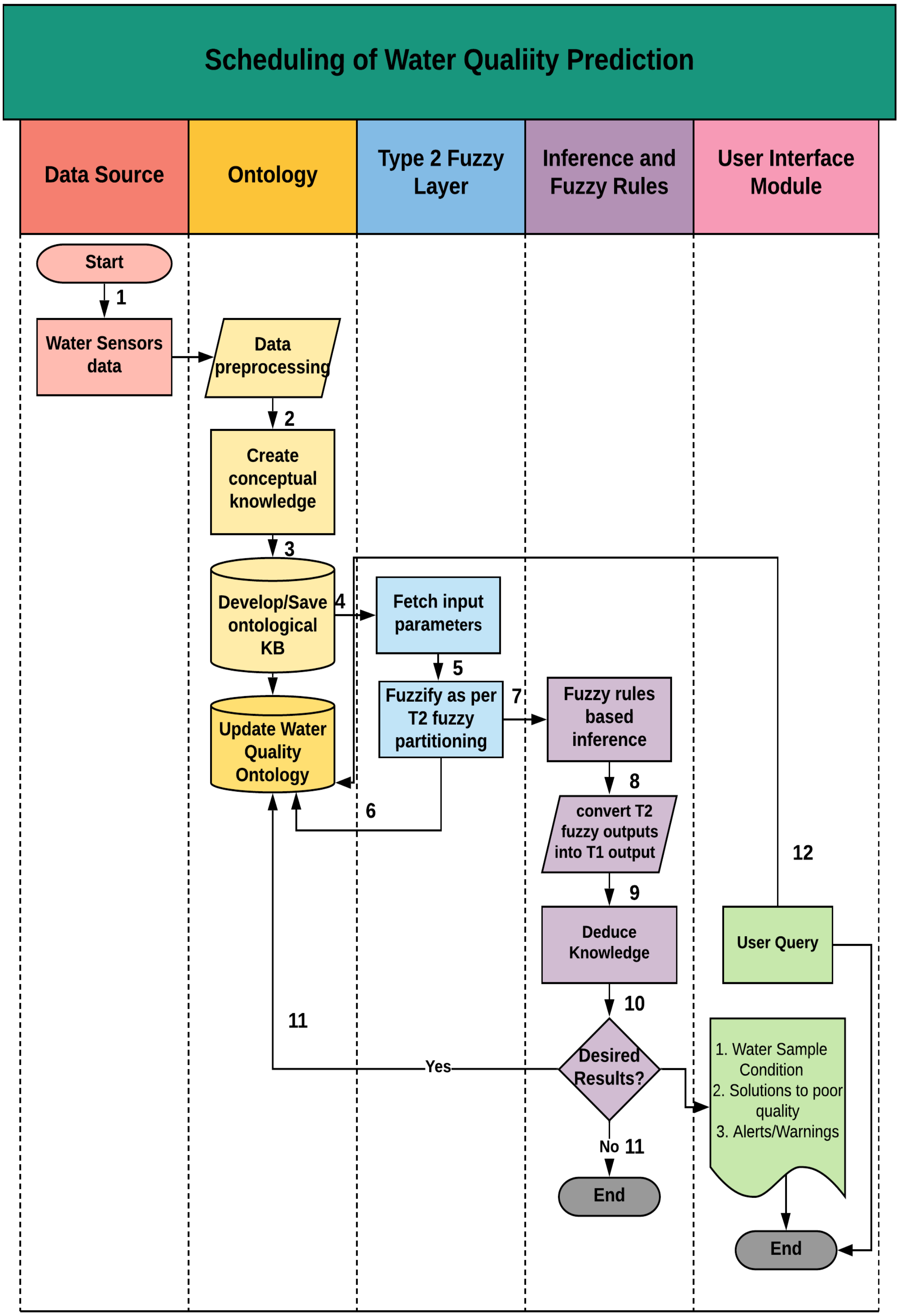


Figure 7.2: Scheduling of Activities involved in Water Quality Prediction System

7.3.2 Fuzzy Linguistic Variables and Fuzzy Partitioning of Water Quality Parameters

This step involves choosing the water quality parameters which are crucial and affect the accuracy of water quality prediction due to the fuzziness and uncertainty they hold within them are considered to be fuzzy sets or variables. These variables are of prime importance and affect the interpretability of the proposed system [151]. Further, these fuzzy sets undergo fuzzy partitioning on the basis of the variable range they take. Each fuzzy partition of the fuzzy set is known to be fuzzy linguistic variable, for instance, very low, low, moderate, high, very high. During the reasoning mechanism, the partitioned fuzzy sets are referred to by fuzzy linguistic terms and correspond to numerical values. In type-1 fuzzy system logic, each input value is mapped to a crisp membership value $\in$[0,1].The proposed system, on the other hand, deploys type-2 fuzzy system FS logic, assigning an interval i.e. [LMF, UMF] to each input $x_i \in X$,where X denotes the input space or the universe of discourse. The upper bounded and lower bounded values of membership function are called UMF and LMF respectively. The process of creating interpretable fuzzy partitions, requires the definition of fuzzy sets satisfying at least the following constraints [152] :

- ***Distinguishability:*** The fulfillment of semantic integrity requires all membership functions, representing different linguistic concepts, as being easily understandable and readable.

- ***A justifiable number of fuzzy sets:*** The number of fuzzy sets corresponding to each attribute is typically maintained between two to nine owing to the limitations of human capacity to process too much of information.

- ***Uniformity of coverage:***Each input data point x_i should belong significantly, at least to one fuzzy set, and at most to two fuzzy sets.

- ***Overlapping and orthogonality:*** It is significant that only two adjacent fuzzy sets may overlap.

The task of fuzzy partitioning is performed in the following steps:

1. As an initial step, as per the domain knowledge gathered through the literature survey and various water regulation reports published by reputable organizations, non-uniform partitions are selected which segregate the input space of the variable into intervals.

2. Each fuzzy partition is associated with a membership function i.e. triangular MF. The selected MF bounds the membership values of the variables belonging to the fuzzy partition. The MF provides an interval with the boundary values $\in$ [0,1]. Each value of the input domain within the corresponding fuzzy partition is mapped to an interval given by lower membership value (LMF) and upper membership value (UMF). For each such fuzzy set, there exists a middle value of its corresponding interval. This produces triangular shape with respect to both lower membership values and upper membership values.

3. This partitioning constitutes of c fuzzy sets and is represented as: (x_c, y_{cL}, y_{cU}), where c is the number of input points and x_c is a c-dimensional vector denoting domain of the input parameter, x_c. With each partition, the fuzzy set contains a mid value i.e. $x_{(c,mid)}$.

4. If new water regulation parameter knowledge is gathered, then above steps can be iterated and the fuzzy partitioning is updated along with the new fuzzy set being assigned to each partition.

The proposed system can be modeled as a function displayed in the equation (7.8):

$$f : \mathcal{X} \longrightarrow \omega;\ \mathcal{X} = \mathcal{X}_1\ X\ \mathcal{X}_2\ X ... X\ \mathcal{X}_n\ and\, \omega = y_1, y_2, ..., y_m \tag{7.8}$$

Where, $\mathcal{X}$ refers to the universe of discourse i.e. the set of water quality parameters and ω denotes the set of output class variables i.e. the water quality classes. The designing of fuzzy linguistic variables for each parameter is a first and foremost step in implementation of the system. Each of the n input water quality parameters is represented as in equation (7.9):

$$Q_j = (q_j, X_j, V_j, L_j, U_j, F_j) \tag{7.9}$$

Where, q_j denoted the name of the quality parameter, X_j refers to the domain of values taken up by the quality parameter, $V_j = v_{j1}, v_{j2}, \ldots, v_{jc}$ represents the set of total c fuzzy linguistic variables associated with the input quality parameter, $L_j = l_{j1}, l_{j2}, \ldots, l_{jk}$ consists of ordered pairs denoting fuzzy sets of lower membership grades on set X_j, $L_j : X_j \longrightarrow [0,1]$, $U_j = u_{j1}, u_{j2}, \ldots, u_{jk}$ is a set of ordered pairs representing fuzzy sets of upper membership grades on input space $X_j, U_j : X_j \longrightarrow [0,1]$, and F_j associates a bounded interval i.e. $[L_j, U_j]$ with each fuzzy linguistic variable i.e. V_j.

The membership function is a mapping function that maps an input value from set X_j to a value $\in$ [0,1]. The efficiency of the proposed system also depends on the choice of the fuzzy membership function implemented[153]. So, it becomes essential to implement functions that are computationally efficient to generate powerful fuzzy sets and fuzzy partitions.

The LMF and UMF represented as triangular MF's as displayed in equation (7.10) and (7.11) respectively, are implemented in our work:

$$LMF = f(x, L_L, m_L, R_L) = \begin{cases} 0; & x \leq L_L \\ \frac{\beta(x - L_L)}{m_L - L_L}; & L_L \leq x \leq m_L \\ \frac{\beta(R_L - x)}{R_L - m_L}; & m_L \leq x \leq R_L \\ 0; & R_L \leq x \end{cases} \tag{7.10}$$

$$UMF = f(x, L_U, m_U, R_U) = \begin{cases} 0; & x \leq L_U \\ \frac{(x-L_U)}{m_U-L_U}; & L_U \leq x \leq m_U \\ \frac{(R_U-x)}{R_U-m_U}; & m_U \leq x \leq R_U \\ 0; & R_U \leq x \end{cases} \tag{7.11}$$

Alternatively, the triangular MF can be represented in the form of min-max operators in equation (7.12):

$$f(x, L, m, R) = max(min(\frac{x-L}{m-L}, \frac{R-x}{R-m}), 0) \tag{7.12}$$

The overlap between adjacent membership functions is depicted in the figure 7.3. Also, abiding by the criteria of fuzzy partitioning mentioned above, no more than two membership functions can overlap with each other.

Algorithm 13 Fuzzification of Water Quality Ontology

Require: Set I of input/output parameters, F of fuzzy sets for each input/output and M of membership functions.

Ensure: Fuzzified Ontology ($WQstatus_{Onto}$) output file in manchester OWL format.

```
for each param_i ∈ I do
    Instantiate OWLClass with param_i
    Map IRI to param_i
    for each Fuzzy_lv ∈ F do
        Add OWLObjectProperty wq:hasFLV(param_i, wq : Fuzzy_lv)
        Evaluate membership function value i.e. literal_umf & literal_lmf
        Add OWLDataProperty wq:haslmf(wq:Fuzzy_lv, literal_lmf)
        Add OWLDataProperty wq:hasumf(wq:Fuzzy_lv, literal_umf)
    end for
end for
call saveOntology
```

7.3.3 Development of Semantic Knowledge Base

The semantic knowledge base development constitutes of two developed ontologies. The developed ontologies consists of terminal knowledge (TBox) and assertion knowledge (ABox) combined with role centric knowledge (RBox). The task of ontology development is done on the basis of design criteria defined by Gruber[58]. This step constitutes the development of two ontologies which model the concepts of water quality status and semantically annotates the sensor observations of the water sensors. Algorithm 13 depicts the fuzzification of concepts of Water Quality Ontology. The Water Sensor Observations ontology is developed to model the temporal and spatial observations captured by the water sensors deployed at various geographically separated locations. The Water Quality Ontology is a fuzzy type 2 ontology developed to model the domain knowledge in the form of water quality status parameters and represent their fuzziness. It constitutes semantically annotated consensual domain knowledge with a set of axioms defined over the concepts and relationships between them. Both the ontologies are developed using Ontology Web Language (OWL2). Also, Protégé OWL, an open source editor and tool, is used to edit the proposed ontology. Further, this step is accomplished as follows:

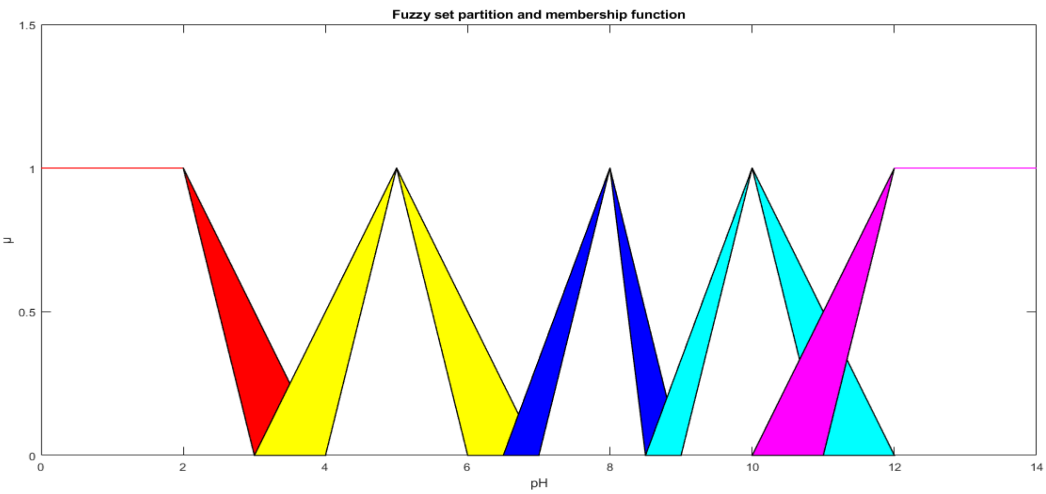


Figure 7.3: Fuzzy Membership Function for pH

Algorithm 14 Ontological Instantiation

Require: Set of water sensor observations with M monitoring stations.
Ensure: Instantiated ontology with observational water sensor data.

Load ontology '$Output'_{Onto}$ with OWLOntologyManager object
Instantiate BufferedReader class with path+monitoringstation name
Initialize global variables
Instantiate DataFactory class
for each $monStation_i \in$ W **do**
    Add datatype axiom for monitoring station code
    for each $sample_{loc} \in L$ **do**
        Add datatype axiom for sample location code
        for each $obs_j \in O$ **do**
            Add datatype axioms for start and end timestamp
            for each $sensor_i \in S$ **do**
                Add feature of interest axiom
                Add observed property axiom
                Add datatype axiom
                hasSensorValue($sensor_i$,wso:SensorResult)
            end for
        end for
    end for
end for
call saveOntology with args

- **Water Sensor Observations Ontology:** It is a measurement ontology which consists of information about the geographically distributed water sensors and their specifications, along with the timestamped data captured by them. Figure 7.4 graphically represents the structure of the water sensor observations ontology. It enriches the heterogeneous sensor data semantically and gives significance and meaning to the observational data. This ontology is uniquely identified by a namespace denoted as "senObs:".The classes that form the gist of this ontology are represented in Table 7.1. Algorithm 14 depicts the process of ontology instantiation with sensor observations.

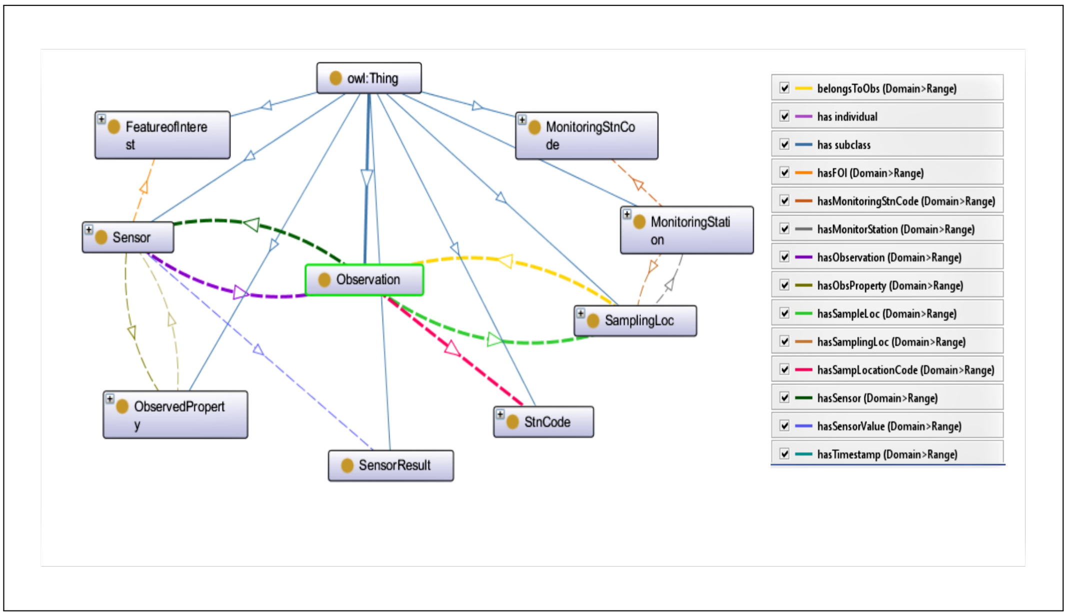


Figure 7.4: Water Sensor Observation Ontology

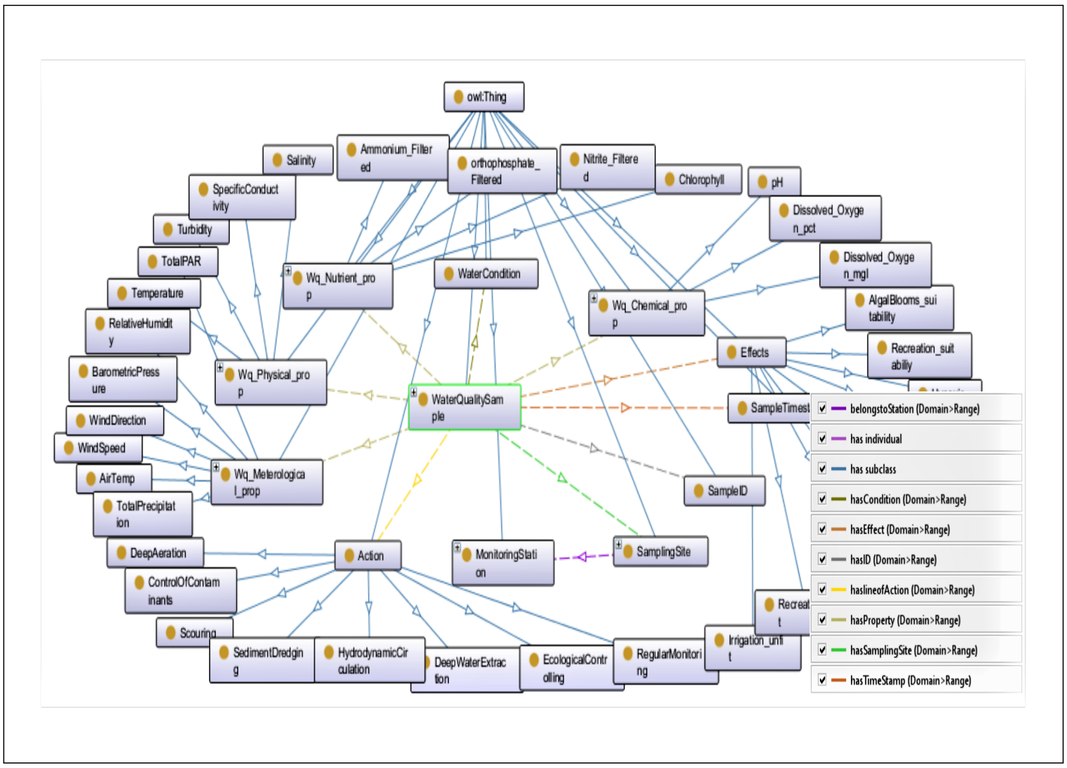


Figure 7.5: Water Quality Ontology

Table 7.1: Entities of Water Sensor Ontology

Classes	Description
senObs:Sensor	Represents the concept that provides a measurement upon change in corresponding stimuli
senObs:Observation	Models a concept that has rows of observations uniquely identified by a timestamp and has sensor results of different sensors
senObs:SensorResult	Models sensor output with respect to each instance of class Sensor
senObs:ObservedProperty	Denotes the class that models the property sensed by each sensor
senObs:FeatureofInterest	Represents the class that models the feature around which the sensor observations revolve
senObs:MonitoringStation	Models the entity that represents the sensor monitoring station that includes various sampling locations for each sensor deployment
senObs:MonitoringStnCode	Represents the class that encapsulated the unique ID for each monitoring station
senObs:SamplingLoc	Models the class that models the location for each sensor deployment
senObs:StnCode	Represents the entity that covers ID for each sensor sampling location
senObs:TimeStamp	Denotes the class that models time and date of sensor sampling

These classes are linked through two types of properties namely, owl:ObjectProperty and owl:DatatypeProperty. This forms the Tbox portion of the developed ontology. The instantiation of the ontology gives the ABox statements and the roles and axioms form the RBox.

- **Water Quality Ontology:** Water Quality Ontology formally represents the knowledge regarding the parameters that are crucial and defines certain axioms-based upon which the water quality of a particular sample/observation is predicted. Figure 7.5 graphically depicts the structure of the Water Quality Ontology. The proposed ontology contains crisp classes as concepts, data types, fuzzy data types, data and object properties, individuals and fuzzy axioms. Fuzzy data types describe the interval of the membership degrees. In order to create an efficient ontology, the proposed ontology is first developed as a crisp ontology, and then fuzzy terms are added into the ontology through Fuzzy OWL2[154].

This ontology is uniquely identified by a namespace denoted as “wq:” The core classes are linked through relationships modeled as properties (objectproperty and datatype property). The crisp ontology for water quality detection is fuzzified with fuzzy sets. Figure 7.6 shows graphical representation of the fuzzified ontology.

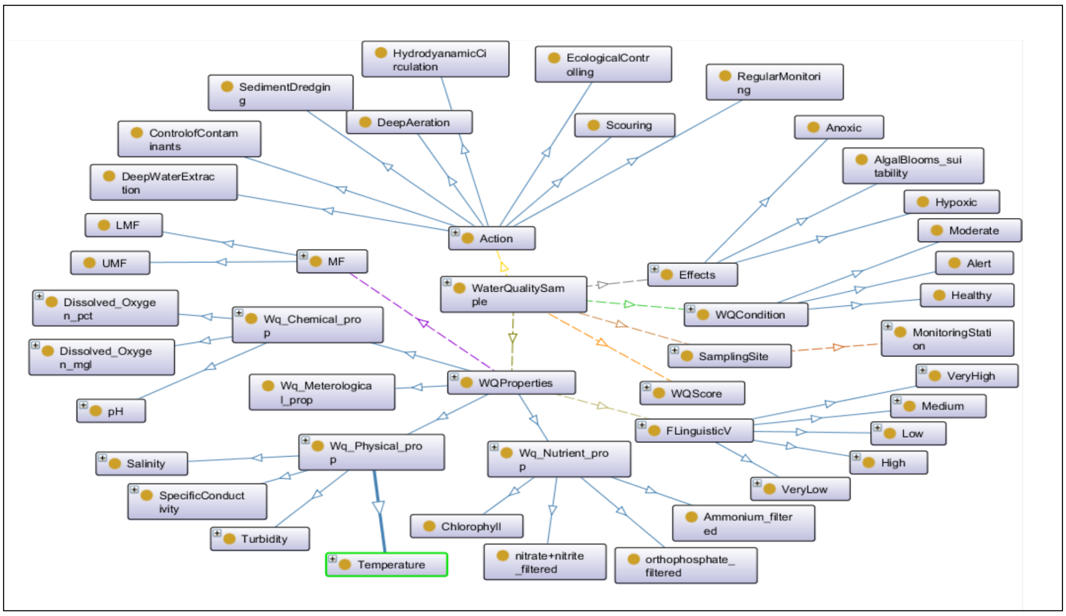


Figure 7.6: Fuzzy Water Quality Ontology

7.3.4 Rules and Inference Mechanism

The domain knowledge and fuzzy rule base form the core component of ontology-based interval type-2 driven inference system. In the proposed system, fuzzy rules over the Tbox statements are designed and created to

drive the inference mechanism. The rules have various input variables like temperature, turbidity, DO_mgl, salinity, pH, specific conductivity, Nitrate+Nitrite, orthophosphate, chlorophyll and output variable as the water sample quality condition. After the variables are fuzzified and fuzzy sets are created, these rules are extracted from the knowledge base to act upon the instantiated fuzzy interval membership data.

Algorithm 15 Interval Type-2 Fuzzy Inference Mechanism

Require: Set of 'W' water quality sensor samples, Set of 'S' sensors, Set 'M' of membership functions and set 'R' of fuzzy rules.

Ensure: The computed water quality output value, effects and line of actions w.r.t. samples.

Load water quality ontology file
Import pyIT2FLS and numpy
Instantiate IT2FLS():wqIT2
Set $domain_{out}$:linspace(a,b,c) for wq:WaterCondition class//a,b,c $\in$ Z: set of integers
for each mf_p $\in$ M **do**// mf_p can be LMF and UMF: lower and upper membership function
    Initialize and define type mf_p
end for
Add class wq:waterCondition as output variable to wqIT2
for each $sample_i \in W$ **do**// Read all water samples one by one for classification
    for each $inputvar_j \in S$ **do**
        Set $domain_j$: linspace(d,e,f) // d,e,f $\in$ Z: set of integers
        Call add_input_variable() w.r.t wqIT2
    end for
    for each $rule_k \in R$ **do**
        if $rule_k \notin$ R **then**
            Call add_rule();
        end if
    end for
    Instantiate the operators w.r.t wqIT2
    Implication: "min_t_norm"
    Aggregation: "max_s_norm"
    Type Reduction: "karnikmendel"
    Defuzzification: "centroid"
    Add the operators to wqIT2
    Call evaluate() //evaluate is a predefined classifier function
    Calculate crisp output C_{sample_i}
    Add axiom hasWQscore(wq:WaterQualitySample,C_{sample_i})
    Call Update_Inference() // Using Algorithm 16
end for

The result of this process will be the condition of the water sample which is assigned to the corresponding instance in the ontology. The rule base consists of multiple fuzzy rules and crisp rules. The fuzzy rules consisting of various input variables and a single output variable along with their linguistic values is generalized as:

Algorithm 16 Function to update fuzzy Water Quality Ontology

Require: $WQstatus_{Onto}$ ontology file, set 'W' of water samples, set 'F'of computed fuzzy output, Set 'E' of Effects and 'A' of Actions.

Ensure: Fuzzified Inferred Ontology ($WQstatus_{Onto}$) output file in manchester OWL format.

1. Function Update_Inference($WQstatus_{Onto}$, F<integer>, E <string>, A<string>)
Load water quality ontologyy
Add axiom wq:hasWQscore($Sample_i$, C_{sample_i}) // $Sample_i$ ∈ W ; C_{sample_i}) ∈ F
if C_{sample_i} ∈ [0,4] **then**
    Add axiom wq:hasCondition($Sample_i$, "healthy")
end if
if C_{sample_i} ∈ [2,6] **then**
    Add axiom wq:hasCondition($Sample_i$, "moderate")
end if
if C_{sample_i} ∈ [5.5,10] **then**
    Add axiom wq:hasCondition($Sample_i$, "Alert_Condition")
end if
Add axiom wq:haslineofAction($Sample_i$, A_m) ; m=1 to all actions
Add axiom wq:hasEffect($Sample_i$, E_n ; n=1 to all effects
Save water quality ontology
End Function

$$Fuzzy\,Rule_n : IF \underbrace{X_1}_{Input_1} \text{ is } v_{1x} \text{ T } \underbrace{X_2}_{Input_2} \text{ is } v_{2x} \text{ T } \underbrace{X_3}_{Input_3} \text{ is } v_{3x} \text{ T... T } THEN \underbrace{\omega}_{Output} \text{ is } y_m$$

Figure 7.7 shows some rules that a part of the rule base of the proposed fuzzy ontology. After the inference task from the well defined rules is performed, the type 2 fuzzy output is reduced into type 1 fuzzy output. This mapping is performed through Karnik–Mendel Algorithm (KMA) [155]. This algorithm when implemented results in type 1 fuzzy result interval representing the water sample quality condition.

Further, the end points of the reduced fuzzy interval is averaged out to compute a numerical value that corresponds to the final condition of the water sample in terms of quality. After the inference mechanism is performed, the instantiated water sample is annotated with a new property defining its quality condition for information extraction purposes. Algorithm 15 represents the pseudo code of the fuzzy output computation from sensor samples and observations.

7.3.5 Querying and Information Retrieval

The developed Semantic Knowledge Base can be queried for instantiated data and information can be retrieved with respect to water samples. The inferencing is further carried out as per the steps mentioned above to come to a conclusion regarding the water sample condition. The ontology is then annotated to add new properties defining the sample quality condition. The users/agents can query the ontologies developed to gain information to predict quality of the sample under consideration.

7.4 Experimental Results and Discussion

This step discusses the experimental setup and evaluation of the proposed system. The proposed system is implemented using OWL2 and RDF4j framework for the knowledge base. It also uses matlab tool for computation of sample quality condition value. It deploys DL and SPARQL query to extract information by the user/agent.

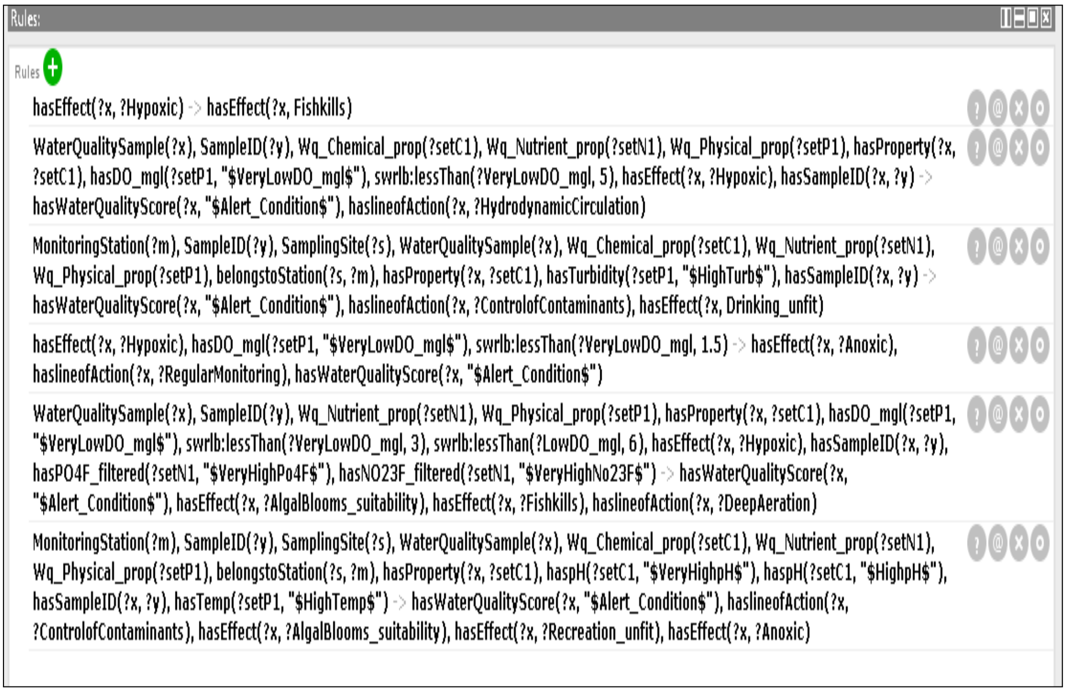


Figure 7.7: Fuzzy Rule Base

Table 7.2: Dataset Parameters

Type	Parameters
Quality	Temperature, water depth, salinity, pH, dissolved oxygen, and turbidity
Meteorological	Air temperature, wind speed and direction, relative humidity, barometric pressure, rainfall, and photosynthetically active radiation (PAR)
Nutrient	Ammonium, nitrate, nitrite, and ortho-phosphate.

7.4.1 Initialization and Instantiation

The implementation of the proposed system required developing two ontologies: Water Sensor Observations Ontology and Water Quality Ontology. The ontology modeling the water quality aspects is a type 2 fuzzy ontology. The knowledge base is developed with OWL2, RDF, RDFS and RDF4j framework.

The inference mechanism classifies the water sample based upon well defined fuzzy rules into a class that signifies its quality. Table 7.3 shows the tools and technology used in the implementation of the proposed model. Let's say we wish to predict the quality of the water sample identified by the ID “acebbwq11885_14:00” with timestamp “09/04/2004 14:00:00”. The following steps are involved in the desired task:

Table 7.3: Tools and Technologies used

Component	Tools and Technologies used
Ontological Layer	RDF4j, Protégé
Fuzzy Inference	pyIT2FS
Querying and Retrieval	SPARQL, DL query

- **Fuzzification and Ontology Instantiation:** Firstly, the sensor sample is semantically annotated in correspondence to the Water Sensor Observations Ontology as shown in figure 7.8. The values of quality, nutritional and meterological parameters are linked with the water sample. Secondly, the fuzzification of the sensor sample “acebbwq11885_14:00” is performed in correspondence to the Water Quality Ontology. The values of all the parameters are fuzzified according to the designed fuzzy partitions.

  So, sample under consideration,“Sample11885”, is linked to numerical values of the parameters with their respective data properties.The fuzzification of the “Sample11885” results in the computation of membership values as per the LMF and UMF. This gives an interval representing the membership degrees of a parametric value belonging to the corresponding linguistic variable. Similarly, each of the parameter value is assigned a linguistic variable and an interval computed from the lower and upper membership functions.

  Figure 7.9 shows a snapshot of instantiation and computation process. Table 7.4 shows the fuzzy set partitioning parameters for LMF and UMF computation.

- **Inference Mechanism:**The inference task is performed based upon the fuzzy rule base defined in the proposed system. The “Sample11885” is classified into the class “Alert_Condition” based upon the mamdani min-max system as part of the proposed system. After the sample is inferred, the ontology is updated, the considered sample to be in the "Alert_Condition". Figure 7.10 shows computation of the fuzzy output.

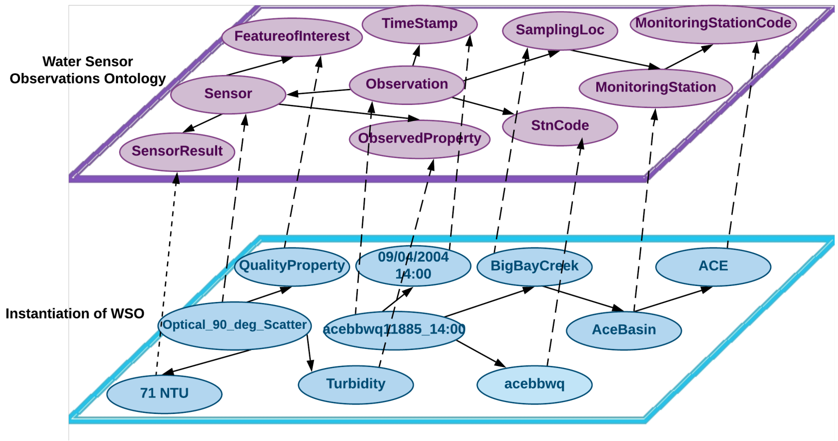


Figure 7.9: Water Quality Ontology with Sensor Data Instance

- **Querying and Information Retrieval:** A SPARQL/DL query can be fired over the ontology fetching details about the sample and the water quality condition so that the required action can be taken in order to avoid any harm to mankind or marine animals due to the toxicity of the water at that particular instant.

7.4.2 Result Evaluation

The experimental evaluation is done by annotating the observations captured by water sensors and performing analysis based on the proposed system. The experimental setup is implemented on Intel Core processor with 8 GB RAM and windows 10 operating system.

Once the system is implemented, it is of utmost importance to test and compare the system's accuracy, consistency, and coherency for overall system evaluation. Hence, the evaluation process is segregated into 2 distinct steps, each with a different evaluation criteria: ontology testing and validation and evaluation of the proposed system's overall performance, efficiency and accuracy.

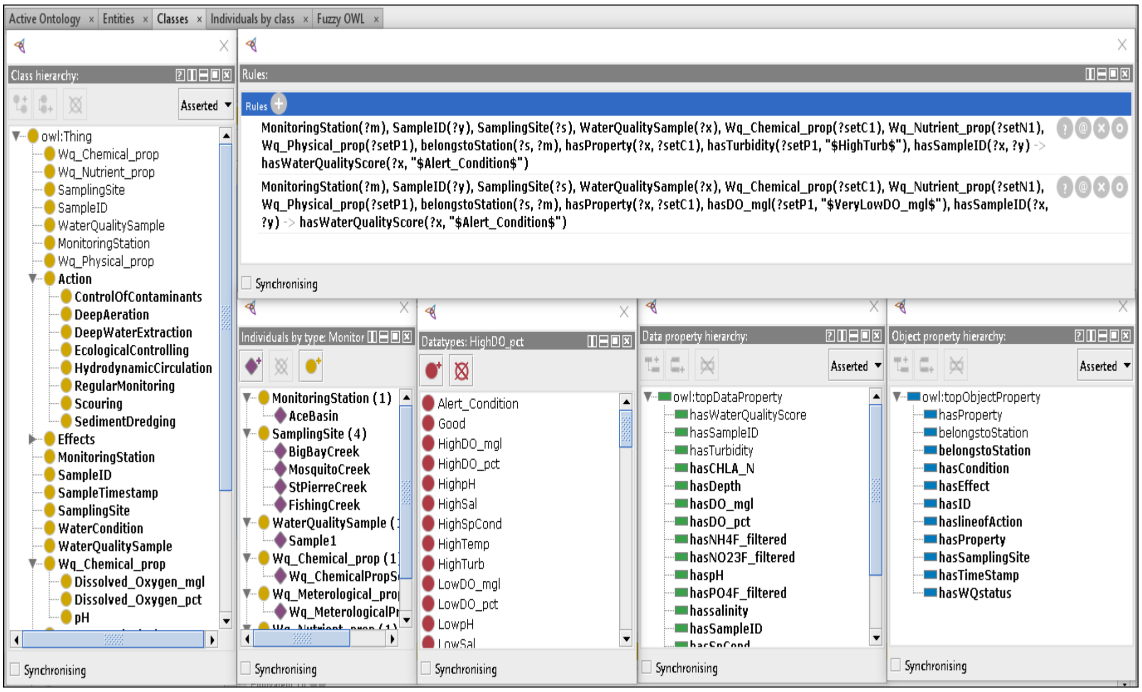


Figure 7.10: Water Sensor Observation Ontology Instantiation

- **Ontology Testing and Validation**

  The performance based evaluation of the proposed system after implementation is performed in order to measure the enhancement in the overall system's accuracy. The verification and validation of the system is of crucial importance as any update to the ontology can lead to redundant instances or inconsistency that can disturb the overall performance of the system. A Question/Answer methodology is deployed to test the performance and evaluation[86]. Protege tool allows launching DL and SPARQL [55] queries on the developed ontology and provides flexibility along with a user-friendly environment for validation. Additionally, it constructs valid instances or individual members as a result of the input query in the required time. DL/SPARQL[87] queries are constructed in OWL language in combination with variables and properties to retrieve the required knowledge.

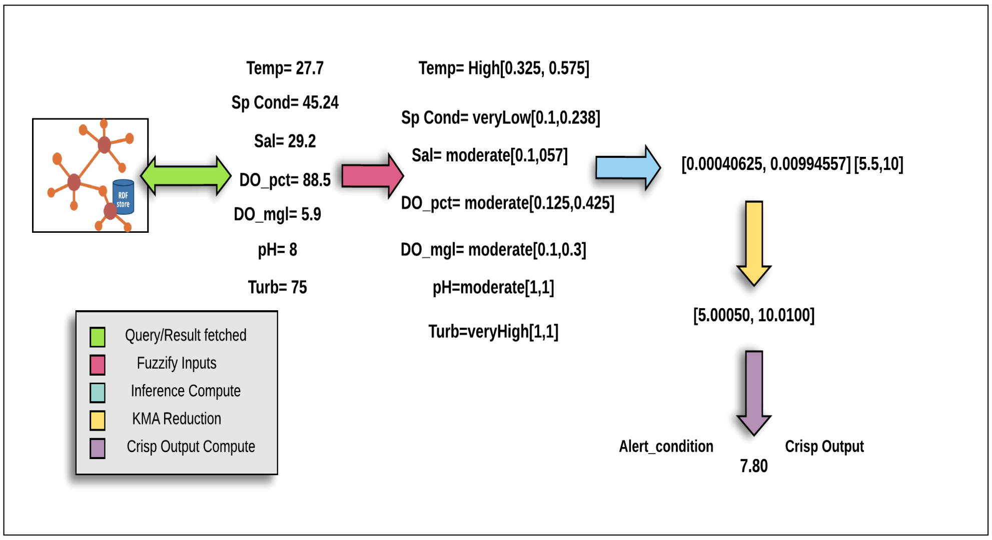


Figure 7.11: Fuzzy Sets Computation

Table 7.4: Fuzzy Partitioning of some parameters

Fuzzy Variable	Fuzzy Linguistic Term	$[LMF_L, M_{LMF_l}, M_{LMF_r}, LMF_R]$	$[UMF_L, M_{UMF_l}, M_{UMF_r}, UMF_R]$
Temperature	veryLow	[-10, 0, 0, 5]	[-10,0,0,6]
	Low	[6, 8, 8, 10]	[5, 8, 8, 11]
	Moderate	[11, 15, 15, 19]	[9, 15, 15, 21]
	High	[21, 25, 25, 29]	[20, 25, 25, 30]
	veryHigh	[28, 34, 40, 40]	[30, 34, 40, 40]
pH	veryLow	[0, 0, 2, 3]	[0, 0, 2, 4]
	Low	[4, 5, 5, 6]	[3, 5, 5, 7]
	Moderate	[7, 8, 8, 8.5]	[6.5, 8, 8, 9]
	High	[9, 10, 10, 11]	[8.5, 10, 10, 12]
	veryHigh	[11, 12, 14, 14]	[10, 12, 14, 14]
Turbidity	veryLow	[0, 0, 1, 2]	[0, 0, 1, 3]
	Low	[2.5, 3, 3 , 3.5]	[2, 3, 3, 4]
	Moderate	[4, 6, 6, 9]	[3, 6, 6, 10]
	High	[12, 16, 16, 18]	[10, 16, 16, 20]
	veryHigh	[20, 30, 150, 150]	[18, 30, 150, 150]
DO_mgl	veryLow	[0, 0, 1, 2]	[0, 0, 1, 3]
	Low	[2.5, 3, 3, 3.5]	[2, 3, 3, 4]
	Moderate	[7, 8, 8, 11]	[5, 8, 8, 13]
	High	[12, 16, 16, 18]	[10, 16, 16, 20]
	veryHigh	[20, 28, 40, 40]	[18, 28, 40, 40]

DO_pct	veryLow	[0, 35, 35, 40]	[0, 35, 35, 50]
	Low	[45, 58, 58, 62]	[40, 58, 58, 65]
	Moderate	[6, 7, 7, 12]	[5, 7, 7, 13]
	High	[120, 135, 135, 140]	[150, 160, 160, 200]
	veryHigh	[150, 160, 200, 200]	[140, 160, 200, 200]

Salinity	veryLow	[0, 0, 1, 2]	[0, 0, 1, 1.5]
	Low	[3, 4, 4, 5]	[2, 4, 4, 6]
	Moderate	[10, 16, 16, 25]	[5, 16, 16, 30]
	High	[35, 40, 40, 45]	[30, 40, 40, 50]
	veryHigh	[50, 70, 100, 100]	[40, 70, 100, 100]

SpCond	veryLow	[0, 0, 30, 40]	[0, 0, 30, 50]
	Low	[60, 70, 70, 80]	[50, 70, 70, 90]
	Moderate	[90, 100, 100, 110]	[80, 100, 100, 120]
	High	[125, 130, 130, 135]	[120, 130, 130, 140]
	veryHigh	[145, 150, 200, 200]	[145, 150, 200, 200]

nitrite+nitrate	veryLow	[0, 0, 0.04, 0.05]	[0, 0, 0.04, 0.075]
	Low	[0.075,0.085, 0.085, 0.095]	[0.05, 0.085, 0.085, 0.15]
	Moderate	[0.12, 0.15, 0.15, 0.17]	[0.1, 0.15, 0.15, 0.2]
	High	[1, 2, 8, 8]	[0.2, 2, 10, 10]
	veryHigh	[10, 20, 70, 70]	[9, 20, 70, 70]

ammonium	veryLow	[0, 0, 0, 0.12]	[0, 0, 0, 0.05]
	Low	[0.07, 0.1, 0.1, 0.15]	[0.05, 0.1, 0.1, 0.2]
	Moderate	[0.3, 0.4, 0.4, 0.7]	[0.2, 0.4, 0.4, 0.8]
	High	[0.85, 0.9, 0.9, 0.95]	[0.8, 0.9, 0.9, 0.1]
	veryHigh	[2, 3, 5, 5]	[1, 3, 5, 5]

orthophosphate	veryLow	[0, 0, 0, 0.05]	[0, 0, 0, 0.15]
	Low	[0.15, 0.18, 0.18, 0.2]	[0.1, 0.18, 0.18, 0.25]
	Moderate	[0.3, 0.35, 0.35, 0.4]	[0.2, 0.35, 0.35, 0.5]
	High	[0.55, 0.6, 0.6, 0.65]	[0.5, 0.6, 0.6, 0.7]
	veryHigh	[0.8, 1, 20, 20]	[0.6, 1, 20, 20]

chlorophyll	veryLow	[10, 12, 12, 15]	[0, 12, 12, 25]
	Low	[27, 33, 33, 38]	[20, 33, 33, 45]
	Moderate	[45, 50, 50, 55]	[40, 50, 50, 65]
	High	[55, 60, 60, 65]	[50, 60, 60, 70]
	veryHigh	[68, 70, 70, 70]	[65, 70, 70, 70]

Few queries designed and executed over the implemented system to perform evaluation for consistency checking are as follows:

SPARQL Query 1: Show the water quality condition and water quality score of water samples taken between 11 am and 6 pm on 04/08/2004.

Query Syntax:

PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX wqo: <http://www.semanticweb.org/ontoWEB/WaterQuality/wqo##>

SELECT ?sample ?timestamp ?ID ?score ?condition WHERE {

?sample rdf:type wqo:WaterQualitySample .

?sample wqo:hasTS ?timestamp .

FILTER (?timestamp > 110000 && ?timestamp < 180000) .

?sample wqo:hasSampleID ?ID .

?sample wqo:hasScorevalue ?score .

FILTER(?score > 5) ?sample wqo:hasQualitycond ?condition . ?sample wqo:hasSamplingDate "04/08/2004"

}

ORDER by ?sample

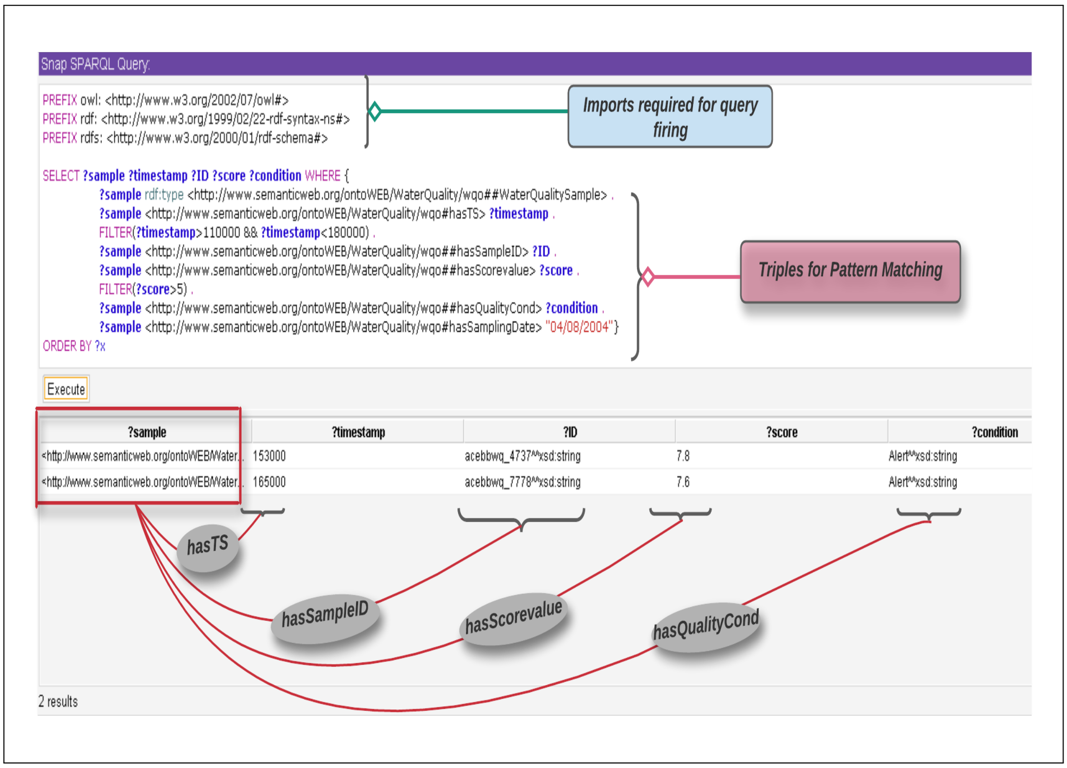


Figure 7.12: Query to list instances with their water quality condition and score

Explanation: While firing the above query on the proposed ontological system, the analyst needs to find out the water quality condition and the computed water quality score of the all water samples that have sampling date "04/08/2004" and are captured between 11 am and 6 pm with computed score greater than 5.

DL Query 2: Query to list all instances of Water Quality Samples under consideration for water quality condition as "Alert".

Query Syntax:

WaterQualitySample ***and*** *(hasQualitycond* ***value*** ***"Alert")*** *and hasSamplingDate* ***value "04/08/2004"***.

Explanation: The analyst retrieves all instances of water quality samples which have unhealthy water quality condition.

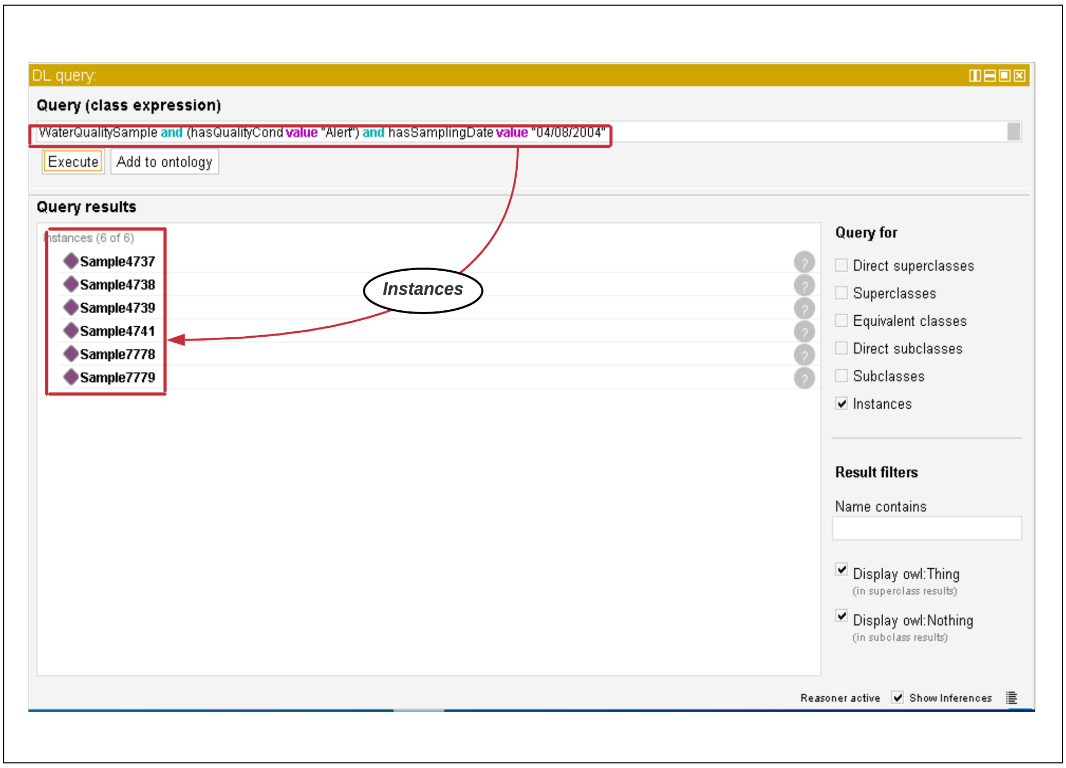


Figure 7.13: Query to list all instances with Alert as the water quality condition

Overall System's Efficiency and Accuracy:

In order to evaluate the overall accuracy and performance, sensor-based water samples data with varying features value is used to instantiate the proposed system with varying scenarios and conditions to make different inquiries to get the desired results. For this evaluation, the ontological system was instantiated with all the samples of "ACE Basin" and "Apalachicola" monitoring station, each having a set of sampling sites capturing feature-based water samples each separated with an epoch interval of 30 mins. Each water sample has parameters-based data which is semantically annotated with the fuzzy ontology and then interval type-2 fuzzy logic inference mechanism is performed. The results in the form of a water quality score is stored in the ontology to be retrieved for decision making. Table 7.5 shows a subset of results with each row showing the computed values of LMF, UMF and water quality score corresponding to each samples. The results are compared with the domain expert knowledge and the precision is figured out to be 100%. Each of the samples, upon classification got a score which accurately assigned a water quality condition.

Table 7.5: Performance for Queries to the Proposed System

No.	Temp.	Sp. Cond	Sal	DO %	DO mgl	pH	Turb.	Score
1687	M[0.12, 0.36]	VL[0.2, 0.32]	VL[0.3, 0.57]	H[0.25, 0.45]	VH[0.3, 0.6]	VL[0.45, 0.74]	VH[0.11, 0.52]	8.2
1990	VH[0.17, 0.63]	L[0.2, 0.3]	VL[0.4, 0.62]	L[0.15, 0.55]	VL[0.1, 0.7]	VH[0.52, 0.62]	H[0.31, 0.32]	6.9
2980	H[0.32, 0.56]	VL[0.1, 0.23]	L[0.1, 0.57]	L[0.125, 0.425]	L[0.2, 0.4]	L[0.425, 0.78]	VH[0.31, 0.62]	7.7
3667	VH[0.21, 0.45]	L[0.3, 0.45]	L[0.2, 0.46]	M[0.5, 0.65]	L[0.3, 0.5]	VL[0.25, 0.68]	H[0.22, 0.54]	7.4
4555	M[0.52, 0.86]	M[0.3, 0.53]	M[0.4, 0.77]	M[0.152, 0.22]	L[0.1, 0.3]	M[0.45, 0.68]	M[0.53, 0.82]	1.8
5540	M[0.32, 0.56]	L[0.2, 0.33]	M[0.2, 0.67]	M[0.125, 0.25]	L[0.3, 0.5]	H[0.225, 0.65]	H[0.24, 0.42]	6.5
5990	H[0.38, 0.55]	VL[0.1, 0.33]	VH[0.1, 0.46]	VL[0.15, 0.49]	VL[0.4, 0.6]	VL[0.35, 0.48]	VH[0.56, 0.6]	9.1

Table 7.6 shows the confusion matrix of all the water samples which belong to two different monitoring stations and distinct sampling sites. There are 4 different performance parameters computed based upon the classification done and known expert domain knowledge, namely: TP (water sample correctly classified as an alert condition), FP (water sample predicted as alert but is actual a healthy sample), TN (water sample correctly predicted as healthy) and FN (unhealthy water quality sample is predicted as healthy).

Additionally, in order to compare the overall performance of the proposed interval type-2 fuzzy-based system with the classical ontology which was obtained before the fuzzification process.The NOAA dataset is used in the result evaluation and is divided into multiple folds of data records and the water sample quality value and condition is computed based on various quality, nutrient and meteorological parameters of the data. The output condition of the water sample under consideration is divided into 3 output fuzzy variables: Healthy (0 - 4), Moderate (2 – 6) and Alert_Condition (5.5-

10).

Table 7.6: Confusion Matrix for Water Samples

Station	Sampling Site	TP	FP	FN	TN	Normal Samples	Unhealthy Samples	Total Samples
ACE Basin	Big Bay	544	6	4	2998	4450	550	5000
	Fishing creek	97	2	2	2998	3001	99	3100
	Mosquito Creek	15	5	1	2029	2030	20	2050
	St. Pierre	177	3	0	3220	3220	180	3400
Apalachicola	Cat Point	48	2	1	1449	1450	50	1500
	Dry Bar	20	0	4	2266	2270	20	2290
	East Bay Bottom	136	4	0	2860	2860	140	3000
	East Bay Surface	57	3	4	1337	1340	60	1400

In the experimental evaluation, various sensor samples having multiple attributes are subjected to the proposed system. Additionally, different types of queries were fired for each water sample, uniquely identified by it's sampleID, in order to extract the desired knowledge from the semantic framework. The proposed system is evaluated for performance based upon the following metrics: precision, recall, accuracy, and function measure. These metrics have been used to evaluate the performance of prediction models and methods used in experiments. Precision determines the degree of closeness of two or more conclusions to one another, and accuracy shows the measure of closeness of a predicted value to a known value. Firstly, a classic ontology is used for the process of water quality prediction and further line of action recommendation, and all the performance metrices i.e. precision, recall, accuracy, and function measure are computed. Then, the proposed Interval Type-2 Fuzzy Logic-based fuzzy ontology is used to compute the values of the performance parameters. In each case, the water sample condition values and query results are compared and the results are shown in Table 7.7.

The entire dataset is segregated into multiple folds. For each fold, a set of

Table 7.7: Results comparison of Proposed System with the Classical Ontology-based Model

Water Samples	Classic Ontology-based Quality Prediction System				The proposed Interval Type-2 Fuzzy Logic Ontology-based system			
	Precision (%)	Recall (%)	Accuracy (%)	FM (%)	Precision (%)	Recall (%)	Accuracy (%)	FM (%)
1^{st}fold	57	40	66	47	80	88	80	84
2^{nd}fold	50	37	60	43	90	81	80	85
3^{rd}fold	62	50	46	55	75	90	75	81
4^{th}fold	40	66	46	50	75	88	75	80

15 queries are fired. The results are evaluated then graphically represented in figure 7.11.

7.4.3 Discussion

The proposed interval type-2 fuzzy logic-based ontology system is capable of handling fuzziness and uncertainty in the data, which cannot be effectively addressed by classical ontology models. This approach has two main advantages. The first one is that it formally represents the dynamic observational information in the form of interval type-2 fuzzy partitions and linguistic variables so as to address the fuzziness at the very fundamental and primitive step. The fuzzy reasoning takes place based upon the semantic fuzzy rule-base developed over the represented knowledge. Secondly, the integration of fuzzy reasoning with the knowledge-based decision making in an intelligent manner efficiently models the water quality prediction process.

Additionally, in the experimental evaluation of the proposed system, the computed value and conditions of various water samples are compared with the classical ontology-based model and the proposed system. It is evident that the classical ontology-based model can predict water quality of samples to a limited extent.

It fails to show positive results in case of uncertain data. On the contrary, the proposed system shows promising results, with good accuracy in the prediction of water quality and average precision of 80%. It can be used as a water quality prediction system which suggests proper actions and effects as per the quality of the sample under consideration.

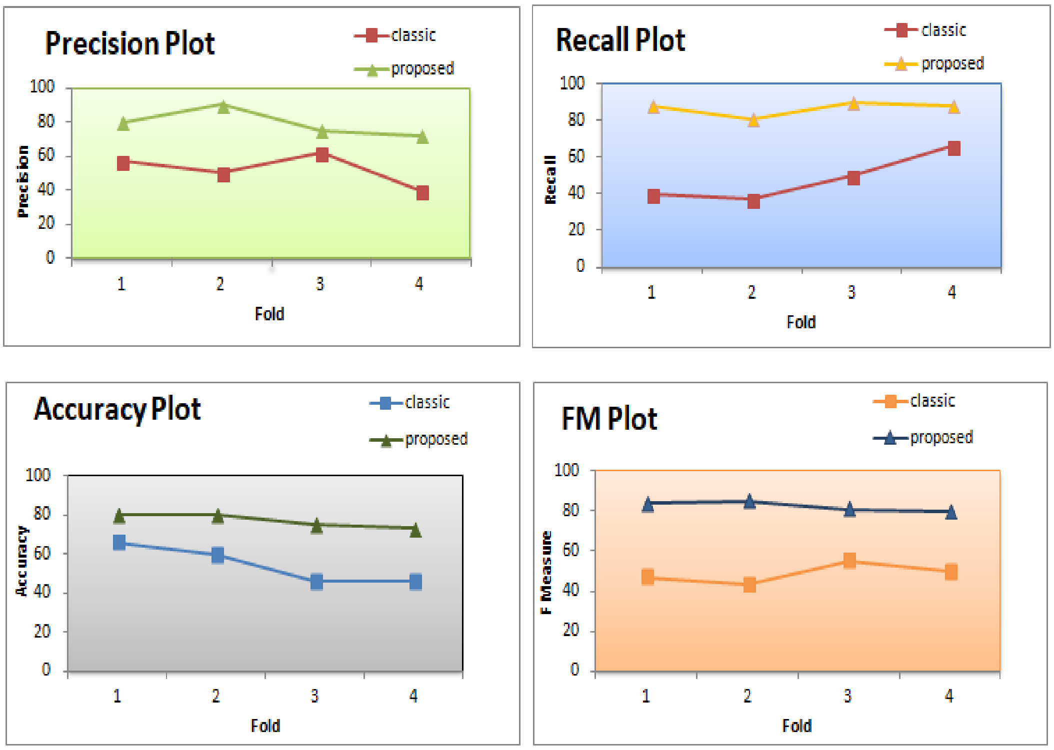


Figure 7.14: Graphical representation of performance metrics

Chapter 8

CONCLUSION AND FUTURE DIRECTIONS

This chapter comprises of the crucial findings, research contributions and outcomes and the scope of future work in the domain of Sensor data analysis and Ontology-driven Knowledge Systems. The further sectioning of this chapter is: Section 8.1 discusses the contributions and outcomes, Section 8.2 contains the prospective future works in this direction.

8.1 Conclusion

The backbone of this thesis is to design and develop Ontological Knowledge-based System for multimodal/heterogeneous sensor data analysis. Since, this domain requires efficient and powerful approaches, Semantic Web Technology is used to gain deeper insights and knowledge from the captured data. The first ontology-based knowledge model is implemented for data from heterogeneous sensors. The proposed model exploits semantic technologies to conceptualise domain knowledge and develop ontologies to become a part of the Ontological Knowledge Base. The instantiation of the developed model is done on data captured by sensors embedded in the smartphones and smartwatches while daily living activities are being performed by various subjects. The semantic annotation to the heterogeneous sensor observational data is provided in the sensor measurements ontology, while the inference tasks are carried out over the activity recognition ontology through dependencies between a complex activity, which is to be inferred, and a combination of simple activities. The semantic web querying service, SPARQL fetches data and reasoning tasks are performed for knowledge mining. The proposed model facilitates easy reuse and exchange of knowledge across multiple domains. It promotes interoperability through fusion of heterogeneous sensor data in a common standardized format. Additionally, due to the modular design of the developed ontologies, it is easier to integrate new information into the knowledge base without changing the underlying structure. The ontology-based knowledge system is validated with a real-life dataset which contains sensor measurements of 60 subjects or users with more than 3,00,000 samples from various sensors. The results classify the sensor measurements at a particular instant with respect to a particular user/subject into a high level activity or inferred act. The resultant knowledge model provides effective analysis and promising results. However, the proposed system has certain limitations in modeling domain knowledge accurately. The uncertainty associated in the domain knowledge gathered can sometimes lead to erroneous results.

In order to overcome this, another approach that enriches domain facts represented as ontological knowledge with probabilistic methods. It semantically annotates the multimodal (heterogeneous) sensor data captured by heterogeneous sensors. To add on, the proposed data segmentation method segregates continuous sensor data into patches allowing full exhaustive utilization of the heterogeneous multi format input data. The proposed knowledge model is dynamically instantiated with a real-world dataset of daily living human activities being performed, containing temporal sensor observations in heterogeneous formats. For further enrichment, data-driven spatial context and temporal relationships are added to enhance the overall representation of real world activity performing scenarios. Earmarking a large and a real-world scenario of heterogeneous sensor data collected while human activities are being done, a probabilistic ontology is implemented, flexible to any number of activities. The performed experimental evaluation illustrates that the system shows desirable performance even in the presence of partial/uncertain data and information compared to existing baseline techniques that are driven by traditional supervised learning. Additionally, the system provides temporal boundaries with respect to the activities more precisely. Further, spatial context also enhances the activity representation as a set of different observed concepts. Consequently, it showcases the integrated implementation of the ontological formalism and its probabilistic axiomatic representation, further driven by Pronto probabilistic reasoning over SPARQL-based querying service. The resultant system overcomes the rigidness of existing approaches. It effectively handles ambiguous and uncertain ontological representation of activities being performed, on the other hand benefiting from the expressive potential and rules driven inference mechanism of the OWL language, a shortcoming of elementary knowledge-driven techniques.

Another pillar in dealing with the heterogeneity associated with sensor data is data captured by digital sensors. These heterogeneous sensors generate output in the digitized format which needs to be analyzed for knowledge required to perform decision making. Another model that is proposed presents an ontology driven knowledge model to deal with digital heterogeneous sensor data to extract knowledge. It is validated and instantiated with digital sensor ECG data belonging to patients. The digitized heterogeneous sensor data is preprocessed and converted into a format that is machine interpretable and understandable, so that Semantic Web technology be deployed for Knowledge Engineering. The proposed models the arrhythmia domain knowledge and the conceptual relationships relevant to classification of heartbeat into corresponding cardiac arrhythmias and to facilitate decision making with respect to the patients. The newly developed arrhythmia ontology consists of three different modules, each semantically annotating a different aspect of the arrhythmia detection process. A SWRL(Semantic Web Rule Language) based ontology classifier performs classification of patient's ECG data into corresponding cardiac arrhythmia types. The inference process is driven by in the ontology reasoning in an effective manner to achieve desired results. Hence, this integration improved the semantic intelligence of the proposed system and its interoperability to promote reuse and exchange of the knowledge and data. The proposed system is experimentally implemented and evaluated with a real world dataset with different test cases producing accurate results.

There exists alot of uncertainty in sensor data captured by heterogeneous sensors. At the decision-making level, classical ontologies are not considered as powerful tools when there is alot of vagueness in sensor data for real-world applications. On the contrary, fuzzy ontological system for heterogeneous sensor data deals with the issue of haziness and uncertainty for effective analysis to give promising results. In the proposed work, fuzzy partitioning, fuzzy sets and fuzzy rules are developed with respect to a real world dataset. Secondly, fuzzy inference process is integrated in the ontology reasoning in an effective manner to achieve desired results. Hence, this integration improved the semantic intelligence of the proposed system and its interoperability to promote reuse and exchange of the knowledge and data. The proposed system is experimentally implemented and evaluated with a real-world dataset with different test cases producing accurate results.

8.2 Future Directions

The research work described in this thesis has a number of prospective promising future directions for further research.

- **Distributed Knowledge Systems:** The distributed nature of Knowledge-based systems can provide more flexibility and robustness. It is even more crucial to achieve workload balancing among peers in case of distributed infrastructure and implementation.

- **Result Ranking:** There is scope of devising efficient result ranking of user queries fired at the system for effective and impressive knowledge sharing with the end users. This enhances the overall experience of the end user and optimized results are obtained for further processing and analysis.

- **Data Visualization:** Sensor data Ontologies are often represented as a graph where nodes represent objects within given ontology and edges represent relations between those objects. Such graphs can be very complex even for medium sized ontologies so a clean and complete graphical representation is needed to visualize heterogeneous sensor data. Existing solutions do not include all elements or produce unreadable results for big ontologies developed for sensor data analysis.

- **Query Optimization:** Query optimization can play a crucial role so as to get better results for sensor data analysis with much less complexity with respect to time. This may be extended to refine results in case of ambiguous user queries fired at the system to fetch sensed information.

References

[1] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge acquisition*, vol. 5, no. 2, pp. 199–220, 1993.

[2] W. N. Borst, "Construction of engineering ontologies for knowledge sharing and reuse.," 1999.

[3] R. Studer, V. R. Benjamins, and D. Fensel, "Knowledge engineering: principles and methods," *Data & knowledge engineering*, vol. 25, no. 1-2, pp. 161–197, 1998.

[4] P. Shvaiko and J. Euzenat, "Ontology matching: state of the art and future challenges," *IEEE Transactions on knowledge and data engineering*, vol. 25, no. 1, pp. 158–176, 2011.

[5] J. Zhang, W. Zhao, G. Xie, and H. Chen, "Ontology-based knowledge management system and application," *Procedia Engineering*, vol. 15, pp. 1021–1029, 2011.

[6] N. F. Noy, M. Crubézy, R. W. Fergerson, H. Knublauch, S. W. Tu, J. Vendetti, and M. A. Musen, "Protégé-2000: an open-source ontology-development and knowledge-acquisition environment.," in *AMIA... Annual Symposium proceedings. AMIA Symposium*, pp. 953–953, 2003.

[7] M. Compton, P. Barnaghi, L. Bermudez, R. Garcia-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, *et al.*, "The ssn ontology of the w3c semantic sensor network incubator group," *Journal of Web Semantics*, vol. 17, pp. 25–32, 2012.

[8] L. Baresi, E. Di Nitto, and C. Ghezzi, "Toward open-world software: Issues and challenges," *Computer*, vol. 39, no. 10, pp. 36–43, 2006.

[9] N. F. Noy and M. Klein, "Ontology evolution: Not the same as schema evolution," *Knowledge and information systems*, vol. 6, no. 4, pp. 428–440, 2004.

[10] U. Cisco, "Cisco annual internet report (2018–2023) white paper," *Cisco: San Jose, CA, USA*, 2020.

[11] H. Chaouchi and T. Bourgeau, "Internet of things: Building the new digital society," *IoT*, vol. 1, no. 1, p. 1, 2018.

[12] B. Mandler, J. Marquez-Barja, M. E. M. Campista, D. Cagáňová, H. Chaouchi, S. Zeadally, M. Badra, S. Giordano, M. Fazio, A. Somov, *et al.*, *Internet of Things. IoT Infrastructures: Second International Summit, IoT 360° 2015, Rome, Italy, October 27-29, 2015. Revised Selected Papers, Part I*, vol. 169. Springer, 2016.

[13] N. Bilandi, H. K. Verma, and R. Dhir, "Ahp–neutrosophic decision model for selection of relay node in wireless body area network," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 3, pp. 222–229, 2020.

[14] S. T. Moghadam, N. Hooman, and A. Sheikhtaheri, "Patient safety classification, taxonomy and ontology systems: A systematic review on development and evaluation methodologies," *Journal of Biomedical Informatics*, p. 104150, 2022.

[15] E. Manzini, J. Garrido-Aguirre, J. Fonollosa, and A. Perera-Lluna, "Mapping layperson medical terminology into the human phenotype ontology using neural machine translation models," *Expert Systems with Applications*, vol. 204, p. 117446, 2022.

[16] C. Blankenberg, B. Gebel-Sauer, and P. Schubert, "Using a graph database for the ontology-based information integration of business objects from heterogenous business information systems," *Procedia Computer Science*, vol. 196, pp. 314–323, 2022.

[17] J. Ingram and P. Gaskell, "Searching for meaning: Co-constructing ontologies with stakeholders for smarter search engines in agriculture," *NJAS-Wageningen Journal of Life Sciences*, vol. 90, p. 100300, 2019.

[18] G. Chen, T. Jiang, M. Wang, X. Tang, and W. Ji, "Modeling and reasoning of iot architecture in semantic ontology dimension," *Computer Communications*, vol. 153, pp. 580–594, 2020.

[19] J. Reynolds, Y. Rezgui, and J.-L. Hippolyte, "Upscaling energy control from building to districts: Current limitations and future perspectives," *Sustainable Cities and Society*, vol. 35, pp. 816–829, 2017.

[20] J. Warner, "Enschede cries–restoring ontological security after a fireworks disaster," *International Journal of Disaster Risk Reduction*, vol. 59, p. 102171, 2021.

[21] A. R. Muljarto, J.-M. Salmon, B. Charnomordic, P. Buche, A. Tireau, and P. Neveu, "A generic ontological network for agri-food experiment integration–application to viticulture and winemaking," *Computers and electronics in agriculture*, vol. 140, pp. 433–442, 2017.

[22] K. Bagherifard, M. Rahmani, M. Nilashi, and V. Rafe, "Performance improvement for recommender systems using ontology," *Telematics and Informatics*, vol. 34, no. 8, pp. 1772–1792, 2017.

[23] O. N. Oyelade, I. E. Aghiomesi, O. Najeem, and A. A. Sambo, "A semantic web rule and ontologies based architecture for diagnosing breast cancer using select and test algorithm," *Computer Methods and Programs in Biomedicine Update*, vol. 1, p. 100034, 2021.

[24] T. Sobral, T. Galvão, and J. Borges, "An ontology-based approach to knowledge-assisted integration and visualization of urban mobility data," *Expert Systems with Applications*, vol. 150, p. 113260, 2020.

[25] B. J. Cortés, J. A. Vera-Ramos, R. C. Lovering, P. Gaudet, A. Laegreid, C. Logie, S. Schulz, M. del Mar Roldán-García, M. Kuiper, and J. T. Fernández-Breis, "Formalization of gene regulation knowledge using ontologies and gene ontology causal activity models," *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1864, no. 11-12, p. 194766, 2021.

[26] D. Meškelė and F. Frasincar, "Aldonar: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model," *Information Processing & Management*, vol. 57, no. 3, p. 102211, 2020.

[27] G. Santipantakis, K. Kotis, and G. A. Vouros, "Obdair: Ontology-based distributed framework for accessing, integrating and reasoning with data in disparate data sources," *Expert Systems with Applications*, vol. 90, pp. 464–483, 2017.

[28] F. Ullah, M. A. Habib, M. Farhan, S. Khalid, M. Y. Durrani, and S. Jabbar, "Semantic interoperability for big-data in heterogeneous iot infrastructure for healthcare," *Sustainable cities and society*, vol. 34, pp. 90–96, 2017.

[29] M. Serrano, A. Gyrard, E. Tragos, and H. Nguyen, "Fiestaiot project: federated interoperable semantic iot/cloud testbeds and applications," in *Companion Proceedings of the The Web Conference 2018*, pp. 425–426, 2018.

[30] A. Gyrard, C. Bonnet, K. Boudaoud, and M. Serrano, "Lov4iot: A second life for ontology-based domain knowledge to build semantic web of things applications," in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud)*, pp. 254–261, IEEE, 2016.

[31] L. Buoncompagni, S. Y. Kareem, and F. Mastrogiovanni, "Human activity recognition models in ontology networks," *IEEE Transactions on Cybernetics*, 2021.

[32] N. Sharma, M. Mangla, S. N. Mohanty, D. Gupta, P. Tiwari, M. Shorfuzzaman, and M. Rawashdeh, "A smart ontology-based iot framework for remote patient monitoring," *Biomedical Signal Processing and Control*, vol. 68, p. 102717, 2021.

[33] T. Elsaleh, S. Enshaeifar, R. Rezvani, S. T. Acton, V. Janeiko, and M. Bermudez-Edo, "Iot-stream a lightweight ontology for internet of things data streams and its use with data analytics and event detection services," *Sensors*, vol. 20, no. 4, p. 953, 2020.

[34] G. A. K. Doukoure and E. Mnkandla, "Facilitating the management of agile and devops activities: Implementation of a data consolidator," in *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*, pp. 1–6, IEEE, 2018.

[35] O. G. Troyanskaya, K. Dolinski, A. B. Owen, R. B. Altman, and D. Botstein, "A bayesian framework for combining heterogeneous data sources for gene function prediction (in saccharomyces cerevisiae)," *Proceedings of the National Academy of Sciences*, vol. 100, no. 14, pp. 8348–8353, 2003.

[36] I. K. Ibrahim, R. Kronsteiner, and G. Kotsis, "A semantic solution for data integration in mixed sensor networks," *Computer Communications*, vol. 28, no. 13, pp. 1564–1574, 2005.

[37] J. Dai, M. Zhang, G. Chen, J. Fan, K. Y. Ngiam, and B. C. Ooi, "Fine-grained concept linking using neural networks in healthcare," in *Proceedings of the 2018 International Conference on Management of Data*, pp. 51–66, 2018.

[38] S. Hachour, F. Delmotte, and D. Mercier, "A robust credal assignment solution based on the generalized bayes' theorem," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 25, no. 06, pp. 947–971, 2017.

[39] P. Poekaew and P. Champrasert, "Adaptive-pca: An event-based data aggregation using principal component analysis for wsns," in *2015 International Conference on Smart Sensors and Application (ICSSA)*, pp. 50–55, IEEE, 2015.

[40] M. Stocker, M. Rönkkö, and M. Kolehmainen, "Making sense of sensor data using ontology: A discussion for residential building monitoring," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pp. 341–350, Springer, 2012.

[41] X. Zhang, Y. Zhao, and W. Liu, "A method for mapping sensor data to ssn ontology," *International Journal of u-and e-Service, Science and Technology*, vol. 8, no. 9, pp. 303–316, 2015.

[42] X. Su, H. Zhang, J. Riekki, A. Keränen, J. K. Nurminen, and L. Du, "Connecting iot sensors to knowledge-based systems by transforming senml to rdf," *Procedia Computer Science*, vol. 32, pp. 215–222, 2014.

[43] S. Zander, N. Merkle, and M. Frank, "Enhancing the utilization of iot devices using ontological semantics and reasoning," *Procedia Computer Science*, vol. 98, pp. 87–90, 2016.

[44] G. Santipantakis, K. I. Kotis, and G. A. Vouros, "Accessing and reasoning with data from disparate data sources using modular ontologies and obda," in *Proceedings of the 11th International Conference on Semantic Systems*, pp. 41–48, 2015.

[45] B. Kryza, J. Kitowski, *et al.*, "Integration of heterogeneous data sources in an ontological knowledge base," *Computing and Informatics*, vol. 31, no. 1, pp. 189–223, 2012.

[46] H. Ouksili, Z. Kedad, S. Lopes, and S. Nugier, "Pattern oriented rdf graphs exploration," *Data & Knowledge Engineering*, vol. 113, pp. 171–183, 2018.

[47] C. Nikolaou and M. Koubarakis, "Querying incomplete information in rdf with sparql," *Artificial Intelligence*, vol. 237, pp. 138–171, 2016.

[48] H. Ouksili, Z. Kedad, and S. Lopes, "Theme identification in rdf graphs," in *International Conference on Model and Data Engineering*, pp. 321–329, Springer, 2014.

[49] R. Angles and C. Gutierrez, "Querying rdf data from a graph database perspective," in *European semantic web conference*, pp. 346–360, Springer, 2005.

[50] S. Ketu and P. K. Mishra, "Performance analysis of machine learning algorithms for iot-based human activity recognition," in *Advances in electrical and computer technologies*, pp. 579–591, Springer, 2020.

[51] S. Ali, S. Khusro, A. Rauf, and S. Mahfooz, "Sensors and mobile phones: evolution and state-of-the-art," *Pakistan journal of science*, vol. 66, no. 4, p. 385, 2014.

[52] P. Korpipää and J. Mäntyjärvi, "An ontology for mobile device sensor-based context awareness," in *International and Interdisciplinary Conference on Modeling and Using Context*, pp. 451–458, Springer, 2003.

[53] S. Ketu and P. K. Mishra, "Internet of healthcare things: A contemporary survey," *Journal of Network and Computer Applications*, vol. 192, p. 103179, 2021.

[54] D. J. Russomanno, C. Kothari, and O. Thomas, "Sensor ontologies: from shallow to deep models," in *Proceedings of the Thirty-Seventh Southeastern Symposium on System Theory, 2005. SSST'05.*, pp. 107–112, IEEE, 2005.

[55] C. A. Henson, J. K. Pschorr, A. P. Sheth, and K. Thirunarayan, "Semsos: Semantic sensor observation service," in *2009 International Symposium on Collaborative Technologies and Systems*, pp. 44–53, IEEE, 2009.

[56] H. Neuhaus and M. Compton, "The semantic sensor network ontology," in *AGILE workshop on challenges in geospatial data harmonisation, Hannover, Germany*, pp. 1–33, 2009.

[57] N. F. Noy, D. L. McGuinness, *et al.*, "Ontology development 101: A guide to creating your first ontology," 2001.

[58] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing?," *International journal of human-computer studies*, vol. 43, no. 5-6, pp. 907–928, 1995.

[59] J. Pérez, M. Arenas, and C. Gutierrez, "Semantics and complexity of sparql," *ACM Transactions on Database Systems (TODS)*, vol. 34, no. 3, pp. 1–45, 2009.

[60] Y. Vaizman, K. Ellis, and G. Lanckriet, "Recognizing detailed human context in the wild from smartphones and smartwatches," *IEEE pervasive computing*, vol. 16, no. 4, pp. 62–74, 2017.

[61] Y. Vaizman, K. Ellis, G. Lanckriet, and N. Weibel, "Extrasensory app: Data collection in-the-wild with rich user interface to self-report behavior," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–12, 2018.

[62] N. F. Noy, M. Sintek, S. Decker, M. Crubézy, R. W. Fergerson, and M. A. Musen, "Creating semantic web contents with protege-2000," *IEEE intelligent systems*, vol. 16, no. 2, pp. 60–71, 2001.

[63] J. R. Beard, A. M. Officer, and A. K. Cassels, "The world report on ageing and health," 2016.

[64] P. Laukkanen, E. Leskinen, M. Kauppinen, R. Sakari-Rantala, and E. Heikkinen, "Health and functional capacity as predictors of community dwelling among elderly people," *Journal of Clinical Epidemiology*, vol. 53, no. 3, pp. 257–265, 2000.

[65] M. L. Lee and A. K. Dey, "Sensor-based observations of daily living for aging in place," *Personal and Ubiquitous Computing*, vol. 19, no. 1, pp. 27–43, 2015.

[66] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Information Fusion*, vol. 35, pp. 68–80, 2017.

[67] K. Colpaert, S. Vanbelleghem, C. Danneels, D. Benoit, K. Steurbaut, S. Van Hoecke, F. De Turck, and J. Decruyenaere, "Has information technology finally been adopted in flemish intensive care units?," *BMC medical informatics and decision making*, vol. 10, no. 1, pp. 1–7, 2010.

[68] C. Orwat, A. Graefe, and T. Faulwasser, "Towards pervasive computing in health care–a literature review," *BMC medical informatics and decision making*, vol. 8, no. 1, pp. 1–18, 2008.

[69] R. Jha, V. Bhattacharjee, and A. Mustafi, "Iot in healthcare: A big data perspective," in *Smart Healthcare Analytics in IoT Enabled Environment*, pp. 201–211, Springer, 2020.

[70] M. Tentori, D. Segura, and J. Favela, "Monitoring hospital patients using ambient displays," in *Mobile Health Solutions for Biomedical Applications*, pp. 143–158, IGI Global, 2009.

[71] J. Burgelman, Y. Punie, E. Aarts, and J. Encarnaceo, "Close encounters of a different kind: ambient intelligence in europe," *True Vision: The Emergence of Ambient Intelligence*, pp. 19–35, 2006.

[72] D. L. McGuinness, F. Van Harmelen, *et al.*, "Owl web ontology language overview," *W3C recommendation*, vol. 10, no. 10, p. 2004, 2004.

[73] F. Baader, D. Calvanese, D. McGuinness, P. Patel-Schneider, D. Nardi, *et al.*, *The description logic handbook: Theory, implementation and applications*. Cambridge university press, 2003.

[74] D. Hooda and R. Rani, "Ontology driven human activity recognition in heterogeneous sensor measurements," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 12, pp. 5947–5960, 2020.

[75] F. Ongenae, F. De Backere, K. Steurbaut, K. Colpaert, W. Kerckhove, J. Decruyenaere, and F. De Turck, "Appendix b: overview of the existing medical and natural language ontologies which can be used to support the trans-lation process," *BMC Med Inform Decis Mak*, vol. 10, no. 3, p. 4, 2011.

[76] A. L. Rector, J. Rogers, P. E. Zanstra, and E. Van Der Haring, "Opengalen: open source medical terminology and tools," in *AMIA Annual Symposium Proceedings*, vol. 2003, p. 982, American Medical Informatics Association, 2003.

[77] A. Burger, D. Davidson, and R. Baldock, *Anatomy ontologies for bioinformatics: principles and practice*, vol. 6. Springer Science & Business Media, 2007.

[78] J. A. Blake and M. A. Harris, "The gene ontology (go) project: structured vocabularies for molecular biology and their application to genome and expression analysis," *Current protocols in bioinformatics*, no. 1, pp. 7–2, 2003.

[79] A. Valls, K. Gibert, D. Sánchez, and M. Batet, "Using ontologies for structuring organizational knowledge in home care assistance," *international journal of medical informatics*, vol. 79, no. 5, pp. 370–387, 2010.

[80] F. Paganelli and D. Giuli, "An ontology-based system for context-aware and configurable services to support home-based continuous care," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 324–333, 2010.

[81] J. Bernardo and A. Smith, "Bayesian theory, 405," 2009.

[82] J. Joyce, "Bayes' theorem," 2003.

[83] Y. Shoham, "Temporal logics in ai: Semantical and ontological considerations," *Artificial intelligence*, vol. 33, no. 1, pp. 89–104, 1987.

[84] P. Klinov and B. Parsia, "Pronto: Probabilistic ontological modeling in the semantic web.," in *International Semantic Web Conference (Posters & Demos)*, vol. 401, Citeseer, 2008.

[85] A. Karakostas, A. Briassouli, K. Avgerinakis, I. Kompatsiaris, and M. Tsolaki, "The dem@ care experiments and datasets: a technical report," *arXiv preprint arXiv:1701.01142*, 2016.

[86] J. Hartmann, P. Spyns, A. Giboin, D. Maynard, R. Cuel, M. Suarez-Figeroa, and Y. Sure, "Methods for ontology evaluation.," 2005.

[87] H. Jung and W. Kim, "Automated conversion from natural language query to sparql query," *Journal of Intelligent Information Systems*, vol. 55, no. 3, pp. 501–520, 2020.

[88] F. Negin, M. Koperski, C. F. Crispim, F. Bremond, S. Coşar, and K. Avgerinakis, "A hybrid framework for online recognition of activities of daily living in real-world settings," in *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 37–43, IEEE, 2016.

[89] F. Negin, S. Cogar, F. Bremond, and M. Koperski, "Generating unsupervised models for online long-term daily living activity recognition," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 186–190, IEEE, 2015.

[90] F. Negin and F. Brémond, "An unsupervised framework for online spatiotemporal detection of activities of daily living by hierarchical activity models," *Sensors*, vol. 19, no. 19, p. 4237, 2019.

[91] K. Avgerinakis, A. Briassouli, and Y. Kompatsiaris, "Activity detection using sequential statistical boundary detection (ssbd)," *Computer Vision and Image Understanding*, vol. 144, pp. 46–61, 2016.

[92] S. Elloumi, S. Cosar, G. Pusiol, F. Bremond, and M. Thonnat, "Unsupervised discovery of human activities from long-time videos," *IET Computer Vision*, vol. 9, no. 4, pp. 522–530, 2015.

[93] S. Wilk, M. Kezadri-Hamiaz, D. Amyot, W. Michalowski, C. Kuziemsky, N. Catal, D. Rosu, M. Carrier, and R. Giffen, "An ontology-driven framework to support the dynamic formation of an interdisciplinary healthcare team," *International Journal of Medical Informatics*, vol. 136, p. 104075, 2020.

[94] Y. Chen, C. Yu, X. Liu, T. Xi, G. Xu, Y. Sun, F. Zhu, and B. Shen, "Pclion: An ontology for data standardization and sharing of prostate cancer associated lifestyles," *International Journal of Medical Informatics*, vol. 145, p. 104332, 2021.

[95] Y. Deng, Z. Gao, S. Xu, P. Ren, Y. Wen, Y. Mao, and Z. Li, "St-net: Synthetic ecg tracings for diagnosing various cardiovascular diseases," *Biomedical Signal Processing and Control*, vol. 61, p. 101997, 2020.

[96] M. Beers and R. Berkow, "The merck manual of diagnosis and therapy. whitehouse station, nj: Merck & co," *Inc.[Google Scholar]*, 1999.

[97] D. H. Bennett, *Bennett's Cardiac Arrhythmias: Practical notes on interpretation and treatment*. John Wiley & Sons, 2012.

[98] R. J. Prineas, R. S. Crow, and Z.-M. Zhang, *The Minnesota code manual of electrocardiographic findings*. Springer Science & Business Media, 2009.

[99] T. Tuncer, S. Dogan, P. Pławiak, and U. R. Acharya, "Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ecg signals," *Knowledge-Based Systems*, vol. 186, p. 104923, 2019.

[100] A. K. Sangaiah, M. Arumugam, and G.-B. Bian, "An intelligent learning approach for improving ecg signal classification and arrhythmia analysis," *Artificial intelligence in medicine*, vol. 103, p. 101788, 2020.

[101] S. Kiranyaz, T. Ince, J. Pulkkinen, and M. Gabbouj, "Personalized long-term ecg classification: A systematic approach," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3220–3226, 2011.

[102] G. Guizzardi and T. Halpin, "Ontological foundations for conceptual modelling," *Applied Ontology*, vol. 3, no. 1-2, pp. 1–12, 2008.

[103] P. Shvaiko and J. Euzenat, "A survey of schema-based matching approaches," in *Journal on data semantics IV*, pp. 146–171, Springer, 2005.

[104] S. Borgo and C. Masolo, "Foundational choices in dolce," in *Handbook on ontologies*, pp. 361–381, Springer, 2009.

[105] D. Brickley and L. Miller, "Foaf vocabulary specification 0.91," 2007.

[106] M. Q. Stearns, C. Price, K. A. Spackman, and A. Y. Wang, "Snomed clinical terms: overview of the development process and project status.," in *Proceedings of the AMIA Symposium*, p. 662, American Medical Informatics Association, 2001.

[107] J. Kunze and T. Baker, "The dublin core metadata element set," tech. rep., 2007.

[108] J. G. Breslin, S. Decker, A. Harth, and U. Bojars, "Sioc: an approach to connect web-based communities," *International Journal of Web Based Communities*, vol. 2, no. 2, pp. 133–142, 2006.

[109] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, M. Dean, *et al.*, "Swrl: A semantic web rule language combining owl and ruleml," *W3C Member submission*, vol. 21, no. 79, pp. 1–31, 2004.

[110] G. B. Moody and R. G. Mark, “The impact of the mit-bih arrhythmia database,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[111] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals,” *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[112] J. Hartmann, P. Spyns, A. Giboin, D. Maynard, R. Cuel, M. C. Suárez-Figueroa, and Y. Sure, “D1. 2.3 methods for ontology evaluation,” *EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB Deliverable D*, vol. 1, 2005.

[113] F. Ali, D. Kwak, P. Khan, S. R. Islam, K. H. Kim, and K. S. Kwak, “Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling,” *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 33–48, 2017.

[114] Z. Dokur and T. Ölmez, “Ecg beat classification by a novel hybrid neural network,” *Computer methods and programs in biomedicine*, vol. 66, no. 2-3, pp. 167–181, 2001.

[115] O. T. Inan, L. Giovangrandi, and G. T. Kovacs, “Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features,” *IEEE transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2507–2515, 2006.

[116] O. Sayadi, M. B. Shamsollahi, and G. D. Clifford, “Robust detection of premature ventricular contractions using a wave-based bayesian framework,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 2, pp. 353–362, 2009.

[117] R. J. Martis, U. R. Acharya, K. Mandana, A. K. Ray, and C. Chakraborty, "Application of principal component analysis to ecg signals for automated diagnosis of cardiac health," *Expert Systems with Applications*, vol. 39, no. 14, pp. 11792–11800, 2012.

[118] R. J. Martis, U. R. Acharya, K. Mandana, A. K. Ray, and C. Chakraborty, "Cardiac decision making using higher order spectra," *Biomedical Signal Processing and Control*, vol. 8, no. 2, pp. 193–203, 2013.

[119] T. Li and M. Zhou, "Ecg classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, no. 8, p. 285, 2016.

[120] V. Mondéjar-Guerra, J. Novo, J. Rouco, M. G. Penedo, and M. Ortega, "Heartbeat classification fusing temporal and morphological information of ecgs via ensemble of classifiers," *Biomedical Signal Processing and Control*, vol. 47, pp. 41–48, 2019.

[121] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ecg beat classifier using a mixture of experts approach," *IEEE transactions on biomedical engineering*, vol. 44, no. 9, pp. 891–900, 1997.

[122] Z. Dokur, T. Ölmez, and E. Yazgan, "Comparison of discrete wavelet and fourier transforms for ecg beat classification," *Electronics Letters*, vol. 35, no. 18, pp. 1502–1504, 1999.

[123] Y. Sun, *Arrhythmia recognition from electrocardiogram using non-linear analysis and unsupervised clustering techniques*. PhD thesis, Nanyang Technological University, School of Electrical and Electronic . . . , 2002.

[124] Z. Zhang, J. Dong, X. Luo, K.-S. Choi, and X. Wu, "Heartbeat classification using disease-specific feature selection," *Computers in biology and medicine*, vol. 46, pp. 79–89, 2014.

[125] M. M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.

[126] P. Pławiak and U. R. Acharya, "Novel deep genetic ensemble of classifiers for arrhythmia detection using ecg signals," *Neural Computing and Applications*, vol. 32, no. 15, pp. 11137–11161, 2020.

[127] X. Yang, K. G. Ong, W. R. Dreschel, K. Zeng, C. S. Mungle, and C. A. Grimes, "Design of a wireless sensor network for long-term, in-situ monitoring of an aqueous environment," *Sensors*, vol. 2, no. 11, pp. 455–472, 2002.

[128] Q. Wang, Y. Li, T. Obreza, and R. Munzo-Carpena, "Monitoring stations for surface water quality: Sl218/ss438, 9/2004," *EDIS*, vol. 2004, no. 18, 2004.

[129] H. Liu, X. Chu, Y.-W. Leung, and R. Du, "Minimum-cost sensor placement for required lifetime in wireless sensor-target surveillance networks," *IEEE transactions on parallel and distributed systems*, vol. 24, no. 9, pp. 1783–1796, 2012.

[130] M.-K. Kazi, F. Eljack, and E. Mahdi, "Data-driven modeling to predict the load vs. displacement curves of targeted composite materials for industry 4.0 and smart manufacturing," *Composite Structures*, vol. 258, p. 113207, 2021.

[131] K. Rajnish, V. Bhattacharjee, and M. Gupta, "A novel convolutional neural network model to predict software defects," *Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools and Applications*, pp. 211–235, 2022.

[132] D.-A. Sitar-Tăut, D. Mican, and R. A. Buchmann, "A knowledge-driven digital nudging approach to recommender systems built on a modified onicescu method," *Expert Systems with Applications*, vol. 181, p. 115170, 2021.

[133] L. Chen and C. Nugent, "Ontology-based activity recognition in intelligent pervasive environments," *International Journal of Web Information Systems*, 2009.

[134] F. Doctor, H. Hagras, and V. Callaghan, "A fuzzy embedded agent-based approach for realizing ambient intelligence in intelligent inhabited environments," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 35, no. 1, pp. 55–65, 2004.

[135] A. Nowak-Brzezińska and A. Wakulicz-Deja, "Exploration of rule-based knowledge bases: a knowledge engineer's support," *Information Sciences*, vol. 485, pp. 301–318, 2019.

[136] S. El-Sappagh and M. Elmogy, "A fuzzy ontology modeling for case base knowledge in diabetes mellitus domain," *Engineering Science and Technology, an International Journal*, vol. 20, no. 3, pp. 1025–1040, 2017.

[137] K. Taylor and L. Leidinger, "Ontology-driven complex event processing in heterogeneous sensor networks," in *Extended Semantic Web Conference*, pp. 285–299, Springer, 2011.

[138] R. G. Raskin and M. J. Pan, "Knowledge representation in the semantic web for earth and environmental terminology (sweet)," *Computers & geosciences*, vol. 31, no. 9, pp. 1119–1125, 2005.

[139] P. Wang, J. G. Zheng, L. Fu, E. W. Patton, T. Lebo, L. Ding, Q. Liu, J. S. Luciano, and D. L. McGuinness, "A semantic portal for next generation monitoring systems," in *International Semantic Web Conference*, pp. 253–268, Springer, 2011.

[140] M. K. Kasi, A. Hinze, C. Legg, and S. Jones, "Sepsen: Semantic event processing at the sensor nodes for energy efficient wireless sensor networks," in *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems*, pp. 119–122, 2012.

[141] V. Sharma and H. K. Verma, "Optimized fuzzy logic based framework for effort estimation in software development," *arXiv preprint arXiv:1004.3270*, 2010.

[142] H. Baazaoui, M.-A. Aufaure, R. Soussi, R. Laboratoy, and E. C. U. de la Manouba, "Towards an on-line semantic information retrieval system based on fuzzy ontologies," *Journal of digital information management*, vol. 6, no. 5, p. 375, 2008.

[143] L. A. Zadeh, "Information and control," *Fuzzy sets*, vol. 8, no. 3, pp. 338–353, 1965.

[144] E. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic," *Readings in Fuzzy Sets for Intelligent Systems Dubois, Morgan Kaufmann Publishers, Inc., Los Altos, CA*, pp. 283–289, 1993.

[145] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—ii," *Information sciences*, vol. 8, no. 4, pp. 301–357, 1975.

[146] J. M. Mendel, "Type-2 fuzzy sets and systems: an overview," *IEEE computational intelligence magazine*, vol. 2, no. 1, pp. 20–29, 2007.

[147] E. Sanchez, *Fuzzy logic and the semantic web*. Elsevier, 2006.

[148] F. Bobillo Ortega *et al.*, "Managing vagueness in ontologies," 2008.

[149] O. Castillo, P. Melin, and W. Pedrycz, "Design of interval type-2 fuzzy models through optimal granularity allocation," *Applied Soft Computing*, vol. 11, no. 8, pp. 5590–5601, 2011.

[150] R. A. Aliev, W. Pedrycz, B. G. Guirimov, R. R. Aliev, U. Ilhan, M. Babagil, and S. Mammadli, "Type-2 fuzzy neural networks with fuzzy clustering and differential evolution optimization," *Information Sciences*, vol. 181, no. 9, pp. 1591–1608, 2011.

[151] S. Guillaume and B. Charnomordic, “Learning interpretable fuzzy inference systems with fispro,” *Information Sciences*, vol. 181, no. 20, pp. 4409–4427, 2011.

[152] C. Mencar and A. M. Fanelli, “Interpretability constraints for fuzzy information granulation,” *Information Sciences*, vol. 178, no. 24, pp. 4585–4618, 2008.

[153] J. M. Alonso, C. Castiello, and C. Mencar, “Interpretability of fuzzy systems: Current research trends and prospects,” *Springer handbook of computational intelligence*, pp. 219–237, 2015.

[154] F. Bobillo and U. Straccia, “Fuzzy ontology representation using owl 2,” *International journal of approximate reasoning*, vol. 52, no. 7, pp. 1073–1094, 2011.

[155] J. M. Mendel, “On km algorithms for solving type-2 fuzzy set problems,” *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 3, pp. 426–446, 2012.

List of Research Publications

1. D. Hooda and R. Rani , "Ontology driven human activity recognition in heterogeneous sensor measurements", *Journal of Ambient Intelligence and Humanized Computing* , 11(12), 5947-5960, 2020. **[SCI Indexed, Impact Factor - 7.10]**

2. D. Hooda and R. Rani, "An Ontology driven model for detection and classification of cardiac arrhythmias using ECG data", *Journal of Intelligent Information Systems*, 58(2), 405-431, 2022. **[SCIE Indexed, Impact Factor - 2.5]**

3. D. Hooda and R. Rani, "An interval type-2 fuzzy ontological model: Predicting water quality from sensory data", *Concurrency and Computation: Practice and Experience*. **[SCIE Indexed, Impact Factor - 1.8],** (Accepted)

4. D. Hooda and R. Rani, "A Probabilistic Ontological Model for Recognition of Daily Living Activities ", *Engineering Applications of Artificial Inteligence*. **[SCIE Indexed, Impact Factor - 6.2]**(Under Review)

5. D. Hooda and R. Rani, "Semantic driven healthcare monitoring and disease detection framework from heterogeneous sensor data", In *Proceedings of Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC),IEEE,* ,415-420, 2020.

6. D. Hooda and R. Rani,"Ontology-based Querying from Heterogeneous Sensor Data for Heart Failure" In *Congress on Intelligent Systems 2022,* Springer. [Accepted]