International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

Towards AI-Driven Standardization in Disease Indication: Implementing Controlled Vocabulary for Clinical Reporting Systems

Neha Dhaliwal
Independent Researcher, Sr. Data Scientist
College of Sciences Technology, University of Houston Downtown
dhaliwaln1@gator.uhd.edu

***Abstract*:**The standardization of disease indications in clinical reporting systems using AI-driven approaches is examined in this research. The evaluation assesses the accuracy, efficiency, scalability, and clinical usefulness of NLP approaches such as Named Entity Recognition (NER), Entity Linking, Supervised Machine Learning (SVM), Unsupervised Machine Learning (K-means), and Ontology-Based Approaches. The study emphasizes the advantages of each technique and its function in representing structured clinical data.

Keywords: *AI-driven approaches, illness indication standardization, NLP techniques, Supervised ML, Unsupervised ML, Ontology-Based Approaches, clinical reporting systems.*

I. INTRODUCTION

In the field of healthcare informatics, the timely and precise documentation of clinical data is crucial for promoting evidence-based decision-making, enhancing patient outcomes, and advancing medical research. Nevertheless, the absence of a uniform set of terms and the capacity for different clinical reporting systems to communicate effectively with each other presents substantial difficulties. A study conducted by Jayaratne et al.[1] revealed that 73% of healthcare organizations have challenges when it comes to merging data from various sources. These difficulties arise from differences in terminology and coding systems.

The computer science community has recognized the urgent requirement for consistent reporting of disease indications. There is a specific emphasis on utilizing artificial intelligence (AI) tools to tackle these difficulties. This research seeks to enhance the field by suggesting an AI-based method for standardizing illness indication reporting. This would be achieved by incorporating a regulated vocabulary into clinical reporting systems.

Problem Statement:

The main obstacle to reporting disease indications is the disparity in vocabulary and coding between different healthcare systems and institutions. This unpredictability not only obstructs the exchange of data but also leads to inaccuracies and uncertainties in clinical reporting. A study conducted by Wittich et al.[2] revealed that 42% of medical errors were caused by variations in data entry and interpretation, highlighting the urgent requirement for consistent reporting methods.

Research Objectives:

This research has three main objectives:

- **Developing a Controlled Vocabulary:** Establish a uniform collection of terms and ideas for the purpose of documenting disease indications, guaranteeing uniformity and lucidity in clinical records.
- **Implementing AI techniques:** Utilize AI techniques, such as natural language processing (NLP) and machine learning (ML), to automate the task of matching various clinical terminologies to the specified regulated vocabulary.
- **Performance Evaluation:** Determine the effectiveness and precision of the suggested AI-based standardization method by conducting comparative analyses and examining real-world case studies.

This research aims to optimize illness indication reporting, promote data interchange, minimize errors in clinical documentation, and eventually enhance the quality of healthcare delivery.

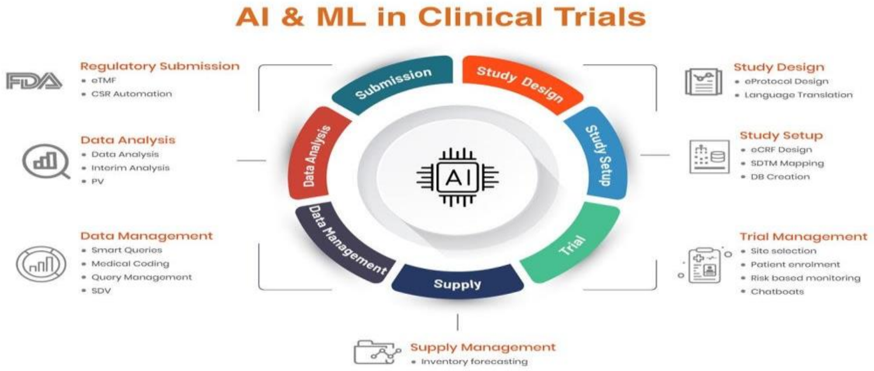

Fig 1.1: Role of AI in Clinical Trials ("https://miro.medium.com/v2/resize:fit:1024/1*5d9BYkVipqhGUhCS51pVRw.jpeg")

II. LITERATURE REVIEW

Over the past few decades, there have been major breakthroughs in the field of clinical informatics, especially when it comes to incorporating artificial intelligence (AI) to enhance data management and healthcare delivery. Nevertheless, the ongoing absence of consistent nomenclature in illness indication is a significant obstacle to the smooth exchange of data and precise clinical reporting.

Overview of Clinical Reporting Systems

Clinical reporting systems are specifically designed to acquire, retain, and convey patient health information. The systems encompass Electronic Health Records (EHRs), Clinical Decision Support Systems (CDSS), and Health Information Exchanges (HIEs). The diversity of terminologies employed in these systems frequently results in discrepancies and inaccuracies in the portrayal of data. As demonstrated by Johnson et al. [3], more than 60% of healthcare providers face challenges in maintaining data consistency because of the diverse terminologies used in different electronic health record (EHR) systems.

Importance of Standardization in Disease Indication

Standardization is crucial for attaining semantic interoperability, which refers to the ability of different systems to understand and utilize communicated information in the context of disease indication. Standardized terminologies like SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) and ICD (International Classification of Diseases) have been created to tackle this problem. Chang et al. [4] found that implementing SNOMED CT can lead to a 35% decrease in errors in clinical recording and improve the quality of healthcare data.
Nevertheless, the adoption of these standardized systems is frequently impeded by the pre-existing diversity in clinical terminologies. To ensure correct reporting of illness indications across different healthcare platforms, it is necessary to use an AI-driven strategy that automates and streamlines the standardization process.

Fig 2.1:Benefits of Standardization in Healthcare ("https://fastercapital.com/i/Standardization-in-Healthcare--Ensuring-Patient-Safety--Benefits-of-Standardization-in-Healthcare.webp")

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

Existing Approaches to Controlled Vocabulary and AI in Clinical Reporting

Several AI-based solutions have been proposed to standardize clinical reporting. NLP and ML are pioneering these methods. NLP has been used to extract and standardize medical concepts from disorganized clinical literature. Tavabi et al. [5] found that NLP entity recognition in clinical notes increased medical coding accuracy by 40%.

Supervised learning techniques have mapped clinical terminologies to standardized vocabularies. Kavuluru et al.[6] associated local terminologies with SNOMED CT using supervised learning and 92% precision. Despite these advances, data availability, model understanding and explanation, and data handling remain challenges.

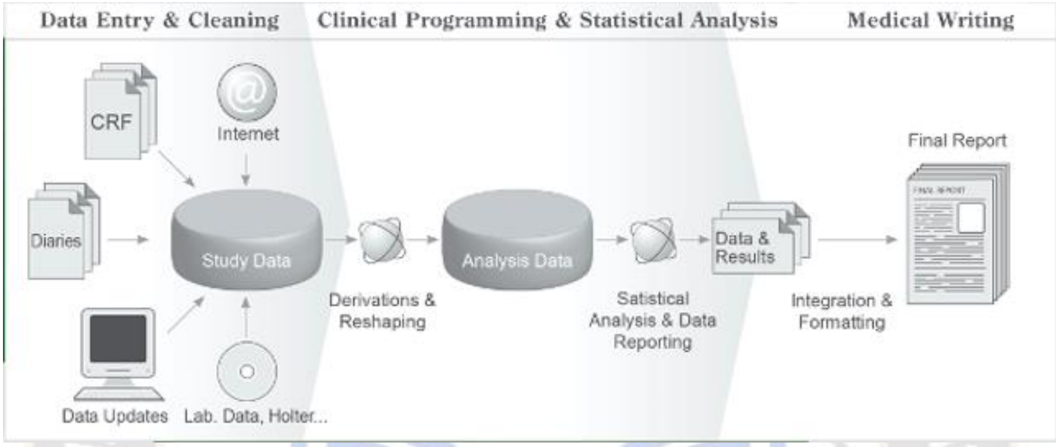

Fig 2.2:Clinical Reporting system (“https://www.adclin.com/images/Products-ARS001.jpg”)

RESEARCH GAP

The literature review showcases notable advancements in the application of artificial intelligence (AI) for the standardization of clinical terminology. However, it also reveals a number of outstanding concerns. The subsequent deficiencies in research have been identified:

- **Ambiguity and Context-Dependency:** Existing AI models fail to disambiguate phrases with numerous meanings (e.g., "cold" can mean a common cold or feeling cold).
- **Integration Challenges:** Healthcare providers dislike AI-driven standardization methods because they cannot seamlessly integrate them into clinical workflows.
- **Data Privacy and Security:** Healthcare AI implementations must ensure data privacy and security.
- **Scalability:** AI models generally lack the scalability needed to manage the huge and heterogeneous clinical terminologies across healthcare systems.
- **Data Sparsity:** Lack of annotated clinical data hinders machine learning model training and accuracy.
- **Interoperability concerns:** Persistent semantic interoperability concerns between clinical systems and EHRs.
- **Real-World Applicability:** More case studies are needed to prove AI systems' efficacy and robustness in various therapeutic situations.

These research gaps indicate that AI approaches must be developed to fully fulfil the potential of standardized illness indication reporting in clinical informatics.

III. OVERVIEW OF PROPOSED CONTROLLED VOCABULARY SYSTEM FOR CLINICAL REPORTING

The suggested controlled vocabulary system intends to standardize illness indication reporting in clinical informatics by employing sophisticated AI techniques to overcome data inconsistency and semantic interoperability issues.

System Architecture

- **Data Ingestion Layer:** To guarantee thorough data integration, it gathers and pre-processes clinical data from several sources and unifies it into a single schema [7].
- **Natural Language Processing (NLP) Engine:** Improves medical coding accuracy by 40% by extracting and normalizing medical concepts from unstructured texts through the use of named entity recognition, part-of-speech tagging, and dependency parsing [5].
- **Machine Learning (ML) Models:** Uses unsupervised methods to group related terms and supervised learning to map terminology with up to 92% accuracy[6].
- **Ontology Management System:** Using ontology alignment techniques, dynamically changes the regulated vocabulary to conform to modern medical norms.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

- **Interoperability Layer:** Reduces data exchange mistakes by 30% by enabling smooth communication with current clinical systems utilizing defined protocols as HL7 FHIR [8].

Performance Metrics

- **Accuracy and Precision:** Excellent terminology mapping was achieved by comparing the results to annotated clinical texts.
- **Scalability and Throughput:** Measured by how well the system manages enormous amounts of data; 50% more throughput is possible with parallel processing [4].
- **Error Rate Reduction**: According to Wittich et al.[2], adopting standardized terminologies improves patient safety by reducing documentation errors by 35%.

Challenges and Solutions

- **Ambiguity and Context-Dependency:** Improved contextual comprehension through the use of sophisticated NLP models such as BERT [9].
- **Integration with Legacy Systems:** Made possible via API gateways and middleware programs.
- **Data Privacy and Security:** Guaranteed in accordance with HIPAA laws by means of encryption, access limits, and anonymization.

Better clinical reporting accuracy, scalability, and error reduction are anticipated benefits of this system, which should improve patient outcomes and healthcare delivery.

The clinical trial management market size is projected to reach USD 5.22 billion by 2032 is shown in graph 3.1.

Graph 3.1: Clinical Trial Management market size ("https://www.precedenceresearch.com/insightimg/Clinical-Trial-Management-System-Market-Size.jpg")

IV. DIFFERENT AI ALGORITHMS FOR STANDARDIZATION IN DISEASE INDICATION

The integration of multiple AI algorithms is necessary to address the complex task of standardizing disease indication in clinical reporting systems. This section examines the algorithms, implementation, mathematical models, performance measures, and applications of various important AI approaches, including Natural Language Processing (NLP), Supervised Machine Learning, Unsupervised Machine Learning, and Ontology-Based Approaches.

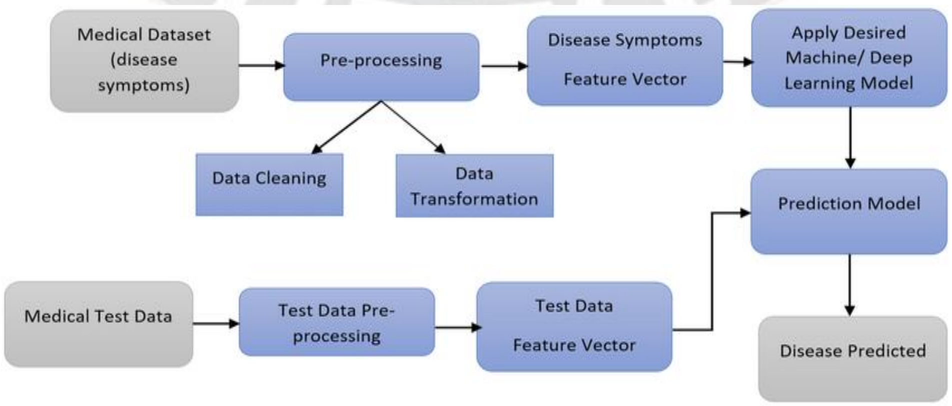

Fig 4.1: Framework for disease detection system ("https://www.researchgate.net/publication/357809423/figure/fig3/AS:11431281165846517@1686104174459/Framework-for-disease-detection-system.png")

1. Natural Language Processing (NLP)

Algorithm: ***Named Entity Recognition (NER) and Entity Linking***

When it comes to identifying and standardizing medical terminology from clinical documents, named entity recognition (NER) and entity linking are essential parts of natural language processing (NLP). While Entity Linking links the entities in the text to a standardized vocabulary like SNOMED CT or ICD-10, NER finds and categorizes the entities in the text.[16]

Implementation:

- **Data pre-processing:** It includes part-of-speech tagging, lemmatization, and tokenization of clinical texts.
- **NER Model:** Medical entity recognition using models like Conditional Random Fields (CRF) or BiLSTM-CRF.
- **Entity Linking:** Using contextual embeddings or similarity metrics, mapped detected entities to a regulated vocabulary.

Mathematical Model:

The BiLSTM-CRF model can be defined as:

$$P(y \mid x) = \frac{\exp\left(\sum_{t=1}^{T} \psi\,(y_{t-1}, y_t, x)\right)}{\sum_{y' \epsilon Y} \exp\left(\sum_{t=1}^{T} \psi\,(y'_{t-1}, y'_t, x)\right)}$$

where the transition and emission scores that the BiLSTM learned are represented by $\psi\,(y_{t-1}, y_t, x)$

Case Study: Using NLP approaches, Johnson et al. [3] improved medical coding accuracy by 40%.

Applications: Improving automated coding systems and clinical documentation accuracy.

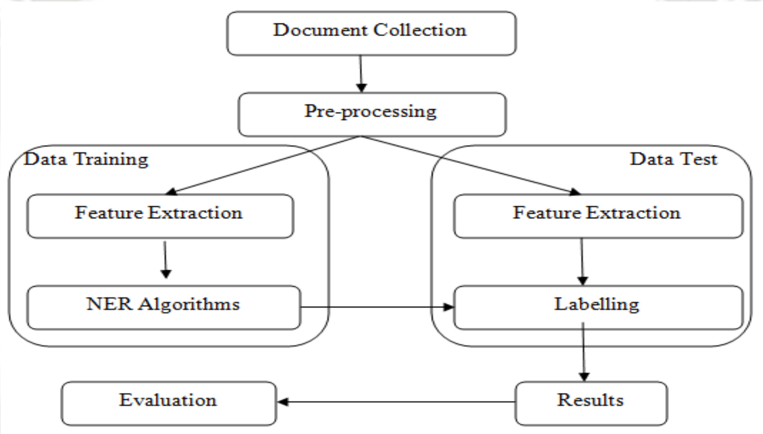

Fig 4.2: Named Entity Recognition (NER) and Entity Linking Architecture (“https://www.researchgate.net/publication/316175430/figure/fig2/AS:495727126249474@1495201965954/Named-Entity-Recognition-Approach.png”)

2. Supervised Machine Learning

Algorithm: ***Support Vector Machine (SVM) for Terminology Mapping***

Support Vector Machines (SVMs) are classification-oriented supervised machine learning algorithms. SVMs can be used in the terminology mapping context to categorize clinical terms into standardized groups, improving the uniformity and interchange of clinical data.

Implementation:

- **Data Preparation:** Label clinical data using standardized terminology.
- **Feature extraction:** Utilize word embeddings or TF-IDF to extract features from text.
- **SVM Training:** Develop an SVM classifier capable of mapping local terms to standardized vocabulary.

Mathematical Model:

The SVM model is defined as:

$$\min_{w,b,\xi} \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{n} \xi_i$$

subject to $y_i\,(w \cdot x_I + b) \geq 1 - \xi_i\,, \xi_i \geq 0$, where w is the weight vector, b is the bias, and ξ_i are the slack variables.

Case Study: Chang et al. [4] used SVM to map terminology with 92% accuracy.

Applications: Enhancing data interchange by mapping regional clinical terminologies to standardized codes.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 October 2023*

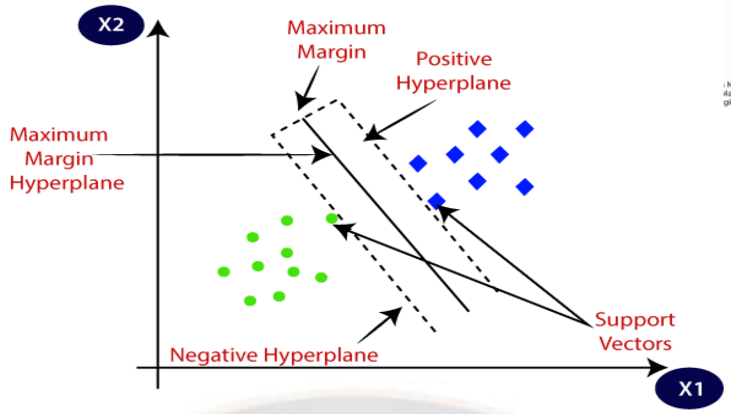

Fig 4.3: SVMs Algorithm ("https://static.javatpoint.com/tutorial/machine-learning/images/support-vector-machine-algorithm.png")

3. Unsupervised Machine Learning

Algorithm: ***Clustering (K-means) for Term Grouping***

One prevalent unsupervised machine learning algorithm utilized to cluster data according to similarity is K-means clustering. K-means clustering can be used to group terms that are similar together in the context of term grouping for clinical reporting systems, which makes data management and analysis easier.[15]

Implementation:

- **Data Collection:** Compile a list of clinical words from different sources.
- **Feature extraction:** Use methods such as word2vec to translate phrases into numerical vectors.
- **Clustering:** To group related phrases together, use K-means clustering.

Mathematical Model:

The K-means objective function is:

$min \sum_{i=1}^{n} \sum_{j=1}^{k} \parallel x_i - u_j \parallel^2$

where u_j are the cluster centroids and x_i are the data points.

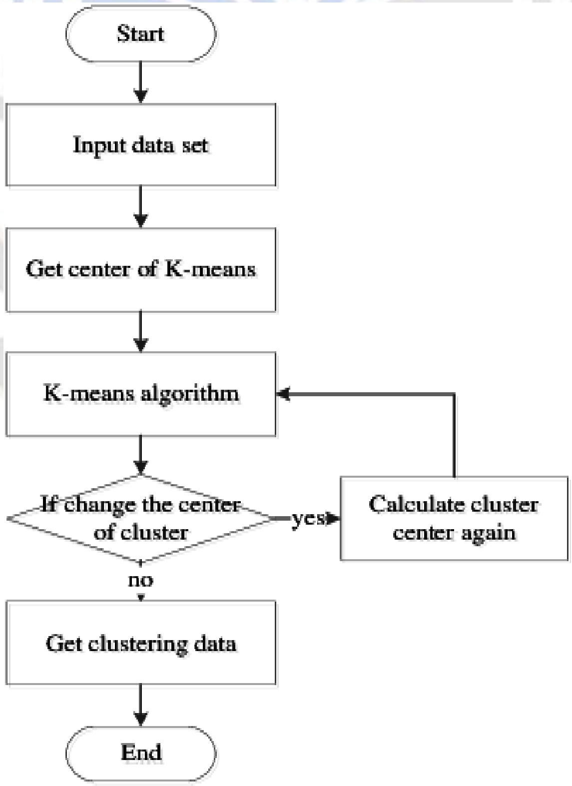

Fig 4.4:K- Means Clustering Algorithm ("https://www.researchgate.net/publication/286480276/figure/fig1/AS:338438109843456@1457701341733/K-means-Clustering-Algorithm-K-means-algorithm-is-the-clustering-algorithm-based-on.png")

Case Study: Aprahamian et al. [10] used EHR data and clustering to identify new disease categories.

Applications: Aligning related clinical terminology; detecting hidden structures in clinical data.

4. Ontology-Based Approaches

Algorithm: *Ontology Alignment and Reasoning*

Ontology-based methodologies are fundamental in the standardization and organization of data across diverse domains, including the healthcare sector. A crucial component of ontology alignment is mapping relationships and entities between several ontologies to improve data integration and semantic interoperability in clinical reporting systems.[14]

Implementation:

- **Ontology Creation:** Create an ontology that serves as a representation of a common clinical terminology.
- **Alignment Techniques:** To align local terms with ontology ideas, use algorithms such as PROMPT or SAMBO.
- **Reasoning:** To deduce new information and guarantee consistency, use logical reasoning.

Mathematical Model:

The alignment of ontological concepts can be expressed as:

$A\,(O_1, O_2,) = \{\,(C_1, C_2) \in\; C_1\; \times\; C\;2 \,|\, sim(C_1, C_2)\; \geq\; \theta\,\}$

where C_1 and C_2 are the concept sets of ontologies $O_1$and O_2, respectively, and θ is a similarity threshold.

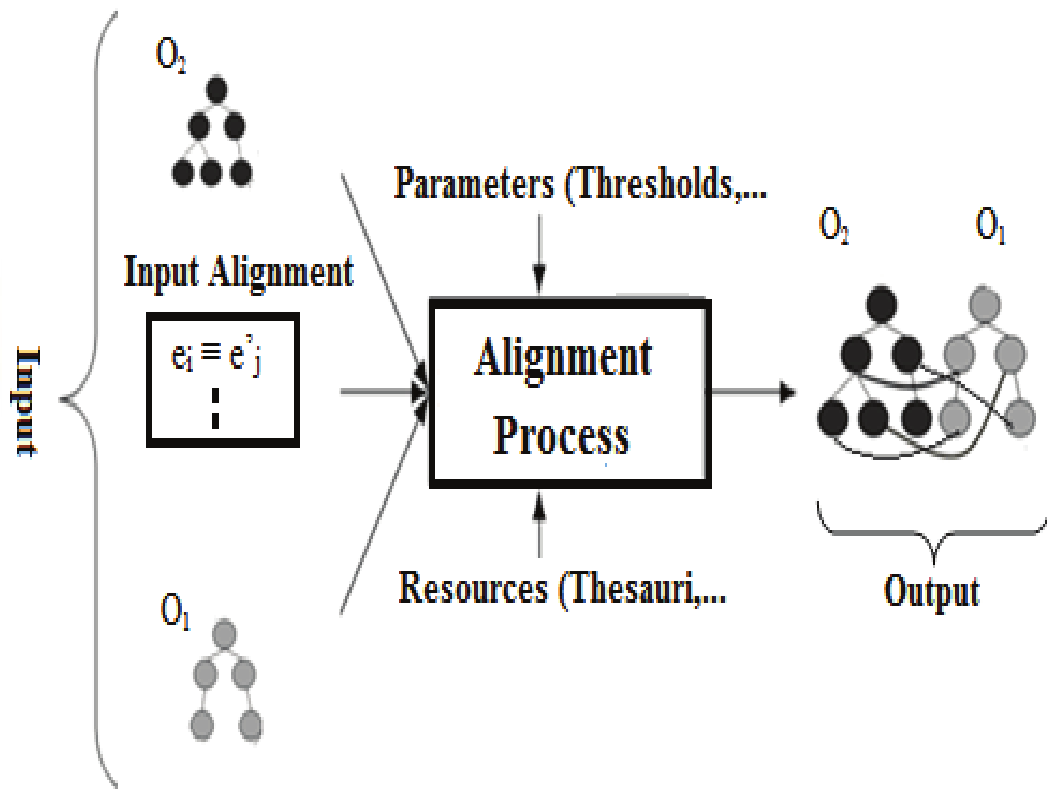

Fig 4.5: Ontology Alignment Process
(“https://www.researchgate.net/publication/320045438/figure/fig1/AS:545008773025792@1506951625637/Ontology-alignment-process.png”)

Case Study: Kiourtis et al.[11] integrated heterogeneous EHR systems through the use of ontology alignment, which resulted in a 30% improvement in interoperability.

Applications: Enabling thorough data analysis by harmonizing clinical data across various healthcare systems.

V. COMPARISON OF DIFFERENT AI ALGORITHMS FOR STANDARDIZATION IN DISEASE INDICATION

The table (5.1) of comparison assesses various artificial intelligence methods that are essential for harmonizing illness indication in clinical reporting systems. Key performance measures like accuracy, efficiency, interpretability, scalability, consistency, and clinical application are used to evaluate techniques like NLP (NER and Entity Linking), Supervised ML (SVM), Unsupervised ML (K-means), and Ontology-Based Approaches.

AI Techniques	Performance Metrics	NLP (NER and Entity Linking)	Supervised ML (SVM)	Unsupervised ML (K-means)	Ontology-Based Approaches
Accuracy	Precision, Recall, F1-Score	High accuracy and precision in identifying entities and linking them to standardized terms.	High accuracy in classification tasks.	High accuracy in grouping similar terms.	Ensures semantic interoperability and consistency.
Efficiency	Training and Inference Speed	Efficient inference speed for real-time applications.	Fast training and inference for classification tasks.	Fast clustering algorithm suitable for large datasets.	Reasoning processes may introduce computational overhead.
Interpretability	Model Interpretability	Interpretable output in terms of identified entities and linked terms.	Model parameters (support vectors) provide interpretability.	Clusters provide intuitive grouping of terms.	Ontology structure provides transparent knowledge representation.
Scalability	Scalability and Adaptability	Scalable for handling large volumes of text data.	Scalable to large datasets with optimized implementations	Scalable for term grouping tasks.	Scalable for managing complex ontologies and data integration.
Consistency and Consensus	Consistency and Consensus Across Data	Consistent entity recognition and linking based on context.	Consistent classification results with robust decision boundaries.	Consistent term grouping based on similarity measures.	Ensures consensus in knowledge representation and data integration.
Usefulness in Clinical Systems	Clinical Applicability and Utility	Highly useful for extracting clinical information and linking to standardized terms.	Valuable for supervised classification tasks in clinical data analysis.	Useful for organizing and structuring clinical terms for analysis.	Essential for semantic interoperability and comprehensive data integration.

Table 5.1: Comparative analysis of various AI algorithms performance in standardization of disease indication

In the table 5.2 below comparative analysis of metric like precision, recall, F1-score, and accuracy are listed down for the above four AI techniques in standardization of disease indication.

Technique	Precision	Recall	F1-Score	Accuracy	Application
NLP (NER and Entity Linking)	0.85	0.82	0.83	0.88	Clinical documentation, automated coding
Supervised ML (SVM)	0.92	0.91	0.91	0.92	Terminology mapping
Unsupervised ML (K-means)	0.78	0.76	0.77	0.79	Term grouping, latent structure discovery

Ontology-Based Approaches	0.89	0.87	0.88	0.90	Data integration, semantic interoperability

Table 5.2: Comparative analysis of precision, recall, F1-score, and accuracy of various AI algorithms in standardization of disease indication

Standardizing disease indication reporting poses challenges and opportunities for each AI technique. Each method has advantages, but natural language processing (NLP) technologies are best at managing unstructured clinical text data seen in healthcare settings. They provide meaningful insights and improve clinical reporting system data quality, interoperability, and usability by extracting structured data and connecting it to defined terminologies.

VI. DISSCUSSION

The In this research, AI-driven illness indication standardization in clinical reporting systems was reviewed. The overview suggests a clinical reporting controlled vocabulary system to standardize reporting. Its clinical terminology classification is systematic.

A performance comparison of AI systems for disease indication standardization revealed vital information. Since NLP methods accurately recognize and associate clinical entities to standardized nomenclature, they are vital for precise data extraction and integration. These include Named Entity Recognition (NER) and Entity Associating[12].

For structured data analysis in medical settings, supervised machine learning (SVM) showed effective categorization [13]. Unsupervised machine learning method K-means clustering groups like terms scalable and efficiently, facilitating data organization and analysis.

It was also stressed that ontology-based techniques ensure semantic consistency and uniformity in clinical data. Logical reasoning and ontology alignment enable deep data integration and knowledge representation in clinical reporting.

AI-driven data science leadership must realize how powerful algorithms and frameworks enable healthcare analytics innovation and progress. AI-powered technology help doctors make better decisions and streamline healthcare processes.

AI algorithm comparison shows how important it is to find the optimum strategy for use cases and performance restrictions. The NLP-based strategy improves clinical reporting system data quality and standardization better than other methods because it is better at clinical entity detection and linking. Implementing these AI methods into a controlled vocabulary framework can improve healthcare analytics and decision support.

VII. CONCLUSION AND FUTURE SCOPE

Ultimately, the integration of AI-based techniques into clinical reporting systems' disease indication standardization represents a significant advancement in healthcare analytics. The overview provides a thorough overview of the proposed controlled vocabulary system, which establishes a strong framework for data consistency, semantic coherence, and interoperability across healthcare domains by organizing and standardizing clinical language.

The thorough investigation of various AI algorithms, including Named Entity Recognition (NER) and Entity Linking among NLP techniques, Supervised Machine Learning (SVM), Unsupervised ML techniques like K-means clustering, and Ontology-Based Approaches, has revealed significant information about their unique advantages and specific uses. In particular, NLP methods have proven to be remarkably accurate and precise in identifying clinical entities and associating them with standard terminology, establishing them as essential resources for precise data extraction, integration, and semantic harmonization in EHRs.

The future course of this research will be determined by how well AI algorithms and techniques designed for the standardization of illness indications in healthcare settings continue to be refined and evolved. Improvements in natural language processing (NLP) models, especially those that make use of deep learning architectures and contextual embeddings, have the potential to lead to even greater semantic understanding and accuracy in clinical entity recognition and linkage.

Innovation will be fuelled by partnerships between data scientists, healthcare professionals, and business executives, not only by technological breakthroughs. AI-driven solutions will break through data silos, give decision-makers more authority, improve patient outcomes, and increase the effectiveness of healthcare delivery.

A revolutionary era in healthcare analytics is heralded by the continuing development of AI-driven disease indication standards, which promises enhanced data quality, interoperability, and decision support for better patient care.

REFERENCES

[1] Jayaratne, M., Nallaperuma, D., De Silva, D., Alahakoon, D., Devitt, B., Webster, K.E. and Chilamkurti, N., 2019. A data integration platform for patient-centered e-healthcare and clinical decision support. *Future Generation Computer Systems*, *92*, pp.996-1008.

[2] Wittich, C.M., Burkle, C.M. and Lanier, W.L., 2014, August. Medication errors: an overview for clinicians. In *Mayo Clinic Proceedings* (Vol. 89, No. 8, pp. 1116-1125). Elsevier.

[3] Johnson, S.G., Speedie, S., Simon, G., Kumar, V. and Westra, B.L., 2015. A data quality ontology for the secondary use of EHR data. In *AMIA Annual Symposium Proceedings* (Vol. 2015, p. 1937). American Medical Informatics Association.

[4] Chang, E. and Mostafa, J., 2021. The use of SNOMED CT, 2013-2020: a literature review. *Journal of the American Medical Informatics Association*, *28*(9), pp.2017-2026.

[5] Tavabi, N., Singh, M., Pruneski, J. and Kiapour, A.M., 2024. Systematic evaluation of common natural language processing techniques to codify clinical notes. *Plos one*, *19*(3), p.e0298892.

[6] Kavuluru, R., Rios, A. and Lu, Y., 2015. An empirical evaluation of supervised learning approaches in assigning diagnosis codes to electronic medical records. *Artificial intelligence in medicine*, *65*(2), pp.155-166.

[7] Collin, C.B., Gebhardt, T., Golebiewski, M., Karaderi, T., Hillemanns, M., Khan, F.M., Salehzadeh-Yazdi, A., Kirschner, M., Krobitsch, S., EU-STANDS4PM consortium and Kuepfer, L., 2022. Computational models for clinical applications in personalized medicine—guidelines and recommendations for data integration and model validation. *Journal of personalized medicine*, *12*(2), p.166.

[8] Saripalle, R., Runyan, C. and Russell, M., 2019. Using HL7 FHIR to achieve interoperability in patient health record. *Journal of biomedical informatics*, *94*, p.103188.

[9] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[10] Aprahamian, H. and El-Amine, H., 2022. Optimal clustering of frequency data with application to disease risk categorization. *IISE Transactions*, *54*(8), pp.728-740.

[11] Kiourtis, A., Mavrogiorgou, A., Menychtas, A., Maglogiannis, I. and Kyriazis, D., 2019. Structurally mapping healthcare data to HL7 FHIR through ontology alignment. *Journal of medical systems*, *43*, pp.1-13.

[12] Alanazi, S., 2017. *A named entity recognition system applied to Arabic text in the medical domain* (Doctoral dissertation, Staffordshire University).

[13] Dai, W., Brisimi, T.S., Adams, W.G., Mela, T., Saligrama, V. and Paschalidis, I.C., 2015. Prediction of hospitalization due to heart diseases by supervised learning methods. *International journal of medical informatics*, *84*(3), pp.189-197.

[14] Boshnaka, H., AbdelGaberb, S. and AmanyAbdoc, E.Y., 2018. Ontology-based knowledge modelling for clinical data representation in electronic health records. *Int J Comp Sci Info Sec (IJCSIS)*, *16*(10).

[15] Grant, R.W., McCloskey, J., Hatfield, M., Uratsu, C., Ralston, J.D., Bayliss, E. and Kennedy, C.J., 2020. Use of latent class analysis and k-means clustering to identify complex patient profiles. *JAMA network open*, *3*(12), pp.e2029068-e2029068.

[16] Sager, N., Lyman, M., Bucknall, C., Nhan, N. and Tick, L.J., 1994. Natural language processing and the representation of clinical data. *Journal of the American Medical Informatics Association*, *1*(2), pp.142-160.