This is a preprint draft. The published article can be found at: https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525

Please cite this paper as:

Naser M.Z. (2021). "Observational Analysis of Fire-induced Spalling through Data Science and Machine Learning." ASCE Journal of Materials in Civil Engineering. https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525.

Observational Analysis of Fire-induced Spalling of Concrete through Ensemble Machine Learning and Surrogate Modeling

M.Z. Naser, PhD, PE

Assistant professor, Glenn Department of Civil Engineering, Clemson University, E-mail: mznaser@clemson.edu, m@mznaser.com, Website: www.mznaser.com

Abstract

Despite ongoing research efforts, we continue to fall short of arriving at a consistent representation of fire-induced spalling of concrete. This is often attributed to the complexity and randomness of spalling as well as our persistence in favoring traditional approaches as a sole mean to examine this phenomenon. With the hope of bridging this knowledge gap, this paper demonstrates how utilizing surrogate modeling via data science and machine learning algorithms can provides us with valuable insights into fire-induced spalling. In this study, nine algorithms namely; Naive Bayes, generalized linear model, logistic regression, fast large margin, deep learning, decision tree, random forest, gradient boosted trees, and support vector machine, are applied to analyze observations obtained from 185 fire tests (collected over the last 65 years). The same algorithms were also applied to identify key features that govern the tendency of fire-induced spalling in reinforced concrete columns and to develop tools for instantaneous prediction of spalling. The results of this comprehensive analysis highlight the merit in utilizing modern computing techniques in structural fire engineering applications given their extraordinary ability to comprehend multi-dimensional phenomena with ease, high predictivity, and potential for continuous improvement.

Keywords: Concrete; Fire; Spalling; Machine learning; Artificial intelligence.

This is a preprint draft. The published article can be found at: https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525

Please cite this paper as:

Naser M.Z. (2021). "Observational Analysis of Fire-induced Spalling through Data Science and Machine Learning." ASCE Journal of Materials in Civil Engineering. https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525.

1.0 Introduction

Fire-induced spalling continues to be a problem of high interest to the structural fire engineering and safety community (Kodur 2018; Missemer et al. 2019). This is attributed to the fact that not only spalling has the potential to cause damage at the member level, but may also trigger partial, and in few instances complete, collapse of structural systems (Meacham et al. 2009). Given the resiliency of concrete to extreme loadings; and hence its favorable use in various constructions, the propensity of concrete to spall when heated complicates the design of concrete structures. Such complications arise from: 1) inadequate recognition to this phenomenon in codal provisions, and 2) the lack of calculation/prediction methods that can be applied to examine the vulnerability of concrete structures to spalling (ACI216.1 2014; BSI and European Committee for Standardization 2004). The above two observations can be credited to the absence of a comprehensive understanding of fire-induced spalling – a notion that has been duly noted by a number of notable studies (Ali et al. 2004; Kalifa et al. 2001; Kodur and Naser 2020; Liu et al. 2018; Naser 2019a; Phan 2008).

On the positive side, a collection of observations from previous works have qualitatively demonstrated few generalizations of spalling phenomenon. For instance, structural members made of high strength concrete made (of compressive strength about or exceeding 45 MPa) seem to be more vulnerable to spall than those made of normal strength concrete (Kodur et al. 2001). Similarly, axially loaded members have been noted to spall more so than flexurally loaded members (Dotreppe et al. 1997; Naser 2019b). Further, columns with conventional ties (hooked at 90°) seem to have lower resistance to spalling than columns with improved hooked configurations (i.e. hooked at 135°) (Kodur et al. 2013). More recently, the use of steel and polypropylene fibers were also noticed to have a positive influence on limiting spalling of concrete (Kalifa et al. 2001). Based on the aforementioned generalizations, some attempts presented methodologies to evaluate fire-induced spalling in concrete structures (Dwaikat and Kodur 2009; Jansson 2008; Kodur and Dwaikat 2008; Zhao et al. 2014). However, these methods remain: 1) only applicable to certain scenarios, 2) require development of highly complex/multi-stage finite element models, 3) lack proper validation, and 4) involve a number of assumptions that oversimplify the phenomenon of spalling (Bazant 1997; Liu et al. 2018; Peng 2000; Phan and Carino 2002; Sanjayan and Stocks 1993). As such, the applicability of such approaches remains limited and inadequate to practical/real scenarios and to this date, we still lack a well-established approach that can be followed to examine the tendency of a reinforced concrete (RC) member to spall (Ali et al. 2004; Kalifa et al. 2001; Liu et al. 2018; Naser 2019a; Phan 2008).

On another note, the lack of a general understanding and/or an approach that can be applied to predict fire-induced spalling of concrete also stems from the complexity and randomness of this phenomenon. On one side, spalling is expected to be governed by a multitude of factors spanning a multi-dimensional paradigm (i.e. covering: material, geometric, and heating/loading features). This brings in issues on two fronts. The first, the uniqueness of this phenomenon implies the need for a state-of-the-art, comprehensive, and collaborative research that is properly designed and executed – sadly, reports from recent efforts proved that pursuing such a program is challenging to arrange or plan (Hertz 2003; Kalifa et al. 2001; RILEM 1994). The second front can be summed by the fact that the majority of available incremental works, which primarily applied traditional engineering methods, do not seem to properly converge – due to differences in testing set-ups,

This is a preprint draft. The published article can be found at: https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525

Please cite this paper as:

Naser M.Z. (2021). "Observational Analysis of Fire-induced Spalling through Data Science and Machine Learning." ASCE Journal of Materials in Civil Engineering. https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525.

materials compositions, assumptions used in modeling/deriving theories etc. (Bazant 1997; Gawin et al. 2018; Liu et al. 2018; Peng 2000; Phan and Carino 2002; Sanjayan and Stocks 1993).

This opens up an opportunity to attempt to examine fire-induced spalling phenomenon through a perspective that is built on a modern hypothesis. This proposition states that if spalling observations are compiled from actual fire tests, then it is possible to link such observations to material, geometric, and heating/loading characteristics of fire-tested structural members via surrogate modeling (often used when the outcome of interest cannot be easily directly measured, so a model of the outcome is applied instead). In this scenario, this linkage (relation) between all involved factors is highly nonlinear and hence arriving at such a relation may not be possible using traditional engineering methods; however, could still be arrived at using data science and machine learning (DS+ML) as these techniques are specifically designed to unbox hidden relations/patterns embodied in large sets of data (Naser 2020). At the time of this work, very few studies applied technologies such as artificial neural networks (ANNs) to examine the fire-induced spalling phenomenon (i.e. (McKinney and Ali 2014)). Unfortunately, these studies share common features: 1) applied outdated approaches, 2) examined limited number of specimens, and 3) did not fully utilize a variety of algorithms or contemporary analysis solutions.

This work utilizes advanced computations (data science) and machine learning algorithms namely; Naive Bayes, generalized linear model, logistic regression, fast large margin, deep learning, decision tree, random forest, gradient boosted trees, and support vector machine, to analyze spalling observations from 185 fire tests carried out on full scale reinforced concrete (RC) columns. This work also utilizes the above algorithms to identify critical parameters/features that govern the fire-induced spalling of RC columns as to enable developing tools that can instantaneously predict such phenomenon with high accuracy. Given that the collected data: 1) was obtained from actual fire tests (rather than simulated/controlled responses), and 2) was analyzed through novel algorithms, then the results of this comprehensive analysis is expected to truly capture actual spalling tendency in RC columns.

2.0 An Overview to Data Science and Machine Learning Algorithms

This section provides an overview on data science and machine learning algorithms given that the majority of structural fire and safety researchers/practitioners may not be commonly exposed to such techniques given the nature of their training and work atmosphere. Thus, a succinct overview is presented herein and an in-depth review can be found elsewhere (Jordan and Mitchell 2015; Schmidhuber 2015).

Data science, often referred to as data mining and/or data analytics, is a multi-disciplinary field that applies novel scientific methods and frameworks to process observations and to develop algorithms and systems that can be efficiently used to extract meaningful knowledge and insights from collected datapoints; often relating to actual/real phenomena (Dhar and Vasant 2013). The process of data science can be summarized by the flowchart shown in Fig. 1. In this flowchart, utilizing data science starts by formulating a hypothesis and then collecting raw data on a given phenomenon (i.e. fire-induced spalling). This data is then pre-processed (cleansed) to remove outliners/noise. Afterwards, the cleansed data is investigated using a (or a collection of) algorithm(s)/technique(s) to arrive at a suitable model to predict the phenomenon in hand. When necessary, the developed model can be further enhanced for improved optimization and predictability. Then, the validated model can be applied into real world applications. Finally, the applied model can undergo a series of upgrades (tunings) given new datapoints and/or through addition of new observations etc.

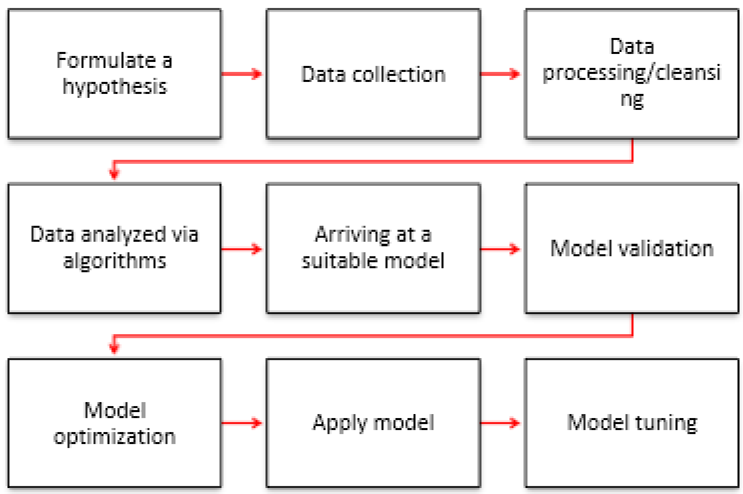


Fig. 1 Typical data science "machine learning" process

Machine learning is a subset of data science and primarily focuses on the ability of machines to receive, understand and learn datapoints to identify key features in order to arrive at a suitable representation that best demonstrates the phenomenon embodied within a dataset (Sayad et al. 2019). Machine learning can come in handy in practical scenarios, where mathematical or conventional modelling approaches become obsolete as a result of limitation of precise reasoning in modeling multi-dimensional problems and uncertainties arising from the complexity of a given phenomenon etc. Machine learning can be broadly grouped into supervised, unsupervised and semi-supervised learning based on the type of available datapoints (i.e. labeled/not labeled etc.) as well as type of phenomenon under investigation (regression, classification etc.) (Bishop 2006). A number of machine learning algorithms have been developed over the past few years and those of interest to this work are highlighted herein. It is worth noting that predicting spalling tendency falls into a supervised classification investigation.

2.1 Naive Bayes

Naive Bayes (NB), an algorithm commonly used in supervised classification problems, is easy to develop, construe and apply to large dataset. This algorithm allows constructing rules that serve for predicting observations (i.e. column spalled/did not spall) and does not require complicated iterative procedure (but still involve evaluating a series of conditional probabilities). In this algorithm, a set of attributes, i.e. $x_1, \ldots, x_n$ (such as compressive strength of concrete etc.), for a fire-exposed RC column can be grouped under a class observation – Y (e.g. column spalls). This arrangement aims to maximize the posterior probability of the class variable given the set of attributes through the following relations (Shiri Harzevili and Alizadeh 2018):

$$arg\ maxP_{c\in C}(Y|x_1, \ldots, x_n) \quad (1)$$

The application of Bayes rule in classification is formulated as:

$$P_{c\in C}(Y|x_1, \ldots, x_n) = \frac{P(Y)P(x_1, \ldots, x_n|Y)}{P(x_1, \ldots, x_n|Y)} \quad (2)$$

and further simplifies to:

$$arg\ maxP_{c\in C}(Y)\prod_{i=1}^{n} P(x_i|Y) \quad (3)$$

2.2 Generalized linear model

The generalized linear model (GLM) extends traditional linear models through fitting generalized linear models by maximizing the log-likelihood of a dataset. The GLM fitting computation is quick, and efficiently scales for phenomena with limited predictors; especially those having nonzero coefficients. In this algorithm, the outcome class (Y) of a phenomenon is assumed to be a linear combination of the coefficients (β) and attributes ($x_1, \ldots, x_n$) such that:

$$Y = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_n x_{n,i} \quad (4)$$

This model is often supplemented with a function that determines how the mean depends on the linear predictors and a variance function that describes how the variance depends on the mean.

2.3 Logistic regression

Logistic regression chooses to maximize the likelihood of observing an event and hence is often used in scenarios where the outcome of a phenomenon is dichotomous (binary). This algorithm approximates the multi-linear regression function shown below:

$$logit(p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \quad (5)$$

where p is the probability of presence of the event (i.e. spalled/did not spall). The logit transformation is defined as the logged odds:

$$odds = \frac{p}{1-p} \quad (6)$$

and,

$$logit(p) = \ln\left(\frac{p}{1-p}\right) \quad (7)$$

2.4 Fast large margin

Fast large margin (FLM) algorithm seeks to minimize error rate as well as to separate the outcome of an observation (i.e. spalling, no spalling) by the largest probable margin. As such, this algorithm often achieves a good generalization on new datapoints. FLM can be applied through the following relation (Cheng et al. 2009):

$$\forall s \neq y, D(x, y) > D(x, s) + pH(s, y) \tag{8}$$

where, *(x, y)* denotes an observation sequence and its ground correct label, *H(s, y)* is the Hamming distance between two hidden state sequences of the same length, and ρ is a constant margin scaling factor that is greater than zero.

2.5 Deep learning

Deep learning (DL) is another supervised learning algorithm that learn iteratively. This technique consists of a number of layers: an input layer, multi-intermediate layers, and an output layer (Shahin et al. 2009). Each of these layers contains a number of neurons that process datapoints. DL mimics human brain and cognitive processing and hence has the ability to work on incomplete data and to perform analysis in a parallel computing platform. DL can be applied in binary and multi-outcome problems as can be seen below (Behnood and Golafshani 2018).

$$net_j = \sum_{i=1}^{n} In_i w_{ij} + b_j \tag{9}$$

$$Y = f(net_j) \tag{10}$$

where, In_i and b_j are the ith input signal and the bias value of jth neuron, respectively, w_{ij} is the connecting weight between ith input signal and jth neuron and f is an activation function such as hyperbolic tangent sigmoid.

2.6 Decision tree

A decision tree (DT) is a support tool that utilizes a tree-like model comprising of decisions and their possible consequences. Such a tree is produced by splitting a dataset into branch-like arrangement where a decision tree starts at a root node. DT is favorable in classification problems as it provides schematic representation of the outcome of analysis, thus becoming of high value to trace specifics in a given DT simulation (Safavian and Landgrebe 1991). Depending on the observation in hand, impurity measures (e.g. Gini) can be used to process datapoints. For example, for a node t, Gini index $g(t)$ is defined as (Chou et al. 2014):

$$g(t) = \sum_{j \neq i} p(j|t)p(i|t) \quad (11)$$

where i and j are target field categories.

$$p(j|t) = \frac{p(j,t)}{p(t)}; p(jt) = \frac{\pi(j)N_j(t)}{N_j}; \text{ and } p(t) = \sum_j p(j,t) \quad (12)$$

where, $\pi(j)$ is the prior probability for category j, $N_j(t)$ is the number of records in category j of node t, and N_j is the number of records of category j in the root node.

2.7 Random forest

Random forest (RF) is an algorithm that capitalizes on principles of ensemble learning (in which a specific algorithm is applied multiple times in an analysis, and/or where different types of algorithms are joined together to form a more powerful prediction model) (Liaw and Wiener 2002). A typical formulation of RF is presented herein:

$$Y = \frac{1}{J}\sum_{j=1}^{J} c_{j,full} + \sum_{k=1}^{K}\left(\frac{1}{J}\sum_{j=1}^{J} contribution_j(x,k)\right) \quad (13)$$

where, J is the number of trees in the forest, k represents a feature in the observation, K is the total number of features, c_{full} is the average of the entire dataset (initial node).

2.8 Gradient boosted trees

Gradient boosted trees (GBT) is a machine learning technique that forms an ensemble of DT models of low prediction ability through optimization of an arbitrary-developed differentiable loss function (see Eq. 9). GBT only uses a small part of training datasets for increasing computation speed and accuracy of prediction. GBT iteratively corrects developed ensembles by comparing iterative predictions against true observations. As such, the next iteration will help correct for previous mistakes.

$$Y = \sum_{k=1}^{M} f_k(x_i) \, , f_k \in F = \{f_x = w_{q(x)}, q: R^p \to T, w \in R^T\} \tag{14}$$

where, M is additive functions, T is the number of leaves in the tree, w is a leaf weights vector, w_i is a score on i-th leaf, and $q(x)$ represents the structure of each tree that maps an observation to the corresponding leaf index (Intel 2019).

2.9 Support vector machine

Support vector machine (SVM) is an algorithm that can be used in classification. SVM determines the best method to distinguish between classes in the training data. SVM is very accurate and this accuracy comes as a result of intensive calculations (Hou et al. 2018). An interesting feature of SVM is that errors smaller than a set threshold (hinge loss) ε do not contribute to the overall error measure such that:

$$L(Y_i - \widehat{Y}_i) = \begin{cases} 0 & \text{if} |Y_i - \widehat{Y}_i| < \epsilon \\ |Y_i - \widehat{Y}_i| - \epsilon & \text{if} |Y_i - \widehat{Y}_i| > \epsilon \end{cases} \tag{15}$$

SVM seeks to fit a model of the form,

$$\widehat{Y}(x) = \sum_{i=1}^{N} c_i k(x, x_i) \tag{16}$$

where, the parameters c_i are referred to as choice coefficients, and $k(x, x_i)$ is defined as the Gaussian kernel function (Young et al. 2019).

3.0 Rationale and Database Development

Performing a fire-based data science/machine learning (DS+ML) analysis is quite different than traditional analysis methods (i.e. hand calculations, finite element/difference simulations etc.), in which the former does not require discretization, temperature-dependent material properties/constitutive models nor a multi-stage (hydro, thermal and structural) analysis. In DS+ML, fire-induced spalling can be evaluated through intelligent algorithms that analyze fire test observations to arrive at an understanding of this phenomenon.

The rationale behind utilizing DS+ML to examine fire-induced spalling of RC columns stems from the following hypothesis, "*if spalling observations are collected from fire tests, then it is possible to apply intelligent algorithms to analyze such observations to arrive at an understanding of spalling – or at the very least to identify the key factors that influence this phenomenon*". Since a number of factors (i.e. compressive strength, restraint conditions etc.) have already been shown to influence occurrence of spalling in RC columns, and yet we do not actually know the quantitative importance of such factors (from spalling point of view), then analyzing this dilemma through DS+ML becomes attractive as such techniques are primarily developed to solve complex real world phenomena. As utilizing DS+ML to evaluate a phenomenon (which in this case is fire-induced spalling) requires the availability of a well-prepared database, thus a comprehensive literature review was carried out to locate commissioned fire testing reports as well as research works that tested RC columns under standard fire conditions (Dotreppe et al. 1997; Hass 1986; Hertz 2003; Khoury 2000; Kodur et al. 2001; Kodur 2018; Kodur and McGrath 2003; Lie and Woollerton 1988; Liu et al. 2018; Myllymaki and Lie 1991; Rodrigues et al. 2010; Schneider 1988) (see Fig. 2).

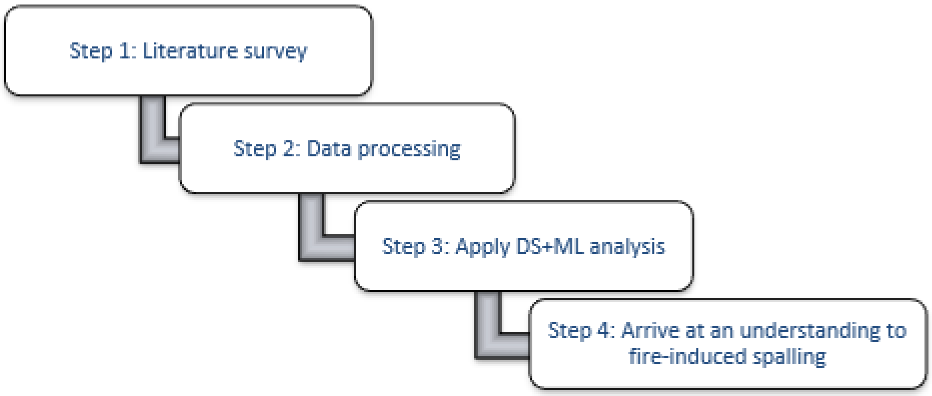


Fig. 2 Framework of proposed methodology

This literature survey focuses on collecting datapoints on material, geometric, and loading aspects of each of the tested columns as well as occurrence of spalling. The developed database compiled data on 248 fire tests, all of which were conducted on full scale RC columns and spanned the time period between 1953-2018. Due to differences between researchers' backgrounds, as well as norms of documentation at the time of reporting outcome of fire tests – some studies did not report information on certain features, and thus only 185 RC columns were deemed suitable for analysis. For the sake of this study, all selected columns were tested under standard fire conditions, thus neutralizing the effect of varying thermal/heating loading. In addition, this work maintains the common notion of identifying spalling qualitatively and with binary notion (spalling/no spalling) due to the absence of actual measurements during collected tests and/or tools to quantitatively measure fire-induced spalling.

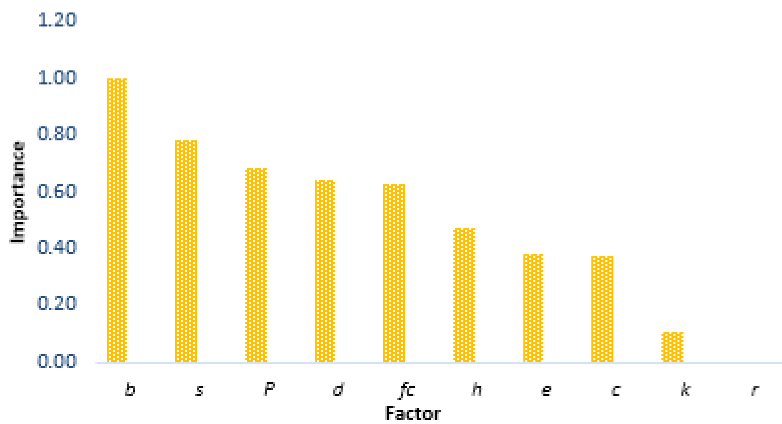
The collected data on these columns covered 10 independent parameters: strength of concrete, f_c, cross sectional breadth and height, b and d, boundary conditions, k, tie spacing, s, tie diameter, d, steel reinforcement ratio, r, magnitude, P, and eccentricity of applied loading, e. This collection of observations is then arranged into a database. For convenience, the compiled database is provided herein and is listed in Table 1. This database is compiled from the works of Lie and Woollerton (Lie and Woollerton 1988), Buch and Sharma (Buch and Sharma 2019), Shah and Sharma (Shah and Sharma 2017), Myllymi and Lie (Myllymaki and Lie 1991), Rodrigues et al. (Rodrigues et al. 2010), Kodur et al. (Kodur et al. 2001, 2005, 2003), Thomas and Webster (Thomas and Webster 1953), as well as Davey and Ashton (Davey and Ashton 1953). Full details on these specimens can be found in their respective references.

4.0 Data Science and Machine Learning (DS+ML) Analysis

Now that the database is compiled, this database is ready to be analyzed using an observational DS+ML based approach. First, the database was randomly arranged to eliminate any biasness arising from a particular study/factor (feature). Then, the dataset was split into a model development set (for training and validation purposes) (80%) and a testing set (20%) which was used for evaluating performances of applied ML algorithms (Barber 2012). The database was then analyzed using the collection of ML algorithms listed in Sec. 2 (through commonly available codes (Brownlee 2019)). The outcome of this analysis is presented herein from two perspectives: 1) identified critical factors that triggers spalling from each algorithm's point of view, and 2) comparison between algorithms' accuracy in predicting spalling of RC columns. The importance of this analysis is to identify "which are the key factors with highest impact on spalling" to allow fire researchers/designers from easily evaluating the tendency of spalling with a level of confidence that is not currently available for them. For parallel works that use a much larger spectrum of factors, the readers may refer to the following (McKinney and Ali 2014; Naser 2019b; c; Naser and Seitllari 2019; Seitlllari and Naser 2019).

4.1 Critical factors governing fire-induced spalling phenomenon

The DS+ML analysis can provide a preliminary view that is independent of the modeling algorithm into the global importance of each of the selected input variables as to predict spalling. For instance, Fig. 3a shows that breadth, *b*, tie spacing, *s*, applied loading, *P*, tie diameter, *d*, and compressive strength of concrete, f_c are the main factors that affect spalling. Another visualization that can also be arrived at through this preliminary analysis is the correlation matrix. This matrix shows how each of the selected variables correlates to the occurrence of spalling (i.e. positively – increases likelihood of spalling, or negatively – reduces likelihood of spalling). For example, it can be seen that if the geometric size of a column increases, then such a column is more likely to spall due to the positive correlation between *b* and *h* with tendency to spalling (highlighted in green). Similarly, if tie diameter, *d*, or ratio of steel reinforcement, *r*, increases, then the column is expected to be less likely to spall (as there is a negative correlation between *d* and *r* with tendency to spalling – highlighted in red). Overall, the factors that seem to have a positive correlation with spalling (i.e. spalling is likely to occur if these factors increase) include: breadth and height of column, compressive strength of concrete, pinned restraint conditions, loading magnitude eccentricity and tie spacing. On the other hand, the factors with negative correlation to occurrence of spalling (i.e. an increase in these factors would reduce the tendency to spall and hence increase the tendency not to spall) include: cover distance, ratio of longitudinal reinforcement, and tie diameter.



(a) Importance of input factors

	b	*c*	*d*	*e*	f_c	*h*	*k*	*P*	*r*	*s*	*SP = yes*
b	1.00	-0.02	0.07	0.00	0.20	0.76	-0.12	0.57	-0.20	0.11	0.28
c	-0.02	1.00	0.33	0.26	0.39	0.01	0.19	0.31	0.19	0.47	-0.13
d	0.07	0.33	1.00	-0.04	0.02	0.10	-0.23	0.14	0.02	0.24	-0.19
e	0.00	0.26	-0.04	1.00	0.30	-0.02	0.57	0.14	0.00	0.10	0.13
f_c	0.20	0.39	0.02	0.30	1.00	0.16	0.10	0.71	0.01	0.30	0.19
h	0.76	0.01	0.10	-0.02	0.16	1.00	-0.13	0.46	-0.20	0.11	0.15
k	-0.12	0.19	-0.23	0.57	0.10	-0.13	1.00	-0.07	0.02	0.18	0.07
P	0.57	0.31	0.14	0.14	0.71	0.46	-0.07	1.00	0.07	0.27	0.20
r	-0.20	0.19	0.02	0.00	0.01	-0.20	0.02	0.07	1.00	0.18	-0.04
s	0.11	0.47	0.24	0.10	0.30	0.11	0.18	0.27	0.18	1.00	0.23
SP = yes	0.28	-0.13	-0.19	0.13	0.19	0.15	0.07	0.20	-0.04	0.23	1.00

(b) Correlation matrix

Fig. 3 Basic outcome of DS+ML analysis

In DS+ML analysis, it is common to cleanse a database in order to minimize effects of noise and outliners as well as to reduce the number of governing input parameters – while maintaining high: 1) accuracy in understanding the phenomenon on hand, and 2) prediction capability. This process is often referred to as feature engineering and comprises of two components; feature generation and feature extraction. Feature generation is the process of combining two (or more) input parameters to yield a new parameter that has a much greater influence on undersetting or predicting the given phenomenon. On the other hand, feature extraction is a reduction process in which the total dimension (i.e. 10 inputs) of spalling phenomenon is reduced to manageable groups (features) with the condition that these extracted features can still accurately and fluently describe the original dataset (listed in Table 1). Based on feature engineering process, interactions between extracted features and spalling tendency may turn much more important than those with highest correlations (i.e. shown in Fig. 3b) and hence the effect of feature engineering was investigated here. While the application of the feature engineering can be complex and resource intensive, fortunately, there is a number of pre-developed codes and software that can be applied to help facilitate carrying out such analysis (Matlab 2019).

The outcome of the DS+ML analysis, when supplemented with feature engineering, is listed in Table 2 and is also plotted in Fig. 4. This outcome shows that the main re-occurring inputs between all applied algorithms are compressive strength of concrete and diameter of ties; followed by breadth of column and spacing of ties. It can be also seen that this analysis yields a slightly different outcome than that shown in Fig. 3a as it accounts for all interactions within each of the inputs (and not just inputs with the spalling as a phenomenon). By normalizing the results obtained from the above two analyses (with and without feature engineering), the outcome of this work shows that diameter of ties, compressive strength of concrete, breadth of column, and spacing of ties are the governing factors of fire-induced spalling in the RC columns examined herein – with all other factors having minor contributions (see Fig. 4). In other words, this analysis infers that it is possible to predict spalling in a RC column with high confidence through evaluating the identified four factors listed in Table 2 and highlighted in grey – rather than all 10 factors listed in Table 1; hence further simplifying prediction of fire-induced spalling as will be shown in Sec. 4.2.

Table 2 Outcome of DS+ML analysis with feature engineering

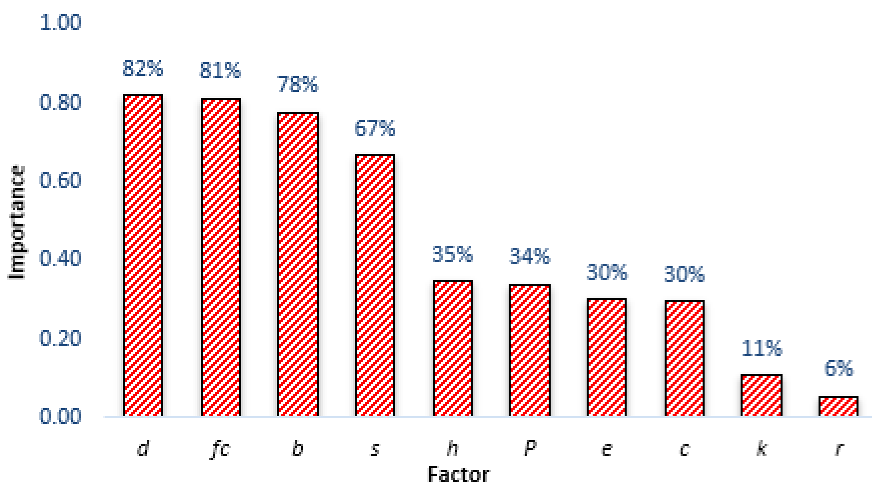


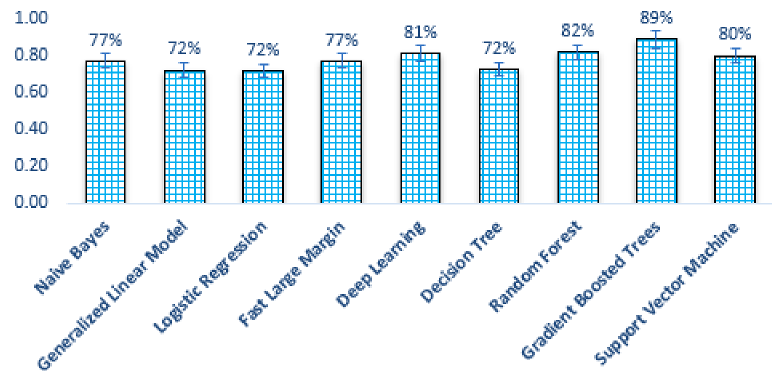
Fig. 4 Overall importance of selected parameters to fire-induced spalling phenomenon

4.2 Prediction of fire-induced spalling

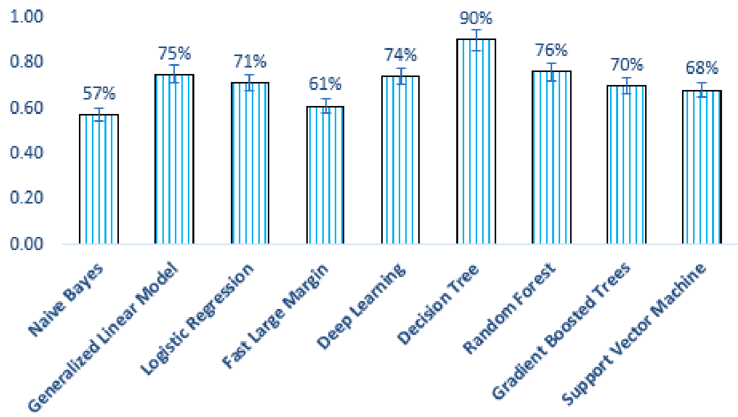
The DS+ML approach can also be used to develop tools that can instantaneously predict the occurrence of spalling in a given RC column using the ML algorithms discussed in Sec. 2.0. The development of such tools can be attractive in evaluating the tendency of a RC column to spall given the set of input parameters employed and shown in Table 1 or resulting from analysis using feature engineering. Overall, the selected algorithms achieved reasonable (72%) to high (89%) accuracy in predicting spalling phenomenon. A look into Fig. 5a shows that the gradient boosted trees (GBT) algorithm achieved the highest accuracy in predicting spalling of RC columns, followed by deep learning (DL) and support vector machine (SVM) algorithms.

An interesting exercise is to assume that there is a RC column with features equal to that of average values of all observations (listed at the bottom of Table 1), and then predict spalling tendency of this column using each of the applied nine algorithms. These predictions are listed in Fig. 5b and show that it is very likely that this RC column is going to spall under fire conditions (with 71.3%

tendency to spall). In case fire designers/engineers employ similar tools, such a tool can be of tremendous help not only by providing a mean to evaluate propensity of a RC column to spall but also by allowing designers to identify cost-effective and efficient solutions to mitigate such spalling. For example, the GBT algorithm expects that this column would not spall if tie diameter is increased to 10 mm (from 8 mm), or if compressive strength is reduced to 26 MPa (from 45.1 MPa) etc. given that all other parameters stay the same. In this case, any of these solutions can mitigate spalling and the chief engineer would have the flexibility to decide given specific aspects in his/her project (i.e. cost/availability/constructability of each solution). It should be noted that access to the developed models and database will be freely available at a permanent and dedicated webpage.



(a) Accuracy of ML algorithms in predicting fire-induced spalling phenomenon



(b) Spalling tendency for a RC column with features resembling average values from database

Fig. 5 Prediction outcome and capabilities of DS+ML

4.3 A Note on DS+ML – from a fire engineering perspective

It goes without saying that the outcome of a DS+ML analysis remains highly dependent on the available and quality of collected datapoints. Given that the database developed herein collects observations from 185 fire tests on RC columns, the outcome of this analysis is expected to properly represent fire-induced spalling in RC columns of various characteristics and configurations. Still, the size of this database is much smaller than those commonly used in other fields (i.e. medical etc.). While we may not be able to develop such massive databases of 1000's of observations, due to the complex and restricted nature of fire testing, the developed database can still be improved both vertically (i.e. by adding more observations from fire tests) as well as horizontally (by adding additional input parameters such as aggregate/fiber type etc.). A collaboration between various research groups and industry partners is foreseen as a mean to improve such database and hence is encouraged.

While this work examined 10 independent parameters, other parameters such as mix proportions, moisture content, porosity, use of admixtures etc. which were also identified to have an influence on spalling phenomenon (Hertz 2003; Kodur 2018; Kodur et al. 2003; Maluk et al. 2017; Rickard et al. 2018), were not examined herein as: 1) information on these parameters were not available/reported, and 2) it is unlikely that RC columns tested by Thomas and Webster (Thomas and Webster 1953), as well as Davey and Ashton (Davey and Ashton 1953) which were carried out in early 1950s incorporated any modern additives or fibers nor measured porosity of concrete. Furthermore, it is worth noting that the bulk of the tested columns were of square shape, had similar length, grade of reinforcement, and tie configuration (90°) etc. and hence these parameters were not also examined as their influence is expected to be normalized across all specimens. It is worth noting that the phenomenon of spalling was also examined in companion studies (Naser 2019b) that mainly considered genetic programing (GP) as the main a tool for analysis. These studies still did not examine importance of input parameters nor application of other ML algorithms and were mainly interested in developing predictive expressions that can predict occurrence of spalling through GP-derived expressions. These expressions were derived through analysis of about 100 RC columns and incorporated other parameters that were not considered herein (i.e. humidity, aggregate type etc.) as information on these inputs was not provided for the additional 85 of the RC columns utilized herein.

A final note is directed towards the fact that DS+ML analysis can be undertaken using commercially available workstations and software and hence does not require sophisticated/special processing units. In fact, this analysis can be carried out in a matter of seconds/minutes – depending on the algorithm selected for analysis and specifications of computing workstation. This allows a much faster and efficient prediction of spalling phenomenon via surrogate modeling. For example, the analysis shown herein was carried out on an Intel Core i7-9700K @ 3.6 GHz powered machine and took about 28 min from start to completion (including database organization and processing time for all nine algorithms). In this analysis, the fastest solution was obtained by Naive Bayes (8.7 seconds), and the longest was by Gradient Boosted Trees (466.9 seconds). It should be noted that while analysis speed is often regarded as a metric for evaluating performance of ML algorithms in the field of computer science, this metric is perhaps of limited relevance to this study. This metric could be of importance when applying DS+ML in future works to optimize fire design and/or predict response of large scale structural systems/buildings.

With continuous improvement in software and hardware engineering, traditional assessment methods are expected to improve. However, this improvement may still fall short of reaching that obtained by DS+ML (given the versatility/accuracy/simplicity of the presented approach) or the notion that correlation always insinuates causation (which may or may not be true in all scenarios). At this point in time, it is expected that both assessment methods can be used in parallel and in conjunction. In all cases, readers of this work are advised to steer away from overfitting DS+ML models or pursue DS+ML analysis through "black box" software.

5.0 Conclusions

This paper presents insights into the application of data science (DS) and machine learning (ML) algorithms to identify critical parameters governing fire-induced spalling of RC columns as well as developing assessment tools to predict this phenomenon. This study applied nine algorithms namely; Naive Bayes, generalized linear model, logistic regression, fast large margin, deep learning, decision tree, random forest, gradient boosted trees, and support vector machine, to analyze data from 185 fire tests carried out over the last 65 years on full scale reinforced concrete (RC) columns. The results of this comprehensive analysis show the potential of utilizing modern computing techniques in analyzing structural fire engineering phenomena given their high accuracy, ease of applications, and potential for continuous improvement. The following conclusions could also be drawn from the results of this study:

- Diameter of ties, compressive strength of concrete, geometric features, and spacing of ties are the main governing factors of fire-induced spalling in RC columns.
- Gradient boosted trees (GBT) algorithm achieved the highest accuracy (of 89%) in predicting spalling of RC columns, followed by deep learning (DL) and support vector machine (SVM) algorithms of 81% and 80%, respectively. Thus, optimizing these algorithms may lead to better examination of fire-induced spalling.
- A number of challenges continue to limit the integration of DS+ML in the field of fire engineering and safety, such as scarcity of fire tests etc. Future works are encouraged to develop approaches and techniques (i.e. big data/small data analysis, advanced transformation of inputs etc.) to overcome such challenges.

Data Availability

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request. (List items.)

Acknowledgment

The author would like to thank the support and constructive comments of the Editor and Reviewers.

Conflict of Interest

The author declares no conflict of interest.

References

ACI216.1. (2014). *Code requirements for determining fire resistance of concrete and masonry construction assemblies*. American Concrete Institue. Farmington Hill, MI.

Ali, F., Nadjai, A., Silcock, G., and Abu-Tair, A. (2004). "Outcomes of a major research on fire resistance of concrete columns." *Fire Safety Journal*, 39(6), 433–445.

Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge University Press, London. ISBN: 0521518148

Bazant, Z. (1997). "Analysis of Pore Pressure, Thermal Stress and Fracture in Rapidly Heated Concrete." *NIST special publication no. 919*.

Behnood, A., and Golafshani, E. M. (2018). "Predicting the compressive strength of silica fume concrete using hybrid artificial neural network with multi-objective grey wolves." *Journal of Cleaner Production*, Elsevier, 202, 54–64.

Bishop, C. (2006). *Pattern recognition and machine learning*. Springer. New York, NY.

Brownlee, J. (2019). "Start Here With Machine Learning." <https://machinelearningmastery.com/start-here/> (Apr. 11, 2019).

BSI, and European Committee for Standardization. (2004). *Design of concrete structures - Part 1-2: General rules - Structural fire design*. *Eurocode 2*.

Buch, S. H., and Sharma, U. K. (2019). "Fire Resistance of Eccentrically Loaded Reinforced Concrete Columns." *Fire Technology*, Springer US, 1–36.

Cheng, C.-C., Sha, F., and Saul, L. K. (2009). *A Fast Online Algorithm for Large Margin Training of Continuous Density Hidden Markov Models*. In Tenth Annual Conference of the International Speech Communication Association. Brighton, United Kingdom.

Chou, J.-S. S., Tsai, C.-F. F., Pham, A.-D. D., and Lu, Y.-H. H. (2014). "Machine learning in concrete strength simulations: Multi-nation data analytics." *Construction and Building Materials*, Elsevier, 73, 771–780.

Davey, N., and Ashton, L. (1953). *Investigations on Building Fires: Part V.: Fire Tests on Structural Elements*.

Dhar, V., and Vasant. (2013). "Data science and prediction." *Communications of the ACM*, ACM, 56(12), 64–73.

Dotreppe, J.-C., Franssen, J.-M., Bruls, A., Baus, R., Vandevelde, P., Minne, R., van Nieuwenburg, D., and Lambotte, H. (1997). "Experimental research on the determination of the main parameters affecting the behaviour of reinforced concrete columns under fire conditions." *Magazine of Concrete Research*.

Dwaikat, M. B., and Kodur, V. K. R. (2009). "Hydrothermal model for predicting fire-induced spalling in concrete structural systems." *Fire Safety Journal*, Elsevier, 44(3), 425–434.

Gawin, D., Pesavento, F., and Castells, A. G. (2018). "On reliable predicting risk and nature of thermal spalling in heated concrete." *Archives of Civil and Mechanical Engineering*, Elsevier, 18(4), 1219–1227.

Hass, R. (1986). *Practical rules for the design of reinforced concrete and composite columns submitted to fire*. *Technical Report (in German)*, T.U. Braunschweig.

Hertz, K. D. D. (2003). "Limits of spalling of fire-exposed concrete." *Fire Safety Journal*, Elsevier, 38(2), 103–116.

Hou, Q., Liu, L., Zhen, L., and Jing, L. (2018). "A novel projection nonparallel support vector machine for pattern classification." *Engineering Applications of Artificial Intelligence*, Pergamon, 75, 64–75.

Intel. (2019). "Gradient boosted tree (GBT)." <https://software.intel.com/en-us/daal-programming-guide-details-24> (Apr. 9, 2019).

Jansson, R. (2008). *Material properties related to fire spalling of concrete*. Division of Building Materials, Lund Institute of Technology, Lund University.

Jordan, M. I., and Mitchell, T. M. (2015). "Machine learning: Trends, perspectives, and prospects." *Science (New York, N.Y.)*, American Association for the Advancement of Science, 349(6245), 255–60.

Kalifa, P., Chéné, G., and Gallé, C. (2001). "High-temperature behaviour of HPC with polypropylene fibres - From spalling to microstructure." *Cement and Concrete Research*, Pergamon, 31(10), 1487–1499.

Khoury, G. A. (2000). "Effect of fire on concrete and concrete structures." *Progress in Structural Engineering and Materials*, 2(4), 429–447.

Kodur, V., Cheng, F., Wang, T., Latour, J., and Leroux, P. (2001). "Fire resistance of high-performance concrete columns." *National Research Council Canada*, 136.

Kodur, V. K. R. (2018). "Innovative strategies for enhancing fire performance of high-strength concrete structures." *Advances in Structural Engineering*, 21(11), 1723–1732.

Kodur, V. K. R., Cheng, F.-P., Wang, T.-C., and Sultan, M. A. (2003). "Effect of Strength and Fiber Reinforcement on Fire Resistance of High-Strength Concrete Columns." *Journal of Structural Engineering*, 129, 253–9.

Kodur, V. K. R., and Dwaikat, M. (2008). "A numerical model for predicting the fire resistance of reinforced concrete beams." *Cement and Concrete Composites*.

Kodur, V., Khaliq, W., and Raut, N. (2013). "An approach to account for tie configuration in predicting fire resistance of reinforced concrete columns." *Engineering Structures*, 56, 1976-1985.

Kodur, V., and McGrath, R. (2003). "Fire endurance of high strength concrete columns." *Fire Technology*, 39, 73–87.

Kodur, V., McGrath, R., Leroux, P., and Latour, J. (2005). *Experimental studies for evaluating the fire endurance of high-strength concrete columns*. National Research Council Canada, Internal Report, (197).

Kodur, V., and Naser, M. Z. (2020). *Structural fire engineering*. McGraw Hill. New York, ISBN: 1260128598.

Liaw, A., and Wiener, M. (2002). *Classification and Regression by RandomForest*. R News, ISSN 1609-3631

Lie, T., and Woollerton, J. (1988). *Fire Resistance of Reinforced Concrete Columns - NRC Publications Archive - National Research Council Canada*.

Liu, J.-C. C., Tan, K. H., and Yao, Y. (2018). "A new perspective on nature of fire-induced spalling in concrete." *Construction and Building Materials*, Elsevier, 184, 581–590.

Maluk, C., Bisby, L., and Terrasi, G. P. (2017). "Effects of polypropylene fibre type and dose on the propensity for heat-induced concrete spalling." *Engineering Structures*. 141, 584-595.

Matlab. (2019). "Feature Detection and Extraction - MATLAB & Simulink." <https://www.mathworks.com/help/vision/feature-detection-and-extraction.html> (Apr. 10, 2019).

McKinney, J., and Ali, F. (2014). "Artificial Neural Networks for the Spalling Classification & Failure Prediction Times of High Strength Concrete Columns." *Journal of Structural Fire Engineering*, 5(3), 203–214.

Meacham, B., Park, H., van Straalen, Ij., Maljaars, J., and Netherlands Kees van Weeren, T. (2009). *Fire and Collapse, Faculty of Architecture Building, Delft University of Technology: Data Collection and Preliminary Analyses*. In 8th International Conference on

Performance-Based Codes and Fire Safety Design Methods, Lund University, Sweden.

Missemer, L., Ouedraogo, E., Malecot, Y., Clergue, C., and Rogat, D. (2019). “Fire spalling of ultra-high performance concrete: From a global analysis to microstructure investigations.” *Cement and Concrete Research*, Pergamon, 115, 207–219.

Myllymaki, J., and Lie, T. (1991). “Fire Resistance Test of a Square Reinforced Concrete Column.” *National Research Council Canada*, Report No. 36.

Naser, M. Z. (2019a). “Autonomous and resilient infrastructure with cognitive and self-deployable load-bearing structural components.” *Automation in Construction*, Elsevier, 99, 59–67.

Naser, M. Z. (2019b). “AI-based cognitive framework for evaluating response of concrete structures in extreme conditions.” *Engineering Applications of Artificial Intelligence*, 81, 437–449.

Naser, M. Z. (2019c). “Heuristic machine cognition to predict fire-induced spalling and fire resistance of concrete structures.” *Automation in Construction*, Elsevier, 106, 102916.

Naser, M. Z. (2020). “Autonomous Fire Resistance Evaluation.” *ASCE Jounral of Structural Engineering*. 146(6).

Naser, M. Z., and Seitllari, A. (2019). “Concrete under fire: an assessment through intelligent pattern recognition.” *Engineering with Computers*, Springer London, 1–14.

Peng, G. (2000). *Evaluation of fire damage to high performance concrete*. The Hong Kong Polytechnic University

Phan, L., and Carino, N. (2002). “Effects of test conditions and mixture proportions on behavior of high-strength concrete exposed to high temperatures.” *ACI Materials*, 99(1), 54-66

Phan, L. T. (2008). “Pore pressure and explosive spalling in concrete.” *Materials and Structures/Materiaux et Constructions*, 41(10), 1623–1632.

Rickard, I., Bisby, L., and Deeny, S. (2018). “Explosive spalling of concrete in fire: novel testing to mitigate design risk.” *The Structural Engineer*. 96(1), 42-47.

RILEM. (1994). “Draft recommendation for damage classification of concrete structures.” *Materials and Structures*, Kluwer Academic Publishers, 27(6), 362–369.

Rodrigues, J. P. C., Laím, L., and Correia, A. M. (2010). “Behaviour of fiber reinforced concrete columns in fire.” *Composite Structures*, 92, 1263–8.

Safavian, S. R., and Landgrebe, D. (1991). “A survey of decision tree classifier methodology.” *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3), 660–674.

Sanjayan, G., and Stocks, L. (1993). “Spalling of high-strength silica fume concrete in fire.” *ACI Materials Journal*. 90(2), pp.170-173.

Sayad, Y. O., Mousannif, H., and Al Moatassime, H. (2019). "Predictive modeling of wildfires: A new dataset and machine learning approach." *Fire Safety Journal*. 104, 130-146.

Schmidhuber, J. (2015). "Deep learning in neural networks: An overview." *Neural Networks*, Springer, 61, 85–117.

Schneider, U. (1988). "Concrete at high temperatures - A general review." *Fire Safety Journal*, 13(1), 55–68.

Seitlllari, A., and Naser, M. Z. (2019). "Leveraging artificial intelligence to assess explosive spalling in fire-exposed RC columns." *Computers and Concrete*. 24(3), 271-282.

Shah, A. H., and Sharma, U. K. (2017). "Fire resistance and spalling performance of confined concrete columns." *Construction and Building Materials*, 156, 161–74.

Shahin, M., Jaksa, M., Systems, H. M.-A. in A. N., and 2009, U. (2009). "Recent advances and future challenges for artificial neural systems in geotechnical engineering applications." *Advances in Artificial Neural Systems*, 5.

Shiri Harzevili, N., and Alizadeh, S. H. (2018). "Mixture of latent multinomial naive Bayes classifier." *Applied Soft Computing*, 69, 516–527.

Thomas, F., and Webster, C. (1953). *Investigations on Building Fires: Part VI.: the Fire Resistance of Reinforced Concrete Columns*.

Young, B. A., Hall, A., Pilon, L., Gupta, P., and Sant, G. (2019). "Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods." *Cement and Concrete Research*, Pergamon, 115, 379–388.

Zhao, J., Zheng, J., Peng, G., and van Breugel, K. (2014). "A meso-level investigation into the explosive spalling mechanism of high-performance concrete under fire exposure." *Cement and Concrete Research*, Pergamon, 65, 64–75.

This is a preprint draft. The published article can be found at: https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525

Please cite this paper as:

Naser M.Z. (2021). "Observational Analysis of Fire-induced Spalling through Data Science and Machine Learning." ASCE Journal of Materials in Civil Engineering. https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525.

List of Tables:

Table 1 Compiled database used for DS+ML analysis

Study	Sp. Ref.	***b*** (mm)	***h*** (mm)	***r*** (%)	f_c (MPa)	***k***	***C*** (mm)	***e*** (mm)	***P*** (kN)	***s*** (mm)	***d*** (mm)
(Lie and Woollerton 1988)	*1a*	305	305	2.19	34	FF	38	0	0	305	10
(Lie and Woollerton 1988)	*2a*	305	305	2.19	37	FF	38	0	1333	305	10
(Lie and Woollerton 1988)	*3a*	305	305	2.19	34	FF	38	0	800	305	10
(Lie and Woollerton 1988)	*4a*	305	305	2.19	35	FF	38	0	711	305	10
(Lie and Woollerton 1988)	*5g*	406	406	2.47	41	FF	38	0	0	406	10
(Lie and Woollerton 1988)	*6g*	203	203	2.75	42	FF	38	0	169	203	10
(Lie and Woollerton 1988)	*7a*	305	305	2.19	36	FF	38	0	1067	305	10
(Lie and Woollerton 1988)	*8a*	305	305	2.19	35	FF	38	0	1778	305	10
(Lie and Woollerton 1988)	*9a*	305	305	2.19	38	FF	38	0	1333	305	10
(Lie and Woollerton 1988)	*10b*	305	305	2.19	41	FF	38	0	800	305	10
(Lie and Woollerton 1988)	*11b*	305	305	2.19	37	FF	38	0	1067	305	10
(Lie and Woollerton 1988)	*12b*	305	305	2.19	40	FF	38	0	1778	305	10
(Lie and Woollerton 1988)	*1e*	305	305	2.19	42	PP	38	0	342	305	10
(Lie and Woollerton 1988)	*2e*	305	305	2.19	44	FF	38	0	1044	305	10
(Lie and Woollerton 1988)	*3e*	305	305	2.19	35	FF	38	0	916	305	10
(Lie and Woollerton 1988)	*4d*	305	305	2.19	53	FF	38	0	1178	305	10
(Lie and Woollerton 1988)	*5d*	305	305	2.19	50	FF	38	0	1067	305	10
(Lie and Woollerton 1988)	*6c*	305	305	2.19	47	FF	38	0	1076	305	10
(Lie and Woollerton 1988)	*7c*	305	305	2.19	43	FF	38	0	947	305	10
(Lie and Woollerton 1988)	*8f*	305	305	4.38	43	FF	38	0	978	305	10
(Lie and Woollerton 1988)	*9f*	305	305	4.38	37	FF	38	0	1333	305	10
(Lie and Woollerton 1988)	*10g*	406	406	2.47	39	FF	38	0	2418	406	10
(Lie and Woollerton 1988)	*11g*	406	406	3.97	38	FF	38	0	2795	406	10

(Lie and Woollerton 1988)	*12g*	406	406	3.97	46	FF	64	0	2978	406	10
(Lie and Woollerton 1988)	*1i*	305	305	2.19	40	PP	38	0	800	305	10
(Lie and Woollerton 1988)	*2i*	305	305	2.19	39	PP	38	0	1000	305	10
(Lie and Woollerton 1988)	*3k*	305	305	2.19	40	FF	38	25	1000	305	10
(Lie and Woollerton 1988)	*4j*	305	305	2.19	38	FF	38	0	1067	305	10
(Lie and Woollerton 1988)	*5h*	305	457	2.22	43	FF	38	0	1413	305	10
(Lie and Woollerton 1988)	*6h*	203	914	1.22	42	FF	38	0	756	203	10
(Lie and Woollerton 1988)	*14k*	305	305	2.19	38	FF	38	25	1178	305	10
(Kodur et al. 2001)	*HSC 1*	406	406	2.42	92	FF	38	0	0	406	8
(Kodur et al. 2001)	*HSC 2*	406	406	2.42	127	PF	38	0	2913	406	8
(Kodur et al. 2001)	*HSC 3*	406	406	2.42	100	FF	38	0	3080	406	8
(Kodur et al. 2001)	*HSC 4*	406	406	2.42	90	PF	38	0	2934	406	8
(Kodur et al. 2001)	*HSC 5*	406	406	2.42	86	FF	38	0	2406	406	8
(Kodur et al. 2001)	*HSC 6*	406	406	2.42	96	FF	38	0	4919	406	8
(Kodur et al. 2001)	*HSC 7*	305	305	1.72	120	FF	41	0	1979	152	6
(Kodur et al. 2001)	*HSC 8*	305	305	1.72	120	FF	41	0	2363	76	6
(Kodur et al. 2001)	*HSC 9*	305	305	1.72	120	FF	41	0	2954	76	6
(Kodur et al. 2001)	*HSC 10*	305	305	2.42	120	PF	41	25	2954	76	6
(Kodur et al. 2005)	*TNC 1*	305	305	2.18	40	FF	40	0	930	145	10
(Kodur et al. 2005)	*TNC 2*	305	305	2.18	40	FF	40	0	1500	145	10
(Kodur et al. 2005)	*TNC 3*	305	305	2.18	40	PP	40	25	1000	145	10
(Kodur et al. 2005)	*THC 4*	305	305	2.18	100	FF	40	0	2000	145	10
(Kodur et al. 2005)	*THC 5*	305	305	2.18	100	FF	40	0	2000	145	10
(Kodur et al. 2005)	*THC 6*	305	305	2.18	100	FF	40	0	3000	145	10
(Kodur et al. 2005)	*THC 7*	305	305	2.18	73	FF	40	0	1300	145	10

This is a preprint draft. The published article can be found at: https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525

Please cite this paper as:

Naser M.Z. (2021). "Observational Analysis of Fire-induced Spalling through Data Science and Machine Learning." ASCE Journal of Materials in Civil Engineering. https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525.

(Kodur et al. 2005)	*THC 8*	305	305	2.18	73	FF	40	0	2000	145	10
(Kodur et al. 2005)	*THC 9*	305	305	2.18	73	PF	40	25	1200	145	10
(Kodur et al. 2005)	*THS 10*	305	305	2.18	73	FF	40	0	1800	145	10
(Kodur et al. 2005)	*THS 11*	305	305	2.18	73	FF	40	0	2200	145	10
(Kodur et al. 2005)	*THS 12*	305	305	2.18	73	PF	40	25	1500	145	10
(Kodur et al. 2005)	*THP 13*	305	305	2.18	69	FF	40	0	1800	145	10
(Kodur et al. 2005)	*THP 14*	305	305	2.18	69	FF	40	0	2200	145	10
(Kodur et al. 2005)	*THP 15*	305	305	2.18	69	PF	40	25	1500	145	10
(Shah and Sharma 2017)	*M3S 50*	300	300	1.78	34	FF	40	0	1170	200	10
(Shah and Sharma 2017)	*M3S 75*	300	300	1.78	34	FF	40	0	1170	150	10
(Shah and Sharma 2017)	*M3S 100*	300	300	1.78	34	FF	40	0	1170	75	10
(Shah and Sharma 2017)	*M3S 150*	300	300	1.78	34	FF	40	0	1170	150	10
(Shah and Sharma 2017)	*M3S T150*	300	300	1.78	34	FF	40	0	1170	100	10
(Shah and Sharma 2017)	*M3S 200*	300	300	1.78	34	FF	40	0	1170	50	10
(Shah and Sharma 2017)	*M6S 150*	300	300	1.78	63	FF	40	0	1858	150	10
(Shah and Sharma 2017)	*M6S T150*	300	300	1.78	63	FF	40	0	1858	150	10
(Kodur et al. 2005)	*HS2-1*	406	406	2.47	85	FF	40	0	3895	203	10
(Kodur et al. 2005)	*HS2-2*	406	406	2.47	85	FF	40	0	4328	305	10
(Kodur et al. 2005)	*HS2-3*	406	406	2.47	85	FF	40	0	4328	406	10
(Kodur et al. 2005)	*HS2-4*	406	406	2.47	114	FF	40	0	4567	203	10
(Kodur et al. 2005)	*HS2-5*	406	406	2.47	114	FF	40	0	5373	305	10
(Kodur et al. 2005)	*HS2-6*	406	406	2.47	114	FF	40	0	3546	406	10
(Kodur et al. 2005)	*HS2-7*	406	406	2.47	138	PF	40	27	4233	203	10
(Kodur et al. 2005)	*HS2-8*	406	406	2.47	138	PF	40	27	4981	305	10
(Kodur et al. 2005)	*HS2-9*	406	406	2.47	138	PF	40	27	4981	305	10

(Kodur et al. 2005)	*HS2-10*	406	406	2.47	138	PF	40	27	4981	406	10
(Buch and Sharma 2019)	*NSC 0*	300	300	2.28	28	PP	40	0	544	300	8
(Buch and Sharma 2019)	*NSC 1*	300	300	2.18	28	PP	40	20	532	300	8
(Buch and Sharma 2019)	*NSC 2*	300	300	2.28	32	PP	40	20	579	300	8
(Buch and Sharma 2019)	*NSC 3*	300	300	2.28	31	PP	40	40	567	300	8
(Buch and Sharma 2019)	*NSC 4*	300	300	2.28	27	PP	40	20	544	150	8
(Buch and Sharma 2019)	*NSC 5*	300	300	2.28	31	PP	40	40	567	150	8
(Buch and Sharma 2019)	*HSC 0*	300	300	2.28	69	PP	40	0	1008	300	8
(Buch and Sharma 2019)	*HSC 1*	300	300	2.18	58	PP	40	20	892	300	8
(Buch and Sharma 2019)	*HSC 2*	300	300	2.28	69	PP	40	20	973	300	8
(Buch and Sharma 2019)	*HSC 3*	300	300	2.28	67	PP	40	20	996	150	8
(Buch and Sharma 2019)	*HSC 4*	300	300	2.28	60	PP	40	40	892	150	8
(Rodrigues et al. 2010)	*C1*	250	250	3.14	24	PP	30	0	686	187	8
(Rodrigues et al. 2010)	*C2*	250	250	3.14	27	PP	30	0	686	187	8
(Rodrigues et al. 2010)	*C3*	250	250	3.14	25	PP	30	0	686	187	8
(Rodrigues et al. 2010)	*C4*	250	250	3.14	29	PP	30	0	686	187	8
(Myllymaki and Lie 1991)	*C*	300	300	0.89	38	PP	30	0	1400	240	6
(Davey and Ashton 1953)	*C27*	152	152	2.19	29	FF	25	0	209	152	8
(Davey and Ashton 1953)	*C41*	152	152	6.71	28	FF	25	0	346	152	8
(Davey and Ashton 1953)	*C11*	254	254	0.79	28	FF	25	0	463	152	8
(Davey and Ashton 1953)	*C13*	254	254	0.79	15	FF	25	0	448	152	8
(Davey and Ashton 1953)	*C15*	254	254	0.79	17	FF	25	0	508	152	8
(Davey and Ashton 1953)	*C21*	254	254	3.97	28	FF	29	0	725	152	8
(Davey and Ashton 1953)	*C23*	254	254	0.79	38	FF	25	0	623	152	8
(Davey and Ashton 1953)	*C24*	254	254	0.79	36	FF	25	0	657	152	8

This is a preprint draft. The published article can be found at: https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525

Please cite this paper as:

Naser M.Z. (2021). "Observational Analysis of Fire-induced Spalling through Data Science and Machine Learning." ASCE Journal of Materials in Civil Engineering. https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525.

(Davey and Ashton 1953)	*C30*	254	254	0.79	39	FF	25	0	465	152	8
(Davey and Ashton 1953)	*C31*	254	254	0.79	36	FF	25	0	463	152	8
(Davey and Ashton 1953)	*C32*	254	254	0.79	37	FF	25	0	463	152	8
(Davey and Ashton 1953)	*C35*	254	254	0.79	16	FF	25	0	465	152	8
(Davey and Ashton 1953)	*C46*	254	254	4.91	29	FF	32	0	918	152	9
(Davey and Ashton 1953)	*C20*	279	279	1.02	26	FF	38	0	711	114	8
(Davey and Ashton 1953)	*C33*	279	279	1.02	33	FF	38	0	586	178	9
(Davey and Ashton 1953)	*C34*	279	279	1.02	34	FF	38	0	858	178	9
(Davey and Ashton 1953)	*C36*	279	279	1.02	22	FF	38	0	586	178	9
(Davey and Ashton 1953)	*C37*	279	279	1.02	26	FF	38	0	586	178	9
(Davey and Ashton 1953)	*C38*	279	279	1.02	31	FF	38	0	0	178	9
(Davey and Ashton 1953)	*C39*	279	279	1.02	27	FF	38	0	586	178	9
(Davey and Ashton 1953)	*C42*	279	279	1.02	29	FF	38	0	711	178	9
(Davey and Ashton 1953)	*C82*	279	279	4.07	36	FF	38	0	909	178	9
(Davey and Ashton 1953)	*C86*	279	279	4.07	38	FF	38	0	909	191	9
(Davey and Ashton 1953)	*C87*	279	279	4.07	36	FF	38	0	911	191	9
(Davey and Ashton 1953)	*E25/ S3*	279	279	4.07	25	FF	38	0	906	191	9
(Davey and Ashton 1953)	*C89*	279	279	4.07	29	FF	38	0	608	191	9
(Davey and Ashton 1953)	*C90*	279	279	4.07	30	FF	38	0	608	191	9
(Davey and Ashton 1953)	*C88*	279	279	4.07	29	FF	38	0	608	191	9
(Davey and Ashton 1953)	*C12*	305	305	0.85	18	FF	51	0	779	114	8
(Davey and Ashton 1953)	*E16/ S8*	356	356	0.90	30	FF	38	0	857	229	6
(Davey and Ashton 1953)	*C28*	406	406	1.88	27	FF	25	0	1230	152	9
(Davey and Ashton 1953)	*C44*	406	406	4.65	23	FF	35	0	2092	152	8
(Davey and Ashton 1953)	*C47*	406	406	4.65	34	FF	35	0	2361	152	8

This is a preprint draft. The published article can be found at: https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525

Please cite this paper as:

Naser M.Z. (2021). "Observational Analysis of Fire-induced Spalling through Data Science and Machine Learning." ASCE Journal of Materials in Civil Engineering. https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525.

(Davey and Ashton 1953)	*C45*	483	483	1.33	27	FF	25	0	747	152	9
(Davey and Ashton 1953)	*C48*	508	508	2.40	30	FF	44	0	3257	152	9
(Davey and Ashton 1953)	*C29*	514	514	0.97	29	FF	44	0	1982	127	9
(Davey and Ashton 1953)	*C49*	305	305	1.09	24	FF	13	0	784	38	8
(Davey and Ashton 1953)	*C53*	305	305	1.09	30	FF	13	0	876	38	8
(Davey and Ashton 1953)	*C50*	356	356	0.80	34	FF	13	0	896	51	8
(Davey and Ashton 1953)	*C51*	406	406	0.96	38	FF	13	0	1235	44	8
(Davey and Ashton 1953)	*C54*	406	406	0.96	44	FF	13	0	1394	44	8
(Davey and Ashton 1953)	*C52*	508	508	0.61	35	FF	13	0	1833	44	8
(Davey and Ashton 1953)	*C55*	508	508	0.88	44	FF	13	0	2102	44	8
(Thomas and Webster 1953)	*1*	305	305	1.67	22	FF	22	0	747	178	10
(Thomas and Webster 1953)	*2*	280	280	1.01	28	FF	25	0	747	178	10
(Thomas and Webster 1953)	*3*	254	254	1.77	32	FF	25	0	747	178	8
(Thomas and Webster 1953)	*4*	203	203	6.22	46	FF	25	0	747	178	8
(Thomas and Webster 1953)	*6*	381	381	1.40	25	FF	29	0	997	152	9
(Thomas and Webster 1953)	*7*	356	356	1.22	29	FF	25	0	1495	152	9
(Thomas and Webster 1953)	*8*	305	305	3.33	44	FF	25	0	1495	178	10
(Thomas and Webster 1953)	*10*	483	483	1.64	27	FF	25	0	747	152	8
(Thomas and Webster 1953)	*12*	356	356	3.20	43	FF	35	0	2243	152	9
(Thomas and Webster 1953)	*15*	483	483	2.20	29	FF	29	0	2990	152	8
(Thomas and Webster 1953)	*16*	406	406	1.56	35	FF	29	0	2990	178	8
(Thomas and Webster 1953)	*17*	229	229	6.04	26	FF	29	0	747	152	8
(Thomas and Webster 1953)	*18*	279	279	11.7	24	FF	32	0	1495	229	8
(Thomas and Webster 1953)	*19*	356	356	9.79	28	FF	38	0	2243	152	8
(Thomas and Webster 1953)	*20*	406	406	1.56	23	FF	44	0	2990	229	9

This is a preprint draft. The published article can be found at: https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525

Please cite this paper as:

Naser M.Z. (2021). "Observational Analysis of Fire-induced Spalling through Data Science and Machine Learning." ASCE Journal of Materials in Civil Engineering. https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525.

(Thomas and Webster 1953)	*21*	381	381	1.77	25	FF	29	0	1495	152	9
(Thomas and Webster 1953)	*22*	381	381	1.77	33	FF	29	0	1495	152	9
(Thomas and Webster 1953)	*23*	381	381	1.77	41	FF	29	0	1495	152	9
(Thomas and Webster 1953)	*26*	381	381	1.77	23	FF	29	0	747	152	9
(Thomas and Webster 1953)	*27*	381	381	1.77	29	FF	29	0	498	152	9
(Thomas and Webster 1953)	*28*	381	381	1.77	25	FF	29	0	1495	152	9
(Thomas and Webster 1953)	*A1*	381	381	1.77	26	FF	29	0	947	152	9
(Thomas and Webster 1953)	*A2*	381	381	1.77	27	FF	29	0	1495	152	9
(Thomas and Webster 1953)	*A3*	381	381	1.77	28	FF	29	0	1495	152	9
(Thomas and Webster 1953)	*A4*	381	381	1.77	28	FF	29	0	299	152	9
(Thomas and Webster 1953)	*A5*	381	381	1.77	24	FF	29	0	1495	152	9
(Thomas and Webster 1953)	*A6*	381	381	1.07	27	FF	29	0	997	152	9
(Thomas and Webster 1953)	*A7*	305	305	0.85	17	FF	25	0	249	178	9
(Thomas and Webster 1953)	*A9*	254	254	1.23	31	FF	25	0	249	178	9
(Thomas and Webster 1953)	*A10*	254	254	2.40	35	FF	25	0	249	178	9
(Thomas and Webster 1953)	*A11*	483	483	0.66	18	FF	25	0	2243	152	8
(Thomas and Webster 1953)	*A12*	483	483	1.10	21	FF	25	0	1794	152	8
(Thomas and Webster 1953)	*A13*	483	483	1.10	27	FF	29	0	997	152	8
(Thomas and Webster 1953)	*A14*	483	483	0.66	22	FF	29	0	2193	152	8
(Thomas and Webster 1953)	*A18*	406	406	0.31	19	FF	25	0	1495	152	9
(Thomas and Webster 1953)	*C23*	254	254	0.79	38	FF	25	0	623	152	8
(Thomas and Webster 1953)	*C24*	254	254	2.40	36	FF	25	0	658	152	8
(Thomas and Webster 1953)	*C28*	406	406	0.31	27	FF	25	0	1231	152	9
(Thomas and Webster 1953)	*C30*	254	254	0.79	39	FF	25	0	465	152	8
(Thomas and Webster 1953)	*C31*	254	254	0.79	36	FF	25	0	463	152	8

(Thomas and Webster 1953)	*C32*	254	254	1.23	37	FF	25	0	463	152	8
(Thomas and Webster 1953)	*C33*	279	279	1.02	33	FF	38	0	586	178	9
(Thomas and Webster 1953)	*C34*	279	279	0.65	34	FF	38	0	858	178	9
(Thomas and Webster 1953)	*C35*	254	254	4.91	16	FF	25	0	465	152	8
(Thomas and Webster 1953)	*C46*	254	254	4.91	29	FF	32	0	919	152	8
(Thomas and Webster 1953)	*C82*	279	279	4.07	36	FF	32	0	912	191	9
(Thomas and Webster 1953)	*C86*	279	279	4.07	38	FF	32	0	912	191	9
(Thomas and Webster 1953)	*C87*	279	279	4.07	36	FF	32	0	912	191	9
(Thomas and Webster 1953)	*C88*	279	279	4.07	29	FF	32	0	608	191	9
(Thomas and Webster 1953)	*C89*	279	279	4.07	29	FF	32	0	608	191	9
(Thomas and Webster 1953)	*C90*	279	279	4.07	30	FF	32	0	608	191	9
(Thomas and Webster 1953)	*E25*	279	279	4.07	25	FF	32	0	907	191	9
Database statistics	*Average*	*323.8*	*328.5*	*2.3*	*45.0*	-	*34.1*	*2.8*	*1373.2*	*201.2*	*8.9*
	Std. div.	*69.0*	*81.5*	*1.5*	*27.9*	-	*7.6*	*8.4*	*1080.8*	*87.9*	*0.9*
	Minimum	*152.0*	*152.0*	*0.3*	*15.0*	-	*13.0*	*0.0*	*0.0*	*38.1*	*6.0*
	Maximum	*514.0*	*914.0*	*11.7*	*138.0*	-	*64.0*	*40.0*	*5373.0*	*406.0*	*10.0*
	Skewness	*0.75*	*2.40*	*2.66*	*1.83*	-	*-0.57*	*2.90*	*1.72*	*0.80*	*-0.97*

FF: fixed-fixed, PP: pined-pined, PF: pined-fixed boundary conditions
Yes: column spalled, No: column did not spall.

Table 2 Outcome of DS+ML analysis with feature engineering

<table>
<tr><th></th><th colspan="4">Number of identified critical parameters</th><th colspan="3">Identified critical parameters</th><th colspan="3">Feature(s) generated</th></tr>
<tr><td>Naive Bayes</td><td colspan="4">4</td><td colspan="3">e, s, d, f_c</td><td colspan="3">$e–s$</td></tr>
<tr><td>Generalized Linear Model</td><td colspan="4">4</td><td colspan="3">b, f_c, s, d</td><td colspan="3">-</td></tr>
<tr><td>Logistic Regression</td><td colspan="4">3</td><td colspan="3">d, f_c, c</td><td colspan="3">-</td></tr>
<tr><td>Fast Large Margin</td><td colspan="4">5</td><td colspan="3">$s, f_c. e, d, h, b$</td><td colspan="3">$e–d$, $abs(f_c)$</td></tr>
<tr><td>Deep Learning</td><td colspan="4">5</td><td colspan="3">b, d, f_c, k, s</td><td colspan="3">-</td></tr>
<tr><td>Decision Tree</td><td colspan="4">3</td><td colspan="3">d, f_c, h</td><td colspan="3">$d \times f_c \times h$</td></tr>
<tr><td>Random Forest</td><td colspan="4">5</td><td colspan="3">d, r, f_c, b</td><td colspan="3">f_c/r, $b \times r$</td></tr>
<tr><td>Gradient Boosted Trees</td><td colspan="4">5</td><td colspan="3">f_c, c, d, s, b</td><td colspan="3">-</td></tr>
<tr><td>Support Vector Machine</td><td colspan="4">2</td><td colspan="3">f_c, d</td><td colspan="3">$exp(d)- f_c$</td></tr>
<tr><td rowspan="2">Overall reoccurring</td><td>d</td><td>f_c</td><td>b</td><td>s</td><td>c</td><td>e</td><td>h</td><td>k</td><td>r</td><td>P</td></tr>
<tr><td>9</td><td>9</td><td>5</td><td>5</td><td>2</td><td>2</td><td>2</td><td>1</td><td>1</td><td>0</td></tr>
</table>

List of Figures: