Safety Science 166 (2023) 106260

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

# Exploring the impact of safety culture on incident reporting: Lessons learned from machine learning analysis of NHS England staff survey and incident data

G.K. Kaya [a,*], S. Ustebay [b], J. Nixon [a], C. Pilbeam [a], M. Sujan [c]

[a] Safety and Accident Investigation Centre, School of Aerospace, Transport and Manufacturing, Cranfield University, UK
[b] Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Istanbul Medeniyet University, Turkiye
[c] Human Factors Everywhere, Woking, UK

ARTICLE INFO

*Keywords:*
Safety culture
Safety
Incident analysis
Healthcare
Incident reporting
Machine learning

ABSTRACT

Safety culture is one of the key factors contributing to safety, even though limited evidence supports its impact on safety outcomes. This study uses supervised machine learning algorithms to explore the association between safety culture and incident reporting. The study used National Health Service (NHS) England annual staff survey data as a proxy of safety culture to predict eighteen incident reporting variables. The study did not achieve high accuracy rates in the prediction models. The highest association was found between safety culture and the number of incidents reported in class low, medium and high. LightGBM was the best-performed algorithm. SHAP plots were used to explain the model. Findings suggest that compassionate culture, violence and harassment and work pressure are critical in predicting the number of incidents reported. More specifically, the violence and harassment had a more significant impact on predicting the number of incidents reported in class high than in class medium and low. The involvement had more effect on predicting class low. The results demonstrated different behaviours in predicting different incident reporting classes. The findings facilitate lessons learned from staff surveys and incident reporting data in NHS England. Consequently, the findings can contribute to improving the safety culture in hospitals.

## 1. Introduction

Safety culture has a significant impact on the safety management (Amalberti et al., 2005; Reason, 1998). Poor safety culture was identified as a contributory factor in many major industrial accidents (Aburumman et al., 2019; Goncalves Filho and Waterson, 2018), as well as in healthcare, as evidenced, for example, by the recent independent investigation into maternity deaths at East Kent, UK (Kirkup, 2022). In other words, positive safety culture contributes to improved safety behaviours, outcomes as well as safety management (Agnew et al., 2013; Maneechaeye and Potipiroon, 2022; Shi et al., 2022).

NHS England defines positive safety culture as "one where the environment is collaboratively crafted, created, and nurtured so that everybody (individual staff, teams, patients, service users, families, and carers) can flourish to ensure brilliant, safe care by continuous learning and improvement of safety risks, supportive, psychologically safe teamwork, and enabling and empowering speaking up by all." (NHS England, 2022a, 2022b). Here, it is useful to note that NHS's approach to safety culture is broader than other safety–critical industries. One reason can be given as the definition of the incident in healthcare includes psychological harm.

The NHS focuses on developing a positive safety culture to remove blame and encourage staff to speak up, raise safety concerns and report incidents (Brennan et al., 2021). Moreover, NHS England promotes a safety culture that aims to contribute to learning and focuses on repairing relationships and meeting the needs of those affected by patient safety incidents. This latter dimension of safety culture is called 'restorative culture' (Dekker, 2020). Considering all these, it is expected to observe a relationship between safety culture and safety outcomes. Exploring such relationships is a priority research area (Dunstan et al., 2019; Groves, 2014). Despite the generally accepted theoretical links between the two, there is limited empirical evidence to associate them

* Corresponding author at: School of Aerospace, Transport and Manufacturing, Building 300 (Martell House) F046, 1st Floor, West Wing, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK.

*E-mail address:* kubra.kaya@cranfield.ac.uk (G.K. Kaya).

https://doi.org/10.1016/j.ssci.2023.106260
Received 21 March 2023; Received in revised form 8 June 2023; Accepted 10 July 2023
Available online 13 July 2023

and explain the 'how' (Flin, 2007; Simsekler et al., 2021; Simsekler and Qazi, 2022).

Sorra et al. (2012) explored the relationship between patient safety measures and safety culture in large, teaching, and private hospitals by examining bivariate correlations, and they found no significant relationship. Groves (2014) examined the relationship between safety culture and patient outcomes through *meta*-analysis, and no significant association was identified. This likely results from the culture being a multi-dimensional construct, and its assessment is complicated (Guldenmund, 2010).

Safety culture can be measured quantitatively, semi-quantitatively or qualitatively (Kalteh et al., 2020). Semi-quantitative and qualitative approaches aim to facilitate discussions on safety culture, such as the MaPSaF (The University of Manchester, 2006) and the safety culture discussion cards (EUROCONTROL, 2018). Quantitative methods are to assess the level of safety culture, which is through conducting surveys. Investigators often use 'safety climate' surveys as a proxy measure of the 'safety culture' (Morello et al., 2013; The Health Foundation, 2011). Safety climate refers to members' perceptions and attitudes toward safety in their working environment (Halligan and Zecevic, 2011). Despite the safety climate only reflecting on the surface level of the safety culture, safety climate surveys are used to measure safety culture (Aburumman et al., 2019; Guldenmund, 2007).

Safety Attitudes Questionnaire (SAQ) (Sexton et al., 2006) and Hospital Survey on Patient Safety Culture (HSOPSC) (AHRQ, 2016) are two of the most commonly used surveys for assessing and measuring safety culture (Simsekler et al., 2020). SAQ comprises of 64 survey items under six main themes: teamwork climate, job satisfaction, perceptions of management, safety climate, working conditions and stress recognition (Sexton et al., 2006). HSOPSC has 42 survey items, which are grouped under 12 safety culture dimensions: teamwork within units, supervision, organisational learning, management support, the overall perception of patient safety, feedback and communication about the error, communication openness, frequency of events reported, teamwork across units, staffing, handoffs and transitions, and nonpunitive response to errors (AHRQ, 2016). National Health Service (NHS) staff survey is another commonly used survey that explores staff experiences in the organisation, and it is used to measure safety culture (Simsekler et al., 2021; Simsekler and Qazi, 2022).

Despite the availability of various tools/instruments to measure safety culture, selecting the tools and data analysis methods would influence the results and lessons learned. Studies predominantly used surveys to measure safety culture and conventional statistical analysis methods to analyse survey data (Amarneh and Al Nobani, 2022; Steyrer et al., 2013; Weaver et al., 2014). However, some researchers claimed that conventional statistical analysis methods are limited in uncovering complex relationships and suggested using machine learning (ML) algorithms (Alkaissy et al., 2023; Simsekler et al., 2020; Ustebay et al., 2022; Zhang and Haghani, 2015). ML has the potential to provide valuable insights into larger datasets and to discover links that linear correlational methods may not reveal. Given the complexity of the construct of safety culture, we find this likely.

Simsekler et al. (2020) used ML algorithms to analyse the relationship between safety culture survey dimensions using the HSOPSC data. They found safety perception, supervisor expectations and management support as leading drivers of patient safety. Another study by Simsekler et al. (2021) used ML algorithms to analyse the impact of six NHS staff survey themes (team working, quality of care, immediate managers, health and well-being, safety culture and equality, diversity and inclusion) on the two survey items related to errors and near misses. They found that 'health and well-being' is the most critical driver in predicting errors and near misses.

This study uses ML algorithms to explore the association between safety culture and safety outcome: incident reporting. The study is structured around two key aims. Firstly, to analyse the relationship between safety culture and incident reporting, data drawn from annual NHS staff surveys and NHS incident reporting data. Secondly, to determine the candidate ML algorithm for treating these data. Consequently, this study provides lessons learned from staff survey findings and contributes to improving hospital safety culture.

## 2. Methods

### 2.1. Data collection

This study collected annual NHS staff surveys and NHS incident reporting data from Acute Trusts in England. The study collected five-year data from 109 Acute Trusts between 2017 and 2021. This study only included hospitals with active operations and no name changes (e.g., no hospital merge) during these five years. Both datasets are publicly available on NHS England's web page (NHS England, 2022a, 2022b). Fig. 1 explains the links between staff surveys, safety culture, incidents, and incident reporting.

The NHS staff survey is revised each year. That is why this study only included survey questions available during all five years, and the study used the most recent categorisation scheme. After removing missing data, the collected NHS staff survey data contained 41 items (see Appendix A). The included survey items are categorised under eleven categories: compassionate culture (5 items), diversity and equality (3 items), recognition and rewards (4 items), autonomy and control (3 items), involvement (3 items), raising concerns (2 items), work pressure (3 items), violence and harassment (11 items), team working (2 items), line management (2 items), and motivation (3 items). The NHS England provides the individual survey item results and overall categorical results. Since this study aimed to conduct a comprehensive analysis and, in turn, learn from survey findings, the study used all individual survey items as safety culture variables to develop prediction models.

NHS incident reporting data include the median number of days between incidents that occurred and were reported, the total number of reported incidents, the number of incidents per 1000 bed days, the percentage of incidents based on the degree of harm (i.e., none, low, moderate, severe and death), incident types and care settings of occurrence (see Appendix B).

This study uses staff survey data as input variables to predict all incident reporting variables. However, we only presented the findings from the model that achieved the highest accuracy rate. Fig. 2 illustrates the block diagram of the overall study design.

### 2.2. Data pre-processing

Missing data were managed before model training. We deleted the row if there were any missing values at any features. 1.98 per cent of data were removed at this data cleaning stage.

This study used sixty variables characterised as features in the ML lexicon. A year feature, forty-one staff survey features, and eighteen

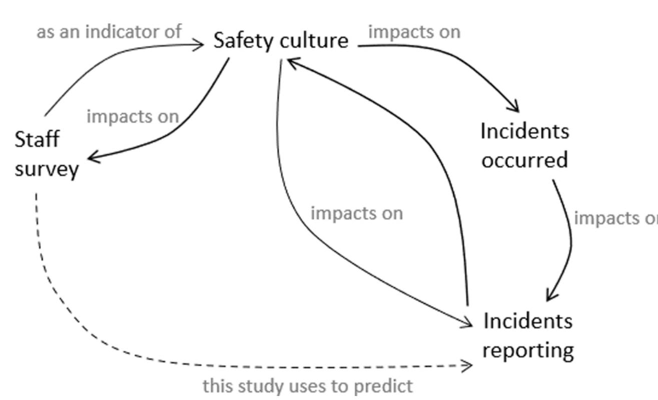


**Fig. 1.** Links between staff survey and incident reporting.

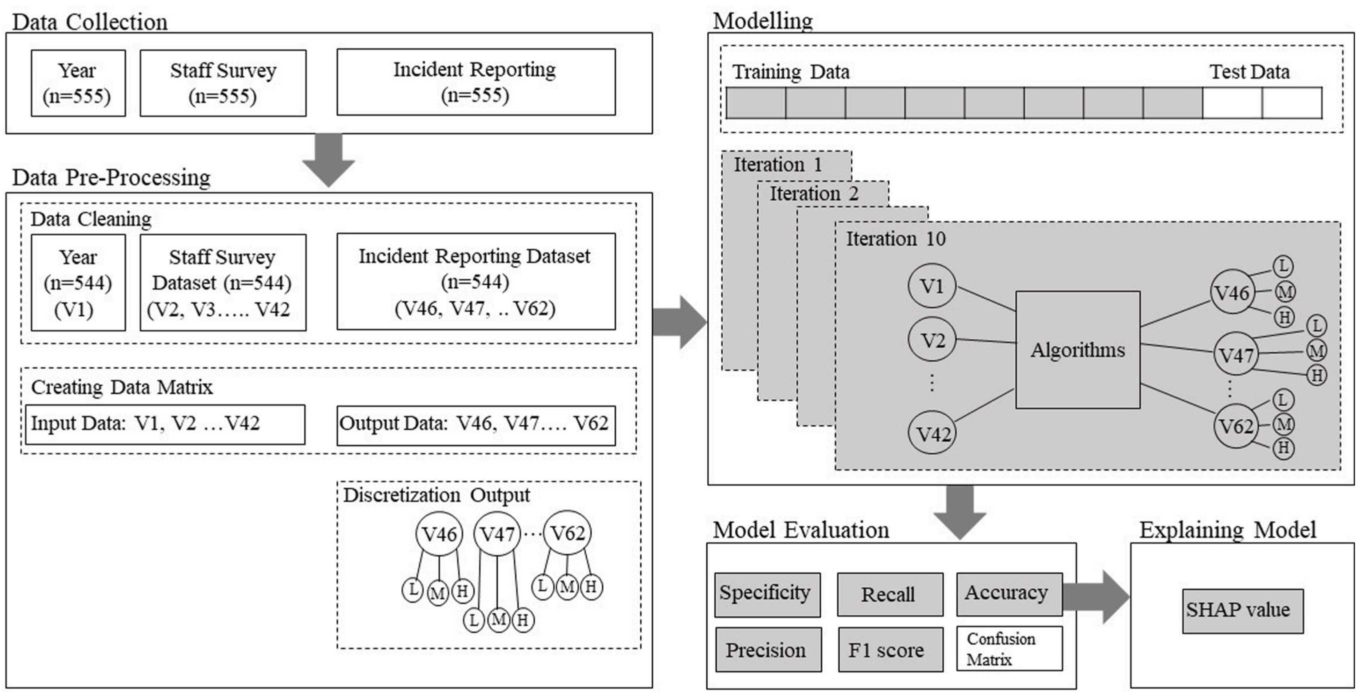


**Fig. 2.** An overall study design.

incident reporting features were used to develop ML models. All features are numerical data. Staff survey features and the data collection year feature were assigned as inputs to make predictions. The eighteen incident reporting features were set as separate targets/outputs for the model to predict.

This study converted the numerical incident reporting data into categorical data. Using a categorical data model in an analysis has benefits. One is the capacity to swiftly spot trends, changes and patterns based on connected variables. We initially tested model performances by using two, three and five classes, and then we selected three classes as they achieved the best prediction performance. Using a quantile-based discretisation method, incident reporting variables were discretised into three equal-sized classes based on rank. For instance, we converted the total number of reported incidents data into three classes: low, medium and high. Each class represents 33.3% of the data. In this case, the first class (low) represents the lowest 33.3% of the data, the second class (medium) represents the middle 33.3% of the data, and the third class (high) represents the highest 33.3% of the data. Dividing a continuous variable into discrete categories allows us to examine the relationship between two variables more easily, as it simplifies the data and makes it easier to interpret. Then, the model was developed to predict each class rather than the numerical value.

### *2.3. Modelling*

This study used fourteen supervised machine learning (ML) algorithms: logistic regression, k-neighbours, naïve Bayes, decision tree, support vector machine (SVM) linear and radial kernel, gaussian process, multilayer perceptron (MLP), ridge, random forest, linear discrimination analysis, extra tree, extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM). No single algorithm works best in every situation, especially considering the constraints of real-life systems. This study selected the commonly used fourteen supervised ML algorithms to represent algorithms built on different approaches, including conventional and new analysis algorithms. The authors aimed to determine the best-performed ML algorithm for treating these data.

The dataset was split for training and testing. Stratified random sampling has been applied to split the data into homogenous strata between the training and testing split with a training size of 80 per cent and a testing size of 20 per cent. Although there is no hard rule on selecting the training and test split ratio, this study used the 80/20 ratio as it is a common one and follows the Pareto Principle. 10-fold cross-validation was adopted. The scikit-Learn Randomised Search CV method optimised hyperparameters (Pedregosa et al., 2011).

This study developed models to predict the labels of eighteen incident reporting variables, comparatively using fourteen supervised ML algorithms. Model performance was evaluated with six standard metrics: AUROC, accuracy, recall, specificity, precision, F1 and confusion matrix, using a training dataset. All these performance metric values that are closer to 1 imply a better-performing model, while values that are closer to 0.5 often indicate a model that performs no better than random guessing. The best-performed algorithm was selected based on the accuracy rate. AUROC demonstrates the ability of a binary classifier system at various decision thresholds. The accuracy rate is the portion of correctly predicted observations in the total observations. Recall is the capacity of a model to find all pertinent cases in a data set. The model's specificity indicates how much of the negative outcomes it accurately detects. Precision is the quality of a positive prediction made by the model. The F1 score measures model performance by calculating the harmonic mean of precision and recall (Panesar, 2021). The classification outcomes for each activity are summarised in the confusion matrix, which offers a comprehensive summary.

The model output was explained by calculating Shapley Additive exPlanations (SHAP) based on the cooperative game-theoretic concept. SHAP diagrams are a way to visualise the importance of different features in a machine learning model's prediction (Lundberg and Lee, 2017). It can be applied as a feature attribution approach in the context of ML. The strength of the input factors' influence on the predictions could be decoded by SHAP. By comparing what a model predicts with and without the feature, SHAP values determine the importance of the feature. However, the order in which a model sees features can affect its predictions. That is why it uses game theory concepts to fairly compare features. With SHAP, prediction models can be deeply understood (Mohanty et al., 2022). In this paper, SHAP explains individual contributions of a staff survey question item (e.g., in the last 12 months, how many times have you personally experienced harassment, bullying or abuse at work from other colleagues?) on predicting incident reporting. In other words, this study investigates the impacts of the indicators of safety culture elements on incident reporting rather than safety culture holistically.

## 3. Results

### 3.1. Model performance

This study developed prediction models using staff surveys and incident reporting datasets (n = 544). The findings revealed a low accuracy rate from all models, with the best-performed model receiving an accuracy rate of 0.585. Models received an accuracy rate between 0.4 and 0.585 (see Appendix C for the accuracy rates of all models). As models receiving an accuracy rate of 0.5 and below is not acceptable, this study only presented the findings from the best-performed model, which predicts the number of reported incidents (V47) in class low, medium and high. Table 1 provides all algorithm performances for predicting V47. Results showed that the LightGBM algorithm performed the best in all performance metrics, including an accuracy rate of 0.585, an AUROC rate of 0.688 and a specificity rate of 0.822.

A confusion matrix was built using the test dataset (see Fig. 3). The confusion matrix shows the predicted labels in comparison to the true labels. For instance, the model predicted the high class true in 25 samples, whereas it predicted it to be as in class medium in 8 samples and in class low in 4 samples.

### 3.2. Model explanation

The SHAP values were calculated for the model using the best-performing algorithm. One common type of SHAP diagram is a 'bar plot', as shown in Fig. 4. The plot consists of vertical bars, where each bar's length represents the corresponding feature's impact on the model's output. Features are ordered according to their importance, with the most important feature at the top. The SHAP values are normalised to sum up the model's output so that the length of the bar represents the contribution of each feature to the overall prediction. The plot helps to identify which features have the most significant impact on the model's predictions and can be used to understand how the model is making its decisions. Here, experiencing harassment, bullying or abuse at work *'from other colleagues'* (V35) and *'from patients/ service users, their relatives, or other members of the public'* (V33), and *'being happy with the care provided by the organisation'* (V6) are found as the top three critical features for the prediction model.

Considering the survey item categories, compassionate culture, violence and harassment, work pressure, and motivation were crucial categories for the overall model prediction. However, the findings suggest that features and feature categories have different degrees of impact

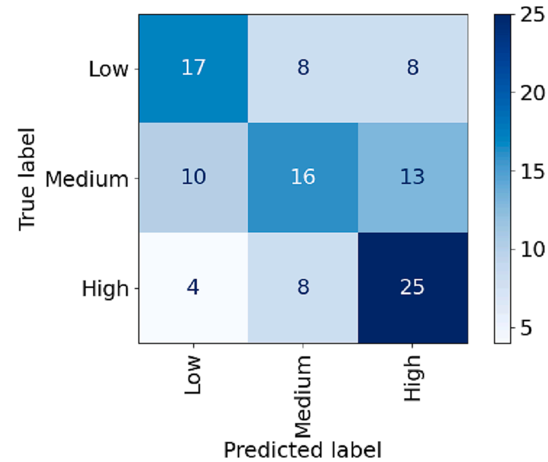


**Fig. 3.** Confusion matrix for predicting reported incidents in low, medium or high classes.

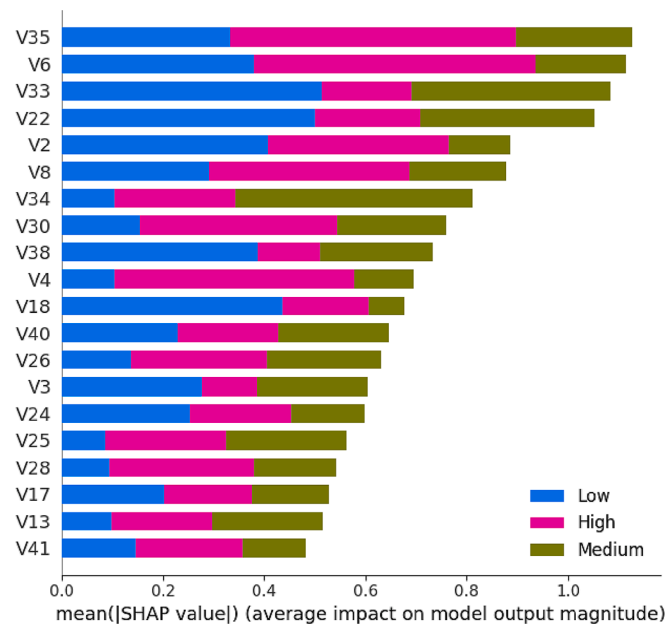


**Fig. 4.** SHAP summary plot for predicting the total number of reported incidents (V47) in class low, medium or high.

**Table 1**
Algorithm performances for predicting reported incidents in class low, medium or high.

| Algorithm | AUROC | Accuracy | Recall | Specificity | Precision | F1 |
|---|---|---|---|---|---|---|
| Logistic Regression | 0.568 | 0.426 | 0.423 | 0.684 | 0.401 | 0.395 |
| K Neighbors Classifier | 0.574 | 0.432 | 0.431 | 0.779 | 0.432 | 0.423 |
| Naive Bayes | 0.587 | 0.451 | 0.450 | 0.687 | 0.444 | 0.436 |
| Decision Tree Classifier | 0.572 | 0.430 | 0.430 | 0.7 | 0.429 | 0.424 |
| SVM - Linear Kernel | 0.552 | 0.403 | 0.402 | 0.494 | 0.357 | 0.318 |
| SVM - Radial Kernel | 0.55 | 0.405 | 0.401 | 0.577 | 0.271 | 0.321 |
| Gaussian Process Classifier | 0.544 | 0.393 | 0.391 | 0.622 | 0.375 | 0.362 |
| MLP Classifier | 0.595 | 0.461 | 0.459 | 0.742 | 0.450 | 0.447 |
| Ridge Classifier | 0.606 | 0.476 | 0.474 | 0.729 | 0.467 | 0.460 |
| Random Forest Classifier | 0.622 | 0.497 | 0.496 | 0.784 | 0.496 | 0.489 |
| Linear Discriminant Analysis | 0.633 | 0.513 | 0.511 | 0.749 | 0.508 | 0.504 |
| Extra Trees Classifier | 0.669 | 0.559 | 0.558 | 0.821 | 0.567 | 0.555 |
| XGBoost | 0.669 | 0.559 | 0.559 | 0.78 | 0.570 | 0.555 |
| LightGBM | 0.688 | 0.585 | 0.584 | 0.822 | 0.600 | 0.584 |

on predicting different output classes. For example, experiencing harassment, bullying or abuse at work from managers (V34) appeared to be the seventh most important feature in the overall model (see Fig. 4). In contrast, it was the most important feature in predicting incidents reported in the class medium. Still, it was not even on the top 15 list for predicting incidents reported in class low (see Fig. 5).

Fig. 5 demonstrates SHAP summary plots for predicting each class by listing the top 15 key drivers. The values in Fig. 5 are coloured according to the feature value and illustrate the direction and magnitude of the feature. The variable name is shown on the y-axis, with the top values being the most significant. The SHAP value is shown on the x-axis. It specifies the magnitude of the shift in log odds. Each dot represents a single instance or observation in the dataset, and the position of the dot along the x-axis represents the magnitude of the feature's effect on the prediction. The dots are coloured based on the feature value: red dots indicate high feature values, while blue dots indicate low feature values. Based on the distribution of the red and blue dots, we may generalise the directionality influence of the characteristics. Positive SHAP values indicate a positive influence on the variable, whereas negative SHAP values indicate a negative influence. When a feature has a SHAP value of 0, the model will disregard that feature's value for the purposes of this example.

In Fig. 5, features were ranked in different orders in predicting different output classes, or new features appeared to predict different classes. For instance, results suggest that experiencing harassment, bullying or abuse at work *'from patients/ service users, their relatives, or other members of the public'* (V33) appeared as the most critical driver in predicting reported incidents in class low (see A in Fig. 5), *'from managers'* (V34) appeared in class medium (see B Fig. 5) and *'from other colleagues'* (V35) appeared in class high (see C in Fig. 5). Here, it should be noted that all these features are under the violence and harassment category. The greater they are, the more incidents are reported. However, conflicted findings were obtained from features less impacting the prediction models. For example, more staff experiencing harassment, bullying or abuse at work from other colleagues (V35) leads to fewer reported incidents in class low and medium.

SHAP plots suggest that the more staff meeting conflicted demands (V22), the fewer incidents reported in class low and more incidents reported in class medium and high. More staff reporting physical violence at work (V25) leads to more incidents reported in class medium, whereas fewer incidents reported in class high. Another example is that the more staff believing their role is making a difference to patients/ service users (V2) and feel happy with the standard of care provided by the hospital (V6), the higher number of reported incidents in class high, and the more staff believing their hospitals acting on concerns raised by patients/service users (V4) leads to fewer reported incidents in class

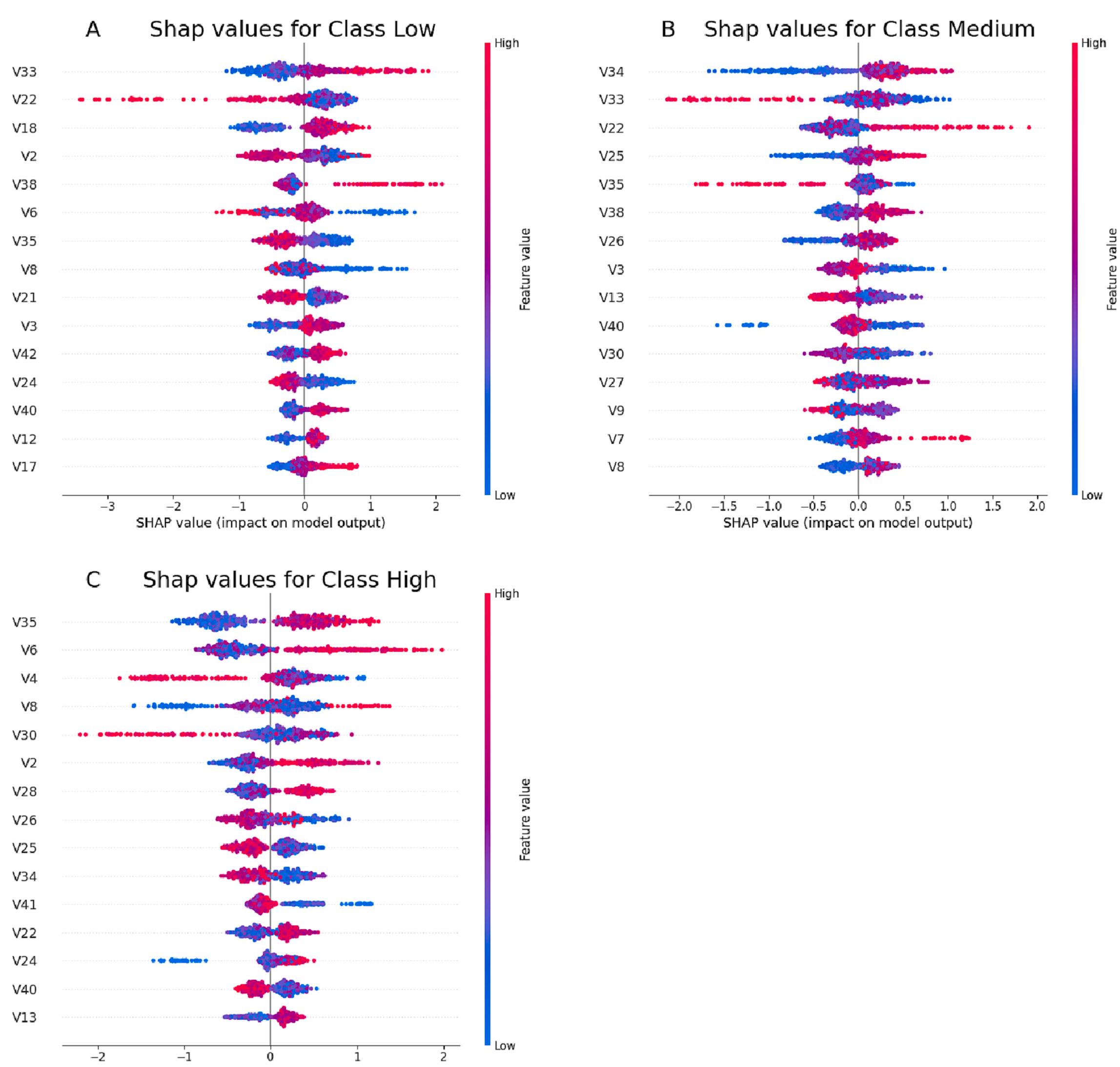


**Fig. 5.** SHAP values for predicting reported incidents in class low (A), medium (B), and high (C).

high. Moreover, the higher values of staff experiencing discrimination at work from patients/service users (V8) and feeling unwell due to work-related stress (V28) leads to a higher number of reported incidents in class high. More staff experiencing physical violence at work from patients/service users (V30) leads to fewer reported incidents in class high.

When looking into commonality, SHAP plots suggest there were four survey items which are *'In the last 12 months, have you personally experienced discrimination at work from patients/service users, their relatives or other members of the public?'* (V8), *'I am able to meet all the conflicting demands on my time at work.'* (V22), *'In the last 12 months, how many times have you personally experienced harassment, bullying or abuse at work from other colleagues?'* (V35) and *'I look forward to going to work.'* (V40), which appeared in the top 15 features for predicting incidents in all classes. Fig. 6 shows how the top 15 features vary with incident reporting classes.

## 4. Discussion

Safety culture is key in safety management (Reason, 1998). Based on this understanding, poor safety culture is expected to be associated with more incidents (Aburumman et al., 2019). Or we can interpret that good safety culture is associated with more incidents reported, especially reporting near misses. Reporting near misses is considered a reflection of an effective safety culture (Hudson, 2007). This is because safety culture impacts the staff's attitude and participation in safety actions (Maneechaeye and Potipiroon, 2022; Saedi et al., 2020). However, despite the theoretical foundations, little evidence shows such an association (Flin, 2007; Simsekler et al., 2021; Simsekler and Qazi, 2022). The findings from our study can be interpreted differently depending on the performance metrics used. While this study may indicate a slight association between safety culture and the number of incidents reported based on the accuracy rate, the use of AUROC and specificity values indicate an acceptable level of relationship. This is due to the characteristics of the performance metrics. What is more, safety culture is only one of the factors that affect incidents reported. That is why it is difficult to achieve outstanding model performances by solely using safety culture data.

The highest association was between safety culture and the total number of reported incidents (V47), and no association was found between safety culture and the percentage of reported incidents resulting in no harm (V48) (representing near misses) or death (V52) (representing actual incident numbers). Here, it must be noted that the number of incidents might only represent a portion of the actual incidents; it might instead reflect reporting behaviour. Having a good safety culture does not necessarily mean that there will not be any incidents. However,

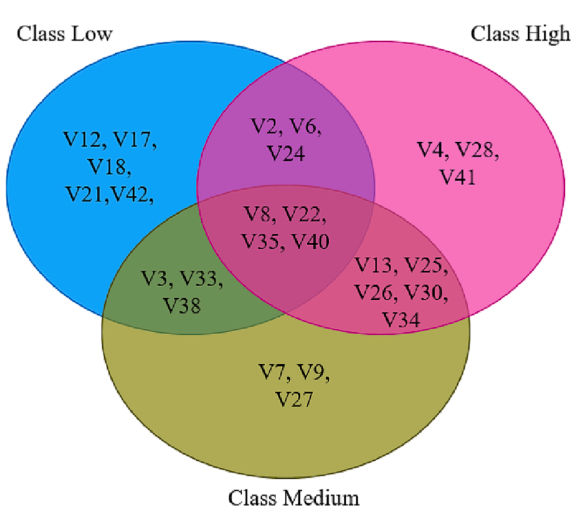


**Fig. 6.** A Venn diagram summarising the top 15 features in predicting each class.

we expect to see some impact on safety outcomes.

Our findings suggest that compassionate culture, violence and harassment, and work pressure are key to predicting the number of incident reporting. Compassionate culture creates a workplace where the staff goes to work, promoting kindness, respect, teamwork and inclusion (de Zulueta, 2021; NHS England, 2023). Compassionate culture is required to deliver safe and high-quality care (West, 2021). It enhances staff well-being and improves patient outcomes by reducing errors (de Zulueta, 2021). In this study, more staff saying that their role makes a difference (V2), which is a survey item under compassionate culture, is linked with fewer incidents reported for hospitals with low reported incidents. In contrast, it is associated with a higher number of reported incidents for hospitals with high-level reported incidents. Here, it could be interpreted that hospitals with low reported incident rates reflect more on the actual number of incidents. In contrast, hospitals with high reported incident rates reflect more on the reporting culture. Although these two should also be correlated, our findings demonstrated different behaviours from different classes, which shows the complexity of the relationship. So, lessons learned can be various depending on the output classes.

Violence and harassment occur when staff experience physical assault, bullying, aggression, sexual harassment, and verbal abuse or threat (NHS England, 2023). This can be done by managers, colleagues, patients, and patient relatives or due to work activities. Workplace violence is associated with a poor quality of care delivered, reduced job satisfaction and higher absenteeism (Khamisa et al., 2015; Liu et al., 2019; Yang et al., 2018). The 2021 NHS staff survey revealed that 29.2 per cent of staff experienced harassment, bullying or abuse at work from patients/ service users, their relatives of other members of the public, and 27.6 per cent from managers or colleagues (Shore et al., 2011). Those experiencing harassment, bullying and abuse are expected to speak up less (Brennan et al., 2020). That would also influence their actions and, in turn, patient safety outcomes (Mustapha et al., 2019). The findings in this study suggest that the violence and harassment category was the key driver in predicting incidents. Violence and harassment had both positive and negative impacts on the number of incidents reported (see Fig. 5). Some features under the violence and harassment category had an increasing impact on the level of reporting which might suggest that they were considered incidents and, in turn, reported. In some other cases, it had a reducing impact on the level of reporting, which might suggest that experiencing violence, harassment, bullying, and abuse has a negative impact on reporting habits.

Work pressure affects an individual's mental health and increases psychological risks, contributing to unsafe behaviours (Schwendimann et al., 2013). Several studies provided evidence of the impact of work pressure on unsafe work behaviour (Amponsah-Tawaih and Adu, 2016; Brown et al., 2000; Seo, 2005). As a result, it would be expected to contribute to incidents. The findings from this study revealed different patterns depending on the output classes. Work pressure seemed to have an increasing impact on the number of reported incidents for hospitals with low reported incident rates. In contrast, the opposite case was observed for hospitals with high reported incident rates. Here, it must be mentioned that work pressure-related variables had a more significant impact on predicting reported incidents in class low than in class high.

The findings suggest that involvement is a key category in using safety culture features to predict reported incidents in class low. Involvement refers to staff engagement in decision-making regarding the change needed in the working environment and improvements required, and staff actions to make the change. Engagement, 'involvement in one's work', affects an individual's behaviour and influences overall performance (The Kings Fund, 2014). Staff engagement is linked to safety and quality of care in various industries, including the healthcare (Bakker et al., 2008; Nahrgang et al., 2011; West and Dawson, 2012). In NHS hospitals, a higher level of engagement is linked to a lower level of mortality, a lower level of staff absenteeism in the workplace (West and Dawson, 2012) and a higher number of safety concerns raised by staff (The Kings Fund, 2014). The findings from this study claimed that involvement contributes to the number of reported incidents for the hospitals classified as having a low number of reported incidents. In contrast, no impact has been observed for hospitals with a medium and high number of reported incidents.

This study acknowledges that finding a direct and significant association between safety culture and incident reporting is challenging. The low accuracy rate of the prediction model also revealed this. This study observed different trends in predicting incidents reported in low, medium and high classes. The SHAP graphs helped explain such differences by revealing the individual contribution of each feature in predicting each output class. The study tested multiple ML algorithms and used the findings from the best-performed algorithm to address complexity. LightGBM revealed the best accuracy rate among all algorithms, and ensemble methods (e.g., XGBoost, LightGBM and Extra Tree) outperformed others. A similar conclusion is reached by other researchers, where ensemble methods outperformed regression-based (e.g., logistic regression), instance-based (e.g., k-NN) and margin-based (e.g., support vector machine) methods in a classification problem (Ustebay et al., 2022). This shows the importance of algorithm selection in ML prediction problems.

Using ML in this study offered an in-depth analysis of the relationship between safety culture and incident reporting. However, this study has several limitations. First, this study achieved low accuracy rates on the models despite receiving high rates in some other performance measurement metrics. The low accuracy rate can be explained by data, the nature of the relationship between safety culture and incident reporting, and the way of measuring safety culture. This study used a limited size of data; ML requires large datasets. Also, the COVID-19 period might impact the data and the model performance. The study used staff survey data as a proxy for safety culture. However, the staff survey questions do not cover all components of safety culture: informed, reporting, just, flexible and learning culture, as well as restorative culture. Despite the availability of some safety culture surveys covering some of these components, the characteristics of safety culture can still be challenging to observe and measure, such as leadership commitment (Strauch, 2015). Surveys cannot reflect well on organisations' shared values, beliefs, and behavioural norms (Guldenmund, 2007); surveys hardly provide a holistic view of the level of safety culture as the surveys measure individual attitudes and perceptions (Kilcullen et al., 2022). Those surveys often measure safety climate rather than safety culture (Guldenmund, 2007).

Second, the findings might be influenced by the survey design. The study used survey items under eleven categories with an imbalanced number of survey items under each category. The study could have used the overall category scores instead of taking them individually. However, the study aimed to show the individual impact of each item on the safety outcome by SHAP analysis. To minimise such limitations, ethnographic techniques, such as interviews and observations, are suggested to further understand the level of safety culture in organisations (Baram and Schoebel, 2007; Strauch, 2015).

Third, incident reporting in NHS England hospitals is often voluntary, except if it results in severe harm and death. As it is voluntary, the number of reported incidents might indicate incident reporting behaviour rather than the actual number of incidents (Kaya, 2019). This would make it difficult to interpret findings. Fourth, this study collected data from NHS England; different patterns could have been observed in other countries. This limits the generalisability of the results. Even hospitals in NHS England might have different strategies for managing their safety culture. That is why this study only included data from acute trusts. However, we must note that ML can reveal complex relationships and provide various lessons learned for each type, still needing to use ethnographic techniques to understand the findings further. Despite the limitations, this study facilitates a better understanding of the impacts of safety culture on incident reporting.

## 5. Conclusion

This study explored the association between safety culture and incident reporting using Machine Learning (ML) algorithms to reveal complex relationships. The performances of ML algorithms showed significant differences in predicting the same model, highlighting the importance of algorithm selection in developing the prediction models. With the best-performed algorithm- LightGBM, the highest association was found between safety culture and the number of incidents reported.

This study found that compassionate culture, violence and harassment, and work pressure are critical in predicting reported incident levels. With the explanatory power of SHAP graphs, the findings showed different behavioural patterns on the impact of these categories in predicting incident reporting for different output classes (low, medium and high). This demonstrated the complexity of the relationship. In this study, ML was useful in revealing complex relationships in safety culture; however, its use is limited. We suggest integrating ethnographic techniques to analyse the ML findings further. Future studies can focus on understanding the underlying reasons for the model prediction differences in different output classes, and studies can benefit from the findings to understand safety culture in hospitals.

## CRediT authorship contribution statement

**G. K. Kaya:** Conceptualization, Methodology, Validation, Investigation, Writing- Original draft preparation, Visualization, Writing- Review & Editing. **S. Ustebay:** Methodology, Software, Validation, Formal Analysis, Writing- Review & Editing. **J. Nixon:** Conceptualization, Writing- Review & Editing. **C. Pilbeam:** Conceptualization, Writing- Review & Editing. **M. Sujan:** Conceptualization, Writing- Review & Editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Appendix A. Descriptive statistics for input features

| Features (FeatureID, category and survey items) | Descriptive Statistics(Min, median, max) (+- standard deviation) |
|---|---|
| V1_Year<br>2017 = 1, 2018 = 2, 2019 = 3, 2020 = 4 and 2021 = 5 | (1, 2, 5) (+-1.41) |

(continued on next page)

(*continued*)

| Features (FeatureID, category and survey items) | Descriptive Statistics(Min, median, max) (+- standard deviation) |
|---|---|
| V2_SS_Compassionate culture<br>I feel that my role makes a difference to patients/ service users | (0.81, 0.89, 0.94) (+-0.018) |
| V3_SS_Compassionate culture<br>Care of patients/service users is my organisation's top priority | (0.48, 0.77, 0.91) (+-0.064) |
| V4_SS_Compassionate culture<br>My organisation acts on concerns raised by patients/service users | (0.45, 0.73, 0.87) (+-0.062) |
| V5_SS_Compassionate culture<br>I would recommend my organisation as a place to work | (0.37, 0.62, 0.84) (+-0.085) |
| V6_SS_Compassionate culture<br>If a friend or relative needed treatment, I would be happy with the standard of care provided by this organisation. | (0.41, 0.71, 0.92) (+-0.1) |
| V7_SS_Diversity and equality<br>Does your organisation act fairly with regard to career progression/promotion, regardless of ethnic background, gender, religion, sexual orientation, disability or age? | (0.4, 0.57, 0.72) (+-0.05) |
| V8_SS_Diversity and equality<br>In the last 12 months, have you personally experienced discrimination at work from patients/service users, their relatives or other members of the public? | (0.02, 0.06, 0.17) (+-0.03) |
| V9_SS_Diversity and equality<br>In the last 12 months, have you personally experienced discrimination at work from the manager/team leader or other colleagues? | (0.02, 0.076, 0.17) (+-0.027) |
| V10_SS_Recognition and rewards<br>The recognition I get for good work | (0.41, 0.55, 0.68) (+-0.046) |
| V11_SS_Recognition and rewards<br>The extent to which my organisation values my work | (0.29, 0.45, 0.61) (+-0.058) |
| V12_SS_Recognition and rewards<br>My level of pay. | (0.24, 0.34, 0.47) (+-0.046) |
| V13_SS_Recognition and rewards<br>My immediate manager values my work. | (0.6, 0.71, 0.8_) (+-0.032) |
| V14_SS_Autonomy and control<br>I always know what my work responsibilities are | (0.79, 0.87, 0.93) (+-0.021) |
| V15_SS_Autonomy and control<br>I am trusted to do my job. | (0.86, 0.92, 0.96) (+-0.015) |
| V16_SS_Autonomy and control<br>There are frequent opportunities for me to show initiative in my role | (0.59, 0.73, 0.79) (+-0.029) |
| V17_SS_Involvement<br>I am able to make suggestions to improve the work of my team/department | (0.63, 0.73, 0.83) (+-0.035) |
| V18_SS_Involvement<br>I am involved in deciding on changes introduced that affect my work area/team/department. | (0.4, 0.51, 0.61) (+-0.038) |
| V19_SS_Involvement<br>I am able to make improvements happen in my area of work | (0.43, 0.55, 0.65) (+-0.04) |
| V20_SS_Raising concerns<br>I would feel secure raising concerns about unsafe clinical practice. | (0.57, 0.71, 0.83) (+-0.038) |
| V21_SS_Raising concerns<br>I am confident that my organisation would address my concern. | (0.38, 0.58, 0.76) (+-0.057) |
| V22_SS_Work pressure<br>I am able to meet all the conflicting demands on my time at work. | (0.35, 0.46, 0.63) (+-0.044) |
| V23_SS_Work pressure<br>I have adequate materials, supplies and equipment to do my work. | (0.32, 0.55, 0.74) (+-0.069) |
| V24_SS_Work pressure<br>There are enough staff at this organisation for me to do my job properly | (0.18, 0.31, 0.53) (+-0.059) |
| V25_SS_Violence and harassment<br>The last time you experienced physical violence at work, did you or a colleague report it? | (0.52, 0.67, 0.85) (+-0.053) |
| V26_SS_Violence and harassment<br>The last time you experienced harassment, bullying or abuse at work, did you or a colleague report it? | (0.33, 0.47, 0.60) (+-0.034) |
| V27_SS_Violence and harassment<br>In the last 12 months, have you experienced musculoskeletal problems (MSK) as a result of work activities?" | (0.18, 0.29, 0.39) (+-0.035) |
| V28_SS_Violence and harassment<br>During the last 12 months, have you felt unwell as a result of work-related stress? | (0.28, 0.41, 0.54) (+-0.049) |
| V29_SS_Violence and harassment<br>"In the last three months, have you ever come to work despite not feeling well enough to perform your duties? | (0.38, 0.55, 0.64) (+-0.049) |
| V30_SS_Violence and harassment<br>"In the last 12 months how many times have you personally experienced physical violence at work from…Patients/service users, their relatives or other members of the public? | (0.06, 0.14, 0.24) (+-0.026) |
| V31_SS_Violence and harassment<br>"In the last 12 months how many times have you personally experienced physical violence at work from…Managers? | (0, 0.006, 0.022) (+-0.0044) |
| V32_SS_Violence and harassment<br>In the last 12 months how many times have you personally experienced physical violence at work from…Other colleagues? | (0, 0.016, 0.05) (+-0.0075) |
| V33_SS_Violence and harassment<br>In the last 12 months how many times have you personally experienced harassment, bullying or abuse at work from…Patients/service users, their relatives or other members of the public?" | (0.18, 0.27, 0.4) (+-0.037) |
| V34_SS_Violence and harassment<br>In the last 12 months how many times have you personally experienced harassment, bullying or abuse at work from…Managers | (0.062, 0.13, 0.24) (+-0.031) |
| V35_SS_Violence and harassment<br>In the last 12 months how many times have you personally experienced harassment, bullying or abuse at work from…Other colleagues? | (0.12, 0.19, 0.28) (+-0.03) |
| V36_SS_Team working<br>The team I work in has a set of shared objectives." | (0.62, 0.72, 0.82) (+-0.0326) |

(*continued on next page*)

(*continued*)

| Features (FeatureID, category and survey items) | Descriptive Statistics(Min, median, max) (+- standard deviation) |
|---|---|
| V37_SS_Team working<br>The team I work in often meets to discuss the team's effectiveness | (0.44, 0.58, 0.70) (+-0.046) |
| V38_SS_Line management<br>My immediate manager gives me clear feedback on my work. | (0.48, 0.61, 0.71) (+-0.035) |
| V39_SS_Line management<br>My immediate manager takes a positive interest in my health and well-being | (0.55, 0.68, 0.77) (+-0.035) |
| V40_SS_Motivation<br>I look forward to going to work. | (0.43, 0.58, 0.68) (+-0.045) |
| V41_SS_Motivation<br>I am enthusiastic about my job | (0.6, 0.73, 0.82) (+-0.038) |
| V42_SS_Motivation<br>Time passes quickly when I am working | (0.69, 0.76, 0.84) (+-0.027) |

## Appendix B. Descriptive statistics for output features

| Features (FeatureID, category and survey items) | Class Low (n = 182)(min, median, max) (+-std) | Class Medium (n = 181)(min, median, max) (+-std) | Class High (n = 181)(min, median, max) (+-std) |
|---|---|---|---|
| V46: Median number of days between incidents occurring and being reported to the NRLS | (2444, 6193.5, 7866) (+- 1146.6) | (7869, 9711, 12327) (+- 1302.6) | (12338, 15576, 45740) (+-5678) |
| V47: Total number of incidents occurring | (43.56,75.8, 91.3) (+-10.55) | (44.5, 77.9, 94.1) (+-9.59) | (47.3, 76.6, 91.9) (+-8.81) |
| V48: % of the degree of harm being- none | (6.57, 21.62, 50.97) (+-10.39) | (5.38, 18.98, 50.59) (+-9.21) | (6.69, 20.71, 51.59) (+-8.4) |
| V49: % of the degree of harm being- low | (0.15, 2.38, 10.14) (+-1.62) | (0.24, 1.7, 9.96) (+-1.36) | (0.22, 1.27, 8.83) (+-1.53) |
| V50: % of the degree of harm being- moderate | (0, 0.26, 1.36) (+-0.23) | (0, 0.21, 1.27) (+-0.18) | (0.004, 0.18, 1.36) (+-0.16) |
| V51: % of the degree of harm being- severe | (0, 0.08, 1.83) (+-0.22) | (0, 0.09, 1.31) (+-0.14) | (0, 0.07, 0.65) (+-0.095) |
| V52: % of the degree of harm being- death | (3.19, 11.13, 34.7) (+-4.7) | (4.55, 11.28, 28.14) (+-3.85) | (3.45, 11.52, 80.1) (+-9.81) |
| V53: % incidents related to access, admission, transfer, discharge | (0.96, 5.78, 25.73) (+-3.71) | (1.61, 5.87, 16.2) (+-2.88) | (1.1, 5.5, 14.8) (+-2.95) |
| V54: % incidents related to clinical assessment | (0.57, 3.59, 12.54) (+-2.2) | (0.61, 3.85, 12.37) (+-2.17) | (0.72, 3.62, 10.7) (+-2.07) |
| V55: % incidents related to consent, communication, confidentiality | (1.16, 5.52, 17.35) (+-2.63) | (1.77, 5.85, 18.2) (+-2.87) | (0.55, 5.75, 33.78) (+-3.03) |
| V56: % incidents related to documentation | (2.32, 13.64, 42.98) (+-8.66) | (0.18, 11.89, 48.24) (+-9.86) | (0.07, 16.75, 47.79) (+-11.19) |
| V57: % incidents related to implementation of care and ongoing monitoring/ review | (0.35, 6, 18.5) (+-4.39) | (0.45, 5.47, 17.95) (+-3.81) | (0.32, 4.17, 28.43) (+-3.66) |
| V58: % incidents related to infrastructure (including staffing, facilities and environment) | (3.7, 10.04, 18.84) (+-2.74) | (5.13, 9.94, 40.01) (+-4.49) | (2.04, 9.4, 46.55) (+-5.29) |
| V59: % incidents related to medication | (6.76, 16.16, 42.3) (+-5.41) | (7.76, 15.52, 35.32) (+-4.72) | (3.48, 14.02, 26.06) (+-4.08) |
| V60: % incidents related to the patient accident | (2.95, 10.9, 33.95) (+-4.67) | (4.07, 11.49, 35.55) (+-6.2) | (1.95, 10.17, 52.89) (+-6.19) |
| V61: % incidents related to the treatment procedure | (0.58, 6.28, 34.26) (+-4.32) | (0.96, 6.43, 31.4) (+-4.63) | (1.54, 6.54, 34.45) (+-5.67) |
| V62: % incidents related to all other categories | (20.99, 42.56, 100.74) (+-13.88) | (25.38, 42.7, 88.71) (+-11.58) | (26.78, 49.79, 118.74) (+-16.18) |

## Appendix C. Output feature predictions based on best algorithm accuracy rates

| Feature Prediction | Best Algorithm | Accuracy |
|---|---|---|
| V46: Median number of days between incidents occurring and being reported to the NRLS | Extra Trees Classifier | 0.476216 |
| V47: Total number of incidents occurring | Light Gradient Boosting Machine | 0.584725 |
| V48: % of the degree of harm being- none | Extra Trees Classifier | 0.519609 |
| V49: % of the degree of harm being- low | Light Gradient Boosting Machine | 0.483192 |
| V50:% of the degree of harm being- moderate | Extra Trees Classifier | 0.443763 |
| V51: % of the degree of harm being- severe | Linear Discriminant Analysis | 0.420243 |
| V52: % of the degree of harm being- death | Gaussian Process Classifier | 0.441331 |
| V53: % incidents related to access, admission, transfer, discharge | Light Gradient Boosting Machine | 0.491755 |
| V54: % incidents related to clinical assessment | Light Gradient Boosting Machine | 0.466808 |
| V55: % incidents related to consent, communication, confidentiality | Extreme Gradient Boosting | 0.521723 |
| V56: % incidents related to documentation | Light Gradient Boosting Machine | 0.409514 |
| V57: % incidents related to implementation of care and ongoing monitoring/ review | Extra Trees Classifier | 0.489112 |
| V58: % incidents related to infrastructure (including staffing, facilities and environment) | Extra Trees Classifier | 0.512579 |
| V59: % incidents related to medication | Linear Discriminant Analysis | 0.452537 |
| V60: % incidents related to the patient accident | Extra Trees Classifier | 0.510095 |
| V61: % incidents related to the treatment procedure | Light Gradient Boosting Machine | 0.530708 |
| V62: % incidents related to all other categories | Extra Trees Classifier | 0.454863 |
| V63: incident rate per 1,000 bed days | Extra Trees Classifier | 0.542865 |

## References


Aburumman, M., Newnam, S., Fildes, B., 2019. Evaluating the effectiveness of workplace interventions in improving safety culture: A systematic review. Saf. Sci. 115, 376–392.

Agnew, C., Flin, R., Mearns, K., 2013. Patient safety climate and worker safety behaviours in acute hospitals in Scotland. J. Safety Res. 45, 95–101.

AHRQ, 2016. Hospital survey on patient safety culture: user's guide.

Alkaissy, M., Arashpour, M., Golafshani, E.M., Hosseini, M.R., Khanmohammadi, S., Bai, Y., Feng, H., 2023. Enhancing construction safety: Machine learning-based classification of injury types. Saf. Sci. 162, 106102.

Amalberti, R., Auroy, Y., Berwick, D., Barach, P., 2005. Five system barriers to achieving ultrasafe health care. Ann. Intern. Med. 142, 756–764.

Amarneh, B.H., Al Nobani, F., 2022. The influence of physician-nurse collaboration on patient safety culture. Heliyon 8, e10649.

Amponsah-Tawaih, K., Adu, M.A., 2016. Work Pressure and Safety Behaviors among Health Workers in Ghana: The Moderating Role of Management Commitment to Safety. Saf. Health Work 7, 340–346.

Bakker, A.B., Schaufeli, W.B., Leiter, M.P., Taris, T.W., 2008. Work engagement: An emerging concept in occupational health psychology. Work Stress 22, 187–200.

Baram, M., Schoebel, M., 2007. Safety culture and behavioral change at the workplace. Saf. Sci. 45, 631–636.

Brennan, P.A., Shakib, K., Parmar, S., Dhariwal, D., Oeppen, R.S., 2020. Equality, diversity, culture and ethnicity — we cannot stay silent. Br. J. Oral Maxillofac. Surg. 58, 735–737.

Brennan, P.A., Dalal, A., Knighton, J., Jones, R., Oeppen, R.S., 2021. Pilot study to evaluate safety culture perception in the operating theatres of an acute NHS Trust using the National Air Traffic Services (NATS) App. Br. J. Oral Maxillofac. Surg. 59, 1085–1089.

Brown, K.A., Willis, P.G., Prussia, G.E., 2000. Predicting safe employee behavior in the steel industry: Development and test of a sociotechnical model. J. Oper. Manag. 18, 445–465.

de Zulueta, P., 2021. How do we sustain compassionate healthcare? Compassionate leadership in the time of the COVID-19 pandemic. Clin. Integr. Care 8, 100071.

Dekker, S.W.A., 2020. Safety after neoliberalism. Saf. Sci. 125, 104630.

Dunstan, E., Cook, J.L., Coyer, F., 2019. Safety culture in intensive care internationally and in Australia: A narrative review of the literature. Aust. Crit. Care.

England, N.H.S., 2022a. National patient safety incident reports. NHS Engl. https://www.england.nhs.uk/patient-safety/national-patient-safety-incident-reports/#data-workbooks-and-commentary-official-statistics.

NHS England, 2022. Safety culture: learning from the best practice. URL https://www.england.nhs.uk/long-read/safety-culture-learning-from-best-practice/.

NHS England, 2023. We are compassionate and inclusive. URL https://www.england.nhs.uk/looking-after-our-people/the-programme-and-resources/we-are-compassionate-and-inclusive/.

EUROCONTROL, 2018. Safety Culture Discussion Cards.

Flin, R., 2007. Measuring safety culture in healthcare: A case for accurate diagnosis. Saf. Sci. 45, 653–667.

The Health Foundation, 2011. Does improving safety culture affect patient outcomes? London.

Goncalves Filho, A.P., Waterson, P., 2018. Maturity models and safety culture: A critical review. Saf. Sci. 105, 192–211.

Groves, P.S., 2014. The Relationship Between Safety Culture and Patient Outcomes: Results From Pilot Meta-Analyses. West. J. Nurs. Res. 36, 66–83.

Guldenmund, F.W., 2007. The use of questionnaires in safety culture research - an evaluation. Saf. Sci. 45, 723–743.

Guldenmund, F.W., 2010. (Mis)understanding Safety Culture and Its Relationship to Safety Management. Risk Anal. 30, 1466–1480.

Halligan, M., Zecevic, A., 2011. Safety culture in healthcare: a review of concepts, dimensions, measures and progress. BMJ Qual. Saf. 20, 338–343.

Hudson, P., 2007. Implementing a safety culture in a major multi-national. Saf. Sci. 45, 697–722.

Kalteh, H.O., Salesi, M., Cousins, R., Mokarami, H., 2020. Assessing safety culture in a gas refinery complex: Development of a tool using a sociotechnical work systems and macroergonomics approach. Saf. Sci. 132, 104969.

Kaya, G.K., 2019. The Relationship Between Risk Management and Patient Safety Incidents in Acute Hospitals in NHS England, in: Calisir F., Cevikcan E., C.A.H. (Ed.), Industrial Engineering in the Big Data Era. Lecture Notes in Management and Industrial Engineering. Springer, Cham, pp. 471–479.

Khamisa, N., Oldenburg, B., Peltzer, K., Ilic, D., 2015. Work related stress, burnout, job satisfaction and general health of nurses. Int. J. Environ. Res. Public Health 12, 652–666.

Kilcullen, M.P., Bisbey, T.M., Ottosen, M.J., Tsao, K., Salas, E., Thomas, E.J., 2022. The Safer Culture Framework: An Application to Healthcare Based on a Multi-Industry Review of Safety Culture Literature. Hum. Factors 64, 207–227.

Kirkup, B., 2022. Reading the signals. Maternity and neonatal services in East Kent- the Report of the Independent Investigation, HC 681.

Liu, J., Gan, Y., Jiang, H., Li, L., Dwyer, R., Lu, K., Yan, S., Sampson, O., Xu, H., Wang, C., Zhu, Y., Chang, Y., Yang, Y., Yang, T., Chen, Y., Song, F., Lu, Z., 2019. Prevalence of workplace violence against healthcare workers: A systematic review and meta-analysis. Occup. Environ. Med. 76, 927–937.

Lundberg, S., Lee, S.I., 2017. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774.

Maneechaeye, P., Potipiroon, W., 2022. The impact of fleet-level and organization-level safety climates on safety behavior among Thai civilian pilots: The role of safety motivation. Saf. Sci. 147, 105614.

Mohanty, S.D., Lekan, D., McCoy, T.P., Jenkins, M., Manda, P., 2022. Machine learning for predicting readmission risk among the frail: Explainable AI for healthcare. Patterns 3, 100395.

Morello, R.T., Lowthian, J.A., Barker, A.L., McGinnes, R., Dunt, D., Brand, C., 2013. Strategies for improving patient safety culture in hospitals: a systematic review. BMJ Qual. Saf. 22, 11–18.

Mustapha, T., Ho, Y., Andrews, J.S., Cullen, M.J., 2019. See No Evil, Hear No Evil, Stop No Evil: Institutional-Level Tracking to Combat Mistreatment of Residents and Fellows. J. Grad. Med. Educ. 11, 601–605.

Nahrgang, J.D., Morgeson, F.P., Hofmann, D.A., 2011. Safety at Work: A Meta-Analytic Investigation of the Link Between Job Demands, Job Resources, Burnout, Engagement, and Safety Outcomes. J. Appl. Psychol. 96, 71–94.

Panesar, A., 2021. Machine Learning and AI for Healthcare, Second, edi. ed. Apress, Coventry, UK, Machine Learning and AI for Healthcare.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Skit-learn: machine learning in python. JMLR 12, 2825–2830.

Reason, J., 1998. Achieving a safe culture: Theory and practice. Work Stress 12, 293–306.

Saedi, A.M., Ab. Majid, A., Isa, Z., 2020. Relationships between safety climate and safety participation in the petroleum industry: A structural equation modeling approach. Saf. Sci. 121, 240–248.

Schwendimann, R., Milne, J., Frush, K., Ausserhofer, D., Frankel, A., Sexton, J.B., 2013. A Closer Look at Associations Between Hospital Leadership Walkrounds and Patient Safety Climate and Risk Reduction: A Cross-Sectional Study. Am. J. Med. Qual. 28, 414–421.

Seo, D.C., 2005. An explicative model of unsafe work behavior. Saf. Sci. 43, 187–211.

Sexton, J.B., Helmreich, R.L., Neilands, T.B., Rowan, K., Vella, K., Boyden, J., Roberts, P. R., Thomas, E.J., 2006. The Safety Attitudes Questionnaire: Psychometric properties, benchmarking data, and emerging research. BMC Health Serv. Res. 6, 1–10.

Shi, B., He, Y., Lee, J., Huang, Y. hsiang, Li, Y., 2022. Safety climate profiles in remote workers: Association with key predictors and outcomes at the team level. Saf. Sci. 145, 105477.

Shore, L.M., Randel, A.E., Chung, B.G., Dean, M.A., Ehrhart, K.H., Singh, G., 2011. Inclusion and diversity in work groups: A review and model for future research. J. Manage. 37, 1262–1289.

Simsekler, M.C.E., Qazi, A., 2022. Adoption of a Data-Driven Bayesian Belief Network Investigating Organizational Factors that Influence Patient Safety. Risk Anal. 42, 1277–1293.

Simsekler, M.C.E., Qazi, A., Alalami, M.A., Ellahham, S., Ozonoff, A., 2020. Evaluation of patient safety culture using a random forest algorithm. Reliab. Eng. Syst. Saf. 204, 107186.

Simsekler, M.C.E., Rodrigues, C., Qazi, A., Ellahham, S., Ozonoff, A., 2021. A comparative study of patient and staff safety evaluation using tree-based machine learning algorithms. Reliab. Eng. Syst, Saf, p. 208.

Sorra, J., Khanna, K., Dyer, N., Mardon, R., Famolaro, T., 2012. Exploring relationships between patient safety culture and patients' assessments of hospital care. J. Patient Saf. 8, 131–139.

Steyrer, J., Schiffinger, M., Huber, C., Valentin, A., Strunk, G., 2013. Attitude is everything?: The impact of workload, safety climate, and safety tools on medical errors: A study of intensive care units. Health Care Manage. Rev. 38, 306–316.

Strauch, B., 2015. Can we examine safety culture in accident investigations, or should we? Saf. Sci. 77, 102–111.

The Kings Fund, 2014. Improving NHS Care by Engaging Staff and Devolving Decision-Making, The Kings Fund.

The University of Manchester, 2006. Manchester Patient Safety Framework (MaPSaF) Acute, NPSA.

Ustebay, S., Sarmis, A., Kaya, G.K., Sujan, M., 2022. A comparison of machine learning algorithms in predicting COVID-19 prognostics. Intern. Emerg. Med.

Weaver, S.J., Weeks, K., Pham, J.C., Pronovost, P.J., 2014. On the CUSP: Stop BSI: Evaluating the relationship between central line-associated bloodstream infection rate and patient safety climate profile. Am. J. Infect. Control 42. S203–S208.

West, M.A., 2021. Compassionate leadership sustaining wisdom, humanity and presence in health and social care. The Swirling Leaf Press.

West, M.A., Dawson, J.F., 2012. Employee engagement and NHS performance. London.

Yang, B.X., Stone, T.E., Petrini, M.A., Morris, D.L., 2018. Incidence, Type, Related Factors, and Effect of Workplace Violence on Mental Health Nurses: A Cross-sectional Survey. Arch. Psychiatr. Nurs. 32, 31–38.

Zhang, Y., Haghani, A., 2015. A gradient boosting method to improve travel time prediction. Transp. Res. Part C Emerg. Technol. 58, 308–324.

# Exploring the impact of safety culture on incident reporting: lessons learned from machine learning analysis of NHS England staff survey and incident data

Kaya, Gulsum Kubra

2023-10-01



Kaya GK, Ustebay S, Nixon J, et al., (2023) Exploring the impact of safety culture on incident reporting: lessons learned from machine learning analysis of NHS England staff survey and incident data. Safety Science, Volume 166, October 2023, Article Number 106260

https://doi.org/10.1016/j.ssci.2023.106260