Research Article

**Program Evaluation using Data Science approaches to associate student grades to course learning outcomes in Graduate Health Informatics**

**Akshita Patri** 1**, Vedasree Yerrabandi** 1**, Sadia Afreen** 1**, Saptarshi Purkayastha** 1

1 Dept. of BioHealth Informatics, Indiana University, Indianapolis,  
Indianapolis, IN, USA

**Abstract**

**Purpose:** The study aimed to evaluate student competency in the Health Informatics program at the Department

of BioHealth Informatics, Indiana University-Purdue University Indianapolis (IUPUI). Specifically, we sought to

determine the alignment of course assessments with learning objectives and to identify areas for improvement to

enhance student achievement of these objectives.

**Methods:** We analyzed data from over 300 Health Informatics students enrolled between 2017 and 2021.

Instructors aligned course assessments with learning objectives to assess student competency in specific topics.

A comprehensive data-driven approach was employed to objectively discuss potential assessment modifications,

course learning outcomes, and teaching methods with instructors.

**Results:** Our findings indicated that, on average, 74% of students met the course learning objectives. The

analysis was presented in a Program Review and Assessment Committee (PRAC) report, which highlighted

areas for improvement and provided valuable insights for enhancing the Health Informatics program.

**Conclusion:** This study underscores the importance of assessing student competencies and aligning course

objectives with assessments. By identifying areas for improvement and implementing changes, we aim to

increase the percentage of students achieving at least 80% of the course learning objectives to 80% or higher.

Regular program reviews are essential for ensuring that the program meets the needs of its students and can

serve as a model for other programs seeking to perform quantitative program assessments.

**Keywords**

Competencies; Health Informatics; Learning Objectives; Program assessment;

**Introduction**

**Background/Rationale**

The dynamic field of Health Informatics (HI), which blends healthcare, information technology, and social sciences, is rapidly expanding, with the intent to enhance patient outcomes and healthcare delivery [1]. To equip students to  
confront multifaceted healthcare challenges using technology and data, graduate programs in HI are gaining momentum [2]. As with any graduate program, it is essential to evaluate the effectiveness of Health Informatics programs to ensure that it meets the needs of students and employers.

Graduate programs play a pivotal role in higher education, offering specialized knowledge and skills to those seeking to advance their expertise. Program evaluation is not only crucial for satisfying student and stakeholder needs, but also for highlighting potential areas for improvement [3]. Traditional evaluation methods primarily depend on surveys, interviews, and other qualitative techniques. While these can yield significant insights, they can be time-consuming and potentially fail to comprehend the full complexity of program outcomes. Data science approaches, such as machine learning and predictive analytics, have emerged as promising alternatives for program evaluation [4]. These approaches enable the analysis of large datasets to identify patterns and trends that may be difficult to detect using traditional methods. In recent years, data science approaches have emerged as a powerful tool for evaluating graduate programs [5]. By leveraging advanced analytical techniques and large datasets, data science approaches can provide insights into program effectiveness, student outcomes, and the impact of interventions [6].

**Objective**

This paper delivers the results of a study on the competencies of graduate Health Informatics students at Indiana  
University-Purdue University Indianapolis (IUPUI) Department of BioHealth Informatics, within the Luddy School of Informatics, Computing, and Engineering. Our research focused on the correlation between course objectives and learning outcomes, using data from over 300 students from 2017-2021. The evaluation allowed us to measure the percentage of students who met the course learning objectives and detect areas of the program that need improvement. The findings suggest enhancing assessments, course learning outcomes, and pedagogy. The study highlights the importance of analyzing student competencies, aligning course objectives with assessments, and conducting regular program reviews.

Program description

The Master of Science in Health Informatics (MS in HI) at IUPUI is a comprehensive, 36-credit-hour program, providing multidisciplinary knowledge that impacts the health, safety, and efficiency of those involved in the healthcare system. The MS in HI program and its preceding Bachelor of Science in Health Information Management (BS in HIM) program are both accredited by CAHIIM. This distinct BS-MS-PhD pathway is one of only five such fully accredited pathways nationwide. Alongside the Bachelor of Science in Biomedical Informatics program (with tracks in Health Informatics and Bioinformatics), this pathway stands unique in the Midwest and across the United States.  
Our graduates are at the forefront of medical technology, driving the creation of electronic health records and clinical information systems, safeguarding the privacy and security of health data, and supporting clinical teams in harnessing digital devices and emerging technologies to reach more patients. They are actively involved in diverse projects across hospitals, doctors’ offices, insurance companies, government agencies, and health IT software companies. As the adoption of Electronic Health Records (EHRs) has escalated, and healthcare provider reimbursement is increasingly tied to the demonstrable quality of care, the demand for our graduates has seen a notable surge in recent years.  
The MS in HI program offers both fully online and on-campus delivery formats. The program includes two accelerated degree pathways, one from the BS in Health Information Management (4+1 degree), and another from BS degrees in Biology, Public Health, or Health & Human Sciences.

The program further encompasses five graduate certificate programs:  
1. Clinical Informatics.  
2. Public Health.  
3. Health Information Management and Exchange.  
4. Health Information Security.  
5. Health Information Systems Architecture.

Program structure:

The 36-credit hours are split into 12 courses of 3 credit hours with at least 1 core and 2 electives that can be taken as variable credits ranging from 1-3 in a semester. 7 core courses cover a broad range of competencies, and the analysis presented in the paper is on these courses.

• INFO-I 501: Includes analyzing research and data, designing efficient programs, investigating questions, and  
employing machine learning techniques.  
• INFO-B 505: Teaches project management for computer-based knowledge-work projects, covering terminology,  
techniques, and technologies. It emphasizes productive team dynamics and optional professional certification.  
• INFO-B 518: Teaches students to analyze biomedical data, using statistical methods to solve cutting-edge prob-  
lems in genomics, proteomics, and health data.  
• INFO-B 530: Covers Health Informatics fundamentals and how computer technology enhances healthcare  
research and practice. It explores informatics principles for communication systems, clinical decisions, and  
evidence-based medicine.

A table of research paper

Description automatically generated

• INFO-B 535: Covers human-computer interface, health care decision support, system selection, project management, system evaluation, regulatory policies, economic impacts, and future trends.  
• INFO-B 581: Helps students explore the design process of data standards, including domain analysis, medical  
terminology, conceptualization, modeling, and common methods and tools.  
• INFO-B 642: Covers design principles, math foundations, clinical vocabularies, legal/ethical issues, patient-  
centered clinical decision support systems, and applications in clinical practice.

Program characteristics

The MS in Health Informatics program has grown tremendously in the last 10 years. As seen in Figure 1a, the program had steady growth until 2018 and showed rapid growth in the last 3 years. From equal gender distribution, now the program has more than double the number of female students compared to male students. This is similar to the trend seen in other HI graduate programs in the US, but the ratio 2:1 female-male ratio is not as skewed in other programs.

Figure 1b shows the race/ethnicity distribution of the student body. The racial and ethnic distribution of the program is also uncommon among many other HI graduate programs in the nation. While a large percentage of international students is common in many programs, until 2018, the program had a fairly large White and Hispanic/Latino student body. Due to the large international student interest in the last 3 years, the ratio of White and Hispanic/Latino students has rapidly decreased. Like the Health Informatics discipline, our program struggled to attract African American students. Hopefully, with concerted DEI efforts, more African American students will join our program.

Curricular philosophy

Our curriculum is closely aligned with the HI discipline curriculum requirements defined by the CAHIIM accreditation requirements. We have undergone a curricular review for the accreditation, which has helped align our program outcomes with the CAHIIM/AMIA accepted HI core competencies. The curricular design philosophy for health informatics competencies recognizes that healthcare is becoming increasingly data-driven, and that health informatics plays a critical role in leveraging data to improve patient outcomes and healthcare delivery. The philosophy aims to equip students with the knowledge, skills, and competencies necessary to effectively manage and analyze health data, design and implement health information systems and use technology to improve clinical decision-making and patient care.

A graph with numbers and a line

Description automatically generatedA graph of ethnic distribution over last 10 years

Description automatically generated

**Figure 1:** (a) Gender of students in MS HI program (b) Ethnicity of students in MS HI program

Some key elements of the philosophy include:  
• Using technology to improve patient outcomes: The philosophy emphasizes using technology to collect, manage, and analyze health data to improve patient outcomes, reduce costs, and enhance the quality of care.  
• An interdisciplinary approach: Health informatics draws on principles and practices from multiple disciplines,  
including health sciences, computer science, information technology, and management. Philosophy recognizes the importance of this interdisciplinary approach and seeks to provide students with a comprehensive understanding of these fields.  
• Flexibility and customization: The curricular design philosophy for health informatics competencies recognizes  
that different healthcare settings and populations have unique needs and challenges with continuous changes in  
policies and legislative changes for health systems. As such, it is designed to be flexible and customizable to  
meet these needs.  
• Emphasis on ethics and privacy: The philosophy recognizes the importance of ethical considerations and patient  
privacy in health informatics. It emphasizes the need for students to understand and adhere to relevant ethical  
and legal standards in their work.  
• Emphasis on lifelong learning: The philosophy recognizes that healthcare is constantly evolving and that health  
informatics professionals must be prepared to continually update their skills and knowledge. As such, it emphasizes the importance of lifelong learning and provides opportunities for ongoing professional development.  
We strive to offer a comprehensive yet deep preparation that touches the field’s fundamental methodological and  
theoretical areas, emphasizing the professional knowledge needed to succeed in the industry and selected research and application areas connected to the strengths of our faculty.

A diagram of a diagram

Description automatically generated A circular diagram with different colored lines

Description automatically generated

**Figure 2.** Health Informatics Competencies **Figure 3:** Linkage of Core courses and Program Learning Outcomes

1.4 Program Learning Outcomes

We have taken a data science approach to evaluate the courses. We analyze the curriculum by relating the Program Outcomes (PO) with the Course Learning Outcomes from the 6 core courses and then map the outcomes and any missing gaps to achieve the CAHIIM competencies. The following are our Program Learning Outcomes, with Figure 3 showing the linkages between the core courses and Program Learning Outcomes.:

1. PO1: Analyze problems: Analyze, understand, abstract, and model a specific biomedical problem regarding its data, information, and knowledge components.

2. PO2: Produce solutions: Use the analysis to identify and understand the space of possible solutions and generate designs that capture essential aspects of solutions and their components.

3. PO3: Implement, evaluate, and refine: Carry out the solution (including obtaining necessary resources and managing projects), evaluate it, and iteratively improve the solution.

4. PO4: Innovate: Create new theories, typologies, frameworks, representations, methods, and processes to address biomedical informatics problems.

5. PO5: Work collaboratively: Team effectively with partners within and across disciplines.

6. PO6: Understand the fundamentals of the field in the context of the effective use of biomedical data, information, and knowledge.

7. PO7: For substantive problems related to scientific inquiry, problem-solving, and decision-making, apply, analyze, evaluate, and create solutions based on biomedical informatics approaches.

8. PO8: Apply, analyze, evaluate, and relate biomedical information, concepts, and models spanning molecules to individuals to populations.

9. PO9: Analyze and evaluate complex biomedical informatics problems in terms of data, information, and knowledge.

10. PO10: Apply, analyze, and create data structures, algorithms, programming, mathematics, statistics.’

11. PO11: Apply, analyze, and create technological approaches in the context of biomedical problems.

12. PO12: Apply and evaluate methods of inquiry and criteria for selecting and using algorithms, techniques, and methods to solve substantive health informatics problems.

**Methods**

Ethics statement: This was not a study with human subjects; therefore, neither approval by the institutional board nor obtainment of the informed consent was required

Study Design: This study follows a descriptive analysis.

2.1 Data Collection

To evaluate student competencies in specific Health Informatics topics, we contacted instructors. They were asked to indicate which assignments or quizzes were used to measure student understanding as per the course learning objectives. To aid the process, we shared a Word document with a table containing each learning outcome, sourced from the university website. Instructors were then asked to identify the correlation between assessments and learning objectives by mapping outcomes to corresponding assignments/quizzes.

2.2 Data Preprocessing

After receiving inputs from instructors, we connected the course learning outcomes (CLO) to various assessments like assignments, discussions, quizzes, projects, etc. This enabled us to pinpoint assessments employed to measure specific topics. Next, student grades on these assessments were analyzed to gauge their competencies. To ensure student privacy, we anonymized the data and shuffled the rows to eliminate any name-based or grade-based sorting. Data cleaning was performed to exclude null values, and the dataset was divided into semesters using a naming convention. Each row in the dataset represents a student and their final course grade, with a specific column for the CLO. Our analysis focused on calculating the percentage of students scoring above 80% for each CLO in every semester. While maintaining privacy, data was presented semester-wise, covering Fall 2020, Spring 2021, Fall 2021, Spring 2022, and Fall 2022.

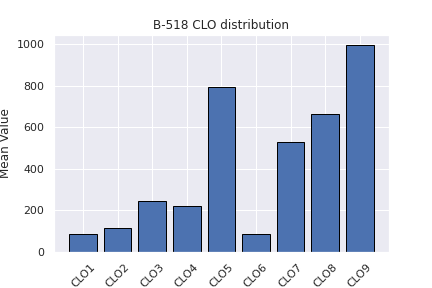
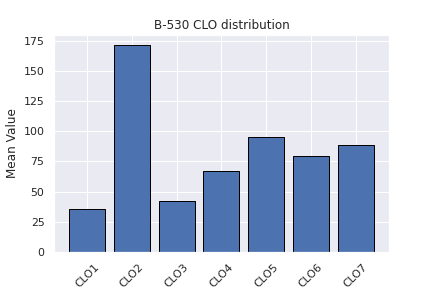
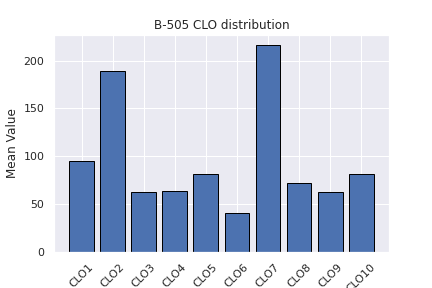
2.3 Exploratory Data Analysis

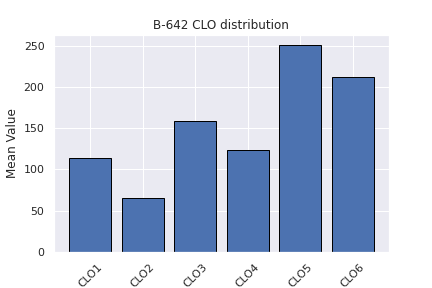
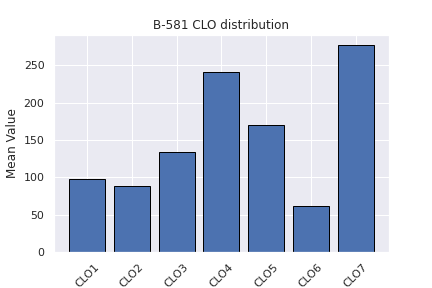
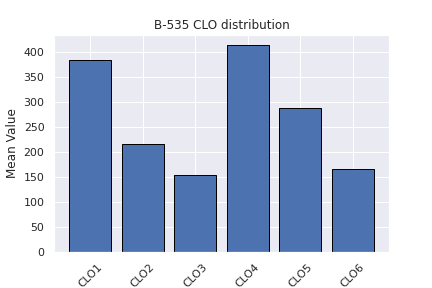
We carried out exploratory data analysis (EDA) for each learning outcome to understand grade distribution and identify trends. We employed descriptive statistics, including mean, median, standard deviation, and quartiles, to examine grade distribution. Visualization tools such as bar charts were utilized to identify potential trends within a specific course’s learning objectives. This analysis enabled us to spot areas of student strengths and areas needing improvement.

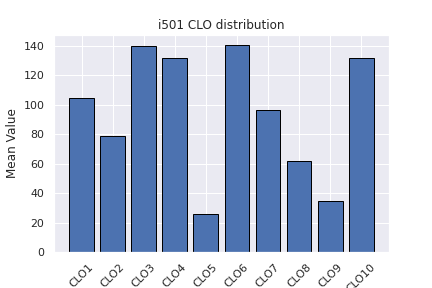
Correlation Mapping: A correlation mapping was performed to discern the relationship between each CLO and the final grade. The correlation table displayed correlation coefficients between each CLO and the final grade, with the latter being the dependent variable. This process helped identify the most significant CLOs influencing the final grade, and those needing further attention. For B530, the final grades showed a strong positive correlation of either 0.9 or 0.91 with all the CLOs. In B505, the final grades exhibited a stronger positive correlation ranging from ≥ 0.89 to ≤ 0.91, with the highest correlation of 0.91 observed for multiple CLOs. In I501, the final grades displayed a stronger positive correlation ranging from≥ 0.9 to≤ 0.94, with the highest correlation of 0.94 found for specific CLOs. For B535, the final grade showed a uniform strong positive correlation of 0.9 with all the CLOs. In B518, the final grade exhibited a strong positive correlation, with a coefficient of 0.94 for multiple CLOs. B581 showed a stronger positive correlation of 0.91 or 0.92 between the final grades and the CLOs. Lastly, in B642, the final grades displayed highly positive correlations of either 0.95 or 0.94 with the CLOs. Many of these were milestone assignments in the course, or after the program review, they were made to be milestone assignments. The correlations we performed highlight the importance of the CLOs in influencing the final grades, emphasizing the significance of addressing each objective for overall course performance.

2.4 Existing tools for Program Assessment

CAHIIM Self-Evaluation Tool Assessment (CSET): The CAHIIM Self-Evaluation Tool Assessment (CSET) is an all-inclusive, standardized tool, developed by the Commission on Accreditation for Health Informatics and Information Management Education (CAHIIM). It is designed to evaluate the effectiveness of Health Informatics and Information Management (HIIM) programs. By utilizing the CSET, HIIM programs can identify their strengths and weaknesses and devise strategies to enhance their educational programs. This tool is crucial for healthcare organizations seeking CAHIIM accreditation or maintaining their current accreditation,







**Figure 4:** Bar Charts depicting the distribution of CLOs per course

as it helps highlight areas that may not meet CAHIIM standards, facilitating proactive improvement measures.

The CSET, also known as the Comprehensive Program Review for Teacher Education, consists of six main areas: curriculum, faculty, students, resources, outcomes assessment, and programmatic accreditation. Each of these areas is subdivided into more specific domains, evaluated via a blend of quantitative and qualitative methods. The evaluation is undertaken by program faculty and staff to gauge their program’s performance against pre-established standards and benchmarks. The ’curriculum’ domain scrutinizes the HIIM program’s course content and structure, verifying alignment with the program’s mission, goals, and outcomes. The ’faculty’ domain assesses the qualifications, credentials, and professional development of faculty members teaching in the program, along with their contribution to its development and evaluation. The ’students’ domain evaluates the qualifications and characteristics of the students enrolled, their performance, and outcomes. It encompasses the assessment of admission criteria, procedures, student retention rates, and student achievements through program-specific assessments and licensure exams. The ’resources’ domain assesses the availability and adequacy of resources and infrastructure that support the program, including classrooms, laboratories, technology resources, and other essential equipment.

Lastly, the ’outcomes assessment’ domain focuses on evaluating the HIIM program’s effectiveness in realizing its mission, goals, and outcomes. It includes the assessment of student learning outcomes, programmatic outcomes, and the overall impact of the program on graduates and the profession.

In conclusion, The CAHIIM Self-Evaluation Tool Assessment (CSET) provides healthcare organizations with a comprehensive method to assess their HIIM programs, ensuring compliance with CAHIIM standards. It helps identify improvement areas, implement changes, and prepare for accreditation.

A graph showing the difference between receiver and receiver

Description automatically generated

**Figure 5:** Combined Receiver Operating Curve

A table of text with numbers and a course

Description automatically generated with medium confidence

**Table 2: Macro-average AUC scores and Support for all the courses**

2.5 Predictive Analysis

We predicted the likelihood of students meeting specific learning objectives using logistic regression and random forest models. Initially, logistic regression was used to predict if a student would score above 80. We plotted AUROC curves to gauge the performance of our random forest classifier model. The model strongly predicted outcomes across various courses, with individual AUROC scores ranging from 0.94 to 0.99. This indicates the model’s robust ability to distinguish different performance levels, making it a valuable tool for educational decision-making. We plotted ROC curves to further understand the performance measure of our random forest classifier model. We determined individual AUC scores for each of the courses, which are shown in TableXX.

In summary, our random forest classifier model has yielded satisfactory performance for predicting the outcomes of various courses. The individual AUROC scores for each course range from 0.94 to 0.99, indicating the model’s strong ability to distinguish between different performance levels. These results demonstrate the model’s effectiveness in predicting course performance, offering valuable insights for educational decision-making.

3. Results and Discussion

Each course evaluates the performance of its students based on their grades for each Course Learning Outcome (CLO). It is noteworthy to mention that all students have excelled with scores above 80 percent in their respective CLO assessments, resulting in a highly positive learning environment and academic success for

everyone. The scores in the CLO assessments directly impact the final grades, achieving above 80 percent. The results for the Random Forest Classifier (RFC) in the B530 course indicate an accuracy of 0.869, with a perfect training accuracy of 100.0. Precision, recall, and F1 scores are provided for each class (A, A-, B, B+). The macro average F1-score of 0.87 indicates good overall performance, while the weighted average F1-score of 0.87 considers the class distribution in the dataset, which includes 14 instances of class A, 22 instances of class A-, 19 instances of class B, and 29 instances of class B+. In summary, the model performed well across all classes, with a good overall performance. Based on the reported cross-validation accuracy score of 0.85694, it appears that the model achieved an average accuracy of around 85.7 percent across various cross-validation folds. The B530 course shows that CLO2 has the highest importance score of 0.23, which means it has the biggest impact on the model’s predictions. CLO6 comes in second with a score of 0.21. CLO4, CLO5, CLO7, CLO1, and CLO3 all have decreasing importance scores, indicating that their influence on the model’s predictions is decreasing.

4. Conclusion

The analysis of student grades in Graduate Health Informatics reveals a strong positive relationship between academic performance and the understanding and application of the course’s

learning outcomes. This suggests that the curriculum design and pedagogical strategies employed in the program are effective in promoting student learning and achievement. The findings from this data-driven approach to program evaluation provide evidence for making informed decisions regarding curriculum alignment and educational outcomes, enabling continuous improvement to ensure students acquire the necessary knowledge and skills for success in the field of health informatics.

5. Authors’ contributions

Conceptualization: . • Data curation: . • Formal analysis: . •

Funding acquisition: . •Methodology: . • Project administration:.

• Visualization: . •Writing – original draft: . •Writing – review

and editing: .

6. Conflict of interest

No potential conflict of interest relevant to this article was reported.

7. Funding

None

8. Data availability

9. Acknowledgments

We want to thank Dr. Josette Jones, the previous program director of BHI, and Dr. Mathew Palakal, the previous dean of BHI, for their exceptional leadership. Their guidance and vision have played a significant role in shaping the program’s success and providing valuable insights for our program evaluation. We also appreciate the support of the program assistants, Basil Wilson and Sriha Yalamanchi, whose contributions were invaluable throughout the project. The collective efforts of the BHI program’s faculty and staff, including the program directors and program assistants, were crucial in completing this paper, and we feel honored to have worked with such a dedicated and talented team.

10. Supplementary Materials

There are no supplementary materials.

References

*1. Annette L Valenta, Eta S Berner, Suzanne A Boren, Gloria J Deckard, Christina Eldredge, Douglas B Fridsma, Cynthia Gadd, Yang Gong, Todd Johnson, Josette Jones, et al. Amia board white paper: Amia 2017 core competencies for applied health informatics education at the master’s degree level. Journal of the American Medical Informatics Association, 25(12):1657–1668, 2018.*

*2. Robert E Hoyt and Ann K Yoshihashi. Health informatics: practical guide for healthcare and information technology professionals. Lulu. com, 2014.*

*3. George Siemens and Phil Long. Penetrating the fog: Analytics in learning and education. EDUCAUSE review, 46(5):30, 2011.*

*4. Yazan A Alsariera, Yahia Baashar, Gamal Alkawsi, Abdulsalam Mustafa, Ammar Ahmed Alkahtani, and Nor’ashikin Ali. Assessment and evaluation of different machine learning algorithms for predicting student performance. Computational Intelligence and Neuroscience, 2022, 2022.*

*5. John P Campbell, Peter B DeBlois, and Diana G Oblinger. Academic analytics: A new tool for a new era. EDUCAUSE*

*review, 42(4):40, 2007.*

*6. Juan L Rastrollo-Guerrero, Juan A Gómez-Pulido, and Arturo Durán-Domínguez. Analyzing and predicting students’*

*performance by means of machine learning: A review. Applied sciences, 10(3):1042, 2020.*

*7. Irmgard U Willcockson, Craig W Johnson, William Hersh, and Elmer V Bernstam. Predictors of student success in graduate*

*biomedical informatics training: introductory course and program success. Journal of the American Medical Informatics*

*Association, 16(6):837–846, 2009.*

*8. Tahereh Saheb andMohammad Saheb. Analyzing and visualizing knowledge structures of health informatics from 1974*

*to 2018: a bibliometric and social network analysis. Healthcare informatics research, 25(2):61–72, 2019.*

*9. Melanie A Meyer. Healthcare data scientist qualifications, skills, and job focus: a content analysis of job postings. Journal*

*of the American Medical Informatics Association, 26(5):383–391, 2019.*