Predicting Attrition among Software Professionals: Antecedents and Consequences of Burnout and Engagement

BIANCA TRINKENREICH and FABIO SANTOS, Colorado State University, USA
KLAAS-JAN STOL, Lero, Ireland and University College Cork, Ireland

In this study of burnout and engagement, we address three major themes. First, we offer a review of prior studies of burnout among IT professionals and link these studies to the Job Demands-Resources (JD-R) model. Informed by the JD-R model, we identify three factors that are organizational job resources, and posit that these (a) increase engagement, and (b) decrease burnout. Second, we extend the JD-R by considering software professionals' intention to stay as a consequence of these two affective states, burnout and engagement. Third, we focus on the importance of factors for intention to stay, and actual retention behavior. We use a unique dataset of over 13,000 respondents at one global IT organization, enriched with employment status 90 days after the initial survey. Leveraging partial least squares structural equation modeling and machine learning, we find that the data mostly support our theoretical model, with some variation across different subgroups of respondents. An importance-performance map analysis suggests that managers may wish to focus on interventions regarding burnout as a predictor of intention to leave. The Machine Learning model suggests that engagement and opportunities to learn are the top two most important factors that explain whether software professionals leave an organization.

CCS Concepts: • **Social and professional topics** → *Computing organizations*; *Professional topics*; • **Human-centered computing** → **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: Organizational leadership, Leadership support, Learning, burnout, engagement, culture, attrition

ACM Reference Format:
Bianca Trinkenreich, Fabio Santos, and Klaas-Jan Stol. 2024. Predicting Attrition among Software Professionals: Antecedents and Consequences of Burnout and Engagement. *ACM Trans. Softw. Eng. Methodol.* 1, 1 (August 2024), 46 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Work-related anxiety and mental disorders are common in the IT sector [77]. Software developers are more likely to feel fatigued, anxious, experience burnout, and stressed than those who perform mechanical tasks [77]. Early work by Glass et al. demonstrated that software development tasks have high intellectual demands [28]. Deteriorating mental health threatens workers' well-being which may lead to an increase in attrition. This in turn can lead to reduced productivity of an organization due to disruption of ongoing work and costs involved in recruiting and onboarding new employees [103]. Apart from concerns for the overall well-being of an organization's staff, the negative outcomes for an organization, such as attrition, should be a major managerial concern.

Authors' addresses: Bianca Trinkenreich, bianca.trinkenreich@colostate.edu; Fabio Santos, fabio.deabreusantos@colostate.edu, Colorado State University, Fort Collins, CO, USA; Klaas-Jan Stol, Lero, Cork, Ireland and University College Cork, Cork, Ireland, k.stol@ucc.ie.

1049-331X/2024/8-ART $15.00
https://doi.org/XXXXXXX.XXXXXXX

Catalyzed by the Covid-19 pandemic, professionals worldwide have started to reflect on their work-life balance, leading to a trend that some have termed the "Great Resignation" [20]. Retaining software professionals, who can work remotely quite easily due to the nature of the job, can be challenging. To retain staff, organizations must create a work environment to allow their workforce to flourish, to increase staff engagement without inducing high levels of stress and burnout [30]. Organizations that invest in the health and safety of their workforce benefit from this in terms of organizational commitment and retention among employees [72]; studies have suggested a return on investment of up to 200% [83, 118].

Two important psychological states in relation to workplace wellbeing are burnout and engagement [15, 65]. Burnout refers to an individual's experiences of exhaustion on physical, emotional, and cognitive levels [84]. Research in other disciplines suggests that burnout is associated with employees' dissatisfaction and intention to leave their job [120]. There are numerous studies of burnout of software professionals [117], emphasizing its importance to software organizations. Engagement, however, has not been the focus of much research in the software engineering literature. While burnout is a result of prolonged stressors, engagement has been defined as *"a positive, fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption"* [65]. Engagement is a desired goal for burnout interventions [69]. Rather than merely the opposite of burnout, or a momentary and specific state, engagement is a positive affective-motivational state of fulfillment in employees [99], characterized by individual perceptions of energy, effectiveness, and motivation at work [100]. Engagement and burnout are thus not opposites but are distinct, yet closely related concepts that require separate measurement [113]. The positive psychology movement shows good health is a great deal but is also relevant to keeping away disease [53, 98].

An influential theory of burnout is the Job Demands-Resources (JD-R) model, which posits that burnout is the result of two processes [15]. The first process is a continuous overtaxing of employees leading to exhaustion. The second, parallel, process is a continuous lacking of resources for employees to do their job, leading to a level of disengagement. Whereas prior work in this area has considered a range of antecedents of burnout, most of these can be characterized as *job demands* [15], i.e., the demands that are made of an employee as part of their job: job overload, work stress, role conflict, and so on. Several, but fewer, studies have focused on job resources that lead to disengagement.

In this article, we build on the JD-R model as follows. First, we identify three organizational job resources (discussed in more detail in Sec. 2): leadership support, organizational culture, and opportunities to learn, and posit that these are (a) positive antecedents of engagement, and (b) negative antecedents of burnout. Second, we consider the role of both engagement and burnout as antecedents to people's intention to stay. Hence, we started this investigation guided by the following research question:

RQ1. How are leadership support, organizational culture, and opportunities to learn related to burnout, engagement, and intention to stay?

To answer RQ1, we develop a theoretical model that proposes leadership support, organizational culture, and opportunities to learn as antecedents of burnout and engagement, and intention to stay as a consequence of burnout and engagement. We tested this model using Partial-Least Squares Structural Equation Modeling (PLS-SEM) with a dataset of over 13,000 responses. The responses were gathered within a single but globally distributed organization, varying across different demographic variables. To determine whether there are any differences among demographic subgroups within this large sample, we ask:

RQ2. Do these factors vary across gender, country, age, and tenure?

To answer RQ2, we conduct a Multi-Group Analysis (MGA) on the variables listed above. A MGA runs the model for each subgroup, so that results for each group, and differences among groups can be inspected.

A central focus of this study is the ability to predict people's retention. The PLS analysis to address RQ1 does not provide a straightforward answer regarding the importance of the factors studied (in terms of their total effects). Some factors might be more important than others (i.e., a higher effect on the variable of interest), but may have a lower 'performance' as measured by the average construct scores. Contrasting importance and performance can help managers to identify those constructs that require more attention. Managers can then implement policies that aim at increasing the performance of constructs that currently 'underperform.' Thus, we ask:

RQ3. Which factors could be improved towards increasing software professionals' intention to stay?

To answer RQ3, we conducted an Importance-Performance Map Analysis (IPMA) [36]. The results of an IPMA facilitate visualization of relative importance and performance of factors, allowing managers to determine which factors to focus on [66].

Many previous studies considered the dependent variable 'intention to stay' or the reverse, 'intention to leave,' rather than measuring actual behavior, i.e., people leaving or staying. A key reason for this is that tracing respondents' actual behavior is very challenging. However, our unique research setting afforded us to investigate whether the extrinsic job characteristics we considered in this study play a different role for those who stayed, and for those who left. Such an analysis therefore goes beyond what can be normally studied; such insights could simply confirm whether studying people's intention is sufficient, or add novel insights if intentions diverge from actual behavior. Hence, we ask:

RQ4. Do these factors vary for those people who leave the organization, and for those who stay?

To answer RQ4, we captured for each respondent their employment status 90 days after the initial survey, and used a MGA to establish whether there are significant differences between those who stayed and those who left voluntarily.

Finally, we sought to establish whether the factors investigated for RQ1 can predict actual attrition; that is, to what extent are the identified factors good indicators of whether or not employees will leave. Using Machine Learning (ML) to train and test a classifier, we addressed the following question:

RQ5. Can these factors predict actual attrition?

We answer these questions within the context of software delivery teams at SoftTech, a global company employing over 26,000 people, and with a global presence in 36 cities in 17 countries across five continents. SoftTech provides services in digital transformation and automation. SoftTech invests in continuous training of its workforce on technical and social skills and has several initiatives in place to retain talent and avoid attrition. SoftTech places the well-being of its employees at the forefront, investing in research to identify and proactively implement strategies to increase engagement while seeking to reduce employee burnout and attrition.

The results of this study show that burnout is a stronger predictor of actual attrition than engagement and intentions to stay. Both engagement and burnout are associated with intentions to stay, varying depending on the tenure of employees within the company. Results also showed that leadership support, organizational culture, and opportunities to learn have different associations with engagement and burnout, and also differ for past and current employees, genders, country of residence, and organizational tenure, while not differing for age nor job roles. We found considerable differences between past and present employees in terms of the relationship between organizational culture, opportunities to learn, and burnout. Further, leadership support was only positively associated with engagement for women and employees from Argentina and Colombia, while only negatively associated with burnout for present employees, not for those who left, regardless of the demographic group.

This article makes a number of novel contributions to the field of behavioral software engineering. First, it contributes to the literature on burnout and engagement by identifying three antecedents that have not previously been studied within the IT domain. Second, it contributes to the very scarce literature on engagement in the IT field. Third, this study measures not only IT workers intentions to leave, but also their actual behavior in the three months that followed the initial data collection. The inclusion of attrition (or, retention) data is novel, and allows the study of actual turnover of survey respondents within the 90-day period after the survey. While several studies have studied turnover in open source communities, frequently measured as "absence of activity" for a given period of time [57], studying actual turnover in the software industry is very rare (cf. [8]). Fourth, this study goes beyond the testing of a theoretical model using PLS-SEM, by leveraging Machine Learning techniques to test whether the identified factors can be used as predictors.

The remainder of this article is organized as follows. In Sec. 3 we review prior work and derive a series of hypotheses which together form the theoretical model. Sec. 4 presents the research design. Results are presented in Sec. 5. Finally, we discuss limitations of this study and implications of the findings in Sec. 6.

2 BACKGROUND AND RELATED WORK

2.1 Prior Work on Burnout among IT Professionals

Recent years have seen an increasing focus on human aspects of software developers and other IT staff, including a considerable stream of work on emotions [80]. Whereas human emotion has long been studied in the social sciences, there is an increasing awareness that the software engineering discipline can also benefit from understanding the specific role and impact of emotion on daily work, including collaboration and productivity, as well as IT professionals' intention to leave [cf. 30, 93].

In this article we focus specifically on two closely-related emergent states among IT professionals in software development and delivery teams: burnout and engagement. Research in the late nineties positioned them as opposites, which suggested that work engagement could be measured using reverse scores of burnout [65]. While there is clearly a considerable inverse correlation between burnout and engagement, it has since been established that the relationship between them is not one of opposites, but that they are distinct concepts [65]. Whereas burnout refers to a negative state of exhaustion and cynicism toward work, engagement is defined as a positive motivational state of vigor, dedication, and absorption. However, although engagement is a positive state and burnout is negative, the absence of one does not represent the others' presence. Thus, studies that focus on these two concepts must operationalize these as distinct concepts [99].

There is a considerable body of research on burnout, both in general and in the software engineering domain [117]. A mapping study by Tulili et al. [117] identified 92 studies that address one or more aspects of burnout among IT staff including software developers which provides an overview of this growing body of work. Appendix A presents an overview of selected studies (sorted by year of publication) that studied antecedents and consequences of burnout or related concepts, such as work fatigue or exhaustion, that focus on IT staff including software developers but also IT managers (see Table 13). While there are some qualitative works that mention burnout as consequences (notably, work by Graziotin et al. [30–32]), we focus here on quantitative studies because we are not aware of any qualitative studies that specifically study burnout as a central concept.

We observe a number of shortcomings of prior work. First, whereas there is considerable research on burnout and related concepts [117], very little work has focused on developer engagement in software engineering, with only a few exceptions [23, 110]; quantitative studies of antecedents of developer engagement are largely missing, for example.

Second, most studies of burnout have focused on its antecedents, with only a few studies considering the consequences of burnout. Clearly, understanding what might cause burnout is very important, because this can help in designing interventions or taking measures so as to prevent burnout in the first place. But, when burnout is not prevented, realizing what may be the consequences is important for organizations so that they can prioritize attention for this topic. The few studies that have studied consequences provide evidence that burnout can lead to an increase in staff turnover intention [cf. 73].

Third, almost all prior studies (see Table 13 in Appendix A) proposed what Demerouti et al. have classified as *job demands*, discussed in more detail below. For example, job overload (or perceived workload) means that people feel they have too many things to do in their job [25, 73, 105, 108]. Other examples include role ambiguity: people may feel that it is not clear what their role and responsibilities are [73, 101, 105], and this imposes a mental 'effort' on behalf of an employee. Far less attention has been paid to what has been labeled *job resources* (also discussed in detail below). Job resources are those aspects that allow people to do their job or stimulate growth. There are a few exceptions; for example, Fujigaki et al.'s study [25] which considered work environment.[1] Other examples of job resources that have been studied are the quality of social interactions [108] and fairness of rewards [73].

Fourth, a few studies have considered intention to leave [73, 106]. While intention is a good predictor of actual behavior, they are not the same. While intention to leave (or the reverse, intention to stay) is commonly studied, very few studies capture whether respondents actually leave due to the difficulty of collecting such information, as this requires researchers to be able to track respondents.

Finally, while there are numerous studies on burnout (see Appendix A), there appears to be little evidence of a cumulative tradition in terms of theory development in these studies. This is surprising, given that burnout has been studied extensively in fields such as psychology. A highly influential theory of burnout is the Job Demands-Resources (JD-R) model [15], which we discuss next.

2.2 The Job Demands-Resources Model of Burnout

Proposed by Demerouti et al. [15], the JD-R model is a general model of burnout that ties several related theories together. The JD-R model posits that burnout manifests primarily as exhaustion and disengagement, which are the result of a too high level of job demands, and a too low level of job resources, respectively. Figure 1 (right-hand side) presents the JD-R model.

Demerouti et al. [15] define *job demands* as: *"those physical, social, or organizational aspects of the job that require sustained physical or mental effort and are therefore associated with certain physiological and psychological costs (e.g. exhaustion)."* These aspects include such things as workload, time pressure, but also shift work [15]. In short, any factor that would require a person to exert effort on, including those factors that make a job more challenging to perform. If job demands

[1]We note that some of the items are rather dated, such as concerns that "the machines at my workplace are inefficient"; efficiency of developers' machines would have been a bigger issue in the early nineties when this study was conducted than it is today. Other items remain as relevant as ever such as "There are often human-relationships problems within the project team."

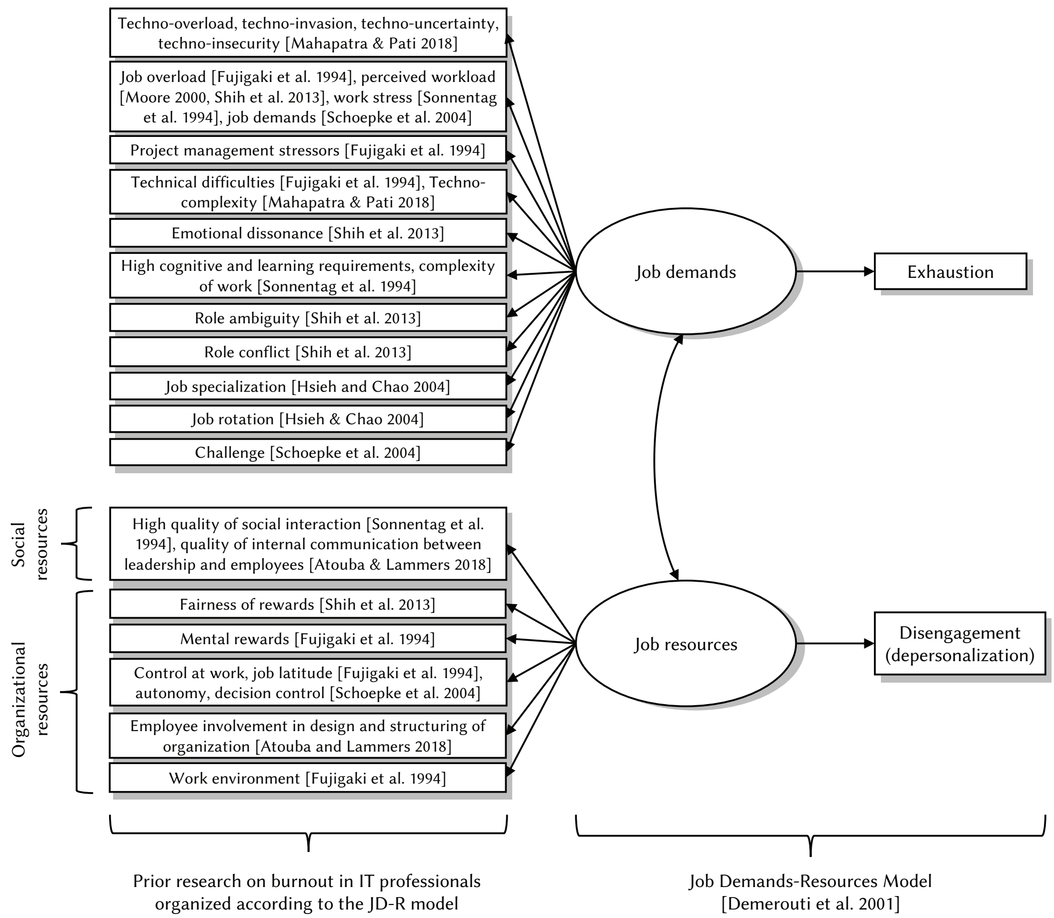

Fig. 1. Prior work on burnout among IT professionals mapped to the Job Demands-Resources model [15]

continue to be too high, exhaustion may follow. *Job resources*, on the other hand, are defined as those aspects that allow a person to achieve work goals, reduce job demands, or stimulate personal growth and development [15]. Resources can be further categorized as social and organizational resources [15]. Social resources include support from colleagues or other people; organizational resources refer to factors such as job latitude and autonomy [15]. A sustained lack of such resources can lead to disengagement. Job demands and resources are closely interlinked, or correlated in statistical terms, which is indicated by the double-headed arrow in Figure 1. The JD-R model, then, can be summarized as capturing two processes. First, sustained overtaxing of employees in terms of job demands leads to exhaustion. Second, sustained lack of job resources leads to disengagement.

Whereas the right-hand side of the figure shows the JD-R model as proposed by Demerouti et al. [15]; the left-hand side shows how prior work on burnout among IT professionals is related to job demands and job resources. We can observe that prior work has studied both job demands and job resources. Further, the JD-R model acknowledges that exhaustion as a key dimension of burnout is separate from (dis)engagement.

3 THEORY DEVELOPMENT

Given the shortcomings of prior work outlined above, we set out to conduct a study to address these. In this section we develop a theoretical model (see Figure 2) that provides the foundation to answer RQ1 and RQ2, namely, to investigate three organizational job resources as antecedents of engagement and burnout as affective states, and intention to stay as a consequence of these states.

3.1 Leadership Support, Opportunities to Learn, and Organizational Culture

Supportive leadership may manifest as emotional or instrumental support [109]. Instrumental support refers to leadership making available resources and information; in this study we refer to emotional support, which includes to a sense of being cared for, and that leadership recognize and value people [109]. A sample study of members of a commerce association (n=283, 54% female) suggests a relationship between psychological climate and employee engagement [107], whereby psychological climate comprised several variables including supportive management, which we equate here with supportive leadership. A previous review of literature also lends support to the importance of leadership for employee engagement [13]; this literature review identified different conceptualizations of leadership, with 'transformational leadership' as a recurring theme. Bass distinguished transformational leadership from transactional leadership, and argued that the former style can be effective by, among others, meeting emotional needs of employees [9].

Leadership support also affects software delivery teams in several ways. Supportive leaders can bring positive effects to the team by reducing friction between developers and helping them to be more productive [30]. Supportive leadership behaviors are directed toward the satisfaction of subordinates' needs and preferences, such as displaying concern for subordinates' well-being, and creating a friendly and psychologically supportive work environment [49]. By enabling subordinates, setting examples, and rewarding desirable behaviors, leaders also contribute to engagement and job satisfaction by bringing role clarity and inducing self-efficacy [54]. Employees are more likely to remain with an organization if they believe that their managers genuinely show interest and care for them [61]. On the other hand, a lack of supportive leadership can lead to additional stress, burnout, and increase a person's intention to leave a team [5, 54]. Thus, in line with prior literature, we propose the following hypotheses.

> Hypothesis 1 (H1): A higher level of leadership support are associated with a higher level of engagement.

> Hypothesis 2 (H2): A higher level of leadership support is associated with a lower level of burnout.

Argote et al. argued that an individual's performance is dependent on ability, motivation, and opportunity [6]. Within an IT context, we do not have reason to question the role of ability as staff usually are well educated. Prior research on motivation in software engineering also suggests that "learning, exploring new techniques and problem solving appear to be the motivating aspects of SE" [10]. Learning appears to be an important factor for software developers; previous work has shown that software developers tend to have a higher need for cognition, which can be described as a tendency to engage in and enjoy effortful thinking [93]. Opportunity, then, is also needed to learn. Wiersma showed that 'excess capacity,' or *slack* in resources, facilitates this opportunity [122]. We argue that such slack in resources can be created and facilitated by an organization's management who value the well-being of IT staff, suggesting a relationship between leadership support and having opportunities to learn. Hence, we propose:

Hypothesis 3 (H3): A higher level of leadership support is associated with more opportunities to learn.

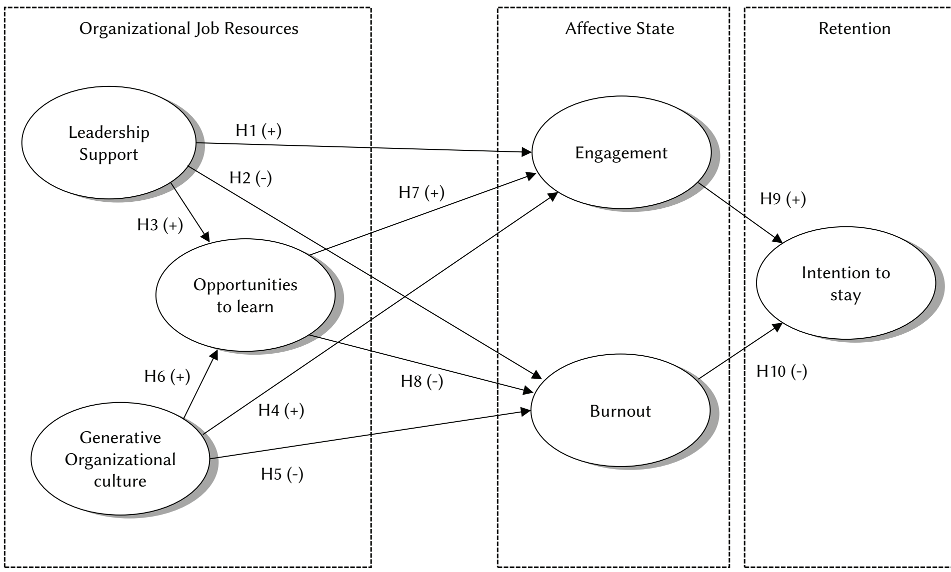

Fig. 2. Research model for research questions 1 and 2

An organization's culture affects people's daily work activities. A team's culture can influence software delivery performance [21, 29], staff well-being, and retention [16]. Westrum developed a typology of organizational cultures based on human factors in system safety, particularly in the context of accidents in technological domains, such as aviation and healthcare [121]. The typology defines three types of organizations in terms of information flow and psychological safety. *Pathological* organizations exhibit low levels of cooperation across groups and a culture of blame. *Bureaucratic* organizations emphasize rules and hierarchy and compartmentalize responsibilities by departments, which in turn inhibits information flow. Generative organizations are performance-oriented, with good information flow, high levels of cooperation and trust, and bridging between teams.[2] A generative organizational culture can be achieved by creating cross-functional teams to improve cooperation, holding non-judgmental postmortems, sharing risks and responsibilities, breaking down organizational 'silos,' and encouraging finding areas to collaborate and improve processes (e.g., DevOps), experimentation, and novelty. An organizational culture in which members of the team cooperate with each other and share responsibilities [121] positively impacts engagement [107] while negatively impacts burnout [21].

Hypothesis 4 (H4): A generative organizational culture is associated with a higher level of engagement.

Hypothesis 5 (H5): A generative organizational culture is associated with a lower level of burnout.

[2]Bridging refers to making connections between different roles and identifying areas to collaborate, and seeks to close the gap that typically exists between organizational 'silos'.

Further, an organization that exhibits a culture for learning makes resources available for continued education and offers continuous encouragement to teams to learn by providing them space and time to acquire new knowledge and explore ideas [29]. A healthy organizational culture fosters the process of learning [24]. When holding 'blameless' (non-judgmental) retrospectives and having out-of-the-box thinking sessions, a generative organizational culture [123] creates more opportunities to learn [58, 116] as instead of punishing, the team is trained to learn from failures. Based on the role of organizational culture outlined above, we propose the following hypotheses:

Hypothesis 6 (H6): A generative organizational culture is associated with more opportunities to learn.

Learning opportunities have been linked to job satisfaction and job-related well-being in other fields [18]. Having opportunities to learn new skills is central to the motivational process for employees to thrive at work. Employees with opportunities to stimulate personal growth, learning, and development will be more engaged to work [26, 110]. The strength of a person's motivation by opportunities to learn is embedded in the desire for personal accomplishment by receiving challenging work to have feedback on actual performance [23]. Nonetheless, stagnation and ineffective learning opportunities contribute to disappointment and burnout [61, 81]. Thus, in line with prior literature, we posit that:

Hypothesis 7 (H7): Having more opportunities to learn is associated with a higher level of engagement.

Hypothesis 8 (H8): Having more opportunities to learn is associated with a lower level of burnout.

Engagement represents an attitude toward the work, an active positive concept that translates into having lower intentions to leave the job. Only a handful of studies have studied the consequences of burnout; two studies considered turnover intention [73] or more generally an intention to leave the IT field [106]. Both studies demonstrated that burnout does indeed correlate with increased levels of respondents' intentions to leave an organization (or leave the IT field altogether). In both studies, the focus was on respondents *intentions* rather than determining whether they actually left, which is considerably more difficult to do. One other study that focused on consequences considered job satisfaction, which in turn is known to correlate strongly to intention to leave, depersonalization and indirectly (through mediation), a sense of personal accomplishment.

Work exhaustion and the effects of stress can influence the decision to leave the job [106], also for professionals in the software industry [70, 73]. Work exhaustion is linked to low job satisfaction for software professionals [70, 73]. Hence, we propose:

Hypothesis 9 (H9): A higher level of engagement is associated with a higher level of intention to stay.

Hypothesis 10 (H10): A higher level of burnout is associated with a lower level of intention to stay.

4 RESEARCH DESIGN

Figure 3 presents an overview of the research design, with an outline of the five research questions presented in Sec. 1, data sources and analysis procedures used for each. We first test establish a foundation by testing the hypotheses presented in Sec. 3. We do so using data collected through a large-scale survey at SoftTech, a global organization whose management invests considerably in employee well-being. The hypotheses were tested using PLS-SEM. We then investigate how these results might differ across different cohorts (RQ2), by evaluating the role of gender, country of location, job role, age, and organizational tenure. To do so, we conducted a multi-group analysis within the PLS framework. We then focus on establishing which of these factors might be the most important ones (RQ3), using an Importance-Performance Map Analysis (IPMA), which is a technique within the PLS-SEM framework.

These first three research questions provide a conceptual foundation, focusing on the relationships between extrinsic job factors, affective state, and an intention to stay with the organization. More important is, of course, whether or not employees actually will stay, rather than knowing their intention, and the ability to predict this type of behavior. Hence, research questions 4 and 5 address these issues. We investigate how the factors analyzed for RQ1 vary for current and past employees. To be able to determine this, we collected additional data from SoftTech's HR department 90 days after the initial data collection on whether respondents had *voluntarily* (i.e., not terminated by SoftTech) left their jobs in that period. This allowed us to divide the respondents into two groups: those who left and those who stayed. We analyzed the extended dataset using a multi-group analysis. Finally, RQ5 seeks to establish whether we can predict employee attrition based on the same set of factors. Leveraging the same additional data from HR, we trained and tested a Machine Learning model.

The remainder of this section discusses how the conceptual variables in this study were measured (Sec. 4.1), how data were collected and analyzed (Sec. 4.2), and the development of the Machine Learning (ML) model (Sec. 4.3).

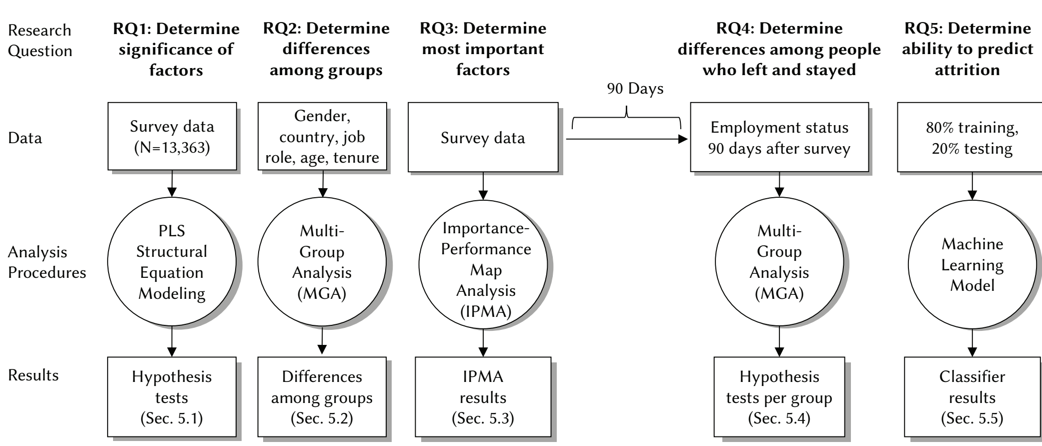

Fig. 3. Research design

4.1 Measurement Model and Data Collection

The hypotheses investigate a number of theoretical concepts that cannot be directly measured (e.g., organizational culture); instead is measured through a set of indicators or manifest variables, from which a proxy is calculated. In this study, we use PLS-SEM, which creates those proxies as weighted composites (using PLS Mode A, which uses a correlation weighting scheme; this is in contrast to PLS Mode B, which produces regression weights. Mode A offers superior out-of-sample predictions [89]).[3] PLS is a suitable approach for predictive studies, more so than CB-SEM due to factor indeterminacy that is inherent in the latter approach [38, 89].[4]

For the constructs in this study, we adapted existing measurement instruments where possible. The survey was co-designed with SoftTech's HR department; consequently, pragmatic decisions were made, including an attempt to restrict the length of the questionnaire. We define the constructs below; Appendix B provides all survey items pertaining to these constructs.

- Leadership Support (LS): items were adapted from the Supervisory Scale [71] and the Emotional Support [109] instruments to better fit the context of SoftTech. Items included caring about well-being and work-life balance, recognizing achievements, having meaningful conversations about the employees' career interests, and employees' preference to work with a leader again.
- Opportunities to Learn (OL) was measured by questions about employees' belief in the readiness and possibility of learning. Questions were inspired by an instrument of Employee's Learning Opportunity [119], and included statements about experiences to increase their skills and being offered opportunities to grow in their career.
- Generative Organizational Culture (OC) was measured by questions inspired by Westrum's typology [121] (see Sec. 3), which has previously been used to measure organizational culture in software delivery teams [21, 22].
- Engagement (EN) was measured by questions adapted from the UWES-3 instrument [100], which includes the dimensions of vigor, dedication, and absorption.
- Burnout (BT) was measured by questions adapted from the Maslach Burnout Inventory (MBI) that includes the dimensions of cynicism, exhaustion, efficacy (reverse-coded), and inefficacy [68].
- Intention to Stay (IS) was measured by reversed questions about turnover intentions [103].

The survey included three demographic questions: age, organizational tenure, and country of residence to be answered in ranges by participants. We captured age as a range as respondents might be uncomfortable or reluctant to share their precise age. Tenure was also captured as a range rather than a precise number because the company acquired many other startups, and some people could not determine precisely in which year they joined the company. The demographic data about role and gender were used from the company's pre-existing demographic data for its reporting requirements under government laws. SoftTech's HR department recorded gender as a binary variable; we acknowledge that some respondents may not identify as a man or woman.

Adult development studies suggest that, as they age, individuals pass through different development stages affecting job priorities. Employees beyond the early stage of their career may have more constraints in their ability to leave their job due to family responsibilities (e.g., care for children or parents) or financial concerns (e.g., mortgage payments); this may play a role in their intention to leave their job. Finally, the survey instrument also included an open question that invited respondents to share their thoughts and experiences about working at SoftTech.

SoftTech's HR department administered the online questionnaire using an internal survey tool, which was answered by members of software delivery teams throughout the company. The survey was sent to respondents by email using a corporate address and was available for one month. All team leaders encouraged their team members to fill out the questionnaire during regular meetings.

[3] In covariance-based SEM (CB-SEM), such proxies are created as common factors, which assumes that there is a common factor structure underlying the data. In this study, we align with Rigdon's realist perspective that acknowledges that the constructs (calculated either as a common factor, or, as in this study, as a composite) representing the theoretical concepts are not equivalent [89].

[4] Factor indeterminacy refers to the problem that there is no single, unique set of values to solve for the model; the number of sets of values is infinite to fit the model equally well [74, 89, 102].

Table 1. Demographics of respondents (n=13,343)

Attribute	N	Percentage
Country of Residence		
Argentina	3,014	22.6%
Colombia	2,906	21.8%
India	2,366	17.7%
Mexico	1,655	12.4%
Chile	743	5.6%
Peru	630	4.7%
Uruguay	533	4.0%
Brazil	466	3.5%
USA	324	2.4%
Spain	261	2.0%
Romania	106	0.7%
Ecuador	84	0.6%
UK	82	0.6%
Costa Rica	76	0.6%
Belarus	55	0.4%
Others	42	0.4%
Gender		
Men	9,554	71.6%
Women	3,789	28.4%

Attribute	N	Percentage
Age Range		
18-24	835	6.3%
25-34	6,919	51.9%
35-44	4,357	32.7%
45+	1,135	8.5%
Prefer not to Say	97	0.6%
Organizational Tenure Range		
Less than 6 months	1,749	13.1%
6 months to 1 year	2,776	20.8%
1-3 years	5,598	42.0%
3-5 years	1,568	11.8%
5 years or more	1,598	12.0%
Did not answer	54	0.3%
Managerial roles		
Managerial roles	1,839	13.8%
Non-managerial roles	11,504	86.2%
Employment Status after 90 days		
Employees who left	474	3.6%
Current employees	12,869	96.4%

A total of 15,762 responses were submitted, including incomplete (partial) responses. As our analysis techniques require complete responses, we remove responses with missing values. While imputation methods could be used, more than 10% of responses had missing values; in such a case, imputation methods could bias results for multi-group analyses [34]. Given the very large number of responses, the trade-off between introducing a potential bias versus increasing the usable sample size suggested taking a more conservative approach.

The survey was not anonymous, but the research team had no access to data that could identify specific respondents. Ninety days after the survey was closed, the HR department collected additional data on the survey respondents' employment status to mark those who left the company and those who had not left. This information was used to answer RQ4, and in the training and testing of the Machine Learning model (for RQ5) (see Fig. 3). The ability to enrich the dataset was only possible because the study was conducted at one organization; capturing such information in an industry-wide survey is practically not possible.

4.2 Data Analysis

After removing incomplete responses, we used a sample of 13,343 complete responses for analysis. Table 1 presents some demographics. The most common job roles of respondents are software developers (n=5,230), testers (n=1,956), team leaders (n=809), business analysts (n=551), and designers (n=456). Software developers are specialized in the following technologies: Java, .Net, Salesforce, Android, NodeJS, iOS, Python, PHP, Mulesoft, C++, Drupal, Go, Ruby, Sharepoint, and Magento. Managerial roles (n=1,839) include product and project managers.

Job psychological states and involvement vary per country culture [79, 82], and people who live in different countries feel a stronger turnover intention-behavior [124]. We performed pairwise comparisons for the five countries with the most respondents (Argentina, Colombia, India, Mexico, and Chile; see Table 1). Moreover, based on the respondents' country of residence, we used the six dimensions of Hofstede's classification of national culture [48] as control variables in the model: Power Distance, Individualism/Collectivism, Long Term Orientation, Masculinity/Femininity, Indulgence, and Uncertainty Avoidance. This, of course, ignores the migration of expatriates. However, there is no practical alternative to measure national culture. As a result, this could introduce measurement error, which in turn would reduce the statistical power of the study.

We compared all ranges of tenure and age in Sec. 5.2 for the multi-group analysis of the model. Previous research showed that organizational tenure has a moderate, significant, and negative relationship with the intention to leave a company [52].

We used SmartPLS version 4.1.0.3 [91] for the analyses to answer RQ1 to RQ4, and Python's scikit-learn package[5] for RQ5. We share the dataset online which allows replication [115].

4.2.1 Multi-Group Analysis. We address RQ2 and RQ4 using a Multi-Group Analysis (MGA) procedure using SmartPLS. The goal of a MGA is to understand how the model varies for different subsets of respondents. To conduct a multi-group analysis, Hair et al. [39] proposed three steps: (1) group creation; (2) invariance test; and (3) result analysis.

Step 1. Group Creation. We grouped respondents to observe any potential heterogeneity among the following sets of groups (see Table 1 for demographics):

- Gender (two groups): men and women.
- Job roles (two groups): Managerial Roles (project and product managers) vs. Non-Managerial Roles (all other roles).
- Age (four ranges): 18-24; 25-34; 35-44; 45+.
- Organizational tenure (five ranges): Less than 6 months, 6 months - 1 year; 1-3 years; 3-5 years; 5+ years.
- Country of residence (5 countries): We considered the five countries with the most respondents (Argentina, Colombia, India, Mexico, and Chile), and compared them pairwise.

Step 2. Evaluation of Measurement Invariance of Composite Models (MICOM). Measurement invariance evaluation is a mechanism to assess whether or not the loadings of the items that represent the latent variables vary significantly across different groups. In other words, a MICOM analysis can be conducted to establish whether any differences can be attributed to the constructs that make up the theoretical model, and not to how those constructs were measured [39]. If differences can be observed in the measurement model for two groups, then differences in the model cannot be fully attributed to the theoretical constructs.

Comparing group-specific model relationships for significant differences using a multi-group analysis requires establishing what is called *configural* and *compositional* invariance [39, 47]. Configural invariance does not involve a test but is a qualitative assessment of ensuring that all of the composites are equally defined ("configured") for each of the groups, such as equivalent indicators per measurement model, equivalent treatment of the data, and equivalent algorithm settings or optimization criteria. Configural invariance is established in our model as no different settings or treatments were applied to the groups.

Compositional invariance exists when the composite scores are the same across both groups, despite possible differences in the indicator weights [97]. While small differences will naturally happen for different groups, we can test whether those differences are significant. For this purpose, the MICOM procedure examines the correlation between the composite scores of both groups and requires that the correlation equals 1. We ran the permutation tests in SmartPLS between all pairs of each group. We verified that compositional invariance is established for all latent variables in

[5] http://scikit.ml

the PLS path model for the groups of attrition, genders, and job roles. However, compositional invariance was not established for the countries Colombia and Mexico, so we excluded this pair from the analysis by country given that compositional invariance is a prerequisite. Compositional invariance was established for all composites in three (out of six) pairs of age ranges, and for seven (out of ten) pairs of tenure ranges, which we included in the analysis of the next step. We established partial measurement invariance, and thus multi-group analysis is suitable for part of the groups we defined [90]; the pairs for which compositional invariance was not established were discarded and not analyzed.

Step 3. Group Comparison and Analysis. Path coefficients generated from different samples are usually numerically different, but the question is whether those differences are statistically significant. We analyzed the differences between the coefficients' paths for the various groups. Significant differences can be interpreted as moderation effects; that is, a difference in the coefficients could be attributed to respondents belonging to a different group (e.g., a different age category).

4.2.2 Importance-Performance Map Analysis. RQ3 asks which of the factors are most important for the dependent variable, Intention to Stay. We employed an Importance-Performance Map Analysis (IPMA), which combines the analysis of two dimensions: importance (represented by total effects, i.e., the sum of direct and indirect effects) and performance (represented by construct scores, rescaled on a scale from 0 to 100) [90]. Sec. 5.3 reports the results of this analysis.

4.3 Machine Learning

Machine Learning (ML) has previously been used to predict employees who are likely to leave an organization, either through demographics, (lack of) salary increase [2], and work-related withdrawal indicators such as lateness and absenteeism [76]. To determine whether we can predict retention (RQ5), we developed an ML model to predict whether an employee will leave or not, using the latent variables identified for RQ1 as predictors.

4.3.1 Dataset. The dataset contains only values for the specific items, which together represent the latent (unobservable) theoretical constructs. To use the values of these theoretical constructs, we exported the latent variable scores of the PLS-SEM model (leadership support, organizational culture, opportunities to learn, engagement, burnout, and intention to stay) which can be used for further analysis. The PLS algorithm generates such scores as weighted linear combinations of a latent variable's indicators. Latent variable scores are unique to each respondent [88] and were used to train ML models and predict the binary outcome of leaving the organization; as mentioned earlier, this is the additional information collected from SoftTech's HR department 90 days after the initial survey.

4.3.2 Classifiers. Highly correlated data can bias machine learning training, and because of that, we used Spearman's correlation algorithm to eliminate similar features. A Spearman correlation coefficient above 0.8 is considered a strong correlation [55]. The dataset had pairs of variables that correlated to a maximum of 0.65; therefore, we retained these six variables in the model.

We selected various commonly-used supervised classification algorithms from scikit-learn[6] that had previously been used to predict attrition [17, 76, 85, 112]: Decision Tree, Random Forest (ensemble classifier), K-NeighborsC, Gaussian, LinearVC, MLPC Classifier (neural network multilayer perceptron), Logistic Regression, and a dummy classifier with a strategy labeled "most_frequent." Dummy classifiers are typically used as a baseline [19, 95].

[6]https://scikit-learn.org/stable/

4.3.3 Training and Test Datasets. Each classifier was trained using the binary attribute as to whether or not people left during the 90-day period after the survey. The number of people who left is far smaller than those who stayed, which means that the distribution of these two groups is imbalanced. We performed a sensitivity analysis of balancing the training data using the Synthetic Minority Over-Sampling (SMOTE) technique [14].

To avoid overfitting, we ran each analysis ten times, using ten different training and test sets to match a 10-fold cross-validation, using the StratifiedShuffleSplit method from the scikit-learn Python package.[7] The StratifiedShuffleSplit keeps the same distribution across training and testing datasets and is a model selection technique that has been widely applied in software engineering research [42, 125]. The original dataset was divided into two parts with an 80:20 ratio (80% of the data used for training, 20% used for testing).

4.3.4 Model Tuning. To prevent overfitting, the two models with the best results (Decision Tree and Random Forest) were regulated using different procedures. The Decision Tree was submitted to a post-pruned procedure using the cost complexity parameter (CCP) [104]. Decision Trees with maximum depth tend to overfit [11]. As the CCP increases, the tree is pruned, generating a much better Decision Tree that can be generalized. The list of CCP obtained was used in the Randomized Search[8] Cross Validation procedure along with a set based on similar studies [64].

We applied Out-of-bag (OOB) and early-stopping techniques for the Random Forest Classifier. OOB is the average error calculated by each training data. For each observation, we predict them using only trees that do not contain the training data observed [27, 60, 86]. We got the Random Forest number of estimators where the OOB found the minimum error. Early stopping uses a number of interactions to measure the test error. Once the error reaches the minimum level and goes up again, the algorithm picks the number of estimators [67]. We obtained the number of estimators from early-stopping and OOB and used them as input to the Randomized Search.

Model performance may vary depending on the dataset and the values of the algorithm parameters, known as hyperparameters. We can test multiple combinations of hyperparameters to obtain better results, but manually doing so is very time-consuming. Tuning is the task of finding optimal hyperparameters for a learning algorithm for a given dataset [87].

For this purpose, we used the Randomized Search Cross Validation (CV)[9] for the two algorithms that performed best (Random Forest and Decision Tree) in the Scikit-Learn3 Python library. Randomized Search randomly generates a set of combinations to consider and evaluate [1]. The scoring function looks for the best setup for the F-measure. Randomized search can lower computational cost, particularly in scenarios where an extensive array of possible configurations is being considered [4]. The algorithm seeks to optimize permutations of hyperparameters and chooses samples randomly. Several models are generated for each permutation of hyperparameters, and their performance is recorded to identify the best model.

Randomized Search does not search for all possible setups as the Grid Search procedure does, but usually obtains similar results in a fraction of the time.[10] We opted to increase the number of parameters tested with Randomized Search instead of reducing the options to run the Grid Search procedure. In large datasets, Grid Search may be unfeasible [60].

For Decision Tree, we used 'criterion,' 'max depth,' 'min samples split,' 'min samples leaf,' 'min weight fraction leaf,' and 'max features.' We used the hyperparameters suggested by Yang and Shami

[7]https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedShuffleSplit.html
[8]https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html
[9]https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html#sklearn.model_selection.RandomizedSearchCV
[10]https://scikit-learn.org/stable/auto_examples/model_selection/plot_randomized_search.html

[126] for Random Forest: 'max depth,' 'min samples split,' 'min samples leaf,' and 'n estimators.' We briefly describe these parameters. *Criterion* determines the quality of splits and can be set to either 'gini' or 'entropy'; gini uses the probability of a random wrong classification for a feature as split criteria. Features with low 'gini' are chosen for split. Entropy measures the node disorder; nodes with more variable content regarding the dependent variable have higher entropy and are better candidates for splitting; 'max depth' represents the maximum number of nodes in each tree; 'min samples split' is the minimum number of data points required to split a decision node; 'min samples leaf' represents the minimum number of data points required to form a leaf node; 'max features' specifies the number of considered features for generating the best split, and 'n estimators' is the number of combined trees in the forest. 'class weights ' allows Decision Trees to balance unbalanced data by attributing weights for each class. Since we are using SMOTE, mixing both would create unexpected balances, so we fixed it to a 'balanced' value. Finally, 'min weight fraction leaf' controls the minimum weight (0-1) of all sum of weights to be used when selecting a leaf node.

4.3.5 Classifier Evaluation. To evaluate the classifiers, we used the following metrics (calculated using the scikit-learn package mentioned above):

- Precision, calculated as the percentage of the correctly classified data samples by the model, i.e., the sum of true positive (TP) and true negative (TN) per the total number of data samples in the dataset.
- Recall corresponds to the percentage of correctly predicted labels among all truly relevant labels.
- F-Measure, calculated as the harmonic mean of precision and recall. F-measure is a weighted measure of the number of relevant labels that are predicted, and the number of the predicted labels that are relevant.
- Accuracy, the percentage of correct classifications, obtained by dividing the number of correct predictions by the total number of predictions in all classes.

We compared the differences observed in algorithm predictions using the non-parametric Mann-Whitney U test, followed by Cliff's delta (δ), a non-parametric effect size test. The magnitude of Cliff's delta was assessed using the thresholds provided by Romano et al. [92], i.e., negligible ($|\delta| < 0.147$), small ($|\delta| < 0.33$), medium ($|\delta| < 0.474$), and large otherwise.

We focused our analysis on the F-Measure. The precision metric has an unbalanced penalty for the models with a large number of false positives, which means it assigns a cost or negative impact to models that produce a high number of incorrect positive predictions (false positives). The recall metric penalizes false negatives more, that is, it assigns a cost or negative impact to models that fail to identify actual positive instances (false negatives). The F-Measure metric is a harmonic mean of the precision and recall metrics, and is a way to penalize false positives and false negatives equally. In this case, the F-Measure metric equally penalizes (a) employees who left and were not predicted to leave and (b) employees who did not leave and were predicted to leave.

4.4 Measurement Validity

The theoretical model in Fig. 2 contains a number of constructs, each of which is measured using specific items. We now discuss the results of several procedures to assess the validity of the measurement model introduced in Sec. 4.1.

4.4.1 Internal Consistency Reliability. Second, we verified how well the different indicators are consistent with one another and able to reliably and consistently measure the constructs. A high degree of consistency suggests that indicators refer to the same construct. There are several metrics to measure internal consistency reliability. A traditional metric is Cronbach's α, but this assumes an equal loading across items and tends to underestimate, which is why this metric is no longer recommended. A more appropriate metric is the composite reliability (ρ_c), which takes the outer loadings of the items into account. Because ρ_c tends to overestimate, an alternative metric is the reliability coefficient ρ_a [38]; we report both ρ_a and ρ_c (see Table 2). A desirable range of values for both metrics's ρ_a and ρ_c is between .7 and .9 [37]. Values below .6 suggest a lack of internal consistency reliability, whereas values over .95 suggest that indicators are too similar and thus are not desirable. All values for ρ_a and ρ_c values fell between .72 and .90.

Table 2. Internal consistency reliability

	ρ_A	ρ_C	AVE
Leadership Support	.88	.91	.58
Generational Organizational Culture	.80	.86	.50
Opportunities to Learn	.74	.85	.66
Engagement	.76	.85	.66
Burnout	.80	.86	.55
Intention to Stay	.72	.86	.75

4.4.2 Convergent Validity. We assessed convergent validity, which assesses whether the indicators that represent the theoretical concepts are understood by the respondents in the same way as they were intended by the designers of the questions. The assessment of convergent validity relates to the degree to which a measure correlates positively with alternative measures of the same construct. We used two metrics to assess convergent validity: the Average Variance Extracted (AVE) and the loading of an indicator onto its construct (see Table 2).

The AVE is equivalent to a construct's communality [37], which is the proportion of variance that is shared across indicators. The AVE should be at least .50, indicating that it explains most of the variation (i.e., 50% or more) in its indicators [37]. All AVE values are at least .50 (see Table 2).

An outer loading of .7 is considered a minimum, though .6 is considered sufficient for exploratory studies [37]. We followed an iterative process to evaluate the outer loading of the constructs. The indicators of all constructs exceeded .60, but burnout initially had six indicators, with one indicator's outer loading below .6 (*BT6: I feel ineffective at work*). We decided to remove the indicator, leaving five indicators; the AVE value of burnout increased from .49 to .55, and all constructs had outer loadings of .69 or higher.

4.4.3 Discriminant Validity. Third, we verified whether each construct represented different concepts or entities, through tests of discriminant validity. The primary means to assess discriminant validity is to investigate the Heterotrait-monotrait (HTMT) ratio of correlations [46]. The discriminant validity could be considered problematic if the HTMT ratio exceeds .9 [46]. The HTMT ratio between the four constructs ranged between .60 and .89. We also assessed the Fornell-Larcker criterion and cross-loadings of indicators (see Tables 15 and 14 in Appendix B). Both procedures indicated that discriminant validity did not pose a threat in this study.

4.4.4 Assessing Collinearity. The theoretical model has three different exogenous variables: Leadership Support, Organizational Culture, and Opportunities to Learn, as well as several control variables. We hypothesized that the exogenous variables and the control variables are associated with the endogenous variables Engagement, Burnout, and Intention to Stay. To ensure that the four exogenous constructs are independent, we calculated their collinearity using the Variance Inflation Factor (VIF). A widely accepted cut-off value for the VIF is 5 [37]; all VIF values were below 2.2.

4.5 Model Evaluation

We report three measures relevant to evaluating the PLS model. We assessed the relationship between constructs and the predictive capabilities of the theoretical model. The R^2 values of the endogenous variables in our model (engagement, burnout, and intention to stay) ranged between .33 and .53. While some scholars have suggested thresholds to evaluate such values, there is considerable debate about setting such thresholds. Other factors play a role in engagement and burnout, for example, and so it is unlikely to get values close to 1.0, nor should that be the goal, as doing so would make the theoretical model overly complex. We consider the R^2 values of .53 (opportunities to learn), .46 (engagement), .45 (burnout) and .33 (intentions to stay) high.

We also inspected the model's predictive relevance by means of Stone-Geisser's Q^2 [111] value, which is a measure of external validity [37]. This measure can be obtained through the PLS-Predict procedure (available within the SmartPLS software). PLS-Predict is a holdout sample-based procedure that generates point predictions on both the item level and the construct level, dividing the sample data into k subgroups ('folds') of roughly the same size and combining k-1 folds into a training sample that is used to estimate the model. The remaining fold serves as a holdout sample that is used to assess the model's predictive power [35]. Q^2 values are calculated only for endogenous variables: opportunities to learn, engagement, burnout, and intention to stay, which led to respective values of .52, .39, .41, and .25. Values larger than 0 indicate the construct has predictive relevance, while negative values show the model does not perform better than the simple average of the endogenous variable would do.

Finally, we report the Standardized Root Mean Square Residual (SRMR) as a common fit measure that is appropriate to detect misspecification of PLS-SEM models [44, 94]. A value of 0 for SRMR would indicate a perfect fit, and values less than .08 (conservative) or .10 (more lenient) are considered a good fit [45]. Our results suggest a good fit of the empirical data with the theoretical model (SRMR = .059).

5 ANALYSIS AND RESULTS

We now present the results to the research questions. Section 5.1 presents the results of the hypothesis testing, and shows which of the hypotheses are supported by the data. Section 5.2 presents the results of multi-group analyses to determine whether the results of the hypothesis testing vary across different groups of respondents (gender, country, age, and tenure). Section 5.3 presents the results of an Importance-Performance Map Analysis (IPMA), which shows the importance of each factor on intention to stay. Section 5.4 presents the results of a multi-group analysis to determine whether the results vary for those people who left SoftTech in the 90 days following the initial survey. Finally, Section 5.5 presents an ML model to predict attrition.

5.1 Hypothesis Evaluation

Table 3 shows the results for the hypotheses, including the mean of the bootstrap distribution (*B*), the standard deviation (*SD*), the 95% confidence interval, and p-values. The path coefficients in Fig. 4 and Table 3 are standardized regression coefficients, indicating the direct effect of one variable on another. Based on these results, we found support for all hypotheses ($p < .001$), except for H1. In PLS-SEM, significance of a parameter estimate is established through a bootstrapping procedure, which generates a confidence interval (CI). If the CI contains the value 0, then the parameter estimate is deemed insignificant, which is the case for H1, despite the p-value being just below .05. None of the control variables was significantly associated with intention to stay, burnout, or engagement.

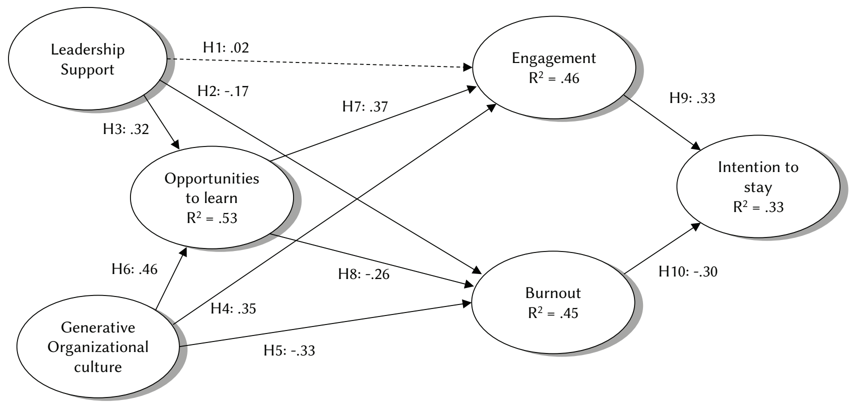

Fig. 4. Significant path coefficients ($p < 0.05$) indicated by a full line). Non-significant links are indicated with a dashed line. None of the control variables were significant, and are thus not shown.

A higher level of leadership support was not significantly associated with a higher level of engagement for the overall population (H1, *B*=0.02), but was negatively associated with burnout (H2, *B*=−0.17) and positively associated with having opportunities to learn (H3, *B*=0.46). Several respondents shared comments regarding a lack of leadership support. For example, one respondent wrote: *"People don't quit companies, we quit bad leaders,"* suggesting that leadership plays an important role indeed in people's decisions to stay or leave. Similarly, another respondent suggested that a person's perception of an organization is affected by leadership: *"Having a bad leader quickly changes the good image you are building in your mind about the company."* While these comments highlight the importance of supportive leadership, it also suggests that people do not perceive leadership to be a constant factor, and that leadership support can change, which in turn would re-calibrate people's relations with the organization. For example, one respondent wrote that: *"[the] recent change of the leader on my project make me feel more comfortable."*

We found support for the positive association between a generative organizational culture and engagement (H4, *B*=.35) on the one hand, and a negative association with burnout (H5, *B*=−0.33) on the other. A generative organizational culture was also positively associated with opportunities to learn (H6, *B*=.46). The generative culture was characterized by one of the respondents as: *"there is no failing, just another moment of learning."* The *"motivating culture"* of SoftTech was mentioned with enthusiasm by several respondents using phrases such as the company being *"inclusive," "employee-oriented," "collaborative," and "innovative,"* and as having the *"power to make changes, to contribute, to be kind, to be epic, to stay curious and always seek reinvention."*

Having opportunities to learn has a positive association with engagement (H7, *B*=.37) and a negative association with burnout (H8, *B*=−0.26). We observed many open-question responses having "learn" and some hedonic terms (e.g., *"nice, calm, learning," "new learning experience, joyful"*). An analysis of open-question responses showed that the word "learn" appeared over 1,000 times, whereas the word "growth" appeared 732 times. While such analyses should be considered with caution, this observation does suggest that respondents appear to value learning on the job; this corresponds to findings in previous work, and indeed hypothesis H7 (cf. [23]).

Finally, the data also lend support to H9 and H10. Engagement is positively associated with having an intention to stay (H9, *B*=.33); burnout, on the other hand, is negatively associated with the intention to stay (H10, *B*=−0.30). When experiencing the excitement of being engaged, team

Table 3. Standardized path coefficients, standard deviations, confidence intervals, and p-values

Hypothesis		*B*	SD	95% CI	p
H1 Leadership Support → Engagement		.02	.01	(−.00[a], .05)	.049[b]
H2 Leadership Support → Burnout		−.15*	.01	(−.20, −.15)	.000
H3 Leadership Support → Opportunities to Learn		.46*	.01	(.44, .48)	.000
H4 Organizational Culture → Engagement		.35*	.01	(.32, .37)	.000
H5 Organizational Culture → Burnout		−.33*	.01	(−.34, −.30)	.000
H6 Organizational Culture → Opportunities to Learn		.46*	.01	(.68, .70)	.000
H7 Opportunities to Learn → Engagement		.37*	.01	(.35, .40)	.000
H8 Opportunities to Learn → Burnout		−.27*	.01	(−.28, −.24)	.000
H9 Engagement → Intention to Stay		.33*	.01	(.31, .35)	.000
H10 Burnout → Intention to Stay		−.30*	.01	(−.32, −.28)	.000
Control Variables					
Power Distance	→ Opportunities to Learn	.01	.02	(−.02, .03)	.691
	→ Engagement	−.01	.01	(−.04, .02)	.548
	→ Burnout	.01	.02	(−.02, .04)	.648
	→ Intentions to Stay	.01	.02	(−.03, .04)	.703
Individualism	→ Opportunities to Learn	−.01	.01	(−.03, .02)	.508
	→ Engagement	.00	.01	(−.02, −.03)	.321
	→ Burnout	.07	.01	(−.02, .03)	.604
	→ Intentions to Stay	−.01	.01	(−.04, .02)	.538
Masculinity	→ Opportunities to Learn	−.02	.01	(−.03, .01)	.049[b]
	→ Engagement	−.01	.01	(−.02, .01)	.532
	→ Burnout	.00	.01	(−.02, .02)	.987
	→ Intentions to Stay	−.01	.01	(−.03, .01)	.244
Uncertainty Avoidance	→ Opportunities to Learn	−.01	.01	(−.02, .01)	.499
	→ Engagement	−.01	.01	(−.02, .01)	.626
	→ Burnout	−.01	.01	(−.03, .01)	.181
	→ Intentions to Stay	−.02	.01	(−.04, .01)	.170
Indulgence	→ Opportunities to Learn	−0.00[a]	.01	(−.02, .02)	.931
	→ Engagement	.00	.01	(−.02, .03)	.774
	→ Burnout	−0.00[a]	.01	(−.02, .02)	.815
	→ Intentions to Stay	.02	.01	(.00, .04)	.091
Long Term Orientation	→ Opportunities to Learn	−.02	.02	(−.05, .02)	.348
	→ Engagement	.00	.02	(−.03, .03)	.994
	→ Burnout	−.01	.02	(−.05, .02)	.467
	→ Intentions to Stay	−0.00[a]	.02	(−.04, .03)	.911

Notes:
[1] Coefficients marked with * are statistically significant.
[a] The actual value is < −.001 but we report only 2 digits precision.
[b] Significance must be determined based on the confidence interval when using bootstrapping, not the p-value. While the p value is technically <.05, we cannot draw conclusions based on this. Further, the p value is very close to .05, which would further cast doubt on any suggestion of significance.

members plan to stay, as mentioned by one of the respondents who did not leave during the 90 day period following the survey (see Sec. 4):

> "It is a pretty good organization, very good atmosphere and working culture I feel very good to work with this company and not at all thinking to leave SoftTech... feeling happy to be a part of this..."

However, some respondents experienced burnout due to challenges within projects, despite the company's efforts to foster a healthy environment. When faced with limited opportunities to switch projects, members of the team felt 'stuck' and consider leaving SoftTech, as mentioned by one of the respondents (who ultimately left SoftTech during the 90-day period since the survey):

> "Even when SoftTech [as an organization] tries to keep a healthy environment, projects don't. My team has reported burnout all year and was encouraged to give extra effort, but the project always minimizes the efforts, and that situation lowers our enthusiasm. We are stuck in this kind of project with little chance to change to another project, so our only change is to leave SoftTech. That's the reason a lot of my coworkers left the company this year."

This quote clearly illustrates that whatever initiatives to support staff an organization might create at the executive level, it may be challenging for such initiatives to trickle down throughout the organization at the operational level. The observation that an organization is not a homogeneous and uniform environment, but that circumstances vary across departments or unit, is an important issue that we discuss further in Section 6.

5.2 Heterogeneity Across Groups

We now turn to the second research question to investigate any differences across subgroups of respondents; this section reports the results of several multi-group analyses.

5.2.1 Gender. Table 4 shows the results of a multi-group analysis between men and women; the parametric tests [38] showed a statistical difference between women and men for H1. The association between Leadership Support and Engagement is significant only for women, though we note a very small coefficient (*B*=.07).

Table 4. Multi-Group Analysis between men and women

	Women	Men
Sample size (N)	3,789	9,584
Opportunities to Learn R^2	.51	.53
Engagement R^2	.46	.46
Burnout R^2	.44	.46
Intention to Stay R^2	.32	.34
H1 Leadership Support → Engagement	**.07***	**.00**
H2 Leadership Support → Burnout	−.19*	−.16*
H3 Leadership Support → Opportunities to Learn	.31*	.32*
H4 Generative Organizational Culture → Engagement	.35*	.35*
H5 Generative Organizational Culture → Burnout	−.33*	−.32*
H6 Generative Organizational Culture → Opportunities to Learn	.46	.46
H7 Opportunities to Learn → Engagement	.34*	.39*
H8 Opportunities to Learn → Burnout	−.23*	−.27*
H9 Engagement → Intention to Stay	.34*	.33*
H10 Burnout → Intention to Stay	−.28*	−.31*

Notes:
[1] Coefficients marked with ∗ are statistically significant.
[2] Rows highlighted in gray indicate a significant difference between groups (women and men).

5.2.2 Tenure. Appendix C (see Table 16) presents the results of multi-group analysis for organizational tenure (in ranges 0-6 months, 6-12 months, 1-3 years, 3-5 years, and 5+ years). Parametric tests showed statistical differences for H9 and H10, referring to the relationship between engagement respectively burnout and intention to stay. New employees (<6m) exhibit a stronger relationship between engagement and intention to stay (H9, B=.37) than employees who have been working at the company for 1-3 years (B=.31). Employees who have been working at the company for 1-3 years (B=.32) exhibited a lower level of intention to stay when feeling burned out (H10) than those who have been working at the company for more than five years (B=.24).

One respondent with 1-3 years tenure who left SoftTech during the 90-day post-survey period commented they felt exhausted and frustrated. When achieving a tenure range of 1-3 years, employees may have surpassed the excitement about having a new job to start thinking about future goals and career progression as reasons to stay in the company, as mentioned by one respondent in this tenure range (who did not leave during the 90-day post-survey period):

> "When I was trying to break into this industry, [SoftTech] was the only company that gave me a chance to show what I'm capable of. During my time here, I feel like it tried hard to be the company I always wished I worked for. However, I still feel a little apprehensive about my future here. It feels like, for all the push to grow my career, the actions to be taken are all about developing soft skills without clear guidance on how to develop those soft skills, and a strong dependence on being lucky enough to be noticed in the crowd."

5.2.3 Country of residence. Appendix C (see Table 17) presents the results of a multi-group analysis across countries of residence. Parametric tests showed a statistical difference between employees in Argentina and Mexico (AR-MX) for H1. Leadership Support is positively associated with engagement only for employees in Argentina (B=.06). No other significant differences were observed. We note the coefficients are very low, and would argue these are not meaningful and significance is likely an effect from the large sample size.

5.2.4 Managers vs. Non-Managers. Table 5 presents the results of the multi-group analysis between job roles. There were no significant differences between managers and non-managers.

5.2.5 Age. Finally, Table 6 presents the results of the multi-group analysis between different age groups; as described earlier, we captured in ranges, rather than respondents' exact age. Table 6 presents the three groups that could be included after the MICOM tests (see Sec. 4.2.1). Again, we found no significant differences among these age groups.

5.3 Importance of Factors

The PLS-SEM analysis that addresses RQ1 sheds light on the magnitude of the effects of leadership support, opportunities to learn, organizational culture in engagement, burnout, and the intention to stay at the company. To address the question which of these factors is the most important, we report on an Importance-Performance Map Analysis (IPMA). The concept of an importance-performance analysis is not new or exclusive to PLS-SEM, but is a more general technique that can be traced back decades ago, and has been applied in many other domains [43, 66]. In the context of PLS-SEM, it maps the performance of the constructs (using construct scores representing the theoretical variables), against the importance of those constructs. The result is a matrix-like structure (see Figure 5) that can serve as a decision tool [43, 66, 90].

Figure 5 shows the impact-performance map for the constructs. The map has four quadrants, divided by the average importance (vertical line, at 0.31), and the average performance (horizontal

Table 5. Multi-Group Analysis between managers and non-managers

	Managers	Non-Managers
Sample size (N)	1,839	11,504
Opportunities to Learn R^2	.47	.48
Engagement R^2	.47	.46
Burnout R^2	.44	.45
Intention to Stay R^2	.31	.33
H1 Leadership Support → Engagement	.06	.02
H2 Leadership Support → Burnout	−.20*	−.17*
H3 Leadership Support → Opportunities to Learn	.33*	.32*
H4 Organizational Culture → Engagement	.35*	.35*
H5 Organizational Culture → Burnout	−.32*	−.32*
H6 Organizational Culture → Opportunities to Learn	.46*	.46*
H7 Opportunities to Learn → Engagement	.35*	.38*
H8 Opportunities to Learn → Burnout	−.23*	−.26*
H9 Engagement → Intention to Stay	.34*	.33*
H10 Burnout → Intention to Stay	−.27*	−.30*

Notes:
[1] Coefficients marked with ∗ are statistically significant.
[2] We observed no significant differences between managers and non-managers for any of the hypotheses.

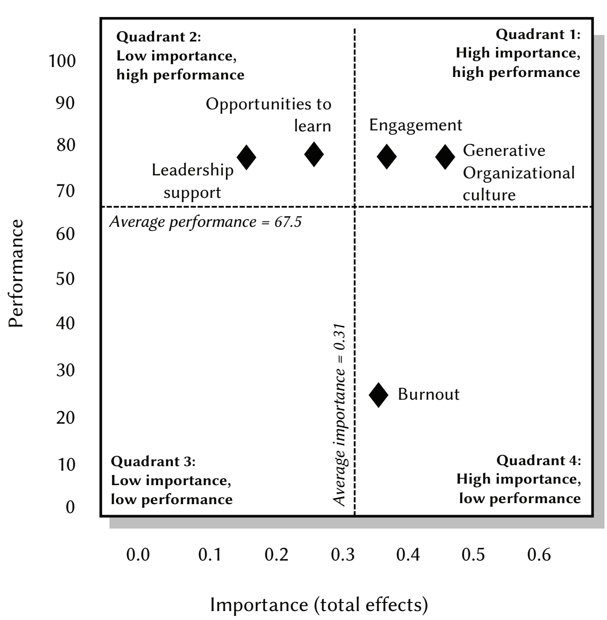

Fig. 5. Importance-Performance Map Analysis (IPMA) of constructs to the target construct Intention to Stay

Table 6. Multi-Group Analysis between age ranges: group 1: age 25-34, group 2: age 35-44, group 3: age 45+

	Groups 1 vs. 2		Group 1 vs. 3		Group 2 vs. 3	
	25-34	35-44	25-34	45+	35-44	45+
Sample size (N)	6,919	4,357	6,919	1,135	4,357	1,135
Opportunities to Learn R^2	.48	.47	.48	.50	.47	.50
Engagement R^2	.47	.44	.47	.48	.44	.48
Burnout R^2	.47	.43	.47	.46	.43	.46
Intention to Stay R^2	.33	.32	.33	.32	.32	.32
H1 Leadership Support → Engagement	.04*	.01	.04*	.02	.01	.02
H2 Leadership Support → Burnout	−.17*	−.17*	−.17*	−.22*	−.17*	−.22*
H3 Leadership Support → Opportunities to Learn	.31*	.33*	.31*	.34*	.33*	.34*
H4 Organizational Culture → Engagement	.34*	.37*	.34*	.31*	.37*	.31*
H5 Organizational Culture → Burnout	−.34*	−.32*	−.34*	−.33*	−.32*	−.33*
H6 Organizational Culture → Opportunities to Learn	.47*	.45*	.47*	.48*	.45*	.48*
H7 Opportunities to Learn → Engagement	.37*	.35*	.37*	.42*	.35*	.42*
H8 Opportunities to Learn → Burnout	−.26*	−.26*	−.26*	−.22*	−.26*	−.22*
H9 Engagement → Intention to Stay	.32*	.34*	.32*	.34*	.34*	.34*
H10 Burnout → Intention to Stay	−.31*	−.29*	−.31*	−.28*	−.29*	−.28*

Notes:
[1] Coefficients marked with ∗ are statistically significant.
[2] There were no statistically significant differences between age groups.
[3] 18-24 and 25-24, 18-24 and 35-44 and 18-24 and 45+ were not included because the MICOM test showed no compositional invariance (see Section 4.2.1)

line, at 67.5).[11] Constructs within Quadrant 1 have both a high importance, indicating a total effect that is larger than average, and high performance. Performance here is measured as the average construct score, re-scaled to a scale of 1-100. Martilla and James labeled Quadrant 1 as "keep up the good work," suggesting little need for intervention by decision makers [66]. Constructs in Quadrant 2 are characterized by low importance and high performance—Martilla and James characterized this as "possible overkill" [66]. That is, they have a similar performance as those in Quadrant 1, but have a modest or small effect. We note that of the three organizational job resources that we proposed as antecedents of engagement and burnout (see Hypotheses 1-3), leadership support and opportunities to learn fall within Quadrant 2. Constructs in Quadrant 3 are characterized as having low importance and low performance. None of the constructs in this study fell within this quadrant. Martilla and James labeled this quadrant as "low priority" [66]. Finally, constructs within Quadrant 4 are characterized as high importance and low performance; Martilla and James labeled this quadrant with a directive to "concentrate here" [66]. In this study, we observe that burnout performs considerably lower than average, but has a higher than average importance.[12] In conclusion, the importance-performance map analysis suggests decision makers direct their focus on preventing burnout.

[11] As pointed out by Henseler [43, p. 289], the placement of these lines that define the quadrants is up to the analyst, and should not be interpreted as a hard decision rule. Constructs fall within one of the quadrants, but, as Henseler commented, the analyst "*should not regard an attribute's placement in a certain quadrant as definite.*"
[12] We took the absolute value of the total effect of burnout, which SmartPLS reports to be −0.36; the sign of the effect depends on the hypothesis, but should be ignored to compare to other effects.

Table 7. Multi-group analysis between employees who left and employees who did not leave

	Employees who left	Employees who did not leave
Sample size (N)	474	12,869
Opportunities to Learn R^2	.61	.52
Engagement R^2	.47	.46
Burnout R^2	.45	.45
Intention to Stay R^2	.32	.33
H1 Leadership Support → Engagement	.03	.02
H2 Leadership Support → Burnout	**.03**	**−.17***
H3 Leadership Support → Opportunities to Learn	.27*	.32*
H4 Organizational Culture → Engagement	.32*	.35*
H5 Organizational Culture → Burnout	**−.47***	**−.32***
H6 Organizational Culture → Opportunities to Learn	.55*	.46*
H7 Opportunities to Learn → Engagement	.39*	.37*
H8 Opportunities to Learn → Burnout	−.28*	−.26*
H9 Engagement → Intention to Stay	.33*	.33*
H10 Burnout → Intention to Stay	−.28*	−.30*

Notes:
[1] Coefficients marked with ∗ are statistically significant.
[2] Rows highlighted in gray indicate a significant difference between groups (i.e. employees who left and employees who did not leave).

5.4 Difference Between Current vs. Past Employees

We now address Research Question 4, which seeks to establish whether the factors analyzed for RQ1 and RQ2 vary among those respondents who remained with the company and those who left in the 90-day period since the survey. Table 7 presents the results of a multi-group analysis. Parametric tests showed a statistical difference regarding employees who left and did not leave for H2 and H5. While current employees had a negative association between Leadership Support and Burnout (H2, B=−0.17), this association was not significant for those who left. Employees who left had a negative and 40.6% higher association between Organizational Culture and Burnout (H5) than current employees (B=−0.47 vs. −0.32).

These last findings resonate with the comment above, namely that people do not quit companies, but rather they quit teams. Compared to the findings for H2, this suggests that a generative organizational culture is far more important to avoid burnout than leadership support. As we noted earlier, while leadership could change more easily, changing the teams' culture could prove much more difficult.

5.5 Prediction of Attrition

We now turn to the last research question (RQ5) which focuses on the ability of the factors we have investigated to predict actual attrition. To answer this, we again rely on the additional data that indicated whether or not staff members left the organization during the 90-day period after the survey. As briefly outlined in Sec. 4.3, we trained and tested several ML classifiers, and analyzed strategies to balance the training data and to tune the ML hyperparameters.

We first balanced the training data using the SMOTE technique (Synthetic Minority Oversampling Technique) [14] and obtained a preprocessed dataset for model building. We then ran the models with default parameters of ML algorithms using the balanced data. Table 9 shows the accuracy, precision, recall, and F-measure obtained for each of the classifiers. The best performing ML algorithms based on the F-Measure were Random Forest and Decision Tree, which we selected for running the Randomized Search Cross Validation for hyperparameter tuning.

Both models were tuned using the Randomized Search. The Randomized Search Cross Validation algorithm identified and tested 30,000 different configurations for Decision Tree and 13,440 for Random Forest. Table 10 shows the results of Random Forest and Decision Tree after hyperparameter tuning to predict employees who left. Hyperparameters values selected by Randomized Search Cross Validation (CV) for Random Forest are presented in Table 8.

Table 8. Hyperparameter values for Random Forest and Decision Tree

Hyperparameter	Random Forest	Decision Tree
max_depth	50	50
n_estimators	99	–
bootstrap	False	–
ccp_alpha	–	9.043346969111224e-05
criterion	–	entropy
class_weight	–	balanced

After running procedures to verify and prevent overfit and procedures for hyperparameter tuning, Random Forest outperformed Decision Tree in all metrics for employees who left. The Random Forest precision slightly increased from 0.835 to 0.836, while the recall and F-Measure slightly decreased from 0.944 to 0.941 and 0.881 to 0.880 compared with the non-tuned model. However, the tuned model brings a more generalized model, able to outperform the original model in processing time, using less memory and storage [67].

Decision Tree suffered from the pruning procedure and was observed to have a high metric degradation, except for the precision for employees who left, being able to increase the precision from 0.792 to 0.821. Hence, we selected Random Forest to analyze *feature importance* of factors that predict attrition (see Table 11).

In ML, understanding the significance of features is important for identifying the best predictors. Feature importance, also known as feature detection, attribution, or model interpretability, is linked to statistical concepts like estimation and attribution. This process yields a specific score or metric, facilitating the ranking of features based on their contribution to the machine's predictions from largest to smallest. Typically, this involves systematically permuting features to assess the impact of each on predictive power. The result is an importance score for each feature, enabling the creation of a ranked list [75]. The closer a feature is to 1.0, the better the prediction capacity, while the features that have the highest rank have the most predictive power and are selected as inputs into the final model [96].

We analyzed importance of the feature to evaluate which variables in this study primarily impacted the attrition prediction. Table 11 presents the results. We found that Opportunities to Learn and Engagement had the top importance as features to predict attrition, followed by Organizational Culture, Leadership Support, Intention to Stay, and Engagement. We note that Burnout has the lowest score on feature importance, consistent with the results of the importance-performance analysis (see Fig. 5).

Table 9. Predictions metrics from Decision Tree and Random Forest after tuning using Randomized Search

	All employees				Employees who left		
Algorithm	Accuracy	Precision	Recall	F-Measure	Precision	Recall	*F*-Measure
RandomForest	**.991**	**.917**	.968	**.938**	**.835**	.944	**.881**
DecisionTree	.989	.895	**.976**	.930	.792	**.962**	.865
KNeighborsClassifier	.925	.657	.940	.718	.317	.956	.476
MLPClassifier	.960	.747	.959	.811	.495	.958	.644
GaussianNB	.445	.504	.526	.338	.039	.614	.073
LinearSVC	.569	.505	.536	.398	.041	.499	.076
LogisticRegression	.566	.504	.532	.396	.041	.496	.075
DummyClassifier	.503	.501	.506	.364	.036	.510	.068

Table 10. Predictions metrics from Decision Tree and Random Forest after tuning using Randomized Search

	All employees				Employees who left		
Algorithm	Accuracy	Precision	Recall	F-Measure	Precision	Recall	*F*-Measure
RandomForest	**.966**	**.917**	**.966**	**.938**	**.836**	**.941**	**.880**
DecisionTree	.954	.907	.901	.901	.821	.809	.809

Table 11. Feature importance of predictors based on Random Forest

Predictor		Feature Importance
EN	Engagement	.201
OL	Opportunities to Learn	.191
OC	Organizational Culture	.166
SL	Leadership Support	.156
IS	Intention to Stay	.144
BT	Burnout	.143

6 DISCUSSION AND CONCLUSION

Before we discuss the implications of the findings of this study, we first discuss a number of threats to validity that should be considered.

6.1 Threats to Validity

External Validity. We conducted this study at a single organization, SoftTech, and this may affect its external validity. There are two main reasons why this potential threat to validity may be limited. First, given the very large sample of respondents who were distributed across the globe, it is likely that different divisions of SoftTech have their own organizational sub-culture, each of which may be influenced by the national culture of the country where a division is located [48]. A multi-group analysis using country of residence did not show any meaningful variation in the results (see Table 17). The second reason is that, even if there is a specific "SoftTech culture" that sets the company apart from all other IT companies, respondents clearly had different perceptions of this supposed organizational culture as reflected in the variation of their scores. Some variation is necessary for the statistical procedures to generate a result. Notwithstanding, we cannot make any claims of generalizability beyond the population from which we sampled, which is software professionals at SoftTech. Future studies could replicate this study, or parts thereof, at other organizations or as a cross-sectional survey within the IT sector. It should be noted that the current study design included the additional recording of employees' employment status 90 days after the initial survey, and this would be very challenging to achieve in a sector-wide survey.

Internal Validity. The current analysis that is based on a cross-sectional survey dataset does not support causal claims. A few notes are in order regarding causality. First, the hypotheses we tested were posed as associations, rather than causal relationships. Establishing relationships such as these is meaningful because it allows managers to identify potentially important factors. Second, in most cases, the relationships align with common sense: it is reasonable to expect that leadership support might *cause* engagement (though, we found that it doesn't in the general case), rather than engagement causing leadership support; there is no good theoretical or practical justification why it is reasonable to believe that engaged employees would cause or lead to more supportive leadership. The same is true for the other antecedents: there is little reason to believe that either engagement or burnout would *cause* organizational culture to be more generative, or opportunities to learn to increase. Third, while we cannot prove causality with the current research design, it would be practically impossible to conduct experimental studies to establish these relationships. It is not possible to "vary" organizational culture in a controlled setting to assess the resulting level of burnout or engagement among employees. A potentially viable approach is to identify settings that act as a natural experiment, whereby two or more similar organizations are compared that would have distinct organizational cultures. Identifying those, however, could be quite challenging.

The decision made by SoftTech to utilize an internal survey system that requires employee authentication sacrifices anonymity and can also be seen as a threat to the validity of the results. For example, respondents could feel constrained in how truthfully they could really answer to certain questions. We acknowledge that without anonymity, respondents may feel inclined to provide answers that align with what they believe the company wants to hear, potentially leading to biased responses. However, respondents at SoftTech are used to non-anonymous surveys and were aware that managers and researchers would not have access to identifiable information and that the data would be aggregated to support company-wide research. Clearly, there is an element of trust involved. Additionally, to help mitigate the risk of biased responses, SoftTech made participation in the survey optional for employees. We note that to answer RQ4 and RQ5, we rely on the additional data that captured whether or not respondents left during the 90-day period after the survey; this data would not be available in an anonymous survey.

The decision to measure employee status after a period of 90 days was one of SoftTech's HR unit. Whether or not 90 days is an appropriate period is difficult to determine, because there appear to be very few studies in general that link intent to actual behavior. Notwithstanding, the IT sector generally faces a high degree of turnover, and long tenure with a single organization is unusual. In light of that, the choice of 90 days post-survey is not an unreasonable point in time to assess respondents' employment status.

A potential threat to the validity of this study arises from reliance on self-reporting surveys, introducing the risk of common method bias. Participants may provide responses influenced by social desirability, memory recall limitations, or personal biases, leading to inaccuracies in the data. The subjective nature of self-reporting surveys may compromise the precision and reliability of the measurements, impacting the overall validity of the study's findings.

While self-reporting methods may not possess the same level of rigor as carefully crafted observational surveys when it comes to measuring behavior, they come with several advantages, which led us to make this trade-off. The gathering of data using self-reporting methods is notably more cost-effective, scalable, quicker to execute, and offers insights that may not be directly attainable through observation—specifically, information regarding the respondent's knowledge, attitudes, and opinions [78].

Although researchers should acknowledge and consider this potential source of measurement error when interpreting and generalizing the results, we used actual (objective) attrition data to drive conclusions about the importance of burnout, rather than self-reporting. Notwithstanding, one method to measure the validity of survey designs is to re-test [12], i.e., to repeat the survey after a period of time, and this is something that the company could undertake.

One notable limitation of this study revolves around the temporal aspect inherent in the data collection process. The study spanned a three-month period to collect attrition data, during which respondents' opinions, experiences, and circumstances may have undergone changes. This temporal evolution introduces a potential source of bias or inaccuracy in the categorization of respondents into groups. Since the group analysis was conducted after the initial 90 days of data collection, it is plausible that shifts in respondents' perspectives occurred during this interval.

Construct validity. This study incorporated several latent variables; to measure these, we adopted and tailored existing measurement instruments when possible and developed measurement instruments for some constructs based on prior literature. The evaluation of the measurement model (see Sec. 4.4) suggests that the reliability of these constructs is good. However, apart from a quantitative assessment of such instruments, it is also good practice to consider their face validity. For each construct, the items that were used to measure that construct can be inspected (see Table 14 in Appendix B). Consider the construct Leadership Support (LS); the argument is that for a respondent to experience a high level of LS, that respondent's score on each of the 5 items would be high. There is a natural variation across respondents: for some, their leaders' recognition of work (item LS5) is more important, whereas for others, leaders' caring about their well-being (item LS4) is more important. The items that we used were, as mentioned, adopted and adapted from existing items used in other studies, but it is equally possible to use a different set of items to measure the same construct. It is not possible to determine whether one instrument can measure a theoretical concept better than another.

In this study we used PLS-SEM instead of covariance-based SEM (CB-SEM). In CB-SEM, the assumption is that a set of items together are observable indicators of an unmeasurable construct, which is statistically measured as a 'common factor'; i.e., a common factor is 'extracted' from the covariance of the items. The assumption underpinning this is that a change in the construct *causes* change in each of the items. For example, if 'leadership support' would increase, then that would mean that all items (LS1 to LS5) would also increase (though some more than others, according to the loadings, see Table 14). In such a model, the items 'reflect' the latent variable. In this study, however, we found that the assumption of a common factor is not tenable. At least some of the instruments that came from previous studies appear to have been analyzed previously using principal component analysis (PCA), which is not a factor analysis method, but rather identifies composites; this is closer to PLS, which itself is based on PCA [56, p.viii]. Thus, we used PLS as an analysis method that relies on composites to serve as 'proxies' to represent the latent variables [89].

6.2 Implications for Practice

The findings of this study have very actionable implications for practice. Table 12 presents a summary of the findings to the five research questions.

This study considers three organizational job resources as antecedents of engagement and burnout. First, we considered Leadership Support as an antecedent of Engagement (Hypothesis 1). The dataset as a whole does not lend support to this hypothesis, but as mentioned in Sec. 5.1, respondents shared several comments in relation to bad leadership. Such comments suggest that bad leadership can drive people to quit, and that a perceived bad leadership affects how people perceive the organization as a whole. In other words, leaders are the representatives of the organization as a whole. This, of course, makes sense given that an organization follows a strategy that is set out by leadership. And yet, leadership was not perceived as a constant, and that new leadership could quickly recalibrate people's opinions about the organization as a whole.

While the dataset as a whole does not lend support to H1, we did find that H1 is supported for women though with a very small coefficient (*B*=.07). There is considerable evidence that women face extensive barriers that inhibit their career advancement [114]. These results align with recent studies that showed when women have a supportive leader they are more engaged and innovative [51]. Leaders' supportive actions can lead to a sequence of 'small wins,' effectively breaking down larger systemic barriers and enhancing engagement [41]. However, given the rather small coefficient, we should question whether this result is meaningful.

Among different categories of respondents, we found that H1 was supported only for those with 1-3 years working at the organization, and also by those in the age bracket 25-34. Perhaps leadership support has a more pronounced impact on relatively new staff, whereas those who have just joined (<6 months) have not had enough time to evaluate the leadership support. This sense of perceived leadership support may fade over time. We note, however, the very small coefficient (*B*=.04) and the fact that over half of respondents fell in this tenure category. Overall, these results do not appear particularly meaningful.

Table 12. Summary of findings

	RQ1	RQ2 (MGA)					RQ4 (MGA)
Hypothesis	All data	Gender	Tenure	Country	Role	Age	Left or stayed
H1. Leadership Support → Engagement	Not significant	Women only	1-3y only	Argentina and Colombia only	Not significant	25-34 only, not for 35-44 and 45+	Not significant
H2. Leadership Support → Burnout	Yes	All	All	All	All	All	Only those who did not leave
H3. Leadership Support → Opportunities to Learn	Yes	All	All	All	All	All	All
H4. Organizational Culture → Engagement	Yes	All	All	All	All	All	All
H5. Organizational Culture → Burnout	Yes	All	All	All	All	All	All, but stronger effect for those who left
H6. Organizational Culture → Opportunities to Learn	Yes	All	All	All	All	All	All
H7. Opportunities to Learn → Engagement	Yes	All	All	All	All	All	All
H8. Opportunities to Learn → Burnout	Yes	All	All	All	All	All	All
H9. Engagement → Intention to Stay	Yes	All	All, stronger effect for group 0-6m than for 1-3y	All	All	All	All
H10. Burnout → Intention to Stay	Yes	All	All, stronger effect for group 1-3y than for group 5+y	All	All	All	All

	RQ3 (IPMA)		RQ5 (ML)
Construct	Performance	Importance (Total Effect)	Feature Importance for actual retention
Leadership Support	77.93	.14	.156
Opportunities to Learn	78.30	.23	.191
Generative Organizational Culture	77.72	.45	.166
Engagement	77.32	.38	.201
Burnout	26.14	.36	.143
Intention to stay	n/a	n/a	.144

We also found support for H1 for those respondents based in Argentina and Colombia, but with a rather moderate coefficient (*B*=.06). Again, we note that over 5,900 respondents, or over 44% of the total sample, are based in these two countries (see Table 1). Like above, we suggest this significance is due to the high power and that this effect is not meaningful.

More clear evidence is lend to H2; a higher level of leadership support implies a lower level of burnout. Interestingly, a multi-group analysis indicates that this relationship holds only for those who did *not* leave, and not for those who left during the 90-day period after the survey. Exactly why this is remains an open question.

The data lend support to Hypotheses H3 and H4, without any discrimination regarding different subgroups (studied as part of RQ2 and RQ4). The same is true for H5, though we observe a significantly stronger effect for those who left than those who stayed; i.e., as respondents perceived the organizational culture to be less generative, they experienced more burnout. Hypotheses H6, H7, and H8 were also supported without variation across subgroups.

Hypotheses H9 and H10, investigating the links between engagement respectively burnout, and intention to stay, are also supported by the data without much distinction between subgroups. We only note a significantly stronger effect for those with 1-3 years tenure than those with over 5 years tenure.

Companies should carefully evaluate how to improve engagement. The term ‘Quiet Quitting’ [20] refers to ceasing to be fully committed to one’s job and doing just enough to meet the requirements of one’s job description [3]. The disruptions caused by the Covid-19 pandemic, including blurred boundaries between work and personal life, have fueled burnout and chronic disengagement [59]. Gallup reports a decline in U.S. employee engagement during the second quarter of 2022, with ’quiet quitters’ constituting at least 50% of the U.S. workforce. Additionally, only a quarter of employees feel connected to their organizational culture, and roughly one in three feel a sense of belonging in their organization [40].

Quiet quitting is influenced by various factors, with the decline in organizational trust being a primary catalyst. Deteriorating trust in leaders and organizations has reached a point where employees prefer trusting strangers over their bosses. This erosion of trust is closely linked to employee commitment, affecting organizational success [62].

We measured trust as part of organizational culture [121]. Our results showed the importance of organizational culture as the second most critical factor associated with intention to stay and the third most important feature to predict attrition.

The relevance of organizational culture was even highlighted when looking at the employees who left (*B*=−0.46 in Table 7), and also for all demographics. SoftTech created a program to have leaders at all levels (team leaders, product and project managers) being trained during a series of bootcamps with sessions about SoftTech’s processes, recognition practices, social awareness, how to talk to the team, how to communicate bad news, and creating psychological safety for members to express their opinions. Moreover, team culture is going to be part of training for all employees involving practices from Westrum’s typology [121] that will seek to create a generative culture that fosters information flow and trust.

Our finding that leadership support was positively associated with engagement for women (though not men), organizations should also focus on training about inclusivity leadership skills to help retain women and decrease the gender gap. Another significant factor for “quiet quitting” is the lack of commitment to career development, where employees feel dissatisfied due to an absence of employer dedication to personal and professional growth [62].

The noteworthy phenomenon of “quiet firing” has garnered increased attention, involving the obstruction of growth opportunities and the neglect of timely feedback to foster an unfavorable work environment, subtly pushing employees toward resignation [3]. Our research findings affirm the significance of these factors, with "supportive leadership" emerging as the second most crucial construct influencing the intention to stay (refer to Section 5.3, Table 7). Moreover, our results underscored that "opportunities to learn" was identified as the most pivotal feature influencing attrition (see Section 5.5). To effectively mitigate attrition, organizations are advised to prioritize the development of supportive leadership styles and ensure ample opportunities for continuous learning. Fostering transparent communication channels and constructive feedback mechanisms can further enhance employee satisfaction and commitment, thereby mitigating the occurrence of "quiet firing."

This finding aligns with an extensive body of work that has established the link between engagement with retention and burnout with attrition. However, the number of studies that test whether the intention to stay translates into actual turnover is far smaller because reliable data on quitting and staying behavior is very challenging to acquire. In this study, we had access to such data. The feature importance analysis (Table 11) showed that perceiving (a lack of) opportunities to learn was the most important feature in predicting people's decision to remain in (or quit) their jobs, closely followed by engagement and burnout. Gamification has been researched as a practice to increase engagement in workplace [33]. When engaging in games, whether card games, board games, sports, or video games, we typically link the experience of playing with positive feelings such as having fun, enjoying social interaction, or feeling motivated to achieve specific goals.

6.3 Implications for Research

This study suggests different links between antecedents and consequences of the two opposite psychological states of burnout and engagement for software delivery teams. Further, the ML analysis detected that engagement was the most important feature in predicting intention to leave and attrition. Future work could explore other antecedents to engagement and burnout, such as compensation and extrinsic rewards, and different consequences, such as productivity and software quality.

The present study relied on turnover data, collected 90 days after the initial survey. This was only possible because the study was conducted at a single organization, which would have precise and reliable data on this. Thus, while the study context was limited to a single organization which poses a threat to external validity (see Sec. 6.1), expanding this line of research to a cross-section of the IT industry is very challenging indeed, because it is very difficult to obtain reliable turnover data. Most studies of turnover in the software engineering field focus on open source communities, where turnover is operationalized as "absence of contributions" for a certain number of days, for example, 180 days [57]. While this is not unreasonable, this is not a fully reliable measure, nor does it fully capture 'leaving' as open source developers may still be "lurking," while no longer actively contributing. We are aware of only one study of turnover in two companies [8]; however, the two studied companies did not keep a record of developers' departure and also relied on an 'absence of activity' to measure turnover.

We found different associations between leadership support and engagement between genders and countries, and different associations between both engagement and burnout to intention to stay across tenure. Leadership Support is positively associated with engagement only for women (not men), and employees from Argentina and Colombia (not the other countries). Novice employees (less than 6 months) are more likely to stay when are engaged, while employees who have been working at SoftTech for one to three years are more likely to stay when they are less burnt out than those who have been there for over five years. To delve into the nuanced dynamics of actions to increase engagement and mitigate burnout, a comprehensive longitudinal analysis can be employed. The study design can involve baseline assessments of the factors associated with engagement and burnout, and subsequent assessments at regular intervals to track changes over time.

6.4 Conclusion

Attention to human factors is critical to software delivery teams' sustainability. We report on a theoretical model that takes the Job Demands-Resources (JD-R) model as a point of departure, and posits three organizational job resources as antecedents of burnout and engagement. We further expand the JD-R model by positing intention to stay as a consequence. A large-scale survey with over 13,000 respondents provided data to test our hypotheses for the whole dataset, and for several different cohorts, distinguishing different categories of tenure, age, country of residence, and gender. Using additional information about the people who left the organization within a 90-day period after the survey, we further investigated differences between past and current employees. Finally, we develop a Machine Learning classifier with good performance that is able to predict who would leave.

The results obtained from the Machine Learning analysis represent a starting point in the development of increasingly efficient employee attrition classifiers. Longitudinal studies and interviews with people who left can bring additional information, improve the overall knowledge of the reasons to leave SoftTech and, consequently, increase the time available to personnel departments to assess and plan the tasks required to mitigate this risk (e.g., retention actions, prepare for turnover, and task redistribution).

Given the international nature of this study, albeit at one company, the findings are of interest to other large organizations. There are clear extension points of our study and opportunities to replicate it, which can contribute to a body of knowledge that considers critical human factors such as engagement and burnout.

ACKNOWLEDGMENTS

This work was supported by Science Foundation Ireland grant no. 15/SIRG/3293 and 13/RC/2094-P2 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero—the Irish Software Research Centre (www.lero.ie).

A SUMMARY OF PRIOR STUDIES ON BURNOUT AMONG IT PROFESSIONALS

Table 13. Prior work on burnout in software development teams

Study	Proposed Antecedents	Proposed Consequences	Method	Key Findings
Fujigaki et al. 1994 [25]	Job overload, Project management, Mental rewards, Job latitude, Communication with users, Career development, Technical difficulties, Work environment	n/a	Sample of 2,296 IS managers (n=300, of which 296 male, 4 female) and 1,996 engineers. Principal Component Analysis yielding 8 factors. Multiple regression. One-way ANOVA to investigate effects of time management.	Project management, mental rewards, and job overload had a significant effect on depressive symptoms. Work environment obstacles were not significant.

Continued on next page

Table 13 *Continued*

Study	Antecedents	Consequences	Method	Findings
Sonnentag et al. 1994 [108]	Work stress, control at work, high requirements (cognitive, learning, communication), complexity of work, and high quality of social interaction (democracy, openness to criticism, competition, dominance).	n/a	Sample of 200 respondents (75% male; full data for 166) from 29 software development projects from 19 companies in Germany and Switzerland. Principal Component Analysis.	Work stress is correlated with burnout; partial support for the relationship between control at work, high requirements, high quality of social interaction, and burnout.
Moore 2000 [73]	Perceived workload, role ambiguity, role conflict, autonomy, fairness of rewards.	Turnover intention	Sample of 252 IT professionals (69% male), covariance-based structural equation modeling (CB-SEM)	Work exhaustion (burnout) partially mediates workplace factors on turnover intention. Work overload was the strongest contributor to exhaustion in IT workers. Insufficient staff and resources is a primary cause of work overload and exhaustion.
Hsieh and Chao 2004 [50]	Job specialization, Job rotation	n/a	Sample of 304 valid responses (185 male, 119 female) from high-tech industry employees in Taiwan. Multiple hierarchical regression analysis.	Job specialization and Job rotation have a negative relation to exhaustion.
Schoepke et al. 2004 [101]	IT demands, Role ambiguity, Decision control, Challenge[a]	n/a	Sample of 624 IT professionals from five companies (54% male). Regression analysis.	IT demands is significantly correlated with fatigue in both men and women. Role ambiguity and Decision control were also significantly correlated with fatigue for women, but not for men. IT demands, role ambiguity, and challenge were predictors for burnout; decision control only for women.
Shropshire and Kadlec 2012 [106]	n/a	Intention to leave the IT field	Sample of 65 IT workers (60% male) in a medium-sized public service organization in the US. PLS structural equation modeling.	Job burnout is linked to an intention to change career. Age did not moderate the relationship.

Continued on next page

Table 13 *Continued*

Study	Antecedents	Consequences	Method	Findings
Shih et al. 2013 [105]	Variables identified from prior work were included to model 'control relationships': Perceived workload, role ambiguity, role conflict, autonomy, fairness of rewards, emotional dissonance	Job satisfaction, Depersonalization, and indirectly Personal accomplishment.	Sample of 504 IT workers (291 male, 213 female) in the Taiwanese manufacturing sector. PLS structural equation modeling.	Primary focus on Job satisfaction and Depersonalization, and Personal accomplishment; Job satisfaction is a negative outcome of work exhaustion; depersonalization goes up, and as a result personal accomplishment is reduced.
Atouba and Lammers 2018 [7]	Internal Communication Adequacy (ICA): quality of internal communication between leadership and employees; Employee Work Participation (EWP): extent to which employees are involved in the design and structuring of their organizations	n/a	Sample of 111 respondents (no gender information) at Technology Management Services, USA. Principal Component Analysis. Hierarchical multiple linear regression.	Negative relationship between EWP and emotional exhaustion, but not between ICA and emotional exhaustion.
Mahapatra and Pati 2018 [63]	Techno-overload, Techno-invasion, Techno-complexity, Techno-insecurity, Techno-uncertainty	n/a	Sample of 163 (129 male, 34 female), of which 133 enrolled in a 2-year public management institute in India, and 30 employees across sectors through personal contacts. Multiple regression.	Techno-invasion and Techno-insecurity are significantly related to burnout.
Trinkenreich et al. 2023 [116]	Work satisfaction	n/a	Sample of 3,281 responses (2,487 male) from Globant. PLS structural equation modeling.	Work satisfaction is significantly negatively correlated with burnout.

Note:

[a] The paper does not specify the items, nor does it elaborate what 'challenge' means.

B ADDITIONAL TESTS OF DISCRIMINANT VALIDITY

Table 14 presents the cross-loadings of all items onto the constructs. The cross-loadings could indicate any issues with discriminant validity. As the table shows, the loadings of all items is highest for the constructs that they purport to measure.

Table 15 presents the correlations among constructs, with the square roots of the AVE on the diagonal in boldface. The Fornell-Larcker criterion for discriminant validity suggests that the square root of the AVE values must be larger than the constructs among the constructs. This is indeed the case in this study. We note that the Fornell-Larcker criterion is merely a heuristic and has been criticized; we report it here for completeness.

Table 14. Cross-loadings of the retained indicators on the constructs

Item	Description	LS	OC	OL	EN	BT	IS
Leadership Support (LS)							
LS1	Leaders encourage healthy balance between personal and professional activities	**.775**	.554	.482	.411	−.542	.358
LS2	I would work with my leaders again	**.812**	.582	.530	.441	−.461	.412
LS3	My leaders and I have meaningful conversations about my career interests and how to reach my career goals	**.751**	.513	.528	.370	−.372	.330
LS4	My leaders care about my well-being	**.830**	.580	.495	.387	−.448	.375
LS5	My leaders recognize and value my work	**.797**	.595	.537	.415	−.416	.378
Organizational Culture (OC)							
OC1	Failures are seen as learning opportunities	.508	**.693**	.493	.396	−.425	.356
OC2	I am empowered to make decisions needed to my job	.519	**.728**	.489	.467	−.417	.333
OC3	I feel encouraged to come up with innovative and disruptive solutions	.511	**.751**	.540	.513	−.487	.397
OC4	I feel safe speaking up and taking risks	.475	**.707**	.433	.406	−.418	.309
OC5	Responsibilities are shared in my team	.532	**.691**	.456	.421	−.439	.326
OC6	There is good teamwork and cooperation between different areas	.485	**.682**	.510	.424	−.459	.385
Opportunities to Learn (OL)							
OL1	I am given different learning experiences and tools to continue boosting my current skills and learning new ones	.484	.525	**.771**	.499	−.446	.389
OL2	I believe I have enough opportunities to develop and grow my career	.552	.582	**.851**	.526	−.517	.546
OL3	I feel my performance results and commitment contribute to my career development	.538	.569	**.809**	.500	−.471	.451
Engagement (EN)							
EN1	I am enthusiastic about my job	.488	.552	.590	**.856**	−.566	.498
EN2	I wake up energized in the morning and am ready to begin a new work day	.397	.495	.485	**.806**	−.644	.416
EN3	Time flies when I'm working	.341	.453	.430	**.765**	−.405	.359
Burnout (BT)							
BT1	I'm becoming less interested in work	−.396	−.437	−.495	−.603	**.681**	−.469
BT2	I feel I can blend personal professional activities in healthy way (r)*	−.494	−.535	−.476	−.470	**.791**	−.385
BT3	I feel mentally and physically exhausted from work	−.338	−.358	−.344	−.434	**.745**	−.324
BT4	I feel my current workload is manageable (r)*	−.427	−.462	−.386	−.375	**.714**	−.354
BT5	I feel well physically, mentally and spiritually (r)*	−.421	−.488	−.457	−.578	**.770**	−.371
Intention to Stay (IS)							
IS1	I rarely think about looking for a job at another company	.351	.372	.420	.389	−.387	**.830**
IS2	I see myself working at this company for the next year	.452	.481	.560	.519	−.503	**.906**

Note:

* (r) indicates the item is reverse-coded.

C MULTI-GROUP ANALYSES OF ORGANIZATIONAL TENURE AND COUNTRY OF RESIDENCE

Tables 16 and 17 present the results of multi-group analyses on organizational tenure and country of residence, respectively.

Table 15. Fornell-Larcker criterion: correlations among the constructs

Variable	BT	EN	IS	OL	OC	LS
Burnout (BT)	**.74**					
Engagement (EN)	−0.67	**.81**				
Intention to Stay (IS)	−0.52	0.53	**.87**			
Opportunity to Learn (OL)	−0.59	0.63	0.57	**.81**		
Organizational Culture (OC)	−0.62	0.62	0.49	0.69	**.71**	
Leadership Support (LS)	−0.57	0.51	0.47	0.65	0.71	**.79**

Note:
[1] Square roots of AVE values are in boldface on the diagonal.

Table 16. Multi-Group Analysis between organizational tenure ranges: group 1: <6 months; group 2: 6m-1y; group 3: 1-3y; group 4: 3-5y; group 5: 5+year.

	Comparison to group <6m						Comparison to group 6m+						1-3y vs. 5y	
	<6m	6m-1y	<6m	1-3y	<6m	5+y	6m-1y	1-3y	6m-1y	3-5y	6m-1y	5+y	1-3y	5+y
Sample size (N)	1,749	2,776	1,749	5,598	1,749	1,598	2,776	5,598	2,776	1,568	2,776	1,598	5,598	1,598
Opportunities to Learn R^2	.49	.47	.49	.47	.49	.48	.47	.47	.47	.49	.47	.48	.47	.48
Engagement R^2	.45	.47	.45	.47	.45	.46	.47	.47	.47	.43	.47	.46	.47	.46
Burnout R^2	.42	.47	.42	.46	.42	.43	.47	.46	.47	.44	.47	.43	.46	.43
Intention to stay R^2	.33	.35	.33	.33	.33	.31	.35	.33	.35	.31	.35	.31	.33	.31
H1. Leadership support → Engagement	−.01	.05	−.01	.04*	−.01	.02	.05	.04*	.05	-.02	.05	.02	.04*	.02
H2. Leadership support → Burnout	−.16*	−.19*	−.16*	−.19*	−.16*	−.14*	−.19*	−.19*	−.19*	−.16*	−.19*	−.14*	−.19*	−.14*
H3. Leadership support → Opportunities to learn	.34*	.32*	.34*	.30*	.34*	.33*	.32*	.30*	.32*	.34*	.32*	.33*	.30*	.33*
H4. Organizational culture → Engagement	.31*	.35*	.31*	.34*	.31*	.35*	.35*	.34*	.35*	.40*	.35*	.35*	.34*	.35*
H5. Organizational culture → Burnout	−.29*	−.33*	−.29*	−.32*	−.29*	−.32*	−.33*	−.32*	−.33*	−.35*	−.33*	−.32*	−.32*	−.32*
H6. Organizational culture → Opportunities to learn	.45*	.45*	.45*	.47*	.45*	.46*	.45*	.47*	.45*	.46*	.45*	.46*	.47*	.46*
H7. Opportunities to learn → Engagement	.42*	.36*	.42*	.37*	.42*	.37*	.36*	.37*	.36*	.33*	.36*	.37*	.37*	.37*
H8. Opportunities to learn → Burnout	−.27*	−.25*	−.27*	−.25*	−.27*	−.28*	−.33*	−.32*	−.25*	−.23*	−.25*	−.28*	−.25*	−.28*
H9. Engagement → Intention to stay	.37*	.34*	**.37***	**.31***	.37*	.37*	.34*	.31*	.34*	.31*	.34*	.37*	.31*	.37*
H10. Burnout → Intention to stay	−.26*	−.30*	−.26*	−.32*	−.25*	−.24*	−.30*	−.31*	−.30*	−.30*	−.30*	−.24*	**-.32***	**-.24***

Notes:
[1] Coefficients marked with ∗ are statistically significant.
[2] Coefficients highlighted in gray and set in boldface show a significant difference between groups.
[3] Groups 1-3y and 3-5y, 3-5y and 5+, and <6m and 3-5y were not included because the MICOM test showed no compositional invariance (see Sec. 4.2.1).

Table 17. Multi-Group Analysis across the top-5 countries of residence.

	AR	CO	AR	IN	AR	MX	AR	CL	CO	IN	CO	CL	MX	IN	IN	CL
Sample size (N)	3,014	2,906	3,014	2,366	3,014	1,655	3,014	743	2,906	2,366	2,906	743	1,655	2,366	2,366	743
Opportunities to Learn R^2	.54	.51	.54	.54	.54	.53	.54	.50	.51	.54	.51	.50	.53	.54	.54	.50
Engagement R^2	.48	.46	.48	.47	.48	.46	.48	.45	.46	.47	.46	.45	.46	.47	.47	.45
Burnout R^2	.44	.49	.44	.48	.44	.43	.44	.42	.49	.48	.49	.42	.43	.48	.48	.42
Intention to stay R^2	.32	.33	.32	.38	.32	.30	.32	.34	.33	.38	.33	.34	.30	.38	.38	.34
H1 Leadership Support → Engagement	.06*	.06*	.06*	.00	**.06***	**-.03**	.06*	.02	.06*	.00	.06*	.02	−.03	.00	.00	.02
H2 Leadership Support → Burnout	−.19*	−.19*	−.19*	−.13*	−.19*	−.13*	−.19*	−.17	−.19*	−.13*	−.19*	−.17*	−.13*	−.13*	−.13*	−.17*
H3 Leadership Support → Opportunities to Learn	.31*	.31*	.31*	.30*	.31*	.33*	.31*	.25*	.31*	.30*	.31*	.25*	.33*	.30*	.30*	.25*
H4 Organizational Culture → Engagement	.33*	.31*	.33*	.37*	.33*	.37*	.33*	.31*	.31*	.37*	.31*	.31*	.37*	.37*	.37*	.31*
H5 Organizational Culture → Burnout	−.32*	−.34*	−.33*	−.35*	−.33*	−.31*	−.32*	−.27*	−.34*	−.35*	−.34*	−.27*	−.31*	−.35*	−.35*	−.27*
H6 Organizational Culture → Opportunities to Learn	.48*	.46*	.48*	.48*	.48*	.46*	.48*	.51*	.46*	.48*	.46*	.51*	.46*	.48*	.48*	.51*
H7 Opportunities to Learn → Engagement	.37*	.39*	.37*	.37*	.37*	.39*	.37*	.40*	.39*	.37*	.39*	.40*	.39*	.37*	.37*	.40*
H8 Opportunities to Learn → Burnout	−.22*	−.26*	−.22*	−.29*	−.22*	−.29*	−.22*	−.30*	−.26*	−.29*	−.26*	−.30*	−.29*	−.29*	−.29*	−.30*
H9 Engagement → Intention to Stay	.34*	.30*	.34*	.34*	.34*	.29*	.34*	.36*	.30*	.34*	.30*	.36*	.29*	.34*	.34*	.36*
H10 Burnout → Intention to Stay	−.28*	−.33*	−.28*	−.32*	−.28*	−.31*	−.28*	−.29*	−.33*	−.32*	−.33*	−.29*	−.31*	−.32*	−.32*	−.29*

Notes:
[1] AR=Argentina; CO=Colombia; IN=India; MX=Mexico; CL=Chile.
[2] Coefficients marked with ∗ are statistically significant.
[3] Coefficients highlighted in gray and set in boldface show a significant difference between groups.
[4] CO-MX was not included because the MICOM test showed no compositional invariance (see Sec. 4.2.1)

D IMPORTANCE-PERFORMANCE MAP ANALYSIS

Table 18 presents the importance and performance scores of all but the target construct (Intention to Stay) in the research model.

Table 18. Importance-Performance Map Analysis (IPMA) for the target construct Intention to Stay

Construct	Performance	Effect
Burnout	26.14	.30
Engagement	77.32	.33
Generational Organizational Culture	77.72	.30
Opportunities to Learn	78.30	.20
Leadership Support	79.93	.12

REFERENCES

[1] Tanay Agrawal. 2021. Hyperparameter optimization using scikit-learn. *Hyperparameter optimization in machine learning: make your machine learning and deep learning models more efficient* (2021), 31–51.

[2] D. Alao and A.B. Adeyemo. 2013. Analyzing employee attrition using decision tree algorithms. *Computing, Information Systems, Development Informatics and Allied Research Journal* 4, 1 (2013), 17–28.

[3] Amitabh Anand, Jessica Doll, and Prantika Ray. 2023. Drowning in silence: a scale development and validation of quiet quitting and quiet firing. *International Journal of Organizational Analysis* (2023).

[4] Sigrún Andradóttir. 2014. A review of random search methods. *Handbook of Simulation Optimization* (2014), 277–292.

[5] Narallynne Araújo, Tiago Massoni, Camila Sarmento, Francielle Santos, and Ruan Oliveira. 2022. Investigating the Relationship between Software Team Leadership Styles and Turnover Intention. In *Proceedings of the XXXVI Brazilian Symposium on Software Engineering*. 106–111.

[6] Linda Argote, Bill McEvily, and Ray Reagans. 2003. Managing Knowledge in Organizations: An Integrative Framework and Review of Emerging Themes. *Management Science* 49, 4 (April 2003), 571–582.

[7] Yannick C. Atouba and John C. Lammers. 2018. Examining the relationships between participative organisational communication practices and burnout among IT professionals. *Total Quality Management & Business Excellence* 31, 7-8 (March 2018), 814–828. https://doi.org/10.1080/14783363.2018.1447367

[8] Lingfeng Bao, Zhenchang Xing, Xin Xia, David Lo, and Shanping Li. 2017. Who will leave the company?: a large-scale industry study of developer turnover by mining monthly work report. In *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*. IEEE, 170–181.

[9] Bernard M Bass. 1990. From transactional to transformational leadership: Learning to share the vision. *Organizational dynamics* 18, 3 (1990), 19–31.

[10] Sarah Beecham, Nathan Baddoo, Tracy Hall, Hugh Robinson, and Helen Sharp. 2008. Motivation in Software Engineering: A systematic literature review. *Information and software technology* 50, 9-10 (2008), 860–878.

[11] Max Bramer. 2007. Avoiding Overfitting of Decision Trees. In *Principles of Data Mining*. Springer London, London, 119–134.

[12] Nancy D Brener, John OG Billy, and William R Grady. 2003. Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: evidence from the scientific literature. *Journal of adolescent health* 33, 6 (2003), 436–457.

[13] Marie Carasco-Saul, Woocheol Kim, and Taesung Kim. 2015. Leadership and employee engagement: Proposing research agendas through a review of literature. *Human Resource Development Review* 14, 1 (2015), 38–63.

[14] Francisco Charte, Antonio J Rivera, María J del Jesus, and Francisco Herrera. 2015. MLSMOTE: approaching imbalanced multilabel learning through synthetic instance generation. *Knowledge-Based Systems* 89 (2015), 385–397.

[15] Evangelia Demerouti, Arnold B Bakker, Friedhelm Nachreiner, and Wilmar B Schaufeli. 2001. The job demands-resources model of burnout. *Journal of Applied psychology* 86, 3 (2001), 499.

[16] Klajkó Dóra, Restás Péter, Szabó Zsolt Péter, and Czibor Andrea. 2019. The Effect of Organizational Culture on Employee Well-Being: Work-Related Stress, Employee Identification, Turnover Intention. *Journal of International Cooperation and Development* 2, 2 (2019), 19–19.

[17] Francesca Fallucchi, Marco Coladangelo, Romeo Giuliano, and Ernesto William De Luca. 2020. Predicting employee attrition using machine learning techniques. *Computers* 9, 4 (2020), 86.

[18] Alan Felstead, Duncan Gallie, Francis Green, and Hande Inanc. 2015. Fits, misfits and interactions: Learning at work, job satisfaction and job-related well-being. *Human Resource Management Journal* 25, 3 (2015), 294–310.
[19] Peter A Flach and Meelis Kull. 2015. Precision-Recall-Gain Curves: PR Analysis Done Right.. In *NIPS*, Vol. 15.
[20] Sandro Formica and Fabiola Sfodera. 2022. The Great Resignation and Quiet Quitting paradigm shifts: An overview of current situation and future research directions. *Journal of Hospitality Marketing & Management* 31, 8 (2022), 899–907.
[21] Nicole Forsgren and Jez Humble. 2016. The role of continuous delivery in IT and organizational performance. In *Proceedings of the Western Decision Sciences Institute (WDSI)*.
[22] Nicole Forsgren, Jez Humble, Gene Kim, A Brown, and N Kersten. 2018. Accelerate: State of DevOps Strategies for a New Economy. *Report. DevOps Research & Assessment (DORA)* (2018).
[23] César França, Fabio QB Da Silva, and Helen Sharp. 2020. Motivation and satisfaction of software engineers. *IEEE Transactions on Software Engineering* 46, 2 (2020), 118–140.
[24] Jörg Freiling and Hanno Fichtner. 2010. Organizational culture as the glue between people and organization: A competence-based view on learning and competence building. *German Journal of Human Resource Management* 24, 2 (2010), 152–172.
[25] Yuko Fujigaki, Takashi Asakura, and Takashi Haratani. 1994. Work Stress and Depressive Symptoms among Japanese Information Systems Managers. *Industrial Health* 32, 4 (1994), 231–238. https://doi.org/10.2486/indhealth.32.231
[26] Jason C Gawke, Marjan J Gorgievski, and Arnold B Bakker. 2018. Personal costs and benefits of employee intrapreneurship: Disentangling the employee intrapreneurship, well-being, and job performance relationship. *Journal of occupational health psychology* 23, 4 (2018), 508.
[27] Robin Genuer and Jean-Michel Poggi. 2020. *Random Forests*. Springer International Publishing, Cham, 33–55.
[28] Robert L Glass, Iris Vessey, and Sue A Conger. 1992. Software tasks: Intellectual or clerical? *Information & Management* 23, 4 (1992), 183–191.
[29] Google. 2020. DORA Research Program. https://www.devops-research.com/research.html/. [Online; accessed 2022-06-14].
[30] Daniel Graziotin and Fabian Fagerholm. 2019. Happiness and the productivity of software engineers. In *Rethinking Productivity in Software Engineering*. Springer.
[31] Daniel Graziotin, Fabian Fagerholm, Xiaofeng Wang, and Pekka Abrahamsson. 2017. Consequences of Unhappiness while Developing Software. In *2017 IEEE/ACM 2nd International Workshop on Emotion Awareness in Software Engineering (SEmotion)*. IEEE. https://doi.org/10.1109/semotion.2017.5
[32] Daniel Graziotin, Fabian Fagerholm, Xiaofeng Wang, and Pekka Abrahamsson. 2018. What happens when software developers are (un) happy. *Journal of Systems and Software* 140 (2018), 32–47.
[33] Hazel Grünewald, Petra Kneip, and Arjan Kozica. 2019. The use of gamification in workplace learning to encourage employee motivation and engagement. *The Wiley handbook of global workplace learning* (2019), 557–575.
[34] Joe Hair, Carole L Hollingsworth, Adriane B Randolph, and Alain Yee Loong Chong. 2017. An updated and expanded assessment of PLS-SEM in information systems research. *Industrial management & data systems* 117, 3 (2017), 442–458.
[35] Joseph F Hair, Claudia Binz Astrachan, Ovidiu I Moisescu, Lăcrămioara Radomir, Marko Sarstedt, Santha Vaithilingam, and Christian M Ringle. 2021. Executing and interpreting applications of PLS-SEM: Updates for family business researchers. *Journal of Family Business Strategy* 12, 3 (2021), 100392.
[36] Joe F Hair, Christian M Ringle, and Marko Sarstedt. 2011. PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory and Practice* 19, 2 (2011), 139–152.
[37] Joseph F Hair, Jeffrey J Risher, Marko Sarstedt, and Christian M Ringle. 2019. When to use and how to report the results of PLS-SEM. *European business review* (2019).
[38] Joseph F Hair Jr, G Tomas M Hult, Christian M Ringle, and Marko Sarstedt. 2022. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (3rd ed.). Sage publications.
[39] Joseph F Hair Jr, Marko Sarstedt, Christian M Ringle, and Siegfried P Gudergan. 2017. *Advanced issues in partial least squares structural equation modeling*. Sage Publications.
[40] Jim Harter. 2022. Is quiet quitting real. *Gallup. com* (2022).
[41] Deneen M Hatmaker and Shahidul Hassan. 2023. When do women receive managerial support? The effects of gender congruence and the manager-employee relationship. *Public Management Review* 25, 1 (2023), 22–41.
[42] Jianjun He, Ling Xu, Meng Yan, Xin Xia, and Yan Lei. 2020. Duplicate bug report detection using dual-channel convolutional neural networks. In *Proceedings of the 28th International Conference on Program Comprehension*. 117–127.
[43] Jörg Henseler. 2020. *Composite-based structural equation modeling: Analyzing latent and emergent variables*. Guilford Publications.
[44] Jörg Henseler, Theo K Dijkstra, Marko Sarstedt, Christian M Ringle, Adamantios Diamantopoulos, Detmar W Straub, David J Ketchen Jr, Joseph F Hair, G Tomas M Hult, and Roger J Calantone. 2014. Common beliefs and reality about PLS: Comments on Rönkkö and Evermann. *Organizational Research Methods* 17, 2 (2014).

[45] Jörg Henseler, Geoffrey Hubona, and Pauline Ash Ray. 2016. Using PLS path modeling in new technology research: updated guidelines. *Industrial Management & Data Systems* 116, 1 (2016), 2–20.

[46] Jörg Henseler, Christian M Ringle, and Marko Sarstedt. 2015. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science* 43, 1 (2015), 115–135.

[47] Jörg Henseler, Christian M Ringle, and Marko Sarstedt. 2016. Testing measurement invariance of composites using partial least squares. *International marketing review* 33, 3 (2016), 405–431.

[48] Geert Hofstede. 2001. *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations.* Sage publications.

[49] Robert J House. 1996. Path-goal theory of leadership: Lessons, legacy, and a reformulated theory. *The Leadership Quarterly* 7, 3 (1996), 323–352.

[50] An-Tien Hsieh and Hui-Yu Chao. 2004. A reassessment of the relationship between job specialization, job rotation and job burnout: example of Taiwan's high-technology industry. *The International Journal of Human Resource Management* 15, 6 (Sept. 2004), 1108–1123. https://doi.org/10.1080/09585190410001677331

[51] Muhammad Jawad, Munazza Naz, and Sohail Rizwan. 2023. Leadership support, innovative work behavior, employee work engagement, and corporate reputation: Examining the effect of female in not government organizations. *Corporate Social Responsibility and Environmental Management* 30, 2 (2023), 708–719.

[52] Damien Joseph and Soon Ang. 2003. Turnover of IT professionals: a quantitative analysis of the literature. In *Proceedings of the 2003 SIGMIS conference on Computer personnel research: Freedom in Philadelphia–leveraging differences and diversity in the IT workforce.* 130–132.

[53] E Kevin Kelloway and Arla L Day. 2005. Building healthy workplaces: what we know so far. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 37, 4 (2005), 223.

[54] Renata Korsakienė, Asta Stankevičienė, Agnė Šimelytė, and Milda Talačkienė. 2015. Factors driving turnover and retention of information technology professionals. *Journal of business economics and management* 16, 1 (2015), 1–17.

[55] Helena Chmura Kraemer, George A Morgan, Nancy L Leech, Jeffrey A Gliner, Jerry J Vaske, and Robert J Harmon. 2003. Measures of clinical significance. *Journal of the American Academy of Child & Adolescent Psychiatry* 42, 12 (2003), 1524–1529.

[56] Hengky Latan, Joseph F. Hair, and Richard Noonan (Eds.). 2023. *Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications* (2nd ed.). Springer.

[57] Bin Lin, Gregorio Robles, and Alexander Serebrenik. 2017. Developer turnover in global, industrial open source projects: Insights from applying survival analysis. In *2017 IEEE 12th International Conference on Global Software Engineering (ICGSE).* IEEE, 66–75.

[58] Peter Lok and John Crawford. 2004. The effect of organisational culture and leadership style on job satisfaction and organisational commitment: A cross-national comparison. *Journal of management development* (2004).

[59] Mingxiao Lu, Abdullah Al Mamun, Xuelin Chen, Qing Yang, and Mohammad Masukujjaman. 2023. Quiet quitting during COVID-19: The role of psychological empowerment. *Humanities and Social Sciences Communications* 10, 1 (2023), 1–16.

[60] Gustavo A. Lujan-Moreno, Phillip R. Howard, Omar G. Rojas, and Douglas C. Montgomery. 2018. Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study. *Expert Systems with Applications* 109 (2018), 195–205.

[61] Ayman Mahmoud Maaitah. 2018. The role of leadership style on turnover intention. *International Review of Management and Marketing* 8, 5 (2018), 24.

[62] Thalmus Mahand and Cam Caldwell. 2023. Quiet Quitting—Causes and Opportunities. *Business and Management Research* 12, 1 (2023), 9–18.

[63] Monalisa Mahapatra and Surya Prakash Pati. 2018. Technostress Creators and Burnout. In *Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research.* ACM. https://doi.org/10.1145/3209626.3209711

[64] Rafael Gomes Mantovani, Tomáš Horváth, Ricardo Cerri, Sylvio Barbon Junior, Joaquin Vanschoren, and André Carlos Ponce de Leon Ferreira de Carvalho. 2018. An empirical study on hyperparameter tuning of decision trees. *arXiv preprint arXiv:1812.02207* (2018).

[65] Laurențiu P Maricuțoiu, Coralia Sulea, and Alina Iancu. 2017. Work engagement or burnout: Which comes first? A meta-analysis of longitudinal evidence. *Burnout research* 5 (2017), 35–43.

[66] John A Martilla and John C James. 1977. Importance-performance analysis. *Journal of marketing* 41, 1 (1977), 77–79.

[67] Gonzalo Martinez-Munoz and Alberto Suárez. 2007. Using boosting to prune bagging ensembles. *Pattern Recognition Letters* 28, 1 (2007), 156–165.

[68] Christina Maslach, Susan E Jackson, and Michael P Leiter. 1997. *Maslach burnout inventory.* Scarecrow Education.

[69] Christina Maslach and Michael P Leiter. 2008. Early predictors of job burnout and engagement. *Journal of applied psychology* 93, 3 (2008), 498.

[70] Tiago Massoni, Nilton Ginani, Wallison Silva, Zeus Barros, and Georgia Moura. 2019. Relating voluntary turnover with job characteristics, satisfaction and work exhaustion-An initial study with Brazilian developers. In *2019 IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*. IEEE, 85–88.

[71] Katherine S McGilton. 2010. Development and psychometric testing of the supportive supervisory scale. *Journal of Nursing Scholarship* 42, 2 (2010), 223–232.

[72] Kathryn Mearns, Lorraine Hope, Michael T Ford, and Lois E Tetrick. 2010. Investment in workforce health: Exploring the implications for workforce safety climate and commitment. *Accident Analysis & Prevention* 42, 5 (2010), 1445–1454.

[73] Jo Ellen Moore. 2000. One Road to Turnover: An Examination of Work Exhaustion in Technology Professionals. *MIS Quarterly* 24, 1 (March 2000), 141. https://doi.org/10.2307/3250982

[74] Stanley A Mulaik. 2010. *Foundations of factor analysis* (2nd ed.). CRC press.

[75] Anthony M Musolf, Emily R Holzinger, James D Malley, and Joan E Bailey-Wilson. 2022. What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. *Human Genetics* 141, 9 (2022), 1515–1528.

[76] Vishnuprasad Nagadevara and Vasanthi Srinivasan. 2008. Early Prediction of Employee Attrition in Software Companies-Application of Data Mining Techniques. *Research and Practice in Human Resource Management* 16 (2008), 2020–2032.

[77] Ramyashilpa D Nayak. 2014. Anxiety and mental health of software professionals and mechanical professionals. *International Journal of Humanities and Social Science Invention* 3, 2 (2014), 52–56.

[78] David E Nelson. 1996. Validity of self reported data on injury prevention behavior: lessons from observational and self reported surveys of safety belt use in the US. *Injury Prevention* 2, 1 (1996), 67–69.

[79] Thomas WH Ng, Kelly L Sorensen, and Frederick HK Yim. 2009. Does the job satisfaction—job performance relationship vary across cultures? *Journal of Cross-Cultural Psychology* 40, 5 (2009), 761–796.

[80] Nicole Novielli and Alexander Serebrenik. 2019. Sentiment and emotion in software engineering. *IEEE Software* 36, 5 (2019), 6–23.

[81] Ruan Oliveira, Tiago Massoni, Narallynne Araújo, Camila Sarmento, and Francielle Santos. 2021. Ants Doing Legwork: Investigating Motivators for Software Development Career Abandonment. In *Proceedings of the XXXV Brazilian Symposium on Software Engineering*. 353–362.

[82] Pamela L Perrewé, Wayne A Hochwarter, Ana Maria Rossi, Alan Wallace, Isabelle Maignan, Stephanie L Castro, David A Ralston, Mina Westman, Guenther Vollmer, Moureen Tang, et al. 2002. Are work stress relationships universal? A nine-region examination of role stressors, general self-efficacy, and burnout. *Journal of International management* 8, 2 (2002), 163–187.

[83] Jack J Phillips, Patricia Pulliam Phillips, and Al Pulliam. 2014. *Measuring ROI in Environment, Health, and Safety.* John Wiley & Sons.

[84] A.M. Pines and Elliot Aronson. 1981. *Burnout: From tedium to personal growth.* The Free Press.

[85] Usha PM and NV Balaji. 2019. Analysing Employee attrition using machine learning. *Karpagam Journal of Computer Science* 13 (2019), 277–282.

[86] Philipp Probst and Anne-Laure Boulesteix. 2018. To tune or not to tune the number of trees in random forest. *Journal of Machine Learning Research* 18, 181 (2018), 1–18.

[87] Philipp Probst, Marvin N Wright, and Anne-Laure Boulesteix. 2019. Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery* 9, 3 (2019), e1301.

[88] Nicole Franziska Richter, Gabriel Cepeda-Carrion, José Luis Roldán Salgueiro, and Christian M Ringle. 2016. European management research using partial least squares structural equation modeling (PLS-SEM). *European Management Journal, 34 (6), 589-597.* (2016).

[89] Edward E Rigdon. 2012. Rethinking partial least squares path modeling: In praise of simple methods. *Long range planning* 45, 5-6 (2012), 341–358.

[90] Christian M Ringle and Marko Sarstedt. 2016. Gain more insight from your PLS-SEM results: The importance-performance map analysis. *Industrial management & data systems* 116, 9 (2016), 1865–1886.

[91] Christian M. Ringle, Sven Wende, and Jan-Michael Becker. 2024. SmartPLS 4. Retrieved from https://www.smartpls.com.

[92] Jeanine Romano, Jeffrey D Kromrey, Jesse Coraggio, and Jeff Skowronek. 2006. Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys. In *annual meeting of the Florida Association of Institutional Research*, Vol. 177. 34.

[93] Daniel Russo, Andres Ramon Masegosa Arredondo, and Klaas-Jan Stol. 2022. From Anecdote to Evidence: The Relationship Between Personality and Need for Cognition of Developer. *Empirical Software Engineering* 27, 71 (2022).

[94] Daniel Russo and Klaas-Jan Stol. 2021. PLS-SEM for Software Engineering Research: An Introduction and Survey. *ACM Computing Surveys (CSUR)* 54, 4 (2021), 1–38.

[95] Takaya Saito and Marc Rehmsmeier. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one* 10, 3 (2015), e0118432.
[96] Fabio Santos, Jacob Penney, João Felipe Pimentel, Igor Wiese, Bianca Trinkenreich, Igor Steinmacher, and Marco A Gerosa. 2023. Tell Me Who Are You Talking to and I Will Tell You What Issues Need Your Skills. In *2023 IEEE/ACM 20th International Conference on Mining Software Repositories (MSR).*
[97] Marko Sarstedt, Christian M Ringle, and Joseph F Hair. 2017. Treating unobserved heterogeneity in PLS-SEM: A multi-method approach. In *Partial least squares path modeling*. Springer, 197–217.
[98] Wilmar B Schaufeli. 2004. The future of occupational health psychology. *Applied Psychology* 53, 4 (2004), 502–517.
[99] Wilmar B Schaufeli, Marisa Salanova, Vicente González-Romá, and Arnold B Bakker. 2002. The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness studies* 3 (2002), 71–92.
[100] Wilmar B Schaufeli, Akihito Shimazu, Jari Hakanen, Marisa Salanova, and Hans De Witte. 2019. An ultra-short measure for work engagement: the UWES-3 validation across five countries. *European Journal of Psychological Assessment* 35, 4 (2019), 577.
[101] Jen Schoepke, Peter L. T. Hoonakker, and Pascale Carayon. 2004. Quality of Working Life among Women and Men in the Information Technology Workforce. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 48, 14 (Sept. 2004), 1576–1580. https://doi.org/10.1177/154193120404801404
[102] Peter H Schönemann. 1971. The minimum average correlation between equivalent sets of uncorrelated factors. *Psychometrika* 36, 1 (1971), 21–30.
[103] Gaurav G Sharma and Klaas-Jan Stol. 2020. Exploring onboarding success, organizational fit, and turnover intention of software professionals. *Journal of Systems and Software* 159 (2020), 110442.
[104] Haijian Shi. 2007. *Best-first Decision Tree Learning*. Ph. D. Dissertation. Citeseer.
[105] Sheng-Pao Shih, James J. Jiang, Gary Klein, and Eric Wang. 2013. Job burnout of the information technology worker: Work exhaustion, depersonalization, and personal accomplishment. *Information & Management* 50, 7 (Nov. 2013), 582–589.
[106] Jordan Shropshire and Christopher Kadlec. 2012. I'm leaving the IT field: The impact of stress, job insecurity, and burnout on IT professionals. *International Journal of Information and Communication Technology Research* 2, 1 (2012).
[107] Brad Shuck, Thomas G Reio Jr, and Tonette S Rocco. 2011. Employee engagement: An examination of antecedent and outcome variables. *Human resource development international* 14, 4 (2011), 427–445.
[108] Sabine Sonnentag, Felix C. Brodbeck, Torsten Heinbokel, and Wolfgang Stolte. 1994. Stressor-burnout relationship in software development teams. *Journal of Occupational and Organizational Psychology* 67, 4 (1994), 327–341.
[109] Maie Stein, Sylvie Vincent-Hoeper, and Sabine Gregersen. 2020. Why busy leaders may have exhausted followers: A multilevel perspective on supportive leadership. *Leadership & Organization Development Journal* 41, 6 (2020), 829–845.
[110] Klaas-Jan Stol, Mario Schaarschmidt, and Shelly Goldblit. 2022. Gamification in software engineering: the mediating role of developer engagement and job satisfaction. *Empirical Software Engineering* 27, 2 (2022), 1–34.
[111] Mervyn Stone. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)* 36, 2 (1974), 111–133.
[112] M Subhashini and R Gopinath. 2020. Employee attrition prediction in industry using machine learning techniques. *International Journal of Advanced Research in Engineering and Technology* 11, 12 (2020), 3329–3341.
[113] Carolyn Timms, Paula Brough, and Deborah Graham. 2012. Burnt-out but engaged: the co-existence of psychological burnout and engagement. *Journal of Educational Administration* 50, 3 (2012), 327–345.
[114] Bianca Trinkenreich, Ricardo Britto, Marco A Gerosa, and Igor Steinmacher. 2022. An empirical investigation on the challenges faced by women in the software industry: A case study. In *Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software Engineering in Society*. 24–35.
[115] Bianca Trinkenreich, Fabio Santos, and Klaas-Jan Stol. 2023. Replication Package. https://figshare.com/s/f36deebdc82cb764ff4e.
[116] Bianca Trinkenreich, Klaas-Jan Stol, Igor Steinmacher, Marco Gerosa, Anita Sarma, Marcelo Lara, Michael Feathers, Nicholas Ross, and Kevin Bishop. 2023. A Model for Understanding and Reducing Developer Burnout. In *Proceedings of the International Conference on Software Engineering (Companion).*
[117] Tien Rahayu Tulili, Andrea Capiluppi, and Ayushi Rastogi. 2022. Burnout in software engineering: A systematic mapping study. *Information and Software Technology* (2022), 107116.
[118] Nilay Unsal, GracieLee Weaver, Jeremy W Bray, Daniel Bibeau, and Garrett Saake. 2021. Return on investment of workplace wellness: Evidence from a Long-Term care company. *Workplace Health & Safety* 69, 2 (2021), 81–90.
[119] Robert J Vandenberg and Jodi Barnes Nelson. 1999. Disaggregating the motives underlying turnover intentions: when do intentions predict turnover behavior? *Human relations* 52, 10 (1999), 1313–1336.
[120] Jacob Weisberg. 1994. Measuring workers' burnout and intention to leave. *International Journal of Manpower* (1994).

[121] Ron Westrum. 2004. A typology of organisational cultures. *BMJ Quality & Safety* 13, suppl 2 (2004), ii22–ii27.
[122] Eelke Wiersma. 2007. Conditions that shape the learning curve: Factors that increase the ability and opportunity to learn. *Management Science* 53, 12 (2007), 1903–1915.
[123] Anja Wölbling, Kira Krämer, Clemens N. Buss, Katrin Dribbisch, Peter LoBue, and Abraham Taherivand. 2012. Design Thinking: An Innovative Concept for Developing User-Centered Software. In *Management for Professionals*. Springer Berlin Heidelberg, 121–136.
[124] Kin Fai Ellick Wong and Cecilia Cheng. 2020. The turnover intention–behaviour link: A culture-moderated meta-analysis. *Journal of Management Studies* 57, 6 (2020), 1174–1216.
[125] Guanping Xiao, Xiaoting Du, Yulei Sui, and Tao Yue. 2020. Hindbr: Heterogeneous information network based duplicate bug report prediction. In *2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 195–206.
[126] Li Yang and Abdallah Shami. 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* 415 (2020), 295–316.

Received July 2023; revised June 2024; accepted August 2024