Personalized Asthma Recommendation System: Leveraging Predictive Analysis and Semantic Ontology-Based Knowledge Graph

Ayan Chatterjee[a]
Dept. of Digital Technology, STIFTELSEN NILU, 2007 Kjeller, Norway

Keywords: Predictive Analysis, Personalized Recommendation System, Asthma Monitoring, Semantic Knowledge.

Abstract: Personalized approaches are required for asthma management due to the variability in symptoms, triggers, and patient characteristics. An innovative asthma recommendation system that integrates automatic predictive analysis with semantic knowledge to provide personalized recommendations for asthma management is proposed by this paper. Asthma exacerbation are predicted and the recommendations are enhanced by the system, which leverages automatic Tree-based Pipeline Optimization Tool (TPOT) and semantic knowledge represented in an OWL Ontology (AsthmaOnto). Furthermore, classifications are explained with Local Interpretable Model-Agnostic Explanations (LIME) to identify feature importance. Tailored interventions based on individual patient profiles are provided by this conceptual model, aiming to improve asthma management. The proposed model has been verified using public asthma datasets, and a public weather air-quality dataset has been utilized to support ontology development and verification. In TPOT, the Gaussian Naive Bayes (GaussianNB) classifier has outperformed other supervised machine learning models with an accuracy of 0.75, for the used dataset. To implement and evaluate the proposed model in clinical settings, further development and validation with more diverse and robust datasets with model calibration are required.

1 INTRODUCTION

Asthma is a significant noncommunicable disease affecting individuals of all ages, particularly children (WHO, 2019). It is characterized by inflammation and constriction of the lung's small airways, leading to coughing, wheezing, breathlessness, and chest constriction (WHO, 2019). In 2019, asthma affected an estimated 262 million people, resulting in 455,000 deaths (WHO, 2019). Avoiding triggers is crucial for symptom reduction. Most asthma-related deaths occur in low- to lower-middle-income countries due to under-diagnosis and under-treatment. Asthma is classified as allergic or non-allergic, triggered by factors like allergens, pollution, weather, tobacco smoke, and food allergens (Ajami et al., 2022). Symptoms vary in frequency and severity based on individual reactions to these triggers (Ajami et al., 2022). The WHO is committed to improving asthma diagnosis, treatment, and monitoring to reduce the global burden of noncommunicable diseases and promote universal health coverage (WHO, 2019). Proper lifestyle management and personalized recommendations can help individuals maintain a normal, active life. Effective asthma management involves identifying triggers, controlling symptoms, and preventing exacerbation. However, the diverse nature of asthma requires tailored approaches, making its management complex. Traditional monitoring often relies on oversimplified systems that do not capture the condition's complexity. Conventional methods classify patients as controlled or uncontrolled based on predetermined benchmarks. Personalized asthma management now includes patient-centered care and tailored self-management support, requiring changes in practice organization, consultation modes, and digital health options (Pinnock et al., 2023). Anantharam et al. (Anantharam et al., 2015) developed kHealth, a system that aggregates multisensory data from sensors and questionnaires from asthma patients to help doctors more precisely determine the cause, severity, and control level of asthma, and to alert patients to seek timely clinical assistance. Alharbi et al. (Alharbi et al., 2023) proposed a multi-layered eHealth framework using telemonitoring, environmental sensors, and machine learning to predict asthma attacks and provide safe routes, updating recommendations with real-time data. Another study by Alharbi et al.

[a] https://orcid.org/0000-0002-1051-2814

Chatterjee, A.
Personalized Asthma Recommendation System: Leveraging Predictive Analysis and Semantic Ontology-Based Knowledge Graph.
DOI: 10.5220/0012894500003838
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 2: KEOD*, pages 127-134
ISBN: 978-989-758-716-0; ISSN: 2184-3228

(Alharbi et al., 2021) presented a smart healthcare framework that visualized asthma triggers and provided attack notifications, using deep learning to analyze patient and environmental data, including a dynamic air pollution map and safe-route recommendations. Morita et al. (Morita et al., 2019) designed Breathe, a web-based mHealth platform for managing asthma, which showed initial strong utilization but later declined, with users reporting high satisfaction and confidence in its assessments. Bose et al. (Bose et al., 2021) developed machine learning models to predict persistent asthma in children, with the XG-Boost model performing best, although direct comparisons were difficult due to a lack of similar studies. Kadariya et al. (Kadariya et al., 2019) introduced kBot, a personalized chatbot for pediatric asthma patients, monitoring medication adherence and tracking health data, which showed high acceptance and usability in preliminary evaluations.

Traditional asthma management often relies on standardized guidelines, which, while informative, may not fully account for individual variations in symptoms, triggers, and treatment responses. Predicting and explaining asthma exacerbations remains challenging due to factors like environmental triggers, patient demographics, and medical history. This research introduces a novel **Asthma Recommendation System** designed for personalized asthma monitoring and evidence-based recommendations. The system integrates an automatic prediction pipeline with a semantic knowledge graph, leveraging predictive analysis and semantic reasoning to enhance decision-making accuracy. By using OWL-based ontology and SPARQL querying, the system captures complex data patterns, enriching personalized recommendations. This approach aims to surpass traditional methods by offering more tailored and actionable insights for optimal asthma management (Chatterjee et al., 2021b)(Chatterjee and Prinz, 2022)(Cima et al., 2017)(Chatterjee et al., 2023)(Chatterjee et al., 2022b). The research questions guiding this study are as follows: **a.** How does integrating automatic predictive analysis with a semantic knowledge graph enhance the development of a personalized, evidence-based asthma recommendation system? **b.** How can the system's predictive analysis be effectively explained? **c.** How can recommendations be systematically generated and clarified within the proposed asthma recommendation system?

This research aims to enhance the theoretical understanding of hybrid recommendation generation (automatic predictive analysis combined with semantic rule-based methods) and to foster practical applications for proactive asthma management. The study serves as a technical proof-of-concept, focusing on the conceptual modeling of a recommendation system for asthma, validated with publicly available datasets. Future research will emphasize technical validation using real-world health monitoring data. The paper is organized as follows: Section 2 discusses the proposed system architecture and modeling approaches. Section 3 details the datasets used for predictive model evaluation and semantic modeling for personalized recommendations. Section 4 covers the experimental setup, outcomes, and research questions. Finally, Section 5 presents conclusions and future directions.

2 PROPOSED WORK

The design and development of the proposed asthma recommendation system involve following aspects.

2.1 System Architecture

The proposed system architecture comprises multiple layers distributed across system components, including the Front-end User Interface (UI) and the Back-end server with a database (see Fig. 1). These components communicate via REST (Representational State Transfer) architecture (Barbaglia et al., 2017), ensuring a scalable, secure, and user-friendly framework, using JSON (JavaScript Object Notation) (Barbaglia et al., 2017) for data interchange. The system is composed of the following layers: *Data Acquisition Layer:* Collects asthma data and demographics from electronic health records, surveys, and monitoring devices. Features dashboards, tools, and widgets for recording data, viewing recommendations, and monitoring progress. *Data Integration Layer:* Integrates with healthcare IT systems like EHR, CDSS, and telemedicine platforms, focusing on scalability and security, with FHIR and HL7 standards support (Chatterjee et al., 2022a). *Data Processing Layer:* Applies preprocessing techniques to clean, transform, and standardize data for predictive models. *Predictive Analysis Layer:* Uses TPOT (Olson and Moore, 2016), an AutoML tool, to predict asthma exacerbations and optimize machine learning pipelines with scikit-learn compatibility. *Knowledge Representation Layer:* Develops an asthma ontology (AsthmaOnto) for domain-specific knowledge, including symptoms and treatments. *Recommendation Generation Layer:* Combines predictive results with ontology knowledge to generate personalized asthma management recommendations, mapped to the AsthmaReco ontology. *Automatic Machine Learning Pipeline:* TPOT auto-

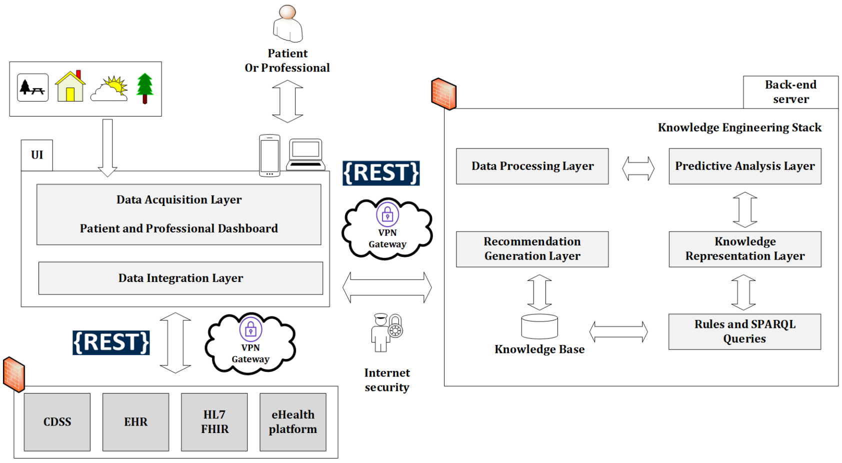


Figure 1: The architecture of the proposed recommendation system.

mates data preprocessing and pipeline construction, reducing manual effort and time.

2.2 Predictive Analysis

TPOT automates the discovery of optimal data preprocessing and machine learning model combinations, improving predictive accuracy and robustness. Using genetic programming, TPOT efficiently evaluates thousands of potential pipelines to identify the best model for a dataset, including hyperparameter optimization. The process includes generating random pipelines, evaluating performance, selecting top performers, and applying crossover and mutation across generations to refine results, ultimately selecting the best pipeline after a set number of generations. The TPOT configuration customizes feature selection methods and multiple sklearn classifiers, including RandomForest, Logistic Regression, SVC, DecisionTree, KNN, and Naive Bayes (Olson and Moore, 2016). The TPOTClassifier is configured with 100 generations and a population size of 100, allowing extensive exploration of the feature space. A high mutation rate of 0.9 and a crossover rate of 0.1 prioritize new solutions over recombination. Models are optimized for accuracy using 5-fold cross-validation. TPOT uses following performance metrics to evaluate and select the best pipelines – Accuracy, Precision, Recall, F1-score, and Matthews Correlation Coefficient (MCC) (Chatterjee et al., 2023)(Chatterjee et al., 2020)(Chatterjee et al., 2021a). Stratification technique, using a Stratified Split (80% training, 20% test), ensures that class proportions in training and test sets match the original dataset (Hutchinson, 1982). This approach reduces bias and enhances reliability in training and evaluation by maintaining consistent performance metrics. While TPOT does not directly provide model explanations, the resulting pipelines has been analyzed with tools like LIME. LIME approximates the model locally around specific predictions, revealing which features significantly influenced the decision (Garreau and Luxburg, 2020). This approach enhances transparency and offers actionable insights to improve patient care and personalized recommendations (Garreau and Luxburg, 2020).

2.3 AsthmaOnto OWL Ontology Model

While traditional databases are effective for structured data storage and retrieval, they fall short in semantic reasoning and integrating diverse data sources. Ontologies offer a more dynamic and interconnected approach, enhancing personalized, context-aware recommendations. This paper presents an OWL ontology for asthma monitoring, encompassing domain knowledge through classes, properties, relationships, individuals, logical operators (AND, OR, NOT), inference rules, and axioms using set theory or first-order predicate logic. The ontology can integrate multiple relevant ontologies, including Asthma from BioPortal, weather and environment from COPDology, food allergens from FoodOn, and symptoms and pollen concepts from SNOMED-CT (Ajami et al., 2022).

Let $G = (V, E)$ be a directed graph where V represents the set of all concepts and E represents the set of all relationships (Chatterjee et al., 2021b)(Chatterjee et al., 2023)(Chatterjee et al., 2022b).

$$V = \{\text{Symptom, Demographic, MedicalHistory, ...}\}$$
$$E = \{(\text{Patient}, \textbf{hasSymptom}, \text{Symptom}), ...)\}$$

SPARQL is a query language for handling RDF data, representing web information in a graph form (Chatterjee and Prinz, 2022). It is vital for extracting and using data from ontologies, particularly in retrieving and managing information from ontology-based knowledge graphs in asthma recommendations (Cima et al., 2017). A basic graph pattern in SPARQL can be represented as:

$$\{?s\,P\,?o\}$$

where $?s$ is a subject variable, P is a predicate, and $?o$ is an object variable.

A triple pattern is an RDF triple where each of the subject, predicate, and object may be a variable.

Let $\mathcal{T}$ represent the set of all triple patterns:

$$\mathcal{T} = \{(?s, P, ?o) \mid ?s, ?o \in I \cup V \cup \mathbb{V} \text{ and } P \in R\}$$

where $\mathbb{V}$ is the set of SPARQL variables. The further terminologies are depicted in Fig. 2.

2.4 Personalized Recommendation

Let X be the feature matrix and y be the label vector. Train a model $f : X \to y$, such as a logistic regression or random forest. For a new patient x, predict the probability of an asthma exacerbation: $\hat{y} = f(x)$. If $\hat{y} > \theta$ (a chosen threshold), generate and recommend preventive measures.

The proposed personalized recommendation generation algorithm (Algo. 1) leverages predictive analysis and semantic knowledge to generate tailored asthma management recommendations. It starts by cleaning and normalizing raw asthma data, followed by training a machine learning model to predict asthma exacerbation probabilities. The system loads an AsthmaOnto ontology, which provides domain-specific semantic knowledge, to enhance the recommendations. For each patient, the system predicts exacerbation likelihood and queries the ontology based on patient attributes, generating personalized asthma management recommendations. The goal is to trigger logical rules of the form **(A IMPLIES B)** or **(NOT(A) OR B)**, providing tailored advice when specific variables are inferred as true.

Time Complexity – Data Preprocessing: $O(n \cdot m)$, where n is the number of records and m is the number of features. Predictive Analysis: $O(t)$ to $O(t \cdot n \cdot m)$, with t as training time. Semantic Knowledge: Ontology loading has $O(1)$. Recommendations: $O(k \cdot q)$, where k is the number of patients and q is query complexity. Overall: $O(n \cdot m + t + k \cdot q)$. **Space Complexity** – Data Preprocessing: $O(s)$ for data. Predictive Analysis: $O(p)$ for the model. Semantic Knowledge: $O(o)$ for the ontology. Recommendations: $O(r)$ for storing results. Overall: $O(s + p + o + r)$. The time and space complexity of Algo. 1 depends on the dataset size, model complexity, ontology size, and query efficiency.

Algorithm 1: Personalized Asthma Recommendation Generation.

Require: • Asthma patient data: $D = \{(x_i)\}_{i=1}^{n}$
• Asthma ontology

Ensure: Personalized recommendations for each patient

Preprocess data: Clean, engineer features, and normalize D to obtain $D_{\text{preprocessed}}$
Train predictive model on $D_{\text{preprocessed}}$ to obtain model M
Load asthma ontology as O
for each patient x_i in $D_{\text{preprocessed}}$ **do**
  Predict exacerbation probability for x_i using M
  Query O for personalized recommendations based on x_i
  Combine prediction and ontology knowledge to generate recommendations for x_i
end for
return List of personalized recommendations for all patients

Potential recommendation messages for personalized asthma management can be represented using the AsthmaOnto ontology and their potential types are mentioned in Table 1.

3 DATA DESCRIPTION

The datasets have been used for predictive analysis and semantic modeling. The proposed hypothesis suggests that amalgamating these datasets could create an asthma recommendation system for ongoing surveillance of health metrics and weather patterns. This system could anticipate potential asthma triggers, generating timely alerts for patients during elevated pollen counts, drastic weather changes, or deviations in vital signs. The "Asthma Disease Prediction" dataset (Dataset, 2024) on Kaggle offers key data for predicting asthma, including demographic details (age, gender), clinical symptoms (tiredness, dry cough, breathing difficulties, sore throat, nasal congestion, runny nose), and medical history (treatment in different medical units, specific medical unit, pneumonia history). In addition to features from the used dataset, weather data from the OpenWeatherMap API (OpenWeather, 2024) is utilized for semantic modeling. The API provides essential weather-related data relevant to asthma management, including temperature (current, feels like, min/max), humidity, air pressure, wind, precipitation, visibility, pollution levels (Air Quality Index, PM2.5, PM10, ozone, nitrogen dioxide, sulfur dioxide, carbon monoxide), and pollen counts. These factors are critical for asthma manage-

The Terminological box (TBox) (Chatterjee et al., 2021b) **contains the vocabulary of the ontology:**

$$\text{Patient}(x) \rightarrow \forall y\,(\text{hasSymptom}(x,y) \rightarrow \text{Symptom}(y))$$
$$\text{Patient}(x) \rightarrow \forall y\,(\text{hasDemographic}(x,y) \rightarrow \text{Demographic}(y))$$
$$\text{Patient}(x) \rightarrow \forall y\,(\text{hasMedicalHistory}(x,y) \rightarrow \text{MedicalHistory}(y))$$
$$\text{Patient}(x) \rightarrow \forall y\,(\text{hasAllergyFactor}(x,y) \rightarrow \text{ExternalAllergyFactor}(y))$$
$$\text{Patient}(x) \rightarrow \forall y\,(\text{hasRecommendation}(x,y) \rightarrow \text{Recommendation}(y))$$
$$\text{Symptom}(x) \rightarrow \exists y\,(\text{Patient}(y) \wedge \text{hasSymptom}(y,x))$$
$$\text{Demographic}(x) \rightarrow \exists y\,(\text{Patient}(y) \wedge \text{hasDemographic}(y,x))$$
$$\text{MedicalHistory}(x) \rightarrow \exists y\,(\text{Patient}(y) \wedge \text{hasMedicalHistory}(y,x))$$
$$\text{ExternalAllergyFactor}(x) \rightarrow \exists y\,(\text{Patient}(y) \wedge \text{hasAllergyFactor}(y,x))$$
$$\text{Recommendation}(x) \rightarrow \exists y\,(\text{Patient}(y) \wedge \text{hasRecommendation}(y,x))$$

The Assertional box (ABox) (Chatterjee et al., 2021b) **contains assertions about "XYZ" individual:**

$$\text{Patient}(\text{XYZ}), \text{hasSymptom}(\text{XYZ}, \text{DryCough}), \text{hasDemographic}(\text{XYZ}, \text{Age 42}), \text{hasMedicalHistory}(\text{XYZ}, \text{Asthma})$$

Logical rules (Chatterjee et al., 2021b) **to infer new knowledge as reasoning can be expressed as:**

$$\forall x(\text{Patient}(x) \wedge \text{hasSymptom}(x,y) \rightarrow \text{Symptom}(y))$$
$$\forall x(\text{Patient}(x) \wedge \text{hasDemographic}(x,y) \rightarrow \text{Demographic}(y))$$
$$\forall x(\text{Patient}(x) \wedge \text{hasMedicalHistory}(x,y) \rightarrow \text{MedicalHistory}(y))$$
$$\forall x(\text{Patient}(x) \wedge \text{hasAllergyFactor}(x,y) \rightarrow \text{ExternalAllergyFactor}(y))$$

Figure 2: Ontology-Based Representation of Asthma Patients' Data.

Table 1: Types of recommendations in Asthma management.

Category	Key Actions
Preventive Measures	Avoid triggers (pollen, dust, pets); use allergen-proof bedding, air purifiers, and close windows during pollen seasons.
Medication Adherence	Take prescribed medications consistently; keep rescue inhaler available.
Environmental Control	Maintain humidity below 50%; clean home regularly; use allergen-proof covers.
Lifestyle Modifications	Exercise regularly; eat a balanced diet; avoid tobacco smoke and pollutants.
Symptom Management	Monitor symptoms; practice relaxation techniques.
Follow-Up Care	Regular appointments; stay current with vaccinations.

ment, as extreme weather conditions, pollution, and high pollen levels can trigger or worsen symptoms.

4 IMPLEMENTATION

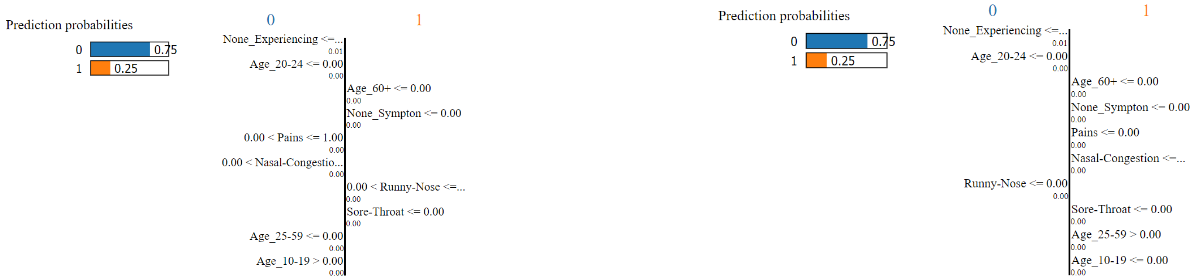
The implementation has been divided into the following subsections:

4.1 Experimental Setup

We used Python 3.9.15 with libraries like pandas (v. 1.5.2), NumPy (v. 1.22.4), SciPy (v. 1.7.3), Matplotlib (v. 3.6.2), Seaborn (v. 0.12.0), and scikit-learn (v. 1.1.3) for data processing and model development. The environment was configured on Windows 10 using Anaconda and Jupyter Notebook 6.5.2. The system had 16 GB RAM and a 64-bit architecture, running models on the CPU due to the dataset's small size. Additionally, we used OWL 2 Protégé in the Protégé editor to design the ontology model, with Hermit reasoner for consistency checking.

4.2 Predictive Analysis

The processed asthma dataset contained 3,16,800 records with 19 features, including 100,820 Class:0 and 33,606 Class:1 entries. Feature selection utilized SelectKBest, which identifies statistically significant features, and VarianceThreshold, which removes low-variance features. Correlation analysis confirmed that features were not highly correlated. TPOT identified the best pipeline, which exclusively used the GaussianNB classifier, achieving an accuracy of 0.75 (Precision: 0.72, Recall: 0.75, F1-score: 0.75, MCC: 0.70). Despite its simplicity, GaussianNB proved effective, likely due to the dataset's alignment with the model's assumption of normally distributed features. Its fewer parameters and simplicity helped prevent overfitting, making it a robust choice for this dataset. The advantages of GaussianNB stem from its appropriate assumption of feature distribution, ef-



(a) Top ten features to predict **No asthma**.

(b) Top ten features to predict **Asthma**.

Figure 3: LIME explanations for predicting class: 0 (No asthma) and class: 1 (Asthma).

$$\forall x,y(\mathrm{Patient}(x) \wedge \mathrm{hasDemographic}(x,y) \rightarrow \mathrm{Demographic}(y))$$
$$\forall x,y(\mathrm{Patient}(x) \wedge \mathrm{hasMedicalHistory}(x,y) \rightarrow \mathrm{MedicalHistory}(y))$$
$$\forall x,y(\mathrm{Patient}(x) \wedge \mathrm{hasAllergyFactor}(x,y) \rightarrow \mathrm{ExternalAllergyFactor}(y))$$
$$\forall x,y(\mathrm{Patient}(x) \wedge \mathrm{hasRecommendation}(x,y) \rightarrow \mathrm{Recommendation}(y))$$
$$\forall x(\mathrm{Symptom}(x) \rightarrow \exists y\,(\mathrm{Patient}(y) \wedge \mathrm{hasSymptom}(y,x)))$$
$$\forall x(\mathrm{Demographic}(x) \rightarrow \exists y\,(\mathrm{Patient}(y) \wedge \mathrm{hasDemographic}(y,x)))$$
$$\forall x(\mathrm{MedicalHistory}(x) \rightarrow \exists y\,(\mathrm{Patient}(y) \wedge \mathrm{hasMedicalHistory}(y,x)))$$
$$\forall x(\mathrm{ExternalAllergyFactor}(x) \rightarrow \exists y\,(\mathrm{Patient}(y) \wedge \mathrm{hasAllergyFactor}(y,x)))$$
$$\forall x(\mathrm{Recommendation}(x) \rightarrow \exists y\,(\mathrm{Patient}(y) \wedge \mathrm{hasRecommendation}(y,x)))$$

Figure 4: Logical Expressions Representing Relationships Between Patients and Associated Data.

fective feature selection, and its probabilistic nature, which supports good generalization without overfitting. This highlights the value of simple models like GaussianNB when they align well with data characteristics. GaussianNB was combined with LIME for predictive explanation, enhancing model transparency and trust. LIME's visualizations use blue color to indicate a push towards the negative class and orange color towards the positive class. The MCC metric provided a balanced evaluation of classifier performance on this imbalanced asthma dataset.

4.3 Semantic Recommendations

The key classes of OWL ontology are shown in Fig. 5 (Axiom: 143, Logical axiom: 82, Declaration axiom: 61, Class: 10, object property: 9, data property: 20, individual count: 23, SubClassOf: 2). Using the Hermit reasoner in Protégé, reasoning was completed in under 40 seconds with no inconsistencies. When loaded in Jena (TTL format, OWL full), reading time was 0.5-1.0 second, and queries for ontology elements executed in under 1.0 second. Each ontology model, as an RDF graph, links to a document manager (default: OntDocumentManager") for document processing. In the ontology API, all ontology classes inherit from OntResource," sharing attributes (versionInfo, comment, label, seeAlso, isDefinedBy, sameAs, differentFrom) and methods (add, set, list, get, has, remove). These logical inferences, based on the ontology, ensure data consistency and validity as depicted in Fig. 4.

The SPARQL queries has been executed successfully to retrieve relevant health conditions from the individual ontology instance to generate personalized recommendations following the criteria: $\mathbf{X} \in \mathbb{R}^n$ be the feature vector representing an individual or item. $\mathbf{w} \in \mathbb{R}^n$ be the weight vector of the predictive model. $\hat{y} \in \mathbb{R}$ be the predicted score obtained from the predictive model. $\theta \in \mathbb{R}$ be the chosen threshold for making recommendations. The predictive score $\hat{y}$ is calculated as: $\hat{y} = f(\mathbf{X}, \mathbf{w})$, where f is the predictive function (e.g., a machine learning model and here, GaussianNB). The decision rule for making recommendations is given by: $\hat{y} = f(\mathbf{X}, \mathbf{w}) > \theta \implies$ Recommend preventive measures. Here are example of successfully executed SPARQL queries in this purpose:

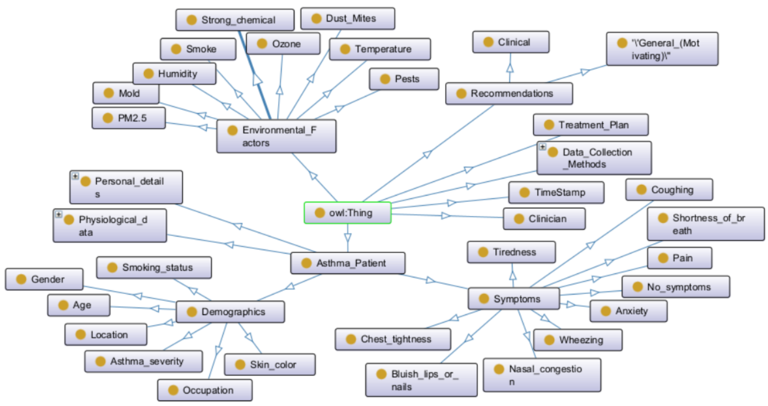


Figure 5: The structure of the proposed AsthmaOnto ontology in Protégé.

Query 1: Patients with specific symptoms, age range, and gender

```
SELECT ?patient
WHERE {
  ?patient :hasSymptom :DryCough .
  ?patient :hasSymptom :DifficultyInBreathing .
  ?patient :hasDemographic ?demographic .
  ?demographic :age ?age .
  ?demographic :gender “female” .
  ?patient :hasMedicalHistory ?medicalHistory .
  ?medicalHistory :pneumonia true .
  ?patient :hasSymptom :NasalCongestion .
  FILTER (?age > 30 && ?age < 50)
}
```

Query 2: Patients with runny nose, asthma prediction and symptom-based recommendations

```
SELECT ?patient ?recommendation
WHERE {
  ?patient :hasSymptom :RunnyNose .
  ?patient :hasDemographic ?demographic .
  ?demographic :gender “male” .
  ?patient :hasAsthmaPredicted ?asthmapredicted .
  FILTER (?asthmapredicted = true)
  ?symptom rdf:type :SoreThroat .
  ?recommendation rdf:type :PreventiveMeasure .
  ?recommendation :relatedTo ?symptom .
}
```

For a 40-year-old asthma patient with symptoms like dry cough and difficulty breathing, a history of pneumonia, and exposure to high pollen levels, the most relevant **RecommendationType** based on the provided SPARQL example would be **“Preventive Measures”**. This involves sending a **RecommendationMessage** that emphasizes proactive steps to minimize asthma triggers and improve indoor air quality, helping the patient manage symptoms and reduce the risk of exacerbations.

The paper integrates automatic predictive analysis with a semantic knowledge graph to enhance personalized asthma monitoring, a novel approach according to the existing literature. It preprocesses a large asthma dataset using SelectKBest and VarianceThreshold to build a robust GaussianNB model, which proved highly accurate. To explain predictions, GaussianNB is combined with LIME, offering visual explanations that increase system transparency. The paper also develops an OWL ontology and uses SPARQL queries for personalized recommendations, ensuring that both predictions and recommendations are precise, personalized, and grounded in semantic knowledge.

5 CONCLUSION

The proposed asthma recommendation system integrates automatic predictive analysis with semantic knowledge representation, offering personalized recommendations to improve asthma outcomes and patient care. TPOT automates machine learning pipeline optimization, identifying the best model and hyperparameters for the asthma dataset, while LIME explanations enhance the interpretability of the GaussianNB model. The use of predictive analysis and the AsthmaOnto OWL ontology provides significant advantages in efficiency, performance, interoperability, and explainability. Future research will focus on refining models, expanding the ontology, and conducting clinical validation to further assess the system's effectiveness.

REFERENCES

Ajami, H., Mcheick, H., and Laprise, C. (2022). First steps of asthma management with a personalized ontology model. *Future Internet*, 14(7):190.

Alharbi, E., Cherif, A., and Nadeem, F. (2023). Adaptive smart ehealth framework for personalized asthma attack prediction and safe route recommendation. *Smart Cities*, 6(5):2910–2931.

Alharbi, E., Nadeem, F., and Cherif, A. (2021). Smart healthcare framework for asthma attack prediction and prevention. In *2021 National Computing Colleges Conference (NCCC)*, pages 1–6. IEEE.

Anantharam, P., Banerjee, T., Sheth, A., Thirunarayan, K., Marupudi, S., Sridharan, V., and Forbis, S. G. (2015). Knowledge-driven personalized contextual mhealth service for asthma management in children. In *2015 IEEE international conference on mobile services*, pages 284–291. IEEE.

Barbaglia, G., Murzilli, S., and Cudini, S. (2017). Definition of rest web services with json schema. *Software: Practice and Experience*, 47(6):907–920.

Bose, S., Kenyon, C. C., and Masino, A. J. (2021). Personalized prediction of early childhood asthma persistence: a machine learning approach. *PloS one*, 16(3):e0247784.

Chatterjee, A., Gerdes, M. W., and Martinez, S. G. (2020). Identification of risk factors associated with obesity and overweight—a machine learning overview. *Sensors*, 20(9):2734.

Chatterjee, A., Gerdes, M. W., Prinz, A., and Martinez, S. G. (2021a). Comparing performance of ensemble-based machine learning algorithms to identify potential obesity risk factors from public health datasets. In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 1*, pages 253–269. Springer.

Chatterjee, A., Pahari, N., and Prinz, A. (2022a). Hl7 fhir with snomed-ct to achieve semantic and structural interoperability in personal health data: a proof-of-concept study. *Sensors*, 22(10):3756.

Chatterjee, A., Pahari, N., Prinz, A., and Riegler, M. (2022b). Machine learning and ontology in ecoaching for personalized activity level monitoring and recommendation generation. *Scientific Reports*, 12(1):19825.

Chatterjee, A. and Prinz, A. (2022). Personalized recommendations for physical activity e-coaching (ontorecomodel): ontological modeling. *JMIR Medical Informatics*, 10(6):e33847.

Chatterjee, A., Prinz, A., Gerdes, M., and Martinez, S. (2021b). An automatic ontology-based approach to support logical representation of observable and measurable data for healthy lifestyle management: Proof-of-concept study. *Journal of Medical Internet Research*, 23(4):e24656.

Chatterjee, A., Prinz, A., Riegler, M. A., and Meena, Y. K. (2023). An automatic and personalized recommendation modelling in activity ecoaching with deep learning and ontology. *Scientific Reports*, 13(1):10182.

Cima, G., De Giacomo, G., Lenzerini, M., and Poggi, A. (2017). On the sparql metamodeling semantics entailment regime for owl 2 ql ontologies. In *Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics*, pages 1–6.

Dataset, A. D. P. (2024). https://www.kaggle.com/datasets/deepayanthakur/asthma-disease-prediction.

Garreau, D. and Luxburg, U. (2020). Explaining the explainer: A first theoretical analysis of lime. In *International conference on artificial intelligence and statistics*, pages 1287–1296. PMLR.

Hutchinson, C. F. (1982). Classification improvement. *Photogrammetric Engineering and Remote Sensing*, 44(1):123–130.

Kadariya, D., Venkataramanan, R., Yip, H. Y., Kalra, M., Thirunarayanan, K., and Sheth, A. (2019). kbot: knowledge-enabled personalized chatbot for asthma self-management. In *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 138–143. IEEE.

Morita, P. P., Yeung, M. S., Ferrone, M., Taite, A. K., Madeley, C., Lavigne, A. S., To, T., Lougheed, M. D., Gupta, S., Day, A. G., et al. (2019). A patient-centered mobile health system that supports asthma self-management (breathe): design, development, and utilization. *JMIR mHealth and uHealth*, 7(1):e10956.

Olson, R. S. and Moore, J. H. (2016). Tpot: A tree-based pipeline optimization tool for automating machine learning. In *Workshop on automatic machine learning*, pages 66–74. PMLR.

OpenWeather (2024). https://openweathermap.org/api.

Pinnock, H., Noble, M., Lo, D., McClatchey, K., Marsh, V., and Hui, C. Y. (2023). Personalised management and supporting individuals to live with their asthma in a primary care setting. *Expert review of respiratory medicine*, 17(7):577–596.

WHO (2019). https://www.who.int/news-room/fact-sheets/detail/asthma.