



UNIVERSIDADE FEDERAL DE MATO GROSSO - UFMT
CAMPUS UNIVERSITÁRIO DO ARAGUAIA - CUA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DE MATERIAIS

CAROLINA LUIZA EMERECIANA PESSOA

**UTILIZAÇÃO DE MACHINE LEARNING PARA PREVISÃO DE
PROPRIEDADES DE CONCRETO AUTO-REPARÁVEL MODIFICADO
COM BACTÉRIA E FIBRA**

Barra do Garças
2023

CAROLINA LUIZA EMERECIANA PESSOA

**UTILIZAÇÃO DE MACHINE LEARNING PARA PREVISÃO DE
PROPRIEDADES DE CONCRETO AUTO-REPARÁVEL MODIFICADO
COM BACTÉRIA E FIBRA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência de Materiais da Universidade Federal de Mato Grosso como parte das exigências para a obtenção do título de Mestre.

Orientador: Dr. Ricardo Stefani

Barra do Garças
2023

Dados Internacionais de Catalogação na Fonte.

P475u Pessoa, Carolina Luiza Emereciana.

Utilização de Machine Learning para Previsão de Propriedades de Concreto Auto-reparável Modificado com Bactéria e Fibra [recurso eletrônico] / Carolina Luiza Emereciana Pessoa. -- Dados eletrônicos (1 arquivo : 46 f., il. color., pdf). -- 2023.

Orientador: Ricardo Stefani.

Dissertação (mestrado) - Universidade Federal de Mato Grosso, Campus Universitário do Araguaia, Programa de Pós-Graduação em Ciência de Materiais, Barra do Garças, 2023.

Modo de acesso: World Wide Web: <https://ri.ufmt.br>.

Inclui bibliografia.

1. Concreto auto-reparável. 2. Machine Learning. 3. Support Vector Machine. 4. Random Forest Regressor. I. Stefani, Ricardo, *orientador*. II. Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

Permitida a reprodução parcial ou total, desde que citada a fonte.



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE MATO GROSSO
PRÓ-REITORIA DE ENSINO DE PÓS GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DE MATERIAIS

FOLHA DE APROVAÇÃO

TÍTULO: Utilização de Machine Learning para Previsão de Propriedades de Concreto Auto-reparável Modificado com Bactéria e Fibra

AUTORA: MESTRANDA **CAROLINA LUIZA EMERENCIANA PESSOA**

Dissertação defendida e aprovada em 25 de julho de 2023.

COMPOSIÇÃO DA BANCA EXAMINADORA

1. **PROF. DR. RICARDO STEFANI** (Presidente Banca / **PRESIDENTE DA BANCA** (Orientador)

INSTITUIÇÃO: UNIVERSIDADE FEDERAL DE MATO GROSSO

2. **[PROF. DR. JONATAS EMMANUEL BORGES** (Membro Interno)

INSTITUIÇÃO: UNIVERSIDADE FEDERAL DE MATO GROSSO

3. **PROFA. DRA. LAYS BATISTA FITARONI** (Membro Externo)

INSTITUIÇÃO: CENTRO DE CURSOS DE TECNOLOGIA E DE EDUCAÇÃO PROFISSIONAL
- REDE CETEPS

BARRA DO GARÇAS, 26/07/2023



Documento assinado eletronicamente por **LAYS BATISTA FITARONI**, **Usuário Externo**, em 16/08/2023, às 08:59, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **JONATAS EMMANUEL BORGES**, **Docente da Universidade Federal de Mato Grosso**, em 16/08/2023, às 10:56, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufmt.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6092361** e o código CRC **E8D4AF54**.

Dedicado esse trabalho aos meus pais,
Meus primeiros professores,
Minha base e também minhas asas.

AGRADECIMENTOS

Agradeço a instituição Universidade Federal de Mato Grosso responsável por permitir e viabilizar minha formação acadêmica. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Agradeço fortemente ao meu orientador, o Professor Dr. Ricardo Stefani pela dedicação com o presente trabalho, por todo o conhecimento compartilhado, paciência e companheirismo durante essa trajetória.

Agradeço ao meu colega de orientação Victor Hugo Peres por todo auxílio e companhia em laboratório. É um privilégio poder dividir as lutas e as glórias.

Sou imensamente grata aos meus pais Ana Pessoa e José Pessoa, que viveram mais uma etapa na minha vida acadêmica comigo. Sorriram com cada vitória e choraram comigo cada tropeço. Sonhamos juntos mais um sonho que agora se realiza.

Agradeço a toda minha família pelo suporte e incentivo, em especial a minha irmã Camila que sempre acreditou de perto e minha avó Anísia, a semente de todos os nossos frutos.

Agradeço por todo apoio e companheirismo de Danillo Camargo Álvares, que me trouxe ânimo e apoio durante mais uma trajetória.

Não podendo esquecer agradecimentos especiais a todos os professores que me capacitaram e a todos os meus amigos e companheiros que tornaram essa caminhada mais suave e menos solitária.

RESUMO

Os concretos auto-reparáveis são estudados há décadas como um material alternativo para contornar problemas como fissuras e baixa durabilidade dos concretos convencionais. No entanto, experimentos laboratoriais podem ser onerosos e demandar muito tempo. O emprego de modelos computacionais preditivos que possam estimar as propriedades do concreto estão sendo pesquisados para desenvolver novos tipos de concreto e suprir os requisitos desafiadores da indústria da construção. Neste trabalho foram desenvolvidos modelos de Machine Learning (ML) utilizando Regressão Linear Composta (MLR), Support Vector Machine (SVM) e Random Forest Regressor (RFG) para prever e analisar a taxa de reparo da área fissurada e a absorção capilar de água em concretos auto-reparáveis modificados com bactérias alcali-resistente, fibras de álcool polivinílico (PVA) e fibras de polipropileno (PP). Os resultados mostram que os modelos de ML obtiveram melhor desempenho na predição da taxa de reparo de área fissurada e que os algoritmos de SVM e RFG entregam modelos mais bem ajustados e eficientes. Além disso, os resultados desse trabalho mostram que o ML pode ser uma ferramenta valiosa para o estudo e desenvolvimento baseado em AI de concretos melhorados.

Palavras-chave: Concreto auto-reparável; Machine Learning, Support Vector Machine; Random Forest Regressor

ABSTRACT

Self-healing concrete has been studied for decades as an alternative material to overcome problems such as cracking and low durability of conventional concrete. However, laboratory experiments can be costly and time-consuming. The use of predictive computational models that can estimate the properties of concrete are being researched to develop new types of concrete and meet the challenging requirements of the construction industry. In this work, Machine Learning (ML) models were developed using Multiple Linear Regression (MLR), Support Vector Machine (SVM) and Random Forest Regressor (RFG) to predict and analyze the repairing rate of the cracked area and capillary water absorption in self-healing concrete modified with alkali-resistant bacteria, polyvinyl alcohol (PVA) fibers and polypropylene (PP) fibers. The results show that the ML models performed better in predicting the repairing rate of the cracked area and that the SVM and RFG algorithms deliver better adjusted and efficient models. Furthermore, the results of this work show that ML can be a valuable tool for the AI-based study and development of improved concrete.

Keywords: Self-healing concrete; Machine Learning, Support Vector Machine; Random Forest Regressor

LISTA DE TABELAS

Tabela 1 – Código numérico para bactérias e fibras.	24
Tabela 2 – Resultados e estatísticas do modelo de MLR para taxa de reparo da área fissurada.	27
Tabela 3 – Resultados e estatísticas do modelo de MLR para absorção capilar de água.	28
Tabela 4 – Parâmetros para cada modelo de SVM.	29
Tabela 5 – Performance dos modelos de SVM para taxa de reparo da área fissurada... ..	29
Tabela 6 – Performance dos modelos de SVM para absorção capilar de água.	31

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO BIBLIOGRÁFICA	15
	2.1 INTELIGENCIA ARTIFICIAL	15
	2.2 MACHINE LEARNING	15
	2.3 TÉCNICAS DE MACHINE LEARNING	16
	2.3.1 Regressão Linear Múltipla	16
	2.3.2 Random Forest Regressor	17
	2.3.3 Support Vector Machine	17
	2.3.3.1 Parâmetros de SVM	19
	2.4 CONCRETOS AUTO-REPARÁVEIS	20
	2.5 UTILIZAÇÃO DE MACHINE LEARNING NA ANÁLISE E PREVISÃO DE MATERIAIS CIMENTÍCIOS	21
3	OBJETIVOS	22
	3.1 OBJETIVO GERAL	22
	3.2 OBJETIVOS ESPECÍFICOS	23
4	MATERIAIS E MÉTODOS	23
	4.1 BASE DE DADOS	23
	4.2 PRÉ-PROCESSAMENTO DE DADOS E SELEÇÃO DE VARIÁVEIS	24
	4.3 DESENVOLVIMENTO DOS MODELOS DE APRENDIZADO DE MÁQUINA	25
	5.4 VALIDAÇÃO DO MODELO	25
5	RESULTADOS E DISCUSSÃO	27
	5.1 REGRESSÃO LINEAR MÚLTIPLA	27
	5.1.1 Taxa de reparo da área fissurada	27
	5.1.2 Absorção capilar de água	28

	12
5.2 SUPPORT VECTOR MACHINE	29
5.2.1 Taxa de reparo da área fissurada	29
5.2.2 Absorção capilar de água	31
5.3 RANDOM FOREST REGRESSOR	32
5.3.1 Taxa de reparo da área fissurada	32
5.3.2 Absorção capilar de água	33
5.4 COMPARAÇÃO DA CAPACIDADE DE PREVISÃO DOS MODELOS DE ML DESENVOLVIDOS	34
5.5 INFLUÊNCIA DAS VARIÁVEIS INDEPENDENTES	36
6 CONCLUSÕES	38
7 TRABALHOS FUTUROS	39
8 REFERÊNCIAS BIBLIOGRÁFICAS	40

1. INTRODUÇÃO

A engenharia civil é um campo diversificado que vai desde os recursos hídricos até o projeto e análise de estruturas, encontrando problemas complexos que requerem intervenção humana na forma de experiência passada, expertise e intuição. Nesse sentido, as técnicas de computação auxiliam as atividades de engenharia, aproveitando o comportamento cognitivo da mente humana para chegar a soluções que economizam tempo e recursos (DEKA, 2019).

À medida que a Indústria 4.0, também chamada de quarta revolução industrial, avança, as máquinas estão sendo conectadas e colaborando entre si para tomar decisões sem interferência humana. Essa nova prática combina diversas tecnologias usadas para automatizar processos, otimizar produtos, reduzir custos, reduzir o desperdício de energia e fornecer informações úteis por meio da análise dos dados coletados (KOTSIPOULOS et al., 2021).

A Inteligência Artificial (IA), apresentada ao público em 1956, é a parte da ciência da computação envolvida no projeto de sistemas computacionais que exibem as características relacionadas à inteligência no comportamento humano, como compreensão, linguagem, aprendizado, raciocínio, resolução de problemas, entre outros. É amplamente definida como “a ciência de fazer os computadores fazerem coisas que o ser humano precisa que a inteligência faça” (KOTSIPOULOS et al., 2021; HUANG; LI; FU, 2019).

Machine Learning (ML) é um subconjunto da IA, sendo a ciência de fazer as máquinas agirem sem interferência humana e sem programação específica, utilizando-se de dados anteriores como forma de aprendizagem. Existem vários algoritmos de ML usados, sendo categorizados como supervisionados ou não supervisionados, dependendo de como aprendem com os dados (KOTSIPOULOS et al., 2021; ELSISI, 2021).

Machine Learning tem provado ser cada vez mais relevante no processo científico e produtivo, auxiliando ou substituindo métodos estatísticos tradicionais e modelos mecânicos em várias aplicações, incluindo na engenharia civil, onde esses modelos tradicionalmente dominam (VADYALA et al., 2022). Com o auxílio de Machine Learning é possível, por exemplo, realizar melhoria de produtividade, prognósticos e predição de falhas, estimar o tempo de vida de uma máquina ou material, reduzir falhas e minimizar a ocorrência de paradas não planejadas em diferentes processos, aumentar a eficiência no uso de recursos financeiros e humanos, entre outras aplicações, utilizando-se de uma grande quantidade de dados para treinar um algoritmo de Machine Learning (KOTSIPOULOS et al., 2021; DALZUCHIO et al., 2020).

A utilização de modelos preditivos de ML se mostra como uma ferramenta eficaz e de

grande viabilidade econômica para estimar propriedades e comportamentos de materiais, inclusive materiais cimentícios, bem como auxiliar a desenvolver novos materiais, e satisfazer parte das grandes demandas da construção civil. Modelos preditivos capazes de estimar as propriedades do concreto estão sendo constantemente desenvolvidos para pesquisar novos tipos de concreto estimulados pelos requisitos cada vez mais desafiadores da indústria da construção (BEN CHAABENE; FLAH; NEHDI, 2020)

Sendo o concreto o material de construção mais popular no mundo, seu uso em diversos tipos de estruturas é mais amplo do que qualquer outro material, bem como a quantidade produzida. Entretanto, a formação de fissuras é um fenômeno comum no concreto, podendo causar redução da resistência à penetração contra cloretos e água e a perda de rigidez nas estruturas de concreto, encurtando sua vida útil e afetando negativamente sua durabilidade (LI et al., 2022; HUANG et al., 2022; HUANG et al., 2020).

Diante disso, a ideia de projetar um concreto com comportamento auto-regenerativo para reparar as fissuras têm atraído grande atenção nas últimas décadas, visto que obter um método de reparo eficiente pode ser capaz de diminuir o custo de manutenção, economizar energia e reduzir o impacto ambiental. Materiais expansíveis como polímeros, fibras, misturas minerais, microencapsulação e microorganismos estão sendo utilizados para recuperar microfissuras (ZHUANG; ZHOU, 2019; KUMAR JOGI; VARA LAKSHMI, 2021).

Experimentos mostram que argamassas cimentícias modificadas com esporos bacterianos microencapsulados obtiveram uma maior área de fissura reparada e apresentaram menor permeabilidade geral à água do que argamassas sem modificações microbiológicas. Fibras sintéticas como álcool polivinílico, álcool etileno-vinílico, poliacetal e polipropileno também têm sido usadas para o controle de fissuras e da permeabilidade à água (WANG et al., 2014; NISHIWAKI et al., 2012; MANVITH; RAMESH; MACRIN, 2020).

Entretanto, experimentos laboratoriais podem ser extremamente onerosos e demandar muito tempo. Além disso, eventuais erros na dosagem do concreto podem afetar ainda mais os recursos financeiros e o uso do material, visto que seu processo de preparo é irreversível (ZIOLKOWSKI; NIEDOSTATKIEWICZ, 2019). Dessa forma, o presente estudo tem como foco implementar algoritmos de Machine Learning de previsão e análise que se utilizam de dados empíricos para prever a taxa de reparo de área fissurada e a absorção capilar de água em concretos auto-reparáveis modificados com bactérias alcali-resistentes, fibras de álcool polivinílico (PVA) e fibras de polipropileno (PP).

2. REVISÃO BIBLIOGRÁFICA

2.1 Inteligência Artificial (IA)

A Inteligência Artificial (IA) foi inicialmente apresentada em uma conferência realizada na Universidade de Dartmouth, no ano de 1956, sendo amplamente definida como “a ciência de fazer os computadores fazerem coisas que o ser humano precisa que a inteligência faça”. Assim, a IA refere-se a capacidade que uma máquina possui de executar as mesmas funções que a mente humana, exibindo características relacionadas à inteligência no comportamento humano, como compreensão, linguagem, aprendizado, raciocínio, resolução problemas, etc (KOTSIPOULOS et al., 2021; HUANG; LI; FU, 2019).

O desenvolvimento das técnicas de IA passaram pelo período de incubação (antes de 1956), o período de formação (1956-1970), o período sombrio (1966-1974), o período de aplicação do conhecimento (1970-1988) e o período de desenvolvimento integrado (1986-presente). Atualmente, com o grande desenvolvimento da computação em nuvem e do poder computacional, é relativamente fácil aplicar IA ou subconjuntos de IA, apresentando um crescimento explosivo (KOTSIPOULOS et al., 2021; HUANG; LI; FU, 2019).

Setores de engenharia, economia, medicina e outros aplicam extensivamente IA, podendo ser utilizada para modelagem, identificação, otimização, previsão, controle de sistemas complexos, etc. No contexto da “Indústria 4.0”, ou quarta revolução industrial, a engenharia civil e a indústria da construção passam por constantes inovações em direção à digitalização e inteligência, a fim de aumentar a automação, produtividade e confiabilidade nos seus processos (HUANG; LI; FU, 2019; PAN; ZHANG, 2021).

2.2 Machine Learning (ML)

O aprendizado de máquina, ou Machine Learning (ML), é um grande avanço da IA para ensinar as máquinas a descobrir padrões ocultos em grandes volumes de dados e realizar previsões ou observações futuras baseadas em dados de aprendizagem. O aprendizado de máquina representa um subconjunto importante da IA para aprender dados suficientemente

robustos de várias fontes e, em seguida, tomar decisões assertivas sem a interferência humana ou uma programação específica (PAN; ZHANG, 2021).

O aprendizado de máquina geralmente se concentra no desenvolvimento de suas habilidades por meio de observações, experiências passadas e amostras de dados de treinamento, contendo diferentes programas e algoritmos para projetar um modelo de aprendizado para auto-aperfeiçoamento quando exposto a novos conjuntos de dados (KHAN et al., 2022; KHAMBRA; SHUKLA, 2021). Eles podem ser categorizados como supervisionados ou não supervisionados, dependendo de como aprendem com os dados. Os supervisionados aprendem a executar tarefas a partir de modelos rotulados – resultados pré-definidos – que apresentem uma condição a ser atingida e que a máquina pode medir sua precisão. Os não supervisionados aprendem a executar tarefas a partir de dados não rotulados, com base nos padrões semelhantes, sem uma condição pré-definida para atingir (NETO, CASCUDO, CARASEK, 2021).

Modelos de Machine Learning possuem um domínio de aplicação considerável, no entanto, previsão e análise são um dos domínios mais adequados, onde conjuntos de dados anteriores são usados para prever e fornecer uma avaliação mais precisa das probabilidades futuras. Um dos pontos importantes no conjunto de dados é que os dados devem conter parâmetros que afetem o comportamento desejado (KHAN et al., 2022; KHAMBRA; SHUKLA, 2021; SUN et al., 2019).

2.3 Técnicas de Machine Learning

Os algoritmos de machine learning se utilizam de diferentes técnicas de ML, como regressão e classificação, podendo apresentar funções diversificadas. Alguns dos diversos algoritmos de ML serão discutidos abaixo.

2.3.1 Regressão Linear Múltipla (MLR)

A regressão linear múltipla é uma versão estendida da regressão linear simples, sendo um modelo estatístico associado a duas ou mais variáveis, usadas para prever o resultado de uma variável de resposta com base em diferentes parâmetros (KHAMBRA; SHUKLA, 2021).

O algoritmo modela uma função linear de aproximação entre um conjunto de variáveis independentes 'x' e a variável dependente 'y', minimizando o erro residual, podendo ser expressa pela Equação 1:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (1)$$

sendo β_1 e β_2 os coeficientes polinomiais de x_1 x_2 , n é o número de variáveis independentes (PRAKASH; SHARMA; SAHU, 2023; NIU et al., 2019).

2.3.2 Random Forest Regression (RFR)

Random Forest Regression combina o desempenho de numerosos algoritmos de árvores de decisão, ou seja, um conjunto de condições ou restrições organizadas hierarquicamente e aplicadas sucessivamente de uma raiz a uma folha da árvore, para classificar ou prever o valor de uma variável (RODRIGUEZ-GALIANO et al., 2019; BREIMAN, 1984).

A partir de um nó da árvore o Random Forest recebe um vetor de entrada (x), um pequeno conjunto de variáveis de entrada selecionadas do conjunto total de uma determinada área de treinamento, sendo considerado aleatoriamente para o particionamento binário. Dessa forma, o Random Forest constrói um número K de árvores de decisão (RODRIGUEZ-GALIANO et al., 2019; WANG et al., 2016).

O valor previsto de uma observação é calculado pela média de todas as árvores. Após o crescimento K dessas árvores $T(x)_1^K$, o preditor de regressão é dado pela Equação 2:

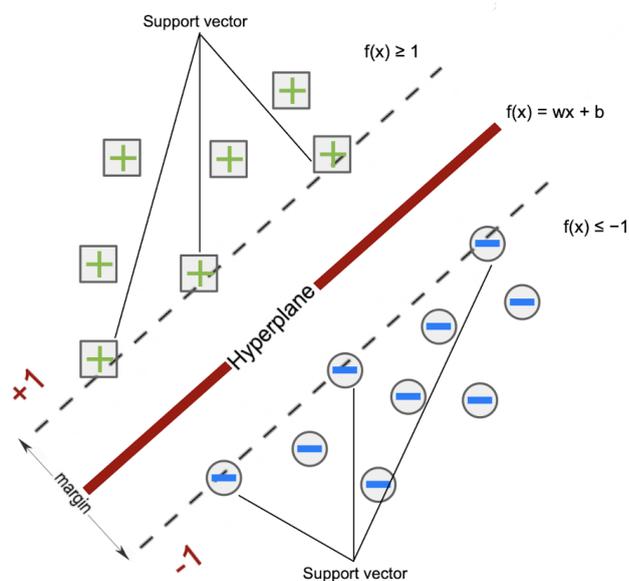
$$\hat{f}_{rf}^K(x) = \frac{1}{K} \sum_{k=1}^K T(x) \quad (2)$$

2.3.3 Support Vector Machine (SVM)

O SVM é um algoritmo de aprendizado de máquina supervisionado, que emprega a teoria de aprendizado estatístico e o princípio de minimização de risco estrutural, analisando dados para classificação. Ele visa encontrar uma margem máxima de separação entre duas classes construindo um hiperplano de classificação no centro da margem máxima, sendo assim considerado como um classificador linear de duas classes (BEN-HUR et al., 2008; ARMAGHANI et al., 2020).

Uma maneira simples de classificação de pontos é traçar uma linha reta de separação e chamar os pontos de um lado positivo e do outro lado negativo, de modo que os dois conjuntos fiquem bem separados, como ilustra a Figura 1. Os pontos de treinamento mais próximos do hiperplano são chamados de vetores de suporte e uma vez que a superfície de decisão é alcançada, a classificação de novos dados pode ser realizada (ARMAGHANI et al., 2020).

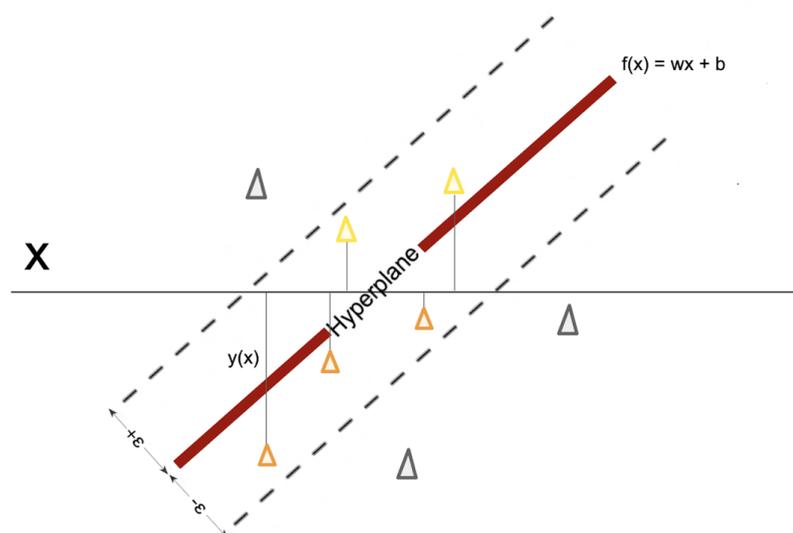
Figura 1 - Esquema do SVM para classificação de dados.



Fonte: Adaptado de Armaghani et al. (2020)

O método de classificação (Support Vector Classifier) pode ser estendido para resolver problemas de regressão (Support Vector Regression), podendo fornecer melhores soluções para otimizar o problema e diminuir o número de erros no modelo projetado, sendo ilustrado pela Figura 2. Para classificação, a minimização da margem é realizada sob a condição de que todos os pontos sejam classificados corretamente, enquanto a regressão segue a condição de que $f(x)$ esteja dentro de uma precisão exigida (ϵ) do valor $y(x)$ (KHAMBRA; SHUKLA, 2021; KLUYVER et al., 2016; ARMAGHANI et al., 2020).

Figura 2 - Esquema do SVM para regressão de dados.



Fonte: Adaptado de Armaghani et al. (2020)

2.3.3.1 Parâmetros do SVM

Para melhorar a precisão do SVM, alguns parâmetros precisam ser ajustados, dentre os principais estão os parâmetros Kernel, C e Gamma.

O algoritmo de Support Vector Machine transforma observações não lineares para o lineares, alterando para um formato simples e processável. Essa formatação de dados é realizada usando uma função matemática conhecida como função kernel, onde sua principal função é transformar um espaço de entrada de baixa dimensão em um espaço de dimensão superior (ARMAGHANI et al., 2020).

Utilizando o parâmetro kernel, diferentes funções podem ser aplicadas para diferentes necessidades. As mais utilizadas são as funções gaussianas (Kernel RBF), as sigmóides (Kernel Sigmóide), as polinomiais (Kernel Polinomial) e as lineares (Kernel Linear).

O parâmetro C representa o erro de classificação, compensando a classificação incorreta de exemplos de treinamento. Valores maiores para o parâmetro C correspondem a uma margem menor, enquanto valores menores correspondem a margens mais largas e abrangentes.

Se os dados possuírem muitas observações ruidosas, é preferível diminuir o C, o que corresponde a uma maior regularização. Um valor mais alto para C é capaz de classificar corretamente todos os pontos dos dados, porém corre o risco de ocorrer um sobreajuste

(quando um modelo estatístico se ajusta bem aos dados já vistos, porém se mostra ineficaz para prever novos resultados) (PEDREGOSA et al., 2011).

O parâmetro gamma define o quanto de influência um único exemplo de treinamento possui. Dessa forma, quanto maior o gamma, mais próximo os pontos devem estar para serem afetados.

2.4 Concretos auto-reparáveis

O concreto é o material de construção mais popular no mundo. Isso ocorre devido a sua alta resistência à compressão, excelente trabalhabilidade e baixo preço, podendo se adaptar a uma vasta gama de mudanças ambientais. Entretanto, a formação de fissuras é um fenômeno comum no concreto, afetando negativamente sua durabilidade (LI et al., 2022; HUANG et al., 2022).

Outro problema evidente é o custo de reparo de estruturas de concreto fissurado, representando uma parte significativa do orçamento da obra devido às operações complexas. Diante disso, a ideia de projetar um concreto com comportamento auto-regenerativo para reparar as fissuras têm atraído grande atenção nas últimas décadas, visto que obter um método de reparo eficiente pode ser capaz de diminuir o custo de manutenção, economizar energia e reduzir o impacto ambiental (HUANG et al., 2022; ZHUANG; ZHOU, 2019).

Fissuras menores que 0,2 mm não são consideradas um caso grave, podendo ser preenchidas pelo próprio concreto, porém fissuras maiores que 0,2 mm e fissuras internas não visíveis durante a inspeção podem afetar drasticamente a durabilidade do concreto. Elas também podem gerar a redução da resistência à penetração contra cloretos e água e a perda de rigidez nas estruturas de concreto, encurtando sua vida útil (HUANG et al., 2022; VIJAY; MURMU; DEO, 2017; HUANG et al., 2020).

A precipitação do carbonato de cálcio tem sido considerada como o fator de maior influência no reparo de fissuras. Com o intuito de desenvolver concretos auto-reparáveis, materiais expansíveis como polímeros, fibras, misturas minerais, microencapsulação e microorganismos estão sendo utilizados para recuperar microfissuras (QIAN et al., 2015; KUMAR JOGI; VARA LAKSHMI, 2021).

Visto que o concreto é um material altamente alcalino, bactérias capazes de suportar ambientes alcalinos são introduzidas, fazendo com que a precipitação de carbonato de cálcio induzida microbiologicamente ajude a preencher as microfissuras e ligar os outros materiais,

como areia, cascalho no concreto (KUMAR JOGI; VARA LAKSHMI, 2021; VIJAY; MURMU; DEO, 2017).

Experimentos mostram que corpos de prova de argamassas cimentícias modificadas com esporos bacterianos microencapsulados obtiveram uma largura máxima de fissura reparada de 970 μm , cerca de 4 vezes maior do que a largura máxima de fissura reparada por agentes não bacterianos (250 μm). Além disso, a permeabilidade geral à água nos corpos de prova modificados com bactéria foi cerca de 10 vezes menor do que na série não bacteriana (WANG et al., 2014).

Fibras sintéticas também têm sido usadas para o controle de fissuras causadas pelo processo de retração do concreto e para redução da permeabilidade, impedindo a penetração de água pelos poros. Experimentos utilizando materiais como álcool polivinílico, álcool etileno-vinílico, poliacetal e polipropileno mostram que a alta polaridade dos compósitos de fibra tem uma maior taxa de precipitação de carbonato de cálcio em fissuras de até 0,3 mm, reduzindo também a permeabilidade à água devido à precipitação em torno das fibras (NISHIWAKI et al., 2012; MANVITH; RAMESH; MACRIN, 2020).

2.5 Utilização de Machine Learning na análise e previsão de materiais cimentícios

Modelos preditivos capazes de estimar as propriedades do concreto estão sendo constantemente desenvolvidos para pesquisar novos tipos de concreto estimulados pelos requisitos cada vez mais desafiadores da indústria da construção. As técnicas de aprendizado de máquina, também chamadas de Machine Learning (ML), têm ganhado atenção na engenharia civil, podendo gerar resultados com alto grau de precisão (LI et al., 2022; BEN CHAABENE; FLAH; NEHDI, 2020).

Polímeros reforçados com fibras são amplamente adotados em confinamento passivo para o concreto armado, melhorando sua resistência à compressão e a ductilidade. Casardi et al. (2017) apresentaram um modelo baseado em rede neural artificial (ANN) para a previsão da resistência à compressão do concreto restrito por polímeros reforçados com fibras.

Estruturas de concreto reforçado com barras de polímero reforçado com fibra de vidro têm maior durabilidade e a avaliação das propriedades de adesão dessas barras ao concreto é de grande importância. Yan et al. (2017) propuseram um modelo híbrido para prever a resistência de aderência de barras de de polímero reforçado com fibra de vidro ao concreto

utilizando a capacidade de mapeamento não linear da Rede Neural Artificial (ANN) com a capacidade de pesquisa global do Algoritmo Genético (GA).

O concreto com agregado reciclado pode reduzir o desperdício de construção, e diferentes técnicas têm sido usadas para prever e avaliar as propriedades do RAC nos últimos anos. Naderpour et al. (2018) aplicaram o modelo ANN para prever a resistência à compressão do concreto com agregado reciclado.

Os algoritmos de Support Vector Machine (SVM) têm sido amplamente utilizados como uma ferramenta para extrair características e detectar danos, embora ainda seja um grande desafio escolher os parâmetros adequados. Gui et al. (2017) combinaram SVM orientado a dados com tecnologia de otimização para aprimorar a extração de recursos e a otimização de parâmetros no monitoramento de integridade estrutural e na detecção de danos.

No monitoramento de integridade estrutural, as técnicas de processamento de imagens são usadas principalmente no processamento de imagens para extrair características defeituosas, como rachaduras em superfícies de concreto e aço. Cha et al. (2017) propuseram um método baseado em reconhecimento visual utilizando redes neurais convolucionais (CNNs) para detectar fissuras de concreto.

Algoritmos de ML têm sido aplicados no campo da engenharia civil há décadas, porém, recentemente, a IA pode aproveitar informações valiosas com as tecnologias de Big Data. Assim, espera-se que a cada ano, apareçam ainda mais exemplos aplicados à engenharia civil [40,44]. (HUANG; LI; FU, 2019; PAN; ZHANG, 2021).

3. OBJETIVOS

3.1 Objetivo Geral

Este estudo tem como objetivo geral a implementação de algoritmos de Machine Learning de previsão e análise para prever a taxa de reparo de área fissurada e a absorção capilar de água em concretos auto-reparáveis modificados com bactérias álcali-resistente, fibras de álcool polivinílico (PVA) e fibras de polipropileno (PP).

3.2 Objetivos específicos

- Implementar algoritmos de Machine Learning como Regressão Linear Múltipla, Support Vector Machine e Random Forest Regressor para prever a taxa de reparo de área fissurada e a absorção capilar de água em concretos auto-reparáveis modificados com bactérias alcali-resistente, fibras de álcool polivinílico (PVA) e fibras de polipropileno (PP);
- Analisar a performance de cada modelo de ML desenvolvido;
- Comparar a capacidade de predição de cada modelo de ML desenvolvido, elucidando qual algoritmo apresenta maior eficácia no estudo em questão;
- Analisar a influência de cada componente (variáveis) na taxa de reparo de área fissurada e a absorção capilar de água em concretos auto-reparáveis modificados.

4. MATERIAIS E MÉTODOS

4.1 Base de dados

A base de dados empírica utilizada foi compilada a partir de dois trabalhos da literatura (FENG; SU; QIAN, 2019; SU et al., 2021). Os autores produziram diversos tipos de amostras de argamassa cimentícia para avaliar a influência auto reparadora de cada material, bem como seu efeito acoplado. Os materiais utilizados para modificar as argamassas cimentícias foram as bactérias alcalino-resistentes *Bacillus alcalophilus* e *Bacillus cereus* GLRT202, fibra de álcool polivinílico (PVA), fibra de polipropileno de 6 mm de comprimento (PP1) e fibra de polipropileno de 9 mm de comprimento (PP2). Um conjunto de dados (planilha) com 684 linhas foi gerado pela extração de dados com o auxílio de um script interno escrito em Object Pascal. Os dados para bactérias e fibras, que são categóricos, foram codificados numericamente conforme mostrado na Tabela 1. Os modelos ML tem como objetivo prever a taxa de reparo de área fissurada e a absorção capilar de água em l/mm.

Tabela 1 - Códigos numéricos para bactérias and fibras

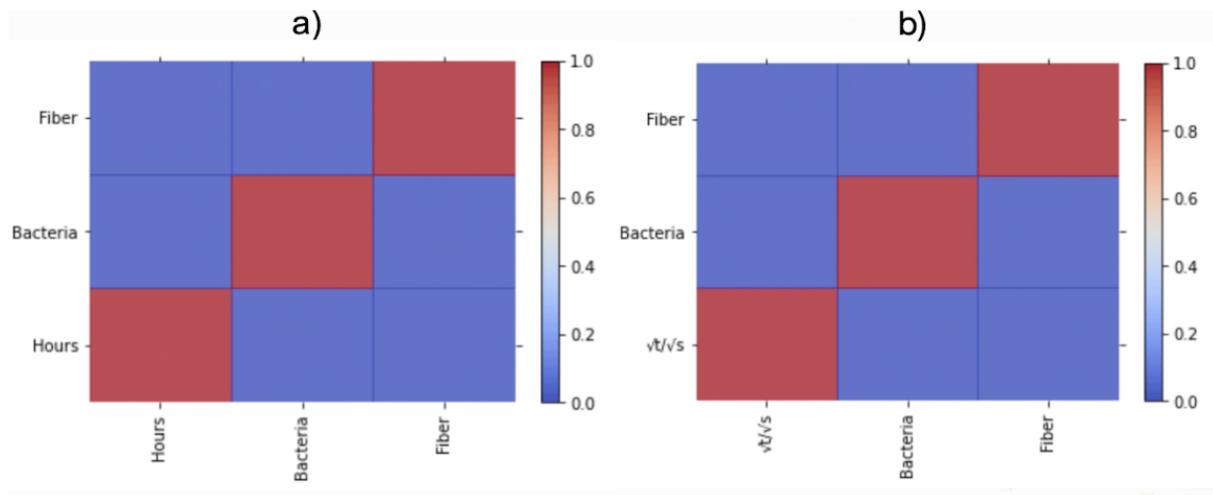
Bactéria	
Nome	Código
<i>Bacillus alcalophilus</i>	1
<i>Bacillus Cereus</i>	2
Sem bactéria	0
Fibra	
Nome	Código
Álcool Polivinílico (PVA)	1
Polipropileno 6 mm (PP1)	2
Polipropileno 9 mm (PP2)	3
Sem fibra	0

Fonte: Autoria própria.

4.2 Pré-processamento de dados e seleção de variáveis

Dados de treinamento contendo variáveis relacionadas à taxa de reparo da área fissurada e variáveis relacionadas à absorção capilar de água foram utilizados para implementar modelos de Machine Learning (ML) utilizando Regressão Linear Múltipla (MLR), Random Forest Regressor (RF) e Support Vector Machine (SVM). A seleção das variáveis foi realizada a partir de matrizes de correlação, conforme mostrado na Figura 3, e pela análise dos dados para encontrar valores constantes ou quase constantes. A análise dos dados mostra que os valores das colunas de cimento e areia são constantes, enquanto os valores da água são quase constantes (apenas 170 ou 270) e como é bem conhecido que constantes ou quase constantes não influenciam o treinamento de aprendizado de máquina, bem como as previsões, eles foram removidos do conjunto de dados final. Finalmente, as seguintes colunas de variáveis foram selecionadas para modelagem de ML: Tempo (horas ou \sqrt{t}/\sqrt{s}), Bactéria e Fibra, então os dados foram usados como um conjunto de treinamento para Modelos de Aprendizado de Máquina. Antes do treinamento e teste, os dados foram divididos em conjuntos de treinamento e teste em uma proporção de 8:2 usando a função scikit-learn `train_test_split` e um estado aleatório de 42. O conjunto de treinamento pode ser acessado através do repositório do GitHub, disponível em <https://github.com/clpessoa/Machine-Learning-Models>.

Figura 3 - Matriz de correlação das variáveis relacionadas a (a) taxa de reparo da área fissurada e (b) absorção capilar de água.



Fonte: Autoria própria.

4.3 Desenvolvimento dos modelos de aprendizado de máquina

Os modelos de ML foram desenvolvidos com a linguagem Python e as seguintes bibliotecas: Scikit-learn versão 1.3.0 (PEDREGOSA et al., 2011), Pandas versão 1.3.4 (MCKINNEY et al., 2010), NumPy versão 1.21.3 (HARRIS et al., 2020), Plotly versão 5.3.1 (INC. PT., 2015), Matplotlib versão 3.3.4 (HUNTER, 2007), Scipy versão 1.7.3 (VIRTANEN, 2020) e Seaborn versão 0.11.1 (WASKOM, 2021). Os modelos de Machine Learning estão disponíveis como Jupyter Notebooks e podem ser acessados através do repositório do GitHub, disponibilizado acima.

4.4 Validação do modelo

A capacidade preditiva dos modelos ML foi avaliada pelo coeficiente de determinação (R^2), erro absoluto médio (MAE) e pela raiz quadrada do erro médio (RMSE).

R^2 é uma medida estatística que indica a proporção da variância na variável dependente que pode ser explicada pela variável independente em um modelo de regressão (bem ajustado). Um valor mais alto de R^2 significa melhor ajuste e desempenho do modelo. Ele representa uma magnitude estatística significativa para avaliar o desempenho de previsão

de modelos de ML variando de 0 a 1, onde valores maiores significam melhor desempenho. O valor de R^2 pode ser calculado pela Equação 3:

$$R^2 = \frac{\sum_{i=1}^n (y'_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

MAE é uma métrica que indica a média da diferença absoluta entre os valores reais e previstos. Essa é uma métrica comum usada para avaliar modelos de regressão e pode avaliar o desempenho do modelo mesmo quando há muitos *outliers* (dados discrepantes) no conjunto de dados. Um valor de MAE tendendo a 0 indica um bom desempenho do modelo. O valor de MAE pode ser calculado pela Equação 4:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (4)$$

O RMSE reflete a precisão da previsão por ser sensível aos erros extremos dos valores de previsão. Um RMSE mais baixo indica uma maior precisão do modelo de ML. O valor de RMSE pode ser calculado pela Equação 5:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y'_i - y_i)^2}{n}} \quad (5)$$

Para as equações acima, tem-se que y'_i é o valor previsto, y_i é o valor real e n é o número de amostras de dados.

Embora todas essas medidas estatísticas forneçam algumas informações úteis sobre o desempenho de um modelo de regressão de Machine Learning, apenas uma medida não deve ser suficiente na avaliação de um modelo estatístico. Portanto, as conclusões devem ser tiradas analisando todas as medidas em conjunto.

5. RESULTADOS E DISCUSSÃO

5.1 Regressão Linear Múltipla

5.1.1 Taxa de reparo da área fissurada

O modelo de Regressão Linear Múltipla para análise e previsão da taxa de reparo da área fissurada apresentou um valor de R^2 igual a 0,47 e um valor de RMSE igual a 0,029. A partir do modelo, a Equação (6) final pode ser obtida:

$$\text{Taxa de reparo} = 0.0006 * \text{Horas} + 0.0916 * \text{Bactéria} - 0.0147 * \text{Fibra} + 0.1359 \quad (6)$$

As estatísticas detalhadas da regressão são apresentadas pela Tabela 2.

Tabela 2 - Resultados e estatísticas do modelo de MLR para taxa de reparo da área fissurada.

Variáveis Independentes	Coefficientes	Standard Error	t-value	p-value
Horas	0,0006	$3,65 \times 10^{-5}$	15,75	< 0,03
Bactéria	0,0916	0,008	11,20	< 0,03
Fibra	-0,0148	0,007	-2,15	0,032
Intercept	0,1359			
Outras estatísticas				
F-Value			127,0	
Kurtosis			2,873	
R-Squared			0,412	
Adjusted R-Squared			0,409	

Fonte: Autoria própria.

Com o auxílio da Tabela 2 e da Equação 6, fica claro que a Regressão Linear Múltipla possui baixo poder preditivo e não será útil para prever a taxa da área de reparo com boa precisão. No entanto, observa-se que a taxa de reparo da área fissurada tem correlação direta com o uso de bactérias e correlação inversa com o uso de fibras. Ou seja, o tipo e a presença

de bactérias afetam positivamente a área de reparo da trinca, enquanto o tipo e a presença de fibras afetam negativamente. Esse achado está de acordo com outros estudos, uma vez que as fibras de PP não são tão eficientes quanto as fibras de PVA para auxiliar na auto cicatrização isoladamente, sem o auxílio de bactérias (FENG; SU; QIAN, 2019; SU et al., 2021; FENG et al., 2021).

5.1.2 Absorção capilar de água

O modelo de Regressão Linear Múltipla para análise e previsão da absorção capilar de água apresentou um valor de R^2 igual a 0,55 e um valor de RMSE igual a 0.088. A partir do modelo, a Equação (7) final pode ser obtida:

$$\text{Absorção} = 0.0055 * \sqrt{t}/\sqrt{s} + 0.161 * \text{Bactéria} - 0.0903 * \text{Fibra} + 0.4641$$

(7)

As estatísticas detalhadas da regressão são apresentadas pela Tabela 3.

Tabela 3 - Resultados e estatísticas do modelo de MLR para absorção capilar de água.

Variável Independente	Coefficientes	Standard Error	t-value	p-value
Horas	0,0055	0,000	26,198	< 0,03
Bactéria	-0,1610	0,013	-12,393	< 0,03
Fibra	-0,0903	0,011	-8,251	0,032
Intercept	0,4641			
Outras estatísticas				
F-Value			322,1	
Kurtosis			2,679	
R-Squared			0,625	
Adjusted R-Squared			0,623	

Fonte: Autoria própria.

É possível perceber com o auxílio da Tabela 3 e da Equação 7 que, similarmente, a Regressão Linear Múltipla não será útil para prever a absorção capilar com boa precisão. Observa-se ainda que a absorção capilar de água tem correlação indireta com o uso de

bactérias e fibras. Ou seja, na presença de bactéria e fibra, a absorção capilar de água nas amostras tende a diminuir, o que representa um resultado satisfatório para o concreto e está em concordância com os estudos da literatura (FENG; SU; QIAN, 2019; SU et al., 2021).

5.2 Support Vector Machine

Neste estudo, modelos SVM foram treinados e testados com funções de kernel gaussianas (RBF), polinomiais e sigmóides e uma iteração máxima de 10.000. Os parâmetros C, gama, epsilon e coeficientes (coef0) para cada modelo são mostrados na Tabela 4.

Tabela 4 - Parâmetros para cada modelo de SVM.

Kernel	C	Gamma	coef0	epsilon
RBF	1	0,1	-	0,1
Sigmóide	0,5	0,01	0,05	-
Polinomial	0,5	0,1	0,15	-

Fonte: Autoria própria.

5.2.1 Taxa de reparo da área fissurada

O desempenho dos modelos de SVM para análise e previsão da taxa de reparo da área fissurada está resumido na Tabela 5.

Tabela 5 - Performance dos modelos de SVM para taxa de reparo da área fissurada.

Kernel	R²		MAE		RMSE	
	<i>Treino</i>	<i>Teste</i>	<i>Treino</i>	<i>Teste</i>	<i>Treino</i>	<i>Teste</i>
RBF	0,916	0,927	0,004	0,053	0,004	0,004
Polinomial	0,815	0,792	0,073	0,079	0,009	0,011
Sigmóide	0,393	0,406	0,133	0,135	0,073	0,032

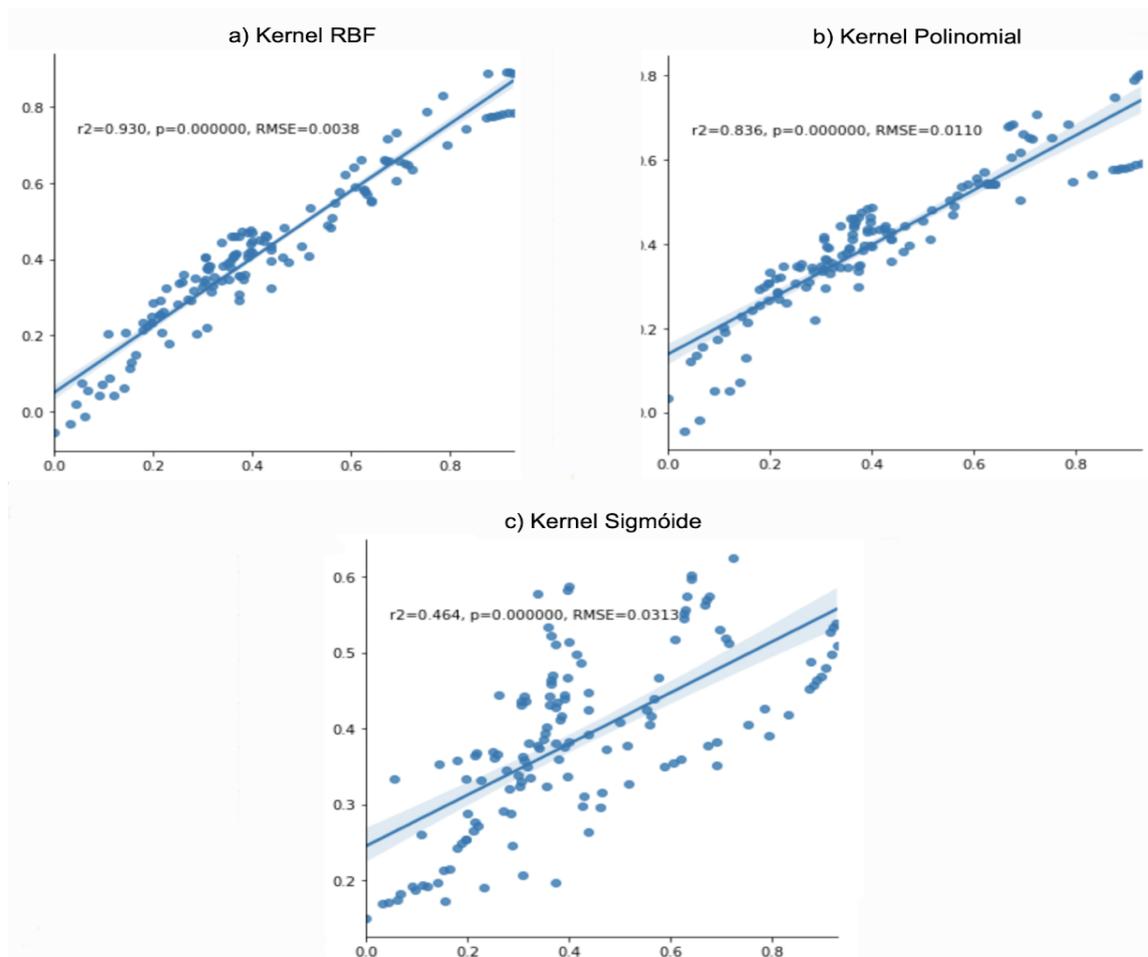
Fonte: Autoria própria.

Como visto na Tabela 5, o melhor modelo preditivo em relação a todas as métricas é o modelo RBF ($R^2 = 0,927$, $MAE = 0,053$ e $RMSE = 0,004$), enquanto o pior modelo preditivo

é o modelo Sigmóide ($R^2 = 0,406$, $MAE = 0,135$ e $RMSE = 0,032$). Curiosamente, os valores de R^2 para o conjunto de teste são ligeiramente maiores do que para o conjunto de treinamento, tanto para o melhor quanto para o pior modelo. As métricas para o Sigmoid também mostram que este modelo possui problemas de underfitting, ou seja, este modelo não é ajustado nem para o conjunto de treinamento e não é capaz de prever valores corretos para a taxa da área de reparo.

Quando ilustrado com o auxílio de gráficos para o conjunto de teste (Figura 4), quanto mais próximos os pontos estiverem da linha de regressão, melhor o ajuste e mais preciso é o modelo. A Figura 4 mostra que os pontos para o modelo sigmóide são muito dispersos, enquanto para os modelos RBF e Polinomial ($R^2 = 0,792$, $MAE = 0,079$ e $RMSE = 0,011$) os pontos estão mais próximos da linha de regressão.

Figura 4 - Performance dos modelos de SVM na previsão da taxa de reparo da área fissurada utilizando a) Kernel RBF b) Kernel Polinomial c) Kernel Sigmóide.



Fonte: Autoria própria.

5.2.1 Absorção capilar de água

O desempenho dos modelos de SVM para análise e previsão da absorção capilar de água está resumido na Tabela 6.

Tabela 6 - Performance dos modelos de SVM para absorção capilar de água.

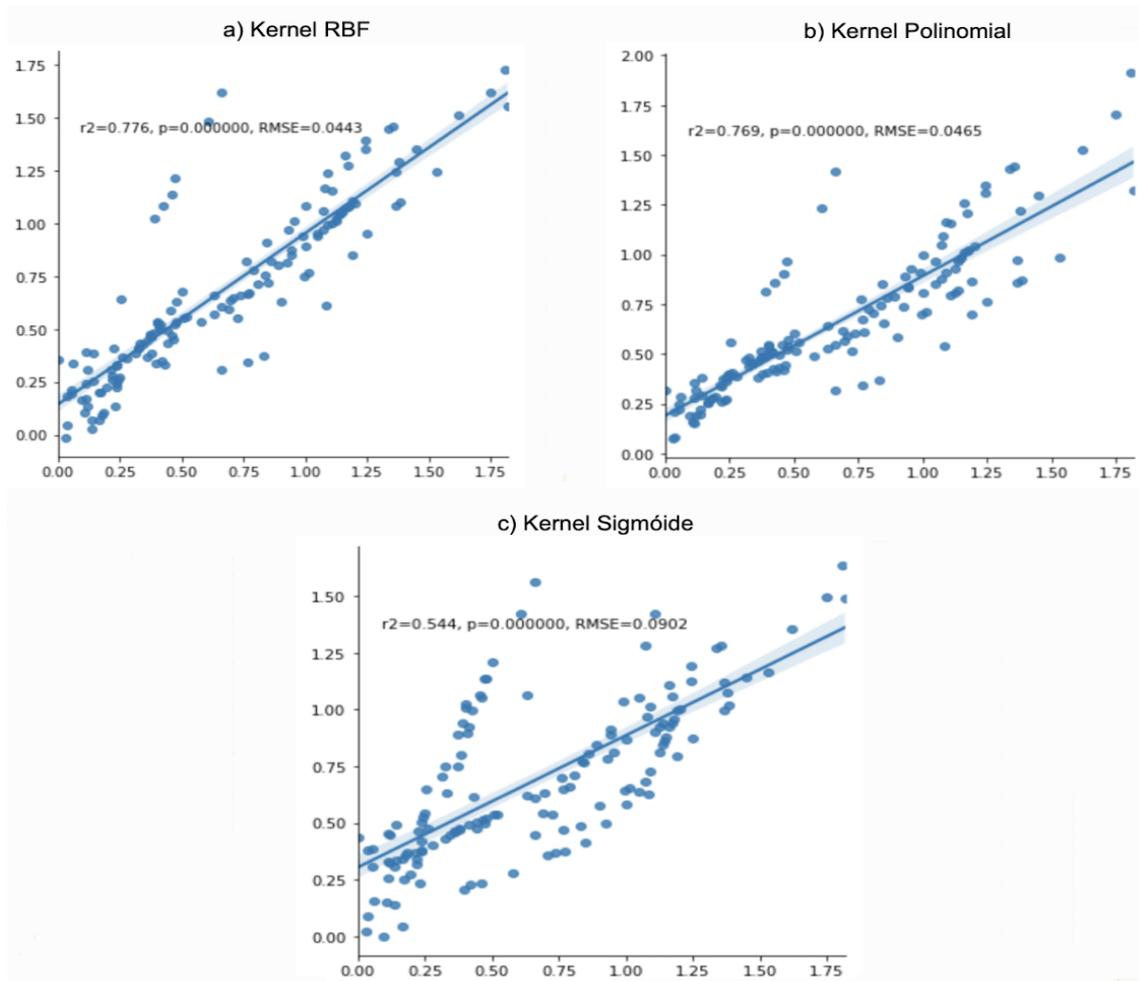
Kernel	R ²		MAE		RMSE	
	<i>Treino</i>	<i>Teste</i>	<i>Treino</i>	<i>Teste</i>	<i>Treino</i>	<i>Teste</i>
RBF	0,796	0,773	0,042	0,139	0,042	0,044
Polinomial	0,780	0,762	0,156	0,161	0,046	0,046
Sigmóide	0,616	0,538	0,224	0,242	0,080	0,090

Fonte: Autoria própria.

A partir da Tabela 6, nota-se que novamente o melhor modelo preditivo em relação a todas as métricas é o modelo RBF ($R^2 = 0,773$, $MAE = 0,139$ e $RMSE = 0,044$), enquanto o pior modelo preditivo é o modelo Sigmóide ($R^2 = 0,538$, $MAE = 0,224$ e $RMSE = 0,090$).

A Figura 5 mostra os pontos para o modelo sigmóide ainda mais dispersos, enquanto para os modelos RBF e Polinomial ($R^2 = 0,762$, $MAE = 0,161$ e $RMSE = 0,046$) estão mais próximos da linha de regressão, apesar de ainda apresentarem grande quantidade de *outliers* – pontos significativamente diferentes do restante do conjunto de dados.

Figura 5 - Performance dos modelos de SVM na previsão da absorção capilar de água utilizando a) Kernel RBF b) Kernel Polinomial c) Kernel Sigmóide.



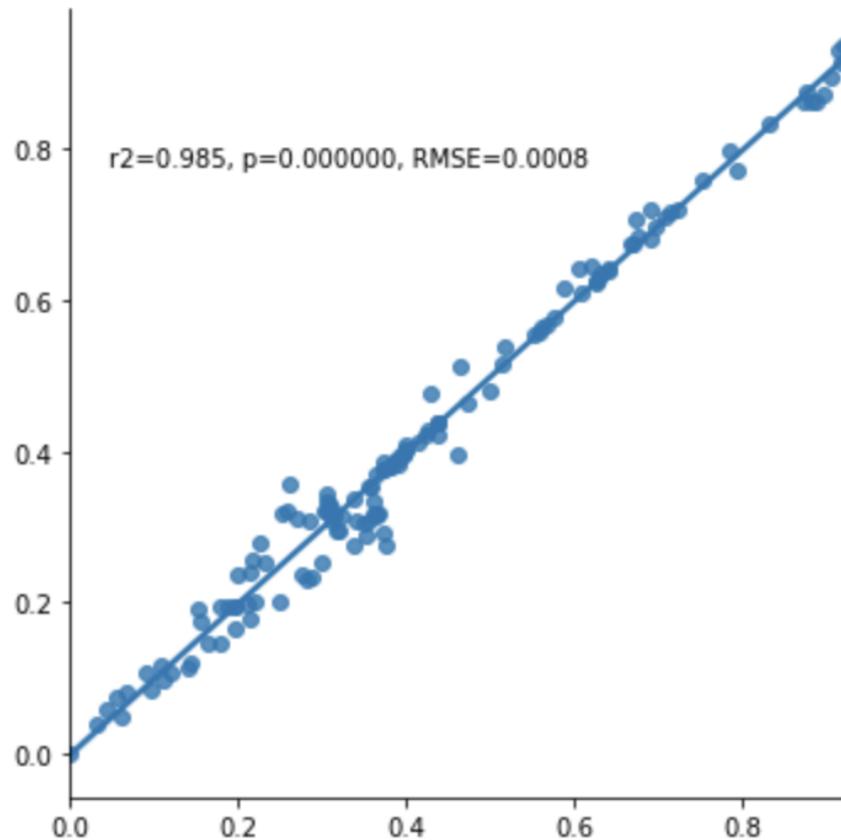
Fonte: Autoria própria.

5.3 Random Forest Regressor

5.3.1 Taxa de reparo da área fissurada

Random Forest Regressor apresentou o seguinte desempenho na previsão da taxa de reparo da área fissurada: $R^2 = 0,995$, $MAE = 0,000$, $RMSE = 0,000$ durante o treinamento e $R^2 = 0,985$, $MAE = 0,019$, $RMSE = 0,000$ durante o teste. A Figura 6 mostra que o modelo utilizando Random Forest Regressor supera os outros modelos preditivos durante o teste na análise e previsão da taxa de reparo. Os pontos representando as previsões feitas pelo modelo estão bem alinhados com a reta de regressão, indicando um modelo bem ajustado.

Figura 6 - Performance do Random Forest Regressor na previsão da taxa de reparo da área fissurada.

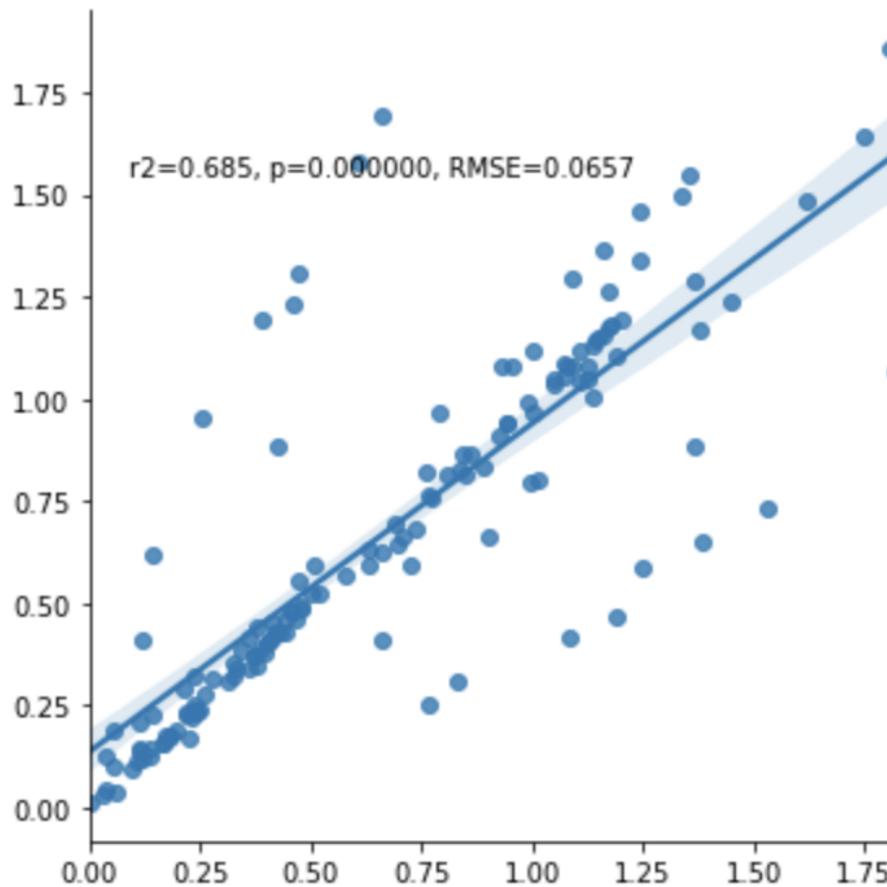


Fonte: Autoria própria.

5.3.2 Absorção capilar de água

Na análise e previsão da absorção capilar de água, o Random Forest Regressor apresentou o seguinte desempenho: $R^2 = 0,866$, $MAE = 0,028$, $RMSE = 0,028$ durante o treinamento e $R^2 = 0,685$, $MAE = 0,124$, $RMSE = 0,066$ durante o teste. A Figura 7 mostra o desempenho do modelo, que por sua vez se mostra menos eficiente na previsão de absorção capilar de água quando comparado ao Support Vector Machine.

Figura 7 - Performance do Random Forest Regressor na predição da absorção capilar de água.



Fonte: Autoria própria.

5.4 Comparação da capacidade de previsão dos modelos de ML desenvolvidos

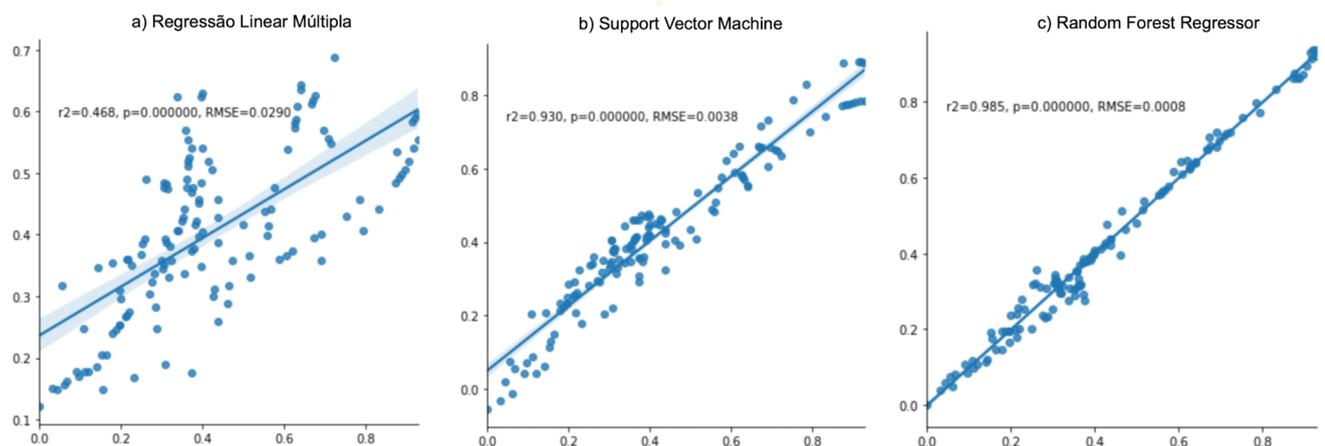
Comparando os resultados de todos os modelos preditivos, a Regressão Linear Múltipla apresentou o pior desempenho, se mostrando ineficaz para o estudo em questão. Support Vector Machine e Random Forest apresentaram modelos preditivos melhores.

O desempenho dos modelos SVM dependem da escolha do kernel e do ajuste dos hiperparâmetros, sendo o que o Kernel RBF mostrou melhor desempenho tanto para os dados de treinamento relacionados a taxa de reparo da área fissurada, quanto para os dados de treinamento relacionados a absorção capilar de água.

O regressor de RF é o modelo ideal para prever a taxa de reparo de fissuras (Figura 8), uma vez que suas métricas de validação (R^2 , MAE e RMSE) são melhores do que as mesmas

métricas para todos os outros modelos, supera até mesmo o SVM, conhecido por ter desempenho satisfatório em conjuntos de dados pequenos e médios (YAO et al., 2004; BARBOSA-DA-SILVA; STEFANI, 2013). Este resultado é semelhante a outros resultados encontrados na literatura, onde o regressor de RF e os regressores de árvore de decisão são melhores do que outras técnicas de ML para prever as propriedades de auto cicatrização do concreto (HUANG et al., 2022; ZHUANG; ZHOU, 2019). Como Random Forest é um método ensemble que representa um conjunto de árvores de decisão, ele pode gerar e balancear aprendizes fortes e fracos (árvores) aumentando o desempenho final (ALABDULJABBAR et al., 2023; AI et al., 2019), explicando o motivo do maior desempenho para o modelo preditivo de RF neste estudo.

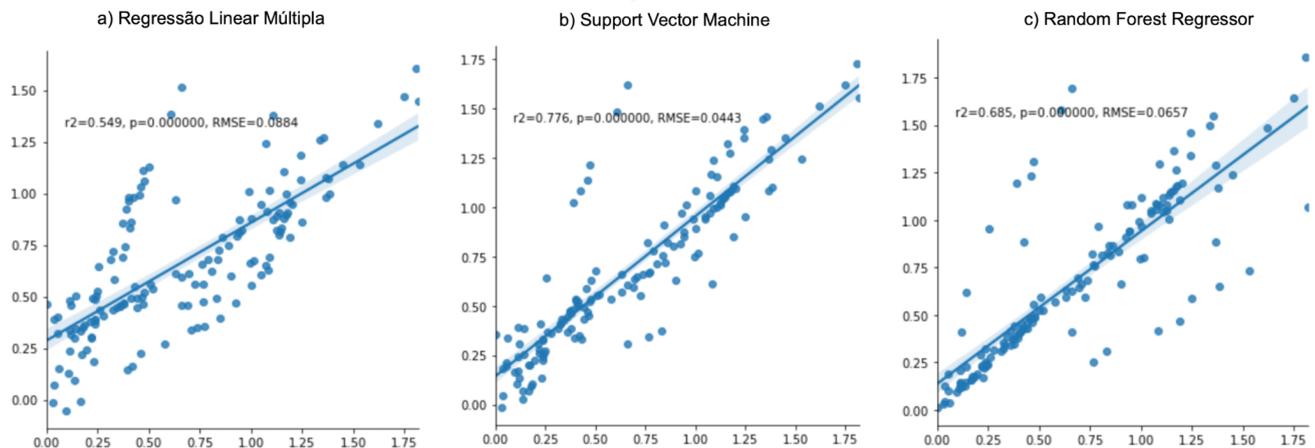
Figura 8 - Performance dos modelos de (a) Regressão Linear Múltipla, (b) Support Vector Machine e (c) Random Forest Regressor na previsão da taxa de reparo da área fissurada.



Fonte: Autoria própria.

Na análise e previsão da absorção capilar de água, o SVM mostrou o melhor desempenho dentre os outros algoritmos, como mostra a Figura 9. Entretanto, todos os modelos de previsão da absorção capilar de água se mostram menos ajustados e precisos quando comparado com o desempenho dos modelos na análise e previsão da taxa de reparo da área fissurada. Ainda que, na previsão da taxa de reparo, o modelo de SVM ($R^2 = 0,927$, $MAE = 0,053$ e $RMSE = 0,004$) tenha apresentado um desempenho inferior ao modelo de RF ($R^2 = 0,995$, $MAE = 0,000$, $RMSE = 0,000$), ele apresenta um desempenho significativamente melhor do que quando aplicado aos dados de treinamento relacionados a absorção capilar de água ($R^2 = 0,773$, $MAE = 0,139$ e $RMSE = 0,044$).

Figura 9 - Performance dos modelos de (a) Regressão Linear Múltipla, (b) Support Vector Machine e (c) Random Forest Regressor na previsão da absorção capilar de água.



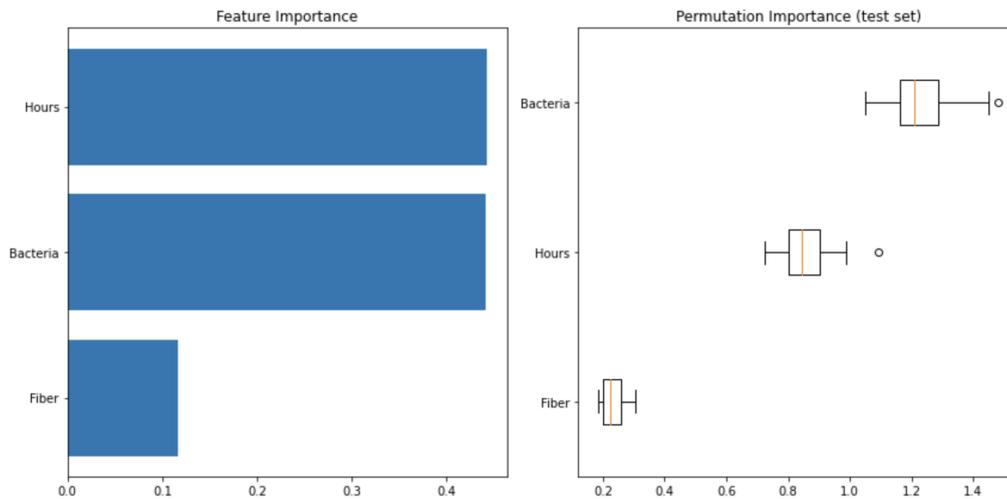
Fonte: Autoria própria.

5.5 Influência das variáveis independentes

Algoritmos de ML como RF e SVM são muitas vezes referidos como caixas pretas, ou seja, não é possível inferir a relação entre as variáveis de entrada (independentes) e de saída (dependentes), ao contrário do que costuma ser feito em modelos estatísticos. No entanto, é um mito que as relações entre as variáveis e como os modelos de ML funcionam não podem ser explicadas (RUDIN, 2019). Assim, modelos de ML como SVM e RF podem ser interpretáveis e usados como ferramenta para entender os fenômenos investigados e descobrir a influência de cada variável na fissuração final.

Assim, para determinar a influência de cada componente (variável) da argamassa na taxa de reparo da área fissurada e na absorção capilar de água, utilizou-se a função `permutation_importance` disponível no pacote `sklearn.inspection`. Esta função embaralha aleatoriamente os valores de uma entrada selecionada, enquanto mantém outras entradas inalteradas e faz uma previsão. A diferença entre a precisão de duas previsões (alterada e inalterada) é definida como a importância de cada variável. Neste trabalho, a função `permutation_importance` foi executada com os seguintes parâmetros: `n_repeats=10`, `random_state=42`, `n_jobs=2`. As Figuras 10 e 11 apresentam um gráfico com a importância de cada variável.

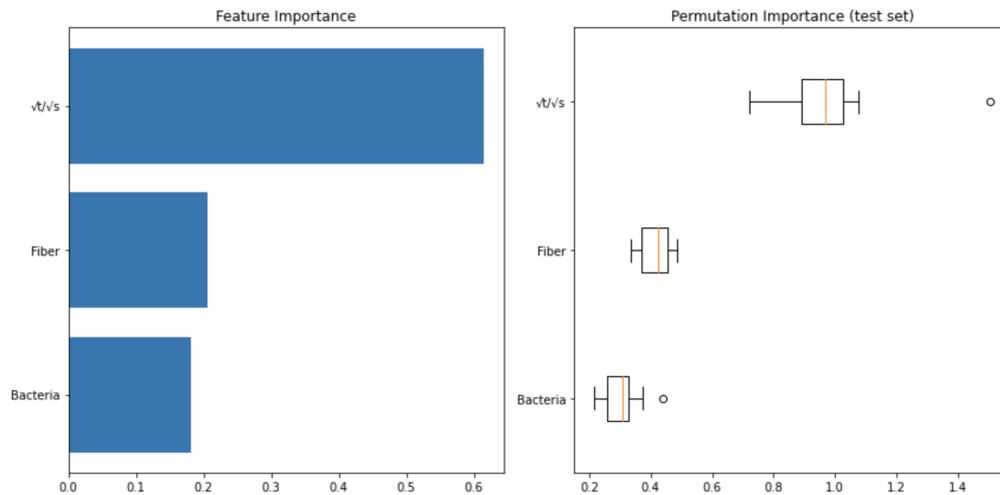
Figura 10 - Importância das variáveis independentes relativas a taxa de reparo da área fissurada nos modelos finais de SVM e RF.



Fonte: Autoria própria.

A Figura 10 mostra que o tempo (horas) tem grande influência na taxa de reparo da fissura, tendo uma influência de 44,20%. A presença de bactérias também têm grande influência (44,08%), enquanto a presença e o tipo de fibra (11,71%) não influenciam tanto na taxa de reparo. Como as bactérias aceleram a taxa de reparo precipitando o carbonato de cálcio (FENG et al., 2021), as fibras podem ajudar na distribuição do carbonato de cálcio nas rachaduras (FENG; SU; QIAN, 2019), esses resultados mostram que a presença de bactérias é um ingrediente principal na reparação das fissuras do concreto, enquanto as fibras são auxiliares e até não obrigatórias. Esses achados estão de acordo com a literatura (ZHUANG; ZHOU, 2019; SU et al., 2021; ALTHOEY et al., 2022).

Figura 11 - Importância das variáveis independentes relativas a taxa de reparo da área fissurada nos modelos finais de SVM e RF.



Fonte: Autoria própria.

A Figura 11 mostra que o tempo (\sqrt{t}/\sqrt{s}) tem grande influência na absorção capilar de água, seguido pela utilização de fibras que apresentam uma influência ligeiramente maior que a utilização de bactérias, isso se deve ao fato de a presença de fibras sintéticas como o álcool polivinílico (PVA) e o polipropileno (PP) tendem a reduzir a permeabilidade à água devido à precipitação do carbonato de cálcio em torno das fibras (FENG; SU; QIAN, 2019; NISHIWAKI et al., 2012; MANVITH; RAMESH; MACRIN, 2020).

6. CONCLUSÕES

Neste estudo, modelos de Machine Learning utilizando Regressão Linear Múltipla, Support Vector Machine e Random Forest Regressor foram desenvolvidos com o objetivo de analisar e prever a taxa de reparo da área fissurada e a absorção capilar de água em concretos auto-reparáveis modificados com bactéria e fibra. Os modelos de SVM com Kernel RBF e RF obtiveram melhor desempenho, enquanto a Regressão Linear Múltipla mostrou-se ineficaz para realizar as previsões no estudo em questão.

Os algoritmos de SVM e RF foram precisos na previsão da taxa de reparo da área fissurada do concreto auto-reparável – em especial o modelo de RF – apresentando desempenhos significativamente maiores quando comparado à previsão da absorção capilar de água.

Como o regressor de RF tem alta precisão, ele pode ser implantado como um modelo de ML para prever a área de reparo para concretos auto-reparáveis. Conseqüentemente, os modelos de Random Forest Regressor e até mesmo Support Vector Machine com Kernel RBF podem ser aplicados para desenvolver e validar formulações de concreto auto-regenerativo de alto desempenho baseadas em fibras poliméricas e bactérias.

Além disso, a análise da influência das variáveis independentes dos conjuntos de treinamento foram capazes de confirmar que as bactérias possuem um impacto significativo na reparação de fissuras e que a utilização de fibras ajuda a reduzir a permeabilidade geral à água em concretos auto-reparáveis.

7. TRABALHOS FUTUROS

Para um melhor entendimento sobre o uso de Machine Learning na análise e previsão de propriedades de concretos auto-reparáveis sugere-se:

- A implementação de conjuntos de dados relacionados a diferentes propriedades de concretos auto-reparáveis, como por exemplo, propriedades mecânicas;
- Melhorar o pré-processamento do conjunto de dados relacionados à absorção capilar de água para reduzir o aparecimento de outliers;
- Utilizar ferramentas de aprimoramento de parâmetros nos modelos de Machine Learning desenvolvidos.
- Testar a eficácia de diferentes algoritmos de Machine Learning na previsão de propriedades de concretos auto-reparáveis;

8. REFERÊNCIAS

AI, Haixin; WU, Xuewei; ZHANG, Li; *et al.* QSAR modelling study of the bioconcentration factor and toxicity of organic compounds to aquatic organisms using machine learning and ensemble methods. **Ecotoxicology and Environmental Safety**, v. 179, p. 71–78, 2019.

ALABDULJABBAR, Hisham; KHAN, Kaffayatullah; AWAN, Hamad Hassan; *et al.* Modeling the capacity of engineered cementitious composites for self-healing using AI-based ensemble techniques. **Case Studies in Construction Materials**, v. 18, p. e01805, 2023.

ALTHOEY, Fadi; AMIN, Muhammad Nasir; KHAN, Kaffayatullah; *et al.* Machine learning based computational approach for crack width detection of self-healing concrete. **Case Studies in Construction Materials**, v. 17, p. e01610, 2022.

BARBOSA-DA-SILVA, Rogério; STEFANI, Ricardo. QSPR based on support vector machines to predict the glass transition temperature of compounds used in manufacturing OLEDs. **Molecular Simulation**, v. 39, n. 3, p. 234–244, 2013.

BEN CHAABENE, Wassim; FLAH, Majdi; NEHDI, Moncef L. Machine learning prediction of mechanical properties of concrete: Critical review. **Construction and Building Materials**, v. 260, p. 119889, 2020.

BEN-HUR, Asa; ONG, Cheng Soon; SONNENBURG, Sören; *et al.* Support Vector Machines and Kernels for Computational Biology. **PLoS Computational Biology**, v. 4, n. 10, p. e1000173, 2008.

BREIMAN, Leo. **Classification and Regression Trees Regression Trees**. [s.l.: s.n.], 1984.

CASCARDI, Alessio; MICELLI, Francesco; AIELLO, Maria Antonietta. An Artificial Neural Networks model for the prediction of the compressive strength of FRP-confined concrete circular columns. **Engineering Structures**, v. 140, p. 199–208, 2017.

CHA, Young-Jin; CHOI, Wooram; BÜYÜKÖZTÜRK, Oral. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. **Computer-Aided Civil and Infrastructure Engineering**, v. 32, n. 5, p. 361–378, 2017.

DALZOCHIO, Jovani; KUNST, Rafael; PIGNATON, Edison; et al. Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. *Computers in Industry*, v. 123, p. 103298, 2020.

DEKA, Paresh Chandra. *A Primer on Machine Learning Applications in Civil Engineering*. [s.l.]: **CRC Press**, 2019.

ELSISI, Mahmoud; MAHMOUD, Karar; LEHTONEN, Matti; et al. Reliable Industry 4.0 Based on Machine Learning and IoT for Analyzing, Monitoring, and Securing Smart Meters. **Sensors**, v. 21, n. 2, p. 487, 2021.

FENG, Jianhang; SU, Yilin; QIAN, Chunxiang. Coupled effect of PP fiber, PVA fiber and bacteria on self-healing efficiency of early-age cracks in concrete. **Construction and Building Materials**, v. 228, p. 116810, 2019.

FENG, Jinpeng; ZHANG, Haowei; GAO, Kang; et al. Efficient Creep Prediction of Recycled Aggregate Concrete Via Machine Learning Algorithms. **SSRN Electronic Journal**, 2022.

FENG, Jun; CHEN, Bingcheng; SUN, Weiwei; et al. Microbial induced calcium carbonate precipitation study using *Bacillus subtilis* with application to self-healing concrete preparation and characterization. **Construction and Building Materials**, v. 280, p. 122460, 2021.

GUI, Guoqing; PAN, Hong; LIN, Zhibin; et al. Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection. **KSCE Journal of Civil Engineering**, v. 21, n. 2, p. 523–534, 2017.

HARRIS, C.R., MILLMAN, K.J., VAN DER WALT, S.J. et al. *Array programming with NumPy*. **Nature** 585, 357–362, 2020. DOI: 10.1038/s41586-020-2649-2.

HUANG, Xu; GE, Jun; KAEWUNRUEN, Sakdirat; *et al.* The Self-Sealing Capacity of Environmentally Friendly, Highly Damped, Fibre-Reinforced Concrete. **Materials**, v. 13, n. 2, p. 298, 2020.

HUANG, Xu; SRESAKOOLCHAI, Jessada; QIN, Xia; *et al.* Self-Healing Performance Assessment of Bacterial-Based Concrete Using Machine Learning Approaches. **Materials**, v. 15, n. 13, p. 4436, 2022.

HUANG, Youqin; LI, Jiayong; FU, Jiyang. Review on Application of Artificial Intelligence in Civil Engineering. **Computer Modeling in Engineering & Sciences**, v. 121, n. 3, p. 845–875, 2019.

Inc. PT. Collaborative data science [Internet]. Montreal, QC: Plotly Technologies Inc. 2015. Available from: <https://plot.ly>

JAHED ARMAGHANI, Danial; ASTERIS, Panagiotis G.; ASKARIAN, Behnam; *et al.* Examining Hybrid and Single SVM Models with Different Kernels to Predict Rock Brittleness. **Sustainability**, v. 12, n. 6, p. 2229, 2020.

J. D. HUNTER, "Matplotlib: A 2D Graphics Environment", **Computing in Science & Engineering**, vol. 9, no. 3, pp. 90-95, 2007.

KHAMBRA, Geetanjli; SHUKLA, Prashant. Novel machine learning applications on fly ash based concrete: An overview. **Materials Today: Proceedings**, 2021.

KHAN, Kaffayatullah; AHMAD, Waqas; AMIN, Muhammad Nasir; *et al.* A Systematic Review of the Research Development on the Application of Machine Learning for Concrete. **Materials**, v. 15, n. 13, p. 4512, 2022.

KLUYVER, T. *et al.* Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt, eds. **Positioning and Power in Academic Publishing: Players, Agents and Agendas**. pp. 87–90, 2016.

KOTSIPOULOS, Thanasis; SARIGIANNIDIS, Panagiotis; IOANNIDIS, Dimosthenis; et al. Machine Learning and Deep Learning in smart manufacturing: The Smart Grid paradigm. **Computer Science Review**, v. 40, p. 100341, 2021.

KUMAR JOGI, Pavan; VARA LAKSHMI, T.V.S. Self healing concrete based on different bacteria: A review. **Materials Today: Proceedings**, v. 43, p. 1246–1252, 2021.

LEE, Yong-Soo; RYOU, Jae-Suk. Self healing behavior for crack closing of expansive agent via granulation/film coating method. **Construction and Building Materials**, v. 71, p. 188–193, 2014.

LI, Yongjian; ZHANG, Qizhi; KAMIŃSKI, Paweł; *et al.* Compressive Strength of Steel Fiber-Reinforced Concrete Employing Supervised Machine Learning Techniques. **Materials**, v. 15, n. 12, p. 4209, 2022.

LI, Zongjin; ZHOU, Xiangming; MA, Hongyan; *et al.* **Advanced concrete technology**. [s.l.]: John Wiley & Sons, 2022.

MANVITH KUMAR REDDY, C.; RAMESH, B.; MACRIN, Dannie. Effect of crystalline admixtures, polymers and fibers on self healing concrete - a review. **Materials Today: Proceedings**, v. 33, p. 763–770, 2020.

MAULUD, Dastan; ABDULAZEEZ, Adnan M. A Review on Linear Regression Comprehensive in Machine Learning. **Journal of Applied Science and Technology Trends**, v. 1, n. 4, p. 140–147, 2020.

MCKINNEY, W. & others. Data structures for statistical computing in python. **Proceedings of the 9th Python in Science Conference**. pp. 51–56, 2010.

NADERPOUR, Hosein; RAFIEAN, Amir Hossein; FAKHARIAN, Pouyan. Compressive strength prediction of environmentally friendly concrete using artificial neural networks. **Journal of Building Engineering**, v. 16, p. 213–219, 2018.

NETO, Tiago Ferreira Campos; CASCUDO, Oswaldo; CARASEK, Helena. Uso do aprendizado de máquina na previsão da frente de carbonatação: uma trajetória dos anos 90 aos dias atuais. **CONCRETO & Construções**, v. XLIX, n. 102, p. 76–83, 2021.

NISHIWAKI, Tomoya; KODA, Marina; YAMADA, Makoto; *et al.* Experimental Study on Self-Healing Capability of FRCC Using Different Types of Synthetic Fibers. **Journal of Advanced Concrete Technology**, v. 10, n. 6, p. 195–206, 2012.

NIU, Wen-Jing; FENG, Zhong-Kai; FENG, Bao-Fei; *et al.* Comparison of Multiple Linear Regression, Artificial Neural Network, Extreme Learning Machine, and Support Vector Machine in Deriving Operation Rule of Hydropower Reservoir. **Water**, v. 11, n. 1, p. 88, 2019.

PAN, Yue; ZHANG, Limao. Roles of artificial intelligence in construction engineering and management: A critical review and future trends. **Automation in Construction**, v. 122, p. 103517, 2021.

PEDREGOSA *et al.*, **Scikit-learn: Machine Learning in Python**, JMLR 12, pp. 2825-2830, 2011.

PRAKASH, Shikha; SHARMA, Animesh; SAHU, Sitanshu Shekhar. Soil Moisture Prediction Using Machine Learning. *In*: **2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)**. [s.l.]: IEEE, 2018. Disponível em: <<http://dx.doi.org/10.1109/icicct.2018.8473260>>. Acesso em: 17 Apr. 2023.

QIAN, Chunxiang; CHEN, Huaicheng; REN, Lifu; *et al.* Self-healing of early age cracks in cement-based materials by mineralization of carbonic anhydrase microorganism. **Frontiers in Microbiology**, v. 6, 2015.

RODRIGUEZ-GALIANO, V.; SANCHEZ-CASTILLO, M.; CHICA-OLMO, M.; *et al.* Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. **Ore Geology Reviews**, v. 71, p. 804–818, 2015.

RONG, Shen; BAO-WEN, Zhang. The research of regression model in machine learning field. **MATEC Web of Conferences**, v. 176, p. 01033, 2018.

RUDIN, Cynthia. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. **Nature Machine Intelligence**, v. 1, n. 5, p. 206–215, 2019.

SEBER, George A. F.; LEE, Alan J. **Linear Regression Analysis**. [s.l.]: John Wiley & Sons, 2012.

SUN, Lei; KOOPIALIPOOR, Mohammadreza; JAHED ARMAGHANI, Danial; *et al.* Applying a meta-heuristic algorithm to predict and optimize compressive strength of concrete samples. **Engineering with Computers**, v. 37, n. 2, p. 1133–1145, 2019.

SU, Yilin; QIAN, Chunxiang; RUI, Yafeng; *et al.* Exploring the coupled mechanism of fibers and bacteria on self-healing concrete from bacterial extracellular polymeric substances (EPS). **Cement and Concrete Composites**, v. 116, p. 103896, 2021.

SVM Hyperparameter Tuning using GridSearchCV. Velocity Business Solutions Limited.

Disponível em:

<<https://www.vebuso.com/2020/03/svm-hyperparameter-tuning-using-gridsearchcv/>>.

Acesso em: 23 Nov. 2022.

VADYALA, Shashank Reddy; BETGERI, Sai Nethra; MATTHEWS, John C.; *et al.* A review of physics-based machine learning in civil engineering. **Results in Engineering**, v. 13, p. 100316, 2022.

VIJAY, Kunamineni; MURMU, Meena; DEO, Shirish V. Bacteria based self healing concrete – A review. **Construction and Building Materials**, v. 152, p. 1008–1014, 2017.

VIRTANEN, P., GOMMERS, R., OLIPHANT, T.E. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nature Methods**, v. 17, p. 261–272, 2020.

WANG, Li'ai; ZHOU, Xudong; ZHU, Xinkai; *et al.* Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. **The Crop Journal**, v. 4, n. 3, p. 212–219, 2016.

WANG, J.Y.; SOENS, H.; VERSTRAETE, W.; *et al.* Self-healing concrete by use of microencapsulated bacterial spores. **Cement and Concrete Research**, v. 56, p. 139–152, 2014.

WASKOM, Michael. seaborn: statistical data visualization. **Journal of Open Source Software**, v. 6, n. 60, p. 3021, 2021.

YAN, Fei; LIN, Zhibin; WANG, Xingyu; *et al.* Evaluation and prediction of bond strength of GFRP-bar reinforced concrete using artificial neural network optimized with genetic algorithm. **Composite Structures**, v. 161, p. 441–452, 2017.

YAO, X. J.; PANAYE, A.; DOUCET, J. P.; *et al.* Comparative Study of QSAR/QSPR Correlations Using Support Vector Machines, Radial Basis Function Neural Networks, and Multiple Linear Regression. **Journal of Chemical Information and Computer Sciences**, v. 44, n. 4, p. 1257–1266, 2004.

ZHUANG, Xiaoying; ZHOU, Shuai. The Prediction of Self-Healing Capacity of Bacteria-Based Concrete Using Machine Learning Approaches. **Computers, Materials & Continua**, v. 59, n. 1, p. 57–77, 2019.

ZIOLKOWSKI, Patryk; NIEDOSTATKIEWICZ, Maciej. Machine Learning Techniques in Concrete Mix Design. **Materials**, v. 12, n. 8, p. 1256, 2019.