Integrating HCI Datasets in Project-Based Machine Learning Courses: A College-Level Review and Case Study

Xiaodong Qu, Matthew Key, Eric Luo and Chuhui Qiu

The George Washington University
Washington, DC 20052
(x.qu, matthewlkey, qluo, chqiu)@gwu.edu

Abstract. This study explores the integration of real-world machine learning (ML) projects using human-computer interfaces (HCI) datasets in college-level courses to enhance both teaching and learning experiences. Employing a comprehensive literature review, course websites analysis, and a detailed case study, the research identifies best practices for incorporating HCI datasets into project-based ML education. Key findings demonstrate increased student engagement, motivation, and skill development through hands-on projects, while instructors benefit from effective tools for teaching complex concepts. The study also addresses challenges such as data complexity and resource allocation, offering recommendations for future improvements. These insights provide a valuable framework for educators aiming to bridge the gap between theoretical knowledge and practical application in ML education.

Keywords: Machine Learning education · Human-Computer Interfaces (HCI) · Real-World Datasets · Project-based learning · Applied Machine Learning

1 Introduction

1.1 Motivation

The rapid advancement of Machine Learning (ML) and Artificial Intelligence (AI) technologies has created a pressing demand for skilled professionals in various industries [27, 72, 40, 3, 59]. In response, higher education institutions have expanded their ML curricula to include diverse courses aimed at imparting AI literacy and problem-solving skills [8, 26, 37, 39]. These courses cater to undergraduates and working professionals seeking to enhance their expertise.

Despite the proliferation of ML courses, significant discrepancies exist in their design and delivery. Project-based ML courses, emphasizing hands-on experience and practical application, show considerable variation in structure and content, creating challenges in achieving consistent educational outcomes.

A critical aspect is the integration of datasets for practical exercises. Human-Computer Interaction (HCI) datasets offer rich opportunities for applying ML techniques. However, effective use requires careful course design and delivery.

This paper addresses key issues by conducting a comprehensive literature review, analyzing related course websites, and performing a detailed case study to identify best practices and provide actionable insights for educators. The primary audience includes computer science professors and students. The focus on learners' engagement and instructors' perspectives aims to bridge the gap between theoretical knowledge and practical application, enhancing ML education for the AI-driven job market.

1.2   Research Questions

The following research questions guide this study and form the basis for its analysis and conclusions:

- **What best practices can be identified from a review of academic papers related to project-based ML course delivery, particularly those incorporating HCI datasets?**
- **What trends in course design, including the integration of math and statistics, and dataset usage emerge from an analysis of in-person ML courses at colleges and universities?**
- **What insights can be gained from a detailed case study of teaching ML courses with specific HCI datasets over the past three years, considering both learner engagement and instructor perspectives?**

By addressing these questions, this paper aims to bridge the gap between theoretical knowledge and practical application in ML education. The findings are intended to guide educators in enhancing the design and delivery of project-based ML courses, ultimately improving student learning outcomes and better preparing them for the demands of the AI-driven job market.

2   Related Works

2.1   Overview

Historically, higher education in technical fields has relied on a blend of lectures, exercises, and practical labs to foster both theoretical understanding and hands-on proficiency with field-relevant methods [1]. In the domain of computer science, it is crucial for graduates to acquire practical skills that align with industrial demands, ensuring they are well-prepared for the workforce [7, 55]. Institutions such as universities of applied sciences emphasize this alignment through curricula designed to bridge academic knowledge and practical application.

Despite these efforts, challenges persist when graduates encounter large, complex real-world projects or when they lack essential teamwork skills [7, 55]. These challenges underscore the growing importance of project-based learning (PBL), which involves industrial topics and aims to better prepare students for professional environments [64, 15, 16]. PBL provides students with opportunities to engage in collaborative, hands-on projects that simulate real-world scenarios, thereby enhancing their problem-solving abilities and teamwork competencies.

2.2   Project-Based Learning in ML Education

In rapidly evolving fields such as deep learning, a subset of Machine Learning (ML), there is a continuous emergence of new trends in algorithms, datasets, and pedagogical approaches [21, 38, 9, 65]. Project-based learning (PBL) has been recognized as an effective pedagogical strategy in these contexts. It facilitates active learning by immersing students in real-world problems and encouraging them to apply theoretical concepts to practical challenges.

Several studies have highlighted the benefits of integrating PBL into ML education. For instance, Huang et al. (2019) emphasize the importance of incorporating contemporary datasets and real-world applications to keep the curriculum relevant and engaging [21]. Miller et al. (2019) discuss how PBL can promote deeper understanding and retention of complex ML concepts by enabling students to work on projects that mirror industrial applications [38]. Brungel et al. (2020) and Wong et al. (2020) further elaborate on how PBL fosters critical thinking and problem-solving skills, which are essential for success in the fast-paced field of ML [9, 65].

2.3   K-12 ML Education

The burgeoning interest in K-12 Machine Learning education has prompted a synthesis of existing research to better understand how ML can be effectively integrated into early education. Several studies have explored resources for ML education at the K-12 level, its integration into existing curricula, and innovative pedagogical strategies [53, 36, 50, 54, 61].

Sanusi et al. (2020) and Marques et al. (2020) investigate various pedagogical approaches that can make ML concepts accessible to younger students, emphasizing the importance of foundational understanding and engagement [53, 36]. Reddi et al. (2021) highlight the need for curriculum development that not only introduces ML concepts but also integrates them seamlessly into subjects already being taught, thereby enriching the overall educational experience [50]. Sanusi (2023) and Van (2023) provide systematic reviews and emerging trends in K-12 ML education, offering insights into the most effective practices and resources for educators [54, 61].

These studies collectively underscore the critical role of early ML education in fostering future generations of tech-savvy individuals who are well-prepared for advanced studies and careers in AI and ML.

3   Methods

This study employed a comprehensive review of academic papers on ML-related courses, supplemented by an extensive survey of relevant course websites. Additionally, a reflective analysis of teaching experiences over the past three years was presented as a case study.

3.1   Keywords

Utilizing the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach, pertinent papers were systematically identified over a two-month period, from August to October 2023. The databases explored included Google Scholar, IEEE Xplore, ACM Digital Library, arXiv, and ERIC. The keyword search comprised: (*'Machine Learning' OR 'Deep Learning' OR 'ML' OR 'DL'*) *AND* (*'project-based' OR 'project-based learning' OR 'PBL'*) *AND* (*'Survey' OR 'Review' OR 'Case Study'*).

Additionally, the search strings used were: *"project-based learning in machine learning", "applied machine learning course design", "teaching strategies for machine learning", "HCI datasets in project-based learning", "challenges in project-based machine learning instruction", "student engagement in project-based machine learning courses"*.

This strategy aimed to pinpoint papers aligning with the research questions. Table 1 and Figure 1 visualize the search trajectory, showcasing the number of papers identified and excluded based on set criteria. To cater to the target audience's time constraints, a concise list of papers encapsulating the prevailing trends in the domain was curated.

3.2   Selection Criteria

To ensure the relevance and quality of the review's content, the following criteria were applied:

- **Project-based:** Papers and courses must emphasize project-based machine learning, detailing their design and execution. This focus ensures that the content is practically relevant and applicable to real-world ML problems.

Table 1. Progression of Paper Search Steps: S1 represents initial search results, S2 indicates potentially relevant findings, S3 highlights confirmed relevant results, and S4 enumerates those results after removing duplicates.

Paper Source (Steps)	S1	S2	S3	S4
Google Scholar	312	230	112	112
ACM DIgital Library	120	90	45	40
IEEE Xplore	66	51	37	35
ERIC	31	21	15	12
arXiv	15	7	5	5
subtotal	544	399	214	204

- **Publication Time Frame:** Papers published from 2017 onwards and courses updated after 2020 were included. This time frame ensures that the review encompasses the most recent advancements and trends in ML education.
- **Machine Learning Focus:** Preference was given to content primarily addressing project-based Machine Learning or Deep Learning. This focus aligns with the goal of enhancing practical ML education for professionals.
- **Target Audience:** Papers and courses should cater to computer science (CS) professors, students, or non-CS majors enrolled in such courses. This criterion ensures that the content is relevant to educators and learners who
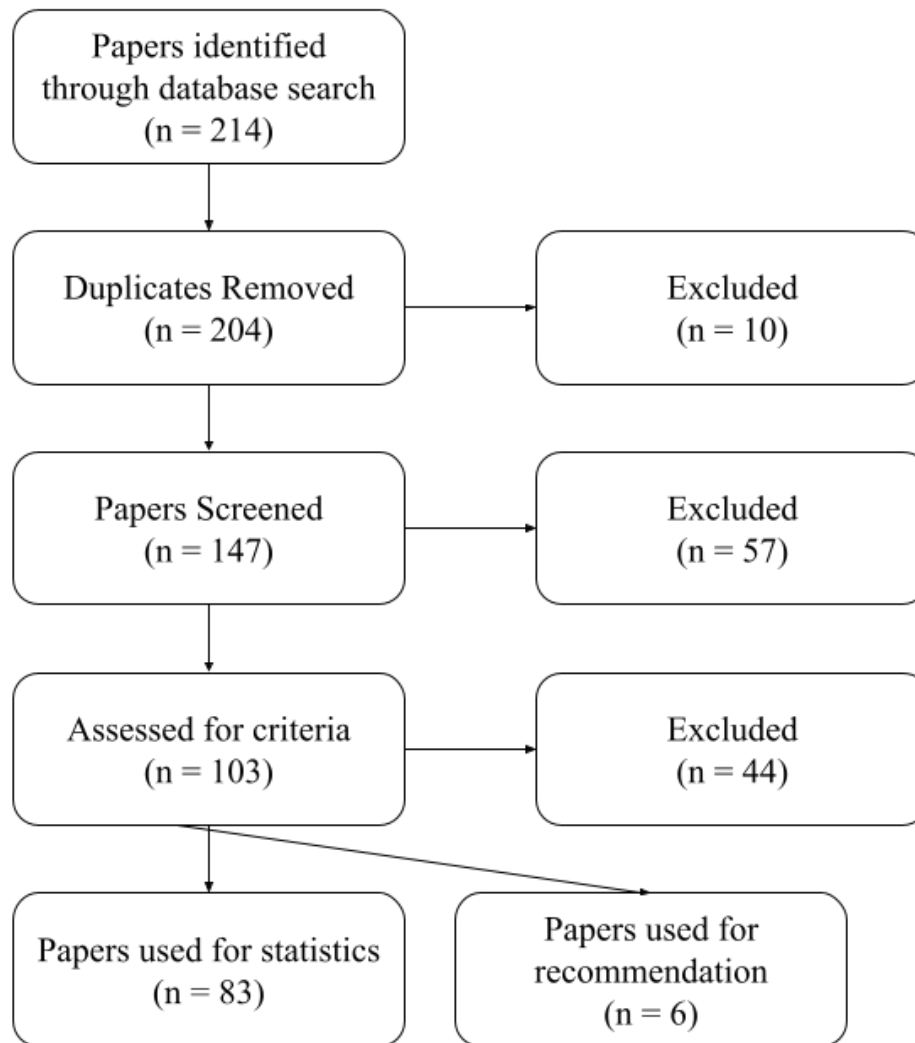

Fig. 1. Selection process for the papers

are directly involved in ML education, as well as those from diverse academic backgrounds seeking to gain practical ML skills.

3.3   Case Study

Over the past three years (2022 to 2024), four project-based machine learning courses were delivered by an assistant professor in computer science at two different institutions. In 2021 and 2022, undergraduate students at Swarthmore College, PA, participated in these courses. In 2023, graduate students at George Washington University, DC, were enrolled in similar project-based machine learning courses. These courses aimed to evaluate the effectiveness of project-based learning in enhancing students' practical machine learning skills and their ability to apply theoretical knowledge to real-world problems. Detailed course outlines, project descriptions, and instructional methodologies were documented and analyzed. These materials are available on the faculty page of the instructor [1].

A 2021 EEG dataset was utilized in these machine learning courses. Electroencephalography (EEG) is extensively used in research fields such as neural engineering, neuroscience, biomedical engineering, and brain-like computing, with particular emphasis on brain-computer interfaces (BCIs). The analysis of EEG signals is crucial for the development of BCIs, providing deep insights into the complex neural activities of the human brain. Over the past decade, a variety of machine learning and deep learning algorithms have been applied to EEG data, leading to significant advancements in numerous applications. These applications include emotion recognition, motor imagery, mental workload assessment, seizure detection, Alzheimer's disease classification, sleep stage scoring, and many more [23, 13, 51, 4, 42, 18, 20, 66, 24, 30, 25, 39, 17, 73, 46, 47, 45, 48, 43, 52, 44, 49, 14, 63, 28, 2].

This case study used a mixed-methods approach to gather comprehensive feedback from students. Thirty-five students were interviewed after completing the courses, providing qualitative data on their learning experiences, challenges, and perceived benefits. Additionally, anonymous course feedback forms were collected and analyzed, offering quantitative insights into student satisfaction and instructional effectiveness.

Key aspects highlighted by student feedback include:

- **Engagement and Motivation:** High levels of engagement and motivation due to the hands-on nature of the projects, showing practical applications of ML concepts.
- **Skill Development:** Significant improvements in technical skills, especially in data preprocessing, model development, and evaluation, with real-world datasets enhancing problem-solving abilities.
- **Collaboration and Teamwork:** Encouraged collaboration, fostering a collaborative learning environment, and developing essential teamwork skills.

---

[1] https://faculty.cs.gwu.edu/xiaodongqu/

  – **Challenges and Areas for Improvement:** Challenges with project com-
    plexity and the steep learning curve of advanced ML techniques. Suggestions
    included more support during initial project stages.

  The analysis of course feedback revealed that the project-based learning ap-
proach was well-received, with students valuing the practical application of the-
oretical knowledge. Insights from this case study inform recommendations for
enhancing project-based ML courses, making them more effective and accessible
for students at various academic levels.
  This case study underscores the importance of integrating practical, hands-on
projects in ML education. It provides valuable insights for educators seeking to
design courses that impart theoretical knowledge and practical skills for success
in the evolving field of ML.

4   Results

4.1   Literature Review

The literature review focused on identifying best practices in Machine Learning
(ML) education, particularly in in-person courses offered by colleges and univer-
sities. The findings were categorized into several key topics, highlighting common
best practices across various studies. Table 2 summarizes these findings.

Table 2. Paper Results Key: U denotes undergrad-only studies, G for graduate-only,
and UG for both levels. R signifies review papers, C indicates case studies, and Best P
stands for best practices.

Paper	Level	Type	Best P
[1,2,4,7,15,19,20]	U	R	1,2,3
[5,9,12,14,17,25]	U	C	2,3,4
[3, 6, 11, 18, 21]	UG	R	3,5
[22, 23,26, 28,31]	UG	C	2,6
[16, 24, 29, 33]	G	R	1,2,3,5
[8, 10, 13, 16, 34]	G	C	1, 3, 4

1. Machine Learning:

- **Comprehensive Curriculum:** Effective ML courses offer a well-rounded curriculum covering fundamental concepts, advanced techniques, and practical applications. This ensures learners gain a broad understanding and can apply knowledge to real-world problems, benefiting instructors with a structured roadmap for teaching complex topics.
- **Integration of Math and Statistics:** A solid foundation in mathematics and statistics is crucial for understanding ML algorithms. Effective courses include relevant topics like linear algebra, calculus, probability, and statistical inference, equipping students with essential analytical skills.
- **Hands-On Projects:** Incorporating hands-on projects enhances ML education. Real-world datasets and practical problems significantly improve learning outcomes by providing practical experience and reinforcing theoretical knowledge, while also helping instructors assess student understanding.

2. Project-Based Teaching and Learning:

- **Engagement and Motivation:** Project-based learning (PBL) increases student engagement and motivation. Working on relevant projects keeps learners invested, leading to more dynamic and interactive classrooms.
- **Collaborative Learning:** Successful PBL courses encourage student collaboration through group projects and peer feedback, promoting deeper understanding and teamwork skills, while easing the instructional burden.
- **Practical Application:** PBL emphasizes applying theoretical knowledge to real-world problems, helping students develop critical thinking and problem-solving skills for their future careers.

3. Exam-Based Teaching:

- **Knowledge Assessment:** Exam-based teaching assesses students' theoretical understanding through written exams and quizzes, effectively measuring knowledge retention and application.
- **Standardization:** Exams provide a consistent evaluation method across students, useful for large classes where individual project assessment is impractical.
- **Individual Focus:** Exams emphasize individual performance, fostering independent study habits, though they may not capture collaborative and practical skills.
- **Certification Preparation:** Exams prepare students for professional certification tests, which often follow similar formats, aiding in obtaining required credentials.

4. Formal Course Structure:

- **Structured Schedules:** In-person ML courses often benefit from structured schedules that ensure consistent progress and regular engagement. These schedules help students balance their academic responsibilities and maintain steady progress. Instructors can design courses with regular milestones and assessments to keep students on track.

– **Support and Resources:** Providing ample support and resources is critical for in-person course settings. This includes access to office hours, tutoring, instructional videos, and supplemental materials that help learners overcome obstacles. For instructors, it means developing comprehensive resources that can guide students through their learning journey and provide additional support as needed.

5. Students' Feedback:

– **Positive Impact of Structured Learning:** Students frequently highlight the benefits of the structured schedule offered by in-person courses. This structure allows them to balance their studies with other academic and personal commitments effectively. Instructors receive positive feedback on course accessibility and organization, which can enhance course ratings and attract more students.

– **Need for Interactive Elements:** Feedback often suggests that incorporating interactive elements, such as quizzes, labs, and real-time feedback, can enhance the learning experience in in-person courses. Instructors can leverage this feedback to design more engaging and interactive course content, improving student satisfaction and outcomes.

6. Professors' Feedback:

– **Importance of Course Design:** Professors emphasize the significance of well-structured course design in in-person learning environments. Clear learning objectives, organized content, and regular assessments are essential for maintaining student engagement and ensuring successful learning outcomes. Effective course design helps instructors manage course delivery more efficiently and ensures that learning goals are met.

– **Challenges in Providing Support:** While in-person courses offer structure, professors note the challenges in providing timely support and feedback to students. Implementing systems and leveraging technology can help address these challenges. Instructors can use tools such as automated grading systems, learning management systems (LMS), and discussion forums to provide timely support.

By addressing both the learners' and instructors' perspectives and comparing project-based teaching with exam-based teaching, this section provides a comprehensive view of the best practices in ML education within in-person college and university settings. It highlights the benefits and challenges of both teaching methods, offering insights that can inform the design and implementation of effective ML courses.

4.2   Course Websites Analysis

Table 3 showcases the findings from from the analysis of existing course websites. This analysis aimed to identify common elements and best practices in Machine

Learning (ML) courses offered in-person at colleges and universities. Several key topics were explored to understand how these courses are structured and what resources they provide to both learners and instructors.

Table 3. Analysis of Course Websites: U represents undergrad-only courses, G for graduate-only, and UG for both. Best P signifies courses emphasizing best practices.

Course	Level	Institution	Best P
[41, 42]	U	Williams	7,8
[43]	U	Amherst	8
[44,45]	U	Swarthmore	7,9,11,12
[46, 47,48]	U	Pomona	7,8,9,10
[49, 50]	UG	Harvard	7,8,9,11
[50, 51]	UG	Upenn	7, 10, 11,12
[52, 53, 54]	UG	Stanford	7,8,9,10,11,12
[55, 56]	UG	MIT	8,9,10,11
[57,58,59]	UG	CMU	8, 9,11
[60,61,62]	UG	UC B	7,8,9,10, 11

7. Course Structure:

- **Modular Structure:** Many analyzed courses feature a modular structure, allowing for systematic progression through the material. This structure is beneficial for organizing the curriculum in a way that builds upon foundational concepts before advancing to more complex topics. For instructors, it provides a clear framework for delivering content effectively.
- **Integration of Math and Statistics:** Effective ML courses integrate essential mathematics and statistics topics, such as linear algebra, calculus, probability, and statistical inference. This integration is crucial for students to understand the theoretical underpinnings of ML algorithms. Instructors can use this foundation to explain complex concepts and ensure students are well-prepared for practical applications.

8. Progress Tracking:

– **Tools for Tracking Progress:** Effective courses often include tools for tracking progress, such as dashboards that display completed modules and upcoming tasks. These tools help learners stay organized and motivated. For instructors, progress tracking tools provide insights into student performance and areas that may require additional attention.

9. Project-Based Teaching and Learning:

– **Hands-On Projects:** A significant number of courses incorporate project-based learning, where students work on real-world projects to apply the concepts they have learned. These projects often involve datasets from industry or research, providing practical experience. For instructors, hands-on projects offer a practical way to assess students' application of theoretical knowledge.
– **Peer Collaboration:** Some courses facilitate peer collaboration through discussion forums or group projects, allowing learners to share insights and provide mutual support. This collaboration helps build teamwork skills, which are valuable in professional settings. Instructors benefit from the collaborative learning environment as it can enhance student engagement and reduce the instructional burden.

10. Sample Code:

– **Code Repositories:** Many courses provide access to code repositories, such as GitHub, where learners can find sample code and scripts used in the course. This is particularly useful for understanding practical implementation details. Instructors can use these repositories to demonstrate coding practices and provide students with resources for independent study.
– **Code Walkthroughs:** Courses that include detailed code walkthroughs, either in written form or through video demonstrations, help learners understand the step-by-step process of developing ML models. For instructors, code walkthroughs are an effective teaching tool to illustrate coding techniques and problem-solving strategies.

11. Lecture Slides:

– **Comprehensive Lecture Slides:** High-quality courses offer comprehensive lecture slides that summarize key concepts and provide visual aids to enhance understanding. These slides are often available for download, allowing learners to review them at their own pace. Instructors can use these slides to structure their lectures and provide students with a consistent reference material.

– **Supplemental Materials:** In addition to slides, some courses provide supplemental materials such as cheat sheets, reference guides, and additional readings to deepen learners' understanding. These materials support instructors in offering a richer educational experience and cater to diverse learning needs.

12. Course Videos:

– **Engaging Video Lectures:** Video lectures are a staple of in-person ML courses, often used to supplement classroom teaching. The best courses feature engaging, well-produced videos that clearly explain complex concepts. These videos often include demonstrations, animations, and real-world examples to illustrate key points. Instructors can leverage these videos to reinforce classroom teaching and provide students with additional learning resources.
– **Interactive Elements:** Some courses incorporate interactive elements within videos, such as embedded quizzes or coding challenges, to reinforce learning and keep learners engaged. These elements provide immediate feedback to students and help instructors gauge understanding in real-time.

By addressing both the learners' and instructors' perspectives, this section provides a comprehensive view of the best practices in ML education within in-person college and university settings. It highlights the importance of integrating foundational math and statistics, hands-on projects, and collaborative learning, offering insights that can inform the design and implementation of effective ML courses.

4.3   Case Study

The case study examines the implementation of project-based learning in Machine Learning (ML) courses from 2022 to 2024 at Swarthmore College and George Washington University. These courses provided students with hands-on experience applying ML techniques to real-world problems using a brain-computer interfaces (BCI) dataset.

Course Context and Structure:

Courses included lectures, hands-on projects, and collaborative activities to bridge theoretical knowledge and practical application, ensuring students could apply ML concepts to complex problems.

– **Instructors' Perspective:** Designed by an assistant professor in computer science to create a challenging yet supportive environment, the courses used a BCI dataset for advanced ML applications in Human-Computer Interaction (HCI).
– **Learners' Perspective:** Undergraduate and graduate students engaged deeply with material, developing practical skills highly valued in the industry.

Implementation of the BCI Dataset:
The BCI dataset consisted of EEG recordings from subjects engaged in tasks. This dataset's complexity and relevance to HCI research made it an ideal choice.

– **Project Design:** Students developed ML models to classify mental states based on EEG data, involving preprocessing, feature extraction, and model training and evaluation.
– **Instructors' Role:** Instructors guided technical aspects, facilitated group discussions, and provided feedback.
– **Learners' Experience:** Students reported high engagement and motivation, finding the project a valuable learning experience that challenged them to think critically and innovate.

Outcomes and Feedback:
Course outcomes were evaluated through student feedback, project assessments, and instructor observations.

– **Engagement and Motivation:** Students showed increased engagement and deeper interest in HCI and ML applications, citing the challenging dataset as a key factor.
– **Skill Development:** Both instructors and students noted significant improvements in technical skills, particularly in data preprocessing, feature extraction, and model development.
– **Collaboration and Teamwork:** Courses facilitated collaborative learning, enhancing understanding of ML concepts and developing teamwork and communication skills.
– **Instructor Insights:** Instructors found the project-based approach effective in promoting deep learning and practical skill acquisition, with the BCI dataset enriching the experience.

Challenges and Recommendations:
Several challenges were encountered during the courses.

– **Data Complexity:** The BCI dataset's complexity challenged students, especially those with limited ML experience. Instructors provided additional support and resources, such as tutorials on EEG data processing.
– **Time Management:** Balancing project workload with other requirements was challenging. Instructors recommended clearer guidelines and structured timelines.
– **Resource Allocation:** The resource-intensive nature of project-based learning required significant instructor time. Future courses could benefit from more teaching assistants or automated tools.

In conclusion, the case study highlights the effectiveness of project-based learning in ML education, particularly using complex datasets like BCI. Both learners and instructors benefited from the hands-on, collaborative approach, facilitating deep learning and practical skill development. These insights provide valuable guidance for educators implementing similar approaches.

5   Discussion

5.1   Real-World Machine Learning Projects with HCI Datasets

A significant finding from this study is the crucial role of real-world machine learning projects with HCI datasets in enhancing both teaching and learning experiences. These projects provide learners with hands-on experience and practical application of the concepts they have learned. Instructors also benefit from these projects as they provide a rich context for teaching complex ML topics and assessing student understanding. Here is a subset of the experiments that learners and instructors have explored and practiced so far [5, 6, 29, 12, 11, 33, 31, 32, 70, 71, 62, 60, 45, 47, 19, 67, 17, 58, 39, 69, 67, 68, 66, 34, 35, 22].

The projects discussed involved various HCI datasets, including brain-computer interfaces (BCI), eye-tracking, and gesture recognition data. These datasets were chosen for their relevance to cutting-edge HCI research and their ability to challenge students to apply ML techniques to real-world problems.

Learners' Perspective Working with HCI datasets offers several benefits:

- **Engagement and Motivation:** Real-world projects significantly increase engagement and motivation, showing the relevance of studies.
- **Skill Development:** Students develop critical skills like data preprocessing, feature extraction, and model development.
- **Critical Thinking and Problem-Solving:** Projects require critical thinking and innovative solutions, building problem-solving skills.
- **Collaboration and Teamwork:** Collaborative projects foster teamwork and communication skills.

A student noted that applied ML courses felt like multiple courses in one, demanding self-guided learning in problem selection, literature review, proposing novel methods, and technical implementation. This comprehensive approach significantly contributed to their skill development.

Instructors' Perspective Real-world projects with HCI datasets offer several advantages:

- **Effective Teaching Tool:** HCI datasets make abstract ML concepts concrete and understandable.
- **Assessment of Student Understanding:** Projects provide a practical means of assessing ML concept application.
- **Enhanced Engagement and Interaction:** Projects lead to dynamic, interactive classrooms with high participation.
- **Resource for Research and Development:** Student projects can contribute to ongoing research, exploring new ideas.

Examples of HCI Projects Examples of projects include:

- BCI: Developing models to classify mental states based on EEG signals.
- Eye-Tracking: Predicting user intent and analyzing gaze patterns to improve UI design.
- Gesture Recognition: Creating models to recognize and interpret human gestures for device control.

These projects provided insights into ML applications in HCI, offering a deeper understanding of the technologies' potential and challenges.

Challenges and Recommendations Challenges encountered:

- **Data Complexity:** HCI datasets can be overwhelming. Instructors should provide tutorials and support materials.
- **Time Management:** Balancing project work with other requirements is challenging. Clear guidelines and structured timelines are essential.
- **Resource Allocation:** Project-based learning requires significant instructor time. Additional teaching assistants or automated tools can help manage the workload.

5.2 Future Work

Future work can build on this study's findings:

- **Expanding Dataset Variety:** Incorporate more HCI datasets, including speech recognition, NLP, and VR interactions, to provide broader exposure.
- **Enhancing Support for Novice Researchers:** Develop structured support for students with little research experience, such as mentorship programs, research guides, and workshops.
- **Leveraging Technology for Support:** Utilize AI-driven tutoring and automated feedback tools to provide timely, personalized support, especially in large classes.
- **Longitudinal Studies on Learning Outcomes:** Conduct longitudinal studies to track the long-term impact of project-based learning on careers and research outputs.
- **Integrating Interdisciplinary Approaches:** Encourage interdisciplinary projects combining ML with fields like psychology, neuroscience, and engineering for innovative solutions.
- **Comparative Studies of ML Techniques:** Explore various machine learning techniques on different datasets for comparative studies [5, 6, 22, 33, 10, 34, 35, 19, 57, 56, 41, 70] to provide valuable insights and further enhance the field.

In conclusion, integrating real-world ML projects with HCI datasets enhances teaching and learning. These projects provide practical skills and critical thinking abilities, while offering instructors effective tools for teaching complex concepts and assessing understanding. The insights from this study guide educators in implementing similar approaches.

6   Conclusion

This study highlights the effectiveness of integrating real-world machine learning (ML) projects with human-computer interfaces (HCI) datasets in enhancing both teaching and learning experiences. By providing hands-on experience and practical application opportunities, these projects significantly increase student engagement, motivation, and skill development. From the instructors' perspective, they offer valuable tools for teaching complex concepts and assessing student understanding. Despite the challenges of data complexity and resource allocation, the benefits of project-based learning are evident. Future work should focus on expanding dataset variety, enhancing support for novice researchers, leveraging technology for support, and conducting longitudinal studies on learning outcomes. The insights gained from this study provide a robust framework for educators seeking to implement similar approaches, ultimately bridging the gap between theoretical knowledge and practical application in ML education.

References

1. Abood, H.G.: E-learning applications in engineering and the project-based learning vs problem-based learning styles: a critical & comparative study. Engineering and Technology Journal **37**(4), 391–396 (2019)
2. Aggarwal, S., Chugh, N.: Review of machine learning techniques for eeg based brain computer interface. Archives of Computational Methods in Engineering **29**(5), 3001–3020 (2022)
3. Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., Martinez-Maldonado, R.: Human-centred learning analytics and ai in education: a systematic literature review. Computers and Education: Artificial Intelligence p. 100215 (2024)
4. Altaheri, H., Muhammad, G., Alsulaiman, M., Amin, S.U., Altuwaijri, G.A., Abdul, W., Bencherif, M.A., Faisal, M.: Deep learning techniques for classification of electroencephalogram (eeg) motor imagery (mi) signals: A review. Neural Computing and Applications **35**(20), 14681–14722 (2023)
5. An, S., Bhat, G., Gumussoy, S., Ogras, U.: Transfer learning for human activity recognition using representational analysis of neural networks. ACM Transactions on Computing for Healthcare **4**(1), 1–21 (2023)
6. An, S., Tuncel, Y., Basaklar, T., Ogras, U.Y.: A survey of embedded machine learning for smart and sustainable healthcare applications. In: Embedded Machine Learning for Cyber-Physical, IoT, and Edge Computing: Use Cases and Emerging Challenges, pp. 127–150. Springer (2023)
7. Beckman, K., Coulter, N., Khajenoori, S., Mead, N.R.: Collaborations: closing the industry-academia gap. IEEE software **14**(6), 49–57 (1997)
8. Bennett, B.T.: Teaching artificial intelligence in a multidisciplinary computing environment. Journal of Computing Sciences in Colleges **33**(2), 222–228 (2017)
9. Brüngel, R., Rückert, J., Friedrich, C.M.: Project-based learning in a machine learning course with differentiated industrial projects for various computer science master programs. In: 2020 IEEE 32nd Conference on Software Engineering Education and Training (CSEE&T). pp. 1–5. IEEE (2020)
10. Chen, J., Hu, Y., Wang, Y., Lu, Y., Cao, X., Lin, M., Xu, H., Wu, J., Xiao, C., Sun, J., et al.: Trialbench: Multi-modal artificial intelligence-ready clinical trial datasets. arXiv preprint arXiv:2407.00631 (2024)

11. Chen, P., Ding, H., Araki, J., Huang, R.: Explicitly capturing relations between entity mentions via graph neural networks for domain-specific named entity recognition. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). pp. 735–742 (2021)

12. Chen, P., Sarkar, S., Lausen, L., Srinivasan, B., Zha, S., Huang, R., Karypis, G.: Hytrel: Hypergraph-enhanced tabular data representation learning. Advances in Neural Information Processing Systems **36** (2024)

13. Craik, A., He, Y., Contreras-Vidal, J.L.: Deep learning for electroencephalogram (eeg) classification tasks: a review. Journal of neural engineering **16**(3), 031001 (2019)

14. Dadebayev, D., Goh, W.W., Tan, E.X.: Eeg-based emotion recognition: Review of commercial eeg devices and machine learning techniques. Journal of King Saud University-Computer and Information Sciences **34**(7), 4385–4401 (2022)

15. Daun, M., Salmon, A., Tenbergen, B., Weyer, T., Pohl, K.: Industrial case studies in graduate requirements engineering courses: The impact on student motivation. In: 2014 IEEE 27th Conference on Software Engineering Education and Training (CSEE&T). pp. 3–12. IEEE (2014)

16. Daun, M., Salmon, A., Weyer, T., Pohl, K., Tenbergen, B.: Project-based learning with examples from industry in university courses: an experience report from an undergraduate requirements engineering course. In: 2016 IEEE 29th International Conference on Software Engineering Education and Training (CSEET). pp. 184–193. IEEE (2016)

17. Dou, G., Zhou, Z., Qu, X.: Time majority voting, a pc-based eeg classifier for non-expert users. In: International Conference on Human-Computer Interaction. pp. 415–428. Springer (2022)

18. Gao, Z., Dang, W., Wang, X., Hong, X., Hou, L., Ma, K., Perc, M.: Complex networks and deep learning for eeg signal analysis. Cognitive Neurodynamics **15**(3), 369–388 (2021)

19. Gui, S., Song, S., Qin, R., Tang, Y.: Remote sensing object detection in the deep learning era—a review. Remote Sensing **16**(2), 327 (2024)

20. Hossain, K.M., Islam, M.A., Hossain, S., Nijholt, A., Ahad, M.A.R.: Status of deep learning for eeg-based brain–computer interface applications. Frontiers in computational neuroscience **16**, 1006763 (2023)

21. Huang, L.: Integrating machine learning to undergraduate engineering curricula through project-based learning. In: 2019 IEEE Frontiers in Education Conference (FIE). pp. 1–4. IEEE (2019)

22. Jiang, C., Hui, B., Liu, B., Yan, D.: Successfully applying lottery ticket hypothesis to diffusion model. arXiv preprint arXiv:2310.18823 (2023)

23. Kastrati, A., Płomecka, M.B., Pascual, D., Wolf, L., Gillioz, V., Wattenhofer, R., Langer, N.: Eegeyenet: a simultaneous electroencephalography and eye-tracking dataset and benchmark for eye movement prediction. arXiv preprint arXiv:2111.05100 (2021)

24. Key, M.L., Mehtiyev, T., Qu, X.: Advancing eeg-based gaze prediction using depth-wise separable convolution and enhanced pre-processing. In: International Conference on Human-Computer Interaction. pp. 3–17. Springer (2024)

25. Koome Murungi, N., Pham, M.V., Dai, X., Qu, X.: Trends in machine learning and electroencephalogram (eeg): A review for undergraduate researchers. arXiv e-prints pp. arXiv–2307 (2023)

26. Kwan, P.: A college freshman's guide to machine learning: short and sweet way to introduce machine learning to college freshman. Journal of Computing Sciences in Colleges **30**(1), 36–37 (2014)
27. Lao, N.: Reorienting machine learning education towards tinkerers and ML-engaged citizens. Ph.D. thesis, Massachusetts Institute of Technology Cambridge, MA, USA (2020)
28. Li, G., Lee, C.H., Jung, J.J., Youn, Y.C., Camacho, D.: Deep learning for eeg data analytics: A survey. Concurrency and Computation: Practice and Experience **32**(18), e5199 (2020)
29. Li, H., Chen, C., Shi, T., Qiu, Y., An, S., Chen, G., Han, X.: Spherehead: Stable 3d full-head synthesis with spherical tri-plane representation. arXiv preprint arXiv:2404.05680 (2024)
30. Li, W., Zhou, N., Qu, X.: Enhancing eye-tracking performance through multi-task learning transformer. In: International Conference on Human-Computer Interaction. pp. 31–46. Springer (2024)
31. Lu, Y., Chen, T., Hao, N., Van Rechem, C., Chen, J., Fu, T.: Uncertainty quantification and interpretability for clinical trial approval prediction. Health Data Science **4**, 0126 (2024)
32. Lu, Y., Sato, K., Wang, J.: Deep learning based multi-label image classification of protest activities. arXiv preprint arXiv:2301.04212 (2023)
33. Lu, Y., Shen, M., Wang, H., Wang, X., van Rechem, C., Wei, W.: Machine learning for synthetic data generation: a review. arXiv preprint arXiv:2302.04062 (2023)
34. Ma, X.: Traffic performance evaluation using statistical and machine learning methods. Ph.D. thesis, The University of Arizona (2022)
35. Ma, X., Karimpour, A., Wu, Y.J.: Data-driven transfer learning framework for estimating on-ramp and off-ramp traffic flows. Journal of Intelligent Transportation Systems pp. 1–14 (2024)
36. Marques, L.S., Gresse von Wangenheim, C., Hauck, J.C.: Teaching machine learning in school: A systematic mapping of the state of the art. Informatics in Education **19**(2), 283–321 (2020)
37. Martins, R.M., Gresse Von Wangenheim, C.: Findings on teaching machine learning in high school: A ten-year systematic literature review. Informatics in Education **22**(3), 421–440 (2023)
38. Miller, E.C., Krajcik, J.S.: Promoting deep learning through project-based learning: A design problem. Disciplinary and Interdisciplinary Science Education Research **1**(1), 1–10 (2019)
39. Murungi, N.K., Pham, M.V., Dai, X.C., Qu, X.: Empowering computer science students in electroencephalography (eeg) analysis: A review of machine learning algorithms for eeg datasets. the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) (2023)
40. Ng, D.T.K., Lee, M., Tan, R.J.Y., Hu, X., Downie, J.S., Chu, S.K.W.: A review of ai teaching and learning from 2000 to 2020. Education and Information Technologies **28**(7), 8445–8501 (2023)
41. Qiu, Y., Zhao, Z., Yao, H., Chen, D., Wang, Z.: Modal-aware visual prompting for incomplete multi-modal brain tumor segmentation. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 3228–3239 (2023)
42. Qu, X.: Time Continuity Voting for Electroencephalography (EEG) Classification. Ph.D. thesis, Brandeis University (2022)
43. Qu, X., Hall, M., Sun, Y., Sekuler, R., Hickey, T.J.: A personalized reading coach using wearable eeg sensors. CSEDU (2019)

44. Qu, X., Hickey, T.J.: Eeg4home: A human-in-the-loop machine learning model for eeg-based bci. In: Augmented Cognition: 16th International Conference, AC 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings. pp. 162–172. Springer (2022)

45. Qu, X., Liu, P., Li, Z., Hickey, T.: Multi-class time continuity voting for eeg classification. In: Brain Function Assessment in Learning: Second International Conference, BFAL 2020, Heraklion, Crete, Greece, October 9–11, 2020, Proceedings 2. pp. 24–33. Springer (2020)

46. Qu, X., Liukasemsarn, S., Tu, J., Higgins, A., Hickey, T.J., Hall, M.H.: Identifying clinically and functionally distinct groups among healthy controls and first episode psychosis patients by clustering on eeg patterns. Frontiers in psychiatry **11**, 541659 (2020)

47. Qu, X., Mei, Q., Liu, P., Hickey, T.: Using eeg to distinguish between writing and typing for the same cognitive task. In: Brain Function Assessment in Learning: Second International Conference, BFAL 2020, Heraklion, Crete, Greece, October 9–11, 2020, Proceedings 2. pp. 66–74. Springer (2020)

48. Qu, X., Sun, Y., Sekuler, R., Hickey, T.: Eeg markers of stem learning. In: 2018 IEEE Frontiers in Education Conference (FIE). pp. 1–9. IEEE (2018)

49. Rasheed, K., Qayyum, A., Qadir, J., Sivathamboo, S., Kwan, P., Kuhlmann, L., O'Brien, T., Razi, A.: Machine learning for predicting epileptic seizures using eeg signals: A review. IEEE reviews in biomedical engineering **14**, 139–155 (2020)

50. Reddi, V.J., Plancher, B., Kennedy, S., Moroney, L., Warden, P., Agarwal, A., Banbury, C., Banzi, M., Bennett, M., Brown, B., et al.: Widening access to applied machine learning with tinyml. arXiv preprint arXiv:2106.04008 (2021)

51. Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T.H., Faubert, J.: Deep learning-based electroencephalography analysis: a systematic review. Journal of neural engineering **16**(5), 051001 (2019)

52. Saeidi, M., Karwowski, W., Farahani, F.V., Fiok, K., Taiar, R., Hancock, P.A., Al-Juaid, A.: Neural decoding of eeg signals with machine learning: a systematic review. Brain Sciences **11**(11), 1525 (2021)

53. Sanusi, I.T., Oyelere, S.S.: Pedagogies of machine learning in k-12 context. In: 2020 IEEE frontiers in education conference (FIE). pp. 1–8. IEEE (2020)

54. Sanusi, I.T., Oyelere, S.S., Vartiainen, H., Suhonen, J., Tukiainen, M.: A systematic review of teaching and learning machine learning in k-12 education. Education and Information Technologies **28**(5), 5967–5997 (2023)

55. Shaw, M., Herbsleb, J., Ozkaya, I.: Deciding what to design: Closing a gap in software engineering education. In: Proceedings of the 27th international Conference on Software Engineering. pp. 607–608 (2005)

56. Tan, J., Shen, X., Zhang, X., Wang, Y.: Multivariate encoding analysis of medial prefrontal cortex cortical activity during task learning. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 6699–6702. IEEE (2021)

57. Tan, J., Zhang, X., Wu, S., Song, Z., Chen, S., Huang, Y., Wang, Y.: Audio-induced medial prefrontal cortical dynamics enhances coadaptive learning in brain–machine interfaces. Journal of Neural Engineering **20**(5), 056035 (2023)

58. Tan, J., Zhang, X., Wu, S., Wang, Y.: State-space model based inverse reinforcement learning for reward function estimation in brain-machine interfaces. In: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 1–4. IEEE (2023)

59. Tan, M., Lee, H., Wang, D., Subramonyam, H.: Is a seat at the table enough? engaging teachers and students in dataset specification for ml in education. Proceedings of the ACM on Human-Computer Interaction **8**(CSCW1), 1–32 (2024)
60. Tang, Y., Song, S., Gui, S., Chao, W., Cheng, C., Qin, R.: Active and low-cost hyperspectral imaging for the spectral analysis of a low-light environment. Sensors **23**(3), 1437 (2023)
61. Van Mechelen, M., Smith, R.C., Schaper, M.M., Tamashiro, M., Bilstrup, K.E., Lunding, M., Graves Petersen, M., Sejer Iversen, O.: Emerging technologies in k–12 education: A future hci research agenda. ACM Transactions on Computer-Human Interaction **30**(3), 1–40 (2023)
62. Wang, J., Chang, R., Zhao, Z., Pahwa, R.S.: Robust detection, segmentation, and metrology of high bandwidth memory 3d scans using an improved semi-supervised deep learning approach. Sensors **23**(12), 5470 (2023)
63. Wang, R., Qu, X.: Eeg daydreaming, a machine learning approach to detect daydreaming activities. In: International Conference on Human-Computer Interaction. pp. 202–212. Springer (2022)
64. Winzker, M.: Semester structure with time slots for self-learning and project-based learning. In: Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON). pp. 1–8. IEEE (2012)
65. Wong, K., Tomov, S., Dongarra, J.: Project-based research and training in high performance data sciences, data analytics, and machine learning. The Journal of Computational Science Education **11**(1) (2020)
66. Yi, L., Qu, X.: Attention-based cnn capturing eeg recording's average voltage and local change. In: Artificial Intelligence in HCI: 3rd International Conference, AI-HCI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings. pp. 448–459. Springer (2022)
67. Yunoki, I., Berreby, G., D'Andrea, N., Lu, Y., Qu, X.: Exploring ai music generation: A review of deep learning algorithms and datasets for undergraduate researchers. In: International Conference on Human-Computer Interaction. pp. 102–116. Springer (2023)
68. Zhang, Z., Tian, R., Ding, Z.: Trep: Transformer-based evidential prediction for pedestrian intention with uncertainty. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 3534–3542 (2023)
69. Zhang, Z., Tian, R., Sherony, R., Domeyer, J., Ding, Z.: Attention-based interrelation modeling for explainable automated driving. IEEE Transactions on Intelligent Vehicles **8**(2), 1564–1573 (2022)
70. Zhao, S., Yang, X., Zeng, Z., Qian, P., Zhao, Z., Dai, L., Prabhu, N., Nordlund, P., Tam, W.L.: Deep learning based cetsa feature prediction cross multiple cell lines with latent space representation. Scientific Reports **14**(1), 1878 (2024)
71. Zhao, Z., Zhou, F., Xu, K., Zeng, Z., Guan, C., Zhou, S.K.: Le-uda: Label-efficient unsupervised domain adaptation for medical image segmentation. IEEE transactions on medical imaging **42**(3), 633–646 (2022)
72. Zheng, C., Yuan, K., Guo, B., Hadi Mogavi, R., Peng, Z., Ma, S., Ma, X.: Charting the future of ai in project-based learning: A co-design exploration with students. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. pp. 1–19 (2024)
73. Zhou, Z., Dou, G., Qu, X.: Brainactivity1: A framework of eeg data collection and machine learning analysis for college students. In: International Conference on Human-Computer Interaction. pp. 119–127. Springer (2022)