ORIGINAL RESEARCH

Machine Learning to Identify Clusters in Family Medicine Diplomate Motivations and Their Relationship to Continuing Certification Exam Outcomes: Findings and Potential Future Implications

David W. Price, MD, Peter Wingrove, MD, and Andrew Bazemore, MD, MPH

Background: **The potential for machine learning (ML) to enhance the efficiency of medical specialty boards has not been explored. We applied unsupervised ML to identify archetypes among American Board of Family Medicine (ABFM) Diplomates regarding their practice characteristics and motivations for participating in continuing certification, then examined associations between motivation patterns and key recertification outcomes.**

Methods: **Diplomates responding to the 2017 to 2021 ABFM Family Medicine continuing certification examination surveys selected motivations for choosing to continue certification. We used Chi-squared tests to examine difference proportions of Diplomates failing their first recertification examination attempt who endorsed different motivations for maintaining certification. Unsupervised ML techniques were applied to generate clusters of physicians with similar practice characteristics and motivations for recertifying. Controlling for physician demographic variables, we used logistic regression to examine the effect of motivation clusters on recertification examination success and validated the ML clusters by comparison with a previously created classification schema developed by experts.**

Results: **ML clusters largely recapitulated the intrinsic/extrinsic framework devised by experts previously. However, the identified clusters achieved a more equal partitioning of Diplomates into homogenous groups. In both ML and human clusters, physicians with mainly extrinsic or mixed motivations had lower rates of examination failure than those who were intrinsically motivated.**

Discussion: **This study demonstrates the feasibility of using ML to supplement and enhance human interpretation of board certification data. We discuss implications of this demonstration study for the interaction between specialty boards and physician Diplomates. (J Am Board Fam Med 2024;37:279–289.)**

Keywords: **ABFM, Accreditation, Certification, Chi-Square Test, Family Medicine, Information Technology, Logistic Regression, Machine Learning, Motivation, Physicians, Specialty Boards**

Introduction

Machine learning (ML) is a type of artificial intelligence (AI) that uses computers to "systematically apply algorithms to synthesize the underlying relationships among data and information."[1] ML is said to be unsupervised when algorithms are applied to data that has not been labeled; the computer identifies patterns or relationships among data without any initial human guidance.[2] Many industries use ML for purposes including text generation and analysis, speech recognition, image recognition, data analytics, and algorithmic recommendations of content to users[3]. Although more validation, research, quality control and oversight are needed, early and proposed uses of ML in health care include reading images,[4–6] making treatment recommendations,[7] patient and population risk-assessment,[8,9] natural

This article was externally peer reviewed.
Submitted 15 October 2023; revised 22 November 2023; accepted 4 December 2023.
From the American Board of Family Medicine, Lexington, KY (DWP, PW, AB); Department of Family Medicine, University of Colorado Anschutz School of Medicine, Aurora, CO (DWP); Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh PA (AB).
Funding: This work was internally funded by the American Board of Family Medicine. The authors report no competing interests.
Conflict of interest: None.
Corresponding author: David W. Price, MD 1648, McGrathiana Parkway, Suite 550 Lexington, KY 40511 (E-mail: DPrice@theabfm.org).

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230369R1 on 13 May 2024. Downloaded from http://www.jabfm.org/ on 9 October 2026 by guest. Protected by copyright.

language processing,[10] and various administrative tasks.[11]

Less explored are potential applications of ML to help medical specialty boards tailor their certification programs to individual physicians (Diplomates). ML could potentially assist medical specialty boards in several areas, including question development for summative and formative (learning) assessments, preliminary analysis of physician data and survey responses, elucidating patterns of physician engagement with and feedback on certification activities, identification of demographic and performance trends, and preliminary identification of physicians at risk of voluntarily or involuntarily failing to meet standards for continuing certification. In a previous study, Peabody and colleagues analyzed motivations for physician involvement with American Board of Family Medicine (ABFM) certification activities.[12] Building on this work, we sought to apply unsupervised ML techniques to identify ABFM Diplomate motivational patterns and contrast the results with a with these previous "human-defined" classifications. To explore whether the generated clusters reflect meaningful aggregations of Diplomates, we examined the association of motivation with rates of failure on the ABFM recertification examination for both the human-defined groups and the ML clusters.

Methods

Data

We combined data from the 2017 to 2021 ABFM Family Medicine continuing certification examination surveys. Because the questionnaire is mandatory, this captures all board-certified family physicians in a yearly cohort seeking to maintain their certification. Prior research has demonstrated its reliability as a representative sample of the overall population of family physicians.[13] Most Diplomates will have only completed 1 continuing certification examination survey; if applicants had multiple demographic surveys (eg, due to a previous examination failure in the time period), we took only the earliest available. Consistent with prior literature, we excluded physicians who did not provide direct outpatient continuity care. Assuming practice characteristics to be important determinants of external motivations,[14] we incorporated data on practice type/ownership/size as well as demographic variables (age, gender, international medical graduate (IMG) status, and degree (MD vs DO) as confounders in our regression analysis. We merged continuing certification survey with performance data for Diplomates taking the 1-day secure continuing certification examination between 2017 and 2021.

Variables

During unsupervised learning, data were limited to 4 questions from the recertification survey. The first question was "Why are you seeking to continue your ABFM certification at this time?" This question was a select-all-that-apply format with 11 possible binary (yes/no) responses, including "other" with free-text response. Because we sought to minimize the influence of human assumptions in this study, we omitted, rather than manually reclassified, the free-text responses. The remaining variables included practice type (11 categories), ownership status (5 responses), and practice size (4 responses) (Table 1).

We limited the analysis only to ABFM Diplomates who selected the 1-day examination, as the continuous certification process (FMCLA) was not a fully approved option until 2021. We selected the year of the first examination attempt within the study period. Physicians with ongoing failed attempts that began before 2017 were excluded. We used the pass/fail determination and the standardized examination score for the first examination attempt for each Diplomate. Our outcome was whether a Diplomate failed their first attempt at continuing certification examination.

Motivational Patterns

We used unsupervised ML (unprompted by the previous human-derived classification) to identify 5 mutually exclusive clusters of Diplomate motivation. (See Appendix for details). For ease of reference, we gave each cluster a descriptive title based on its predominant characteristic.

For comparison to the ML derived clusters, we replicated a previous human-identified pattern that classified the variables a priori into intrinsic vs extrinsic motivations and subdivided the physicians into 3 groups: "Extrinsic Only," "Intrinsic Only," and "Mixed."[13] A physician was classified as "Extrinsic only" if only 1 or more of the following 3 options were selected: Required by my employer; Required for hospital privileges/credentialing; and/or Required by 1 or more payer/insurance company. A physician was classified as "Intrinsic only" if only 1 or more of the following 7 options were selected: Maintain professional image; Personal preference; Professional

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230369R1 on 13 May 2024. Downloaded from http://www.jabfm.org/ on 9 October 2026 by guest. Protected by copyright.

Table 1. American Board of Family Physician Recertification Survey Questions Used in This Study (2017–2021)

Why are you seeking to continue your ABFM certification at this time?

a. Maintain professional image
b. Personal preference
c. Professional advancement
d. Maintain or improve patient satisfaction
e. Patients prefer being treated by board certified physicians
f. Certification program helps me update my medical knowledge
g. Certification program helps me monitor or improve the quality of my patient care
h. Required by my employer
i. Required for hospital privileges/credentialing
j. Required by one or more payer/insurance company
k. Other (please specify)

Which of the following describes your principal practice site:

a. Academic health center / faculty practice (residency or university teaching environment)
b. Federal (Military, Veterans Administration/Department of Defense)
c. Federally Qualified Health Center or Look-Alike
d. Government clinic, non-federal (eg, state, county, city, maternal and child health, public health center, etc.)
e. Hospital-/health system-owned medical practice (not including managed care or HMO)
f. Independently owned medical practice
g. Indian Health Service
h. Managed care / HMO practice
i. Other, please specify:
j. Rural Health Clinic (federally qualified) Indian Health Service
k. Worksite clinic

Which of the following best describes your role in the ownership of your principal practice

a. No official ownership stake (100% employed)
b. Sole owner
c. Partial owner or shareholder
d. Self-employed as a contractor (including locums)
e. Other, please specify:

Which of the following describes your principal practice size:

a. Solo practice
b. 2 to 5 providers
c. 6 to 20 providers
d. >20 providers

advancement; Maintain or improve patient satisfaction; Patients prefer being treated by board certified physician; Certification program helps me update my medical knowledge; and/or certification program helps me monitor or improve the quality of my patient care. All remaining physicians, who definitionally had selected at least 1 motivation from both categories (except a trivial number who selected only "Other"), were assigned to "Mixed motivation." Because the original article did not incorporate information beyond motivation, we did not devise new rules to reassign Diplomates or further subdivide the original 3 groups into subtypes based on differences in practice type/ownership/size.

Statistical Analysis

We used Chi-squared tests to examine the difference in the proportion of Diplomates failing their first examination attempt who endorsed each of the 11 motivations. Average rates of examination failure on the initial attempt were calculated for the 3 human classification groups and the clusters generated from unsupervised ML. We then used logistic regression to examine the effect of motivation holding other demographic variables constant. Age was divided into tertials; gender, degree, and IMG status were used as binary variables.

The ML components of this analysis were conducted with R 4.2.3. All other components were conducted with Stata 17.0.

As part of routine internal program evaluation, this study was conducted with ethical approval from the American Academy of Family Physicians Institutional Review Board.

Results

ML Clusters

Table 2 shows the 5 ML identified clusters, along with the proportion of Diplomates within that

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230369R1 on 13 May 2024. Downloaded from http://www.jabfm.org/ on 9 October 2026 by guest. Protected by copyright.

Table 2. Percentage of Recertification Diplomates (2017–2021) within Each ML Cluster or Human Generated Group Who Endorsed Each Motivation

Cluster	Description	N (%)	Image	Personal	Professional	Patient satisfaction	Patient preference	Medical knowledge	Patient care	Employer	Credential	Payer	Other
Machine learning clusters													
1	Credential + payer	2895 (19.3)	36	24	15	6	22	22	13	15	65	62	3
2	Employer	3208 (21.4)	32	29	17	6	16	25	15	70	27	10	3
3	Mainly extrinsic	2491 (16.6)	33	25	11	2.5	18	28	16	86	92	73	1
4	Broadly motivated	3553 (23.7)	88	86	75	70	83	93	86	73	77	66	1
5	Mainly intrinsic	2821 (18.8)	74	84	39	25	30	83	71	13	20	13	2
Human classification													
Extrinsic only		2626 (17.5)	0%	0%	0%	0%	0%	0%	0%	62%	64%	50%	2%
Intrinsic only		3067 (20.5)	59%	70%	39%	25%	29%	57%	49%	0%	0%	0%	3%
Mixed		9275 (62.0)	68%	60%	41%	31%	48%	65%	52%	66%	72%	58%	2%

Shaded cells indicate major representative objects that different a cluster from others. Underlined cells indicate the allowed motivations for the Extrinsic Only and Intrinsic Only groups; all other combinations of motivations (ie, those that included at least one extrinsic motivation and at least one intrinsic motivation) were assigned to the Mixed group.

cluster who endorsed a particular motivation. Highlighted motivations were used to give each cluster a descriptive title. The table also displays similar metrics for the 3 human-derived subgroups. Human sorting resulted in imbalanced cohort groupings, with more than 60% assigned to the "Mixed" group.

Table 3 shows how the practice characteristics map to each cluster. There is an association of certain practice type/sizes/ownerships with a specific cluster of extrinsic motivation. For example, almost 50% of Diplomates working at independently owned sites were assigned to Cluster 1 (Credential and Payer), but only 4% and 2% were assigned to Cluster 2 (Employer) and Cluster 3 (Predominantly Extrinsic). The opposite trend is seen for nonphysician owned sites like academic medical and Federally Qualified Health Centers. Within almost all responses to practice characteristics there were Diplomates assigned to Cluster 4 (Mixed motivations) because they had extrinsic motivations associated with their practice location but also reported a number of intrinsic motivations.

Motivation and Initial Recertifying Examination Success

Table 4 shows associations between each Diplomate survey question on motivation and failure rate for the initial attempt at the continuing certification examination. Diplomates had statistically significantly higher rates of failure if they had endorsed motivations of professional advancement ($P<.001$), a desire to maintain/improve patient satisfaction ($P<.001$), and a desire to monitor or improve patient care ($P=.013$). They had lower rates of failure when endorsing motivations of image ($P=.041$), patient preference ($P<.001$), and employer ($P<.001$), credential ($P<.001$), and payer ($P<.001$) requirements.

Figure 1 shows the adjusted probability of failing a continuing certification examination attempt after controlling for potential confounding from gender, age, IMG status, and type of degree (MD vs DO). Four of 5 ML- and 2 of 3 human-created classifications achieved a statistically significant separation from the overall rate of failure (7.2%). Diplomates clustered into "Mainly Extrinsic" or "Broadly Motivated" had a significantly lower probability of failing (4.3% ($P<.001$) and 5.1% ($P<.001$, respectively). Diplomates in the clusters of "Credential + Payer" and "Mainly Intrinsic" had statistically significantly higher rates of failure at 8.7% ($P=.002$) and 9.1% ($P<.001$), respectively. The fail rate for

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230369R1 on 13 May 2024. Downloaded from http://www.jabfm.org/ on 9 October 2026 by guest. Protected by copyright.

Table 3. Breakdown of % of 2017–2021 Recertification Diplomates Within Each Practice Site/Ownership/Size Category Assigned to Each Cluster

Cluster	1 (Credential + Payer)	2 (Employer)	3 (Mainly extrinsic)	4 (Broadly motivated)	5 (Mainly intrinsic)
Variable: Type of Site					
Academic (1046)	2%	29%	33%	30%	7%
Federal (570)	3%	37%	13%	20%	28%
FQHC (951)	3%	26%	26%	24%	21%
Govt, non-federal (230)	2%	36%	15%	18%	28%
Hospital/health system (5258)	5%	27%	28%	30%	9%
Independently Owned (4908)	48%	4%	2%	17%	29%
Indian Health Service (91)	3%	26%	19%	27%	24%
Managed care/ HMO (894)	8%	47%	7%	24%	14%
Other (448)	10%	32%	13%	19%	27%
Rural health clinic (345)	17%	20%	19%	21%	22%
Work site clinic (227)	10%	35%	7%	18%	30%
Site Ownership					
Employee (9584)	5%	28%	25%	27%	15%
Other (186)	22%	28%	5%	20%	26%
Partial owner /stakeholder (2863)	44%	13%	2%	20%	21%
Self-employed [contractor] (427)	32%	15%	8%	18%	26%
Sole owner (1908)	51%	1%	0%	16%	32%
Site size					
Solo practice (1687)	42%	8%	3%	18%	29%
2-to-5 (5068)	26%	18%	9%	21%	27%
6-to-20 (4687)	11%	15%	33%	28%	13%
>20 (3526)	11%	42%	12%	25%	10%

Shaded cells for each variable indicate the major practice characteristics of each cluster (column).
Abbreviation: FQHC, federally qualified health center.

the "Employer" cluster was not statistically significantly different from the overall mean (7.6% vs 7.2%, $P = .378$). Among human-created classifications, the "Mixed" subgroup had a lower probability of failure (5.7%, $P < .001$) and the "Intrinsic Only" subgroup had a higher probability of failure (11.1%, $P < .001$). There was no statistically significant difference observed for the "Extrinsic Only" human created group (6.4% vs 7.2%, $P = .129$).

Discussion

This novel study explored the feasibility of a medical specialty board using unsupervised ML to gain insight from complex datasets. We identified new patterns of Diplomate motivation for participating in continuing board certification, compared the results to a previous human-defined characterization, and retrospectively associated both machine and human-derived clusters with the probability of failing the 1-day continuing certification examination.

Despite not being fed prespecified information on the difference between intrinsic and extrinsic motivations, the clusters derived from unsupervised ML align with literature that characterizes motivation as extrinsic or intrinsic.[15] Three clusters were consistent with only extrinsic motivations, 1 with only intrinsic motivations, and the last had a mixture. Though the computer does much of the heavy lifting, obviating the need for a human to manually think through and create algorithms that classify each Diplomate, this is not the only possible outcome. The final analysis is still heavily influenced by human decisions such as selection of model(s) to use, variables to include, and the desired number of clusters to generate. Variations in these human decisions may lead to substantively different cluster assignments for individual Diplomates as well as greater or lower feasibility of implementing interventions commensurate with how many unique clusters are generated.

As demonstrated here, 1 major advantage of ML is the potential to achieve more balanced and

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230369R1 on 13 May 2024. Downloaded from http://www.jabfm.org/ on 9 October 2026 by guest. Protected by copyright.

Table 4. Percent of 2017–2021 American Board of Family Medicine Diplomates Failing Initial Attempt at Recertification Exam, by Endorsement of Motivation

Motivation	Endorses motivation	Does not endorse motivation	P (χ^2)
Intrinsic			
Image	6.8 (547/8093)	**7.6** (524/6875)	0.041*
Medical knowledge	7.3 (569/7760)	7.0 (502/7208)	0.383
Monitor/improve patient care	**7.8** (492/6334)	6.7 (579/8634)	0.013*
Patient preference	6.1 (327/5366)	**7.8** (744/9602)	<0.001***
Patient satisfaction	**8.5** (308/3617)	6.7 (763/11,351)	<0.001***
Personal preference	7.1 (548/7679)	7.2 (523/7289)	0.927
Professional advancement	**8.2** (413/5035)	6.6 (658/9923)	<0.001***
Extrinsic			
Credentialing	5.6 (464/8325)	**9.1** (607/6643)	<0.001***
Employer	4.5 (349/7772)	**10.0** (722/7196)	<0.001***
Payer	5.6 (374/6665)	**8.4** (697/8303)	<0.001***
Other	9.2 (26/283)	7.1 (1045/14,685)	0.181

*$P < .05$, **$P < .01$, ***$P < .001$.
Bold figures indicate a higher probability of failure.

granular groups, especially when working with complicated datasets containing many interrelated variables that are not always intuitive for humans to understand. Given the small number of classification categories and straightforward assignment rules, it is not surprising that roughly 60% of Diplomates were originally assigned to the "Mixed" motivation group in the article by Peabody et al.[13] In this study, as shown in Table 2, many of the cells in the ML-identified clusters had values greater than 70% for individual motivations; several contain values between 80% and 93%. In contrast, only 2 cells in the human-identified clusters contain values more than 70%. This illustrates the potential of ML to generate more specific clusters than human classification.

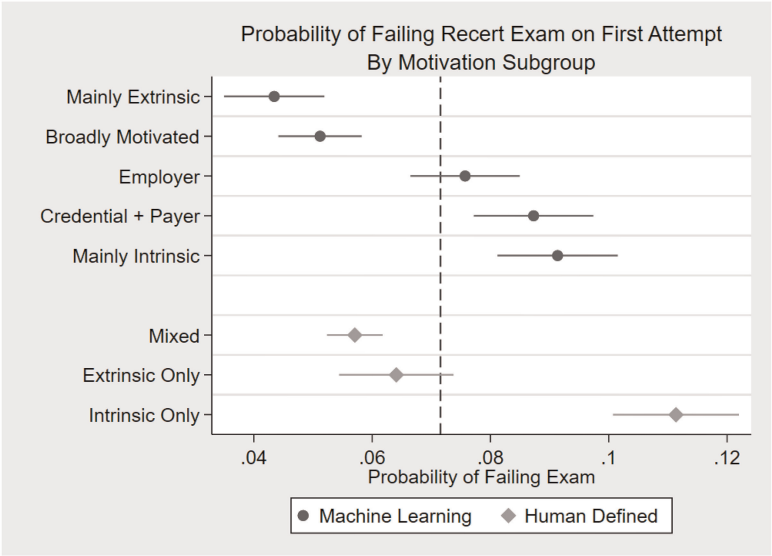
After analysis with unsupervised learning, there were 5 clusters of approximately equal size; roughly 80% of Diplomates clustered into 1 of the 4 groups found to have statistically significantly different rates of failure relative to the baseline. We also found the human-defined "intrinsic only" group was the strongest positive predictor of recertification examination failure, even though the ML clusters seem to have done better at identifying physicians at noticeably below baseline risk. This does not imply that human-defined groups were better or worse than the ML clusters at predicting examination failure. Creating groups to predict examination failure was not a goal of the Peabody et al[13] article. The ML techniques in this article were selected to test the feasibility of identifying coherent clusters of Diplomates reporting similar motivations. However, the higher rate of failure seen in the intrinsic only group suggests that there may be a threshold effect where a single extrinsic motivation is protective. Although self-determination theory and the learning literature note the importance of internal motivation for learning,[16–20] these results reinforce that at least some elements of extrinsic motivation are important for learning outcomes – in other words, "stakes matter."[21–24] It seems plausible that some Diplomates with solely intrinsic motivations are at higher risk of examination failure, perhaps because they may not have immediate external consequences if they fail to recertify. If lack of extrinsic motivation is thought to be a potentially important predictor in future ML models looking at examination failure, a priori designation of extrinsic and intrinsic variables would be required. Research with ML remains an iterative process incorporating human hypotheses and interpretation of results and their implications.

Implications

This pilot illustrates the potential for certifying boards to use ML to create more tailored strategies for diplomate communication. Boards might use motivational pattern differences between Diplomates who perform well on board certification activities with those with lesser performance to develop targeted, proactive outreach and "coaching" of physicians

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230369R1 on 13 May 2024. Downloaded from http://www.jabfm.org/ on 9 October 2026 by guest. Protected by copyright.

Figure 1. Probability of failing recertification examination on first attempt, by motivation subgroup.



Notes: Dashed line represents the overall percentage of American Board of Family Medicine Diplomates who failed their initial attempt at the continuing certification exam, 2017–2021 (7.2%).

at risk for not maintaining or losing their certification. For example, Diplomates assigned to Cluster 1 in our classification (commonly having at least partial ownership in practice and reporting motivations of only credential and/or payer requirements) might respond positively to outreach emphasizing how board certification aligns with expectations of insurance companies and hospitals.[25] Someone in Cluster 5 (commonly an employee at a small independent practice who selected multiple intrinsic motivations and no extrinsic motivations) may be less interested in evidence on economic benefits of board certification and more interested in literature demonstrating how board certification can maintain and enhance the quality of care he/she is able to provide to patients.[26] Medical specialty boards could seek input from their Diplomates on the content and wording of tailored communications and evaluate the effect of these tailored messages on motivation, patterns of engagement with different parts of continuing certification, and outcomes (eg, examination performance, improvement in knowledge) over time.

A broader vision for ML in continuous board certification includes facilitating a bespoke process of continuous learning and improvement. Tailored Diplomate outreach could help increase engagement and retention in continuous learning through board certification. However, motivation patterns may differ between formative/lower-stakes activities and higher stakes activities. Intrinsic motivation might drive participation or performance on formative components, whereas those who are primarily extrinsically motivated might be doing 'just enough to get through' and focus their energy on the summative component. Therefore, associations between the motivational patterns identified in this study and engagement in and performance on other parts of continuing certification should be explored.

Beyond motivational patterns, ML's ability to identify patterns from complex data sets can assist specialty boards in evolving and personalizing other parts of the continuous certification process. For example, ML could help identify individual educational gaps based on self-assessment module, examination performance or real-world practice data, and suggest journal articles, conferences, or other continuing medical education (CME) activities for Diplomates to pursue. These "nudges"[27] could potentially be tailored based on Diplomate motivational profiles. Applying unsupervised ML techniques to data on self-reported scope of practice, Medicare claims, or other registry data (eg, the ABFM PRIME registry) might help boards suggest areas for practice improvement to personalized individual Diplomates. Such microsegmentation of Diplomates by preferences and knowledge gaps, paired with techniques such as spaced repetition,[28] could ultimately

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230369R1 on 13 May 2024. Downloaded from http://www.jabfm.org/ on 9 October 2026 by guest. Protected by copyright.

transform certification online platforms in a manner akin to those currently deployed in the language learning environment, such as Duolingo or Rosetta Stone®. MLs data analysis ability could also help boards prioritize areas for development of self-assessment, secure examination questions, or performance improvement activities. ML could, more efficiently than human analysis, identify predominant educational gaps by Diplomate geographic location that can serve as needs assessment data for organizational or state chapter CME providers.

As in society at large, authors are writing about both MLs potential to transform health care and health care education, but the ethical implications of its use.[29–31] A full discussion of the ethical implications of ML in board certification is beyond the scope of this article. Briefly, however, boards should consider issues such as bias in ML algorithms, transparency of ML assisted decision making processes, implications for question development, and question and data privacy and security as they integrate ML into their processes. Boards should also consider how physicians will use AI and ML to address clinical questions in practice, and to what degree to allow or account for it on summative assessments.[32,33]

Limitations

Although we were able to demonstrate promising patterns in associations between ML-derived motivational clusters and examination performance, several limitations should be considered. Motivation may be a signal for other behaviors linked with (rather than a direct cause of) poor examination outcomes. For example, examination preparation or engagement with certification in general may vary across different motivational clusters. Proximity to retirement and motivation to maintain certification may be related, although we controlled for age (often colinear with time in practice) in this analysis. Using supervised learning on the full breadth of ABFM data (including data such as time spent in self-assessment modules) to proactively predict examination failure to pass an examination is an area of ongoing research. With increasing numbers of Diplomates opting for the continuous Family Medicine Certification Longitudinal Assessment in lieu of the 1-day secure examination, a parallel analysis should be conducted to see if these motivational patterns hold (or how they differ) using this different high-stakes examination format.

Our analysis was limited to family physicians engaged in outpatient continuity care. Further work is needed to assess patterns among physicians in other specialties as well as family physicians working as administrators or in focal settings [for example, emergency, hospital, urgent care, exclusively sports medicine]. We lack data to assess relationships between motivation and study behavior, and we anticipate further effort to understand how motivation relates to certification activities beyond the 1-day, secure, high-stakes recertification examination.

Outputs of ML techniques, as with traditional statistics, reflect the limitations of their underlying data. Diplomate motivation data gathered via the recertification survey is binary. Although we could determine the presence or absence of each motivation, it was not possible to determine the most important or the relative strength of individual motivations for a given individual. It is possible that some Diplomates felt extremely strongly about 1 motivation above the rest; an algorithm on data that included additional granularity about strength of motivation might result in a different cluster assignment. It is possible that some Diplomates did not feel strongly about any motivation and randomly selected just 1 because completing the question was required to advance the survey. Conversely, though there is no current benefit to reporting a large number of motivations, some Diplomates may have chosen to deliberately overstate their motivations to "virtue signal." ML alone could not address these limitations without Diplomate discussions or changes to the Diplomate survey. If this added level of granularity is pursued, it should be done with attention to minimizing Diplomate administrative burden.

Conclusion

Large language models are already impacting clinical care delivery[34,35]; we anticipate a need for commensurate changes to keep the continuous certification paradigm relevant. This study demonstrates both the feasibility and potential impact of applying ML techniques to better understand physician diplomate motivations and preferences. It also reveals how such techniques can supplement and efficiently enhance human judgment in interpreting certification data. The clusters we identified with unsupervised learning

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230369R1 on 13 May 2024. Downloaded from http://www.jabfm.org/ on 9 October 2026 by guest. Protected by copyright.

may help boards better understand certification-related motivational patterns among Diplomates. It could help boards develop targeted strategies for outreach to physicians who may be at risk of failing to maintain board certification. Beyond this initial foray, we imagine how ML-derived motivational clusters, and ML techniques in general, might help boards improve physician engagement across the components of continuous specialty certification. Additional feasibility studies, incorporating Diplomate feedback, will be important to explore how ML techniques can be used to improve the development, applicability, and outcomes of board certification activities and increase the value of board certification to Diplomates.

The authors thank Zachary Morgan, MS, and Shawn Reynolds for their assistance with data access.

To see this article online, please go to: http://jabfm.org/content/37/2/279.full.

References

1. Awad M, Khanna R. (2015). Machine learning. In: *Efficient learning machines*. Berkeley, CA: Apress.
2. Gentleman R, Carey VJ. (2008). *Unsupervised Machine Learning*. In: Bioconductor case studies. Use R!. New York, NY: Springer.
3. Columbia Engineering Boot Camps, Top Machine Learning Applications by Industry: 6 Machine Learning Examples. Available at: https://bootcamp.cvn.columbia.edu/blog/machine-learning-applications/#1640055350333-cb4a9594-ec51. Accessed July 21, 2023.
4. Fakoor R, Ladhak F, Nazi A, Huber M. Using deep learning to enhance cancer diagnosis and classification. Presentation at The 30th International Conference on Machine Learning, 2013, Atlanta, GA, USA. Available at: https://www.researchgate.net/publication/281857285_Using_deep_learning_to_enhance_cancer_diagnosis_and_classification.
5. Vial A, Stirling D, Field M, et al. The role of deep learning and -radiomic feature extraction in cancer-specific predictive modelling: a review. Transl Cancer Res 2018;7:803–16.
6. Loria K. Putting the AI in radiology. Radiology Today 2018;19:10.
7. Lee S-I, Celik S, Logsdon BA, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid-leukemia. Nat Commun 2018;9:42.
8. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. npj Digital Med 2018;1:18.
9. Nait Aicha A, Englebienne G, van Schooten K, Pijnappels M, Kröse B. Deep learning to predict falls in older adults based on daily-life trunk accelerometry. Sensors 2018;18:1654.
10. Le Glaz A, Haralambous Y, Kim-Dufor D, et al. Machine learning and natural language processing in mental health. J Med Internet Res 2021;23:e15708.
11. Javaid M, Haleem A, Pratap Singh R, Suman R, Rab S. Significance of machine learning in healthcare: features, pillars and applications. International Journal of Intelligent Networks 2022;3:58–73.
12. Peabody MR, Peterson LE, Dai M, Eden A, Hansen ER, Puffer JC. Motivation for participation in the American Board of Family Medicine Certification Program. Fam Med 2019;51:728–36.
13. Peterson LE, Fang B, Phillips RL, Jr, Avant R, Puffer JC. The American Board of Family Medicine's data collection method for tracking their specialty. J Am Board Fam Med 2019;32:89–95.
14. Mahmoudi G, Rostami F, Mahmoudjanloo S, Jahani M. Relationship of employees' achievement motivation and quality of working life with their self-efficacy at selected hospitals with a multigroup analysis: moderating role of organizational ownership. Mater Sociomed 2017;29:237–41.
15. Morris LS, Grehl MM, Rutter SB, Mehta M, Westwater ML. On what motivates us: a detailed review of intrinsic v. extrinsic motivation. Psychol Med 2022;52:1801–16.
16. Ten Cate TJ, Kusurkar RA, Williams GC. How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE guide No. 59. Med Teach 2011;33:961–73.
17. Kusurkar RA, Ten Cate TJ, van Asperen M, Croiset G. Motivation as an independent and a dependent variable in medical education: a review of the literature. Med Teach 2011;33:e242-62–e262.
18. Vansteenkiste M, Simons J, Lens W, Sheldon KM, Deci EL. Motivating learning, performance, and persistence: the synergistic effects of intrinsic goal contents and autonomy-supportive contexts. J Pers Soc Psychol 2004;87:246–60.
19. Morris LS, Grehl MM, Rutter SB, Mehta M, Westwater ML. On what motivates us: a detailed review of intrinsic v. extrinsic motivation. Psychol Med 2022;52:1801–16.
20. Chaudhuri JD. Stimulating intrinsic motivation in millennial students: a new generation, a new approach. Anatomical Sciences Ed 2020;13:250–71.
21. Wood T. Assessment not only drives learning, it may also help learning. Med Educ 2009;43:5–6.
22. Putwain DW, Nicholson LJ, Pekrun R, Becker S, Symes W. Expectancy of success, attainment value, engagement, and achievement: a moderated mediation analysis. Learning and Instruction 2019;60:117–25.

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230369R1 on 13 May 2024. Downloaded from http://www.jabfm.org/ on 9 October 2026 by guest. Protected by copyright.

23. Fraundorf SH, Caddick Z, Rottman B, et al. Conceptual foundations for designing continuing certification assessments for physicians. Available at: https://www.theabfm.org/sites/default/files/PDF/Conceptual-Foundations.pdf. Accessed July 21, 2023.
24. Price DW, Wang T, O'Neill TR, Bazemore A, Newton WP. Differences in physician performance and self-rated confidence on high- and low-stakes knowledge assessments in board certification. J Contin Educ Health Prof 2023;Publish Ahead of Print.
25. Payor Enrollment: Understanding the CMS guidelines for hospitals and health plans. Available at: https://www.symplr.com/blog/payor-enrollment-understanding-cms-guidelines-for-hospitals-health-plans. Accessed September 15, 2023.
26. Price DW, Biernacki H, Nora LM. Can maintenance of certification work? Associations of MOC and improvements in physicians' knowledge and practice. Acad Med 2018;93:1872–81.
27. Thaler RH, Sunstein CR. (2008). Nudge: improving decisions about health, wealth, and happiness. Yale University Press.
28. Price DW, Wang T, O'Neill TR, Newton WP. Spaced repetition in a cohort of practicing—methods and preliminary results. Academic Medicine 2022;97: Supplement: S1–S141.
29. Schönberger D. Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. International Journal of Law and Information Technology 2019;27:171–203.
30. Morley J, Machado CCV, Burr C, et al. The ethics of AI in health care: a mapping review. Social Science & Medicine 2020;260:113172.
31. Masters K. Ethical use of artificial intelligence in health professions education: AMEE guide no. 158. Med Teach 2023;45:574–84.
32. Swiecki Z, Khosravi H, Chen G, et al. Assessment in the age of artificial intelligence. Comput. Educ. Artif. Intell 2022;3:100075.
33. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023;9: e45312.
34. Thirunavukarasu AJ, Hassan R, Mahmood S, et al. Trialling a large language model (ChatGPT) in general practice with the applied knowledge test: observational study demonstrating opportunities and limitations in primary care. JMIR Med Educ 2023Apr 21;9:e46599.
35. Forbes Newsletter. The impact of large languages models on healthcare. Available at: https://www.forbes.com/sites/forbestechcouncil/2023/06/20/the-impact-of-large-languages-models-on-healthcare/?sh=781d01e66292. Accessed May 6, 2024.

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230369R1 on 13 May 2024. Downloaded from http://www.jabfm.org/ on 9 October 2026 by guest. Protected by copyright.

Appendix.

Unsupervised Machine Learning Methods Used in This Study

Because our training data consisted solely of categorical variables, we used k-medoids with Hamming distances.[1] A medoid is the representative object from a cluster such that each other member of the cluster is more like that medoid than to the medoid of any other cluster. The Hamming distance between 2 Diplomates is the total number of variables where they have different responses (eg, the Hamming distance between 2 Diplomates is 1 if they have identical responses for each motivation component as well as the same category of practice ownership/size but 1 works at a staff-model health maintenance organization and the other works in a federally qualified health-center). Because we were attempting to perform a cluster analysis on thousands of applicants, we used a sampling approach to reduce computing time and avoid random-access memory storage limitations. We used the ©cluster' package in R to implement CLARA (Clustering Large Applications), which is an extension of sampling to k-medoids methods. To minimize the possibility of cherry-picking results when selecting the number of clusters to use in the final analysis, we used validated clustering approaches including examining the silhouette, gap statistic, and total sum of squares. This used the ©factoextra' package with potential cluster parameter ranging from 2 to 10 and nboot set to 500 iterations. After preliminary examination of cluster silhouettes, gap-statistics, and Within Cluster Sum of Squares to identify the optimal number of clusters, we settled on having the unsupervised learning generate 5 mutually exclusive clusters; for ease of reference and graphical display, we gave each cluster a descriptive title based on the motivations of the medoid of that cluster.

Reference

1. Jin X, Han J. (2011). K-Medoids Clustering. In: Sammut C, Webb GI (eds). *Encyclopedia of machine learning*. Boston, MA: Springer. Available at: https://doi.org/10.1007/978-0-387-30164-8_426.

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230369R1 on 13 May 2024. Downloaded from http://www.jabfm.org/ on 9 October 2026 by guest. Protected by copyright.