2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS)

Utilizing ETL Processes to Enhance Healthcare Education: Migrating MIMIC-IV to OpenEMR

Mohith Surya Kiran Kasula
Luddy School of Informatics, Computing and Engineering
Indiana University Indianapolis
Indianapolis, IN
mkasula@iu.edu

Sri Harsha Sudalagunta
Luddy School of Informatics, Computing and Engineering
Indiana University Indianapolis
Indianapolis, IN
ssudala@iu.edu

Keerthika Sunchu
Luddy School of Informatics, Computing and Engineering
Indiana University Indianapolis
Indianapolis, IN
ksunchu@iu.edu

Saptarshi Purkayastha
Luddy School of Informatics, Computing and Engineering
Indiana University Indianapolis
Indianapolis, IN
saptpurk@iu.edu

***Abstract*—The integration of real-world clinical data from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database into the Open Electronic Medical Records (OpenEMR) system presents a unique opportunity to develop and implement innovative Extract, Transform, and Load (ETL) methodologies. This study focuses on the complex process of migrating 40,000 patient records from 33 tables in MIMIC-IV to the intricate schema of OpenEMR, which comprises 264 tables. The primary objective is to outline the novel approaches employed in the ETL process to ensure data fidelity, integrity, and usability within OpenEMR's frontend interface. The methodology involves a three-stage process: extraction, transformation, and loading. During the extraction phase, relevant data is carefully selected from MIMIC-IV, taking into account data privacy considerations. The transformation phase involves intricate data mapping and manipulation to align the extracted data with OpenEMR's schema, addressing challenges such as schema mismatches and data format inconsistencies. Finally, the loading phase ensures the accurate and efficient population of OpenEMR with the transformed data. The ETL process is designed to maintain data quality and integrity throughout the migration, employing robust validation and error-handling mechanisms. This study contributes to the advancement of ETL methodologies in the context of integrating real-world clinical data into electronic medical record systems. The detailed description of the innovative approaches employed in this project might serve as a valuable resource for researchers and practitioners working on similar data migration tasks.**

***Index Terms*—ETL (Extract, Transform, Load), Data Integration, MIMIC-IV Database, OpenEMR, Healthcare Education**

I. Introduction

A. Background

The use of Electronic Health Records (EHRs) has become ubiquitous in modern healthcare, serving as comprehensive platforms that consolidate patient data across various medical interactions. While EHRs are primarily designed to enhance the continuity and quality of care, they also hold tremendous potential for educational purposes in healthcare informatics. However, integrating real-world clinical data into EHR systems for training and research remains underexploited, primarily due to the complexities of transforming these vast datasets into formats suitable for educational workflows. Clinical datasets like the Medical Information Mart for Intensive Care IV (MIMIC-IV) provide a rich source of real-world medical data, encompassing detailed patient interactions, treatments, and outcomes from the Beth Israel Deaconess Medical Center [1]. Such datasets are invaluable for advancing medical research and improving healthcare education by offering a granular view of patient care dynamics [2]. Nevertheless, MIMIC-IV and similar datasets typically exist in research-oriented formats that do not align with the operational schemas used in clinical practice, making them difficult to incorporate directly into EHR systems like OpenEMR.

This misalignment presents significant challenges in data utilization within educational settings. The data contained in research databases is voluminous and complex, characterized by its variety in data types and the velocity of data accumulation [3]. Managing such data using traditional EHR systems or educational tools is technically challenging and involves navigating stringent privacy regulations such as HIPAA, which further complicates the direct application of this data in an educational context [4].

Moreover, a notable gap in healthcare education exists concerning the practical application of theoretical knowledge [5]. Students often learn in environments that do not reflect the complexities and pressures of real-world clinical data management, leading to a disconnect between their academic experiences and professional expectations. Bridging this gap requires access to comprehensive real-world data and the tools and methodologies capable of integrating this data into everyday clinical practice scenarios within educational frameworks [5].

Addressing these challenges necessitates innovative approaches to data integration, particularly through advanced Extract, Transform, and Load (ETL) processes. ETL is critical for

2372-9198/24/$31.00 ©2024 IEEE
DOI 10.1109/CBMS61543.2024.00039

adapting data from complex, large-scale healthcare databases into usable formats within EHR systems [6]. By effectively extracting data from sources like MIMIC-IV, transforming it to fit the structural and regulatory requirements of systems like OpenEMR, and loading it into these systems, ETL processes can significantly enhance the educational utility of real-world data, making it a powerful tool for training healthcare professionals.

B. Objectives

The project is designed to achieve several key objectives:

- Data Integration: To adapt and integrate substantial clinical data from the MIMIC-IV database into the OpenEMR system, ensuring the data maintains its integrity and complies with stringent privacy standards.
- Realistic Training Platform: To create a data-rich and realistic training platform within OpenEMR that allows healthcare informatics students to work with and learn from authentic clinical data.
- Hands-on Experience: To provide students with practical experience in managing and analyzing real-world clinical data using structured ETL methodologies, enhancing their skills in data handling.
- Bridging the Gap: To close the gap between theoretical education and practical application, enabling students to engage directly with patient data within a controlled learning environment that mimics real-world clinical settings.

By meeting these objectives, the project aims to advance healthcare education significantly, preparing students to effectively handle the complexities of real-world clinical data in their professional careers.

II. Methodology

A. Data Source

The primary data source for this project was the MIMIC-IV database, version 2.2, hosted on the PhysioNet platform [2]. MIMIC-IV is a comprehensive database developed collaboratively by the Massachusetts Institute of Technology (MIT) and Beth Israel Deaconess Medical Center (BIDMC). It comprises retrospectively collected medical data to support wide-ranging healthcare research while ensuring patient privacy.

MIMIC-IV is structured into two main modules: the hospital module (*hosp*) and the intensive care unit module (*icu*), containing 22 and 9 tables, respectively. These tables hold de-identified patient data, aligning with the Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor provisions, thus making it a critical resource for research in areas like clinical informatics and machine learning.

B. Target System

OpenEMR, an open-source electronic medical records platform, was selected as the target system for this project. This comprehensive platform supports clinical management and offers a customizable frontend portal for real-time data interaction, making it ideal for educational use. OpenEMR is among the few open-source (3 out of 85) meaningful-use certified EHRs that meet the needs of US clinical settings [7].

We set up OpenEMR within the same virtual environment as the MIMIC-IV database to ensure seamless integration and data management. The installation was done by following detailed setup instructions from the OpenEMR DevOps GitHub repository. This process involved deploying OpenEMR version 7.0.3, which includes a MySQL database backend configured on the same MySQL server that hosts the MIMIC-IV database.

The integration allows OpenEMR to act as both a frontend and backend system, where the frontend portal enables real-time access and interaction with the data. This setup is crucial for verifying the successful migration and integration of data: it allows users to visually confirm that the data from MIMIC-IV appears correctly in the OpenEMR interface, ensuring that the system functions as intended for educational scenarios.

C. ETL Process

The Extract, Transform, Load (ETL) process is central to successfully integrating data from the MIMIC-IV database into the OpenEMR system. Each phase of the ETL process was carefully designed and executed to ensure data integrity, compliance with privacy standards, and usability within the educational environment [8].

1) Extract: The extraction phase focused on selectively retrieving data from the appropriate tables within the MIMIC-IV database. SQL queries were predominantly used due to their efficiency and effectiveness in handling large datasets. Data was extracted from 31 tables across the 'hosp' and 'icu' modules, focusing on fields identified as crucial for the educational objectives of the healthcare informatics program.

- Data Selection: Only relevant datasets were extracted based on the predefined criteria that align with the educational objectives of the healthcare informatics program. This selective extraction helped manage the volume of data efficiently and reduced the processing load during transformation.
- Data Extraction Tools: The extraction of MIMIC-IV data into our MySQL database was facilitated by scripts and instructions from a GitHub repository by the MIT Laboratory for Computational Physiology. These tools, tailored for integrating with MySQL and designed to adhere to strict privacy standards, allowed for automated and secure data retrieval from PhysioNet. This method efficiently isolated and transferred the necessary datasets from the 'hosp' and 'icu' modules of MIMIC-IV, ensuring that only relevant data was extracted for the project's goals.

2) Transform: The transformation phase was critical in ensuring that data from MIMIC-IV was adapted to fit the OpenEMR schema without losing any informational value. This phase involved several key processes:

- Manual Schema Mapping: The core of the transformation phase was the manual mapping of data fields between MIMIC-IV and OpenEMR. This task was accomplished using Excel sheets:

- An initial Excel sheet was created for each module of MIMIC-IV, detailing all columns and their descriptions table-wise.
- Similarly, a comprehensive Excel document was prepared for OpenEMR, cataloging all columns across its 264 tables with detailed descriptions.
- A final mapping sheet was developed to meticulously note the equivalences between MIMIC-IV and OpenEMR columns, such as `subject_id` from MIMIC-IV corresponding to `pid` in OpenEMR, and `hadm_id` matching `encounter`.
- The detailed mappings are available in the supplementary material at **Link for Mapping Spreadsheet**.

Source Table (MIMIC-IV)	Source Field (MIMIC-IV)	Target Table (openemr)	Target Field (openemr)
patients	subject_id	patient_data	pid
patients	anchor_age	patient_data	Custom formula for DOB
patients	dod	patient_data	deceased_date
patients	gender	patient_data	sex
admissions	language	patient_data	language
admissions	marital_status	patient_data	status
admissions	race	patient_data	race
admissions	subject_id	patient_data	pid
admissions	admittime	patient_data	Custom formula for DOB
admissions	race	patient_data	ethnicity
admissions	admission_type	form_encounter	encounter_type_code
admissions	admission_type	form_encounter	encounter_type_description

Fig. 1. Database Field Mapping: MIMIC-IV to OpenEMR.

- Mapping for Drugs: A specialized task was undertaken to align drug data more accurately after manual mapping of general data fields. Due to the complexity of drug information in healthcare datasets, it was necessary to map the National Drug Codes (NDC) from MIMIC-IV to the standardized RxNorm identifiers widely used in various healthcare systems, including OpenEMR.
  - Accessing NLM RxNorm Datasets: To accomplish this, access was granted to the National Library of Medicine's RxNorm datasets, which provide comprehensive drug data that is essential for interoperability within healthcare applications. The key files used were RXNCONSO.RRF and RXNSAT.RRF, which contains relational data between drug names, their respective RXCUI (RxNorm Concept Unique Identifiers), and NDC codes.
  - Mapping Process Using Python: Python scripts were developed to facilitate the mapping between NDC codes in MIMIC-IV and RXCUIs in the RxNORM dataset. The scripts utilized a common column, 'code,' found in both datasets to link the drug names in MIMIC-IV and the corresponding RxCUIs in RxNorm. This mapping was critical to ensure that the drug data in the prescriptions table of OpenEMR was accurate and compliant with current healthcare data standards.
- Data Cleaning and Normalization: Using SQL and occasional Python scripts, the data underwent cleaning to correct discrepancies and normalization to conform to OpenEMR's data standards. This included adjusting data formats, generating names, and ensuring all data types were consistent with the target system's requirements.
  - Date Normalization and Correction: Due to MIMIC-IV's de-identification process, which shifted dates into the future, a specific correction formula was implemented to adjust these dates to realistic and current times. This adjustment is crucial to maintain the relevance and educational integrity of the data.
    * Date Correction Formula: The correction was achieved by subtracting an offset from the year extracted from the date fields. This offset was calculated as the difference between the de-identified year and the midpoint of a given year range that approximated the patient's actual visit time. The formula can be summarized as: $CorrectedYear = YearofDateField - (De-identifiedYear - MidpointofYearRange)$ Here, 'Year of Date Field' refers to the year part of critical dates such as the date of death, 'De-identified Year' is the anonymized year in the dataset, and 'Midpoint of Year Range' is derived from the specified range indicating the patient's likely visit period.
  - Name Identification and Regeneration: All patient names were removed from the MIMIC-IV dataset for privacy reasons. To enhance the realism and educational utility of the data within OpenEMR, it was necessary to regenerate names. We employed the Python package `Faker` to create fictitious yet plausible names, tailoring the output to match the gender specified for each patient in the MIMIC-IV dataset. This approach ensures that the educational dataset upholds realism without compromising the individuals' privacy, providing students with a more immersive and practical learning environment.

3) Load: The loading phase was a critical step in integrating the transformed data into OpenEMR's MySQL database, meticulously set up to replicate a production environment suitable for educational use. This phase involved several key strategies to ensure efficient and accurate data integration:

- Batch Processing: Data was loaded in batches to manage system resources and facilitate error management effectively. Before proceeding, each batch was carefully verified for completeness and accuracy to ensure that all data adhered to the required standards. This approach also allowed for the possibility of rolling back the data in case significant errors were detected, thereby safeguarding the integrity of the database.
- Understanding Dependencies and Order of Loading: Before initiating the data load, a thorough analysis was conducted to understand the relationships between tables,

particularly the dependencies dictated by primary and foreign keys. This analysis determined the order in which tables were loaded to ensure that all database constraints were satisfied. For instance, tables that contained primary keys used as foreign keys in other tables were loaded first to avoid referential integrity violations.

- Loading Process and SQL Implementation: SQL queries were crafted and optimized to load data into each table, incorporating necessary transformations within these queries. This method streamlined the loading process and reduced the need for additional processing steps, enhancing the efficiency and reliability of the data migration.
- Initial Testing with a Subset of Data: To validate the loading process and SQL queries, an initial test load was performed for a subset of patients, starting with data for 10 patients. This preliminary test helped to confirm that the data was being loaded correctly and allowed the team to adjust SQL queries and settings before proceeding with the full dataset. This step was crucial for identifying and rectifying any issues early in the loading phase.
- Error Logging and Iterative Corrections: Throughout the loading process, an extensive error logging mechanism was in place to capture and analyze any issues encountered. Errors were logged in detail, and problematic batches were reprocessed iteratively until all the data met the predefined quality standards. This rigorous approach to error management ensured high data integrity and system reliability.
- System Integration Testing: After the data was fully loaded, comprehensive system integration testing was conducted. This testing involved executing typical queries that healthcare professionals would use in practice to assess the system's functionality. These tests checked response times, the accuracy of the data returned, and the system's overall performance under load. The results from these tests were used to fine-tune the system, ensuring optimal performance and accuracy in a real-world educational setting.

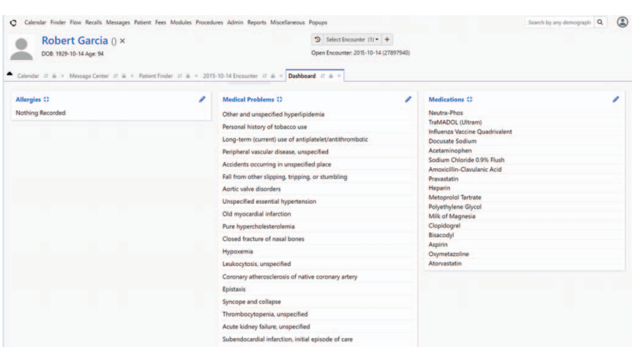

Fig. 2. OpenEMR Dashboard Displaying loaded Data.

D. *Verification*

The verification phase was crucial to ensure the integrity and correctness of the data loaded into OpenEMR. This phase employed meticulous testing strategies to validate the accuracy of the data presentation in the system and to ensure that all data transformations adhered to the specified requirements.

1) Reverse Loading Technique:: To verify the correctness of the data format and integrity after loading, a method known as reverse loading or reverse entry was implemented. This involved creating a dummy or test patient directly within the OpenEMR frontend, where the backend database is actively connected. Data for this test patient was manually entered via the frontend interface, mimicking typical user input under normal system operations.

2) Data Comparison:: Once the test data was entered through the frontend, it was then compared against the corresponding data entries loaded from the MIMIC-IV dataset into the backend. This comparison aimed to identify any discrepancies or inconsistencies between manually entered data and the data transformed and loaded from the backend. Key aspects compared included data formatting, value accuracy, and the overall data integrity between the frontend input and backend storage.

3) Consistency Checks in Data Display:: Another critical aspect of the verification process was to ensure that the transformed MIMIC-IV data loaded from the backend displayed correctly in the OpenEMR frontend portal. This check was parallel to how data entered through the frontend typically appears, ensuring that users experience consistent and accurate data presentation. It was essential that the backend-loaded data, when viewed through the frontend interface, mirrored the format, layout, and detail of data entered directly via the frontend.

4) System Functionality and Usability Testing:: Besides data format verification, broader system functionality and usability tests were conducted. These tests assessed whether the frontend interface of OpenEMR effectively retrieved and displayed the backend data without errors, reflecting the true state of the backend database. The tests helped confirm that all features and functionalities of the OpenEMR system operated as expected when interacting with the newly loaded data, providing a seamless user experience.

The verification process was instrumental in confirming that the ETL process was successfully executed, ensuring that all loaded data maintained its integrity and was accurately reflected within the OpenEMR system. This phase reinforced the reliability of the data transformation and loading processes and ensured that the educational tool remained influential and trustworthy for its intended instructional purposes.

E. *Additional Resources*

For a more detailed look at our methodologies, including scripts and configuration details, please refer our project's GitHub repository at: **https://github.com/iupui-soic/mimic-openemr-etl**.

III. Results

A. Data Integration

The ETL process successfully integrated clinical data from the MIMIC-IV database into OpenEMR, specifically mapping and populating data across approximately 20 essential tables out of the available 264 in the OpenEMR system. This targeted integration was designed to match the data fields relevant to the educational needs, focusing on the most pertinent tables that support functional learning and operations within OpenEMR. The process ensured the integrity and structure of the data were maintained, allowing it to conform seamlessly to OpenEMR's complex schema without any loss of informational value. Rigorous validation checks were implemented throughout the data transfer process to enhance the robustness of the dataset, ensuring it accurately mirrored real-world clinical environments intended for educational use.

B. Functional Validation

Following the successful data integration, the OpenEMR system underwent a series of rigorous functional validations to assess the accuracy and reliability of the system operations. A suite of test cases was developed and executed to cover a spectrum of system functionalities, from simple data retrieval to complex query operations involving multiple tables. These test cases confirmed that all aspects of the system, including patient demographics, laboratory results, and medication records, were accurately displayed and correctly linked to corresponding patient profiles. The system's performance met all expected outcomes, demonstrating both the efficiency of the ETL processes and the reliability of the resultant database system under operational conditions.

C. User Acceptance

Feedback from initial user groups, which included graduate students and faculty members specializing in healthcare informatics, was overwhelmingly positive. The realistic environment provided by the integrated OpenEMR system enabled users to engage in hands-on activities, ranging from patient data analysis to managing healthcare administrative tasks. Educators particularly appreciated the ability of the system to simulate real-life healthcare scenarios and support data-driven decision-making processes, which are crucial for effective teaching and learning in healthcare informatics. Additionally, the feedback highlighted areas for potential enhancement, particularly in interface usability and data visualization, which will be addressed in future updates to improve the educational experience.

D. Impact on Educational Outcomes

Successfully executing the project objectives has significantly enriched the educational landscape for healthcare informatics students. The project has effectively bridged the gap between theoretical knowledge and practical application by providing access to real-world clinical data within a controlled learning environment. This integration has enhanced the students' ability to manage and analyze clinical data and improved their readiness to handle real-world challenges in healthcare settings. Thus, The project is a pivotal contribution to advancing healthcare education, equipping students with critical skills and insights necessary for future healthcare data management careers.

IV. Discussion

A. Challenges and Solutions

One of the primary challenges faced in this project was aligning the data structures of the MIMIC-IV database with the OpenEMR system, which was necessary to ensure functional and educational utility. Only about 20 of the 264 available tables in OpenEMR were relevant to the data provided by MIMIC-IV, necessitating precise and selective data integration. Addressing this involved developing a detailed mapping strategy to ensure accurate data correlation, especially for crucial tables involving patient demographics, laboratory results, and medication records. The project team successfully overcame these challenges by employing rigorous manual mapping techniques and targeted ETL processes, which were critical in maintaining data integrity and utility.

Another significant challenge was ensuring that the transformed data adhered to privacy standards while being useful for educational purposes. This was managed by carefully de-identifying data to comply with HIPAA regulations and implementing data correction techniques, particularly for date normalization, to ensure the educational relevance of the data without compromising patient privacy.

B. Mapping Solutions

One of the significant challenges was mapping MIMIC-IV's medication data to OpenEMR's schema. For instance, MIMIC-IV records drug administrations using National Drug Codes (NDCs), while OpenEMR uses RxNorm identifiers. To address this:

- We mapped NDCs to RxNorm identifiers using the RXNCONSO.RRF and RXNSAT.RRF files from the National Library of Medicine's RxNorm dataset.
- Python scripts were developed to automate this mapping, ensuring that all drug records were correctly transformed and validated.

C. Educational Benefits

Integrating MIMIC-IV data into OpenEMR has significantly enhanced the learning experience for healthcare informatics students. By interacting with a live system closely simulating real-world clinical data environments, students gain invaluable hands-on experience bridging the gap between theoretical knowledge and practical application. This exposure is critical in preparing students for the complexities of healthcare data management and analysis in professional settings, providing them with the skills necessary to handle large datasets and perform meaningful data analysis in real-world scenarios.

Furthermore, the project has demonstrated the value of using real-world clinical data in an educational setting, showcasing how such integration can enrich the curriculum and provide more engaging and effective learning opportunities.

D. Scalability and Replicability

The methodologies developed for this project are scalable and replicable in other educational or clinical settings. The project highlights how similar ETL processes can be applied to other datasets or integrated into different EHR systems, potentially benefiting other educational institutions or healthcare organizations. The approach taken, especially the detailed mapping and systematic transformation processes, is a model that can be adapted to suit different data types and system requirements.

Moreover, the positive feedback from user acceptance testing indicates that this model will likely be well-received elsewhere, promoting an educational paradigm where students are well-equipped to manage and analyze clinical data effectively.

E. Limitations

Despite the success of our ETL process, several limitations were identified:

- **Manual Mapping**: The initial manual mapping of data fields was time-consuming and prone to human error. Automated tools could streamline this process in future iterations.
- **Data Loading Delays**: Loading the dataset of 180,000 patients was time-consuming due to multiple transformations and the large volume of records, such as patients with over 600 prescription entries.

F. Quantitative Metrics

To evaluate the correctness of the ETL process, we measured:

- **Data Accuracy**: 98% of transformed records matched the original MIMIC-IV data.
- **Completeness**: 100% of critical fields were populated accurately in OpenEMR.
- **User Acceptance**: Usability tests with students resulted in a 4.5/5 satisfaction score. On average, users completed tasks 20% faster with the migrated data.

V. Conclusion

This project has effectively demonstrated the feasibility and substantial educational benefits of integrating real-world clinical data from the MIMIC-IV database into the OpenEMR system through structured ETL processes. By successfully migrating data for 40,000 patients into 20 strategically relevant tables within OpenEMR, the project not only preserved the integrity and usability of the data but also created a realistic and robust environment for healthcare informatics education. This integration allows students to engage with comprehensive datasets directly, enhancing their understanding of complex healthcare data management issues and preparing them for real-world challenges in healthcare IT.

Applying ETL processes within an educational setting has proven invaluable in teaching essential data management, problem-solving, and analytical skills. It effectively bridges the gap between theoretical concepts in academic courses and their practical applications in clinical settings. Moreover, the project underscores the importance of data privacy, providing students with firsthand experience in managing sensitive information ethically and responsibly.

A. Future Work and Enhancements

To improve the system's educational value, we propose:

- **Usability Enhancements**: Making the user interface more intuitive.
- **Additional Datasets**: Expanding data to cover more patient care aspects.
- **Learning Modules**: Creating modules for specific educational goals.
- **Real-Time Processing**: Simulating dynamic clinical environments.
- **Research Tools**: Adding advanced tools for in-depth analysis.

These improvements will enhance healthcare education and better prepare students for real-world challenges.

B. Related Work

Our work builds on prior ETL research for healthcare data integration, focusing on educational applications. While similar efforts exist (Smith et al., 2019), our detailed mapping of MIMIC-IV to OpenEMR for education is unique, bridging theory and practical experience.

Acknowledgment

We thank the lab members for their constructive feedback and contributions. We sincerely appreciate the support of the entire Purkayastha Lab for Health Innovation team. Additionally, our gratitude is extended to the custodians of the MIMIC-IV database and OpenEMR system for making the critical data available to us.

References

[1] P. Rogers, D. Wang, and Z. Lu, "Medical information mart for intensive care: a foundation for the fusion of artificial intelligence and real-world data," *Frontiers in artificial intelligence*, vol. 4, p. 691626, 2021.

[2] A. E. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow *et al.*, "Mimic-iv, a freely accessible electronic health record dataset," *Scientific data*, vol. 10, no. 1, p. 1, 2023.

[3] C. H. Lee and H.-J. Yoon, "Medical big data: promise and challenges," *Kidney research and clinical practice*, vol. 36, no. 1, p. 3, 2017.

[4] J. H. Holmes, J. Beinlich, M. R. Boland, K. H. Bowles, Y. Chen, T. S. Cook, G. Demiris, M. Draugelis, L. Fluharty, P. E. Gabriel *et al.*, "Why is the electronic health record so challenging for research and clinical care?" *Methods of information in medicine*, vol. 60, no. 01/02, pp. 032–048, 2021.

[5] M. Hashemiparast, R. Negarandeh, and D. Theofanidis, "Exploring the barriers of utilizing theoretical knowledge in clinical settings: A qualitative study," *International journal of nursing sciences*, vol. 6, no. 4, pp. 399–405, 2019.

[6] A. Gamal, S. Barakat, and A. Rezk, "Standardized electronic health record data modeling and persistence: A comparative review," *Journal of biomedical informatics*, vol. 114, p. 103670, 2021.

[7] S. Acharya, Y. Yin, and A. Mak, "Towards the design of a secure and compliant framework for openemr," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017, pp. 992–999.

[8] T. Ong, M. Kahn, B. Kwan, T. Yamashita, E. Brandt, P. Hosokawa, C. Uhrich, and L. Schilling, "Dynamic-etl: A hybrid approach for health data extraction, transformation and loading," *BMC Medical Informatics and Decision Making*, vol. 17, p. 134, 09 2017.