ORIGINAL ARTICLE

Volume 17 Issue 1 2025

DOI: 10.21315/eimj2025.17.1.2

ARTICLE INFO

Submitted: 23-04-2024
Accepted: 01-11-2024
Online: 26-03-2025

Can We Make Early Predictions to Identify Which Medical Interns are at Risk of Failing a Clinical Posting?

Mohd Salami Ibrahim[1], Nurulhuda Mat Hassan[2], Ong Yu Xuan[2], Norwati Daud[2], Nik Mohd Ronaidi Nik Mahdi[3], Yuzana Mohd Yusop[3]

[1]*Department of Medical Education, Faculty of Medicine, Universiti Sultan Zainal Abidin, Terengganu, MALAYSIA*

[2]*Department of Family Medicine, Faculty of Medicine, Universiti Sultan Zainal Abidin, Terengganu, MALAYSIA*

[3]*Department of Community Medicine, Faculty of Medicine, Universiti Sultan Zainal Abidin, Terengganu, MALAYSIA*

To cite this article: Ibrahim MS, Mat Hassan N, Ong YX, Daud N, Nik Mahdi NMR, Mohd Yusop Y. Can we make early predictions to identify which medical interns are at risk of failing a clinical posting? Education in Medicine Journal. 2025;17(1):3–20. https://doi.org/10.21315/eimj2025.17.1.2

To link to this article: https://doi.org/10.21315/eimj2025.17.1.2

ABSTRACT

Passing a medical internship is a crucial professional milestone. We conducted a prospective study to assess the feasibility of early prediction in identifying interns at risk of failing a clinical posting. We surveyed 496 newly enrolled interns across 26 Malaysian hospitals from January to April 2020, using validated instruments to evaluate various factors related to personal attributes, place of study, and place of practice. After one year, we followed up with the participants to identify those who had failed a clinical posting. Significant predictors were determined using the supervised machine learning (ML) framework to linear discriminant analysis (LDA), with the prediction performance validated through split train-test and cross-validation. The LDA identified a higher risk of clinical posting failure among interns from 8 specific hospitals and 13 medical schools, as well as those with poor interpersonal skills, an avoidant coping style, greater preparedness in information technology, married status, and a shorter gap between graduation and internship commencement. The model achieved 100% sensitivity, 85.7% specificity, 86.7% accuracy, and an area under the curve (AUC) of 0.969 for the training dataset. Cross-validation showed 92.0% sensitivity, 82.9% specificity, 87.4% accuracy, and an AUC of 0.905, while the validation dataset achieved a sensitivity of 75.0%, specificity of 81.0%, accuracy of 80.7%, and an AUC of 0.861. This study shows that surveys of newly enrolled medical interns can effectively predict those at risk of failing a clinical posting within a year. We recommend replicating this study in various countries and practice regions to improve generalisability. Additionally, qualitative studies could provide deeper insights.

Keywords: *Competency, Coping, Interpersonal skills, Medical interns, Resilience*

CORRESPONDING AUTHOR

Nurulhuda Mat Hassan, Department of Family Medicine, Faculty of Medicine, Universiti Sultan Zainal Abidin, 20400 Kuala Terengganu, Terengganu, Malaysia

Email: nurulhudamh@unisza.edu.my

INTRODUCTION

The challenge of medical interns not meeting the expected competency during a clinical posting carries far-reaching consequences. Such situations prompt a thorough examination of the underlying factors contributing to these outcomes. It necessitates an evaluation of complex dynamics and interactions across multi-dimensional aspects of education and training to derive meaningful explanations and to guide a clear trajectory of planning and investment towards quality assurance and quality improvement. Within the context, this phenomenon places supervising clinicians, hospital management, educators, and policymakers in a complex position to elucidate why medical graduates, previously considered competent and qualified, cannot exhibit the requisite competencies during their clinical training.

The immediate interest is in the medical curriculum itself. The global trend has been moving towards competency-based medical education (CBME) since the late 1990s. CBME signifies a renewed emphasis on an education that prioritises competent graduates on multiple pre-determined domains as the overarching outcome of the undergraduate medical curriculum. Some critiques have raised concerns about the philosophy that defines education based on outcomes, especially given the behaviouristic nature of competency that originated from vocational training (1). Nonetheless, the medical education community has recognised the key advantages of a curriculum design that promises clarity that illuminates milestones of achievement to communicate robust, systematic, transparent, relevant, and progressive education (2). For the gaps between competency-based training and competency during the medical internship, we share a similar view with Miller (3), whose seminal work on the hierarchy of assessment has separated competency as an achievement that could be shown during training and simulation against performances, which is the action in actual practice. Thus, while medical schools have fundamental roles in preparing their graduates for real medical practice, we believe there is more to the curriculum that influences medical interns' workplace-based clinical performance.

Burnout is a widely considered topic, sometimes being perceived as the "God term" encompassing issues such as anxiety, depression, substance abuse and cognitive exhaustion (4). Concisely reviewing the key literature, Khan et al. (4) outlined the feelings of exhaustion, mental detachment from one's job, a sense of ineffectiveness, and lack of accomplishment to define burnout. Attempting to remediate this concern, earlier research by Kreitzer and Klatt (5) discussed how medical education has inadequately addressed stressful and occasionally dysfunctional work environments that lead to pathological compassion fatigue. They responded with systematic interventions at Georgetown University School of Medicine on mindfulness, mind-body skills, and resiliency via structured self-assessments, weekly reading and watching videos, reflective writing, peer interactions and guided discussions (5). More research on burnout emerges to investigate the association with resilience (6), emotional intelligence (7), coping styles (8, 9), interest in practising medicine (10, 11), preparatory courses (12), and religious belief (13). Therefore, understanding these attributes for medical interns' performance may prove helpful in resource planning and quality improvement for the training of future medical interns.

The gravity of this understanding consequently signifies broader aspects of medical internship training that are influenced by factors beyond psychological determinants. An interesting systematic review by Klein et al. (14) searched the Medline and Embase databases to determine the extent of evidence related to gender bias in postgraduate medical education. Up to 2018, they identified nine studies assessing gender bias in resident evaluations, including rating scores and qualitative comments. Despite the heterogeneity of the studies preventing a meta-analysis, five studies found that gender significantly influenced traits ascribed to residents, feedback consistency, and performance measures. The authors concluded that gender bias potentially undermines the integrity of resident assessments in graduate medical education. They called for future research to understand how gender bias manifests, its impact on learners, and strategies to mitigate it. In comparison, Sitobata and Mohammadnezhad (15) conducted in-depth interviews among 22 medical interns at a Vanuatu Hospital to explore their challenges during the internship. They insightfully synthesised three themes: the intern's unmet welfare, different medical training institutions, and transitional shock. These themes encapsulated broad issues that affect the quality of internship training, such as poor orientation, working like a registrar, and issues related to accommodation (15). Therefore, both studies concluded the need for policy commitment to address commonly overlooked issues despite their direct impact on medical interns' clinical performance.

In summary, medical interns failing a clinical posting may be explained by reasons related to the place of study, place of practice and individual attributes. This study aims to develop a predictive model to identify interns at risk of failure by determining the key predictors. Identifying these factors can guide early interventions to improve clinical competency during internships. Bridging the gap between competency-based training and real-world performance is crucial to ensuring that medical graduates are not only academically prepared but also capable of excelling in practice. This research offers a data-driven approach to enhancing the effectiveness of internship training programmes.

MATERIALS AND METHODS

Study Design and Instruments for Data Collection

This study conceptualised the following supervised machine learning (ML) research framework to address this research question. Initially, we surveyed the newly enrolled medical interns for potential predictors of competency based on existing literature. Subsequently, we followed them up after one year to identify those who had failed a clinical posting during that period. The follow-up results served as the labelled outcomes for the ML algorithm to train. Then, we conduct variable-selection exploratory analysis to determine the significant predictors, followed by developing a predictive model with two distinct validation approaches.

Consequently, this prospective cohort design study was conducted from January 2020 until April 2021. The online questionnaire via Google Forms consisted of five validated instruments: the Preparedness for Hospital Practice Questionnaire (PHPQ) (12), Connor-Davidson Resilience Scale short version 10 (CD-RISC 10) (16), Brief-COPE questionnaire (17), USM Emotional Quotient Inventory (USMEQ-i) (18), and Duke University Religion Index (DUREL) (19). These validated instruments measured medical interns' preparedness for clinical practice, resilience, coping, emotional intelligence, and religiosity. Participants also provided information on demographics (age, gender, ethnicity, marital status), place of practice and study, history of clinical posting failure during medical school, interest in medical practice, waiting duration (in months) from graduation to internship commencement, and contact details for one-year follow-up.

Recruitment

Through purposive sampling, we invited newly registered medical interns from 26 accredited hospitals for internship training in Malaysia to complete an online questionnaire from January to April 2020. We approached a few representatives from each hospital and asked them to invite their peers through email and WhatsApp. In this study, we included participants who were newly enrolled to commence their internship from January 2020 to April 2020 and who understand both English and Malay. We excluded those who were doing part of their medical internship training abroad. We aimed for 500 participants based on Kyriazos' study (20), who recommends the following guidelines for a stable dimension reduction: 50 (very poor), 100 (poor), 200 (fair), 300 (good), 500 (very good), 1,000 or more (excellent).

We justify aiming for a very good sample size because of three reasons. First, a very good sample size would support successful validation of the prediction due to consistent results. Second, we know from a previous study that most Malaysian interns complete their internship on time (21). Hence, this sample size accounted for a potentially small number of interns who would have failed a clinical posting within a year. Finally, the practical and resource constraints limited the feasibility of recruiting an excellent sample size of 1,000 or higher. Thus, correcting for a potential 20% attrition rate, the final minimum size of recruitment was aimed at ${}^{500}/_{0.8} = 625$. After one year, we followed up with all participants to check if they had completed all their clinical postings on time for the year.

Analyses

Linear discriminant analyses (LDA) – Variable selection and prediction modelling

This study uses LDA as a variable-selection method based on several statistical and methodological considerations that align with the research objectives. When all predictors are numerical, LDA is described by Tabachnick and Fidell (22) as "MANOVA (multivariate analysis of variance) turned around" (p. 377). Thus, LDA leverages the same robust statistical parameters such as F statistics, R^2, and Wilks' Lambda multivariate statistics to determine which predictors are statistically significant and contribute to better class separation at high dimensional space (23, p. 246). Meanwhile, LDA maximises group separability in high-dimensional space as a classifier by discriminating eigenvectors (24). It reduces dimensions by linearly combining predictors via maximising interclass variance-covariance correlation and minimising intraclass variance-covariance correlation, resulting in perfect or near-perfect user-defined group classification (25). The performance of LDA as a model to predict group membership can also be statistically evaluated. These dual benefits make LDA particularly suitable for distinguishing between medical interns who are at risk of failing a clinical posting and those who are not, based on statistically significant predictors.

Moreover, LDA is relatively robust against violations of statistical assumptions typically required in MANOVA, such as multivariate normality and homogeneity of variance-covariance matrices (22, p. 383). The relaxed statistical assumption makes LDA exceptionally suited for our study. LDA's flexibility regarding sample size and design requirements allows for a more inclusive and comprehensive approach to variable selection and classification in high-dimensional datasets. Therefore, LDA serves as a robust statistical analysis for both variable-selection exploratory analysis and classification prediction, which addresses the goal of this research (26).

Training and validation dataset

The train-test split validation is an established data science approach to evaluate the degree of overfitting and, thus, the generalisability of a prediction model (27). This study randomly split the overall dataset into 70% training and 30% validation datasets. The training dataset contained labelled outcomes and was used for exploratory and predictive analyses. The prediction model was then tested using leave-one-out cross-validation on all observations in the training dataset: this statistical technique assessed model overfitting and the generalisability of the classification model (28). In comparison, the validation dataset provided an independent evaluation of the model's overfitting and generalisability by testing it on a new data set it had not encountered before.

Predictor selection

Predictors were added to the model based on *F* statistics, which tested the equality of means between classes for each explanatory variable. The predictors were added in sequence according to their contribution, using a significance threshold of 0.05 (26). The impact of removing the newly added predictor was assessed using Wilk's lambda multivariate statistics starting from third predictor. If the probability of the calculated statistic was higher than the removal threshold of 0.10, the predictor was removed from the model.

Weight correction

We expected an imbalanced dataset due to a prior study that reported most interns in Malaysia managed to complete their internship in due time (21). An imbalanced dataset would undermine the model to train in favour of the majority class. Where W_{ci} was the corrected weight of class *i*, *n* was the total number total samples, and n_1 was the number of samples for class *i*, the formula $W_{ci} = {}^{n}/_{2n_1}$ was applied to balance the total weight of each class (29). The weight correction would penalise the model for misclassifying observations which belonged to the minority class at a higher magnitude, proportionate to the original imbalanced distribution of class frequency.

Data pre-processing

The PHPQ questionnaire assessed medical graduates' preparedness for internship across nine subscales: coping, holistic, ethics, basic skills, patient management, interpersonal skills, information technology skills, clinical skills, and scientific knowledge. The validated Brief-COPE questionnaire measured three types of coping: avoidant, problem-focused and emotion-focused. We included these latent constructs as individual predictors in our analysis to retain discriminating information. Additionally, we used one-hot encoding to transform categorical data, enabling robust training of the discriminant algorithm at higher dimensions.

Software

All statistical and ML analyses were computed using XLSTAT software version 2022.4.1 (Addinsoft, New York, USA).

RESULTS

The Overall Recruitment

A total of 671 newly enrolled interns were identified and invited to participate, and 524 of them completed the initial survey. After one year, another 28 interns were lost to follow-up, leaving a total of 496 interns or a 73.9% participation rate. All participants graduated from 71 medical schools. Additionally, 33 (6.6%) interns reported they had failed a clinical posting after one year of practice. Tables 1a and 1b show the descriptive statistics of qualitative and quantitative predictors among all participants.

Table 1a: Descriptive statistics of qualitative predictors among 496 participants

Predictors	Categories	Frequency	%
Gender	Female	330	66.5
	Male	166	33.5
Ethnicity	Malay	346	69.8
	Chinese	72	14.5
	India	67	13.5
	Others	11	2.2
Currently married	Yes	73	14.7
	No	423	85.3
Fail a clinical posting during medical school	Yes	123	24.8
	No	373	75.2
Attend preparatory course	Yes	266	53.6
	No	230	46.4
Hospital	Tertiary service	300	60.5
	Non-tertiary service	154	31.0
	Tertiary teaching	42	8.5
Medical school by country	Malaysia	218	44.0
	Egypt	147	29.6
	Indonesia	35	7.1
	Russia	33	6.7
	Jordan	28	5.6
	India	16	3.2
	Ireland	6	1.2
	Romania	3	0.6
	United Kingdom	3	0.6
	Czech Republic	2	0.4
	Others	5	1.0

Table 1b: Descriptive statistics of quantitative predictors among 496 participants

Quantitative predictors	Max	Min	Mean	Standard deviation	95% confidence interval
Interest	1	5	4.17	0.84	0.63–0.81
Age	24	35	26.37	1.45	1.87–2.41
Duration_waiting	2	77	11.90	7.16	45.38–58.24
COPING_Avoidant	8	32	15.67	3.87	13.26–17.01
COPING_Emotion_Focus	12	48	33.91	6.32	35.36–45.37
COPING_Problem_Focus	8	32	25.75	4.14	15.21–19.52
RESILIENCE	9	40	28.61	6.25	34.67–44.49
EQ	12	52	40.09	7.57	50.84–65.24
RELIGIOSITY	5	27	21.68	4.67	19.36–24.84
PREP_Coping	5	20	14.79	2.83	7.09–9.09
PREP_Holistic	12	30	23.89	3.64	11.74–15.07
PREP_Ethics	2	10	7.99	1.63	2.35–3.02
PREP_Basic_skills	10	35	28.17	4.14	15.18–19.49
PREP_Patient_Management	3	15	10.46	2.45	5.33–6.84
PREP_IP_Skills	8	35	23.84	5.14	23.4–30.10
PREP_IT_Skills	7	35	25.65	5.29	24.83–31.86
PREP_Clinical_Skills	3	15	9.88	2.78	6.85–8.79
PREP_Scientific_Knowledge	4	20	14.61	2.69	6.39–8.20

Notes: EQ = emotional quotient; PREP = preparation for hospital practice; IP = interpersonal; IT = information technology

Dataset

The training dataset consisted of 346 (69.8%) records that had been randomly selected from the original dataset. Out of the 346, a total of 25 (7.2%) interns had failed a clinical posting. The remaining records were assigned to the validation dataset that featured 150 (30.2%) observations. From the 150, there were 8 (5.3%) interns who had failed a clinical posting.

Predictors Selection and LDA Classification Model

Table 2 demonstrates a selection of 27 predictors from the training dataset and the associated statistics of significance. Meanwhile, Table 3 shows the classification function and standardised discriminant coefficient. The former enabled direct calculations from observed values of predictors to predict whether an intern would fail a clinical posting. Meanwhile, the latter allowed a comparison of contributions among these predictors towards the LDA classification model.

Table 2: Predictors' selection via the forward stepwise method

Number of steps	Number of predictors in the model	Predictor IN/OUT	Status	Partial R^2	*F*	Pr > *F*	Wilks' Lambda	Pr < Lambda
1	1	COPING_Avoidant	IN				0.924	< 0.001
2	2	Hosp_2	IN	0.082	30.51	< 0.001	0.848	< 0.001
3	3	Hosp_1	IN	0.073	26.86	< 0.001	0.787	< 0.001
4	4	Med_School_51	IN	0.070	25.85	< 0.001	0.731	< 0.001
5	5	PREP_IP_Skills	IN	0.077	28.25	< 0.001	0.675	< 0.001
6	6	Hosp_8	IN	0.064	23.05	< 0.001	0.632	< 0.001
7	7	Hosp_16	IN	0.077	28.33	< 0.001	0.583	< 0.001
8	8	Med_School_43	IN	0.033	11.46	< 0.001	0.564	< 0.001
9	9	Hosp_15	IN	0.037	13.01	< 0.001	0.543	< 0.001
10	10	Hosp_7	IN	0.034	11.64	< 0.001	0.525	< 0.001
11	11	Hosp_11	IN	0.038	13.10	< 0.001	0.505	< 0.001
12	12	Med_School_6	IN	0.030	10.46	< 0.050	0.490	< 0.001
13	13	Med_School_58	IN	0.028	9.43	< 0.050	0.476	< 0.001
14	14	Med_School_65	IN	0.033	11.31	< 0.001	0.460	< 0.001
15	15	Med_School_62	IN	0.028	9.35	< 0.050	0.448	< 0.001
16	16	Marriage_status	IN	0.021	7.01	< 0.050	0.438	< 0.001
17	17	Med_School_35	IN	0.022	7.21	< 0.050	0.429	< 0.001
18	18	Med_School_9	IN	0.029	9.89	< 0.050	0.416	< 0.001
19	19	Med_School_7	IN	0.025	8.44	< 0.050	0.406	< 0.001
20	20	Med_School_54	IN	0.025	8.40	< 0.050	0.396	< 0.001
21	21	Med_School_66	IN	0.024	8.00	< 0.050	0.386	< 0.001
22	22	Med_School_5	IN	0.027	8.87	< 0.050	0.376	< 0.001
23	23	Med_School_4	IN	0.030	10.06	< 0.050	0.364	< 0.001
24	24	Med_School_22	IN	0.034	11.45	< 0.001	0.352	< 0.001
25	23	Med_School_6	OUT	0.007	2.32	0.130	0.354	< 0.001
26	25	PREP_IT_Skills	IN	0.034	11.42	< 0.001	0.342	< 0.001
27	26	Med_School_47	IN	0.026	8.66	< 0.050	0.333	< 0.001
28	27	Duration_waiting	IN	0.015	4.88	< 0.050	0.328	< 0.001

Table 3: LDA classification function and the standardised canonical discriminant function coefficients

Category	Predictors	LDA classification function[a]		Standardised canonical discriminant coefficient[b]
		Yes	No	F1
Place of practice	Hosp_1	0.46	−4.62	0.68
	Hosp_2	4.93	1.20	0.50
	Hosp_8	−5.29	−9.77	0.43
	Hosp_16	11.81	6.55	0.41
	Hosp_15	9.81	3.48	0.40
	Hosp_7	7.23	3.24	0.30
	Hosp_4	4.65	2.23	0.22
	Hosp_11	10.88	7.67	0.21
Place of study	Med_School_51	16.58	6.60	0.72
	Med_School_62	−15.22	−28.46	0.69
	Med_School_65	2.05	−11.19	0.69
	Med_School_58	8.23	−3.89	0.58
	Med_School_5	1.63	−3.42	0.50
	Med_School_66	14.77	7.87	0.50
	Med_School_7	22.29	16.63	0.49
	Med_School_35	9.99	4.00	0.48
	Med_School_9	15.73	7.45	0.46
	Med_School_4	7.10	3.75	0.42
	Med_School_22	11.29	6.06	0.31
	Med_School_47	−0.02	−3.93	0.26
	Med_School_54	2.97	−1.23	0.25
	Med_School_43	2.30	6.28	−0.19
Coping style	COPING_Avoidant	1.82	1.48	0.48
Preparedness	PREP_IT_Skills	1.56	1.39	0.29
	PREP_IP_Skills	−0.20	0.17	−0.62
Personal profile	Marriage_status	3.52	1.80	0.24
Months to start internship	Duration_waiting	0.84	0.95	−0.35
	Intercept	−45.68	−38.17	

Notes: [a]The higher value after totalling all predictors' scores times their weights will predict an intern either fail a clinical posting (Yes) or not (No); [b]Standardised comparison of contribution towards prediction

Performance Metrics of the Classification Model

The standardised canonical discriminant function analysis (Table 3) identified place of practice (8 out of 26 hospitals), place of study (13 out of 71 schools), being married, avoidant coping, and good IT skills as significant predictors of failing a clinical posting. Conversely, passing all internship clinical postings could be positively predicted by graduating from one medical school (Med_School_43), higher preparedness in interpersonal skills, and shorter duration between graduation and internship commencement.

The training dataset confusion matrix (Table 4) showed that the prediction model achieved 100.0% sensitivity, 85.7% specificity, and 86.7% accuracy. The following receiver operating characteristics (ROC) curve (Figure 1) indicated an excellent classification model with an area under the curve (AUC) of 0.969.

Table 4: Confusion matrices

			Predict			
			Yes	No	Total	Correct
Training dataset	Obs	Yes	25	0	25	100.00%
		No	46	275	321	85.70%
		Total	71	275	346	
			Predict			
			Yes	No	Total	Correct
Cross-validation of the training dataset	Obs	Yes	23	2	25	92.00%
		No	55	266	321	82.90%
		Total	78	268	346	
			Predict			
			Yes	No	Total	Correct
Validation dataset	Obs	Yes	6	2	8	75.00%
		No	27	115	142	81.00%
		Total	33	117	150	

Note: Yes = Fail a clinical posting.

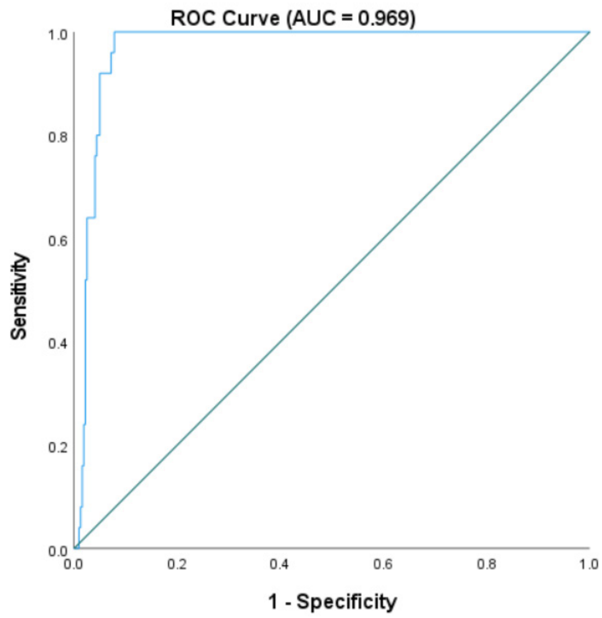

Figure 1: ROC with AUC of the classification model for the training dataset (346 interns with 25 failed clinical postings).

Validation of Prediction

Cross-validation of the training dataset

The cross-validation results (Table 4) indicated a mild and acceptable degree of overfitting, with slightly reduced sensitivity (92.0%) and specificity (82.9%), leading to an overall accuracy of 83.5%. Figure 2, meanwhile, demonstrated an AUC of 0.905.

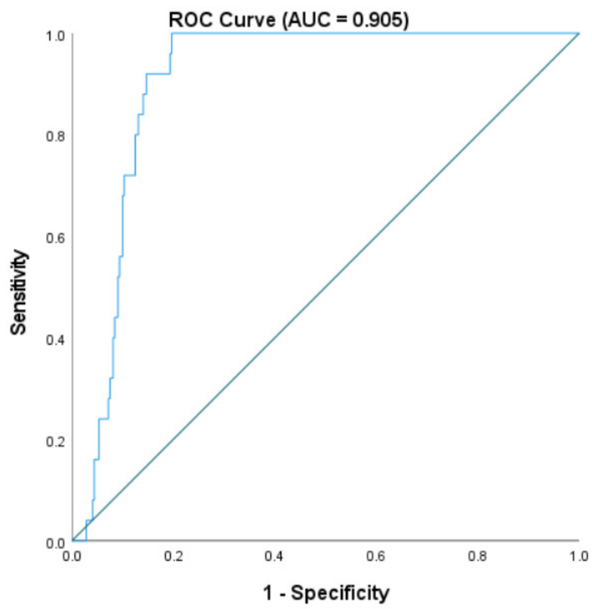

Figure 2: ROC with AUC of cross-validation of the training dataset (346 interns with 25 failed clinical postings).

Validation dataset

These findings were consistent with the classification performance of the model on a new set of observations from the validation dataset (Table 4). Despite the imbalanced class frequency, the model correctly predicted 6 out of 8 interns who failed a clinical posting (75.0% sensitivity) and misclassified only 27 out of 142 interns who passed (81.0% specificity), resulting in an overall accuracy of 80.7%. Figure 3 shows an AUC of 0.861.

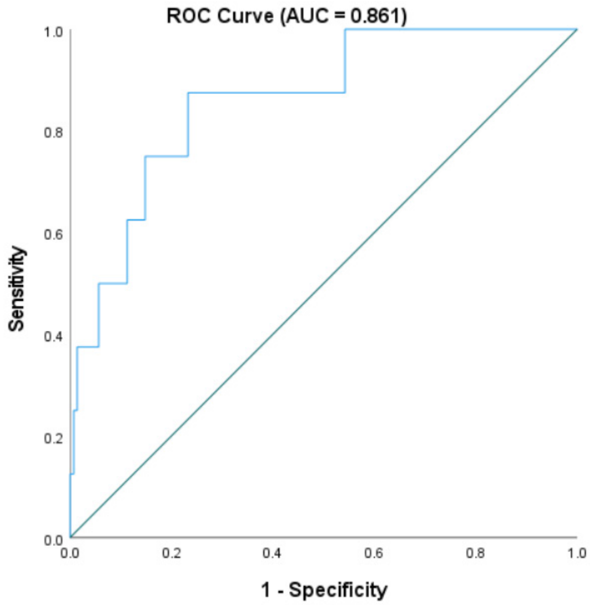

Figure 3: ROC with AUC of the validation dataset (150 interns with 8 failed clinical postings).

Comparison of performance evaluation

Table 5 summarises the comparison of model performance across the training dataset, cross-validation of the training dataset, and validation dataset. These include the model's ability to correctly identify positive cases (sensitivity), correctly identify negative cases (specificity), overall accuracy, precision (the proportion of positive identifications that were actually correct), and the F1 score (a harmonic mean of precision and sensitivity, providing a balance between the two) (30).

Table 5: Performance evaluation of LDA prediction

Performance metrics	Training dataset	Cross-validation of the training dataset	Validation dataset
No. of observations	346	346	150
No. of Interns failed a clinical posting	25	25	8
Sensitivity (%)	100.0	92.0	75.0
Specificity (%)	85.7	82.9	81.0
Accuracy (%)	86.7	83.5	80.7
Precision (%)	35.2	29.5	18.2
F1 score (%)	52.1	44.7	29.3
AUC	0.969	0.905	0.861

DISCUSSION

Novel Contribution

To our knowledge, this is the first nationwide prospective study which successfully demonstrates that an early cross-sectional survey of personal attributes, place of study, and place of practice does bring adequate and valuable information to predict medical interns' failing a clinical posting after one year. Unlike the cross-sectional studies, the longitudinal design of this prospective study allows a prediction that can potentially be conducted before or during the early phase of training. Thus, the results can uniquely contribute to interns' selection and training.

The high-stake decision on interns' selection has motivated Filiberto et al. (31) to survey 521 interns who are graduates of the University of Florida College of Medicine between 2015 and 2018. The single-centre study proves significant associations between clinical performance and scores from the licensing exam in the United States, medical schools and class ranks. Our study supports their findings by showing that interns' place of study does potentially contribute significant information to discriminate interns' clinical performance. Additionally, this multi-centre study extends the understanding further by showing the magnitude of this contribution to prediction is not uniform. Thus, whenever selections involve applicants from a variety of medical schools, scores from the licensing exam or assessments in a standardised setting may prove more valuable to guide decision-making.

This study may also be the first to establish direct roles in policy decision-making with interns' failure in a clinical posting. The lack of significant association between most hospitals and medical schools with poor clinical performance suggests these institutions may have been adequately preparing their graduates and trainees for the internship training. Furthermore, there is one medical school whose graduates are significantly less likely to fail a clinical posting. This evidence communicates accountability that extends beyond individual interns, educators and supervising clinicians. As a result, for the 8 hospitals and 13 medical schools that are significantly associated with interns' poor clinical performance, the results of this study signify mandates on the volume of resource investment for research, education, and broader policy commitments for quality assurance and quality improvement.

Explaining Significant Predictors

These commitments consequently may also manifest as strategies to address avoidant coping styles and to improve interpersonal skills, the two modifiable predictors with the largest contribution. Supporting this call is a study by Ismail et al. (32), who surveyed 431 medical interns in Malaysia to show psychological stress, anxiety and depression have a strong correlation with avoidant coping. On the other hand, an intuitive qualitative study among 18 final-year students and 15 medical interns in Saudi Arabia revealed that participants were defining an unhealthy avoidant copy style, including not wanting to engage in discussions on topics related to medicine (33). Departing from a different perspective, Cruess et al. (34) advocated the concept of identity formation through social learning, where medical trainees must participate in interpersonal engagement to move from legitimate peripheral to full membership of the community of practice. This evidence and theory illuminate the intimate association between avoidant coping and interpersonal skills, yet their contrasting contribution to predicting interns' clinical performance. Therefore, instead of adopting reductionist strategies, we believe addressing both significant predictors may be better served by capitalising on their intrinsic connection via practical but integrated interventions to command strong buy-ins among all relevant stakeholders.

Subsequently, the landmark expectancy theory by Maslow (35) may prove useful in understanding the motivation among medical interns as the main stakeholders. The longer period of unemployment since graduation may have better-motivated medical interns to persevere the physical, psychological and emotional challenges to meet the fundamental needs for safety and security (35). In particular, the economic crisis due to COVID-19 during the conduct of this study may have driven this motivation further. Likewise, the booming of digital industries during the pandemic may have provided a wider career option for interns with good IT skills should the challenges of internship training outweigh the motivation on needs for esteem (35). Similarly, being currently married suggests a competing need for love and belonging that take precedence in the hierarchy of needs (36). Hence, assigning priorities to the family over the career is not an unlikely situation. Therefore, we posit these predictors may represent proxy measures of a common construct of motivation that influences judgement on weighing commitments between personal and professional needs.

The Advantages of a Linear Model

The ability to interpret these findings reflects our choice of employing LDA to leverage the advantages of general linear modelling—it is straightforward for clinicians and policymakers to understand and appreciate. This clarity is crucial, especially when dealing with complex modelling attempting to establish relationships across numerous predictors. The complexity of an advanced ML model for a high-dimensional dataset can often exceed human comprehension. Examples of these models include deep learning, ensemble methods such as XGBoost, other gradient-boosting approaches, and random forest modelling. These models and other advanced ML models often face the "black box" problem, where their complexity obscures the internal workings and information processing from human understanding, reducing these models to mere inputs and outputs (36). Fomin (37) insightfully highlights the challenges posed by this issue. Firstly, human trust in a tool is inherently tied to understanding its processes. Secondly, the inability to interpret these models' predictions hinders human control and oversight. Finally, the lack of transparency raises significant concerns about liability and accountability when erroneous predictions lead to wrong decisions, potentially resulting in complex legal ramifications (37). Therefore, it is crucial to recognise the hidden costs of the black box problem when using advanced ML models to achieve superior accuracy of interns' competency prediction.

Limitations

There were several limitations of this study. Firstly, this study is conducted nationwide in the context of medical internship training in Malaysia. Given the multi-dimensional aspects of workplace-based competency, it is possible that different cultures, policies, and overall medical internship training at the national level may have a significant influence on early predictors of internship competency. Therefore, our interpretation of the roles of significant predictors of this study is limited within the context of Malaysian internship training. However, given the substantial overlap of core competencies between medical internships across different countries, we believe the findings of this study may still prove meaningful in a wider context of postgraduate medical training.

Secondly, we acknowledge competency is a multifaceted construct influenced by various interconnected factors that can change over time. Therefore, although this study does establish predictors which are significant for early prediction, it does not discount the significance, roles, and relevance of other predictors when the interns are actually going through their training. These include factors such as resilience and emotional intelligence, which may prove essential as determinants of successful completion of medical internship training. Ideally, these predictors should be assessed throughout training using multiple sources to provide a more comprehensive understanding of their relationship with competency development over time. However, implementing such assessments is logistically challenging and resource-intensive, making it impractical for the scope of this study. Additionally, resource and time constraints posed challenges in accessing complete manual records of year-long competency assessments from all participating hospitals. These limitations restricted the incorporation of a comprehensive longitudinal assessment, which could have provided further insights into the relationship between predictors and outcomes in this study.

Recommendation

Therefore, we recommend that future studies replicate this study in various countries and regions to assess the generalisability of early predictions of internship competency more broadly. Richer data and detailed comparative analyses across diverse practice contexts could strengthen the evidence for the generalisability of conclusions regarding early predictors of medical internship success. Qualitative research could deepen our understanding of the motivations of interns, the perspectives of other stakeholders, and their relationships to the significant predictors identified in this study. Exploratory analyses could lead to a new theoretical model outlining mechanisms and strategies enabling interns to successfully complete all clinical postings. This understanding could also illuminate strategies within the interns' capacity and define the specific responsibilities of policymakers and educators in enhancing the quality of internship training.

CONCLUSION

In conclusion, this study brings empirical and validated evidence to show that surveying medical interns for a place of study, place of practice, avoidant coping style, IT skills, interpersonal skills, marital status, and duration between graduation and commencing training do bring adequate discriminatory information for high-performance prediction on competency within one year of training. The supervised ML framework using LDA offers four crucial advantages. Firstly, the choice of LDA over more advanced and complex ML modelling removes the intricate dilemmas associated with ML black box problems. Secondly, the survey allows for comprehensive multi-centre data collection, capturing sufficient discriminatory information likely representative of nationwide internship training in Malaysia. Thirdly, the complex, high-dimensional dataset from multiple centres leverages advanced data pre-processing techniques from the data science field to identify significant factors for early prediction. Finally, the framework employs an established ML approach to validation. This validation empirically demonstrates that an early cross-sectional survey, despite its inherent limitations, can predict competency within a year. Thus, the empirical proof removes uncertainties about the generalisability of the conclusions, which is not uncommon to be questioned or challenged when inferentially deduced.

Consequently, the validated findings of this study offer valuable practical implications, guiding evidence-informed strategies and resource allocation for quality medical intern training. Clinicians, educators, and policymakers can leverage these novel findings to formulate timely and targeted interventions, ensuring effective and efficient use of resources for quality improvement. Additionally, further research is needed to understand why only certain predictors are significant for early predictions. We recommend qualitative studies for an in-depth understanding of the quantitative evidence to achieve this goal. Finally, replicating this study in various countries and regions to assess the generalisability of early predictions of internship competency would likely secure broader buy-in for high-stakes decision-making in a wider context.

FUNDING

The authors would like to acknowledge the Ministry of Higher Education Malaysia for funding this research under the Fundamental Research Grant Scheme for Research Acculturation of Early Career Researchers (FRGS-RACER) [RACER/1/2019/SKK01/UNISZA//1].

ETHICAL APPROVAL

This study received research ethics approval from the National Medical Research Register with approval ID NMRR-19-3499-51724 (IIR) and Universiti Sultan Zainal Abidin Human Research Ethics Committee (UHREC) with approval ID UniSZA.C/2/UHREC/628-2Jld 3(4). Additionally, approvals had also been elicited from directors of all hospitals involved in this study.

ACKNOWLEDGEMENTS

We would like to thank all participating medical interns for their volunteer participation and the directors of all participating hospitals for approving the conduct of this study on their premises.

REFERENCES

1. Ten Cate O, Billett S. Competency-based medical education: origins, perspectives and potentialities. Med Educ. 2014;48(3):325–32. https://doi.org/10.1111/medu.12355
2. Hawkins RE, Welcher CM, Holmboe ES, Kirk LM, Norcini JJ, Simons KB, et al. Implementation of competency-based medical education: are we addressing the concerns and challenges? Med Educ. 2015;49(11):1086–102. https://doi.org/10.1111/medu.12831
3. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65(9 Suppl):S63–7. https://doi.org/10.1097/00001888-199009000-00045
4. Khan R, Hodges BD, Martimianakis MA. When I say … burnout. Med Educ. 2023;57(8):704–5. https://doi.org/10.1111/medu.15088
5. Kreitzer MJ, Klatt M. Educational innovations to foster resilience in the health professions. Med Teach. 2017;39(2):153–9. https://doi.org/10.1080/0142159X.2016.1248917
6. Saleem MS, Isha ASN, Mohd Yusop Y, Awan MI, Naji GMA. Agility and safety performance among nurses: the mediating role of mindful organizing. Nurs Rep. 2021;11(3):666–79. https://doi.org/10.3390/nursrep11030063

7. Daud N, Fuad A, Mat Pa M, Ahmad A, Yusof N, Mat Hassan N, et al. Emotional intelligence among medical students and its relationship with burnout. Educ Med J. 2022;14:49–59. https://doi.org/10.21315/eimj2022.14.3.4

8. Liu C-H, Tang W-R, Weng W-H, Lin Y-H, Chen C-Y. The process of coping with stress by Taiwanese medical interns: a qualitative study. BMC Med Educ. 2016;16(1):10. https://doi.org/10.1186/s12909-016-0534-3

9. DeWitt D, Canny BJ, Nitzberg M, Choudri J, Porter S. Medical student satisfaction, coping and burnout in direct-entry versus graduate-entry programmes. Med Educ. 2016;50(6):637–45. https://doi.org/10.1111/medu.12971

10. Daud N, Mat Pa M, Fuad A, Ahmad A, Mat Hassan N. Academic factors associated with burnout in Malaysian medical students: a cross-sectional study. Educ Med J. 2020;12:49–58. https://doi.org/10.21315/eimj2020.12.2.5

11. Daud N, Ahmad A, Fuad A, Mat Pa M, Mat Hassan N. Factors associated with burnout in clinical year medical students in Terengganu, Malaysia. Asian J Med Biomed. 2021;5:1–9. https://doi.org/10.37231/ajmb.2021.5.1.336

12. Manaf N, Ghazali R, Borhan N, Omar A, Mohd D, Adewale A, et al. Preparedness for hospital practice in assuring quality of care. TQM J. 2016;28:834–46. https://doi.org/10.1108/TQM-06-2014-0053

13. Haghnegahdar M, Sharma P, Hubbard KP, White WA. The influence of religious belief on burnout in medical students. Missouri Med. 2021;118(1):63–7.

14. Klein R, Julian K, Snyder E, Koch J, Ufere N, Volerman A, et al. Gender bias in resident assessment in graduate medical education: review of the literature. J Gen Intern Med. 2019;34:712–9. https://doi.org/10.1007/s11606-019-04884-0

15. Sitobata M, Mohammadnezhad M. Transitional challenges faced by medical intern doctors (IDs) in Vanuatu: a qualitative study. Med Educ Online. 2022;27(1):2005458. https://doi.org/10.1080/10872981.2021.2005458

16. Campbell-Sills L, Stein MB. Psychometric analysis and refinement of the Connor-Davidson Resilience Scale (CD-RISC): validation of a 10-item measure of resilience. J Trauma Stress. 2007;20(6):1019–28. https://doi.org/10.1002/jts.20271

17. Dias C, Cruz JFA, Fonseca AM. The relationship between multidimensional competitive anxiety, cognitive threat appraisal, and coping strategies: a multi-sport study. Int J Sport Exerc Psychol. 2012;10(1):52–65. https://doi.org/10.1080/1612197X.2012.645131

18. Arifin WN, Yusoff MSB. Confirmatory factor analysis of the Universiti Sains Malaysia emotional quotient inventory among medical students in Malaysia. SAGE Open. 2016;6(2):2158244016650240. https://doi.org/10.1177/2158244016650240

19. Nurasikin M, Ahmad Adlan A, Cg N. Validity and reliability of the Malay version of Duke University Religion Index (DUREL-M) among a group of nursing students. Malays J Psychiatry. 2010;19(2):68–72.

20. Kyriazos T. Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. Psychology. 2018;09(8):2207–30. https://doi.org/10.4236/psych.2018.98126

21. Han LC. Housemanship programme in Malaysia: availability of positions and quality of training. Penang, Malaysia: Penang Institute; 2017.

22. Tabachnick BG, Fidell LS. Using multivariate statistics. UK: Pearson Education; 2013.

23. Addinsoft. XLSTAT official guide 2022 USA: Addinsoft; 2022; [cited 2024 September 19]. Available from: https://cdn.xlstat.com/helpcentersfdc/XLSTAT_Help.pdf

24. Hassan WM, Al-Dbass A, Al-Ayadhi L, Bhat RS, El-Ansary A. Discriminant analysis and binary logistic regression enable more accurate prediction of autism spectrum disorder than principal component analysis. Sci Rep. 2022;12(1):3764. https://doi.org/10.1038/s41598-022-07829-6

25. Ismail A, Ibrahim MS, Ismail S, Aziz AA, Yusoff HM, Mokhtar M, et al. Development of COVID-19 Health-Risk Assessment and Self-Evaluation (CHaSe): a health screening system for university students and staff during the movement control order (MCO). Netw Model Anal Health Inform BioinformaNetwork Modeling Analysis in Health Informatics and Bioinformatics. 2022;11(1):21. https://doi.org/10.1007/s13721-022-00357-3

26. Addinsoft. Discriminant Analysis (DA) United States of America: Addinsoft; 2022 [cited 2024 September 19]. Available from: https://www.xlstat.com/en/solutions/features/discriminant-analysis-da

27. Xu Y, Goodacre R. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. J Anal Test. 2018;2. https://doi.org/10.1007/s41664-018-0068-2

28. Berrar D. Cross-validation. Encycl Bioinform Comput Biol. 2018;1:542–5. https://doi.org/10.1016/B978-0-12-809633-8.20349-X

29. Singh K. How to improve class imbalance using class weights in machine learning India: analytics Vidhya; 2020 [cited 2024 September 19]. Available from: https://www.analyticsvidhya.com/blog/2020/10/improve-class-imbalance-class-weights

30. Karimi Z. Confusion Matrix. 2021 [cited 2024 September 19]. Available from: https://www.researchgate.net/publication/355096788_Confusion_Matrix

31. Filiberto AC, Cooper LA, Loftus TJ, Samant SS, Sarosi GA, Jr., Tan SA. Objective predictors of intern performance. BMC Med Educ. 2021;21(1):77. https://doi.org/10.1186/s12909-021-02487-0

32. Ismail M, Lee KY, Sutrisno Tanjung A, Ahmad Jelani IA, Abdul Latiff R, Abdul Razak H, et al. The prevalence of psychological distress and its association with coping strategies among medical interns in Malaysia: a national-level cross-sectional study. Asia Pac Psychiatry. 2020;13(2):e12417. https://doi.org/10.1111/appy.12417

33. Abouammoh N, Irfan F, AlFaris E. Stress coping strategies among medical students and trainees in Saudi Arabia: a qualitative study. BMC Med Educ. 2020;20(1):124. https://doi.org/10.1186/s12909-020-02039-y

34. Cruess RL, Cruess SR, Steinert Y. Medicine as a community of practice: implications for medical education. Acad Med. 2018;93(2):185–91. https://doi.org/10.1097/ACM.0000000000001826

35. Maslow AH. A theory of human motivation. Psychol Rev. 1943;50(4):370–96. https://doi.org/10.1037/h0054346

36. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K, et al. Interpreting black-box models: a review on explainable artificial intelligence. Cogn Comput. 2024;16(1):45–74. https://doi.org/10.1007/s12559-023-10179-8

37. Fomin V, Astromskis P. The black box problem. In: John-Stewart Gordon, editor. Future law, ethics, and smart technologies. Leiden: Brill; 20232. p. 112–25. https://doi.org/10.1163/9789004682900_012