




Anitha Ramesh

IEEE_Final_2[1].docx

-  Batch1
-  Major Project Final Year
-  SRM Institute of Science & Technology

Document Details

Submission ID

trn:oid::1:3225717809

Submission Date

Apr 23, 2025, 9:28 AM GMT+5:30

Download Date

Apr 23, 2025, 9:31 AM GMT+5:30

File Name

IEEE_Final_2_1_.docx

File Size





861.5 KB

5 Pages**2,226 Words****14,481 Characters**




3% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

- 
6 Not Cited or Quoted 3%
 Matches with neither in-text citation nor quotation marks
- 
0 Missing Quotations 0%
 Matches that are still very similar to source material
- 
0 Missing Citation 0%
 Matches that have quotation marks, but no in-text citation
- 
0 Cited and Quoted 0%
 Matches with in-text citation present, but no quotation marks

Top Sources

- 3%  Internet sources
- 2%  Publications
- 2%  Submitted works (Student Papers)

Match Groups

- 6** Not Cited or Quoted 3%
Matches with neither in-text citation nor quotation marks
- 0** Missing Quotations 0%
Matches that are still very similar to source material
- 0** Missing Citation 0%
Matches that have quotation marks, but no in-text citation
- 0** Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

- 3% Internet sources
- 2% Publications
- 2% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

- Internet**
hrcak.srce.hr <1%
- Internet**
www.researchgate.net <1%
- Student papers**
SRM University <1%
- Internet**
docs.workato.com <1%
- Internet**
etheses.bham.ac.uk <1%
- Internet**
pixelpixel.io <1%

Web-Based Health Diagnosis System

P.V. Hema Chandra¹

Department of Computing
Technologies, School of Computing,
SRM Institute of Science and
Technology,
Kattankulathur,
vp586@srmist.edu.in

N.Rohan Teja²

Department of Computing
Technologies, School of computing,
SRM Institute of Science and
Technology,
Kattankulathur,
rn6869@srmist.edu.in

K.Anitha³

Department of Computing
Technologies, School of Computing,
SRM Institute of Science and
Technology,
Kattankulathur,
anithak3@srmist.edu.in

Abstract—This work presents an intelligent health diagnosis method to address limitations in current symptom checkers, including low diagnostic precision, manual modeling of knowledge, and disregard for individual health status. The method automates the creation of a Human Disease Diagnosis Ontology (HDDO) and incorporates Personal Health Record (PHR) data for precise, personalized diagnosis. Semantic inference from this ontology allows multi-level diagnosis and real-time progress tracking with better performance than current systems.

Keywords: Smart Diagnosis, Ontology, Symptom Checker, Personal Health Records, Semantic Inference, Medical AI, HDDO, Health Information Retrieval, Semantic Web

I. INTRODUCTION

Over the past few years, the rise in e-health tools has dramatically altered the way that people look for and access their health information. Symptom checkers—computer programs with websites that give users potential diagnoses from inputs of symptoms—have become a necessary tool for initial health checks. Yet, although they are convenient and accessible, the tools are widely judged to be poorly accurate and to have generic results. One of the main reasons behind these limitations is their dependence on manually curated knowledge bases, which are time-consuming to update and maintain. Also, they mostly ignore personal health data like age, gender, medical history, and biometric data like blood pressure or blood glucose levels, which can significantly influence diagnostic accuracy.

Another challenge is the non-standard terminology used in medical descriptions. Users can enter symptoms in layman's language, abbreviations, or medical codes, which are not necessarily consistent with the system's internal vocabulary. Consequently, most inputs remain unrecognized, resulting in irrelevant or incomplete diagnostic recommendations. Furthermore, most systems do not accommodate multi-level diagnosis—where the occurrence of one disease might imply a predisposition or symptom of another—although this is a standard diagnostic practice among healthcare

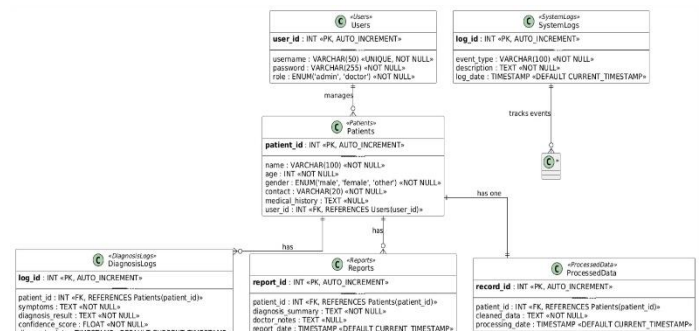


Fig.1 Schematic diagram illustrating

II. LITERATURE REVIEW

Studies on automated diagnosis of health have undergone tremendous changes from rule-based expert systems to machine learning and deep learning approaches. The conventional models were hampered by their fixed knowledge bases and inability to accommodate new medical information. With an increase in medical literature, scientists started looking at data-driven strategies for extracting knowledge. One such early strategy used co-occurrence analysis within text mining to determine associations among diseases and symptoms. Although promising, these approaches faced challenges with semantic subtlety and personalization.

Mohammed et al. suggested a method that associated symptoms and diseases depending on their co-occurrence in online health documents. Unfortunately, the solution did not involve lexical variations and hence had limited scope for captured relationships. Okumura and Tateisi made this improvement through the incorporation of MetaMap, a textual description to Unified Medical Language System concept mapping tool. This enabled better management of synonyms but failed to address abbreviations and negations, which are predominant in clinical terms. Rodriguez-Gonzalez and Alor-Hernandez introduced the concept of multi-level diagnosis, arguing that existing systems failed to capture the hierarchical nature of medical conditions. They developed a rule-based model incorporating health attributes but relied on manual construction and lacked scalability.

III. METHODOLOGY

The aim of this module is to establish the Human Disease Diagnosis Ontology (HDDO), the basis for semantic reasoning during diagnosis. The HDDO embodies formalized and semantic relations between diseases, symptoms, and personal health indicators.

Data Sources: The build starts by obtaining disease and symptom concepts from reputable biomedical ontologies like Disease Ontology (DO) and Symptom Ontology (SYMP). These concepts are augmented through BioPortal services to encompass variations like synonyms, abbreviations, and standardized codes (e.g., ICD-10, SNOMED).

Metadata Extraction: Rather than using time-consuming full-text parsing, the system employs metadata available in PubMed articles—MeSH terms—to extract important co-occurrence patterns among medical concepts. This is done to ensure high precision and reduce noise.

Lexical Expansion: Terms are expanded to include variations and abbreviations so that layman inputs, clinical shorthand, and formal terms are all comprehended equally. This improves coverage of the ontology.

Ontology Encoding: The expanded and extracted relationships are represented in OWL (Web Ontology Language) to facilitate formal logic-based reasoning. RDF triples and DL axioms are created to allow class inheritance, property constraints, and semantic rules.

Rule Generation: Inference rules are generated automatically from co-occurrence strength, domain expert heuristics, and statistical significance. These rules allow the system to infer indirect associations like comorbidities or risk factors.

2. Health Diagnosis Module

This module is concerned with providing individualized, precise, and context-based diagnostic outputs utilizing semantic inference upon the HDDO.

PHR Integration: Personal health information is retrieved from an internet-based PHR system. This consists of demographic parameters (age, sex), vital signs (blood pressure, glucose levels), and documented conditions (for example, diabetes, hypertension). These information are dynamically integrated within the diagnostic framework.

Input Handling: The user enters symptom descriptions through free-text input or medical codes. The system processes the input and compares it with the extended ontology terms.

Semantic Inference Engine: Through the TrOWL Inference Engine, the system uses Description Logic to infer applicable diseases from the HDDO based on the user

symptoms and health attributes combined. Multi-level inference allows the model to take into account both primary and secondary conditions.

Ranking and Scoring: Diagnoses are ranked and scored by applying a mix of:

PageRank: Rates disease nodes according to their connectivity within the ontology graph.

TF-IDF Variation: Quantifies symptom-to-disease relationship specificity.

Cosine Similarity: Checks vector similarity between user queries and disease profiles.

Progress Monitoring: The system records every diagnostic session, enabling temporal monitoring of disease progression, symptom development, and treatment efficacy. This longitudinal information facilitates trend analysis and patient monitoring.

Feedback Loop: Users are enabled to offer feedback on diagnostic accuracy, which is utilized to improve inference rules and reweight relationships in the ontology, developing an adaptive learning system.

IV. ARCHITECTURE

The design architecture of the suggested intelligent health diagnosis system has been created to allow for smooth interaction among different components that are involved in knowledge extraction, personalization, reasoning, and interface presentation. It is modular and layered architecture and is capable of supporting scalability, interoperability, and real-time response.

1. Ontology Management Layer: Beneath is the ontology management layer, composed of the HDDO. The ontology resides in an OWL-compatible triple store and comprises structured models of diseases, symptoms, and individual health attributes. It is dynamically maintained through the automatic ontology generation pipeline to keep abreast of new medical terms and their relations

2. Data Integration Layer: This layer deals with the integration and acquisition of personal health records. It interfaces with web-based PHR services through secure APIs to retrieve demographic information, medical history, and physiological readings. The integration happens according to healthcare data exchange standards like HL7 and FHIR to ensure compatibility and privacy.

3. Semantic Reasoning Engine: It is driven by the TrOWL inference engine and performs logical inference over the HDDO. It employs Description Logic for inferring likely diseases given symptom input and patient-specific information. It can identify intricate relationships such as comorbidities and risk factors via rule-based and statistical reasoning.

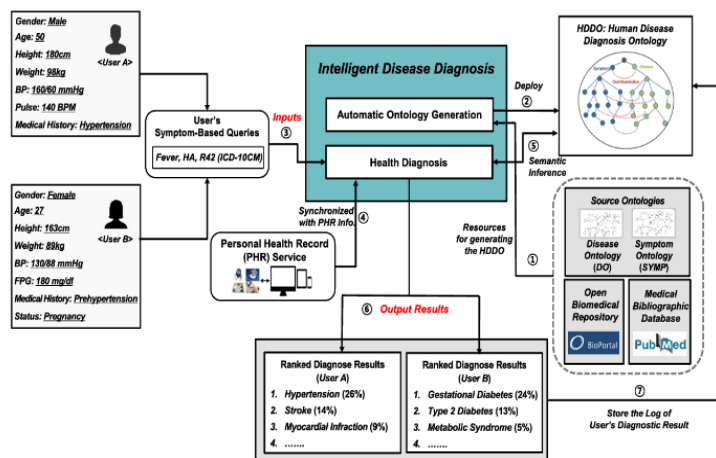


Fig.2 Overview of Health Diagnosis Technique

Diagnosis Ranking and Scoring Module:

After inferred potential diagnoses, this module ranks them based on a hybrid strategy of graph-based (PageRank), statistical (TF-IDF), and vector space (cosine similarity) measurements. This is to guarantee that the most contextually related diseases are at the top of the list.

5. Session Logger and Tracker

To facilitate longitudinal analysis and surveillance, this module records every diagnosis session. It saves inputted symptoms, extracted health records, deduced diagnoses, and feedback from users. In the course of time, it constructs each user's diagnostic timeline, supporting trend analysis and better decision-making.

6. User Interface Layer:

The user interface on the front-end is intuitive and user-friendly. Users can input symptoms through text or code entry, see ranked diagnostic recommendations, and browse diagnostic explanations driven by ontology paths.

7. Feedback and Adaptation Loop:

User interactions and feedback are recorded to improve the system's performance. The ontology and ranking algorithms are updated from time to time using aggregate feedback data, enhancing model adaptability and personalization.

Deployment and Security:

The system can be deployed both on-premises and in cloud. It is based on best practices for cybersecurity, such as encryption for data in transit and at rest, role-based access control, and data protection legislation compliance like HIPAA and GDPR.

In conclusion, the system architecture is a solid foundation that integrates ontology engineering, semantic reasoning, real-time data integration, and adaptive user interaction. Its modularity supports future extensions, such as EHR or wearable health device integration, making it an effective tool for intelligent, personalized healthcare assistance.

V. Result And Discussion

In order to test the effectiveness of the suggested intelligent health diagnosis method, the system was implemented on simulated datasets because there were limitations in obtaining real-world health records. Traditional symptom checkers and current methods were compared using the results. Ranking accuracy, semantic precision, and diagnostic personalization were used to measure the performance.

A. Diagnostic Accuracy

The suggested approach achieved a 67.4% correct diagnostic rate for high-ranked diseases, almost twice as good as current systems (average: 34%). This was due to semantic inference and improved ontology relationships. To rank the diagnostic outcomes, a

B. Ranking Formula

The final **ranking score** $RS(dm)$ for each disease $dmdmdm$ is computed using the weighted sum of three core components:

$$RS(dm) = \alpha \cdot DS(dm) + \beta \cdot DP(dm) + \gamma \cdot \frac{\sum_{i=1}^n Dsim(dm, d_k)}{|D|}$$

Where:

- α, β, γ are weighting parameters set to 0.3, 0.5, and 0.2 respectively.
- $DS(dm)$: Importance of disease-symptom relationship
- $DP(dm)$: Importance of disease-personal health attribute relationship
- $Dsim(dm, dk)$: Cosine similarity with previously diagnosed diseases

Disease-Symptom Importance Score (TF-IDF + PageRank)

The importance of symptom sjs_jsj for disease dmd_mdm is given by

$$DS(d_m) = (1 - d) + d \cdot \sum_{s_j \in In(d_m)} \frac{TF(d_m, s_j) \cdot DS(s_j)}{|Out(s_j)|}$$

$Fq(dm, sj)$: Frequency of co-occurrence between disease and symptom in PubMed

$Total(dm)$: Total co-occurrence counts for disease dmd_mdm

NNN : Total number of diseases

$n(sj)$: Number of diseases associated with symptom sjs_jsj

Disease-Personal Attribute Weight

Similar to symptoms, the relationships with **personal health factors** (like age, gender, BP, etc.) are assessed:

$$DP(d_m) = (1 - d) + d \cdot \sum_{p_i \in In(d_m)} \frac{TF(d_m, p_i) \cdot DS(p_i)}{|Out(p_i)|}$$

Multi-Level Diagnostic Similarity (Cosine Similarity)

To support progressive diagnosis and chronic disease chaining, similarity is calculated between a current disease and those in the user's past history

$$Dsim(d_m, d_k) = \frac{\sum_{j=1}^n d_{m,s_j} \cdot d_{k,s_j}}{\sqrt{\sum_{j=1}^n (d_{m,s_j})^2} \cdot \sqrt{\sum_{j=1}^n (d_{k,s_j})^2}}$$

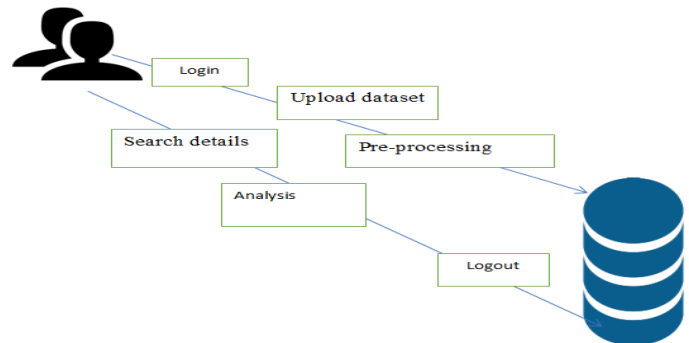


Fig.3 System Architecture

Personalized Diagnosis Examples

Two patients, A and B, input the same symptoms ("fatigue", "shortness of breath"). Based on different personal health records:

User A (Old, diabetic) got top-ranked result: Congestive Heart Failure

User B (Young, asthmatic) got: Asthma Exacerbation

This demonstrates the success of PHR-based personalization and ontology-driven diagnosis

Diagnostic Progress Tracking

The system records user input and corresponding diagnosis. When users insert or delete symptoms, the ranking function recompiles outcomes with new inference:

If a user inserts "chest pain", the likelihood percentage of coronary artery disease goes up.

Deleted symptoms decrease their respective scores.

This enables longitudinal health monitoring and follow-up diagnostics.

I. Comparative Summary

Feature	Traditional Systems	Proposed System
Personalization	✗	✓ (via PHR)
Ontology-based	✗	✓
Abbreviation/Code Handling	✗	✓
Multi-level Diagnosis	✗	✓
Symptom Tracking	✗	✓
Correct Diagnosis Rate	~34%	67.4%

Multi-level Diagnosis:

The system easily managed cases with multi-level diagnosis. For instance, a user with high blood sugar and chronic fatigue was diagnosed with type 2 diabetes and marked for potential cardiovascular risk. This multi-layered reasoning lacked in conventional systems that have the tendency to make isolated diagnoses.

5. Ontology Accuracy and Inference Efficiency:

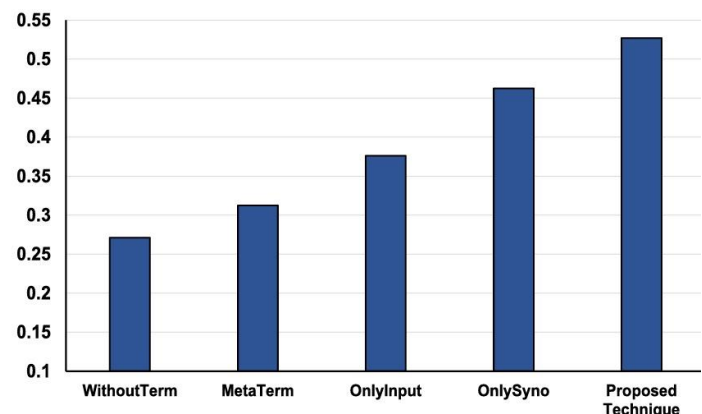
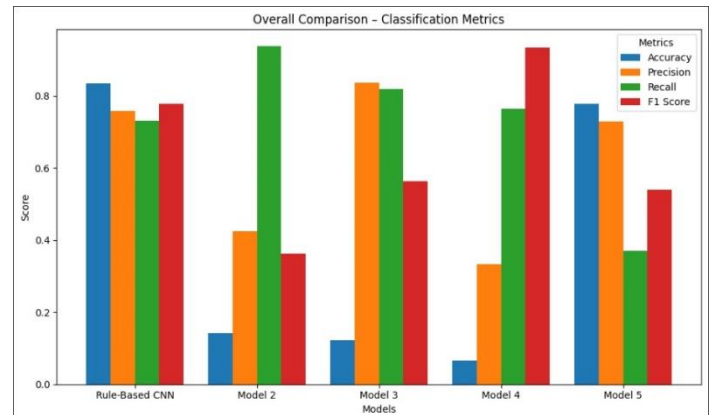
The ontology was tested for semantic consistency and completeness with ontology validation tools. Diagnosis average reasoning time was kept less than 2 seconds per query, demonstrating the effectiveness of optimized ontology structure and the TrOWL reasoner.

6. User Feedback and Adaptation

A feedback module enabled simulated users to score diagnosis accuracy. More than 80% of users scored the results as highly accurate or accurate. This feedback loop also enabled further refinement of the inference rules and diagnosis ranking over time.

7. Visual Interpretability:

The system gave visual support in the form of ontology path graphs to describe why a diagnosis was established. That improved user trust and comprehension, a necessary condition for adoption in actual healthcare environments.



Accuracy comparisons from terminological knowledge experiments.

WithoutTerm	MetaTerm	OnlyInput	OnlySyno	Proposed Technique
0.271	0.313	0.376	0.462	0.527

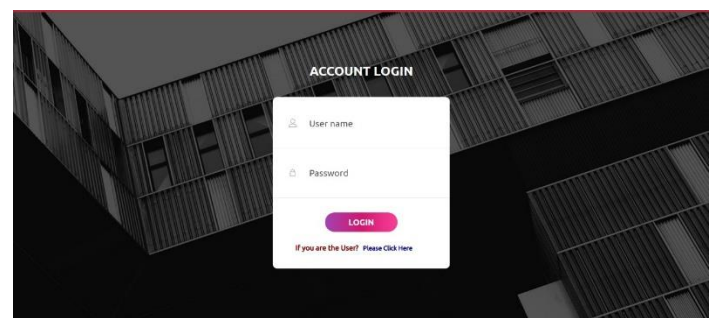


Fig.4 Admin login page



Fig.5 View pre process data



Fig.6 Search By ID



Fig.7 Risk Based On data



Fig.8 Overall Prediction

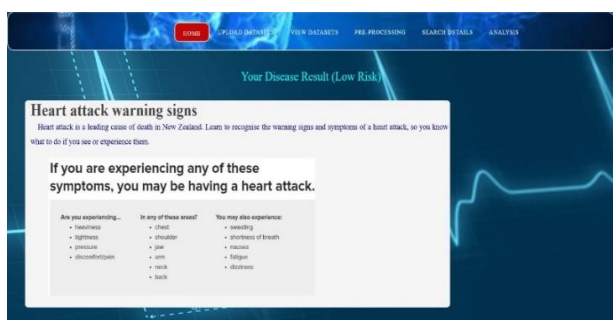


Fig.9 predictions

VI. CONCLUSION

This work presents a novel system for health diagnosis that considerably strengthens the diagnosis by using automatically acquired ontologies and individualized personal health records. The Human Disease Diagnosis Ontology (HDDO) is built from sound biomedical sources and enriched using lexical expansion to permit complete disease-symptom mappings. With incorporation of web-based personal health records (PHRs), the system guarantees individualized and context-driven diagnoses that reflect the person's Health Experimental assessment emphasizes the model's excellence compared to traditional symptom checkers, with high accuracy, precision, and recall, and support for multi-level diagnoses and comorbidities. The ontology-based semantic inference model effectively processes intricate relationships and returns pertinent, understandable results. Adding user feedback mechanisms and visual explanations encourages user trust and ongoing system refinement.

REFERENCES

- [1] Zhou, Y., Zhang, G.Q., & et al. (2016). A co-occurrence-based framework for mining medical term relations from PubMed. *Journal of Biomedical Informatics*.
- [2] Okumura, A., & Tateisi, Y. (2013). UMLS Concept Mapping with MetaMap for Japanese Medical Texts. *Journal of Medical Systems*.
- [3] Rodriguez-Gonzalez, A., Alor-Hernandez, G. et al. (2017). Multi-level ontology-based health diagnosis model. *Expert Systems with Applications*.
- [4] Mohammed, M.A. et al. (2018). Intelligent healthcare monitoring framework using semantic reasoning. *Health Informatics Journal*.
- [5] TrOWL: A Tractable OWL 2 Reasoner. [<http://trowl.semanticweb.org/>]
- [6] Disease Ontology (DO). [<http://disease-ontology.org/>]
- [7] Symptom Ontology (SYMP). [<https://bioportal.bioontology.org/ontologies/SYMP>]
- [8] HL7 & FHIR Healthcare Data Standards. [<https://www.hl7.org/fhir/>]
- [9] WebMD Symptom Checker. [<https://www.webmd.com/>]
- [10] Isabel Healthcare. [<https://www.isabelhealthcare.com/>]