



Aplicação de Regressão Logística para Predição de *Spalling* em Pilares de Concreto Armado Expostos ao Fogo

Marcelo Mesquita do Amaral^{1*}; Viviane Leite Dias de Mattos¹
Mauro de Vasconcellos Real¹

Resumo: O *spalling* é um fenômeno no qual pedaços de concreto se desprendem da superfície de um elemento estrutural quando submetido a condições de aquecimento rápido. Este lascamento ocorre devido ao aumento da poro-pressão, à geração de tensões térmicas ou à combinação desses mecanismos. Vários fatores influenciam sua ocorrência, incluindo as propriedades do concreto, as características do elemento estrutural e as condições de exposição. Este estudo apresenta o desenvolvimento de um modelo de Regressão Logística para a predição de ocorrência de *spalling* em pilares de concreto armado expostos ao fogo. Utilizou-se um banco de dados composto por 119 observações e 10 variáveis. Após o pré-processamento e a análise exploratória de dados restaram 94 observações para a seleção dos modelos de regressão logística. A escolha de variáveis foi realizada com auxílio de uma tabela comparativa de métricas, equilibrando a complexidade do modelo com seu desempenho preditivo. O R^2 de Nagelkerke calculado foi 81,07%, indicando que o modelo obtido explica uma parte significativa da variação de ocorrência de *spalling*. As métricas de erro RMSE, MAE e MAPE foram 0,2917; 0,0851 e 0,0585, respectivamente, sugerindo uma boa qualidade do modelo; que apresentou uma acurácia de 93,62% e um F1-score de 95,08%, indicando também um equilíbrio entre precisão e *recall*. Portanto, este estudo demonstrou que a regressão logística pode ser uma ferramenta útil na predição de *spalling* de estruturas de concreto armado em situação de incêndio.

Palavras-chave: *Spalling*; Concreto Armado; Incêndio; Regressão Logística

*marcelo.amaral@furg.br

¹FURG - Universidade Federal do Rio Grande.

1 INTRODUÇÃO

É intrínseco ao ser humano exigir segurança em seu local de moradia e de trabalho (Seito *et al.*, 2008). O concreto é o material de construção mais utilizado pelo homem, principalmente em conjunto com o aço, formando o concreto armado (Tutikian *et al.*, 2011). No entanto, sob temperaturas elevadas, como no caso de incêndios, o concreto além de sofrer uma redução da resistência, perde área resistente por causa do *spalling*, o que pode comprometer a integridade estrutural da edificação, como evidenciado no incêndio da Torre Madrid Windsor na Espanha em 2005 (figura 1) (So, 2016) .



Figura 1: Colapso da estrutura após o incêndio da Torre Madrid Windsor (Espanha, 2005). (So, 2016)

O *spalling* é um fenômeno no qual pedaços de concreto se desprendem da superfície de um elemento estrutural quando submetido a condições de aquecimento rápido, reduzindo a área resistente e expondo a armadura ao fogo (Silva, 2016).

Atualmente, as disposições normativas carecem de recomendações específicas e métodos de previsão (Al-Bashiti e Naser, 2024). No Brasil, a ABNT NBR 15200 (2024), norma de Projeto de Estruturas de Concreto em Situação de Incêndio, não faz menção alguma ao lascamento do concreto. Portanto, pode-se considerar uma abordagem inovadora a sua predição utilizando regressão logística.

Este estudo visa desenvolver um modelo de regressão logística para prever a ocorrência do *spalling* em pilares de concreto armado expostos ao fogo. Além disso, identificar os principais fatores, dentre as variáveis consideradas, que influenciam a ocorrência do fenômeno.

2 ESTADO DA ARTE

Engenheiros estruturais de concreto em situação de incêndio ainda enfrentam dificuldades para prever com precisão o fenômeno do *spalling*. Até agora, algumas teorias foram propostas para explicar os efeitos do calor sobre o concreto. A maioria dessas teorias foi obtida a partir de testes experimentais e análises empíricas. No entanto, pouco consenso foi alcançado sobre os mecanismos de *spalling* induzidos pelo fogo (Al-Bashiti e Naser, 2022). Isso se deve à complexidade do fenômeno, o que representa uma oportunidade para explorar novas abordagens, como a regressão logística.

Esse fenômeno estocástico combina dois mecanismos principais: (a) o aumento da poro-pressão e (b) a geração de tensões térmicas. Diversos fatores influenciam o *spalling*, incluindo a permeabilidade do concreto, o teor de umidade, a taxa de aquecimento e a carga aplicada. O *spalling* explosivo geralmente resulta da combinação das ações de pressão de poros e tensões térmicas, levando a fissuras paralelas à superfície e uma falha violenta da região aquecida (Khoury, 2008).

Em um estudo teórico-experimental, Kirchhof (2010) concluiu que é fundamental considerar a possibilidade de lascamento explosivos em estruturas suscetíveis ao fenômeno, ou seja, de acordo com o próprio estudo, estruturas de concreto com resistência maior que 40 MPa, que apresentem grau de saturação em torno ou superior a 90%.

Mais recentemente, alguns trabalhos foram desenvolvidos sobre a previsão do *spalling* de concreto, especialmente, utilizando técnicas de aprendizagem de máquina. Por exemplo, Naser e Kodur (2022) utilizaram algoritmos de Aprendizado de Máquina Explicável (RF, ExGBT e DL) para avaliar a resistência ao fogo e o *spalling* em pilares de concreto armado, utilizando 494 observações de testes reais, sintéticos e aumentados.

Ho *et al.* (2024) construíram 6 modelos de aprendizado de máquina e encontraram o AdaBoost como o mais preciso para prever grau de *spalling* em pilares de concreto armado, identificando características críticas como resistência à compressão do concreto, relação carga/resistência, esbeltez e largura do pilar.

Sirisena *et al.* (2024) desenvolveram modelos para prever a profundidade do *spalling* em revestimentos de túneis expostos ao fogo. Utilizaram um banco de dados com 415 resultados, incluindo 16 variáveis de entrada. A pesquisa mostrou que o XGBoost foi o método mais eficaz para essa previsão, contribuindo para melhorar o projeto de revestimentos de túneis com maior resistência ao fogo.

Al-Bashiti e Naser (2024) realizaram uma análise de sensibilidade em 15 classificadores de aprendizado de máquina, destacando que os modelos LightGBM, XGBoost e Random Forest são os mais precisos (72-93%) na previsão de *spalling* e identificaram temperatura máxima de exposição, taxa de aquecimento, resistência à compressão, teor de umidade, conteúdo de sílica e quantidade de fibra de polipropileno como fatores críticos.

Liu *et al.* (2024) examinaram o *spalling* explosivo do Concreto de Ultra-Alto Desempenho (CUAD ou UHPC *Ultra-high Performance Concrete*) sob fogo. Analisaram métodos experimentais, previsões numéricas e consequências na resistência ao fogo, propondo estratégias de intervenção estrutural, identificando lacunas de conhecimento e desafios

atuais, traçando caminhos para futuras pesquisas.

Klimek *et al.* (2024) investigaram a influência da expansão térmica dos agregados no spalling de concreto por meio de testes de fogo em elementos cúbicos e cilíndricos, revelando que a restrição térmica aumenta significativamente a profundidade do *spalling* e que o tipo de agregado (quartzo ou basalto) afeta diretamente o comportamento do *spalling*.

3 FUNDAMENTAÇÃO TEÓRICA

A regressão logística (Cox, 1958) tem como objetivo principal estudar a probabilidade de ocorrência de um evento representado pela variável Y , que é qualitativa e dicotômica (no caso em estudo, $Y = 1$ para descrever a ocorrência de *spalling* e $Y = 0$ para descrever a não ocorrência de *spalling*), com base no comportamento de variáveis explicativas (Fávero e Belfiore, 2017).

A regressão logística é modelada pela função sigmoide, conforme equação (1):

$$f(Z) = \frac{1}{1 + e^{-Z}}, \quad (1)$$

sendo Z , conhecido como logito, dado pela equação (2) :

$$Z = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k, \quad (2)$$

em que p indica a probabilidade de ocorrência de determinado evento de interesse, X representa o vetor de variáveis explicativas (ou independentes), e α e β são os parâmetros do modelo.

Vale destacar que Z não representa a variável de interesse, chamada de Y . O principal objetivo é calcular a probabilidade p de ocorrência do evento, com base no logito Z , em função dos parâmetros estimados para cada variável explicativa.

A função logística pode ser entendida como a probabilidade de a variável dependente ser igual a 1, dado o comportamento das variáveis explicativas X_1, X_2, \dots, X_k . Portanto, isolando p da equação (2), pode-se obter a equação (3).

$$P(Y = 1 | X_1, X_2, \dots, X_k) = \frac{1}{1 + e^{-(\alpha + \sum_{i=1}^k \beta_i X_i)}} \quad (3)$$

A natureza binária da variável dependente (0 ou 1) tem propriedades que violam as suposições da regressão múltipla. Primeiro, o termo de erro de uma variável discreta segue a distribuição binomial ao invés da normal, invalidando assim todos os testes estatísticos que se sustentam nas suposições de normalidade. Segundo, a variância de uma variável dicotômica não é constante, criando casos de heteroscedasticidade também (Hair *et al.*, 2009).

Na regressão logística, para avaliar se uma variável é um preditor significativo da variável de saída, utiliza-se frequentemente a estatística de Wald (Field, 2009), calculada

para cada coeficiente $\hat{\beta}_j$ por meio da equação (4).

$$\text{Wald}_j = \frac{\hat{\beta}_j}{\text{EP}(\hat{\beta}_j)}, \quad (4)$$

onde $\hat{\beta}_j$ é a estimativa do coeficiente do previsor j , $\text{EP}(\hat{\beta}_j)$ é o erro padrão associado, e Wald_j é o valor da estatística de Wald.

Outro teste muito usual para verificar a qualidade do ajuste de modelos de regressão logística é o de Hosmer e Lemeshow, cujo princípio consiste na comparação entre os eventos observados e os esperados, com base na divisão da base de dados em 10 grupos, sendo analisado o número de eventos para cada categoria da variável dependente. Este teste refere-se à aplicação de uma estatística χ^2 para avaliar se há diferenças significativas entre as frequências observadas e esperadas em cada faixa. Um valor de p alto indica uma boa adequação do modelo (Fávero e Belfiore, 2017).

4 METODOLOGIA

A linguagem de programação utilizada foi R (versão 4.3.3 (2024-02-29 ucrt)). Para a análise e modelagem dos dados, foi utilizado o software R Studio (versão 2023.12.1+402).

4.1 Banco de dados

Os dados foram obtidos a partir do artigo de Ho *et al.* (2024). Originalmente continham um total de 119 observações de testes em concreto aquecido. O conjunto de dados original incluía 10 variáveis independentes, conforme tabela 1.

Tabela 1: Descrição das variáveis Independentes.

Variável	Descrição
LeR	Relação da parte do pilar exposta ao fogo com o comprimento total.
SR	Índice de esbeltez do pilar.
b (mm)	Largura da seção transversal do pilar.
ρ_l	Taxa de armadura longitudinal.
c (mm)	Cobrimento de concreto.
e (mm)	Excentricidade da carga aplicada.
f _c (MPa)	Resistência à compressão do concreto.
ρ_s	Taxa de armadura transversal.
LR	Relação entre a carga aplicada e a resistência do pilar.
FRexp (min)	Tempo de Resistência ao Fogo.

A variável dependente foi denominada *SPI* (Identificador de *Spalling*), com as seguintes categorias: *N* - Sem spalling; *MN* - Spalling menor; e *MJ* - Spalling maior.

4.2 Pré-processamento dos Dados

Os dados foram pré-processados para garantir a qualidade e a integridade das análises subsequentes. Primeiramente, foi realizada a eliminação de valores faltantes, pois haviam

21 observações sem valores para *SPI*. Em seguida, procedeu-se com a eliminação de dados duplicados; uma linha continha registros repetidos, foi identificada e removida, assegurando que cada observação fosse única. Posteriormente, foram identificados e removidos 3 outliers (observações n.º 2, 15 e 28), para evitar que dados extremos influenciassem de maneira inadequada os resultados. Finalmente, foi realizado o ajuste da estrutura dos dados (agora com 94 observações), no qual a variável *SPI* foi reclassificada: valores originais com *N* foram transformados em 0 e os valores com *MN* ou *MJ* foram convertidos para 1.

4.3 Análise Exploratória

Após o pré-processamento, efetuou-se uma análise exploratória dos dados através de medidas descritivas das variáveis conforme tabelas 2 e 3 e, posteriormente, através de gráficos.

Tabela 2: Estatísticas descritivas das variáveis

	Mín.	Q1.25%	Mediana	Q3.75%	Máx.	Média	DP
LeR	0,510	0,790	1,000	1,000	1,000	0,911	0,123
SR	16,170	21,640	43,420	54,960	99,770	45,116	22,805
b	127,000	200,000	300,000	305,000	406,000	268,596	74,475
ρ_1	0,980	2,090	2,200	2,800	6,980	2,564	1,160
c	19,000	28,000	28,000	48,000	48,000	33,883	10,535
e	0,000	0,000	0,000	15,000	150,000	17,394	35,505
fc	24,400	31,900	36,650	48,250	120,000	51,680	32,949
ρ_s	0,030	0,054	0,070	0,126	1,000	0,121	0,155
LR	0,200	0,510	0,720	0,853	1,460	0,700	0,252
FRexp	25,000	52,250	105,500	197,500	325,000	128,883	85,797

De modo geral, os dados mostram uma ampla variação em cada variável, implicando um intervalo significativo de abrangência. Além disso, observa-se uma elevada dispersão nos valores, evidenciada pelos desvios padrão.

Tabela 3: Estatísticas adicionais das variáveis.

	Assimetria	Curtose	Valor-p de Shapiro-Wilk
LeR	-1,027	0,250	$7,47 \times 10^{-13}$
SR	0,805	0,142	$2,78 \times 10^{-6}$
b	-0,491	-0,318	$1,19 \times 10^{-9}$
ρ_1	2,351	6,512	$4,92 \times 10^{-12}$
c	0,177	-1,485	$1,72 \times 10^{-8}$
e	2,572	6,195	$1,92 \times 10^{-15}$
fc	1,349	0,088	$4,98 \times 10^{-13}$
ρ_s	4,584	22,047	$1,35 \times 10^{-17}$
LR	0,146	-0,131	$5,26 \times 10^{-1}$
FRexp	0,571	-0,976	$2,64 \times 10^{-6}$

Poucas variáveis (*b*, *c* e *LR*) apresentaram assimetria inferior a 0,5 (valor absoluto), podendo ser consideradas simétricas. Por outro lado, as demais variáveis demonstraram

assimetria significativa. Em muitas situações, a presença de assimetria pode indicar a necessidade de transformações nos dados antes da aplicação de modelos estatísticos (Hair *et al.*, 2009). Foram testadas as seguintes transformações: potenciação, raiz, logarítmica, inversa, recíproca quadrática, exponenciação negativa e recíproca logarítmica. Contudo, por não se observar uma melhora significativa nas métricas de avaliação, optou-se por utilizar os dados originais.

Em relação à curtose, as variáveis e , ρ_1 e ρ_s destacaram-se por apresentarem alta curtose, indicando caudas mais pesadas e picos acentuados. Enquanto que c e $FRexp$ apresentaram curtose negativa, caracterizando distribuições mais achatadas.

O teste de Shapiro-Wilk indicou que a maioria das variáveis, exceto LR , não seguem uma distribuição normal.

As distribuições das variáveis foram exploradas visualmente por meio dos histogramas e boxplots apresentados nas figuras 2 e 3. Um histograma é uma representação gráfica de uma única variável que representa a frequência de ocorrências (valores dos dados) dentro de categorias de dados. As frequências são graficamente representadas para examinar a forma da distribuição de valores (Hair *et al.*, 2009).

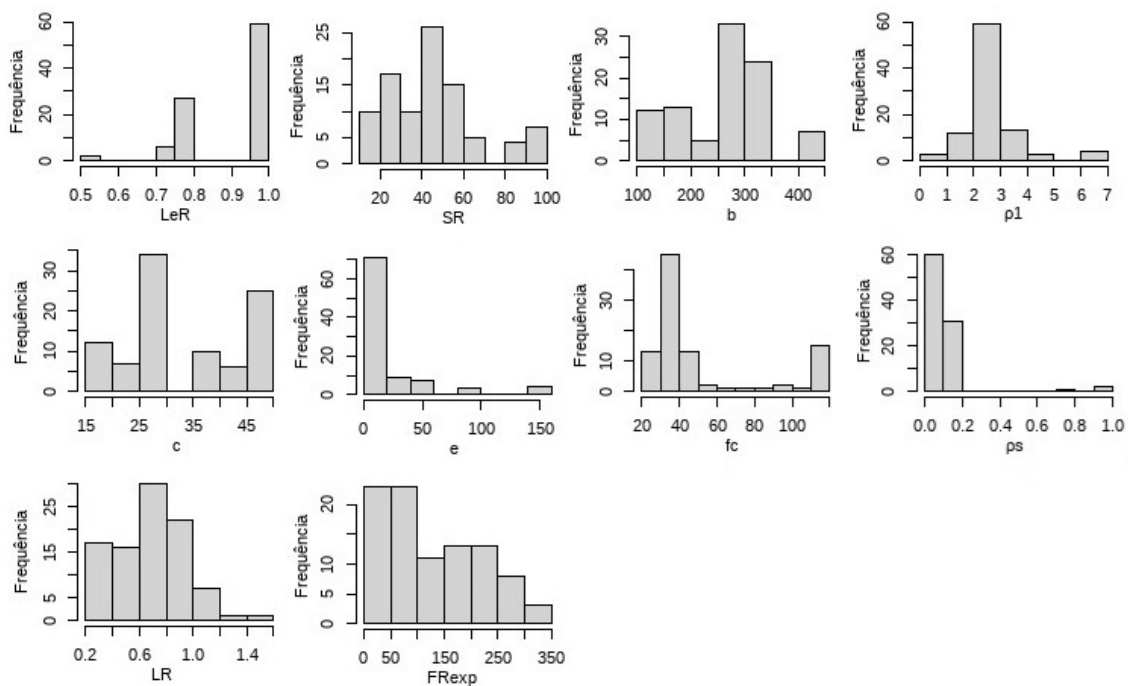


Figura 2: Histogramas

O boxplot, ou diagrama de caixa, é uma representação gráfica que resume cinco medidas principais de uma variável (mínimo, Q1, mediana, Q3 e máximo), permitindo avaliar a simetria e distribuição dos dados. Além disso, ele facilita a identificação visual de outliers, que aparecem fora dos limites superior e inferior. Essas representações gráficas complementam a análise estatística, proporcionando uma análise clara e rápida da estrutura e padrões dos dados (Fávero e Belfiore, 2017).

Procurou-se analisar também a correlação entre variáveis pelo coeficiente de Spearman,

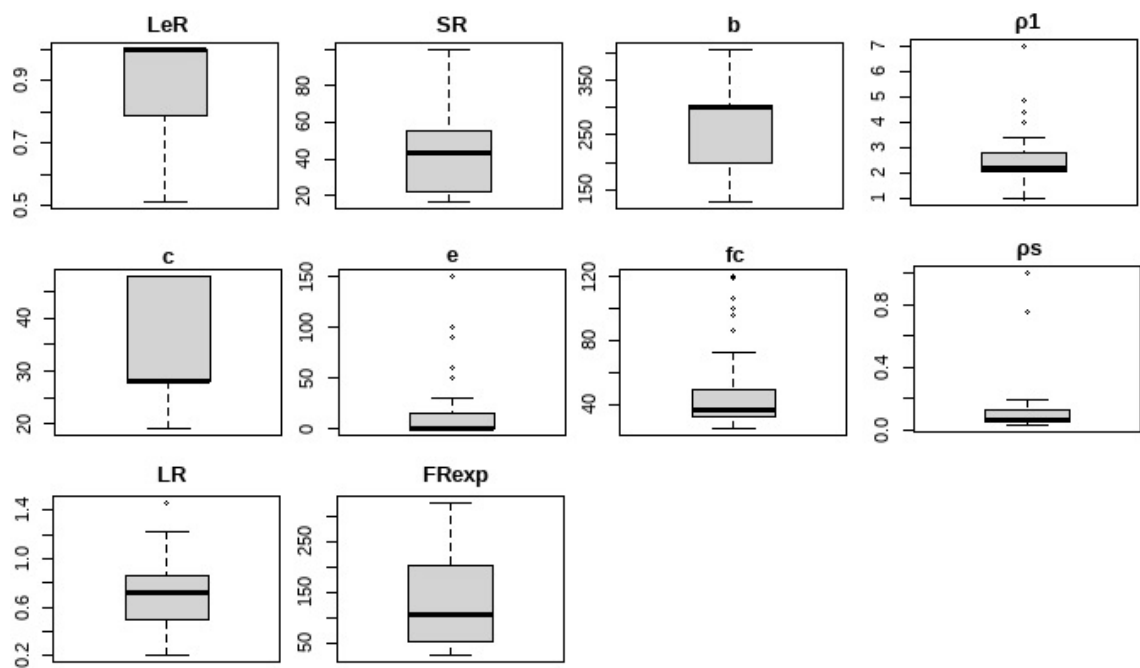


Figura 3: Boxplots

que é uma estatística não-paramétrica, e pode ser usada quando dados não apresentam distribuição normal (figura 4).

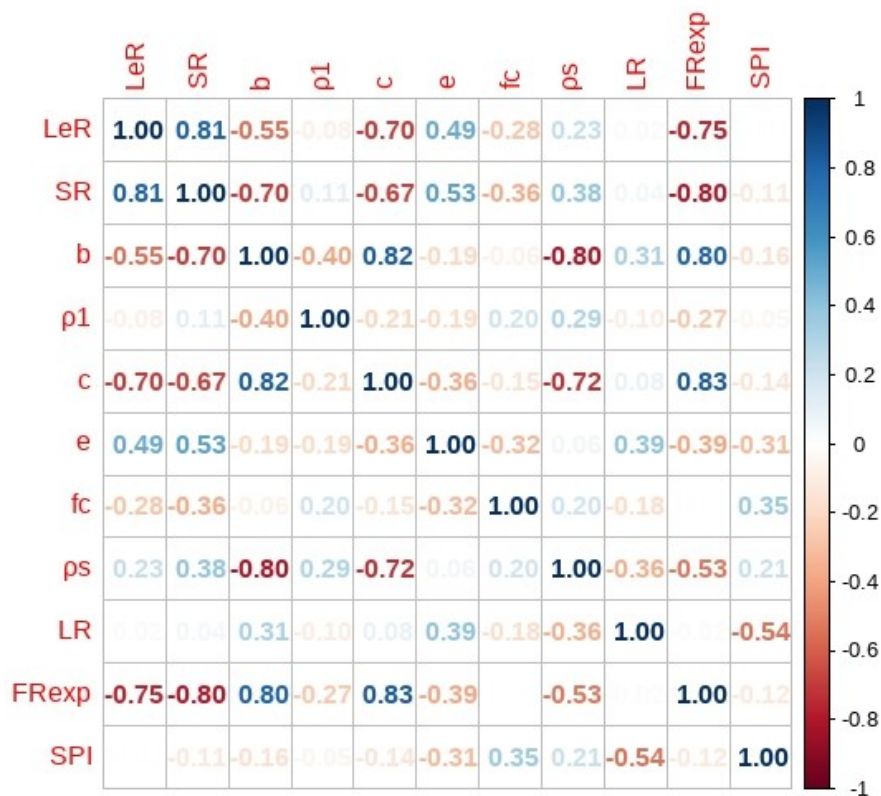


Figura 4: Matriz de Correlação (Spearman)

Observa-se que os valores próximos a zero tem propositadamente a cor esmaecida,

enquanto os valores próximos de ± 1 tem sua cor ressaltada, conforme a escala de cores anexada ao gráfico. A variável *FRexp* foi a que apresentou alta correlação, em módulo, com mais variáveis: *c* (0,83), *b* (0,80), *SR* (-0,80) e *LeR* (-0,75), embora algumas dessas correlações apresentem sinais contrários. Esta forte correlação pode ser explicada fisicamente pelo fato de que *FRexp* representa o tempo de resistência ao fogo, que pode ser dependente de outras características relacionadas ao comportamento do concreto sob condições de incêndio. Especialmente, a relação entre *FRexp* e *c* sugere que o cobrimento e o tempo de resistência ao fogo estão intimamente ligados, em acordo com o estado da arte.

As cinco maiores correlações da variável dependente *SPI*, foram com as variáveis *LR*, *fc*, *e*, ρ_s e *b*, embora nenhuma tenha sido forte, pois o valor máximo não ultrapassou, em módulo, 0,54.

4.4 Modelagem

Para o desenvolvimento dos modelos preditivos, inicialmente foi ajustado um modelo de regressão logística completo (mod1) utilizando todas as variáveis disponíveis, mas sem considerar iterações. A partir deste modelo base, foram aplicadas diferentes abordagens de seleção de variáveis: o método *stepwise forward* (mod2) e *backward* (mod3), ambos baseados no critério de informação de Akaike (AIC), e o método *stepwise both* (mod4), com base no critério de informação bayesiano (BIC). Além disso, foi aplicada uma seleção *stepwise both* considerando o AIC corrigido (AICc) através de uma função customizada (mod5). Subsequentemente, foram testadas várias combinações de variáveis selecionadas manualmente (mod6 a mod15), ajustando-se novos modelos de regressão logística, a fim de identificar as melhores combinações de preditores para o *SPI*. As variáveis de cada modelo estão listadas na tabela 4.

Tabela 4: Variáveis independentes em cada modelo.

Modelo	Variáveis
mod 1	LeR + SR + b + ρ_1 + c + e + fc + ρ_s + LR + FRexp
mod 2	LeR + SR + b + ρ_1 + c + e + fc + ρ_s + LR + FRexp
mod 3	LeR + SR + b + ρ_1 + c + e + fc + ρ_s + LR + FRexp
mod 4	LeR + b + c + e + fc + LR + FRexp
mod 5	LeR + b + ρ_1 + c + e + fc + LR + FRexp
mod 6	LR + fc + b + e + ρ_s
mod 7	LeR + SR + fc + LR + FRexp
mod 8	LR + fc + e + ρ_1
mod 9	LR + fc + b + c + ρ_1
mod 10	LR + fc + b + e + ρ_s + FRexp
mod 11	b + c + e + fc + ρ_s + LR
mod 12	LR + fc + b + e + LeR + FRexp + ρ_1
mod 13	LR + fc + b + e + LeR + FRexp
mod 14	LR + fc + b + e + FRexp
mod 15	LR + fc + b + FRexp

5 RESULTADOS E DISCUSSÕES

O desempenho dos modelos foi avaliado utilizando as seguintes métricas: AIC (Akaike Information Criterion), AICcorrigido, BIC (Bayesian Information Criterion), R² de Nagelkerke, RMSE (*Root Mean Squared Error*), AUC (*Area Under the Curve*), F1-score e Acurácia (ambos com ponto de corte que otimiza o F1-score), conforme a tabela 5 os modelos forma ordenados pelo F1-score.

Tabela 5: Comparação de diferentes modelos com base em múltiplas métricas.

Modelo	AIC	AICc	BIC	RMSE	R ² _N	AUC	F1score	Acur.
mod 1	35,426	38,646	63,403	0,146	0,944	0,996	0,983	0,979
mod 2	35,426	38,646	63,403	0,146	0,944	0,996	0,983	0,979
mod 3	35,426	38,646	63,403	0,146	0,944	0,996	0,983	0,979
mod 4	35,670	37,364	56,017	0,179	0,915	0,991	0,983	0,979
mod 5	34,735	36,878	57,624	0,179	0,929	0,993	0,975	0,968
mod 13	39,348	40,651	57,151	0,231	0,887	0,987	0,975	0,968
mod 12	40,335	42,029	60,682	0,253	0,892	0,986	0,967	0,957
mod 10	44,999	46,302	62,802	0,231	0,858	0,980	0,966	0,957
mod 14	44,359	45,324	59,618	0,253	0,850	0,978	0,958	0,947
mod 6	51,321	52,286	66,580	0,292	0,811	0,968	0,951	0,936
mod 7	54,071	55,037	69,331	0,326	0,794	0,962	0,949	0,936
mod 11	52,084	53,387	69,888	0,273	0,818	0,969	0,941	0,926
mod 15	47,863	48,545	60,580	0,342	0,819	0,973	0,934	0,915
mod 8	61,972	62,654	74,689	0,372	0,731	0,942	0,922	0,894
mod 9	67,085	68,050	82,345	0,372	0,710	0,936	0,906	0,883

Para garantir a qualidade dos resultados, foram realizados os seguintes testes adicionais: avaliação de multicolinearidade, teste de Wald e teste de Hosmer e Lemeshow.

A fim de verificar a presença de multicolinearidade, utilizou-se o Fator de Inflação de Variância (VIF), que mede o grau de correlação de cada variável independente com o conjunto das demais variáveis independentes. Os valores de VIF para cada modelo estão expostos na tabela 6.

A presença de multicolinearidade pode impactar significativamente a estimação e a significância estatística dos coeficientes de regressão. Portanto, foram considerados aceitáveis apenas os modelos com valores de VIF inferiores a 10 (Hair *et al.*, 2009). Os primeiros nove modelos apresentaram valores de VIF elevados, indicando problemas de multicolinearidade. Assim, o modelo com melhores métricas e VIF menor que 10 foi o mod6, cujos valores de VIF foram: 7,6; 3,5; 3,4; 3,6 e 1,8, para as variáveis *LR*, *fc*, *b*, *e* e ρ_s , respectivamente. Portanto, a análise prosseguiu com este modelo, cujas variáveis, inclusive, são as que apresentaram maior correlação com a variável dependente *SPI* (figura 4).

Os resultados do teste de Wald para os coeficientes do modelo estão apresentados na tabela 7. Esta tabela exhibe a estimativa dos coeficientes, o erro padrão, o valor da estatística de Wald e o valor-p, que serve para avaliar a significância estatística dos coeficientes. Valores de *p* inferiores a 0,05 (nível de significância de 5%) indicam que o coeficiente é estatisticamente significativo. Por outro, valores de *p* superiores a 0,05

Tabela 6: Fator de Inflação da Variância (VIF)

	LeR	SR	b	ρ_1	c	e	fc	ρ_s	LR	FRexp
Mod 1	159,2	105,3	443,4	5,8	477,0	9,6	216,8	7,8	375,0	1178,6
Mod 2	159,2	105,3	443,4	5,8	477,0	9,6	216,8	7,8	375,0	1178,6
Mod 3	159,2	105,3	443,4	5,8	477,0	9,6	216,8	7,8	375,0	1178,6
Mod 4	8,6	—	21,6	—	8,7	5,0	21,8	—	51,9	43,0
Mod 5	30,1	—	49,9	3,5	30,4	5,9	54,9	—	115,4	181,8
Mod 6	—	—	3,4	—	—	3,6	3,5	1,8	7,6	—
Mod 7	2,5	5,3	—	—	—	—	5,4	—	5,6	11,4
Mod 8	—	—	—	1,1	—	1,9	2,2	—	2,4	—
Mod 9	—	—	2,3	1,1	1,8	—	1,3	—	2,7	—
Mod 10	—	—	6,4	—	3,2	—	5,4	1,9	13,6	4,4
Mod 11	—	—	4,7	—	2,2	3,4	3,6	2,1	9,7	—
Mod 12	10,7	—	7,6	1,7	3,3	—	9,2	—	20,9	24,9
Mod 13	6,8	—	7,4	—	3,2	—	9,1	—	19,8	17,3
Mod 14	—	—	5,5	—	2,6	—	5,1	—	12,0	4,7
Mod 15	—	—	3,9	—	—	—	3,0	—	6,1	3,9

indicam que não há evidências suficientes para rejeitar a hipótese nula de que o coeficiente é igual a zero.

Tabela 7: Resultados do teste de Wald para os coeficientes do modelo: mod6

Variável	Estimativa	Erro Padrão	Estatística Wald	Valor-p
(Intercepto)	−3,420	2,515	−1,360	0,173 950
LR	−19,426	5,602	−3,468	0,000 525
fc	0,122	0,032	3,820	0,000 134
b	0,033	0,013	2,467	0,013 640
e	0,052	0,018	2,963	0,003 046
ρ_s	−5,423	2,519	−2,153	0,031 325

Conforme pode-se observar na tabela 7, todos os coeficientes (para LR, fc, b, e e ρ_s) do modelo escolhido apresentaram valores de p inferiores a 0,05. Portanto, esses coeficientes têm uma influência estatisticamente significativa na variável dependente, indicando que cada um deles contribui de forma relevante para a explicação da variação observada no modelo.

No presente estudo, o teste de Hosmer e Lemeshow foi aplicado ao modelo escolhido, resultando numa estatística de $\chi^2 = 3,5488$ e um valor de $p = 0,8954$. Este valor elevado de p sugere que não há evidências suficientes para rejeitar a hipótese nula, indicando um bom ajuste do modelo.

A matriz de confusão (figura 5) ilustra a distribuição das previsões do modelo em relação às classes de referência ($Y = 1$ para a ocorrência de *spalling* e $Y = 0$ para a não ocorrência). O ponto de corte foi determinado iterando sobre uma sequência de valores entre 0 e 1, com incrementos de 0,01. Para cada valor, as probabilidades preditas foram convertidas em classes binárias, e o F1-score foi calculado e armazenado. O ponto de corte que maximizou o F1-score foi 0,58, sendo utilizado para gerar as previsões finais.

Como ilustrado na matriz de confusão, o modelo classificou corretamente 58 vezes

		Referência	
		0	1
Previsão	0	58	5
	1	1	30

Figura 5: Matriz de confusão do modelo: mod6.

as não ocorrências e 30 vezes as ocorrências de *spalling*, totalizando 88 acertos em 94 observações. Portanto, o modelo obteve uma acurácia de 93,62%, indicando uma alta proporção de previsões corretas no conjunto de dados analisado. O modelo apresentou uma sensibilidade de 85,71%, indicando boa identificação das ocorrências reais, e uma especificidade de 98,31%, destacando sua excelente capacidade de identificar as não ocorrências. O valor preditivo positivo foi de 96,77%, refletindo a alta qualidade nas previsões de ocorrências, enquanto o valor preditivo negativo foi de 92,06%, demonstrando a eficiência nas previsões de não ocorrências. O F1-score foi de 95,08%, evidenciando um bom equilíbrio entre precisão e *recall*.

Além disso, a área sob a curva ROC (*AUC*) (figura 6) alcançou 96,78%, destacando a elevada capacidade do modelo em distinguir entre as classes.

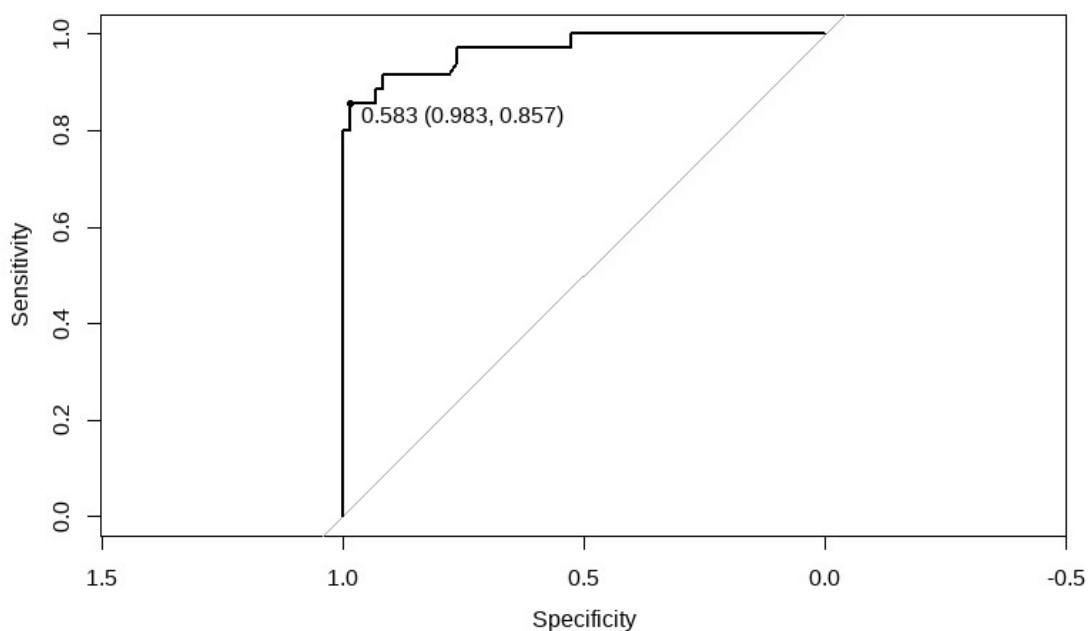


Figura 6: Curva ROC

O R^2 de Nagelkerke calculado foi 81,07%. Esse valor é uma medida pseudo- R^2 usada para avaliar a qualidade de ajuste de modelos de regressão logística. Ao contrário do R^2 tradicional, usado em regressão linear, o R^2 de Nagelkerke ajusta o valor de R^2 de Cox e Snell para que varie entre 0 e 1. Esse valor de 81,07% sugere que o modelo explica uma parte significativa da variação de ocorrência de *spalling*.

Além disso, foram analisadas algumas métricas de erro. Os valores obtidos de RMSE (Erro Quadrático Médio), MAE (Erro Médio Absoluto) e MAPE (Erro Percentual Médio Absoluto) foram 0,2917; 0,0851 e 0,0585, respectivamente, sugerindo que o modelo é preciso, apresentando erros médios relativamente baixos.

6 CONCLUSÃO

Este estudo desenvolveu um modelo de Regressão Logística para prever a ocorrência de *spalling* em pilares de concreto armado expostos ao fogo. O modelo explicou uma porção significativa da variação de ocorrência de *spalling* e atingiu uma elevada acurácia, demonstrando sua eficácia tanto na captura dos fatores críticos que influenciam o fenômeno quanto na sua capacidade preditiva.

A análise dos coeficientes do modelo indicou que as variáveis de relação carga/resistência (LR) e de resistência à compressão do concreto (f_c) têm um papel determinante na manifestação de *spalling*. Variáveis como a excentricidade da carga aplicada (e), a largura da seção do pilar (b) e a taxa de armadura transversal (ρ_s) também contribuíram significativamente para a modelagem, embora em menor grau.

Modelos como o desenvolvido neste estudo, se utilizados por projetistas, permitirão a predição do *spalling*. Com isso, será possível tomar decisões antecipadas e realizar ajustes durante a fase de projeto para evitar esse fenômeno. Além disso, com o crescente estudo nessa área, tais modelos poderão servir para a elaboração de diretrizes normativas sobre o tema.

Portanto, os resultados desta pesquisa podem contribuir com o campo da engenharia estrutural, demonstrando que métodos estatísticos podem ser aplicados como uma alternativa para prever o *spalling* e, consequentemente, na redução de riscos de estruturas de concreto armado em situação de incêndio.

REFERÊNCIAS

- ABNT NBR 15200 (2024). *Projeto de estruturas de concreto em situação de incêndio*. Rio de Janeiro.
- Al-Bashiti, M. K. e M. Z. Naser (2022). “Verifying domain knowledge and theories on Fire-induced spalling of concrete through eXplainable artificial intelligence”. Em: *Construction and Building Materials* 348, p. 128648. DOI: 10.1016/j.conbuildmat.2022.128648.
- Al-Bashiti, M. K. e M. Z. Naser (2024). “A sensitivity analysis of machine learning models on fire-induced spalling of concrete: Revealing the impact of data manipulation on accuracy and explainability”. Em: *Computers and Concrete* 33.4. Special Issue, pp. 409–423. DOI: 10.12989/cac.2024.33.4.409.
- Cox, D. R. (1958). “The regression analysis of binary sequences”. Em: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 20.2, pp. 215–232.

- Field, A. (2009). *Descobrendo a estatística usando o SPSS*. Trad. por L. Viali. 2ª ed. Título original: *Discovering statistics with SPSS*. Artmed.
- Fávero, L. P. e P. Belfiore (2017). *Manual de análise de dados*. 1ª ed. Rio de Janeiro: Elsevier.
- Hair, J. F., W. C. Black, B. J. Babin, R. E. Anderson, e R. L. Tatham (2009). *Análise multivariada de dados*. Trad. por A. S. Sant’Anna. 6ª ed. Bookman.
- Ho, T., T. Nguyen, e G. Truong (2024). “Concrete Spalling Identification and Fire Resistance Prediction for Fired RC Columns Using Machine Learning-Based Approaches”. Em: *Fire Technology* 60, pp. 1823–1866. DOI: 10.1007/s10694-024-01550-8.
- Khoury, G. A. (2008). “Passive fire protection of concrete structures”. Em: *Proceedings of the Institution of Civil Engineers: Structures Buildings* 161.3, pp. 135–145. DOI: 10.1680/stbu.2008.161.3.135.
- Kirchhof, L. D. (2010). “Estudo teórico-experimental da influência do teor de umidade no fenômeno de spalling explosivo em concretos expostos a elevadas temperaturas”. Tese de Doutorado. Porto Alegre: Escola de Engenharia, Universidade Federal do Rio Grande do Sul.
- Klimek, A., L. Stelzner, S. Hothan, *et al.* (2024). “Influence of thermal strain on concrete spalling”. Em: *Materials and Structures* 57.1, p. 15.
- Liu, J.-C., L.-P. Du, Y. Yao, *et al.* (2024). “A close look at fire-induced explosive spalling of ultra-high performance concrete: From materials to structures”. Em: *Archives of Civil and Mechanical Engineering* 24.124. DOI: 10.1007/s43452-024-00942-5.
- Naser, M. Z. e V. K. Kodur (2022). “Explainable machine learning using real, synthetic and augmented fire tests to predict fire resistance and spalling of RC columns”. Em: *Engineering Structures* 253, p. 113824. DOI: 10.1016/j.engstruct.2021.113824.
- Seito, A. I., A. A. Gil, F. D. Pannoni, R. Ono, S. B. Silva, U. D. Carlo, e V. P. Silva (2008). *Segurança contra incêndio no Brasil*. São Paulo: Projeto Editora.
- Silva, V. P. (2016). *Projeto de estruturas de concreto em situação de incêndio: Conforme ABNT NBR 15200:2012*. 2ª ed. São Paulo: Editora Blücher.
- Sirisena, G., T. Jayasinghe, T. Gunawardena, L. Zhang, P. Mendis, e S. Mangalathu (2024). “Machine learning-based framework for predicting the fire-induced spalling in concrete tunnel linings”. Em: *Tunnelling and Underground Space Technology* 153, p. 106000. DOI: 10.1016/j.tust.2024.106000.
- So, H.-S. (2016). “Spalling prevention of high performance concrete at high temperatures”. Em: *InTech*. DOI: 10.5772/64551.
- Tutikian, B. F., G. C. Isaia, e P. Helene (2011). “Concreto de alto e ultra-alto desempenho”. Em: *Concreto: Ciência e Tecnologia*. Ed. por G. C. Isaia. São Paulo, Brasil: IBRACON. Cap. 36.