



UNIVERSIDADE DE PERNAMBUCO
ESCOLA POLITÉCNICA DE PERNAMBUCO
Programa de Pós-graduação em Engenharia Civil

MARÍLIA BEZERRA TENÓRIO CAVALCANTI

**ANÁLISE DE DADOS DE ACIDENTES DE TRABALHO DA
CONSTRUÇÃO COM AUXÍLIO DE APRENDIZADO DE MÁQUINA**

Recife, PE
2022



UNIVERSIDADE DE PERNAMBUCO
ESCOLA POLITÉCNICA DE PERNAMBUCO
Programa de Pós-graduação em Engenharia Civil

MARÍLIA BEZERRA TENÓRIO CAVALCANTI

**ANÁLISE DE DADOS DE ACIDENTES DE TRABALHO DA
CONSTRUÇÃO COM AUXÍLIO DE APRENDIZADO DE MÁQUINA**

Dissertação apresentada ao Curso de Pós-graduação em Engenharia Civil, da Escola Politécnica de Pernambuco da Universidade de Pernambuco para obtenção do título de Mestre em Engenharia.

Área de Concentração: Construção Civil

Orientador: Profa. Dra. Bianca M. Vasconcelos

Recife, PE
2022

Dados Internacionais de Catalogação-na-Publicação (CIP)

Universidade de Pernambuco – Recife

C376a Cavalcanti, Marília Bezerra Tenório
Análise de dados de acidentes de trabalho da construção
com auxílio de aprendizado de máquina. / Marília Bezerra
Tenório Cavalcanti. – Recife: UPE, Escola Politécnica,
2022.

132 f.: il.

Orientadora: Profa. Dra. Bianca M. Vasconcelos

Dissertação (Mestrado - Construção Civil) Universidade
de Pernambuco, Escola Politécnica, Programa de Pós-
Graduação em Engenharia Civil, 2022.

1. Análise de Dados. 2. Acidentes de Trabalho. 3.
Construção Civil. 4. Machine Learning. I. Engenharia Civil
– Dissertação. II. Vasconcelos, Bianca M. (orient.). III.
Universidade de Pernambuco, Escola Politécnica, Mestrado
em Construção Civil. IV. Título.

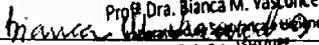
CDD: 690.028

MARÍLIA BEZERRA TENÓRIO CAVALCANTI

**ANÁLISE DE DADOS DE ACIDENTES DE TRABALHO DA
CONSTRUÇÃO COM AUXÍLIO DE APRENDIZADO DE MÁQUINA**

BANCA EXAMINADORA:

Orientadora:


Prof.^a Dra. Bianca M. Vasconcelos
do Trabalho LS117/011
Prof.^a Dr.^a Bianca M. Vasconcelos

Prof.^a Dr.^a Bianca Vasconcelos
Universidade de Pernambuco

Examinadores:



Prof.^a Dr.^a Emilia Rabbani
Universidade de Pernambuco



Prof. Dr. Fabiano Corrêa
Universidade de São Paulo

Recife, PE
2022

DEDICATÓRIA

Dedico esta pesquisa às famílias que sofreram as consequências de um acidente de trabalho.

AGRADECIMENTOS

À Deus, pela minha existência e por dar sentido a tudo.

Aos meus pais, tia, avó e irmã, que sempre me deram amor e apoio incondicional em todos os aspectos da minha vida, sem vocês nada disso seria possível.

Ao meu marido, por todo amor, cuidado, incentivo e parceria de sempre.

Aos amigos, especialmente Luciano, que acreditou e participou intensamente deste projeto desde o início.

Às amigas de infância, pelo incentivo e carinho constante.

Aos professores Dr. Béda Barkokébas Junior (*in memoriam*) e Dr.^a Bianca Vasconcelos, por todo o conhecimento passado, confiança depositada, apoio nos momentos difíceis e pelas orientações, muito além do escopo do mestrado, que me proporcionaram crescer como pessoa e profissionalmente.

Aos amigos do mestrado, por terem compartilhado comigo as dores e vitórias dessa jornada e pelo apoio intenso durante esses últimos anos.

Aos membros da banca, pelas considerações pertinentes para a melhoria deste trabalho.

À UPE, POLI, PEC e LSHT pela oportunidade de poder desenvolver esta pesquisa.

Não me atrevera a dizer que agradei a todos que colaboraram direta ou indiretamente na construção deste trabalho, seria injusto esquecer de alguém. Apenas espero ter a chance de retribuir todo o conhecimento, acolhimento e carinho que me foi repassado.

“Digo o que penso, com esperança.
Penso no que faço, com fé.
Faço o que devo fazer, com amor.
Eu me esforço para ser cada dia melhor,
pois bondade também se aprende.”

Cora Coralina

RESUMO

O alto número de acidentes de trabalho no setor da construção evidencia que a gestão de segurança das obras necessita de melhorias. Nesse contexto, conhecer os fatores gerais relacionados à ocorrência desses eventos é essencial para traçar planos estratégicos específicos com o objetivo de eliminar ou diminuir os riscos nos locais de trabalho. Dessa forma, este trabalho teve o objetivo de analisar dados de acidentes de trabalho da construção com auxílio de técnicas de aprendizado de máquina para identificar padrões de casos. Inicialmente, foi realizada uma Revisão Sistemática da Literatura – RSL a respeito do tema para identificar os principais objetivos e características dos modelos de *Machine Learning* – ML desenvolvidos no mundo. Na segunda etapa do trabalho, foram analisados, por meio de ML, 50.380 dados oficiais de acidentes e doenças ocupacionais do setor da construção, ocorridos no Brasil entre 2018 e 2021 e registrados em Comunicações de Acidente de Trabalho – CAT. Os resultados da revisão mostraram a ausência do Brasil no desenvolvimento de pesquisas relacionadas ao tema. Em relação à análise dos dados das CAT, foram identificados os tipos de acidentes, atividades econômicas, natureza das lesões, partes do corpo atingidas e o perfil do profissional acidentado, com informações sobre a função, idade e gênero. Além disso, a aplicação dos dois algoritmos de *clustering*, *K-Modes* e *K-Prototype*, utilizados para obter grupos prioritários de acidentes relacionados à construção de edifícios, não se mostrou válida para a base de dados da CAT. Assim, foram sugeridas melhorias para o modelo e *insights* para próximos estudos. Por fim, este trabalho tem o potencial de contribuir com a incorporação de tecnologias no gerenciamento da saúde e segurança do trabalho do setor da construção e, conseqüentemente, com a redução de acidentes de trabalho.

Palavras-chave: Análise de dados; Acidentes de Trabalho; Construção Civil; *Machine Learning*.

ABSTRACT

The high number of work-related accidents in construction highlights the necessity to improve safety management in the work sites. In this context, knowing the general factors related to these events is essential to develop strategic plans to eliminate or reduce the risks of the activities. Thus, this research aimed to analyze construction accident data with Machine Learning - ML techniques to identify accident patterns. Initially, a literature review identified the main goals and characteristics of the ML models developed worldwide. Then, the research analyzed 50.380 construction occupational accidents and diseases official data in Brazil between 2018 and 2021 registered in occupational accident communication, known as CAT. The review results showed the lack of Brazil in related research. Also, the analysis results showed general information about the accidents, such as the kinds, economic activities, injury types, parts of the human body affected, and the victim profile with information related to the profession, age, and genre. Furthermore, the application of K-Modes and K-Prototype clustering algorithms, which were used to obtain accident priority groups regarding building construction, proved invalid for the CAT database. Thus, improvements to the model and insights for future studies were suggested. Finally, this paper can contribute to incorporating technologies in construction health and safety management and reducing occupational accidents.

Keywords: Data Analysis. Occupational Accidents. Construction. Machine Learning.

LISTA DE FIGURAS

Figura 2.1 –	Inteligência Artificial e as abordagens de Machine Learning	30
Figura 2.2 –	Classificação de algoritmos de ML de aprendizagem supervisionada e não supervisionada	32
Figura 3.1 –	Metodologia da Revisão Sistemática da Literatura	39
Figura 3.2 –	Fluxograma da aplicação de <i>Machine Learning</i>	43
Figura 4.1 –	Fluxograma PRISMA	50
Figura 4.2 –	Número de publicações identificadas por base de dados	51
Figura 4.3 –	Periódicos com maior número de publicações	52
Figura 4.4 –	Evolução anual das publicações	53
Figura 4.5 –	Relação entre autores, instituições e países/regiões	54
Figura 4.6 –	Autores com maior número de publicações	54
Figura 4.7 –	Instituições com maior número de publicações	55
Figura 4.8 –	Nuvem de palavras com palavras-chave dos estudos	56
Figura 4.9 –	Métodos de ML	58
Figura 4.10 –	Tipos de dados	59
Figura 4.11 –	Quantidade de dados	60
Figura 4.12 –	Métodos de coleta de dados	61
Figura 4.13 –	Abordagens dos dados	62
Figura 4.14 –	<i>Dataset</i> inicial	62
Figura 4.15 –	<i>Dataset</i> final	66
Figura 4.16 –	Tipos de acidentes	67
Figura 4.17 –	Percentual das atividades nos acidentes típicos de acordo com CNAE	68
Figura 4.18 –	Percentual dos agentes causadores dos acidentes típicos	70
Figura 4.19 –	As 10 naturezas das lesões mais comuns nos acidentes típicos	71
Figura 4.20 –	As 10 PCA mais recorrentes nos acidentes típicos	72
Figura 4.21 –	Percentual do gênero de trabalhadores acidentados (acidentes típicos)	73
Figura 4.22 –	Distribuição de acidentes típicos por idade do profissional acidentado	73
Figura 4.23 –	Distribuição de acidentes típicos por dia da semana	74
Figura 4.24 –	Distribuição de acidentes típicos por meses e anos	74
Figura 4.25 –	Percentual das atividades nos acidentes de trajeto de acordo com CNAE	75
Figura 4.26 –	Percentual dos agentes causadores dos acidentes de trajeto	77
Figura 4.27 –	As 10 naturezas das lesões mais comuns nos acidentes de trajeto	77
Figura 4.28 –	As 10 PCA mais recorrentes nos acidentes de trajeto	78
Figura 4.29 –	Percentual do gênero de trabalhadores acidentados (trajeto)	79
Figura 4.30 –	Distribuição de acidentes de trajeto por idade do acidentado	80
Figura 4.31 –	Distribuição de acidentes de trajeto por dia da semana	80
Figura 4.32 –	Distribuição de acidentes de trajeto por meses e anos	81
Figura 4.33 –	Percentual das atividades nas doenças de acordo com CNAE	82
Figura 4.34 –	As 10 naturezas das lesões mais comuns nas doenças ocupacionais	84
Figura 4.35 –	As 10 PCA mais recorrentes nas doenças ocupacionais	85
Figura 4.36 –	Percentual do gênero de trabalhadores acidentados (doenças)	86
Figura 4.37 –	Distribuição de doenças ocupacionais por idade do acidentado	86
Figura 4.38 –	Distribuição de doenças ocupacionais por meses e anos	87
Figura 4.39 –	Percentual de fatalidades	87

Figura 4.40 – Percentual de acidentes fatais de acordo com CNAE	88
Figura 4.41 – Taxa de fatalidade de acordo com CNAE	89
Figura 4.42 – As 10 PCA mais recorrentes nos acidentes fatais	94
Figura 4.43 – Ranking das 10 PCA com maior taxa de fatalidade	94
Figura 4.44 – Percentual do gênero de trabalhadores acidentados (acidentes fatais)	96
Figura 4.45 – Distribuição de acidentes fatais por idade do profissional acidentado	96
Figura 4.46 – Distribuição de acidentes fatais por dia da semana	97
Figura 4.47 – Distribuição de acidentes fatais por meses e anos	98
Figura 4.48 – Atributos selecionados para aplicação do <i>K-Modes</i>	99
Figura 4.49 – Atributos selecionados para aplicação do <i>K-Prototype</i>	99
Figura 4.50 – Custo do K-Modes em função do número de <i>clusters</i> para os acidentes de trabalho	100
Figura 4.51 – Custo do K-Prototype em função do número de <i>clusters</i> para os acidentes de trabalho	101
Figura 4.52 – Gráfico de dispersão CBO, natureza da lesão e parte do corpo atingida	102
Figura 4.53 – Gráfico de dispersão CBO, natureza da lesão e idade	102
Figura 4.54 – Gráfico de dispersão CBO, natureza da lesão e agente causador	102
Figura 4.55 – Gráfico de dispersão CBO, parte do corpo atingida e idade	103
Figura 4.56 – Gráfico de dispersão CBO, parte do corpo atingida e agente causador	103
Figura 4.57 – Gráfico de dispersão natureza da lesão, parte do corpo atingida e idade	103
Figura 4.58 – Gráfico de dispersão natureza da lesão, parte do corpo atingida e agente causador	104
Figura 4.59 – Gráfico de dispersão parte do corpo atingida, idade e agente causador	104
Figura 4.60 – Gráfico de dispersão idade, agente causador e CBO	104
Figura 4.61 – Gráfico de dispersão idade, agente causador e natureza da lesão	105
Figura 4.62 – Histograma da idade por <i>cluster</i>	105
Figura 4.63 – Histograma da natureza da lesão por <i>cluster</i>	106
Figura 4.64 – Histograma da parte do corpo por <i>cluster</i>	107
Figura 4.65 – Histograma do agente causador por <i>cluster</i>	108
Figura 4.66 – Histograma do CBO por <i>cluster</i>	109

LISTA DE QUADROS

Quadro 2.1 –	Abordagens tecnológicas da Indústria 4.0	29
Quadro 4.1 –	Bases de dados utilizadas na RSL	47
Quadro 4.2 –	Protocolo de pesquisa	48
Quadro 4.3 –	Métodos de ML	57
Quadro 4.4 –	Atributos iniciais das CATs	63
Quadro 4.5 –	Lista de códigos CNAEs referentes à Construção Civil	64
Quadro 4.6 –	Atributos dos dados após pré-processamento	66
Quadro 4.7 –	As 10 funções mais afetadas nos acidentes típicos de acordo com CBO ...	69
Quadro 4.8 –	As 5 CID mais comuns nos acidentes típicos	72
Quadro 4.9 –	As 10 funções mais afetadas nos acidentes de trajeto de acordo com CBO	76
Quadro 4.10 –	As 5 CID mais comuns nos acidentes de trajeto	79
Quadro 4.11 –	As 10 funções mais afetadas nas doenças ocupacionais de acordo com CBO	83
Quadro 4.12 –	Os 10 agentes causadores de doenças ocupacionais mais recorrentes	83
Quadro 4.13 –	As 10 CID mais comuns nas doenças ocupacionais	85
Quadro 4.14 –	As 10 funções mais afetadas nos acidentes fatais, de acordo com CBO	90
Quadro 4.15 –	Ranking das 10 funções com maior taxa de fatalidade, de acordo com CBO	90
Quadro 4.16 –	Os 10 agentes causadores mais recorrentes nos acidentes fatais	91
Quadro 4.17 –	Os 10 agentes causadores mais letais	92
Quadro 4.18 –	Total de mortes e taxa de fatalidade a respeito da natureza das lesões	93
Quadro 4.19 –	Quantidade de acidentes e taxa de fatalidade, de acordo com CID	95

LISTA DE ABREVIATURAS E SIGLAS

ACM – Association for Computing Machinery
AED – Análise Exploratória dos Dados
ANFIS – Adaptive Neuro-Fuzzy Inference System
ANN – Artificial Neural Network
ARM – Association Rule Mining
ASCE – American Society of Civil Engineers
BN – Bayesian Network
BT – Boosted Tree
CANPAT – Campanhas Nacionais de Prevenção de Acidentes do Trabalho
CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CART – Classification and Regression Trees
CAT – Comunicação de Acidente de Trabalho
CBIC – Câmara Brasileira da Indústria da Construção
CBO – Classificação Brasileira de Ocupação
CID – Classificação Internacional de Doenças
CLT – Consolidação das Leis do Trabalho
CNAE – Classificação Nacional de Atividades Econômicas
CNN – Convolutional Neural Network
DL – Deep Learning
DT – Decision Tree
EUA – Estados Unidos da América
FIS – Fuzzy Inference System
GSST – Gestão de Saúde e Segurança do Trabalho
IA – Inteligência Artificial
IDE – Ambiente de Desenvolvimento Integrado
IEEE – Institute of Electrical and Electronics Engineers
ILO – International Labour Organization
INSS – Instituto Nacional do Seguro Social
KNN – K-Nearest Neighbor

LAPES – Laboratório de Pesquisa em Engenharia de Software
LCCA – Latent Class Cluster Analysis
LDA – Linear Discriminant Analysis
LR – Logistic Regression
LSTM – Long Short-Term Memory
ML – Machine Learning
MLP – Multilayer Perceptron
MPT – Ministério Público do Trabalho
MTP – Ministério do Trabalho e Previdência Social
NB – Naïve Bayes
NLP – Natural Language Processing
NR – Norma Regulamentadora
NTEP – Nexo Técnico Epidemiológico Previdenciário
OIT – Organização Internacional do Trabalho
OMS – Organização Mundial da Saúde
ONU – Organização das Nações Unidas
OSHA – Occupational Safety and Health Administration
PCA – Principal Component Analysis
PIB – Produto Interno Bruto
PICO – População, Interesse e Contexto
PRISMA – Principais Itens para Relatar Revisões Sistemáticas e Meta-Análises
RF – Random Forest
RSL – Revisão Sistemática da Literatura
SAT – Seguro de Acidente do Trabalho
SENAI – Serviço Nacional de Aprendizagem Industrial
SGSST – Sistema de Gestão em Saúde e Segurança do Trabalho
SSD – Single Shot Detector
SST – Saúde e Segurança do Trabalho
START – State of the Art Through Systematic Review
STRAB – Secretaria do Trabalho
SVM – Support Vector Machine

TF-IDF – Term Frequency – Inverse Document Frequency

UFSCAR – Universidade Federal de São Carlos

WHO – World Health Organization

YOLO – You Only Look Once

SUMÁRIO

1 INTRODUÇÃO	17
1.1 Objetivos	18
<i>1.1.1 Objetivo geral</i>	18
<i>1.1.2 Objetivos específicos</i>	18
1.2 Hipóteses	19
1.3 Impactos e contribuições	19
1.4 Organização do trabalho	20
2 REFERENCIAL TEÓRICO	21
2.1 Segurança na Construção	21
2.2 Acidentes de trabalho no Brasil: legislação e instrumentos	24
2.3 Comunicação de Acidente de Trabalho – CAT	26
2.4 Tecnologias para prevenção de acidentes	28
2.5 Machine Learning	29
2.6 Aplicações de ML na construção	32
3 METODOLOGIA	39
3.1 Revisão Sistemática da Literatura	39
<i>3.1.1 Pesquisa bibliométrica</i>	40
<i>3.1.2 Análise descritiva</i>	42
3.2 Aplicação de Machine Learning	42
<i>3.2.1 Obtenção dos dados</i>	43
<i>3.2.2 Organização e tratamento dos dados</i>	43
<i>3.2.3 Leitura dos dados</i>	44
<i>3.2.4 Seleção dos atributos</i>	44
<i>3.2.5 Seleção do algoritmo</i>	44
<i>3.2.6 Análise dos resultados</i>	45
4 RESULTADOS	46
4.1 RSL	46
<i>4.1.1 Análise Bibliométrica</i>	46
<i>4.1.1.1 Protocolo de pesquisa</i>	46

4.1.1.2 Diagrama de fluxo	49
4.1.1.3 Visualização gráfica	50
4.1.2 Análise descritiva	56
4.1.2.1 Métodos de ML	56
4.1.2.2 Tipos de dados	59
4.1.2.3 Quantidade de dados	60
4.1.2.4 Método de coleta dos dados	60
4.1.2.5 Abordagens dos dados	61
4.2 Análise de dados com <i>Machine Learning</i>	62
4.2.1 Entrada dos dados	62
4.2.2 Pré-processamento dos dados	64
4.2.3 Análise Exploratória dos Dados – AED	67
4.2.3.1 Acidentes típicos	68
4.2.3.2 Acidentes de trajeto	75
4.2.3.3 Doenças ocupacionais	81
4.2.3.4 Acidentes fatais	87
4.2.4 Seleção dos atributos	98
4.2.5 Aplicação dos algoritmos	99
4.2.5.1 K-Modes	100
4.2.5.2 K-Prototype	101
4.2.6 Resultados do modelo	101
5 DISCUSSÃO	110
6 CONCLUSÕES	116
REFERÊNCIAS	118
ANEXO A – Modelo de Comunicação de Acidente de Trabalho	128
ANEXO B – Modelo de ML	130

1 INTRODUÇÃO

A Construção Civil é considerada um dos setores produtivos mais relevantes no mundo, com alta participação na economia dos países. No entanto, também é responsável pela ocorrência de milhares de acidentes de trabalho todos os anos. Os altos números de eventos evidenciam que a forma de conduzir as atividades na construção não está sendo eficaz o suficiente para garantir condições adequadas de saúde e segurança aos trabalhadores. Desse modo, entende-se que melhorar a gestão de segurança nas obras é algo necessário, assim como estudar formas de otimizá-las.

As causas dos acidentes de trabalho no setor da construção são normalmente conhecidas e potencialmente evitáveis (ORGANIZAÇÃO INTERNACIONAL DO TRABALHO – OIT, 2015). Dessa forma, é fundamental realizar o monitoramento abrangente e preciso da saúde e segurança dos trabalhadores, além de atuar em medidas que visem a eliminação ou diminuição dos riscos envolvidos na execução das atividades (ORGANIZAÇÃO MUNDIAL DA SAÚDE – OMS; OIT, 2021). Nesse contexto, também é essencial compreender os fatores gerais relacionados à ocorrência dos acidentes para que seja possível traçar estratégias específicas evitando, assim, novos eventos da mesma natureza.

Atualmente, várias tecnologias vêm sendo adotadas por pesquisadores para analisar e transformar dados em informações úteis, desde a estatística básica às técnicas computacionais automatizadas de alto nível, sendo uma destas o aprendizado de máquinas, ou do inglês *Machine Learning* – ML (CHOKOR *et al.*, 2016). A análise de dados baseada em ML é uma abordagem que vem ganhando popularidade ao longo dos anos, devido ao fato dos sistemas computacionais conseguirem analisar informações de forma mais eficiente do que o processo manual, tanto em relação ao custo e tempo de operação, quanto à qualidade das informações (SARKAR; MAITI, 2020).

A utilização de ML tem apresentado resultados promissores em diversos setores, inclusive na prevenção dos acidentes na construção. A descoberta de informações importantes por meio da identificação de relações e padrões ocultos dos dados e a possibilidade de gerar previsões de eventos, vem contribuindo com a resolução de problemas gerais associados a ocorrência dos

acidentes, aumentando a eficiência da gestão de segurança das obras. Além disso, pesquisas recentes desenvolvidas com tecnologias baseadas em dados, sugerem que os acidentes de trabalho na construção não ocorrem de forma aleatória, dessa forma, dados históricos podem ser utilizados como base para prevenir futuros eventos (TIXIER *et al.*, 2016a; GOH; UBEYNARAYANA, 2017).

No Brasil, o registro oficial dos acidentes de trabalho gera grande quantidade de dados relevantes que muitas vezes não são aproveitados para gerar informações úteis às políticas públicas de Saúde e Segurança do Trabalho – SST (SMARTLAB, 2022). Nesse sentido, é importante desenvolver pesquisas que utilizem esses registros para descobrir padrões de acidentes, contribuindo com a tomada de decisões relacionada à gestão de SST nas obras. Portanto, esta pesquisa teve o objetivo de analisar dados de acidentes de trabalho da construção, registrados no Brasil nos últimos anos, utilizando modelo de ML baseado em aprendizagem não supervisionada para identificar informações relacionadas aos acidentes e compreender os fatores gerais associados à ocorrência desses eventos.

1.1 Objetivos

1.1.1 Objetivo geral

Analisar dados oficiais brasileiros de acidentes de trabalho da Construção Civil com auxílio de aprendizado de máquina, de forma a contribuir com a tomada de decisão na Gestão de Saúde e Segurança do Trabalho dos canteiros de obras.

1.1.2 Objetivos específicos

- Realizar Revisão Sistemática da Literatura – RSL sobre o tema proposto;
- Identificar as aplicações de ML que estão sendo desenvolvidas no mundo e suas principais características;
- Analisar dados oficiais brasileiros de acidentes de trabalho da construção do período de 2018 a 2021;

- Identificar grupos prioritários de acidentes por meio de algoritmos de ML baseados em *clustering* (*K-Modes* e *K-Prototype*).

1.2 Hipóteses

Este trabalho visa investigar as seguintes hipóteses:

- H1: No Brasil, o aprendizado de máquinas é incipiente no contexto da prevenção de acidentes de trabalho na construção;
- H2: Método de *cluster* é eficaz em encontrar padrões de acidentes nos dados de Comunicação de Acidente de Trabalho – CAT.

1.3 Impactos e contribuições

As contribuições deste trabalho incluem:

- Impacto Social: Promoção da segurança e do bem-estar aos trabalhadores da construção, colaborando com a redução dos acidentes no setor, maior qualidade de vida e melhores condições de trabalho;
- Impacto Tecnológico: Colaboração com o desenvolvimento tecnológico da Indústria da Construção brasileira pela utilização de tecnologias para otimização dos processos produtivos;
- Impacto Ambiental: Alinhamento aos Objetivos de Desenvolvimento Sustentável – ODS de número 8 e 9 da Agenda 2030 (ORGANIZAÇÃO DAS NAÇÕES UNIDAS – ONU, 2015), relacionados à promoção do trabalho decente e à industrialização sustentável e inovação.
- Impacto Científico: Estímulo ao intercâmbio de ideias entre as áreas da Engenharia Civil, Saúde e Segurança do Trabalho e Computação, fortalecendo a interdisciplinaridade e a pesquisa científica.
- Impacto Econômico: Colaboração com a redução dos custos empresariais e governamentais provenientes dos acidentes de trabalho, contribuindo com o desenvolvimento econômico do país.

1.4 Organização do trabalho

O presente trabalho segue a estrutura:

- Capítulo 1 – Introdução: aborda e contextualiza o problema da pesquisa, apresentando os objetivos geral e específicos, as hipóteses cujo trabalho visa responder e os impactos/contribuições relacionados ao seu desenvolvimento;
- Capítulo 2 – Metodologia: apresenta os métodos utilizados para desenvolver a pesquisa, divide-se em uma Revisão Sistemática da Literatura – RSL e na aplicação de *Machine Learning* – ML para analisar dados oficiais brasileiros de acidentes de trabalho da construção;
- Capítulo 3 – Resultados: apresenta as descobertas obtidas por meio dos métodos utilizados, divide-se nos resultados da RSL e da aplicação de ML;
- Capítulo 4 – Discussão: traz comentários a respeito dos resultados;
- Capítulo 5 – Conclusão: sintetiza o aprendizado adquirido com o desenvolvimento da pesquisa.

2 REFERENCIAL TEÓRICO

O referencial teórico deste trabalho foi dividido em seis tópicos. Inicialmente, foi abordado como o setor da construção se comporta em relação aos aspectos de Saúde e Segurança do Trabalho. Em seguida, focou-se na legislação e nos instrumentos de comunicação e controle dos acidentes de trabalho no Brasil, passando pela explanação das Comunicações de Acidente de Trabalho – CAT. Ademais, foram abordadas as tecnologias para a prevenção de acidentes que vem sendo utilizadas atualmente, com foco no aprendizado de máquina, trazendo exemplos de aplicações observadas na literatura internacional.

2.1 Segurança na Construção

A Indústria da Construção Civil é um dos setores responsáveis por impulsionar a economia dos países, sendo considerada um importante componente dos investimentos nacionais e muitas vezes o termômetro para a avaliação da situação econômica de um local. Além disso, a construção é capaz de proporcionar desenvolvimento social, visto que diversos setores são influenciados e se relacionam com suas atividades, como comércio, educação, indústria e transporte, devido a necessidade de construção, manutenção e operação de fábricas, hospitais, escolas, rodovias, aeroportos, habitações, entre outros (SERVIÇO NACIONAL DE APRENDIZAGEM INDUSTRIAL – SENAI; CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO – CBIC, 2021). Dados do Banco Mundial (2007) comentam que a indústria movimenta cerca de 1,7 trilhões de dólares anualmente, com percentual de participação no PIB variando entre 5 e 7% na maioria dos países. Além disso, em relação à mão de obra utilizada no setor, a OIT estima que a construção seja responsável por empregar entre 6% e 10% da força de trabalho mundial (SOMAVIA, 2005).

Em contrapartida, o setor da construção normalmente é conhecido pelos altos números de acidentes de trabalho, estando entre as indústrias mais perigosas e ocupando o quarto lugar entre as que mais matam (ALLI, 2001; OIT, 2014). Estima-se que pelo menos 60 mil acidentes fatais ocorram em canteiros de obra ao redor do mundo todos os anos, o que representa 1 fatalidade a cada 10 minutos (SOMAVIA, 2005; LINGARD, 2013). Em países desenvolvidos, a OIT (2014) estima que os

trabalhadores da construção civil estejam de 3 a 4 vezes mais propensos a sofrer um acidente fatal no trabalho do que outros trabalhadores. Nos países em desenvolvimento, no entanto, os riscos associados às obras são de 3 a 6 vezes maiores. No Brasil, também é possível observar taxas alarmantes de acidentes na construção, conforme mostram os dados do Observatório de Segurança e Saúde do Trabalho (SMARTLAB, 2022) que apontam cerca de 130 mil acidentes de trabalho no setor da construção, especificamente a construção de edifícios, registrados entre o período de 2012 e 2021, dentre os quais, 875 foram relacionados aos casos fatais.

Os motivos para a alta incidência de acidentes na construção podem ser explicados pela natureza dinâmica dos canteiros de obra, cujos ambientes de trabalho possuem características específicas, temporárias, complexas e com o envolvimento simultâneo de diferentes entidades, como trabalhadores, materiais e máquinas/equipamentos. Somado a isso, a alta rotatividade, informalidade e ausência de qualificação da mão de obra no setor, além das extensas jornadas de trabalho as quais os trabalhadores se submetem e os processos executivos muitas vezes artesanais tornam os ambientes de trabalho mais propícios à ocorrência dos acidentes (SAMOVIA, 2005; KIM; CHI, 2019; LIMA JÚNIOR, 2019; KHAN *et al.*, 2021). As particularidades da construção, quando aliadas à gama de riscos presentes nas diversas fases do processo construtivo das obras, torna o gerenciamento de segurança do ambiente laboral algo dificultoso e muitas vezes ineficiente, podendo gerar, além dos acidentes, prejuízos financeiros às empresas, tais como situações de retrabalho, desperdícios, diminuição da produtividade, comprometimento da qualidade dos produtos e, inclusive, demandas jurídicas, das esferas trabalhista, previdenciária, civil e penal (LIMA JÚNIOR, 2019, p. 83).

Os acidentes de trabalho são provocados por uma conjunção de fatores, e estes podem ser classificados basicamente em cinco grupos, os físicos, químicos, biológicos, ergonômicos e mecânicos. Os riscos de natureza física são relacionados à exposição ao ruído, vibrações, radiações ionizantes e não ionizantes, altas e baixas temperaturas, pressões anormais, umidade, entre outros. Agentes químicos referem-se ao contato, inalação e/ou absorção de poeiras, fumos, gases, neblinas e substâncias compostas ou produtos químicos em geral. Os biológicos são relacionados à exposição aos vírus, bactérias, fungos, parasitas, etc. Os ergonômicos, por sua vez, são aqueles

relacionados às posturas inadequadas, esforço excessivo, jornadas de trabalho extensas, entre outros. (BRASIL, 1994; ALLI, 2001; OMS, 2010).

Todavia, dentre as principais causas de fatalidades na construção, destacam-se os fatores mecânicos, relacionados geralmente aos trabalhos realizados em altura, à movimentação de materiais, aos serviços de escavação e os com eletricidade, como os acidentes de quedas, soterramento e eletrocussão, responsáveis por milhares de mortes por ano no mundo (CHOKOR *et al.*, 2016; FANG Q. *et al.*, 2018; ZHANG; CAO; ZHAO, 2019; ASSAAD; EL-ADAWAY, 2021). Sabe-se, no entanto, que as causas dos acidentes no setor são normalmente conhecidas e potencialmente evitáveis (ALLI, 2001; OIT, 2015). Dessa forma, a implementação de práticas de Saúde e Segurança do Trabalho – SST tem sido peça fundamental para a eliminação dos perigos e minimização dos riscos das atividades executadas nos canteiros de obras.

O Sistema de Gestão da Saúde e Segurança do Trabalho – SGSST consiste em um conjunto de elementos interrelacionados ou integrantes de uma organização para estabelecer políticas, objetivos e processos para atingir estes objetivos. Os elementos do sistema, por sua vez, incluem a estrutura, funções, responsabilidades, planejamento, operação, avaliação de desempenho e melhoria da organização, cujo resultado principal é prevenir acidentes e qualquer efeito adverso sobre as condições físicas, mentais e cognitivas dos trabalhadores, contribuindo ainda para a promoção de ambientes de trabalho seguros e saudáveis (ABNT, 2018).

De acordo com a OIT (SOMAVIA, 2005), além dos benefícios evidentes para os trabalhadores, relacionados à prevenção dos acidentes e doenças ocupacionais, a implantação do SGSST vem mostrando ser relevante na melhoria do desempenho empresarial. Atualmente, as empresas entendem que a SST é um componente importante do sucesso financeiro de um empreendimento e que o setor caminha em paralelo com a qualidade, produtividade e redução de custos (BADRI *et al.*, 2018). Ademais, a adoção de uma cultura preventiva de segurança no setor, de um ponto de vista macro, auxilia na redução dos custos sociais decorrentes dos acidentes de trabalho, como o caso dos benefícios acidentários e pensões específicas. Dessa forma, entende-se que os ganhos da implantação da gestão de segurança atingem trabalhador, empregador e sociedade em geral e, portanto, a participação em conjunto dessas esferas é algo essencial para o funcionamento efetivo

da SST, seja no estabelecimento de diretrizes de segurança, na implementação e cumprimento de procedimentos de organização do trabalho, na oferta de formação e informação aos trabalhadores ou nas atividades de inspeção.

2.2 Acidentes de trabalho no Brasil: legislação e instrumentos

O Brasil dispõe de um conjunto de leis e normas que estabelecem os requisitos mínimos necessários para garantir aos empregados condições adequadas de saúde e segurança nos ambientes de trabalho, sendo estes dispositivos responsáveis por definir conceitos e determinar responsabilidades (TEIXEIRA; REIS; MENDES, 2019). Assim, de acordo com a Lei 8.213 de 24 de julho de 1991 em seu Art. 19, acidente de trabalho é o que ocorre ao profissional a serviço de empresa ou pelo exercício do trabalho, independente de sua situação empregatícia, provocando lesão corporal ou perturbação funcional que cause a morte ou a perda ou redução, permanente ou temporária, da capacidade para o trabalho (BRASIL, 1991).

Os acidentes de trabalho, pelo Art. 20 da referida lei, podem ser divididos em acidentes e doenças ocupacionais, estas últimas englobam: a doença profissional, produzida ou desencadeada pelo exercício do trabalho peculiar a determinada atividade e constante da respectiva relação elaborada pelo Ministério do Trabalho e da Previdência Social; e a doença do trabalho, adquirida ou desencadeada em função de condições especiais em que o trabalho é realizado e com ele se relacione diretamente. Ademais, o Art. 21 considera acidente as diversas situações: sofrida pelo trabalhador no local e no horário do trabalho, em consequência de situações como agressão, imprudência e fatores decorrentes de força maior (desabamentos, incêndio, etc.); ocorrida pelo trabalhador ao representar os interesses da empresa ou agindo em defesa de seu patrimônio; ocorrida no trajeto da residência para o local de trabalho ou vice-versa, independente do seu meio de locomoção; e outras quando constatado o Nexo Técnico Epidemiológico Previdenciário – NTEP causal entre o trabalho e a lesão (BRASIL, 1991; BRASIL, 2006).

As Normas Regulamentadoras – NRs, disposições complementares da Consolidação das Leis do Trabalho – CLT, constituem em instrumentos para estabelecer as obrigações, direitos e deveres a serem cumpridos pelos empregadores e trabalhadores visando garantir trabalho seguro e sadio. As

primeiras NRs foram publicadas em 1978, no entanto, novas normas vieram sendo criadas ao longo dos anos para tratar de serviços e setores de forma mais específica, levando em consideração suas particularidades em relação à prevenção de acidentes e doenças. O processo de elaboração e atualização das NRs baseia-se nas premissas da OIT, onde é adotado um sistema tripartite paritário com comissões formadas por representantes do governo, empregadores e trabalhadores com o objetivo de atingir um senso comum e justo para todas as partes interessadas (BRASIL, 2022).

Dentre as obrigações frente às NRs, destaca-se o papel da Secretaria do Trabalho – STRAB na formulação de diretrizes, normas de atuação e supervisão das atividades em relação à SST, além da promoção de Campanhas Nacionais de Prevenção de Acidentes do Trabalho – CANPAT, implementação de políticas nacionais de SST, fiscalização do cumprimento das leis e normas de SST no território nacional e aplicação de penalidades relacionadas ao descumprimento. Em relação às outras partes, cabe ao empregador responsabilizar-se por cumprir as disposições de SST, devendo sempre informar aos trabalhadores sobre os riscos ocupacionais presentes nos locais de trabalho e implementar medidas preventivas para redução dos riscos, enquanto cabe ao trabalhador cumprir as normativas e colaborar com sua devida aplicação, entre outros aspectos relevantes (BRASIL, 2020).

Na ocorrência de um acidente de trabalho ou doença ocupacional, a legislação brasileira entende que o trabalhador é o elo mais frágil. Portanto, além da esfera legal, os trabalhadores podem contar com um sistema de proteção oferecido pela esfera institucional da Previdência Social, criado não só com o objetivo de garantir a renda do acidentado, mas também de atuar na prevenção desse tipo de evento relacionado ao trabalho. Dessa forma, existem instrumentos que auxiliam o gestor público no objetivo de alinhar a ação do governo com as empresas: os instrumentos de comando e controle, como as normas de SST; instrumentos econômicos, como o Seguro de Acidente do Trabalho – SAT; e instrumentos de comunicação e informação, como a Comunicação de Acidente de Trabalho – CAT (FERRO, 2019).

O propósito principal dos sistemas de informação sobre acidentes do trabalho é fornecer informações confiáveis sobre seu impacto, sejam relacionadas aos trabalhadores afetados, lesões provocadas, custo gerado aos cofres públicos ou aspectos associados às suas origens. O

conhecimento obtido por meio da análise dos acidentes tem a capacidade de indicar falhas sistêmicas causadas por potenciais fatores de risco, informações que podem ser usadas em políticas públicas e pelos gestores como ferramentas de prevenção de acidentes (WÄNNSTRÖM; LARSSON, 1996; BRASIL, 2006; ZHANG; CAO; ZHAO, 2019).

2.3 Comunicação de Acidente de Trabalho – CAT

Em decorrência de um acidente ou constatação de uma doença ocupacional, seja estes com ou sem afastamento do trabalhador de suas atividades, os empregadores são obrigados a comunicar o evento à Previdência Social. Desse modo, o registro oficial da situação é realizado por meio do preenchimento da CAT (ANEXO A). A comunicação do acidente pode ser realizada de forma imediata, nos casos de lesões fatais, ou em até 1 dia útil após o evento, para os demais. Caso não seja realizada pelo empregador, o comunicado pode ser emitido pelo próprio segurado ou dependente, médico, entidades sindicais ou autoridade pública (INSTITUTO NACIONAL DO SEGURO SOCIAL – INSS, 2022).

A CAT possui campos obrigatórios para preenchimento, referentes: ao tipo da comunicação, se consta de CAT inicial, de reabertura ou refere-se a uma comunicação de óbito; ao tipo de acidente de trabalho sofrido, se é considerado acidente típico, de trajeto ou doença ocupacional; às informações do empregador (razão social, atividade econômica, endereço, etc.); às informações do empregado acidentado (dados pessoais, função, tipo de filiação, etc.); aos dados sobre o acidente (data, local, hora, parte do corpo atingida, agente causador, etc.); aos dados sobre o atendimento recebido; e aos dados médicos referente ao acidente/doença (natureza da lesão, classificação da doença, etc.).

As informações sobre as atividades econômicas do empregador são relacionadas à Classificação Nacional de Atividades Econômicas – CNAE, classificação oficialmente adotada pelos órgãos federais administrativos para a padronização nacional dos códigos de atividades econômicas. A classificação, por sua vez, segue uma estrutura hierárquica, dividida em seções, divisões, grupos, classes e subclasses (BRASIL, 2020). A respeito das funções dos acidentados, a CAT utiliza como base a Classificação Brasileira de Ocupações – CBO, classificação instituída em 2002 por Portaria

Ministerial que tem o objetivo de expor as diversas atividades profissionais existentes no país. Seus dados alimentam as bases estatísticas de trabalho e servem de subsídio para a formulação de políticas públicas de emprego (BRASIL, 2017). Sobre as informações dos problemas relacionados à saúde, a CAT utiliza os códigos da Classificação Internacional de Doenças – CID-10, publicada pela Organização Mundial da Saúde – OMS com o objetivo de fornecer conhecimento crítico sobre a extensão, causas e consequências de doenças e mortes humanas em todo o mundo por meio de dados relatados e codificados. Esta classificação é utilizada no mundo todo, inclusive no Brasil, e seus dados e estatísticas baseiam sistemas de pagamento de benefícios, planejamento de serviços, gestão de qualidade e segurança, além de pesquisas de serviços de saúde (OMS, 2022). As demais informações presentes nas CAT, como natureza da lesão, parte do corpo atingida e agente causador, por exemplo, seguem as listas de tabulação fornecidas pelo governo brasileiro, utilizadas também para o cadastro no e-social, que tem o objetivo de padronizar as informações, possibilitando análises futuras para diversos fins.

Um problema comum associado à comunicação dos acidentes de trabalho, evidenciado pelo Ministério Público do Trabalho – MPT e por muitos estudos brasileiros, refere-se à subnotificação dos casos, como situações em que não houve diagnóstico, casos não registrados devido às consequências consideradas menos graves para a saúde dos acidentados, além das doenças ocupacionais, que muitas vezes são relatadas anos após o período trabalhado (BRASIL, 2006). O percentual de subnotificação dos registros, conforme dados do MPT, pode chegar aos 20% dos casos, algo que dificulta a estimar e acompanhar o real número dos acidentes e doenças e seus impactos na saúde da população brasileira (SMARTLAB, 2022).

Apesar das subnotificações, a CAT é a forma oficial e mais fidedigna de acompanhamento dos casos de acidentes de trabalho no Brasil. A análise do seu conteúdo pode oferecer aos gestores públicos e empresários a possibilidade de identificar informações acerca da natureza dos acidentes mais frequentes, detectando padrões que de alguma forma poderiam permanecer ignorados, além de viabilizar políticas e direcionar procedimentos que visem a prevenção dos acidentes e doenças ocupacionais (GIACOMELLI, 2020). Portanto, orientar as empresas sobre sua necessidade e forma correta de preenchimento, além de procurar utilizar ferramentas para extrair informações desse tipo

de registro, são ações essenciais para colaborar com o processo de gestão de saúde e segurança dos canteiros de obras.

2.4 Tecnologias para prevenção de acidentes

A origem e a evolução dos esforços para melhorar a saúde, a segurança e o bem-estar dos trabalhadores são complexas e é relativamente recente que a promoção da saúde na esfera pública tem sido especificamente ligada ao local de trabalho (OMS, 2010). Com a primeira revolução industrial, que correspondeu às mudanças fundamentais dos processos industriais e de manufatura, deu-se início à chamada primeira era da SST. O período focou principalmente no desenvolvimento da engenharia e de soluções tecnológicas que pudessem tratar dos novos perigos presentes nos ambientes de trabalho. Devido às limitações das abordagens, a segunda era da segurança tratou da interface entre pessoas e tecnologias, focando nos fatores humanos. Já a terceira fase do processo evolutivo da SST buscou a implementação dos sistemas de gerenciamento de segurança. Assim, cada período trouxe ferramentas importantes para o desenvolvimento do setor, criando estratégias novas e mais completas para a resolução dos problemas (HALE; HOVDEN, 1998; LINGARD, 2013).

Atualmente, com a quarta revolução industrial, surge uma nova era da segurança, impulsionada pela mudança de paradigma trazida pela aplicação das TIC aos processos produtivos. A Indústria 4.0, por sua vez, promete aumento de produtividade por meio da integração de sistemas digitais de produção com a análise e comunicação dos dados, gerados dentro de um ambiente inteligente. Dentre os principais benefícios dessas inovações tecnológicas destacam-se a comunicação em tempo real, o *Big Data*, cooperação homem-máquina, sensoriamento remoto, monitoramento e controle automáticos, equipamentos autônomos e interconectividade, funcionalidades que, entre outros aspectos, tem o potencial de colaborar com o desempenho de segurança nos canteiros de obras (BADRI *et al.*, 2018; TEIXEIRA; REIS; MENDES, 2019). O Quadro 2.1 apresenta as principais abordagens tecnológicas trazidas pela Indústria 4.0 e suas respectivas definições.

Quadro 2.1 – Abordagens tecnológicas da Indústria 4.0

Abordagem	Definição
<i>Big Data</i>	Conjuntos de dados extensos que excedem a capacidade de gerenciamento das informações dos sistemas tradicionais de processamento de dados, exigindo tecnologias específicas para processá-los.
<i>Internet of Things – IoT</i>	Trocas de informações e dados que chegam à Internet a partir de dispositivos que realizam tarefas reais no mundo físico.
Sistemas ciber físicos (CPS)	Um sistema no qual elementos computadorizados colaboram para monitorar e controlar entidades físicas.
Robôs colaborativos (<i>Cobotics</i>)	Ramo da tecnologia de design robótico baseado na combinação das ciências da informação, fatores humanos (comportamento, decisão, robustez e monitoramento de erros), biomecânica (modelagem do comportamento e da dinâmica do movimento) e robótica.
Simulação	Representação do comportamento de um processo industrial por meio de um modelo computacional em que os parâmetros e variáveis são reflexos daqueles do processo em estudo. Ex.: Realidade Virtual – RV, Realidade Aumentada – RA, Building Information Modeling – BIM.
Inteligência Artificial – IA	As teorias multidisciplinares, técnicas, conceitos e tecnologias implementadas para desenvolver máquinas capazes de simular inteligência. Ex.: Machine Learning – ML, Deep Learning – DL.

Fonte: Adaptado de Badri *et al.* (2018)

Apesar dos grandes ganhos obtidos com a evolução da SST, as altas taxas de acidentes e fatalidades na construção continuam a evidenciar que os métodos convencionais de gerenciamento de segurança necessitam de melhorias. Dessa forma, investir em ferramentas que possibilitem o aperfeiçoamento das técnicas de gerenciamento de segurança e direcionar esforços para a industrialização do setor da construção, buscando utilizar as tecnologias habilitadoras nos processos produtivos, espelhando-se inclusive nos exemplos de outras indústrias que obtiveram sucesso, poderia gerar um avanço considerável no desempenho da construção, tornando-a mais produtiva e segura.

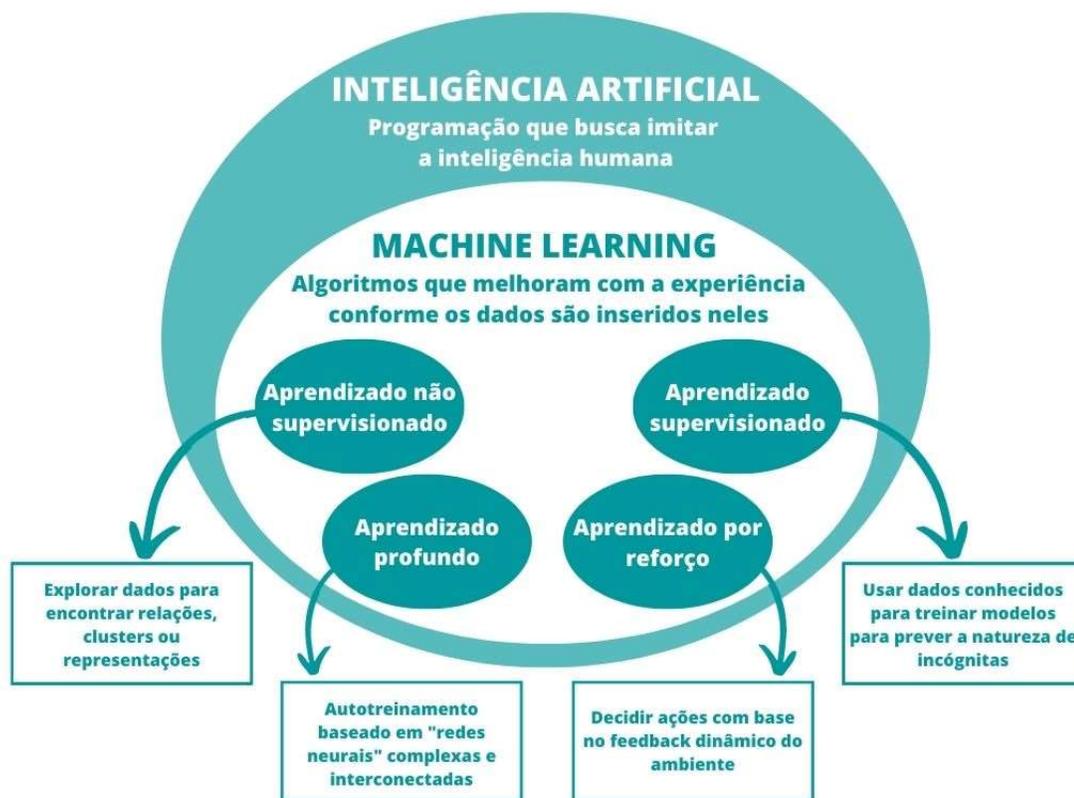
2.5 Machine Learning

Uma das tecnologias mais recentes e populares da Indústria 4.0 refere-se à um campo da IA conhecido por aprendizagem de máquina, ou do inglês *Machine Learning* – ML. Na abordagem, os sistemas computacionais são estimulados a aprender e aprimorar as experiências automaticamente, assim, a tecnologia permite a obtenção de conhecimento de forma autônoma, por meio de uma variedade de algoritmos que aprendem iterativamente com os dados para melhorar, descrever informações e até prever resultados (HURWITZ; KIRSCH, 2018; SARKER, 2021).

O uso de ML tem evoluído nos últimos anos, impulsionado pelo aumento da capacidade de processamento computacional e pelo acesso a dados. Na prevenção de acidentes, as vantagens são evidentes, não só relacionadas à economia de tempo, custo e mão de obra para realizar análises de segurança, mas principalmente pelo fato de os modelos conseguirem aprender com as situações passadas e auxiliar nos processos de tomada de decisão. Nesse contexto, diversas empresas, instituições e órgãos governamentais têm utilizado a tecnologia para auxiliar no acompanhamento de suas atividades e na otimização de suas ações (TOLETO; TIMOTEO; BARBOSA, 2020).

Diversas são as abordagens e métodos utilizados em *Machine Learning*, todavia, existem quatro tipos principais de aprendizagem, como pode ser evidenciado na Figura 2.1, a supervisionada (*supervised learning*), não supervisionada (*unsupervised learning*), por reforço (*reinforcement learning*) e profunda (*deep learning*).

Figura 2.1 – Inteligência Artificial e as abordagens de *Machine Learning*



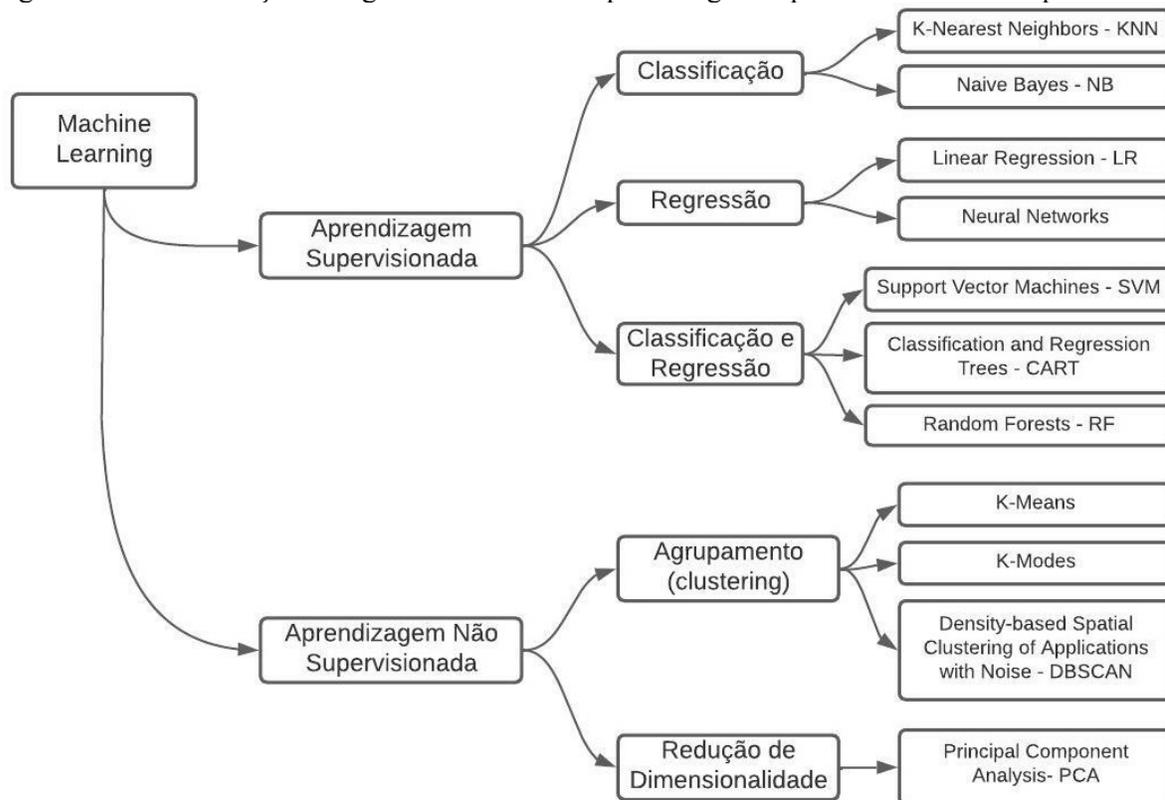
Fonte: Adaptado de Robinson e Akins (2021)

A abordagem supervisionada normalmente é utilizada em conjuntos de dados padronizados, quando há certa compreensão a respeito de sua classificação e os dados apresentam rótulos (*labels*) para definir seus respectivos significados. (HURWITZ; KIRSCH, 2018). O aprendizado por reforço, por sua vez, permite aprender por tentativa e erro em um ambiente interativo, a partir da entrada de ações e experiências. Ao contrário do aprendizado supervisionado, que é baseado em dados de amostra ou exemplos, o método é baseado na interação com o ambiente (SARKER, 2021). Outra abordagem de ML é o aprendizado profundo ou *Deep Learning* – DL, método baseado em redes neurais que simula o funcionamento do cérebro humano. O DL utiliza camadas sucessivas para aprender com os dados de maneira iterativa, sendo normalmente utilizado quando o objetivo é encontrar padrões de dados não estruturados (HURWITZ; KIRSCH, 2018; POUYANFAR *et al.*, 2018).

No aprendizado de máquina não supervisionado, os algoritmos são usados para determinar padrões ou estruturas de um conjunto de dados. A aprendizagem é chamada de “sem supervisão” no sentido de que os dados não são rotulados ou vinculados a resultados específicos, ou seja, os resultados ainda serão descobertos pelos modelos. Nesse contexto, os tipos mais comuns envolvem o agrupamento de dados em grupos (*clustering*) com base nas semelhanças ou diferenças entre os membros do grupo (ROBINSON; AKINS, 2021). A grande vantagem da técnica de *clustering* é poder descrever de forma mais eficiente as características de cada agrupamento, obtendo um maior entendimento do conjunto de dados original e descobrindo relações importantes entre os atributos.

Em geral, a eficácia e a eficiência de uma solução de ML vão depender da natureza do problema abordado, além das características, volume e qualidade dos dados e do desempenho dos algoritmos de aprendizado (FAN, 2020; SARKER, 2021). Em relação aos algoritmos usados em ML, existem técnicas de classificação, regressão, agrupamento (*clustering*), redução de dimensionalidade, entre outras. Como exemplo, a Figura 2.2 apresenta a classificação de alguns algoritmos de ML baseados nas abordagens mais comuns, a supervisionada e não supervisionada.

Figura 2.2 – Classificação de algoritmos de ML de aprendizagem supervisionada e não supervisionada



Fonte: Adaptado de Alsharif *et al.* (2020)

O uso dos algoritmos de forma isolada podem apresentar resultados promissores, no entanto, a combinação de métodos de ML pode conferir resultados mais positivos ao desenvolvimento dos modelos, principalmente relacionados à precisão alcançada. Além disso, problemas complexos e grandes volumes de dados requerem a combinação de técnicas de ML para efetivamente contribuir com a solução do problema (FAN, 2020).

2.6 Aplicações de ML na construção

No contexto da SST, é possível observar diversos tipos de aplicações de ML que vem sendo utilizados na resolução de problemas de segurança de setores econômicos variados. Mais especificamente na construção civil, modelos de ML têm sido aplicados no reconhecimento automático de situações e pessoas, extração de informações de bases de dados de acidentes, predições de gravidade de lesões, entre outros exemplos.

Risco de colisão

Operações envolvendo máquinas e equipamentos geram milhares de acidentes de atropelamentos e esmagamentos, por exemplo, sendo consideradas atividades perigosas quando executadas nos canteiros de obras. Neste sentido, muitos pesquisadores desenvolveram modelos visando evitar colisões entre funcionários e máquinas por meio da detecção e rastreamento de entidades ou previsão de trajetórias (BRILAKIS; PARK; JOG, 2011; ZHANG M. *et al.*, 2020; TANG; ROBERTS; GOLPARVAR-FARD, 2020; SON; KIM, 2021).

Risco de queda

Em relação às situações de quedas, pesquisas buscaram contribuir com a segurança dos trabalhos em altura, desenvolvendo modelos de reconhecimento de quase acidentes, alguns inclusive com capacidade de identificar os elementos perigosos e trabalhadores vulneráveis (YANG *et al.*, 2016; ZHANG F. *et al.*, 2019). Além disso, estudiosos criaram soluções para a gestão de risco nos serviços em altura, como a detecção automática de equipamentos de proteção (EPIs, sistemas de travamento, guarda-corpos e andaimes), verificação de conformidade das proteções em relação às normas de segurança e detecção de comportamento inseguro dos trabalhadores (SIDDULA *et al.*, 2016; KOLAR; CHEN; LUO, 2018; FANG Q. *et al.*, 2018; KHAN *et al.*, 2021).

Ainda sobre trabalhos em altura, pesquisadores criaram modelos para identificar as causas de quedas e as relações de dependência entre as diferentes variáveis (MARTÍN *et al.*, 2009; MISTIKOGLU *et al.*, 2015), prevendo a gravidade do risco (CHEN; LUO, 2016) e mais especificamente prevendo quedas de andaimes (JAHANGIRI; SOLUKLOEI; KAMALINIA, 2019). Um exemplo pode ser evidenciado em Bonifazi *et al.* (2022) que criaram plataforma capaz de identificar ocorrências de quedas, estabelecendo suas causas e severidade, possibilitando ainda, a partir dessas informações, o acionamento de alarmes e a organização e ativação de planos de gerenciamento de resgate. Assim como Sakhakarmi, Park e Cho (2019) que propuseram um modelo para analisar estruturas de andaimes em tempo real para prever condições de segurança por meio da alimentação de dados de deformação contínua.

Risco ergonômico

A necessidade de garantir condições ergonômicas adequadas aos trabalhadores para a execução segura das atividades motivou o surgimento de aplicações que também pudessem reconhecer e monitorar informações de cargas de trabalho e postura corporal. Assim, estudos focaram no reconhecimento de ações inseguras abordados por meio de sistemas de modelagem e classificação de ações humanas utilizando dados de captura de movimento (HAN; LEE; PEÑA-MORA, 2014; DING *et al.*, 2018; ZHAO; OBONYO, 2020).

Um exemplo pode ser visto em Yu *et al.* (2019) que desenvolveram modelo integrado para fornecer avaliações ergonômicas automáticas e detalhadas de trabalhadores da construção. O modelo proposto pelos autores foi combinado com módulo de pontuação de risco ergonômico baseado em avaliação rápida de corpo inteiro e pôde identificar as atividades de alto risco e os cronogramas de descanso adequados para cada atividade ou trabalhador. Ainda nesse contexto, Alwasel *et al.* (2017) desenvolveram modelo para classificar poses de trabalho entre grupos de pedreiros durante a construção de parede de concreto padrão. O trabalho, posteriormente continuado por Ryu *et al.* (2020), identificou com alta precisão que pedreiros com maior experiência realizam as atividades com posturas ergonomicamente mais seguras, além de serem mais produtivos, o que indica menor gasto de energia e menos movimentos desperdiçados.

Fiscalização de obra

A fiscalização é fundamental na promoção da gestão de segurança nas obras, pois a partir da análise das situações, torna-se possível monitorar as atividades e identificar se elas estão sendo executadas de forma segura e eficiente. No entanto, devido ao seu caráter dinâmico e temporário, o canteiro de obras muitas vezes pode apresentar dificuldades em ser fiscalizado, trazendo a necessidade da criação de ferramentas que auxiliem no processo de inspeção nos locais de trabalho (KIM; CHI, 2019; ZHANG M. *et al.*, 2020; KHAN *et al.*, 2021). Neste contexto, são aspectos valorizados reconhecer e monitorar entidades de forma automática, como trabalhadores, materiais e máquinas/equipamentos (ZHANG; CAO; ZHAO, 2019).

Modelos focados para a fiscalização de canteiros de obras foram evidenciados em diversos trabalhos que utilizaram métodos de ML para reconhecimento do uso correto de EPIs diversos (CHEN; DEMACHI, 2021), reconhecimento das cores de capacetes de segurança (SHEN *et al.*, 2021; HUANG *et al.*, 2021), detecção e monitoramento visual de trabalhadores e atividades (TEIZER, 2008; TANG *et al.*, 2020; FANG Q. *et al.*, 2020; LIU *et al.*, 2021), detecção sonora de acidentes e perigos (LEE Y. *et al.*, 2020), detecção automática de recursos de segurança e trabalhadores em obras de edifícios pré-fabricados (LIU *et al.*, 2021), entre outros. Além disso, a necessidade de reduzir o tempo, o custo e a probabilidade de erros na verificação de conformidades em relatórios, fez com que algumas pesquisas desenvolvessem modelos específicos para automatizar este processo (SALAMA; EL-GOHARY, 2013).

Treinamento

A realização de treinamentos e capacitações de trabalhadores de construção é fundamental na manutenção da segurança nos canteiros de obras, pois aumenta o conhecimento e a conscientização sobre SST no local de trabalho (WACHTER; YORIO, 2014). Nesse contexto, Mariscal *et al.* (2019) desenvolveram modelo de ML para compreender a relação entre os treinamentos e informações passadas aos trabalhadores nas taxas de risco de acidente. Os resultados, por sua vez, mostraram que a probabilidade de um trabalhador sofrer um acidente na construção pode dobrar quando as informações sobre os riscos no local de trabalho não são suficientes, confirmando a importância dessas ações. Outra questão importante é identificar se de fato a cultura de segurança da empresa está sendo absorvida, sendo fundamental avaliar a percepção dos trabalhadores, como foi o caso da pesquisa desenvolvida por Singh e Misra (2020).

Análise de fatores de risco

Para que o gerenciamento de segurança seja, de fato, eficaz, é necessário estabelecer medidas de segurança adequadas para cada situação do processo, sendo fundamental analisar os riscos envolvidos nas atividades e compreender suas causas e consequências. Esta necessidade fez com que modelos de ML fossem utilizados para descobrir padrões ocultos relativos aos fatores de risco responsáveis pela ocorrência dos acidentes, devido ao fato de poderem tratar com eficiência a

relação entre variáveis independentes (CIARAPICA; GIACCHETTA, 2009). Sabendo que uma lesão raramente, ou nunca, é o resultado de um único fator de risco, mas de uma gama de fatores, a utilização desses modelos contribui para manipular dados de acidentes de forma eficiente, deduzindo e sistematizando variáveis significativas para abordar de forma racional alguns problemas de segurança na construção (LEE J. *et al.*, 2020).

Neste contexto, muitos trabalhos criaram modelos para classificar e/ou identificar fatores gerais que podem afetar a incidência dos acidentes de trabalho (MATÍAS *et al.*, 2008; CIARAPICA; GIACCHETTA, 2009; LIAO; PERNG; CHIANG, 2009; ABDAT *et al.*, 2014; CHOKOR *et al.*, 2016; KIM; CHI, 2019; BAKER; HALLOWELL; ASSAAD; EL-ADAWAY, 2021). Alguns mais específicos, realizaram simulações para compreender, por exemplo, a relação entre o status de segurança dos projetos e aumento da procura por aluguel (FENG; SHI; ZHANG, 2019). Ademais, autores desenvolveram modelos para classificar relatórios de acidentes na construção (GOH; UBEYNARAYANA, 2017; TIXIER *et al.*, 2017; ZHANG F. *et al.*, 2019), classificar obras de acordo com o risco de segurança do projeto (POH; UBEYNARAYANA; GOH, 2018), determinar o nível de risco das obras (DEBNATH *et al.*, 2016) e classificar informações de quase acidentes presentes em relatórios de segurança (FANG, W. *et al.*, 2020).

Outro exemplo foi a pesquisa realizada por Fan (2020) que desenvolveu modelo para avaliar o risco de não conformidades em obras, estabelecendo seus fatores de risco e relação hierárquica para determinar a ordem de prioridade na gestão de segurança. Zhong *et al.* (2020) também desenvolveram modelo para identificar tópicos implícitos na descrição textual de registros das situações relacionadas à segurança da obra, demonstrando padrões de recorrência de perigos de forma automática e confiável.

Levando em consideração que a ocorrência dos acidentes de trabalho não é algo aleatório, técnicas de ML podem ser utilizadas para revelar e discutir potenciais padrões e tendências, podendo, inclusive, realizar previsões que contribuam com a gestão de segurança das obras (SARKAR *et al.*, 2019; ZHU *et al.*, 2021). A previsão, por sua vez, é uma questão importante para a segurança do trabalho, pois pode fornecer uma oportunidade para revelar problemas antes que eles ocorram (AYHAN; TOKDEMIR, 2019). Modelos precisos de previsão de fatores de risco podem ajudar os

gestores da indústria da construção a identificar riscos potenciais em estágio inicial, tornando possível a aplicação antecipada de medidas adequadas e realização de treinamentos com os trabalhadores (LI X. *et al.*, 2021). Nesse sentido, pesquisas focaram no desenvolvimento de modelos de predição de risco de acidentes (RIVAS *et al.*, 2011; TIXIER *et al.*, 2016b; LI X. *et al.*, 2021), dos níveis de fadiga física por meio do monitoramento das variações fisiológicas dos trabalhadores (ARYAL; GHAHRAMANI; BECERIK-GERBER, 2017), do resultado de incidentes (AYHAN; TOKDEMIR, 2019, 2020), do risco de acidentes fatais (CHOI *et al.*, 2020) e das consequências de lesões (ZHU *et al.*, 2021).

Dificuldades dos modelos

Apesar dos benefícios evidenciados, algumas dificuldades das aplicações de ML são relatadas na literatura, como a falta de padronização nos dados e rótulos de classificação, conjuntos de dados pequenos e incompletos, confiabilidade das informações, entre outros. Entretanto, soluções foram criadas para superar esses problemas e gerar resultados mais confiáveis, como por exemplo a padronização dos dados, a utilização de ferramentas de ontologia nas etapas de pré-processamento, o uso de bases de dados de órgãos oficiais e a fusão de dados de diferentes fontes (ZHANG J. *et al.*, 2020; LEE, J. *et al.*, 2020).

Um exemplo de solução foi proposta por Zhang *et al.* (2019) que utilizaram técnicas de mineração de dados e processamento de linguagem natural – NLP para analisar relatórios de acidentes na construção. O método criado apresentou a vantagem de não precisar de dados classificados, reduzindo demasiadamente o trabalho manual, sendo ideal para extrair padrões de pequeno grupo de documentos com estrutura gramatical similar. Outros autores também buscaram utilizar ferramentas ontológicas e de NLP para o problema dos dados não estruturados, como o caso de Tixier *et al.* (2016a) que criaram modelo capaz de escanear relatórios de acidentes de forma rápida e automática com alta precisão, bem como Chin, Lin e Hsieh (2014) e seu modelo de classificação ontológica para a área de segurança na construção, que contou com uma base para seu entendimento para auxiliar nas análises de risco do trabalho.

Em relação à quantidade de dados, Sakhakarmi, Park e Cho (2019) evidenciaram que a questão tem influência nas performances, pois a precisão de seu modelo aumentou consideravelmente com aumento dos dados. Ademais, a confiabilidade das informações também é um fator essencial para a qualidade dos resultados, dessa forma, pesquisas utilizaram informações oficiais de bases de dados governamentais com casos históricos de acidentes para realizar análises e previsões diversas. Alguns exemplos foram vistos em Liao, Perng e Chiang (2009) que utilizaram dados de acidentes oficiais de Taiwan, Chokor *et al.* (2016) e Li *et al.* (2021) com dados oficiais dos EUA e Zhu *et al.* (2021) com informações oficiais da China.

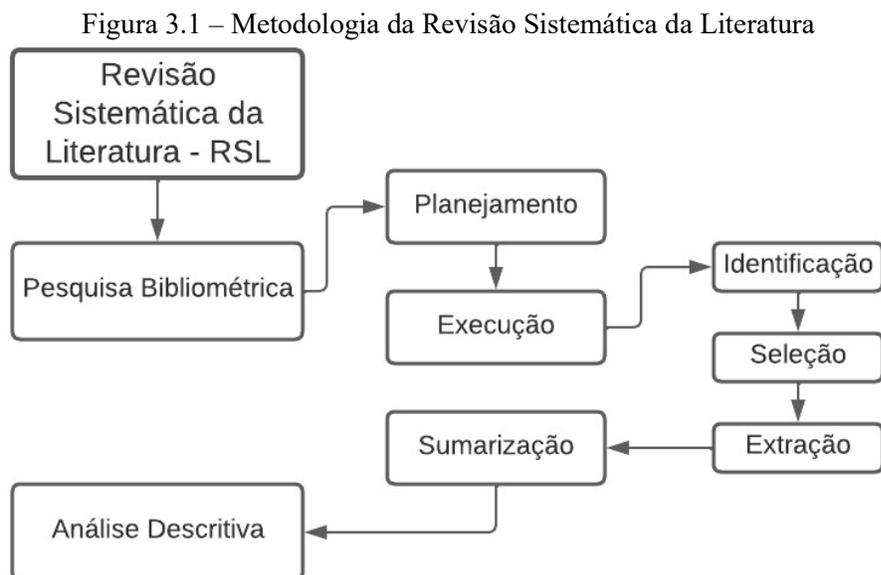
3 METODOLOGIA

A metodologia desenvolvida neste trabalho foi realizada em duas etapas. Inicialmente, foi executada uma RSL sobre o tema proposto para identificar as aplicações de ML que estão sendo desenvolvidas no mundo e suas principais características. Em seguida, foram obtidos dados de acidentes e doenças do trabalho de base de dados oficial brasileira, utilizando o aprendizado de máquina para analisar as informações e buscar padrões de acidentes.

3.1 Revisão Sistemática da Literatura

A RSL teve por objetivo analisar a literatura existente sobre *Machine Learning* aplicado à prevenção de acidentes do trabalho na Indústria da Construção Civil. Dessa forma, foram abordadas informações gerais sobre as publicações encontradas, além da descrição dos métodos e dos dados utilizados nos modelos.

A metodologia utilizada nesta RSL compreendeu duas fases (Figura 3.1). Na primeira etapa, foi realizada pesquisa bibliométrica, onde foram extraídos documentos de bases de dados pré-estabelecidas e que estivessem de acordo com critérios de seleção indicados. Em sequência, foi realizada análise descritiva da produção encontrada, apresentando o resumo dos resultados.



Fonte: Autora

3.1.1 Pesquisa bibliométrica

A pesquisa bibliométrica consistiu em um método de pesquisa utilizado para a investigação de publicações científicas relativas ao tema proposto. A etapa buscou realizar uma análise quantitativa da produção científica encontrada nas bases de dados internacionais. Nesta fase, foram realizados três processos: o planejamento do estudo, a execução da pesquisa e a sumarização dos resultados encontrados, todas realizadas com o auxílio das ferramentas *StArt* e *Excel*.

Planejamento

Inicialmente, foi realizado o planejamento do estudo através da elaboração de protocolo de pesquisa, onde foram especificadas as informações necessárias para o desenvolvimento do trabalho. No protocolo foram definidos os seguintes itens: título do trabalho, objetivos, questões principais de pesquisa, palavras-chave e sinônimos, além dos critérios de seleção de bases e estudos (exclusão e inclusão) e os resultados esperados com a revisão.

Para auxiliar na definição do título, questão de pesquisa e palavras-chave foi utilizada a estratégia *PICO* (AKOBENG, 2005). Apesar de normalmente ser utilizado em revisões sistemáticas da área das ciências naturais, o critério pode ser um aliado no processo de revisão sistemática, visto que as suas iniciais são acrônimos para população (P), interesse (I) e contexto (Co), tópicos que precisam ser identificados inicialmente na pesquisa para que a busca seja realizada de forma coerente.

A partir da definição do critério *PICO* e pergunta da pesquisa, foram selecionados os termos de busca, que consistiram nas palavras-chave e descritores referentes ao tema proposto. Além dos termos, foram estabelecidas as fontes de pesquisa, sendo selecionadas bases de dados de acordo com critérios de seleção e que pudessem ser acessadas pelo Portal de Periódicos da CAPES.

Após a definição das bases, foi gerada uma *string* de busca genérica com as palavras-chave e descritores previamente selecionados para que fosse feita a busca das publicações em cada base de dados. No entanto, como há diferença entre as fontes de pesquisa, em alguns casos a *string* precisou ser complementada com alguns sinônimos e operadores *booleanos*.

Com o objetivo de refinar os resultados da busca e verificar a qualidade das publicações, foram determinados critérios de limitação e seleção para os estudos encontrados. O critério de limitação, por sua vez, relacionou-se ao idioma dos documentos pesquisados, enquanto os critérios de seleção se referiram à inclusão e exclusão de publicações na revisão. Determinadas as questões acima, foi possível gerar um protocolo de pesquisa contendo informações necessárias para o processo de execução do trabalho.

Execução

A fase de execução compreendeu três subetapas: identificação das publicações nas bases; seleção dos estudos; e extração daqueles que poderiam responder à questão de pesquisa. Na fase de identificação dos estudos, foi realizada a busca das publicações a partir das *strings* adaptadas e genéricas inseridas nas bases de dados. Feito este processo, foi iniciada a seleção dos estudos, que levou em consideração os critérios de limitação, inclusão e exclusão (definidos no protocolo) para filtrar as publicações.

Após a definição dos estudos, foi realizado o processo de extração, onde foi feito um refinamento das publicações selecionadas, para que de fato fossem obtidos os trabalhos em concordância com o tema da revisão. Para ilustrar o processo, foi utilizado um modelo baseado no fluxograma dos Principais Itens para Relatar Revisões Sistemáticas e Meta-análises – PRISMA (PAGE et al., 2021).

Sumarização

De posse dos estudos extraídos da pesquisa, foi realizada a sumarização dos dados através do software *Excel*. No processo, foi possível obter um resumo das informações da pesquisa, além de visualizar os resultados através de gráficos. A sumarização também apresentou dados pertinentes para a compreensão do tema da revisão, relacionando as publicações com diversos aspectos, como: evolução anual; meios de publicação; autores; instituições; países e regiões; palavras-chave, etc.

3.1.2 Análise descritiva

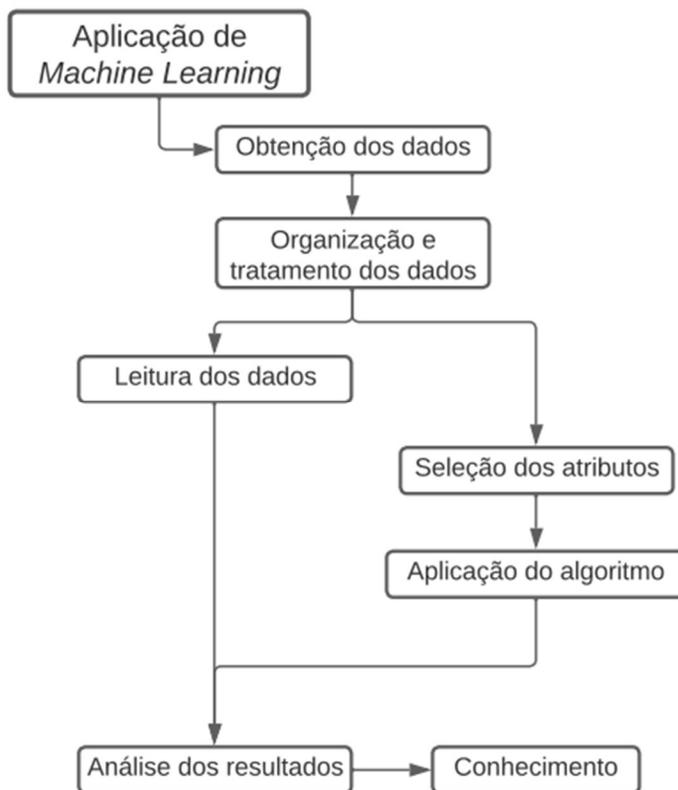
Após o resultado da pesquisa bibliométrica, deu-se início à análise descritiva dos dados encontrados, que teve o objetivo de complementar e trazer outras interpretações para as informações apresentadas. A análise se deu através da catalogação dos dados no *Excel* a respeito de cada publicação, como: finalidade das pesquisas, abordagens de ML utilizadas, método de coleta de dados, tipos de dados, entre outros. A catalogação, por sua vez, foi apresentada por meio de quadros e figuras.

3.2 Aplicação de Machine Learning

A segunda etapa da metodologia deste trabalho consistiu no uso de *Machine Learning* para analisar os acidentes de trabalho ocorridos na Indústria da Construção Civil no Brasil. O desenvolvimento do modelo foi motivado pela necessidade de se obter informações relevantes capazes de contribuir com a gestão de segurança dos canteiros de obras, como padrões de ocorrência de acidentes.

O modelo foi desenvolvido na linguagem de programação *Python*, por meio do *Spyder*, Ambiente de Desenvolvimento Integrado – IDE. O processo metodológico da aplicação foi dividido em 6 etapas, sendo elas: obtenção dos dados, organização e tratamento dos dados, leitura dos dados, seleção dos atributos, aplicação dos algoritmos e análise dos resultados. O fluxograma desta etapa pode ser observado na Figura 3.2 abaixo.

Figura 3.2 – Fluxograma da aplicação de *Machine Learning*



Fonte: Adaptado de Mutlu e Altuntas (2019)

3.2.1 Obtenção dos dados

Para o desenvolvimento do modelo proposto, foram utilizados dados oficiais de acidentes de trabalho apresentados por meio de CATs emitidas entre 2018 e 2021. Os registros foram obtidos de forma gratuita pelo site da Previdência Social.

3.2.2 Organização e tratamento dos dados

Os dados obtidos das CATs são disponibilizados em formato padrão, no entanto, é comum encontrar erros associados ao preenchimento desses documentos, bem como a presença de dados incompletos e outros problemas que podem prejudicar o desenvolvimento do modelo de *ML*. Dessa forma, a organização e tratamento dos dados foi fundamental para garantir menos variações nas informações e consequentemente melhores resultados (LEE, J. *et al.*, 2020).

Inicialmente, os dados presentes nos arquivos foram organizados em um único local, gerando uma base de dados com todos os dados obtidos. Como os dados das CATs foram disponibilizados de forma abrangente, com acidentes ocorridos em todos os setores produtivos, foi necessário também selecionar pela CNAE apenas aqueles que se relacionassem com a construção civil. Em seguida, foram realizados outros processos de correção e tratamento dos dados, onde foram excluídos e/ou modificados dados com informações nulas ou que apresentassem algum tipo de erro. Ademais, alguns atributos foram agrupados em classes para obter menos variações de informações.

3.2.3 Leitura dos dados

Após a organização e tratamento dos dados, foi realizada a leitura das informações obtidas. Esta etapa foi fundamental no processo, pois a partir dela foi possível analisar os dados de forma exploratória, obtendo um entendimento geral sobre o problema para dar continuidade ao desenvolvimento do modelo (MUTLU; ALTUNTAS, 2019).

3.2.4 Seleção dos atributos

A partir da análise exploratória dos dados, foram excluídos atributos não relevantes para o modelo. A seleção das informações foi necessária para inserir as características que vão representar os resultados de forma mais clara e eficaz (SARKAR *et al.*, 2019). Dessa forma, esta etapa colaborou para a obtenção de um *dataset* mais enxuto e mais objetivo.

3.2.5 Seleção do algoritmo

A etapa de seleção dos algoritmos foi necessária devido às particularidades de cada método para a análise e agrupamento de dados, situação que pode interferir diretamente na qualidade dos resultados (SARKAR; MAITI, 2020). Sendo assim, levando em consideração os resultados da revisão de literatura realizada e os tipos de dados utilizados na pesquisa, foi selecionado um método de aprendizagem não supervisionada para o desenvolvimento da ferramenta.

3.2.6 Análise dos resultados

Por fim, com as descobertas obtidas por meio da análise exploratória dos dados, juntamente com os resultados encontrados da aplicação dos algoritmos, pôde-se obter informações relevantes a respeito dos acidentes de trabalho ocorridos no Brasil referentes ao setor da Construção Civil que podem contribuir com o desenvolvimento de políticas de SST na construção.

4 RESULTADOS

Os resultados apresentados neste capítulo foram divididos em duas etapas, sendo a primeira parte referente ao resultado da RSL e a segunda parte relacionada aos resultados do desenvolvimento da aplicação de *ML* na análise dos dados de acidentes.

4.1 RSL

A revisão realizada neste trabalho foi executada em duas etapas, pesquisa bibliométrica e análise descritiva dos dados. A seguir, poderão ser visualizados os resultados referentes a cada processo.

4.1.1 *Análise bibliométrica*

4.1.1.1 *Protocolo de pesquisa*

A etapa de planejamento, conforme descrito na metodologia, foi responsável por gerar o protocolo de pesquisa (Quadro 4.1), documento onde foram apresentadas as informações necessárias para a execução da revisão. Baseado no problema evidenciado na introdução, chegou-se ao questionamento de: Quais aplicações de *Machine Learning (ML)* vêm sendo utilizadas na construção civil para a prevenção de acidentes de trabalho?

Inicialmente, foram determinadas as bases de dados, cujos critérios de seleção (CS) foram: período de cobertura maior que 10 anos; quantidade de títulos maior que 10.000; e relevância para a área de estudo. Sendo assim, optou-se pela escolha de 10 bases de dados, das quais 6 foram multidisciplinares e 4 específicas (áreas de engenharias e computação). As bases e as justificativas utilizadas para a escolha de cada uma delas podem ser visualizadas no Quadro 4.1:

Quadro 4.1 – Bases de dados utilizadas na RSL

<i>Association for Computing Machinery (ACM) Digital Library</i>	Base específica para a área de computação, a <i>ACM</i> possui período de cobertura de mais de 100 anos (início em 1908) e conta com quase 3 milhões de publicações.
<i>American Society of Civil Engineering (ASCE) Library</i>	A associação americana de engenheiros civis possui uma biblioteca com mais de 250 mil publicações na área de engenharia civil, além de contar com um período de cobertura de 1872 até o presente.
<i>Compendex (Engineering Village)</i>	A <i>Engineering Village</i> é uma plataforma de indexação de bases de dados que engloba trabalhos das engenharias e outras áreas relacionadas. A biblioteca <i>Compendex</i> , por sua vez, inclui cerca de 20 milhões de publicações de 190 disciplinas de engenharia, com período de cobertura de 1970 até a presente data.
<i>EBSCO Information Services</i>	A empresa <i>EBSCO</i> foi fundada em meados do século XX e atualmente possui biblioteca que abriga mais de 350 bases de dados. Através desta, é possível acessar publicações de diversas áreas do conhecimento, como tecnologia, ciências da saúde, entre outras.
<i>Institute of Electrical and Electronics Engineers (IEEE) Xplore</i>	O instituto dos engenheiros elétricos e eletrônicos possui uma extensa base de dados com conteúdos científicos e técnicos publicados pelo <i>IEEE</i> e parceiros. A biblioteca engloba mais de 5 milhões de documentos nas áreas de engenharia elétrica, ciência da computação, eletrônica e áreas correlatas, além de possuir período de cobertura de mais de 50 anos.
<i>Science Direct</i>	A <i>Science Direct</i> é uma plataforma que abriga publicações científicas das áreas de engenharias e ciências físicas, ciências humanas, ciências da saúde, entre outras. Criada ao final dos anos 90, a biblioteca apresenta cerca de 18 milhões de artigos e capítulos para consulta.
<i>Scopus</i>	A <i>Scopus</i> é uma base de dados multidisciplinar que abrange publicações das áreas da ciência, tecnologia, medicina, ciências sociais, artes, entre outras. Na biblioteca, podem ser encontrados cerca de 20 milhões de documentos datados a partir de 1788, o que faz com que a base seja de grande relevância para pesquisas no geral.
<i>SpringerLink</i>	A <i>Springer</i> é uma editora multidisciplinar que publica assuntos relacionados às áreas de engenharia, ciências humanas, ciências da saúde, educação, etc. A sua base de dados foi fundada no final dos anos 90 e é conhecida como <i>SpringerLink</i> , onde atualmente é possível consultar mais de 15 milhões de publicações científicas.
<i>Taylor & Francis Online</i>	A <i>Taylor & Francis</i> é uma editora multidisciplinar de publicação científica, que aborda as áreas de humanidades e ciências sociais, ciência, tecnologia, medicina, etc. Sua fundação é datada de 1852 e, atualmente, é possível acessar por ela mais de 100 mil publicações.
<i>Wiley Online Library</i>	A <i>Wiley</i> é uma editora criada ainda no século XIX que atualmente publica livros e periódicos. Nela é possível ter acesso à mais de 1600 periódicos de diversas áreas do conhecimento, como medicina, arquitetura, ciência da computação, etc.

Fonte: Autora

Após determinar as bases, foi utilizada a estratégia *PICo* (AKOBENG, 2005) para auxiliar no desenvolvimento da questão de pesquisa. Para o acrônimo população (P) foi considerada a indústria da construção civil; para o interesse (I), as aplicações de *Machine Learning*; e para o contexto (Co), a prevenção de acidentes de trabalho. A partir disto, foi possível determinar as palavras-chave e sinônimos para gerar as *strings* de busca utilizadas nas bases de dados. Dessa forma, foram utilizados nas *strings* os termos “*Construction Industry*”, “*Machine Learning*” e “*Accident prevention*” e os sinônimos “*Construction*”, “*Civil Engineering*”, “*Deep Learning*”, “*Data Science*”, “*Learning System*”, “*Learning Algorithms*” e “*Injury prevention*”.

Com os termos e sinônimos selecionados, foi gerada uma *string* de busca genérica para encontrar os estudos nas bases de dados. No entanto, para a *Science Direct* e *SpringerLink*, foi necessário adaptar a *string* de acordo com a especificidade de cada biblioteca.

Em relação à pesquisa, também foram considerados fatores de limitação (LI). Sendo assim, foram incluídos apenas documentos que estivessem no idioma inglês, publicados até o mês de julho de 2021, data em que a busca foi realizada. Além disso, foram selecionados para análise estudos dos tipos quantitativo, qualitativo, observacional, descritivo e experimental.

Dando continuidade à pesquisa, foram estabelecidos os critérios de inclusão (CI) e exclusão (CE) dos estudos. A pergunta norteadora, definida no início do protocolo, foi utilizada como critério de inclusão, isto é, os estudos que pudessem responder ao questionamento inicial foram incluídos na revisão. Os critérios de exclusão, por sua vez, se relacionaram a vários fatores, como: publicações duplicadas; artigos de revisão; artigos que não fossem publicados em periódicos; artigos sem acesso disponível; publicações sem contar com os termos de busca no título, resumo e palavras-chave; estudos sem apresentar a performance das abordagens de ML; trabalhos que não fossem sobre a construção civil; e trabalhos que não tivessem o foco da prevenção de acidentes. Após o processo apresentado, obteve-se o protocolo de pesquisa, conforme Quadro 4.2 abaixo.

Quadro 4.2 – Protocolo de pesquisa

Elementos	Definição
Critério PICO	População (P) – Construção Civil; Interesse (I) – Aplicações de <i>Machine Learning</i> ; Contexto (Co) – Prevenção de acidentes de trabalho.
Pergunta de pesquisa	Quais as aplicações de <i>Machine Learning (ML)</i> que vem sendo utilizadas na construção civil para prevenir acidentes de trabalho?
Termos de busca	Construção Civil. <i>Machine Learning</i> . Prevenção de acidentes.
Critério de seleção das bases (CS)	CS-01: Período de cobertura maior que 10 anos; CS-02: Quantidade de títulos maior que 1.000; CS-03: Relevância para a área de estudo.
Bases de dados selecionadas	<i>ACM Digital Library; American Society of Civil Engineering – ASCE; Compendex (Engineering Village); EBSCO; IEEE Xplore; Science Direct; Scopus; SpringerLink; Taylor & Francis Online; Wiley Online Library.</i>
Limitação de pesquisa (LI)	LI-01: Documentos no idioma inglês; LI-02: Documentos até a data de julho de 2021.

Fonte: Autora

Quadro 4.2 – Protocolo de pesquisa (conclusão)

String genérica	("CONSTRUCTION INDUSTRY" OR "CONSTRUCTION" OR "CIVIL ENGINEERING") AND ("MACHINE LEARNING" OR "DEEP LEARNING" OR "DATA SCIENCE" OR "LEARNING SYSTEM" OR "LEARNING ALGORITHMS") AND ("ACCIDENT PREVENTION" OR "INJURY PREVENTION")
String adaptada	<u>Science Direct</u> : ("Construction Industry" OR "construction" OR "civil engineering") AND ("Machine Learning" OR "Deep Learning" OR "Learning System" OR "Learning Algorithm") AND ("Accident prevention" OR "Injury prevention"); <u>SpringerLink</u> : ("Construction Industry" OR "construction") AND ("Machine Learning" OR "Deep Learning" OR "Learning System") AND ("Accident prevention" OR "Injury prevention"). *As demais bases utilizaram a string genérica.
Tipos de estudos	Quantitativo; Qualitativo; Observacional; Descritivo; Experimental.
Critério de inclusão (CI)	CI-01: Documentos que respondam a questão de pesquisa.
Critério de exclusão (CE)	CE-01: Documentos duplicados; CE-02: Documentos de revisão; CE-03: Documentos não publicados em periódicos; CE-04: Documentos sem acesso disponível; CE-05: Documentos sem os termos de busca no título, resumo e/ou palavras-chave; CE-06: Documentos sem apresentar performance de aplicações de ML; CE-07: Documentos que não sejam sobre o setor de construção; CE-08: Documentos sem aplicações focadas na prevenção de acidentes.

Fonte: Autora

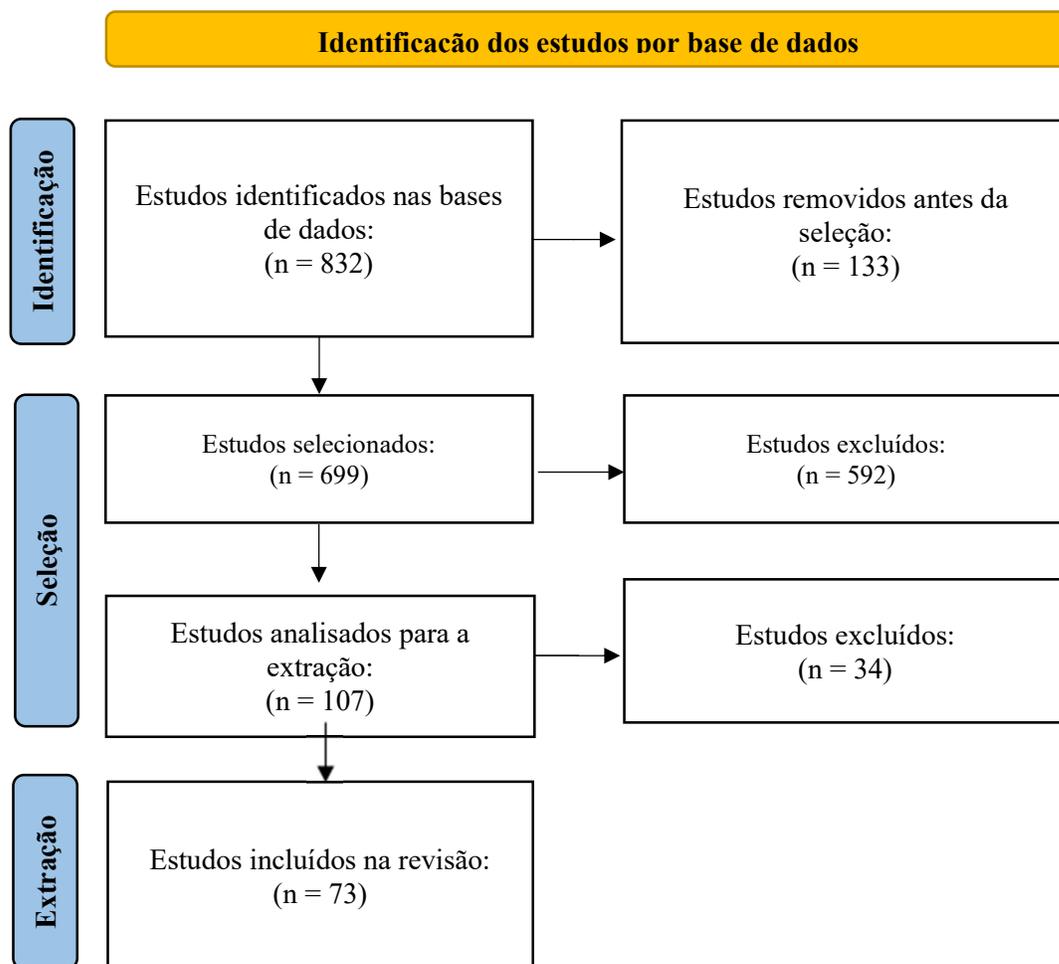
4.1.1.2 Diagrama de fluxo

A etapa de execução da pesquisa, apresentada na metodologia, consistiu em 3 etapas: identificação, seleção e extração. A princípio, os estudos foram identificados nas bases de dados e exportados para o software *StArt*. Inicialmente foram identificados um total de 832 títulos, no entanto, 133 destes foram excluídos manualmente por representarem trabalhos duplicados, restando um total de 699 publicações para a fase de seleção.

Na seleção dos estudos, foram observados o título, resumo e palavras-chave dos documentos e, após a aplicação dos critérios de exclusão, foram excluídas 592 publicações pelo não enquadramento ao proposto, restando 107 estudos. Foi realizada uma leitura completa dos 107 artigos e, após esse processo, outros 34 documentos foram excluídos pelos critérios de exclusão, restando ao final 73 estudos incluídos nesta revisão.

Por fim, os 73 artigos tiveram suas informações extraídas para que fosse possível realizar os processos de sumarização e análise descritiva dos dados. O processo de execução da revisão foi ilustrado no modelo do fluxograma PRISMA, conforme consta na Figura 4.1.

Figura 4.1 – Fluxograma PRISMA



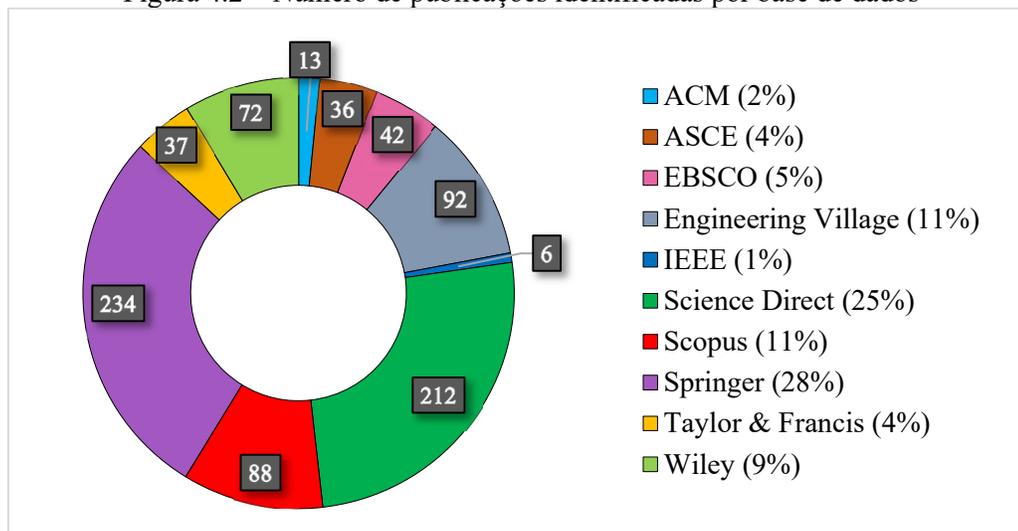
Fonte: Adaptado de Page *et al.* (2021).

4.1.1.3 Visualização gráfica

Na etapa de sumarização, foi possível obter um resumo do processo de extração dos estudos através da visualização de gráficos. Foi possível identificar que a *Springer* (28%) e a *Science Direct* (25%) foram as bases com maior número de publicações identificadas, 234 e 212, respectivamente. Juntas, as duas bases representaram mais da metade dos textos selecionados, um total de 446 (53%) documentos. Também foi observado que a base *ACM* (2%) e a *IEEE* (1%) obtiveram o menor

número de publicações encontradas, representando 19 (3%) documentos. A Figura 4.2 abaixo apresenta o percentual e o número de publicações encontradas em cada base de dados.

Figura 4.2 – Número de publicações identificadas por base de dados

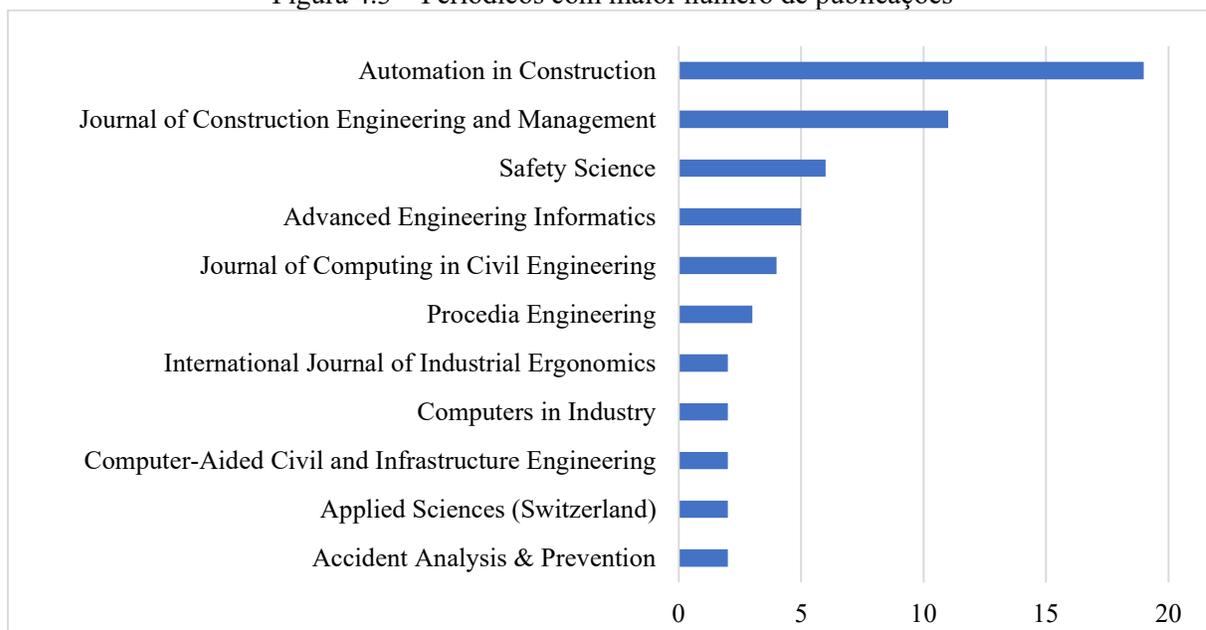


Fonte: Autora

Após a extração dos estudos, foi possível visualizar a quantidade de publicações extraídas por cada base de dados. Sendo assim, foi identificado que a base *Science Direct* foi responsável pelo maior número de documentos, com 29 estudos (40%) incluídos na revisão, seguida da *Engineering Village*, com 15 publicações (21%) e da *Scopus*, com 10 artigos (14%) incluídos. Estas três bases juntas corresponderam a 75% dos documentos extraídos. Também foi possível identificar que a base *Wiley* obteve o menor número de publicações, com apenas 1 estudo (1%) incluído na revisão, enquanto as bibliotecas da *Taylor & Francis*, *ACM* e *IEEE* não obtiveram estudos incluídos.

Também foi observado que 73 artigos incluídos na revisão foram publicados em 26 revistas científicas diferentes. Cerca de 40% dos periódicos (11 revistas) foram responsáveis por mais de uma publicação (Figura 4.3). O maior destaque foi visto no periódico *Automation in Construction*, cujo número de artigos chegou à 19. Também se destacaram os periódicos *Journal of Construction Engineering and Management* e *Safety Science*, responsáveis por 11 e 6 publicações, respectivamente.

Figura 4.3 – Periódicos com maior número de publicações

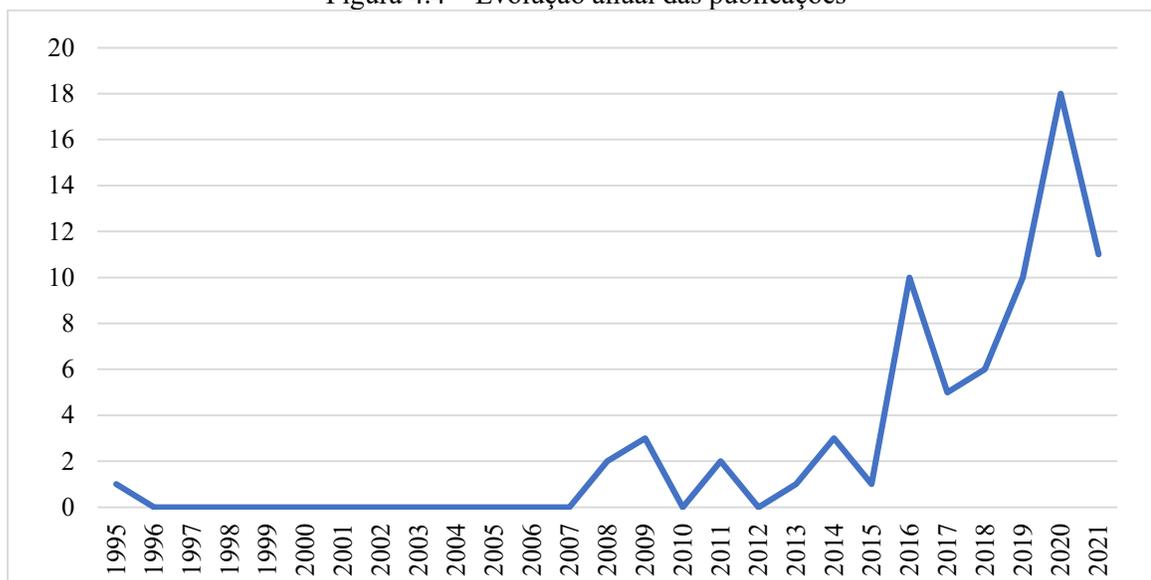


Fonte: Autora

Além das questões acima, verificou-se o ano de publicação dos estudos, onde foi possível identificar que o primeiro artigo sobre a temática foi publicado em 1995 e, após um hiato de 12 anos sem pesquisas publicadas na área, a partir de 2008 houve uma curva crescente na quantidade de artigos. Em 2016, especialmente, verificou-se um pico de publicações, passando de 1 do ano anterior (2015) para 10.

Outra evolução importante se deu no ano de 2020, onde foram encontrados 18 artigos do tema proposto. Como a pesquisa foi realizada até o mês de julho de 2021, não foi possível identificar o total de publicações do ano completo, no entanto, o número de estudos até a metade de 2021 (11) ultrapassou a quantidade de artigos publicados do ano de 2019 (10), o que mostra uma tendência de crescimento das publicações relacionadas ao tema. A Figura 4.4 abaixo apresenta a evolução anual dos artigos incluídos na revisão.

Figura 4.4 – Evolução anual das publicações

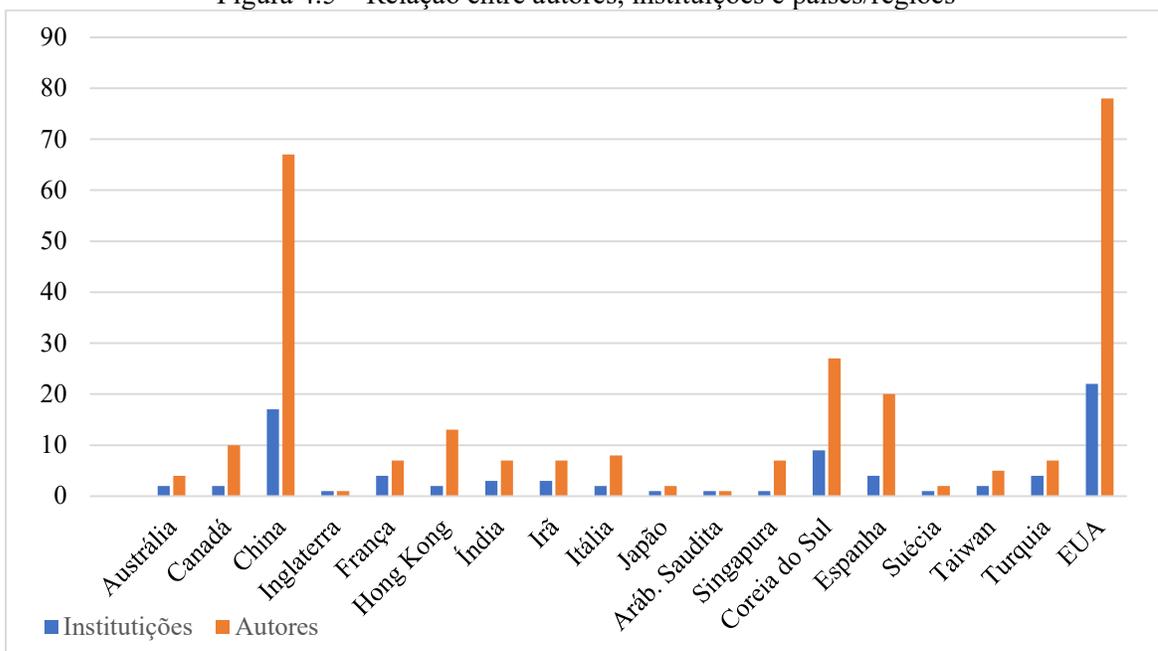


Fonte: Autora

Outra questão levantada, foi a relação entre autores, instituições e países/regiões. Observou-se um total de 18 países/regiões com artigos publicados na temática, sendo os EUA o de maior número de autorias, representando cerca de 30% do total de autores (78). O segundo país com maior taxa de autoria nos trabalhos foi a China, com 40 autores (25%). Os dois países juntos somam 53% das autorias dos estudos incluídos na revisão.

Países como Coréia do Sul, Espanha, Taiwan e a região de Hong Kong também se destacaram na pesquisa, com número de autores variando entre 10 e 20 e percentual de participação entre 5 e 10%, para cada local. Apesar dos dados encontrados apresentarem 18 países/regiões com pesquisas nessa temática, o Brasil não esteve presente na listagem. A Figura 4.5 abaixo apresenta a quantidade de autores e instituições de acordo com os países/regiões.

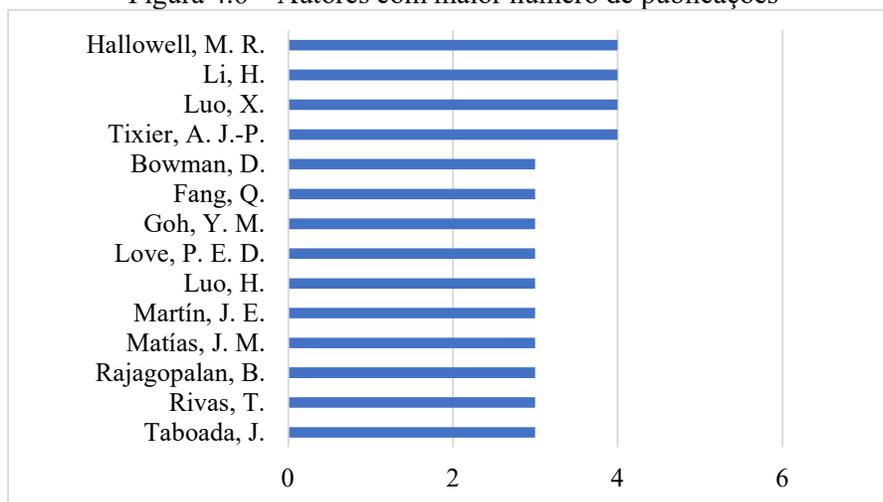
Figura 4.5 – Relação entre autores, instituições e países/regiões



Fonte: Autora

Em relação ao número de autores, observou-se um total de 273 participantes. Cerca de 13% (36) dos autores possuíam mais de uma publicação, dentre estes, 10 autores com 3 artigos cada, e 4 autores com 4 artigos cada. Os autores Hallowell, M. R., Li, H., Luo, X. e Tixier, A. J. -P., representantes de EUA, Hong Kong e EUA, respectivamente, foram os autores com maior número de publicações, com 4 artigos cada. A Figura 4.6 apresenta os autores com mais de 3 publicações.

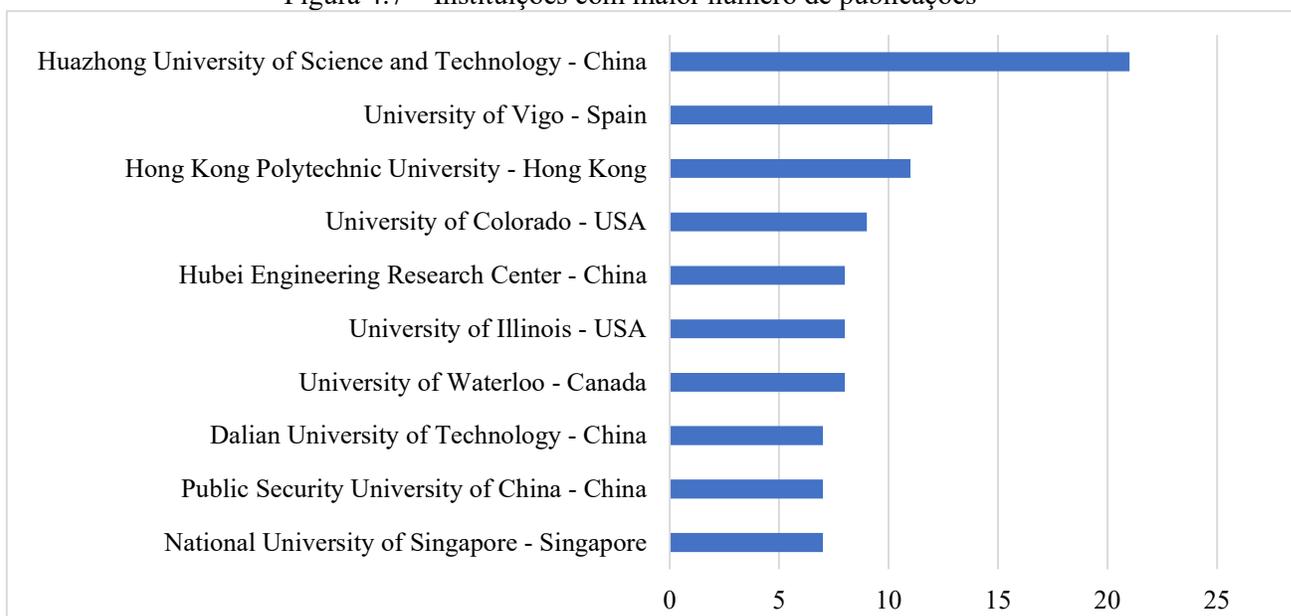
Figura 4.6 – Autores com maior número de publicações



Fonte: Autora

Foram observadas 81 instituições no total, sendo a Universidade de Huazhong na China a mais citada (6 vezes) e com maior número de autores (21). Os destaques também se deram para a Universidade de Vigo, na Espanha, e a Politécnica de Hong Kong, com 12 e 11 autores cada. A Figura 4.7 abaixo apresenta a classificação das 10 instituições com maior número de publicações.

Figura 4.7 – Instituições com maior número de publicações



Fonte: Autora

Por fim, foi gerada uma nuvem com as palavras-chave utilizadas nos trabalhos extraídos. Foi identificado um total de 230 palavras-chave nos artigos, cujo termo mais recorrente foi “*Machine Learning*”, com 19 repetições. Destacam-se ainda os termos “*Construction Safety*”, com 15 repetições, “*Deep Learning*” com 15, além de “*Safety*” com 10 repetições e “*Accident Prevention*”, “*Data Mining*” e “*Natural Language Processing*”, com 6 repetições cada. A Figura 4.8 abaixo apresenta a nuvem de palavras gerada através do site *WordArt*.

Quadro 4.3 – Métodos de ML utilizados nos artigos estudados

Métodos de ML	Artigos	Qtd. ML
SVM, NB, LR, CNN, LSTM, BiLSTM, NLP e C-BLSTM	Zhang <i>et al.</i> (2020)	8
LR, DT, SVM, NB, KNN, RF, MLP e AutoML	Zhu <i>et al.</i> (2021)	8
CART, C4.5, MLP, KNN, LR, NB, SVM e LDA	Bonifazi <i>et al.</i> (2021)	8
DRSA, DT, RF, NB, SVM, LR e MLP	Singh, Misra (2020)	7
SVM, NB, DT, RF, KNN, CNN e LSTM	Zhao, Obonyo (2020)	7
KNN, DT, MLP, NB, SVM, LR e RF	Li <i>et al.</i> (2021)	7
SVM, KNN, DT, LR, NB e SQP	Zhang <i>et al.</i> (2019)	6
DT, RF, LR, KNN, SVM e NB	Goh, Ubeynarayana (2017)	6
LR, BN, SVM, DR e CT	Rivas <i>et al.</i> (2011)	5
DT, RF, LR, KNN e SVM	Poh, Ubeynarayana, Goh (2018)	5
NLP, CNN, HAN, TF-IDF e SVM	Baker <i>et al.</i> (2016)	5
SVM, DT, PCA e LCCA	Lee <i>et al.</i> (2020)	4
GPDM, PCA, SVM e DTW	Han, Lee, Peña-Mora (2014)	4
LR, DT, RF e Adaboost	Choi <i>et al.</i> (2020)	4
Faster RCNN, RPN, RoIAlign e SGD	Tang, Roberts, Golparvar-Fard (2020)	4
DT, SVM, KNN e BT	Lee <i>et al.</i> (2021)	4
NB, KNN, SVM e C4.5	Zhao <i>et al.</i> (2018)	4
BN, ELM, SVM e CT	Matías <i>et al.</i> (2008)	4
YOLO, Faster R-CNN, R-FCN e SSD	Xiao e Kang (2021)	4
K-MODES, CART e RBF	Chen e Luo (2016)	3
BT, DT e SVM	Aryal, Ghahramani, Becerik-Gerber (2017)	3
MAPSAC, SURF e STF	Brilakis, Park, Jog (2011)	3
NLP, PCA e SVM	Chi, Lin, Hsieh (2014)	3
HOG, SVM e KNN	Wu, Zhao (2018)	3
LDA, CNN e SVM	Zhong <i>et al.</i> (2020)	3
SVM, NB e ME	Salama, El-Gohary (2013)	3
MLP, SVM e CNN	Kolar, Chen e Luo (2018)	3
K-MEANS, NLP e TF-IDF	Chokor <i>et al.</i> (2016)	3
YOLO e SN	Son e Kim (2021)	2
CHAID e C5.0	Mistikoglu <i>et al.</i> (2015)	2
K-MEANS e SVM	Alwasel <i>et al.</i> (2017)	2
ANN e FIS	Ciarapica, Giacchetta (2009); Ayhan, Tokdemir (2019)	2
Mask R-CNN e CPN	Fang <i>et al.</i> (2020)	2
ANN e LCCA	Ayhan, Tokdemir (2020)	2
ARM e BN	Fan (2020)	2
LSTM e MDN	Tang <i>et al.</i> (2020)	2
Faster RCNN e FIS	Zhang <i>et al.</i> (2020)	2
CNN e LSTM	Ding <i>et al.</i> (2018)	2
RF e SGTB	Tixier <i>et al.</i> (2016)	2
SSD e CNN	Fang <i>et al.</i> (2018)	2
GM e HCPC	Tixier <i>et al.</i> (2017)	2
CNN	Liu <i>et al.</i> (2021); Li, Luo, Skitmore (2020); Siddula <i>et al.</i> (2016); Shen <i>et al.</i> (2020)	1
ANFIS	Debnath <i>et al.</i> (2016); Jahangiri, Solukloei, Kamalinia (2019)	1

Fonte: Autora

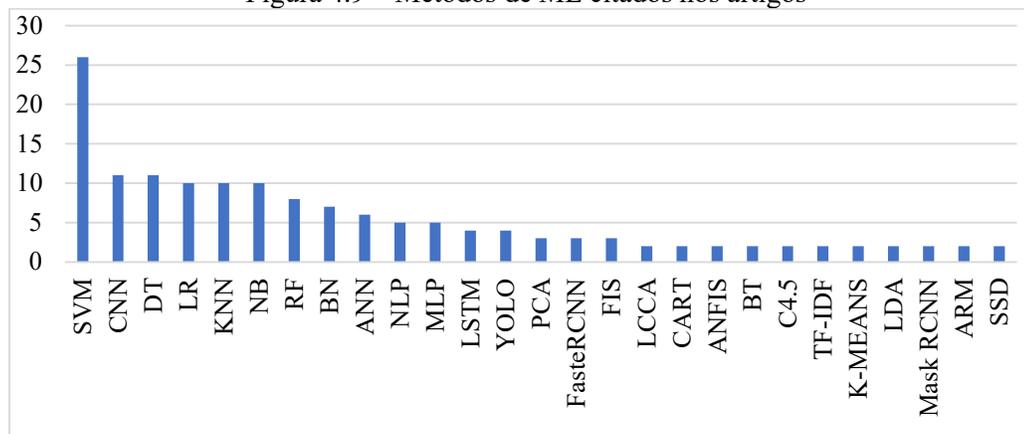
Quadro 4.3 – Métodos de ML utilizados nos artigos estudados (conclusão)

Métodos de ML	Artigos	Qtd. ML
BN	Mariscal <i>et al.</i> (2019); Martín <i>et al.</i> (2009); Mohammadfam <i>et al.</i> (2017); Abdat <i>et al.</i> (2014)	1
NLP	Tixier <i>et al.</i> (2016)	1
TD	Dutta <i>et al.</i> (2020)	1
YOLO	Chen e Demachi (2021); Huang <i>et al.</i> (2021)	1
Clustering	Assaad, El-adaway (2021)	1
SVM	Ryu <i>et al.</i> (2020); Sakhakarmi <i>et al.</i> (2019); Yang <i>et al.</i> (2016)	1
CRF	Kim, Chi (2019)	1
ANN	Lim <i>et al.</i> (2016); Zhang, Cao, Zhao (2019); Cao, Goh (2019)	1
BERT	Fang <i>et al.</i> (2020)	1
KNN	Lee <i>et al.</i> (2020)	1
BP	Feng <i>et al.</i> (2019)	1
HN	Yu <i>et al.</i> (2019)	1
RNN	Park <i>et al.</i> (2017)	1
ROUGH	Arciszewski <i>et al.</i> (1995)	1
Mask R-CNN	Khan <i>et al.</i> (2021)	1
OG	Teizer (2008)	1
ARM	Liao <i>et al.</i> (2009)	1

Fonte: Autora

Foi observado que o método SVM obteve maior recorrência entre os demais, sendo aplicado em 26 pesquisas diferentes. O destaque também se deu nos modelos de CNN e DT, presentes em 11 artigos cada, e LR, KNN e NB, em 10 trabalhos cada. Do total de métodos (67) encontrados, 27 algoritmos foram utilizados em mais de 1 pesquisa, representando um percentual de 40%, como pode ser visualizado na Figura 4.9 abaixo.

Figura 4.9 – Métodos de ML citados nos artigos



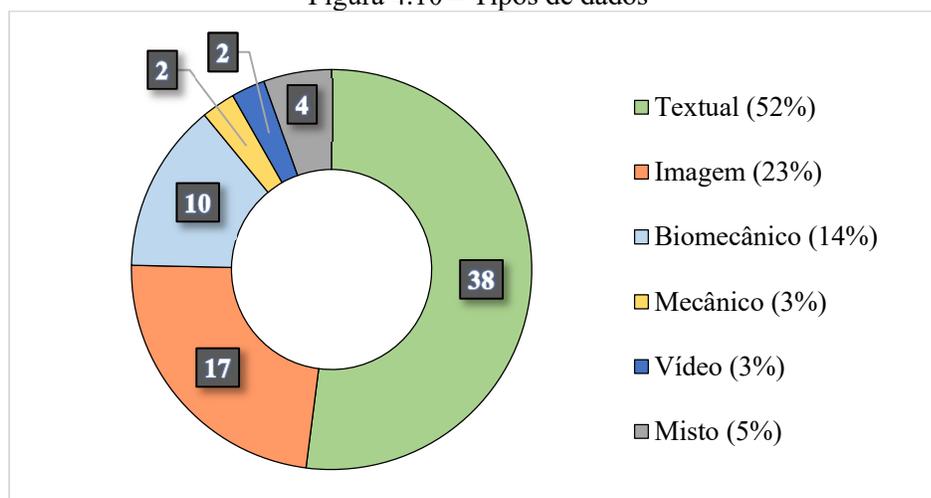
Fonte: Autora

4.1.2.2 Tipos de dados

Os tipos dos dados utilizados nas pesquisas também foram observados durante a análise. Foram encontrados 6 tipos: textual (ex.: comunicação e relatório de acidente, questionário, etc.), imagem (ex.: imagens de máquina, trabalhador, etc.), vídeo (ex.: gravações de canteiros de obras, vídeos de experimentos, etc.), mecânico (ex.: dados de ruptura de estrutura), biomecânico (ex.: dados de movimento na realização de atividade por trabalhador, etc.) e misto (compilado de dados de 2 ou mais tipos diferentes).

Foi observado que mais da metade (52%) das pesquisas realizadas utilizaram dados textuais nos modelos de *ML* aplicados, onde o destaque se deu para as narrativas de acidentes, questionários e relatórios de inspeção de obras. Cerca de 23% das pesquisas, representando 17 artigos, utilizaram dados de imagens (trabalhadores, máquinas, situações de risco, entre outras) para o desenvolvimento dos modelos de *ML*, enquanto 10 trabalhos (14%) utilizaram dados biomecânicos (dados fisiológicos de trabalhadores, dados de movimento, etc.). Também foi observado que os tipos de dados vídeo (3%) e mecânico (3%) foram utilizados em 2 estudos cada, além de 5% das pesquisas utilizarem mais de um tipo de dado. A Figura 4.10 apresenta o percentual encontrado para cada tipo.

Figura 4.10 – Tipos de dados



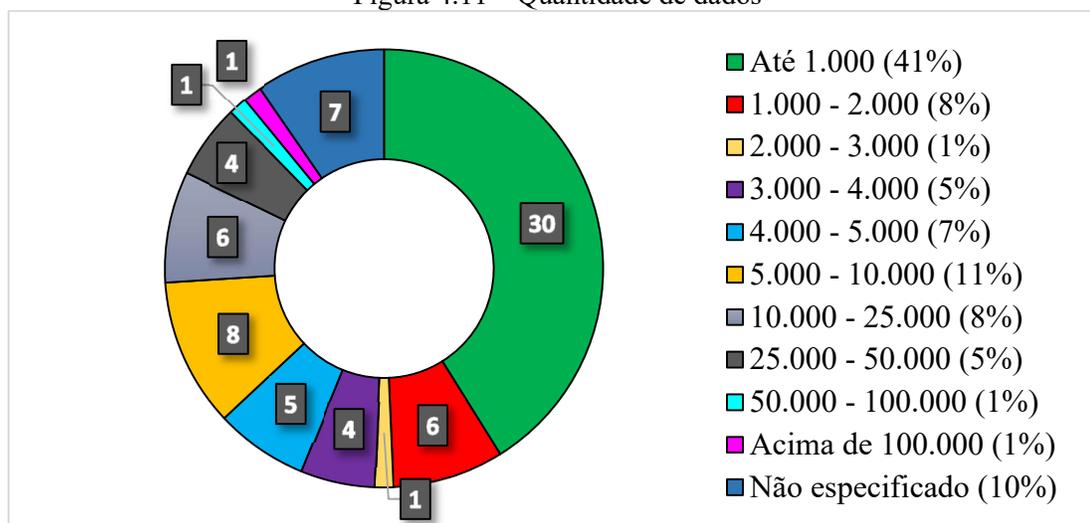
Fonte: Autora

4.1.2.3 Quantidade de dados

Os artigos foram analisados em relação à quantidade de dados utilizada para o desenvolvimento do modelo proposto. Dessa forma, foram estabelecidos intervalos para melhor compreensão dos resultados, sendo eles apresentados na Figura 4.11.

Foi possível identificar que cerca de 41% dos modelos, representando 30 artigos, utilizaram até 1.000 dados na pesquisa, enquanto 11%, segundo maior percentual encontrado (8 artigos), utilizou um número entre 5.000 e 10.000 dados. Também foi observado que apenas 1 estudo utilizou mais de 100.000 dados para o treinamento do modelo.

Figura 4.11 – Quantidade de dados

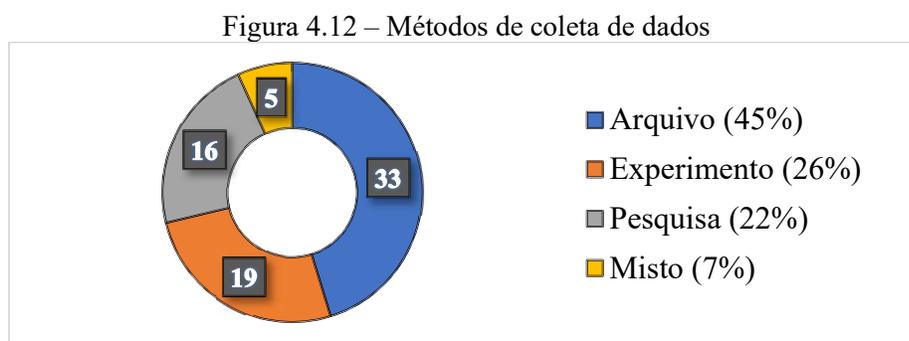


Fonte: Autora

4.1.2.4 Métodos de coleta de dados

As publicações foram analisadas quanto ao método de coleta de dados. Nesta análise, foram encontrados 5 tipos, sendo eles: pesquisa, referente aos dados primários (ex.: pesquisa, relatório publicado por agência, dados de empresa, etc.); arquivo, referente aos dados secundários (ex.: dados coletados antes da pesquisa, arquivos e documentos eletrônicos em bases de dados, etc.); experimento, referente aos casos de experimentos realizados, tanto em laboratório quanto em campo; mistos, referente aos estudos com 2 ou mais métodos diferentes de coleta de dados.

Foi observado que 45% dos estudos, representando 33 artigos, utilizaram dados secundários para o desenvolvimento dos modelos, sendo, em sua maioria, dados eletrônicos obtidos de agências governamentais, como a *Occupational Safety and Health Administration* (OSHA). Também foi observado que 19 artigos (26%) coletaram os dados por meio de experimentos realizados em canteiros de obras e em laboratórios, enquanto 16 artigos (22%) utilizaram dados primários, em sua maioria coletados de pesquisas realizadas em empresas de construção. Além disso, alguns artigos, representando 7% do total, utilizaram mais de um método de coleta de dados. A Figura 4.12 abaixo apresenta o percentual encontrado para cada tipo.



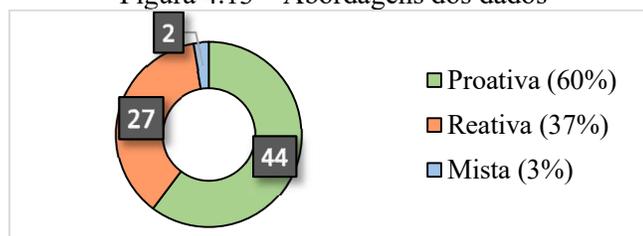
Fonte: Autora

4.1.2.5 Abordagens dos dados

A abordagem dos dados foi analisada e seguiu os conceitos de segurança sobre abordagem proativa e reativa. Dessa forma, foram identificados 3 tipos de abordagem nos estudos, sendo elas: abordagem reativa, referente aos dados coletados após a ocorrência de um acidente (ex.: dados de acidentes de trabalho, frequência de lesões, número de dias perdidos, etc.); abordagem proativa, referente aos dados coletados antes da ocorrência de um acidente (ex.: relatórios de inspeção, relatórios de auditoria, imagens e vídeos de situações perigosas, etc.); abordagem mista, referente aos dados mistos, coletados antes e após a ocorrência de um acidente.

Foi observado que cerca de 60% (44) dos 73 artigos incluídos nesta revisão utilizaram dados proativos para desenvolvimento do modelo proposto. Em relação ao restante dos artigos, cerca de 37% (27) utilizou uma abordagem reativa, enquanto cerca de 3% (2) utilizaram ambas. A Figura 4.13 apresenta o percentual encontrado para cada tipo de abordagem utilizada.

Figura 4.13 – Abordagens dos dados



Fonte: Autora

4.2 Análise de dados com *Machine Learning*

Os resultados da análise dos acidentes de trabalho foram apresentados em seis etapas, sendo elas a entrada dos dados, pré-processamento, análise exploratória, seleção de atributos, aplicação do algoritmo de *clustering* e obtenção de grupos de acidentes. A seguir, poderão ser visualizados os resultados referentes a cada processo.

4.2.1 Entrada dos dados

Para o desenvolvimento do modelo de ML proposto, foram utilizados dados de acidentes de trabalho ocorridos no Brasil no período de julho de 2018 a junho de 2021, registrados por meio de CATs. Os dados foram obtidos de forma gratuita pelo site da Previdência Social, sendo disponibilizados arquivos em formato “csv”. No entanto, devido a indisponibilidade de dados anteriores a 2018, só foi possível utilizar 3 anos de ocorrências.

Inicialmente, os registros foram unificados em uma tabela geral, sendo identificadas 1.331.098 ocorrências. A Figura 4.14 abaixo apresenta o *dataset* inicial utilizado no modelo, enquanto o Quadro 4.4 apresenta as informações sobre cada um dos 25 atributos iniciais observados nas CATs, como descrição, tipo da informação e quantidade de valores possíveis.

Figura 4.14 – *Dataset* inicial

Dimensão dados iniciais: (1331098, 25)
 Registros dados iniciais: 1331098
 Colunas dados iniciais: 25

Fonte: Autora

Quadro 4.4 – Atributos iniciais das CATs

Nº	Atributos	Descrição	Tipo	Valores possíveis
1	Agente Causador Acidente	Indica o agente ou situação que ocasionou o acidente (ex. esforço excessivo, ruído, máquina/equipamento, etc.)	Categórico	573
2	Data Acidente	Indica o mês e ano da ocorrência do evento, segue o formato AAAAMM (ano,mês)	Numérico	56
3	CBO	Indica o código da ocupação do acidentado, baseado na Classificação Brasileira de Ocupação (CBO)	Numérico	5500
4	CBO.1	Indica a descrição da ocupação do acidentado, baseada na Classificação Brasileira de Ocupação (CBO)	Categórico	4166
5	CID-10	Indica o código da doença do acidentado, baseado na 10ª revisão da Classificação Internacional de Doenças (CID-10)	Categórico	9481
6	CID-10.1	Indica a descrição da doença do acidentado, baseada na 10ª revisão da Classificação Internacional de Doenças (CID-10)	Categórico	9174
7	CNAE 2.0 Empregador	Indica o código do setor econômico do empregador, baseado na Classificação Nacional de Atividades Econômicas (CNAE)	Numérico	954
8	CNAE 2.0 Empregador.1	Indica a descrição do setor econômico do empregador, baseado na Classificação Nacional de Atividades Econômicas (CNAE)	Categórico	1014
9	Emitente CAT	Indica o responsável pela emissão da CAT	Categórico	12
10	Espécie do benefício	Indica o tipo de benefício recebido pelo acidentado/dependente	Categórico	12
11	Filiação Segurado	Indica se o acidentado é empregado registrado ou funcionário avulso	Categórico	9
12	Indica acidente	Indica se o evento teve como consequência a morte do acidentado	Categórico	5
13	Munic Empr	Indica o município do empregador	Categórico	9806
14	Natureza da Lesão	Indica o tipo da lesão (ex. fratura, contusão, corte, etc.)	Categórico	61
15	Origem de Cadastramento CAT	Indica a forma de cadastro da CAT	Categórico	3
16	Parte Corpo Atingida	Indica a parte do corpo do acidentado que foi atingida devido ao evento registrado (ex. cabeça, mão, dedo, etc.)	Categórico	90
17	Sexo	Indica o sexo do acidentado	Categórico	9
18	Tipo do Acidente	Indica o tipo de acidente, se doença, típico ou trajeto	Categórico	9
19	UF Munic. Acidente	Indica o estado em que ocorreu o acidente	Categórico	40
20	UF Munic. Empregador	Indica o estado do empregador	Categórico	66
21	Data Afastamento	Indica a data em que o acidentado foi afastado de suas atividades, segue o formato AAAA/MM (ano/mês)	Numérico	44
22	Data Despacho Benefício	Indica a data de despacho do benefício, segue o formato AAAA/MM (ano/mês)	Numérico	49
23	Data Acidente.1	Indica a data do evento, segue o formato DD/MM/AAAA (dia/mês/ano)	Numérico	1277
24	Data Nascimento	Indica a data de nascimento do acidentado, segue o formato DD/MM/AAAA (dia/mês/ano)	Numérico	21233
25	Data Emissão CAT	Indica a data em que a comunicação do acidente foi efetuada, segue o formato DD/MM/AAAA (dia/mês/ano)	Numérico	1096

Fonte: Autora

4.2.2 Pré-processamento dos dados

A etapa de pré-processamento se referiu à organização e tratamento dos dados obtidos das CATs. Inicialmente, os dados foram filtrados de acordo com os códigos da CNAE para abordar apenas os registros que tratassem de acidentes de trabalho do setor da construção. Dessa forma, foi utilizada a seção F (construção) e suas divisões 41 (construção de edifícios), 42 (obras de infraestrutura) e 43 (serviços especializados para construção), obtendo ainda 8 grupos e 16 classes de atividades, que podem ser observadas no Quadro 4.5 abaixo.

Quadro 4.5 – Lista de códigos CNAEs referentes à Construção Civil.

Códigos CNAE – Construção Civil				
Seção	Divisão	Grupo	Classe	Denominação
F	41	41.2	41.20-4	Construção de edifícios
	42	42.1	42.11	Construção de rodovias e ferrovias
			42.12	Construção de obras-de-arte especiais
			42.13	Obras de urbanização - ruas, praças e calçadas
		42.2	42.22-7	Construção de redes de abastecimento de água, coleta de esgoto e construções correlatas
		42.9	42.99-5	Obras de engenharia civil não especificadas anteriormente
	43	43.1	43.11-8	Demolição e preparação de canteiros de obras
			43.12-6	Perfurações e sondagens
			43.13-4	Obras de terraplenagem
			43.19-3	Serviços de preparação do terreno não especificados anteriormente
		43.2	43.21-5	Instalações elétricas
			43.22-3	Instalações hidráulicas, de sistemas de ventilação e refrigeração
			43.29-1	Obras de instalações em construções não especificadas anteriormente
		43.3	43.30-4	Obras de acabamento
		43.9	43.91-6	Obras de fundações
	43.99-1		Serviços especializados para construção não especificados anteriormente	

Fonte: Autora

Em seguida, foram feitos tratamentos nas colunas com informações de datas, que foram organizadas para o formato dia/mês/ano (DD/MM/AAAA). Além disso, foi observada a necessidade de transformar alguns dados com o objetivo de utilizar as informações disponíveis para encontrar outras relevantes ao estudo. Dessa forma, foi criada a coluna “idade” para identificar a idade do profissional no momento da ocorrência, com as informações da data de nascimento do acidentado e da data de emissão da CAT, além disso, as datas dos acidentes também foram transformadas em dias da semana, sendo relacionadas de segunda à sexta.

Entretanto, observou-se que alguns dados não apresentavam coerência, como por exemplo idades muito baixas (p. ex. menores que 9 anos) ou muito altas (p. ex. acima dos 90 anos) que estivessem fora do intervalo de idades produtivas. Sendo assim, optou-se por restringir os dados ao intervalo de 18 a 65 anos de idade, para levar em consideração apenas os profissionais que pudessem exercer, legalmente, atividades na Construção Civil.

Devido a grande quantidade de dados com necessidade de correção, foram utilizadas funções com o objetivo de encapsular os tratamentos e assim facilitar as alterações. Dessa forma, foram corrigidas as informações com erros, como campos com valores não definidos devido aos erros de grafia e em caracteres especiais. Após as correções, foi realizada a remoção dos espaçamentos laterais em branco dos atributos e, por fim, foram excluídos os registros nulos.

Após esse processo, foram realizados agrupamentos nas colunas CBO, CID-10 e Agente Causador para reduzir os valores observados. Na base original, a informação da CBO estava representada pelo código de maior especificidade, sendo assim, os códigos foram convertidos para o nível de subgrupo (3 primeiros dígitos), tornando possível a redução dos valores. O mesmo entendimento seguiu para a CID-10, onde os códigos foram convertidos em intervalos de doenças (ex. doenças do sistema nervoso), de acordo com tabulação específica da classificação. Para o atributo de Agente Causador, no entanto, as informações foram classificadas em 5 grupos, conforme legislação e literatura específica de SST, sendo estes os agentes químicos, físicos, biológicos, ergonômicos e mecânicos.

Feito o agrupamento, foi verificada a necessidade de excluir algumas colunas, sendo elas: "CBO.1", "CID-10.1", "CNAE2.0 Empregador", "Data Afastamento", "Data Emissão CAT", "Data Acidente", por apresentarem informações duplicadas; "UF Munic. Acidente" pela alta quantidade de valores nulos (mais de um terço); "Data Despacho Benefício", por conter valores inconsistentes; "Munic Empr", pela alta variação de valores; "Data Nascimento", que foi transformada na informação de idade do acidentado. Dessa forma, foram selecionadas 16 colunas para a etapa de análise exploratória que podem ser visualizadas no Quadro 4.6.

Quadro 4.6 – Atributos dos dados após pré-processamento

Nº	Coluna	Tipo	Valores possíveis
1	Agente Causador Acidente	Categórico	5
2	CBO	Categórico	145
3	CID-10	Categórico	22
4	CNAE2.0 Empregador.1	Categórico	16
5	Emitente CAT	Categórico	5
6	Espécie do Benefício	Categórico	3
7	Filiação Segurado	Categórico	2
8	Indica Acidente	Categórico	2
9	Natureza da lesão	Categórico	28
10	Origem de Cadastramento CAT	Categórico	1
11	Parte Corpo Atingida	Categórico	42
12	Sexo	Categórico	2
13	Tipo do Acidente	Categórico	3
14	UF Munic Empr	Categórico	27
15	Data Acidente.1	Categórico	7
16	Idade	Numérico	48

Fonte: Autora

Com os ajustes realizados na etapa de pré-processamento, foi observado que a quantidade de dados de acidentes diminuiu consideravelmente, chegando ao total de 50.380 registros. A Figura 4.15 abaixo apresenta o *dataset* final após o tratamento dos dados.

Figura 4.15 – *Dataset* final

#	Column	Non-Null Count	Dtype
0	Agente Causador Acidente	50380 non-null	object
1	CBO	50380 non-null	object
2	CID-10	50380 non-null	object
3	CNAE2.0 Empregador.1	50380 non-null	object
4	Emitente CAT	50380 non-null	object
5	Espécie do benefício	50380 non-null	object
6	Filiação Segurado	50380 non-null	object
7	Indica acidente	50380 non-null	object
8	Natureza da Lesão	50380 non-null	object
9	Origem de Cadastramento CAT	50380 non-null	object
10	Parte Corpo Atingida	50380 non-null	object
11	Sexo	50380 non-null	object
12	Tipo do Acidente	50380 non-null	object
13	UF Munic. Empregador	50380 non-null	int64
14	Data Acidente.1	50380 non-null	int64
15	Idade	50380 non-null	int64

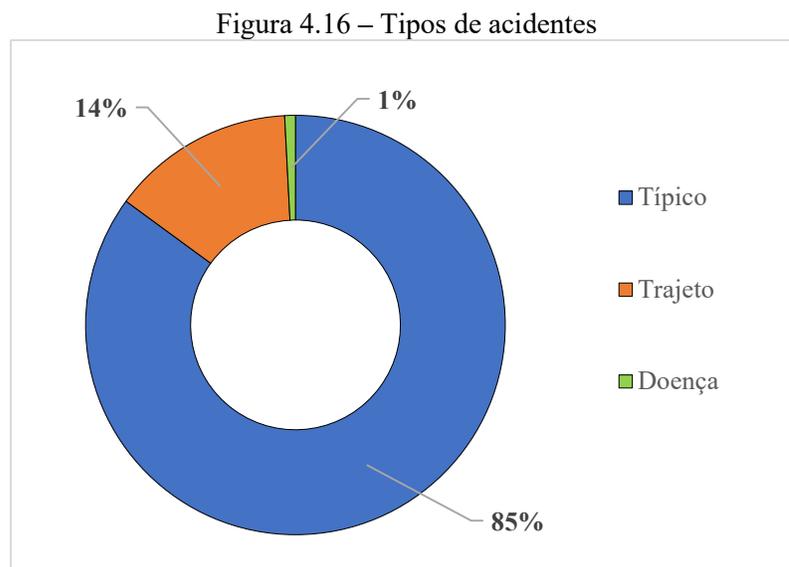
Fonte: Autora

4.2.3 Análise Exploratória dos Dados – AED

Após o pré-processamento dos dados, a etapa de Análise Exploratória dos Dados – AED teve o objetivo de obter um entendimento inicial sobre as informações de acidentes de trabalho identificando características relevantes sobre estes eventos.

Inicialmente, no processo de AED a análise foi dividida conforme a CAT classifica os acidentes de trabalho: típicos, trajeto e doenças. Ademais, foi criado um tópico a respeito das fatalidades observadas, devido a sua relevância para políticas brasileiras de SST.

Dessa forma, foi identificado que os acidentes típicos foram os mais recorrentes dentre os demais tipos, apresentando um total de 42.838 registros, cerca de 85% dos dados. Também foi observado que os acidentes de trajeto apareceram em 7.127 dos casos, com um percentual em torno de 14%. Além disso, observou-se que as doenças ocupacionais estiveram relacionadas a 415 ocorrências, por volta de 1% dos dados analisados. A Figura 4.16 abaixo apresenta o percentual encontrado para os tipos de acidentes de trabalho.



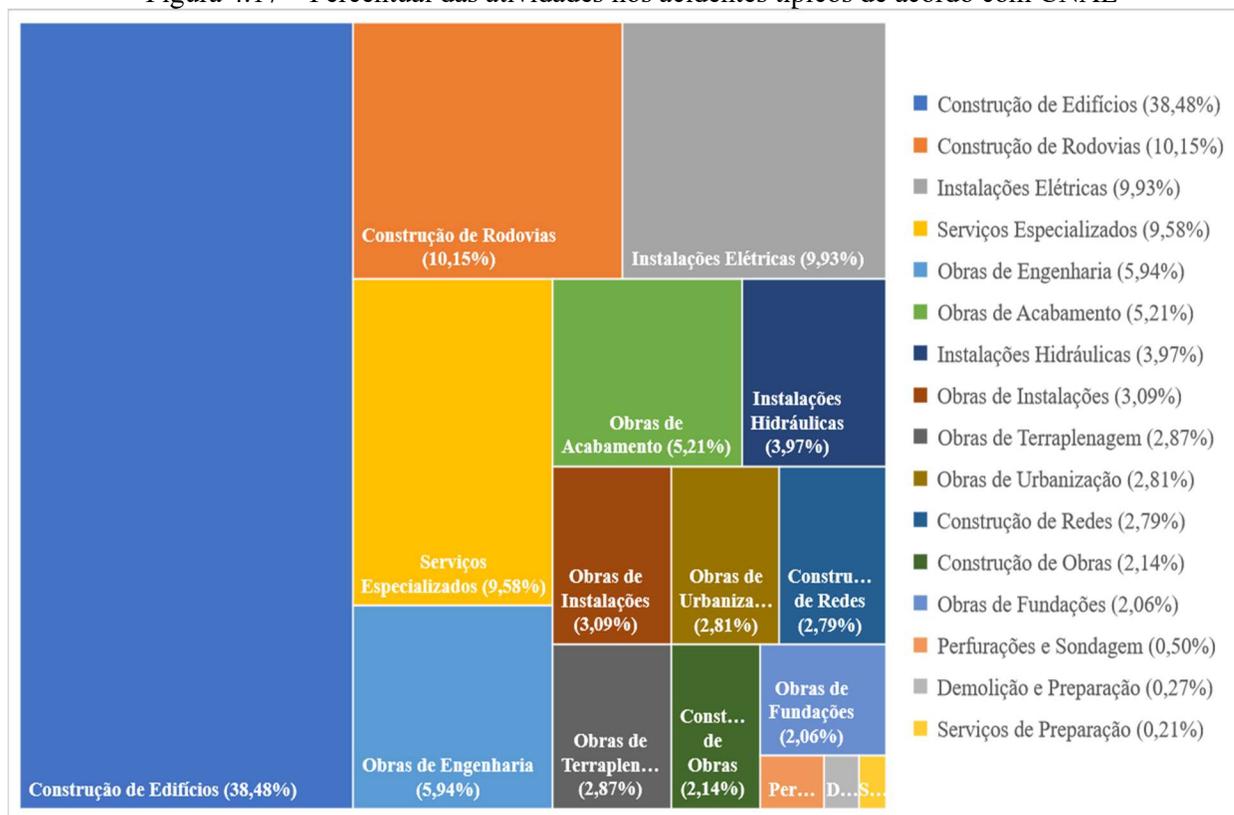
Fonte: Autora

4.2.3.1 Acidentes típicos

CNAE

Na análise CNAE, foi possível identificar as atividades econômicas relacionadas aos acidentes de trabalho típicos (Figura 4.17). O destaque se deu para a atividade de “construção de edifícios”, que apresentou um total de 16.484 acidentes e percentual próximo aos 40% dos registros. Logo abaixo, observou-se que a “construção de rodovias”, com 4.348 acidentes (10,15%) e as atividades de “instalações elétricas”, com 4.255 (9,93%), também foram responsáveis por grande número de acidentes dessa natureza, representando o 2º e 3º, respectivamente.

Figura 4.17 – Percentual das atividades nos acidentes típicos de acordo com CNAE



Fonte: Autora

CBO

Em relação ao CBO, as funções dos acidentados também foram observadas para identificar as mais recorrentes neste tipo de acidente (Quadro 4.7). Dessa forma, observou-se que os “trabalhadores da construção civil e obras públicas” foram os mais afetados por acidentes típicos, com um total de 12.806 registros e percentual de cerca de 30% dos dados analisados. Em seguida, ocupando o segundo lugar, foi visualizada a função de “ajudantes de obras”, que apresentou 11.275 acidentes e percentual de aproximadamente 26%. Logo abaixo, com 2.689 (6,28%) e 2.657 (6,20%), foram observadas as funções de “trabalhadores de montagem de tubulações, estruturas metálicas e de compósitos” e “trabalhadores nos serviços de administração, conservação e manutenção de edifícios e logradouros”, respectivamente.

Quadro 4.7 – As 10 funções mais afetadas nos acidentes típicos de acordo com CBO

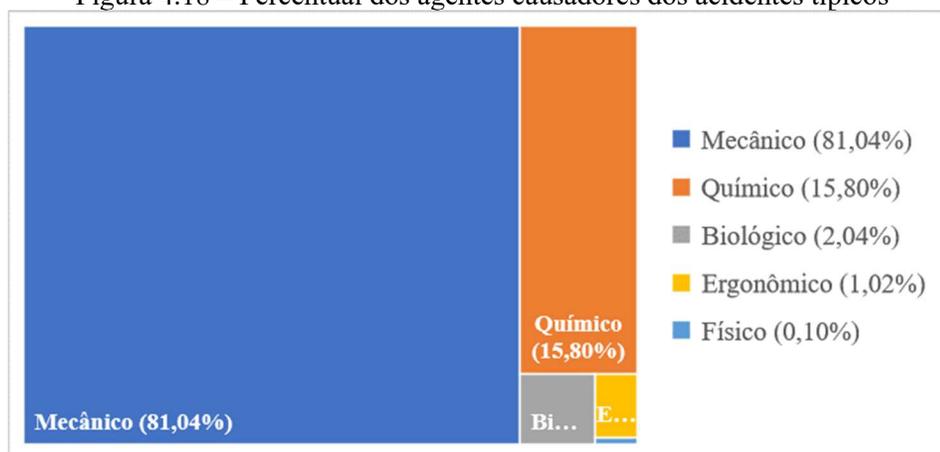
Subgrupo CBO	Descrição	Percentual	Total de acidentes
715	Trabalhadores da construção civil e obras públicas	29.89%	12.806
717	Ajudantes de obras	26.32%	11.275
724	Trabalhadores de montagem de tubulações, estruturas metálicas e de compósitos	6.28%	2.689
514	Trabalhadores nos serviços de administração, conservação e manutenção de edifícios e logradouros	6.20%	2.657
782	Condutores de veículos e operação de equipamentos de elevação e de movimentação de cargas	4.35%	1.863
716	Trabalhadores de acabamento de obras	2.37%	1.016
710	Supervisores da extração mineral e da construção civil	1.82%	778
992	Trabalhadores elementares da manutenção	1.76%	756
911	Mecânicos de manutenção de máquinas e equipamentos industriais, comerciais e residenciais	1.70%	727
951	Eletricistas eletrônicos de manutenção industrial, comercial e residencial	1.66%	710

Fonte: Autora

Agentes causadores

Analisando os dados de agentes causadores (Figura 4.18), foi possível identificar que os agentes mecânicos, referentes a materiais (metais, madeiras, tijolos), ferramentas manuais e elétricas (martelos, marretas, alavancas, serras), equipamentos (escadas, andaimes, plataformas), máquinas (veículos de transporte, rodoviários e de terraplenagem) e situações específicas (quedas sem/com diferença de nível, impactos) foram responsáveis por 34.717 acidentes, representando aproximadamente 81% do total de acidentes típicos analisados. Em seguida, foi observado que os agentes químicos, referentes ao contato, inalação ou absorção de substâncias químicas como ácidos, aerodispersóides, álcalis, asfaltos, gases, entre outros, respondeu por 6.769 registros, cerca de 16% dos casos. Os agentes biológicos relacionados, por exemplo, aos ataques de seres vivos, ingestão de produtos alimentícios de origem animal/vegetal e agentes infecciosos (bactérias, vírus, protozoários), entre outros, foram considerados causas de 874 acidentes, cerca de 2% dos registros. Em relação aos agentes ergonômicos, relacionados com condições de esforço excessivo e mobiliários/acessórios inadequados, por exemplo, observou-se um total de 435 eventos, representando cerca de 1% dos acidentes típicos. Por fim, os agentes de natureza física, como radiações ionizantes e não ionizantes, altas/baixas temperaturas, ruídos e níveis de pressão altos/baixos responderam por 43 acidentes, 0,1% do total de registros.

Figura 4.18 – Percentual dos agentes causadores dos acidentes típicos

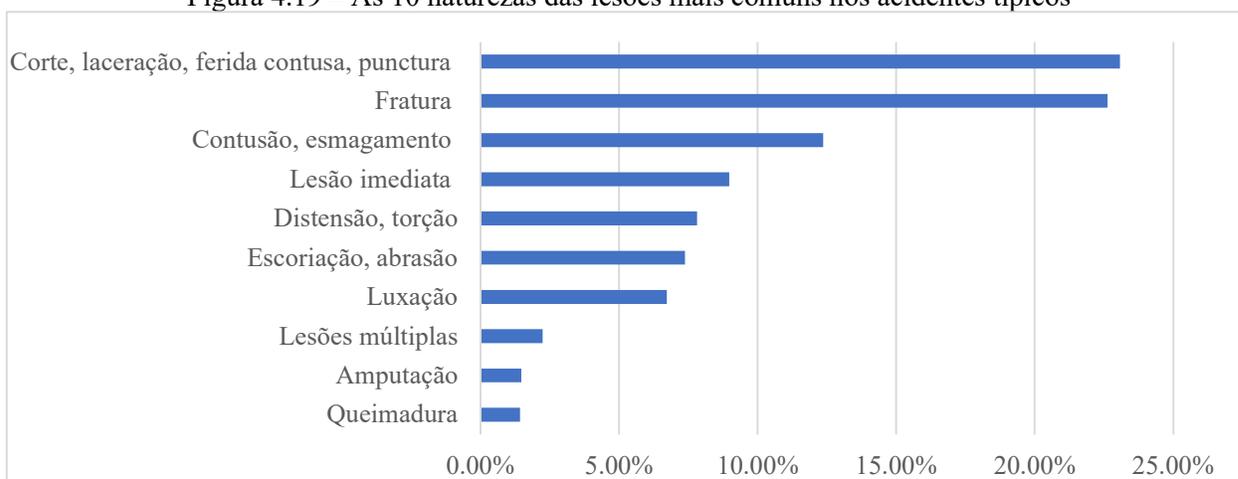


Fonte: Autora

Natureza das lesões

De acordo com a análise conduzida sobre a natureza das lesões (Figura 4.19), observou-se que os “cortes, lacerações, feridas contusas e puncturas” foram evidenciados em 9.887 registros, representando aproximadamente 23% do total de acidentes típicos. Ocupando o segundo lugar dessa categoria, foram visualizadas as “fraturas”, apresentando 9.696 registros, com 22,63%. Logo em seguida, puderam ser observadas as “contusões, esmagamentos”, relacionadas a 5.301 acidentes, em torno de 12%.

Figura 4.19 – As 10 naturezas das lesões mais comuns nos acidentes típicos

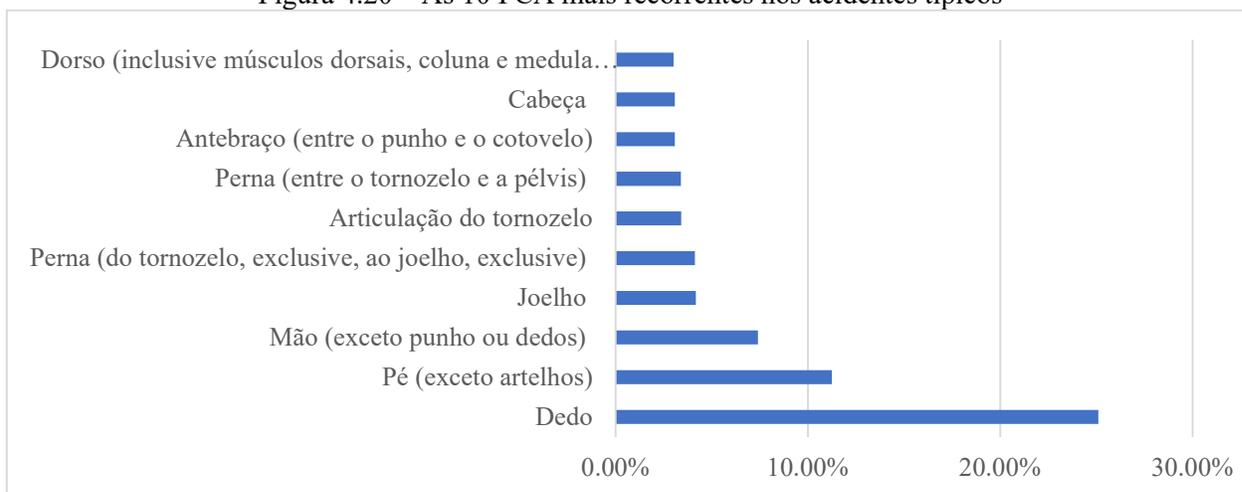


Fonte: Autora

Parte do corpo atingida

Em relação à Parte do Corpo Atingida – PCA, foi observado que os “dedos” foram o local mais afetado em decorrência de um acidente típico, com 10.756 recorrências, cerca de 25% do total. Em seguida, representando aproximadamente 11% dos acidentes, observou-se que os “pés (exceto artelhos)” dos trabalhadores foram afetados em 4.817 casos. Além desses locais, “mãos (exceto punho ou dedos)”, “joelhos” e “pernas (do tornozelo, exclusive, ao joelho, exclusive)” também se destacaram entre as partes do corpo mais atingidas em um acidente típico, representando cerca de 7% (3.173 casos), 4% (1790 casos) e 4% (1770 casos), respectivamente. A Figura 4.20 abaixo apresenta as 10 PCA mais comuns nos acidentes típicos.

Figura 4.20 – As 10 PCA mais recorrentes nos acidentes típicos



Fonte: Autora

CID

Sobre a análise da CID (Quadro 4.8) observou-se que 36.166 acidentes ocorreram devido a “lesões, envenenamentos e algumas outras consequências de causas externas”, o que representou cerca de 85% do total de registros. Além disso, “doenças do sistema osteomuscular e do tecido conjuntivo” apareceram em 2.423 dos casos, aproximadamente 6% dos eventos. Outras classificações para “causas externas de morbidade e mortalidade”, “fatores que exercem influência sobre o estado de saúde e o contato com serviços de saúde” e “sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte” também apareceram nas primeiras colocações, com 4%, 2% e 1%, respectivamente.

Quadro 4.8 – As 5 CID mais comuns nos acidentes típicos

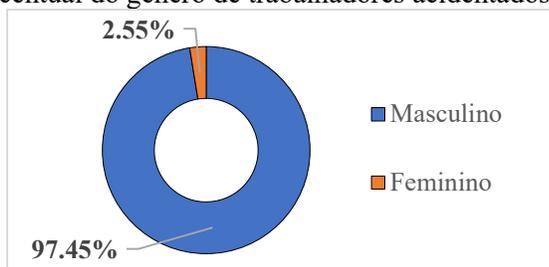
Intervalo CID	Descrição	Percentual	Total de acidentes
S00 - T98	Lesões, envenenamentos e algumas outras consequências de causas externas	84.43%	36.166
M00 - M99	Doenças do sistema osteomuscular e do tecido conjuntivo	5.66%	2.423
V01 - Y98	Causas externas de morbidade e de mortalidade	4.07%	1.742
Z00 - Z99	Fatores que exercem influência sobre o estado de saúde e o contato com serviços de saúde	2.27%	973
R00 - R99	Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte	1.07%	457

Fonte: Autora

Gênero

Os acidentes típicos também foram analisados a respeito do gênero dos acidentados. Dessa forma, foi observado que 41.744 trabalhadores, cerca de 98% do total, pertenciam ao gênero masculino, enquanto 1.094, aproximadamente 2%, ao feminino (Figura 4.21).

Figura 4.21 – Percentual do gênero de trabalhadores acidentados (acidentes típicos)

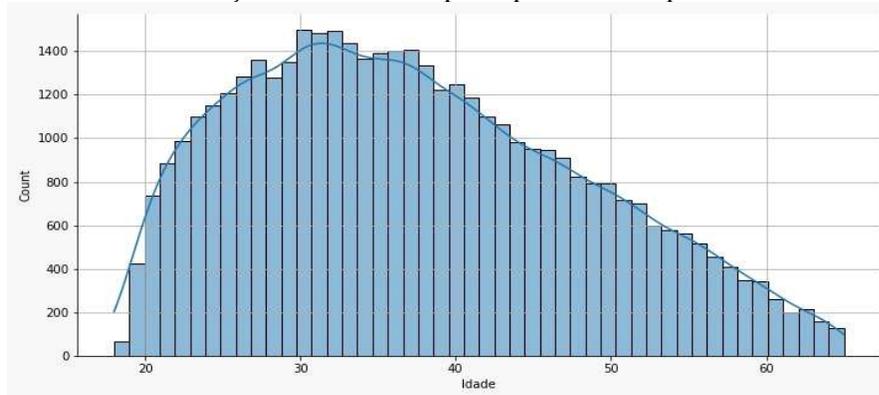


Fonte: Autora

Idade

A idade dos profissionais que sofreram acidentes típicos foi analisada a partir da distribuição presente na Figura 4.22. Dessa forma, observou-se que a idade com maior número de acidentes foi a de 31 anos, enquanto a com menor quantidade foi a de 18 anos. Além disso, observou-se que a frequência de acidentes típicos foi maior na faixa etária entre os 30 e 40 anos e que, com o aumento das idades, os níveis de acidentes tendem a diminuir de forma relativamente constante.

Figura 4.22 – Distribuição de acidentes típicos por idade do profissional acidentado

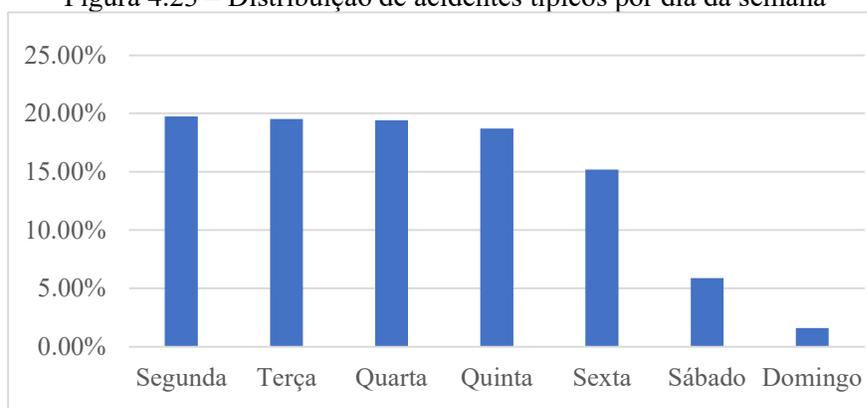


Fonte: Autora

Data do acidente

Em relação aos dias da semana de ocorrência dos acidentes típicos, foi observado que a segunda obteve o maior número de acidentes (8.462, 19,75%) e o domingo o menor número (678, 1,58%). Ao longo da semana, é possível observar também uma pequena variação entre os dias de segunda e quinta, todos com percentuais próximos aos 20%. Os finais de semana, entretanto, apresentaram menor recorrência. A Figura 4.23 apresenta a distribuição dos acidentes ao longo da semana.

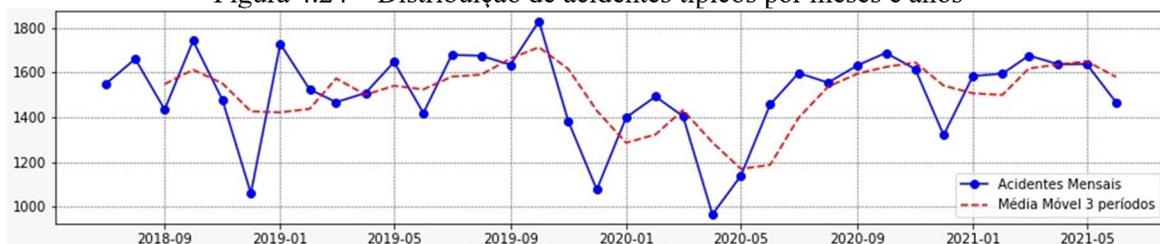
Figura 4.23 – Distribuição de acidentes típicos por dia da semana



Fonte: Autora

Uma outra análise relacionada à data do acidente levou em consideração os meses e os anos dos registros, que variaram entre julho de 2018 e junho de 2021, totalizando 3 anos (Figura 4.24). Nesta análise, foi identificado que o mês com maior frequência de acidentes foi outubro de 2019 e o com menor frequência, abril de 2020. Também foi observado, em relação aos 3 anos de dados, que o mês de outubro foi o que apresentou maiores taxas de acidente, enquanto os meses de dezembro normalmente se relacionam aos menores índices. A distribuição por meses e anos também evidenciou uma queda dos eventos durante o período de pandemia.

Figura 4.24 – Distribuição de acidentes típicos por meses e anos



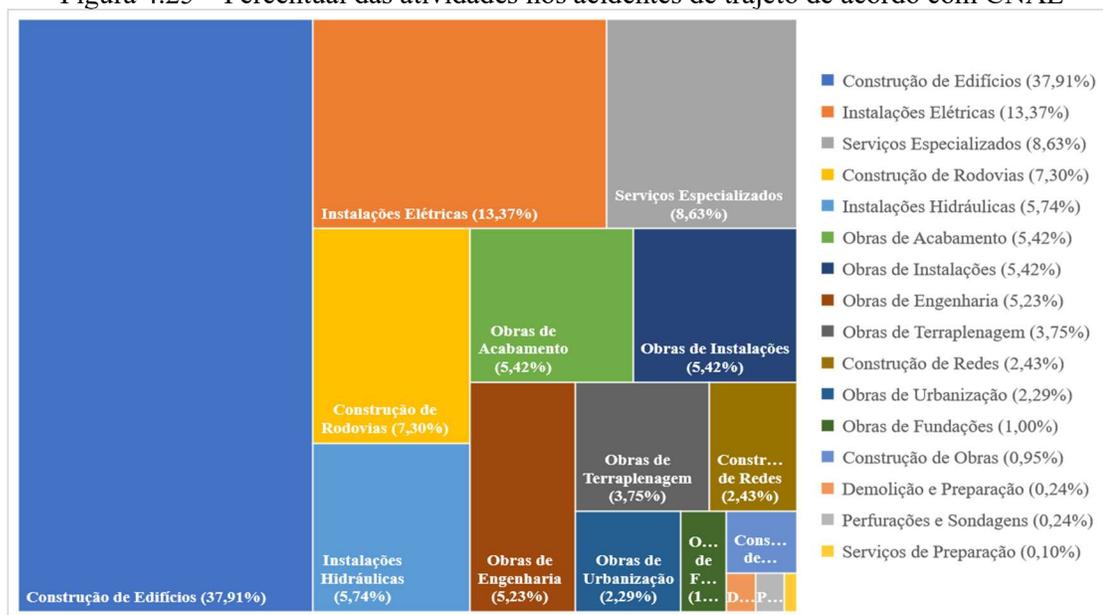
Fonte: Autora

4.2.3.2 Acidentes de trajeto

CNAE

Na análise CNAE, foi possível identificar as atividades econômicas relacionadas aos acidentes de trajeto (Figura 4.25). O destaque se deu para a atividade de “construção de edifícios”, que apresentou um total de 2.702 acidentes e percentual próximo aos 38% dos registros. Logo abaixo, observou-se que atividades relacionadas com “instalações elétricas” apresentaram 953 acidentes, cerca de 13%, e os “serviços especializados”, 615 registros, percentual próximo dos 9%.

Figura 4.25 – Percentual das atividades nos acidentes de trajeto de acordo com CNAE



Fonte: Autora

CBO

Em relação ao CBO, as funções dos acidentados também foram observadas para identificar as mais recorrentes nos acidentes de trajeto (Quadro 4.9). Dessa forma, observou-se que os “trabalhadores da construção civil e obras públicas” foram os mais afetados por acidentes dessa natureza, com um total de 1.771 registros e percentual de cerca de 25%. Em seguida, ocupando o segundo lugar, foi visualizada a função de “ajudantes de obras”, que apresentou 1.302 acidentes e percentual de aproximadamente 18%. Logo abaixo, com 429 (6,02%) e 415 (5,82%), foram observadas as

funções de “trabalhadores nos serviços de administração, conservação e manutenção de edifícios e logradouros” e “condutores de veículos e operação de equipamentos de elevação e de movimentação de cargas”, respectivamente.

Quadro 4.9 – As 10 funções mais afetadas nos acidentes de trajeto de acordo com CBO

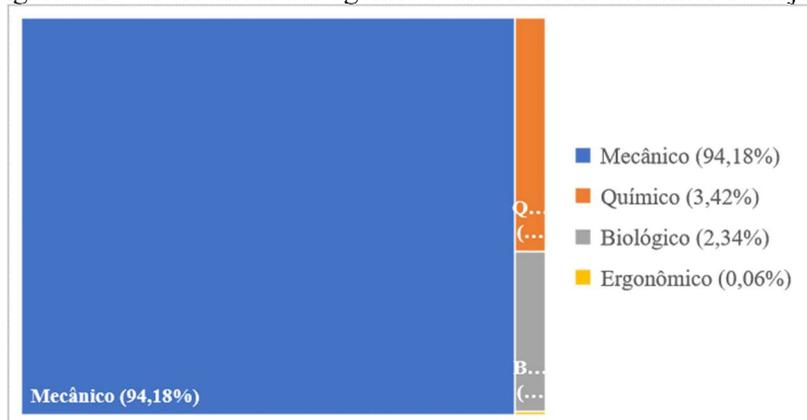
Subgrupo CBO	Descrição	Percentual	Total de acidentes
715	Trabalhadores da construção civil e obras públicas	24.85%	1.771
717	Ajudantes de obras	18.27%	1.302
514	Trabalhadores nos serviços de administração, conservação e manutenção de edifícios e logradouros	6.02%	429
782	Condutores de veículos e operação de equipamentos de elevação e de movimentação de cargas	5.82%	415
724	Trabalhadores de montagem de tubulações, estruturas metálicas e de compósitos	4.36%	311
411	Escriturários em geral, agentes, assistentes e auxiliares administrativos	4.13%	294
716	Trabalhadores de acabamento de obras	2.72%	194
710	Supervisores da extração mineral e da construção civil	2.64%	188
951	Eletricistas eletrônicos de manutenção industrial, comercial e residencial	2.34%	167
732	Instaladores e reparadores de linhas e cabos elétricos e de comunicações	1.89%	135

Fonte: Autora

Agentes causadores

Analisando os dados de agentes causadores (Figura 4.26), foi possível identificar que os agentes mecânicos como motocicletas/motonetas, veículos e bicicletas, além de situações de quedas e locais como calçadas e estradas, por exemplo, foram responsáveis por 6.712 acidentes, representando aproximadamente 94% do total de acidentes de trajeto analisados. Em seguida, foi observado que os agentes químicos, como asfalto e óleo combustível responderam por 244 registros, cerca de 3% dos casos. Os agentes biológicos como animais e vegetação foram considerados causadores de 167 acidentes, cerca de 2% dos registros. Em relação aos agentes ergonômicos, como esforço excessivo, observou-se um total de 4 eventos, representando menos de 1% dos acidentes.

Figura 4.26 – Percentual dos agentes causadores dos acidentes de trajeto

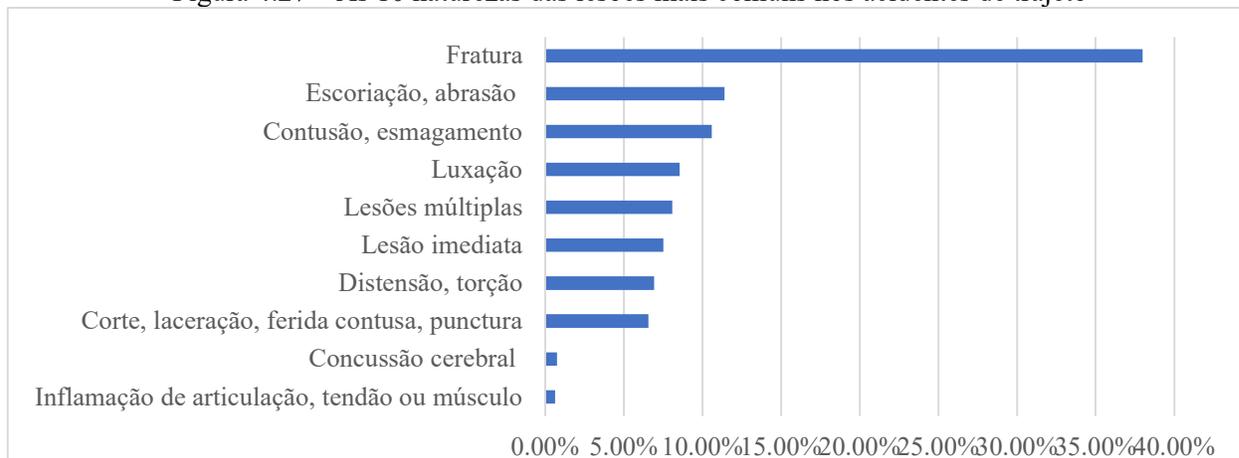


Fonte: Autora

Naturezas de lesões

De acordo com a análise conduzida sobre a natureza das lesões (Figura 4.27), observou-se que as “fraturas” foram evidenciadas em 2.707 registros, representando aproximadamente 38% do total de acidentes de trajeto. Ocupando o segundo lugar dessa categoria, foram visualizadas as “escoriações, abrasões” apresentando 812 registros, cerca de 11%. Logo em seguida, puderam ser observadas as “contusões, esmagamentos”, relacionadas a 754 acidentes, com 10,58% do total.

Figura 4.27 – As 10 naturezas das lesões mais comuns nos acidentes de trajeto

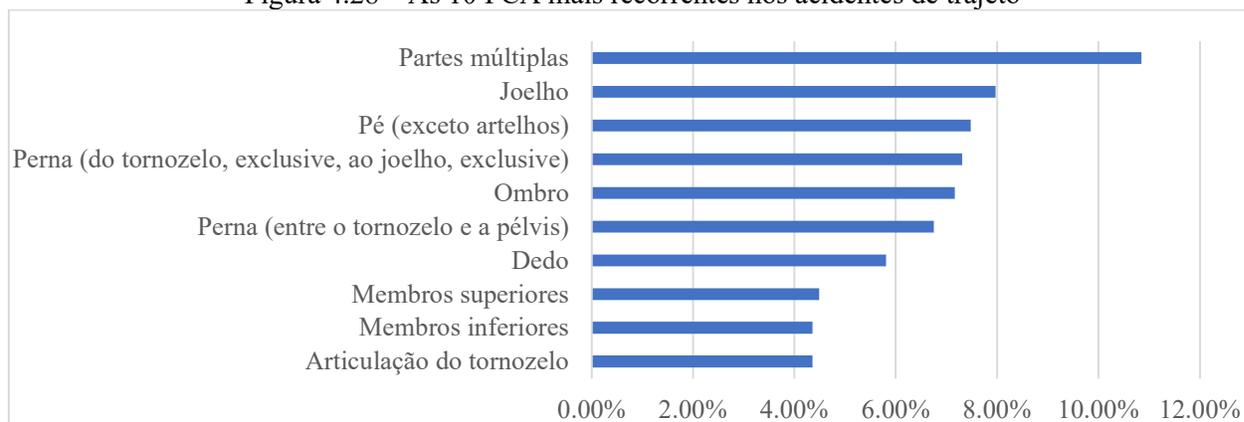


Fonte: Autora

Parte do corpo atingida

Em relação à Parte do Corpo Atingida – PCA de um acidente típico, foi observado que “partes múltiplas” foram mais recorrentes, com 773 registros, cerca de 11% do total. Em seguida, representando aproximadamente 8% dos acidentes, observou-se que os “joelhos” dos trabalhadores foram afetados em 568 casos. Além desses locais, “pés (exceto artelhos)”, “pernas (do tornozelo, exclusive, ao joelho, exclusive)” e “ombros” também se destacaram entre as partes do corpo mais atingidas em decorrência de um acidente típico, representando entre 6% e 7% dos casos analisados. A Figura 4.28 abaixo apresenta as 10 PCA mais comuns nos acidentes de trajeto.

Figura 4.28 – As 10 PCA mais recorrentes nos acidentes de trajeto



Fonte: Autora

CID

Sobre a análise da CID (Quadro 4.10), observou-se que 5.969 acidentes ocorreram devido a “lesões, envenenamentos e algumas outras consequências de causas externas”, o que representou cerca de 84% do total de registros. Além disso, as “causas externas de morbidade e mortalidade” apareceram em 503 casos, aproximadamente 7% dos eventos. Outras classificações para “doenças do sistema osteomuscular e do tecido conjuntivo”, “sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte” e “fatores que exercem influência sobre o estado de saúde e o contato com serviços de saúde” também foram mais evidentes, com cerca de 5%, 1% e 1%, respectivamente.

Quadro 4.10 – As 5 CID mais comuns nos acidentes de trajeto

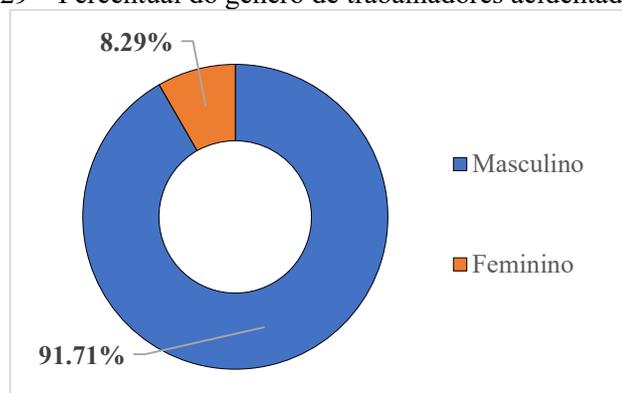
Intervalo CID	Descrição	Percentual	Total de acidentes
S00 - T98	Lesões, envenenamentos e algumas outras consequências de causas externas	83.75%	5969
V01 - Y98	Causas externas de morbidade e de mortalidade	7.06%	503
M00 - M99	Doenças do sistema osteomuscular e do tecido conjuntivo	5.51%	393
R00 - R99	Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte	1.29%	92
Z00 - Z99	Fatores que exercem influência sobre o estado de saúde e o contato com serviços de saúde	1.26%	90

Fonte: Autora

Gênero

Os acidentes de trajeto também foram analisados a respeito do gênero dos acidentados. Dessa forma, foi observado que 6.536 trabalhadores, cerca de 92% do total, pertenciam ao gênero masculino, enquanto 591, aproximadamente 8%, ao feminino (Figura 4.29).

Figura 4.29 – Percentual do gênero de trabalhadores acidentados (trajeto)

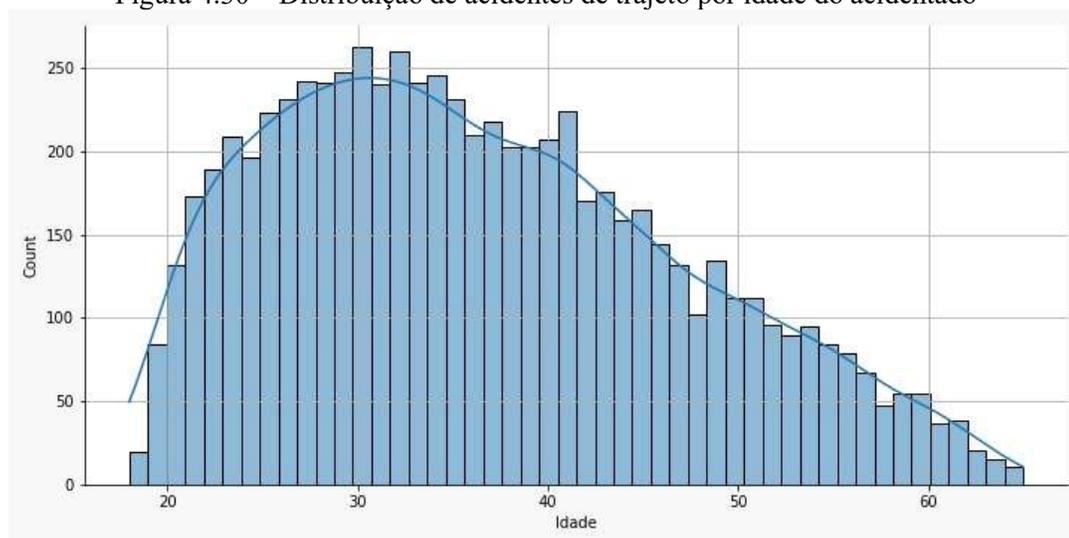


Fonte: Autora

Idade

A idade dos profissionais que sofreram acidentes de trajeto foi analisada a partir da distribuição presente na Figura 4.30. Foi possível observar que a idade com maior taxa de acidentes foi a de 30 anos, enquanto a com menor quantidade foi a de 65 anos. Além disso, observou-se que a frequência de acidentes de trajeto foi maior na faixa etária entre os 25 e 35 anos e que, com o aumento das idades, os níveis de acidentes tenderam a diminuir de forma relativamente constante, salvo as idades de 40 e 41 anos.

Figura 4.30 – Distribuição de acidentes de trajeto por idade do acidentado

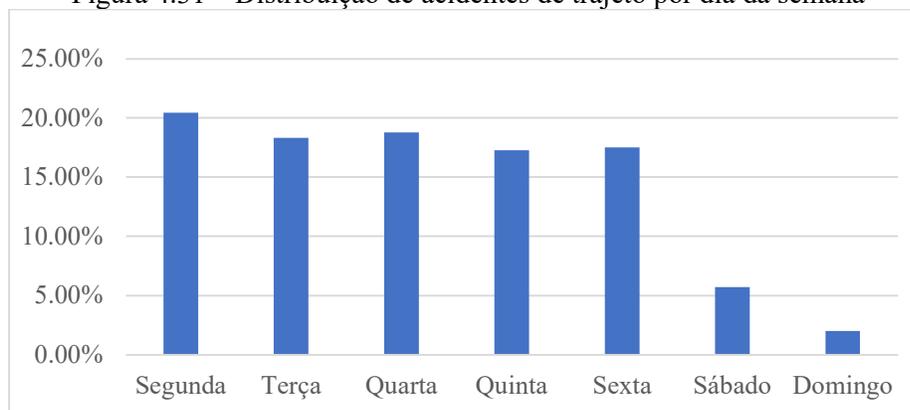


Fonte: Autora

Data do acidente

Em relação aos dias da semana de ocorrência dos acidentes típicos, foi observado que a segunda obteve o maior número de acidentes (1.456, 20,43%) e o domingo o menor número (143, 2,01%), representando cerca de 20% e 2%, respectivamente. Durante a semana, é possível observar uma pequena variação entre os dias, todos com percentuais próximos aos 20%, sendo o menor a sexta com 17,5%. Os finais de semana, no entanto, apresentaram menor recorrência de acidentes de trajeto. A Figura 4.31 apresenta a distribuição dos acidentes ao de acordo com os dias da semana.

Figura 4.31 – Distribuição de acidentes de trajeto por dia da semana



Fonte: Autora

Uma outra análise relacionada à data dos acidentes de trajeto levou em consideração os meses e os anos dos registros (Figura 4.32). Dessa forma, foi identificado que o mês com maior frequência de acidentes foi julho de 2019 e o com menor frequência, março de 2020. Também foi observado, em relação aos 3 anos de dados, que os meses de dezembro normalmente se relacionaram aos menores índices, salvo a queda dos eventos durante o período de pandemia.



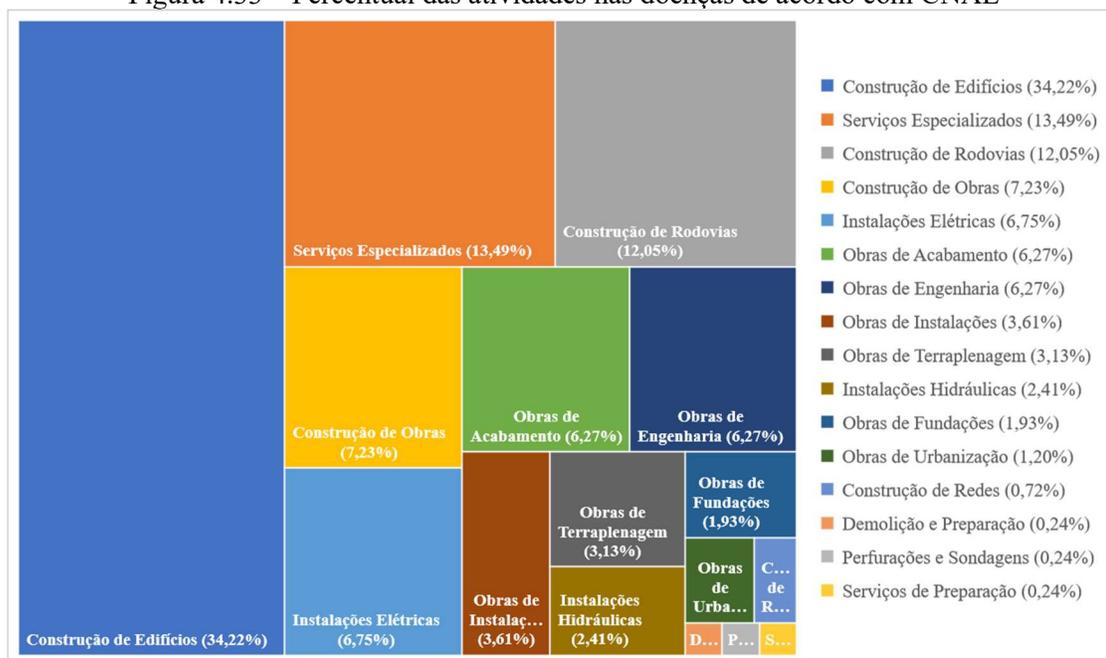
Fonte: Autora

4.2.3.3 Doenças ocupacionais

CNAE

Na análise CNAE, foi possível identificar as atividades econômicas relacionadas às doenças ocupacionais (Figura 4.33). O destaque se deu para a atividade de “construção de edifícios”, que apresentou um total de 142 doenças e percentual de aproximadamente 34% dos registros. Logo abaixo, observou-se que os “serviços especializados”, com 56 registros (13,49%) e a “construção de rodovias”, com 50 (12,05%), também estiveram em evidência dentre as doenças ocupacionais.

Figura 4.33 – Percentual das atividades nas doenças de acordo com CNAE



Fonte: Autora

CBO

Em relação ao CBO, as funções dos acidentados também foram observadas para identificar as mais recorrentes relacionadas às doenças ocupacionais (Quadro 4.11). Dessa forma, observou-se que os “trabalhadores da construção civil e obras públicas” foram os mais afetados, com um total de 110 registros e percentual de cerca de 26% dos dados analisados. Em seguida, ocupando o segundo lugar, foi visualizada a função de “ajudantes de obras”, que apresentou 77 casos e percentual de aproximadamente 19%. Logo abaixo, com 30 (7,23%) e 29 (6,99%) doenças registradas, respectivamente, foram observadas as funções de “condutores de veículos e operação de equipamentos de elevação e de movimentação de cargas” e “trabalhadores de montagem de tubulações, estruturas metálicas e de compósitos”.

Quadro 4.11 – As 10 funções mais afetadas nas doenças ocupacionais de acordo com CBO

Subgrupo CBO	Descrição	Percentual	Total de acidentes
715	Trabalhadores da construção civil e obras públicas	26.51%	110
717	Ajudantes de obras	18.55%	77
782	Condutores de veículos e operação de equipamentos de elevação e de movimentação de cargas	7.23%	30
724	Trabalhadores de montagem de tubulações, estruturas metálicas e de compósitos	6.99%	29
716	Trabalhadores de acabamento de obras	4.34%	18
992	Trabalhadores elementares da manutenção	2.89%	12
710	Supervisores da extração mineral e da construção civil	2.65%	11
514	Trabalhadores nos serviços de administração, conservação e manutenção de edifícios e logradouros	2.41%	10
313	Técnicos em eletroeletrônica e fotônica	2.17%	9
411	Escriturários em geral, agentes, assistentes e auxiliares administrativos	1.69%	7

Fonte: Autora

Agentes causadores

Analisando os dados de agentes causadores (Quadro 4.12), foi possível identificar que os “esforços excessivos” foram responsáveis por 119 doenças ocupacionais, representando aproximadamente 30% do total registros. Em seguida, foi observado que os “ataques de seres vivos” responderam por 42 doenças, cerca de 10% dos casos. A “exposição ao ruído” também destacou-se sendo observados 36 casos, representando cerca de 9% das doenças, enquanto os “agentes infecciosos ou parasitários” também foram considerados causas de 29 doenças ocupacionais, cerca de 7% dos registros.

Quadro 4.12 – Os 10 agentes causadores de doenças ocupacionais mais recorrentes

Agente causador	Total de acidentes	Percentual
Esforço excessivo	119	28,67%
Ataque de ser vivo	42	10,12%
Exposição ao ruído	36	8,67%
Agente infeccioso ou parasitário	29	6,99%
Inalação, ingestão e absorção	24	5,78%
Impacto sofrido por pessoa	15	3,61%
Aprisionamento em, sob ou entre	14	3,37%
Queda de pessoa em mesmo nível	11	2,65%
Queda de pessoa com diferença de nível	9	2,17%
Atrito ou abrasão	9	2,17%

Fonte: Autora

Naturezas de lesões

De acordo com a análise conduzida sobre a natureza das lesões (Figura 4.34), observou-se que as “doenças não informada ou classificada” foram evidenciados em 149 registros, representando aproximadamente 36% do total de doenças. Ocupando o segundo lugar dessa categoria, foram visualizadas as “perdas ou diminuições mediatas de sentido”, com 43 registros, cerca de 10%. Logo em seguida, puderam ser observadas as “inflamações de articulação, tendão ou músculo”, relacionadas a 41 doenças, cerca de 10%.

Figura 4.34 – As 10 naturezas das lesões mais comuns nas doenças ocupacionais

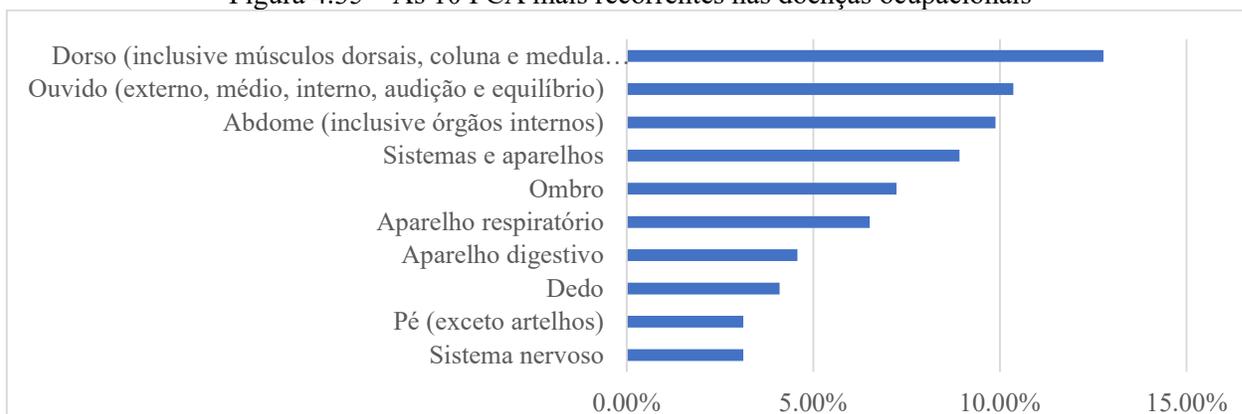


Fonte: Autora

Parte do corpo atingida

Em relação à Parte do Corpo Atingida – PCA, foi observado que o “dorso (inclusive músculos dorsais, coluna e medula espinhal)” foi o local mais afetado em decorrência de uma doença do trabalho, com 53 registros, cerca de 13% do total. Em seguida, representando aproximadamente 11%, observou-se os “ouvidos (externo, médio, interno, audição e equilíbrio)” em 43 casos. Além desses locais, “abdome (inclusive órgãos internos)”, “sistemas e aparelhos” e “ombros” também se destacaram entre as partes do corpo mais atingidas, representando 41 (10%), 37 (9%) e 30 casos (7%), respectivamente. A Figura 4.35 abaixo apresenta as 10 PCA mais comuns nas doenças do trabalho.

Figura 4.35 – As 10 PCA mais recorrentes nas doenças ocupacionais



Fonte: Autora

CID

Sobre a análise da CID (Quadro 4.13), observou-se que 113 registros ocorreram devido a “doenças do sistema osteomuscular e do tecido conjuntivo”, cerca de 27% do total. As “lesões, envenenamentos e algumas outras consequências de causas externas”, representaram em torno de 20% do total de registros. Outras classificações como “algumas doenças infecciosas e parasitárias”, “doenças do ouvido e apófise mastóide” e “doenças do aparelho digestivo” também se destacaram entre as doenças observadas, com aproximadamente 15%, 10% e 9%, respectivamente.

Quadro 4.13 – As 10 CID mais comuns nas doenças ocupacionais

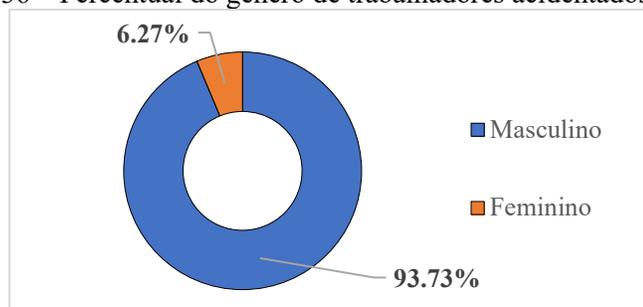
Intervalo CID	Descrição	Percentual	Total de doenças
M00 - M99	Doenças do sistema osteomuscular e do tecido conjuntivo	27.23%	113
S00 - T98	Lesões, envenenamentos e algumas outras consequências de causas externas	19.52%	81
A00 - B99	Algumas doenças infecciosas e parasitárias	14.94%	62
H60 - H95	Doenças do ouvido e da apófise mastóide	9.88%	41
K00 - K93	Doenças do aparelho digestivo	9.4%	39
F00 - F99	Transtornos mentais e comportamentais	3.61%	15
J00 - J99	Doenças do aparelho respiratório	3.61%	15
G00 - G99	Doenças do sistema nervoso	2.65%	11
Z00 - Z99	Fatores que exercem influência sobre o estado de saúde e o contato com serviços de saúde	2.17%	9
L00 - L99	Doenças da pele e do tecido subcutâneo	1.93%	8

Fonte: Autora

Gênero

As doenças ocupacionais também foram analisadas a respeito do gênero dos acidentados. Dessa forma, foi observado que 389 trabalhadores, cerca de 94% do total, pertenciam ao gênero masculino, enquanto 26, aproximadamente 6%, ao feminino (Figura 4.36).

Figura 4.36 – Percentual do gênero de trabalhadores acidentados (doenças)

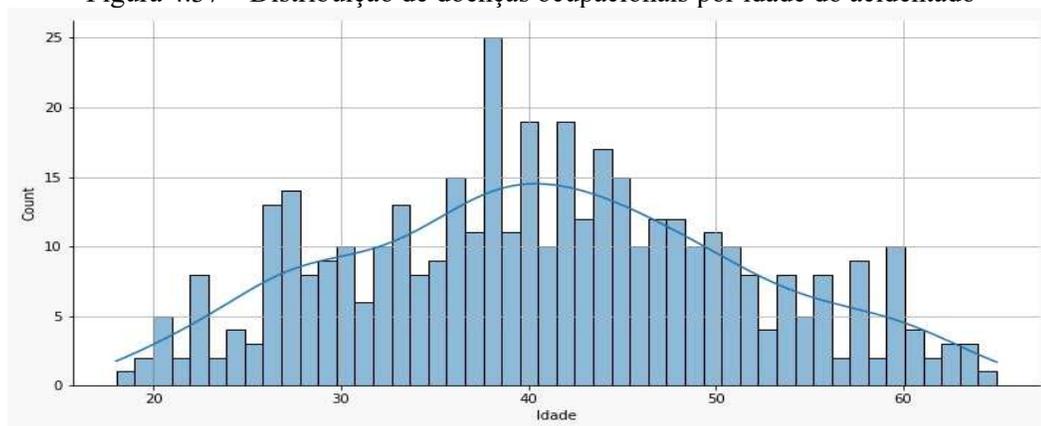


Fonte: Autora

Idade

A idade dos profissionais que sofreram doenças ocupacionais foi analisada a partir da distribuição presente na Figura 4.37. Assim, observou-se que a idade com maior número de doenças foi a de 38 anos, enquanto a com menor quantidade foram as de 18 e 65. Além disso, observou-se que a frequência de doenças foi maior na faixa etária entre os 35 e 45 anos, no entanto, outras idades também apresentaram mais casos de doenças, como idades próximas aos 30 e 60 anos.

Figura 4.37 – Distribuição de doenças ocupacionais por idade do acidentado



Fonte: Autora

Data do acidente

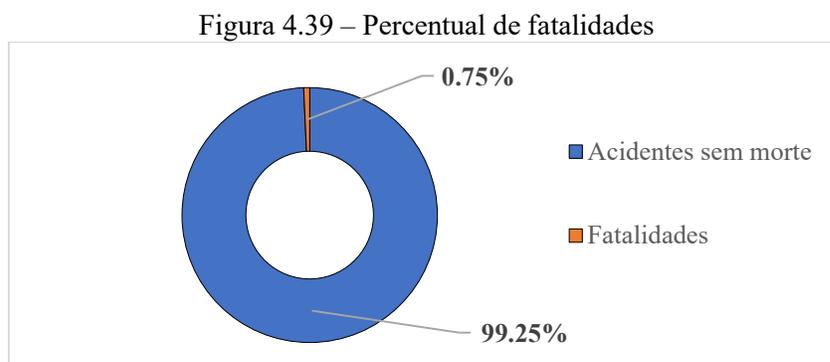
Como as doenças ocupacionais são caracterizadas por exposições prolongadas, ao contrário dos acidentes típicos e de trajeto, a análise por dia da semana não foi conduzida para este tópico. Sendo assim, foi analisada a distribuição de doenças relacionadas aos meses e anos dos dados obtidos (Figura 4.38). Assim, foi identificado que o mês com maior frequência de acidentes foi março de 2021 e o com menor frequência, dezembro de 2020. Observou-se também que o registro das doenças tende a ser menor nos meses próximos ao final do ano.



Fonte: Autora

4.2.3.4 Acidentes fatais

Dentre os 50.380 dados observados, foi possível identificar que 378 foram relacionados a fatalidades na construção, o que representa um percentual de 0,75%, conforme é evidenciado na Figura 4.39.

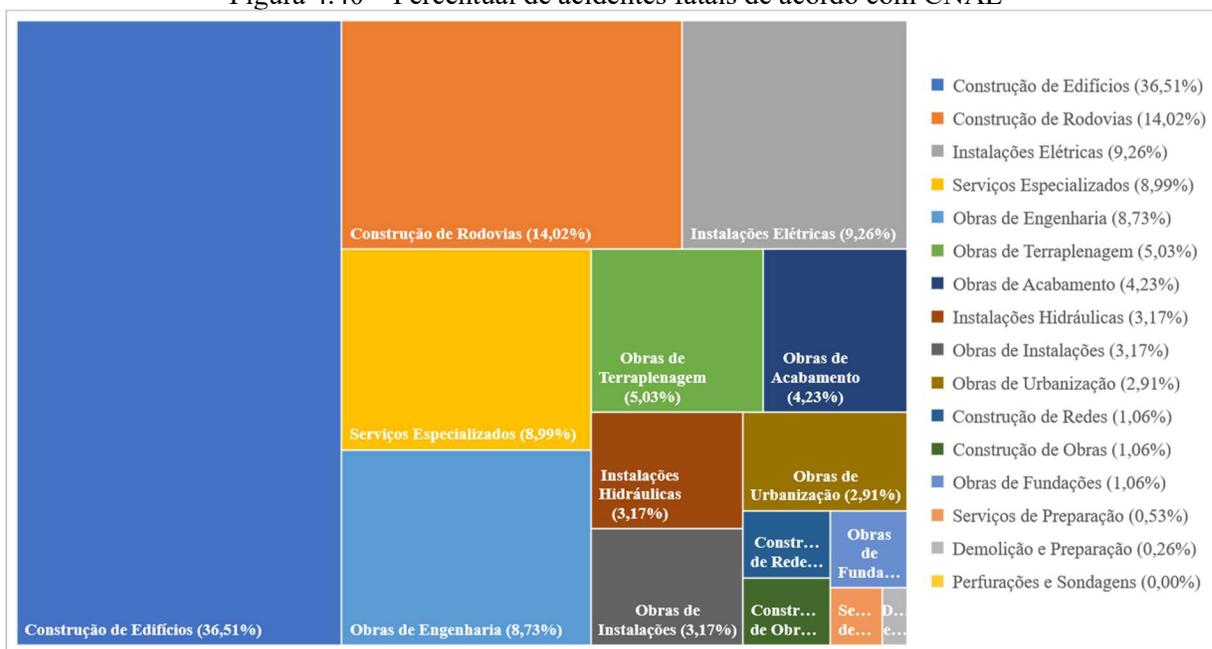


Fonte: Autora

CNAE

Na análise CNAE, foi possível identificar as atividades econômicas relacionadas aos acidentes fatais (Figura 4.40). O destaque se deu para a atividade de “construção de edifícios”, que apresentou um total de 138 mortes e percentual próximo aos 37% dos registros. Logo abaixo, observou-se que a “construção de rodovias”, com 53 fatalidades (14%) e as atividades de “instalações elétricas”, com 35 (9%).

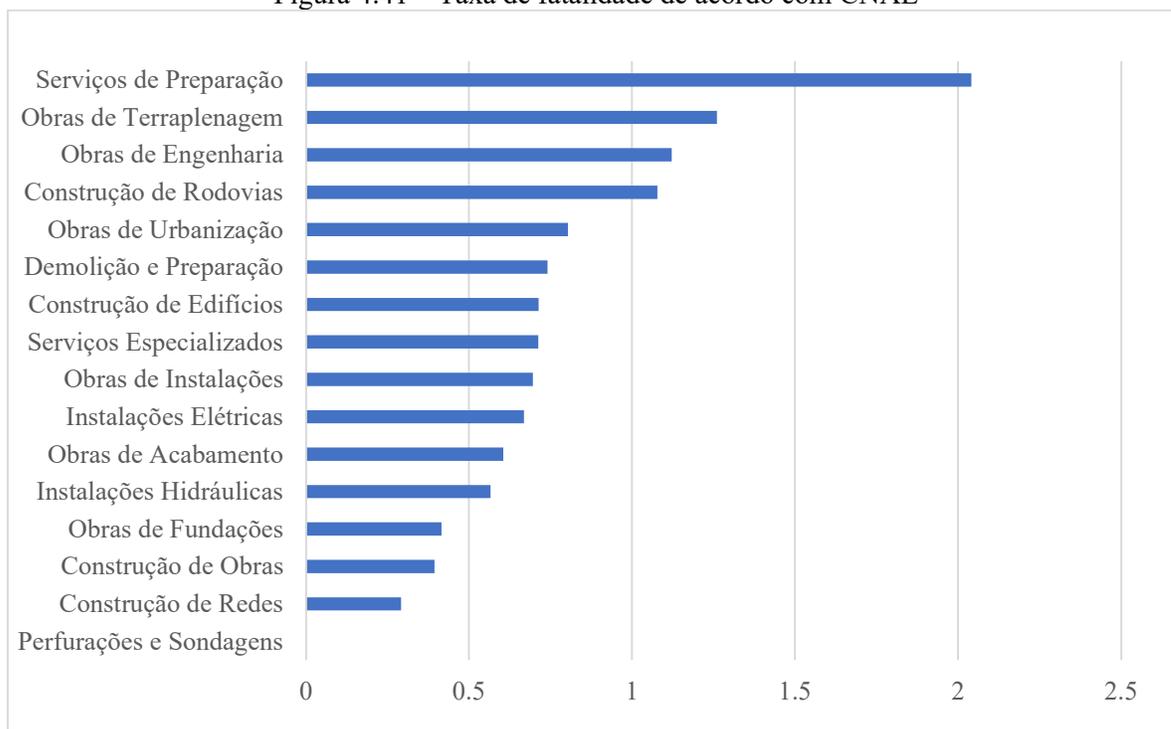
Figura 4.40 – Percentual de acidentes fatais de acordo com CNAE



Fonte: Autora

A taxa de fatalidade por atividade econômica, baseada na classificação CNAE, apresentou a relação entre a quantidade dos acidentes fatais e o total de acidentes. Dessa forma, observou-se que os “serviços de preparação” apresentaram os maiores índices (2,04%), seguidos das “obras de terraplenagem” (1,26%), “obras de engenharia” (1,12%) e atividades de “construção de rodovias” (1,08%). A Figura 4.41 abaixo apresenta o ranking das atividades econômicas de acordo com a taxa de fatalidade.

Figura 4.41 – Taxa de fatalidade de acordo com CNAE



Fonte: Autora

CBO

Em relação à CBO, observou-se que os “trabalhadores da construção civil e obras públicas” foram os mais afetados por acidentes fatais, com um total de 119 registros e percentual de cerca de 31%. Em seguida, ocupando o segundo lugar, foi visualizada a função de “ajudantes de obras”, que apresentou 77 acidentes e percentual de aproximadamente 20%. Logo abaixo, com 27 (7%) e 26 (6,9%), foram observadas as funções de “condutores de veículos e operação de equipamentos de elevação e de movimentação de cargas” e “trabalhadores de montagem de tubulações, estruturas metálicas e de compósitos”, respectivamente. O Quadro 4.14 apresenta as 10 funções que mais sofreram acidentes fatais.

Quadro 4.14 – As 10 funções mais afetadas nos acidentes fatais, de acordo com CBO

Subgrupo CBO	Descrição	Total de acidentes fatais	Percentual
715	Trabalhadores da construção civil e obras públicas	119	31.48%
717	Ajudantes de obras	77	20.37%
782	Condutores de veículos e operação de equipamentos de elevação e de movimentação de cargas	27	7.14%
724	Trabalhadores de montagem de tubulações, estruturas metálicas e de compósitos	26	6.88%
514	Trabalhadores nos serviços de administração, conservação e manutenção de edifícios e logradouros	12	3.17%
732	Instaladores e reparadores de linhas e cabos elétricos e de comunicações	11	2.91%
710	Supervisores da extração mineral e da construção civil	11	2.91%
992	Trabalhadores elementares da manutenção	8	2.12%
312	Técnicos em construção civil, de edificações e obras de infraestrutura	6	1.59%
951	Eletricistas eletrônicos de manutenção industrial, comercial e residencial	5	1.32%

Fonte: Autora

Na análise da taxa de fatalidade relacionada às funções dos acidentados, observou-se que “profissionais em navegação aérea, marítima e fluvial” e “reparadores de instrumentos e equipamentos de precisão” obtiveram percentual de 100%, enquanto os “trabalhadores da mecanização florestal”, “supervisores dos serviços” e “diretores de áreas de apoio” apresentaram taxas de 10%. As 10 funções com maiores taxas de fatalidade podem ser visualizadas conforme o Quadro 4.15 abaixo.

Quadro 4.15 – Ranking das 10 funções com maior taxa de fatalidade, de acordo com CBO

Subgrupo CBO	Descrição	Taxa de fatalidade	Total de acidentados	Total de fatais
215	Profissionais em navegação aérea, marítima e fluvial	100.00%	2	2
915	Reparadores de instrumentos e equipamentos de precisão	100.00%	1	1
642	Trabalhadores da mecanização florestal	10.00%	10	1
510	Supervisores dos serviços	10.00%	10	1
123	Diretores de áreas de apoio	10.00%	10	1
763	Trabalhadores da confecção de roupas	8.33%	12	1
752	Vidreiros, ceramistas e afins	7.14%	14	1
712	Trabalhadores de beneficiamento de minérios e pedras ornamentais	3.92%	51	2
214	Engenheiros, arquitetos e afins	3.74%	107	4
774	Trabalhadores de montagem de móveis e artefatos de madeira	3.03%	33	1

Fonte: Autora

Agentes causadores

De acordo com a análise conduzida sobre os agentes causadores (Quadro 4.16), observou-se que do total de 378 mortes, os “veículos rodoviários motorizados” foram evidenciados em 63 registros, representando aproximadamente 17% do total das fatalidades. Os “veículos não informados ou classificados” e as “motocicletas, motonetas” também se relacionaram à 39 (10,32%) e 25 (6,61%) mortes, respectivamente. Outros agentes como “quedas de pessoas com diferença de nível”, “andaimes, plataformas”, “energia” e “máquinas de terraplenagem e construção de estrada” também se destacaram com 17, 14, 14 e 13 mortes, respectivamente.

Quadro 4.16 – Os 10 agentes causadores mais recorrentes nos acidentes fatais

Agentes causadores	Total de acidentes fatais	Percentual
Veículo rodoviário motorizado	63	16.67%
Veículo não especificado ou classificado	39	10.32%
Motocicleta, motoneta	25	6.61%
Queda de pessoa com diferença de nível	17	4.50%
Andaime, plataforma - edifício ou estrutura	14	3.70%
Energia	14	3.70%
Máquina de terraplenagem e construção de estrada	13	3.44%
Telhado	11	2.91%
Produto mineral não metálico - produto de mineração, escavação, desbarrancamento, etc	10	2.65%
Escavação (para edifício, estrada, etc.)	10	2.65%

Fonte: Autora

Em relação à taxa de fatalidade dos agentes (Quadro 4.17), percentual entre o total de acidentes relacionados ao agente causador e o número de mortes registradas, observou-se que os “diques e barragens” foram os mais letais, com percentual de 35%. Em seguida, observou-se que as “aeronaves”, os “produtos minerais metálicos”, as “pás mecânicas, dragas” e os “elevadores de caçamba” também apresentaram alta letalidade, com percentuais próximos de 29%, 22%, 11% e 10%, respectivamente. Outro destaque se deu para os agentes “energia” e “transformador, conversor”, com cerca de 10% e 7% e a “escavação”, agente que além de estar entre os mais recorrentes, também apareceu entre os mais letais. O percentual dos 10 agentes mais letais pode ser visualizado no Quadro 4.17.

Quadro 4.17 – Os 10 agentes causadores mais letais

Agentes causadores	Percentual	Total de acidentes fatais
Dique, barragem - edifício ou estrutura	35.00%	7
Aeronave	28.57%	2
Produto mineral metálico - produto de mineração em bruto ou beneficiado	22.22%	6
Pá mecânica, draga - equipamento de guindar	10.53%	2
Elevador de caçamba para mineração - equipamento de guindar	10.00%	1
Energia	9.79%	14
Transformador, conversor - equipamento elétrico	7.14%	2
Forno, estufa, retorta, aquecedor de ambiente, fogão, etc.	6.90%	2
Veículo deslizante	6.90%	2
Escavação (para edifício, estrada, etc.)	6.85%	10

Fonte: Autora

Naturezas de lesões

De acordo com a análise conduzida sobre a natureza das lesões, observou-se que as “lesões múltiplas” foram evidenciadas em 164 registros de mortes, com uma taxa de fatalidade de 10,65%. Outros tipos como “lesões imediatas não informadas ou classificadas” e “fraturas” foram evidenciados em muitos registros, 41 e 37, respectivamente, no entanto, suas taxas de fatalidade foram baixas quando comparadas às demais. O destaque da taxa de fatalidade se deu para a categoria “asfixia, estrangulamento”, onde 25 dos registros, representando 53,19% do total, ocasionaram mortes. Ainda em relação à taxa de fatalidade, destacaram-se a “concussão cerebral” e “choque elétrico e eletroplessão (eletrocussão)”, com valores de 13,88% e 10,09%, respectivamente. As demais naturezas e os respectivos números de mortes e taxas de fatalidade podem ser observadas no Quadro 4.18.

Quadro 4.18 – Total de mortes e taxa de fatalidade a respeito da natureza das lesões

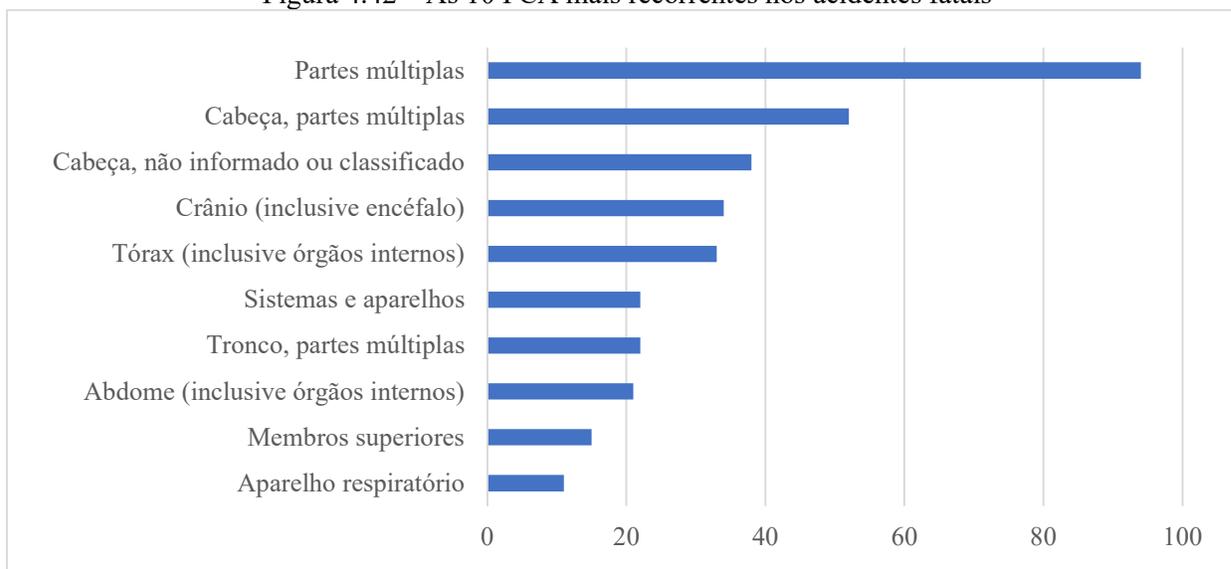
Natureza da lesão	Total de mortes registradas	Taxa de fatalidade
Lesões múltiplas	164	10.65%
Lesão imediata não informada ou classificada	41	0.95%
Fratura	37	0.30%
Choque elétrico e eletroplessão (eletrocussão)	35	10.09%
Concussão cerebral	29	13.88%
Asfixia, estrangulamento, afogamento	25	53.19%
Contusão, esmagamento	24	0.40%
Corte, laceração, ferida contusa, punctura	7	0.07%
Lesão imediata	4	0.61%
Doença não informada ou classificada	4	1.56%
Amputação ou enucleação	3	0.45%
Escoriação, abrasão	2	0.05%
Luxação	1	0.03%
Envenenamento	1	1.11%
Outras lesões não informadas ou classificadas	1	1.79%

Fonte: Autora

Parte do corpo atingida

Em relação à PCA, foi observado que “partes múltiplas” foram os locais mais recorrentes nas fatalidades, com 94 registros, representando cerca de 25% do total das mortes. Em seguida, representando aproximadamente 14% dos acidentes fatais, observou-se “cabeça, partes múltiplas”. Além desses locais, “tórax (inclusive órgãos internos)” e “sistemas e aparelhos” também se destacaram entre as partes do corpo mais atingidas em um acidente fatal, representando cerca de 9% e 6% dos casos, respectivamente. A Figura 4.42 abaixo apresenta as 10 PCA mais comuns nas fatalidades.

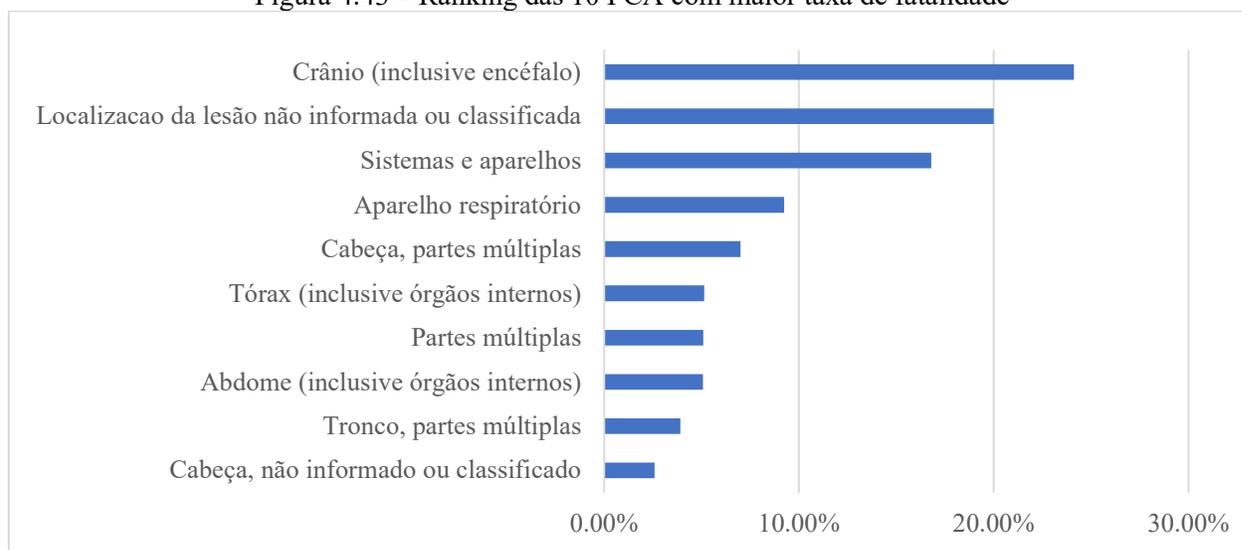
Figura 4.42 – As 10 PCA mais recorrentes nos acidentes fatais



Fonte: Autora

Em relação à taxa de fatalidade das PCA, observou-se que o “crânio (inclusive encéfalo)” apresentou maior taxa, com 24,11%. O destaque também se deu para “sistemas e aparelhos” que apresentaram taxa de fatalidade de aproximadamente 17% e “aparelho respiratório” com 9,24%. O ranking das 10 PCA com maior taxa de fatalidade pode ser visualizado na Figura 4.43.

Figura 4.43 – Ranking das 10 PCA com maior taxa de fatalidade



Fonte: Autora

CID

Sobre a análise da CID, observou-se que 231 mortes ocorreram devido a “lesões, envenenamentos e algumas outras consequências de causas externas”, no entanto, a taxa de fatalidade relacionada a esta classificação foi de 0,55%. Além disso, os “sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte” se relacionaram a 69 casos de mortes, com taxa de 12,5%. O destaque da taxa de fatalidade se deu para as “doenças do sangue e dos órgãos hematopoéticos e alguns transtornos imunitários” que representou percentual de 60% e as “doenças do aparelho circulatório” com cerca de 19%. Os demais números de mortes e taxas de fatalidade de acordo com a CID podem ser observadas no Quadro 4.19.

Quadro 4.19 – Quantidade de acidentes e taxa de fatalidade, de acordo com CID

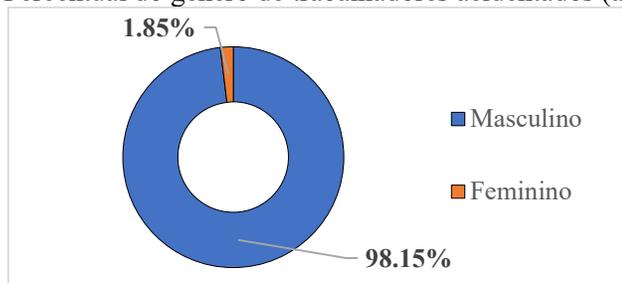
Intervalo CID	Descrição	Total de acidentes fatais	Percentual
S00 - T98	Lesões, envenenamentos e algumas outras consequências de causas externas	231	0.55%
R00 - R99	Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte	69	12.50%
V01 - Y98	Causas externas de morbidade e de mortalidade	51	2.27%
I00 - I99	Doenças do aparelho circulatório	11	18.64%
K00 - K93	Doenças do aparelho digestivo	3	2.48%
G00 - G99	Doenças do sistema nervoso	3	5.08%
D50 - D89	Doenças do sangue e dos órgãos hematopoéticos e alguns transtornos imunitários	3	60.00%
U00 - U99	Códigos para propósitos especiais	2	6.25%
M00 - M99	Doenças do sistema osteomuscular e do tecido conjuntivo	1	0.03%
H00 - H59	Doenças do olho e anexos	1	0.28%
Q00 - Q99	Malformações congênitas, deformidades e anomalias cromossômicas	1	1.02%
H60 - H95	Doenças do ouvido e da apófise mastóide	1	1.64%
J00 - J99	Doenças do aparelho respiratório	1	1.75%

Fonte: Autora

Gênero

Os acidentes fatais também foram analisados a respeito do gênero dos acidentados. Dessa forma, foi observado que 371 trabalhadores, 98,15% do total, pertenciam ao gênero masculino, enquanto 7 (1,85%) ao feminino (Figura 4.44).

Figura 4.44 – Percentual do gênero de trabalhadores acidentados (acidentes fatais)

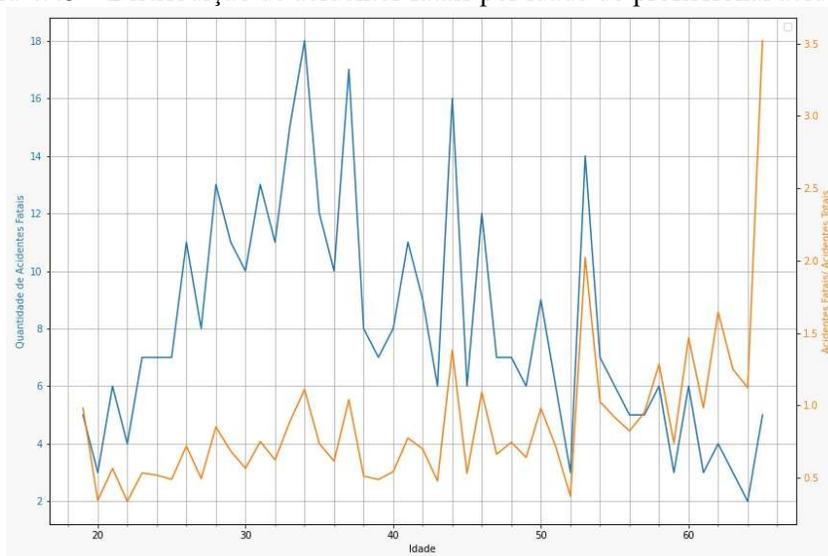


Fonte: Autora

Idade

A análise da idade dos profissionais que sofreram acidentes fatais foi analisada a partir da distribuição presente na Figura 4.45. Assim, observou-se que as idades com maior número de mortes foram a de 34 anos (18), seguida de 37 (17), 44 (16) e 33 anos (15), enquanto a com menor quantidade foi a de 64 (2) e 18 (0). Além disso, observou-se que a taxa de fatalidade registrou maiores valores com o passar das idades, registrando valores de 3,52% para 65 anos e 2,02% para os 53 anos, dando destaque também para os trabalhadores com 34 anos que apresentaram taxas de 1,11%.

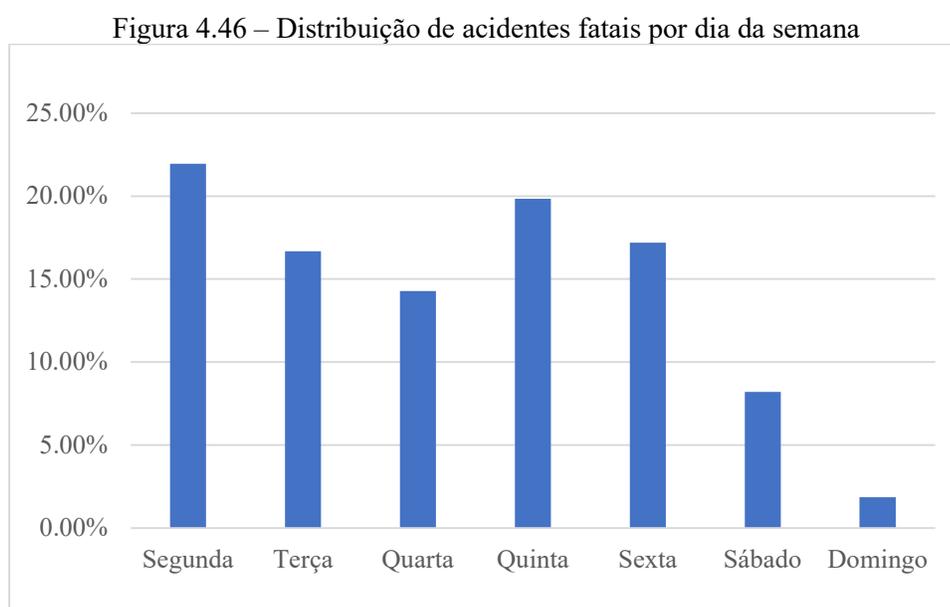
Figura 4.45 – Distribuição de acidentes fatais por idade do profissional acidentado



Fonte: Autora

Data do acidente

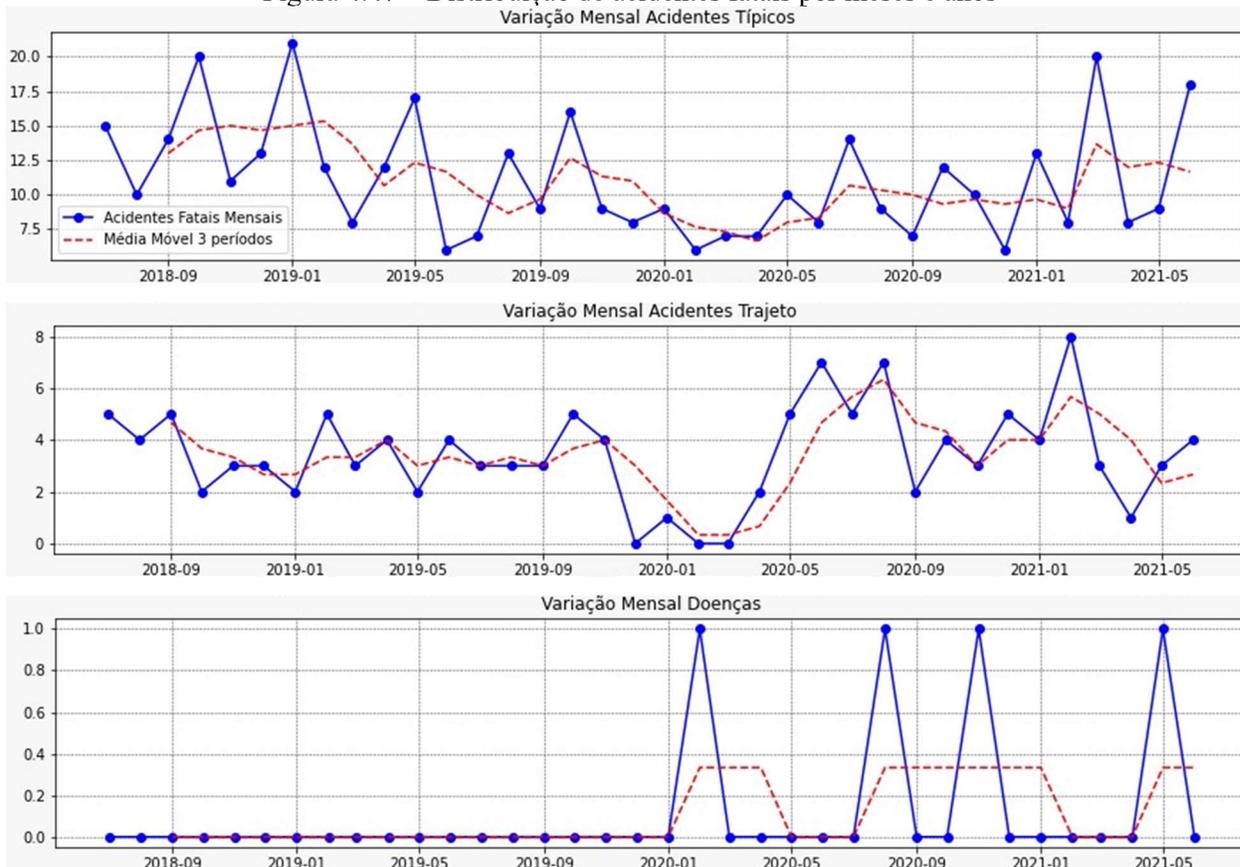
Em relação aos dias da semana de ocorrência dos acidentes fatais, foi observado que a segunda obteve o maior número de acidentes (83, 21,96%) e o domingo o menor número (7, 1,85%). Ao longo da semana, também foi possível observar um pico de acidentes nas quintas (75, 19,84%) e sextas (65, 17,20%), que apresentaram taxas maiores que a terça e quarta. Os finais de semana, entretanto, apresentaram menor recorrência. A Figura 4.46 apresenta a distribuição dos acidentes fatais ao longo da semana.



Fonte: Autora

A análise dos meses que ocorreram acidentes fatais foi dividida de acordo com o tipo do acidente (Figura 4.47). Para os acidentes típicos que geraram mortes, observou-se que os meses de janeiro/19, outubro/18 e março/21 tiveram maior número de fatalidades, enquanto os períodos de junho/19, fevereiro/20 e dezembro/20 apresentaram menor quantidade. No geral, observou-se que os acidentes fatais tiveram uma redução em 2020, comparados aos anos anteriores, mas tenderam a crescer novamente em 2021. Nos acidentes de trajeto fatais, foi observado uma queda expressiva no início de 2020, período relativo à pandemia, havendo um crescimento a partir da metade deste ano e atingindo o ápice em fevereiro de 2021, com 8 fatalidades. Sobre as doenças do trabalho, observou-se 4 mortes ocorridas no ano de 2020 (3) e 2021 (1).

Figura 4.47 – Distribuição de acidentes fatais por meses e anos



Fonte: Autora

4.2.4 Seleção dos atributos

Os resultados da análise exploratória e a compreensão das informações das CAT evidenciaram atributos não tão significativos, sendo assim, optou-se por excluir estas informações para conferir melhores resultados ao modelo. Dessa forma, foram utilizados os atributos: agente causador do acidente, CBO do acidentado, CID, CNAE do empregador, emitente da CAT, espécie do benefício, filiação do segurado, indicador de acidente fatal, município do empregador, natureza da lesão, parte do corpo atingida, gênero, tipo do acidente, estado onde ocorreu o acidente, estado da localização do empregador e idade (apenas utilizado no algoritmo *K-Prototype*). As Figuras 4.48 e 4.49 apresentam os atributos utilizados na aplicação dos algoritmos selecionados.

Figura 4.48 – Atributos selecionados para aplicação do *K-Modes*

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 16179 entries, 560 to 118033
Data columns (total 15 columns):
#   Column                               Non-Null Count  Dtype
---  -
0   Agente Causador Acidente             16179 non-null  object
1   CBO.1                                  16179 non-null  object
2   CID-10                                 16179 non-null  object
3   CNAE2.0 Empregador.1                 16179 non-null  object
4   Emitente CAT                          16179 non-null  object
5   Espécie do benefício                 16179 non-null  object
6   Filiação Segurado                   16179 non-null  object
7   Indica acidente                       16179 non-null  object
8   Munic Empr                            16179 non-null  object
9   Natureza da Lesão                    16179 non-null  object
10  Parte Corpo Atingida                 16179 non-null  object
11  Sexo                                  16179 non-null  object
12  Tipo do Acidente                     16179 non-null  object
13  UF Munic. Acidente                   16179 non-null  object
14  UF Munic. Empregador                 16179 non-null  object
dtypes: object(15)
memory usage: 2.0+ MB

```

Fonte: Autora

Figura 4.49 – Atributos selecionados para aplicação do *K-Prototype*

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 16179 entries, 560 to 118033
Data columns (total 16 columns):
#   Column                               Non-Null Count  Dtype
---  -
0   Agente Causador Acidente             16179 non-null  object
1   CBO.1                                  16179 non-null  object
2   CID-10                                 16179 non-null  object
3   CNAE2.0 Empregador.1                 16179 non-null  object
4   Emitente CAT                          16179 non-null  object
5   Espécie do benefício                 16179 non-null  object
6   Filiação Segurado                   16179 non-null  object
7   Indica acidente                       16179 non-null  object
8   Munic Empr                            16179 non-null  object
9   Natureza da Lesão                    16179 non-null  object
10  Parte Corpo Atingida                 16179 non-null  object
11  Sexo                                  16179 non-null  object
12  Tipo do Acidente                     16179 non-null  object
13  UF Munic. Acidente                   16179 non-null  object
14  UF Munic. Empregador                 16179 non-null  object
15  Idade                                 16179 non-null  int64
dtypes: int64(1), object(15)
memory usage: 2.1+ MB

```

Fonte: Autora

4.2.5 Aplicação do algoritmo

Nesta etapa, foram selecionados e aplicados algoritmos para agrupar acidentes de trabalho de acordo com suas características. Considerando a grande quantidade de dados e o fato de os atributos serem em sua maioria categóricos, optou-se por utilizar o algoritmo de *clustering* K-Modes. O modelo, proposto por Huang (1998), é conhecido por sua eficiência em conjuntos de dados extensos, onde usa uma medida de dissimilaridade para lidar com objetos categóricos, utilizando, no processo de *clustering*, um método baseado em frequência para atualizar as modas. Além disso,

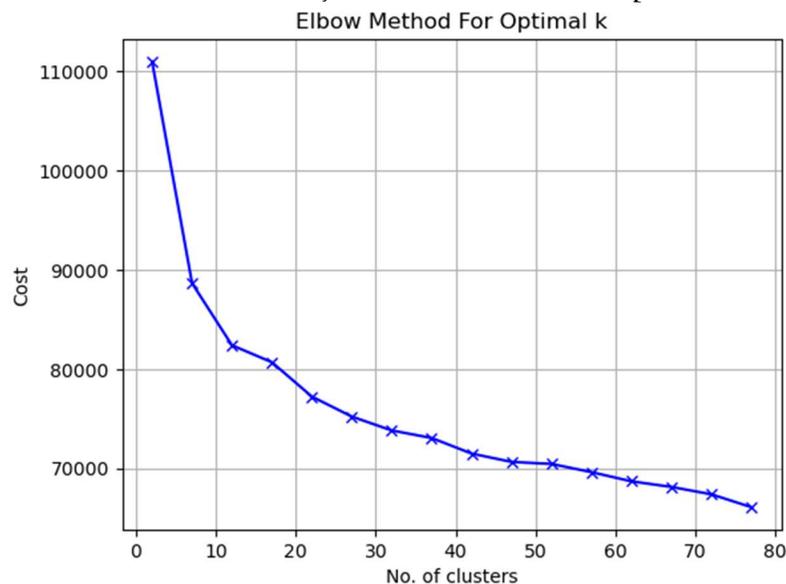
optou-se por utilizar o algoritmo de *clustering* K-Prototype, também proposto por Huang (1997), para lidar com a combinação dos dados categóricos e numéricos. Os códigos dos algoritmos, por sua vez, foram obtidos da biblioteca *Scikit-Learn*.

Os algoritmos foram utilizados para os dados de acidentes relacionados à construção de edifícios, devido a sua representação expressiva nos resultados da AED. Para determinar o número ideal de *clusters*, foi utilizado o método do cotovelo (*Elbow method*), que levou em consideração o custo (soma das diferenças entre os *clusters*) e o intervalo de número de *clusters*, sendo aplicado para os dois algoritmos selecionados.

4.2.5.1 K-Modes

Por meio da análise do gráfico no *K-Modes* (Figura 4.50), verificou-se que os custos apresentaram uma diminuição relevante com o aumento de *clusters*, mas após certo momento, não foi observado um *plateau*. Percebeu-se, então, que o algoritmo não aderiu adequadamente aos dados, visto que o método do cotovelo não apresentou resultados satisfatórios e que, mesmo utilizando a quantidade de *clusters* mais interessante pelo gráfico, próxima ao número de 65, não seria viável proceder a análise com um alto número de *clusters*.

Figura 4.50 – Custo do K-Modes em função do número de *clusters* para os acidentes de trabalho

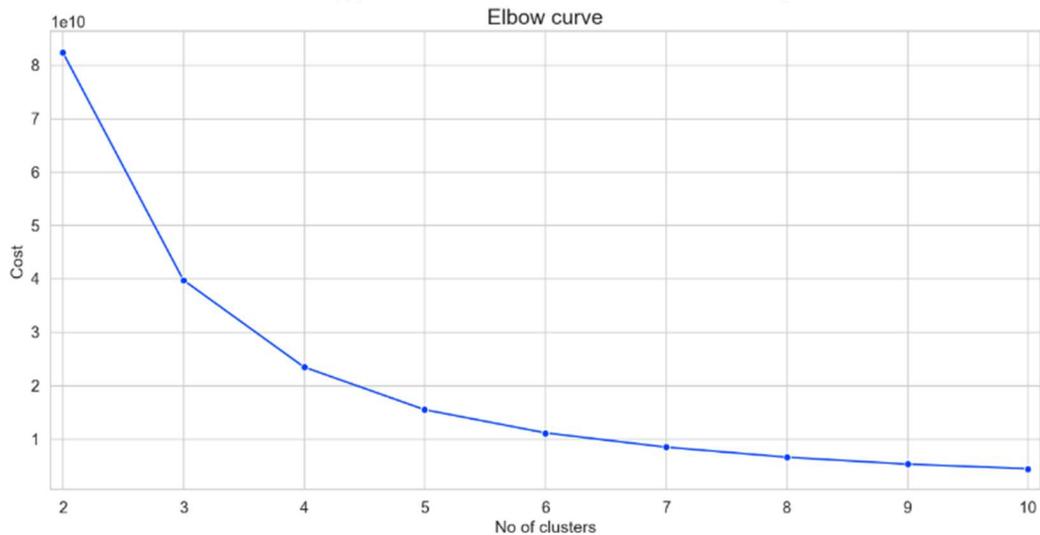


Fonte: Autora

4.2.5.2 K-Prototype

Devido ao resultado do *K-Modes*, optou-se por adicionar os atributos numéricos e aplicar o *K-Prototype* para lidar com os dois tipos de dados. Assim, na análise do método do cotovelo para o *K-Prototype*, conforme pode ser observado na Figura 4.51, verificou-se um resultado mais interessante, onde o número de 5 *clusters* (*k*) foi o mais adequado para representar a quantidade de grupos.

Figura 4.51 – Custo do K-Prototype em função do número de *clusters* para os acidentes de trabalho



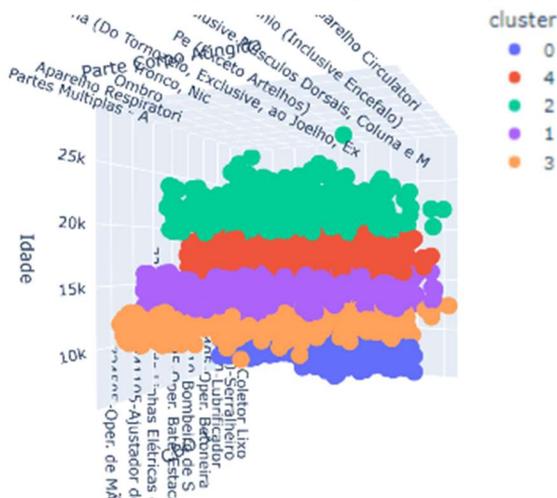
Fonte: Autora

4.2.6 Resultados do modelo

Após a definição do número de *clusters* (*k*) utilizando o algoritmo *K-Prototype*, foram gerados gráficos de dispersão em 3D relacionando as variáveis mais significativas das CATs (CBO, agente causador, natureza da lesão, parte do corpo atingida, idade) para os 5 *clusters* criados, com o objetivo de verificar se os grupos estariam distintos entre si (Figuras 4.52 a 4.61).

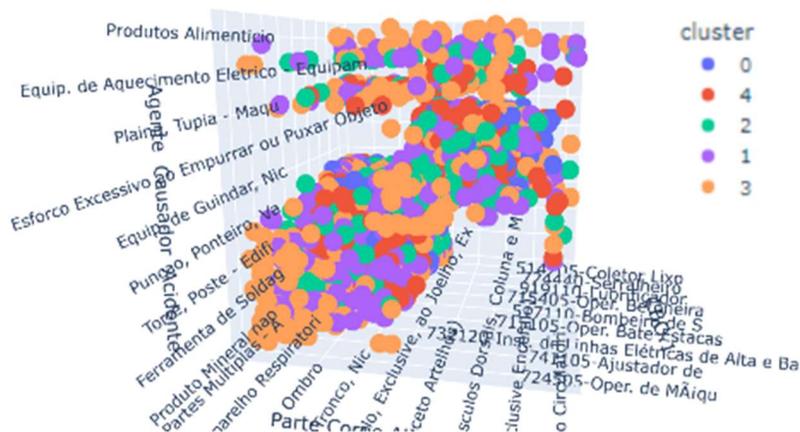
Observou-se que os gráficos com o atributo numérico “idade” (Figuras 4.53, 4.55, 4.57, 4.59, 4.60 e 4.61) evidenciaram clusters bem definidos, indicando um resultado satisfatório para o algoritmo de *clustering*, no entanto, os demais apresentaram grupos bastante heterogêneos, o que refutou a qualidade dos agrupamentos.

Figura 4.55 – Gráfico de dispersão CBO, parte do corpo atingida e idade



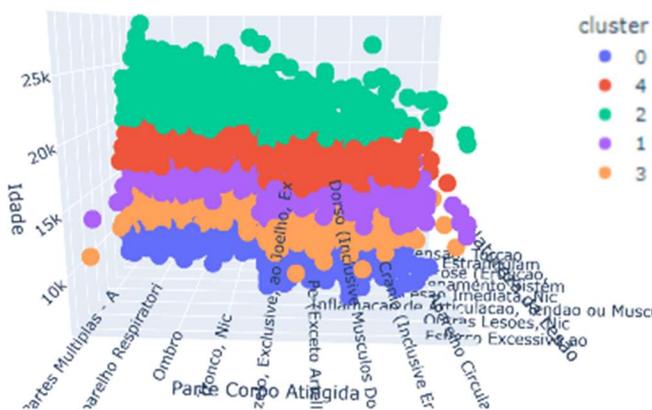
Fonte: Autora

Figura 4.56 – Gráfico de dispersão CBO, parte do corpo atingida e agente causador



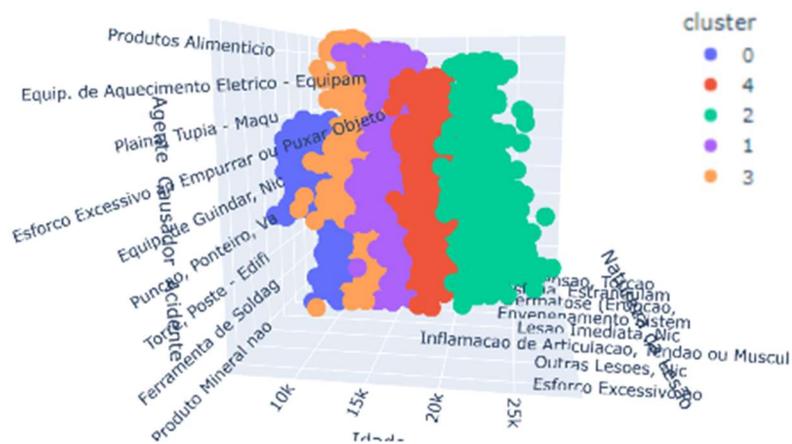
Fonte: Autora

Figura 4.57 – Gráfico de dispersão natureza da lesão, parte do corpo atingida e idade



Fonte: Autora

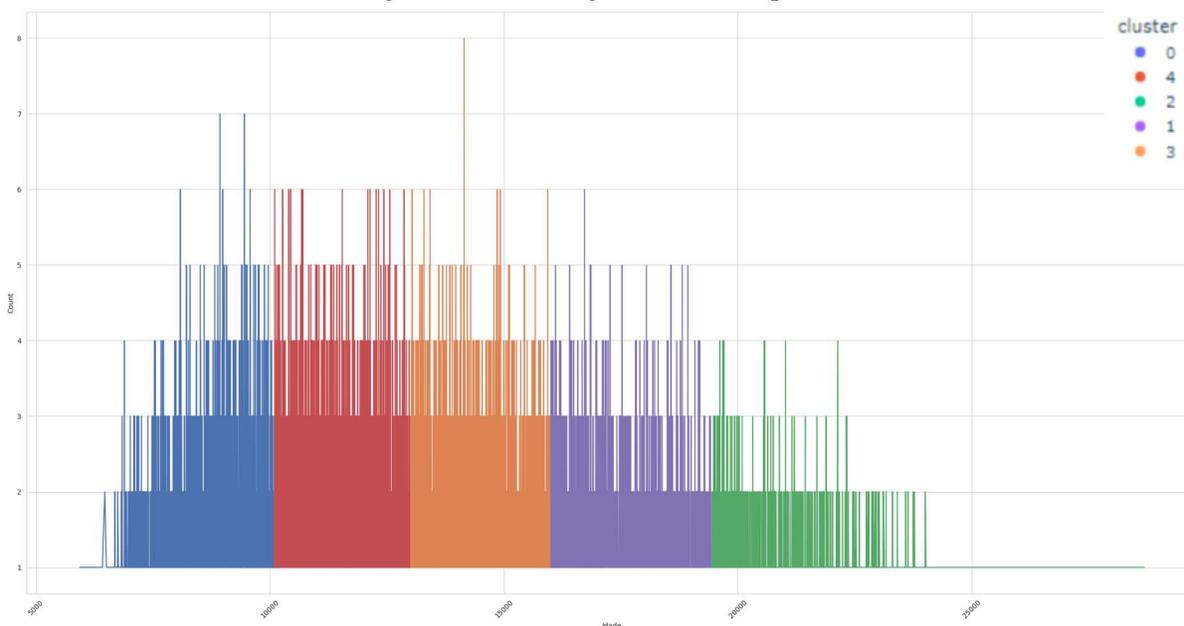
Figura 4.61 – Gráfico de dispersão idade, agente causador e natureza da lesão



Fonte: Autora

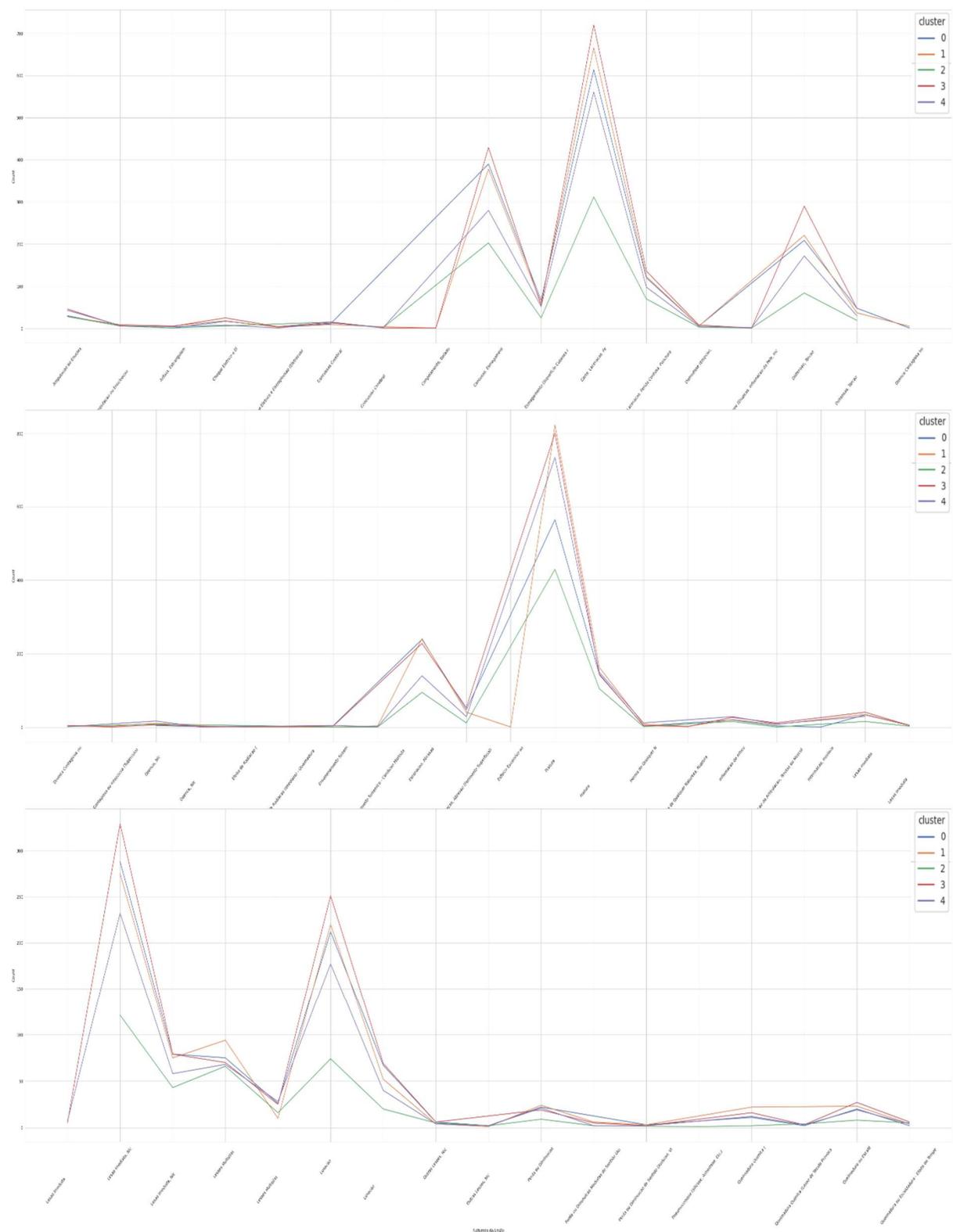
Para confirmar a hipótese acima, foram analisadas as distribuições dentro de cada *cluster* para cada uma das 5 colunas observadas, sendo gerado um histograma para representar as informações (Figuras 4.62 a 4.66), alinhados em 3 linhas para uma melhor visualização. Assim, para a variável “idade”, conforme já havia sido sugerido, a distribuição indicou resultados satisfatórios, no entanto, para os demais atributos, observou-se uma dificuldade do algoritmo em agrupar de forma adequada os dados utilizados, devido aos clusters seguirem basicamente a mesma distribuição, impedindo a identificação de grupos prioritários, hipótese proposta por este trabalho.

Figura 4.62 – Histograma da idade por *cluster*



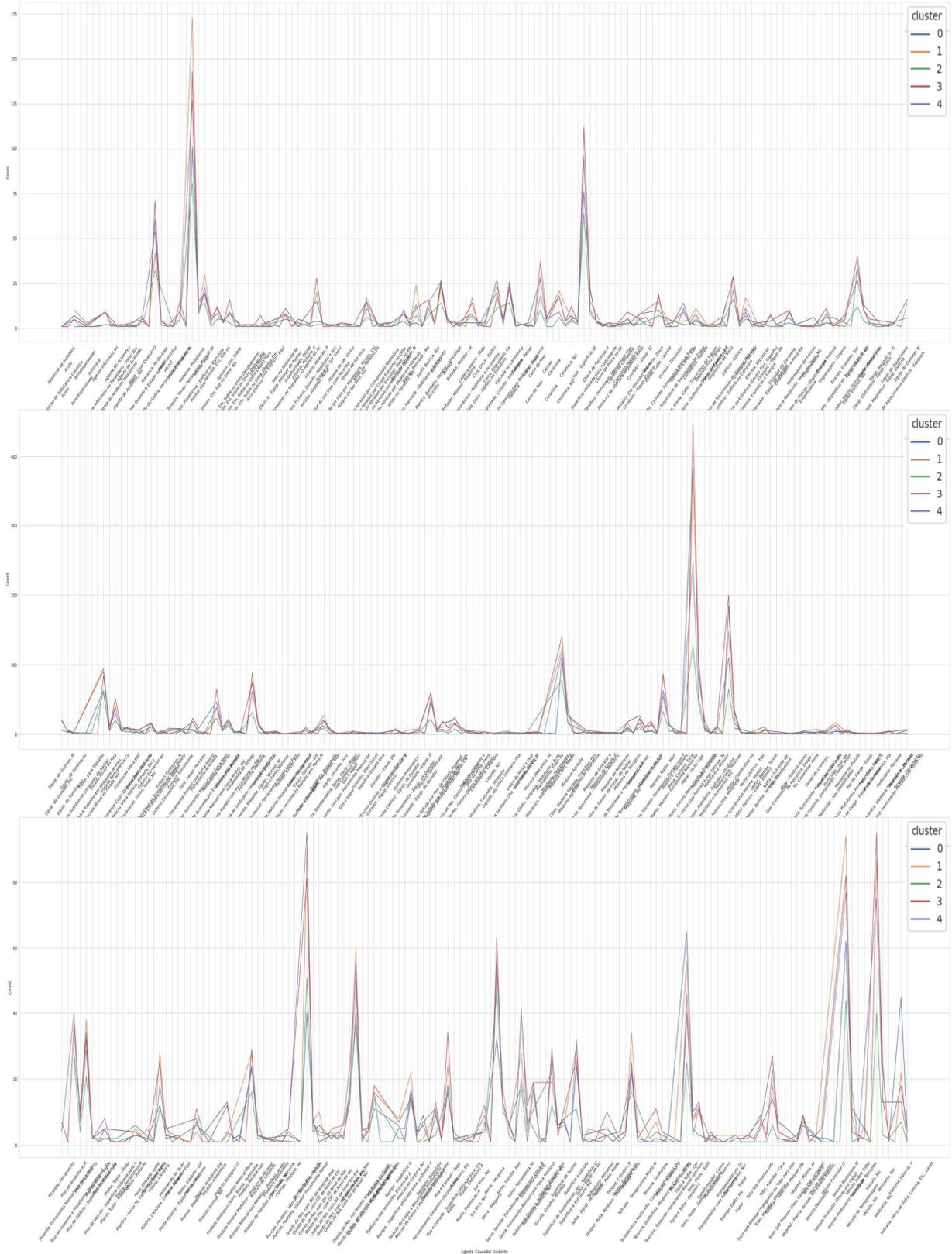
Fonte: Autora

Figura 4.63 – Histograma da natureza da lesão por *cluster*



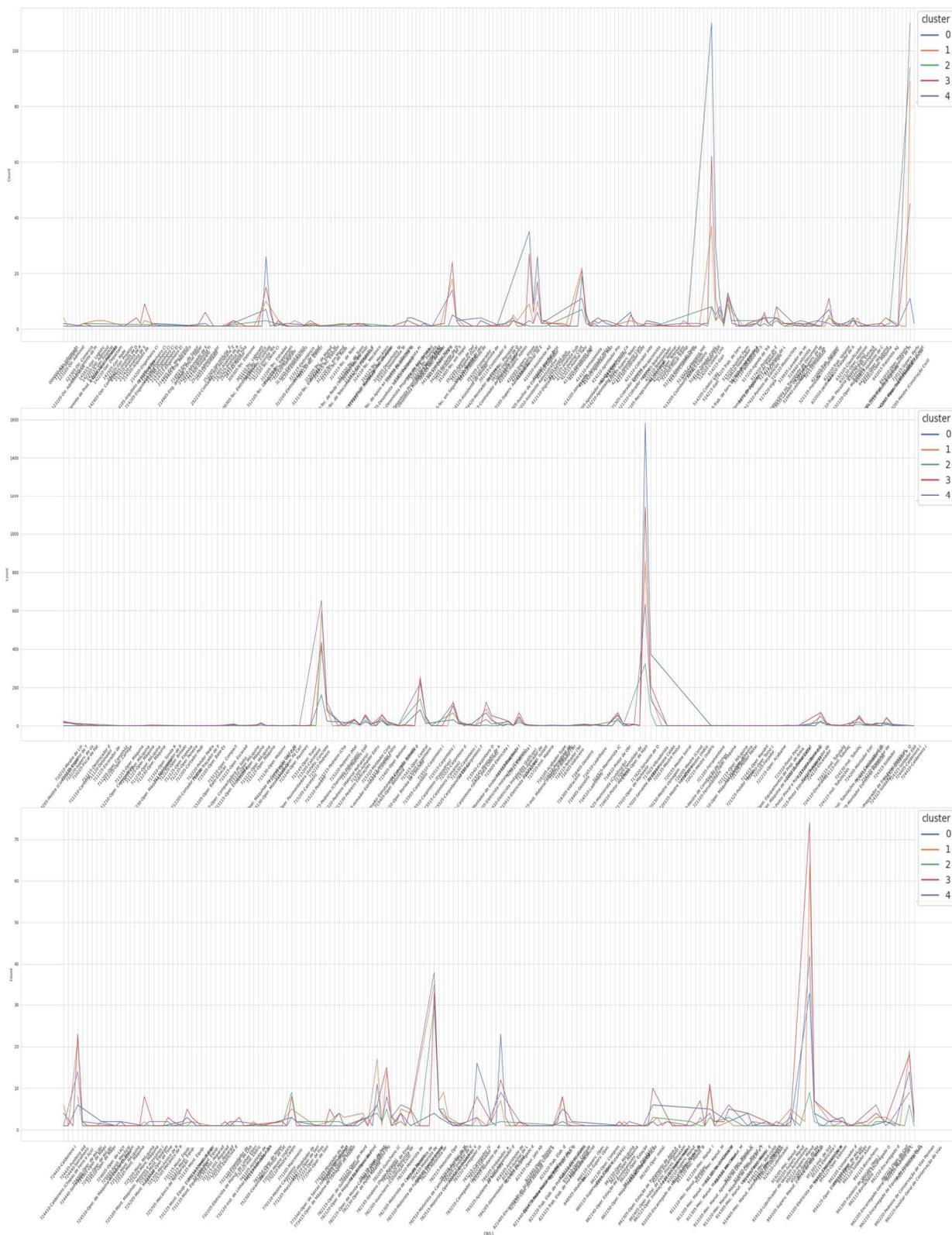
Fonte: Autora

Figura 4.65 – Histograma do agente causador por *cluster*



Fonte: Autora

Figura 4.66 – Histograma do CBO por *cluster*



Fonte: Autora

5 DISCUSSÃO

A análise das características dos acidentes de trabalho da construção é algo fundamental para o entendimento dos fatores gerais responsáveis pela ocorrência desses eventos. Por meio dessas informações, torna-se possível traçar planos e estratégias específicas para eliminar ou diminuir os riscos presentes nas diversas atividades do setor, contribuindo para a redução dos acidentes e doenças ocupacionais. Dessa forma, este trabalho analisou dados de acidentes de trabalho da construção ocorridos no Brasil entre 2018 e 2021, registrados por meio de CAT, com o auxílio de técnicas de aprendizado de máquina. A utilização de métodos automatizados, por sua vez, possibilitou a análise de grande quantidade de dados de forma mais eficiente que o processo manual, com menos tempo para operação e maior qualidade nas informações.

Inicialmente, foram analisados 50.380 registros de acidentes de trabalho relacionados às diversas atividades presentes no escopo da construção civil, como a construção de edifícios, rodovias, obras de acabamento, terraplenagem e instalações. Esta análise, conduzida de forma exploratória, foi realizada levando em consideração a divisão entre os tipos de acidentes de trabalho, conforme consta na legislação brasileira: os acidentes típicos, de trajeto e as doenças ocupacionais. Observou-se que cerca de 85% dos registros foram referentes aos acidentes típicos, enquanto os acidentes de trajeto (14%) e as doenças (1%) apresentaram índices menores. Além dos tipos de acidentes, foram conduzidas análises sobre as fatalidades em decorrência de um acidente de trabalho, assim, observou-se 378 registros de acidentes fatais, o que representou menos de 1% dos casos.

Em relação aos acidentes típicos, foi observado que as atividades de construção de edifícios, rodovias, instalações elétricas e serviços especializados, juntas, representaram a maioria das ocorrências (70%). Observou-se também que cerca de 80% dos acidentes foram causados por agentes mecânicos (ex. atividades com máquinas/equipamentos/ferramentas, situações de quedas/impactos) e que as lesões mais comuns foram cortes/lacerações/feridas contusas/puncturas, fraturas e contusões/esmagamentos, enquanto as partes do corpo mais afetadas foram os dedos, pés e mãos. Sobre o perfil do trabalhador que sofreu um acidente típico, observou-se que os trabalhadores da construção civil/obras públicas e os ajudantes de obras do gênero masculino da faixa etária entre 31 e 40 anos foram os mais afetados. Também foi observado que com o aumento

da idade dos profissionais, os níveis de acidente diminuíram de forma constante, o que pode ser justificado pela experiência profissional adquirida que faz com que os trabalhadores adotem comportamentos mais seguros, fato evidenciado em algumas pesquisas como Alwasel *et al.* (2017).

Na análise dos acidentes de trajeto, assim como nos típicos, foi observado que as atividades de construção de edifícios, instalações elétricas, serviços especializados e construção de rodovias, representaram a maioria das ocorrências. Em relação às causas, observou-se que os agentes mecânicos predominaram entre os demais, chegando a percentual próximo dos 95%, o que pode ser justificado pelo fato deste tipo de acidente ocorrer nos deslocamentos dos trabalhadores entre sua residência e o local de trabalho, sendo motocicletas/mononetas e veículos os meios normalmente adotados para o transporte. As lesões mais recorrentes neste tipo de acidente foram as fraturas, escoriações/abrasões e cortes/esmagamentos, enquanto as partes do corpo mais atingidas foram às partes múltiplas, joelhos, pés, pernas e ombros. O perfil do profissional acidentado mostrou que os trabalhadores da construção civil/obras públicas e os ajudantes de obras do gênero masculino com idade entre 25 e 35 anos foram os que mais se acidentaram, apresentando menor faixa etária comparado aos acidentes típicos. Esta faixa etária, por sua vez, é normalmente relacionada a uma maior acidentabilidade no trânsito, principalmente se o meio de transporte for motocicleta, algo evidenciado em trabalhos como Silva (2008) e Duarte (2011).

As doenças ocupacionais normalmente são fruto de exposições prolongadas a determinados fatores de risco, assim, acabam possuindo características bem específicas que se diferenciam dos demais tipos de acidentes de trabalho. De acordo com a análise conduzida a respeito dos agentes causadores de doenças, observou-se que esforços excessivos, ataques de seres vivos, exposições ao ruído e agentes infecciosos/parasitários foram as mais relatadas. Observando as lesões ocorridas, constatou-se que as perdas ou diminuições mediatas de sentido, inflamações de articulação/tendão/músculo e as hérnias de qualquer natureza foram as mais recorrentes, e as partes do corpo como os dorsos (inclusive músculos dorsais, coluna e medula espinhal), ouvidos (externo, médio, interno, audição e equilíbrio) e abdome (inclusive órgãos internos) foram mais atingidas. As atividades econômicas, no entanto, se assemelharam aos demais tipos de acidentes, destacando-se a construção de edifícios, rodovias, obras e os serviços especializados. A respeito do perfil profissional, foi visto que os trabalhadores da construção civil/obras públicas e os ajudantes de

obras do gênero masculino com faixa etária entre 35 e 45 anos foram os que mais adoeceram. Por sua vez, o aumento da faixa etária pode ter relação com o fato das doenças ocupacionais muitas vezes serem diagnosticadas após anos de exposição e pela chance de desenvolver uma doença deste tipo ser normalmente maior conforme o envelhecimento, como o caso dos transtornos musculoesqueléticos evidenciados por Kines *et al.* (2007).

Dentre os casos de acidentes fatais, foi possível constatar maior participação das atividades econômicas que mais se repetiram nos outros tipos de acidentes, no entanto, observou-se que a taxa de fatalidade (relação entre a quantidade dos acidentes fatais e o total de acidentes) foi mais expressiva para os serviços de preparação, obras de terraplenagem, obras de engenharia e construção de rodovias. Dessa forma, entende-se que por mais que estas atividades tenham apresentado menores números de acidentes no geral, elas são consideradas mais perigosas que as demais, pois têm maior potencial de agravo de lesões, podendo gerar situações extremas como as fatalidades. Sobre os agentes causadores, foi possível observar que os veículos, motocicletas/motonetas e as quedas com diferença de nível foram as principais causas de mortes relatadas, no entanto, os com maior taxa de fatalidade foram as atividades envolvendo equipamentos de guindar, energia e os serviços de escavação. A análise da natureza das lesões identificou que lesões múltiplas/imediatas, fraturas e choques foram as mais recorrentes, todavia, as mais letais foram relacionadas à asfixia/estrangulamento/afogamento e concussão cerebral. A respeito das partes do corpo, partes múltiplas e cabeça foram mais atingidas nos casos de mortes, enquanto o crânio e sistemas/aparelhos apresentaram maior letalidade. Analisando o perfil profissional, foi visto que os trabalhadores da construção civil/obras públicas e os ajudantes de obras do gênero masculino com faixa etária entre 31 e 40 anos foram os que mais sofreram acidentes fatais, no entanto, estas funções apresentaram taxas de fatalidade menores que 1%. Dentre as profissões mais letais, destacaram-se os profissionais ligados a cargos de gestão, como supervisores de serviços, diretores de áreas de apoio, engenheiros e arquitetos.

Além das análises conduzidas neste trabalho, algumas estatísticas utilizadas pelo Ministério do Trabalho e Previdência – MTP, como a Relação Anual de Informações Sociais – RAIS, podem agregar ao estudo dos acidentes de trabalho que ocorrem no Brasil. Em relação ao gênero dos trabalhadores, por exemplo, verificou-se que entre 2018 e 2021, a população masculina representou

cerca de 90% do total de trabalhadores da construção civil. Assim, considerando os dados dos acidentes típicos, de trajeto e das doenças ocupacionais, verificou-se que proporcionalmente os homens se acidentam mais do que as mulheres. Ainda conforme mostra a RAIS nos anos de 2018 a 2021, cerca de 33% da mão de obra da construção civil foi relacionada à construção de edifícios, o que justifica a alta frequência verificada dessa atividade na análise dos acidentes. Somado a isso, durante o período de 2018 a 2021, cerca de 10% das atividades da construção civil se relacionaram com construção de rodovias, ferrovias, obras urbanas e obras de arte especiais, enquanto aproximadamente 4% foram serviços de demolição e preparação do terreno. Essas informações, por sua vez, confirmam que os serviços de infraestrutura, como os de preparação, terraplenagem e construção de rodovias, são de fato mais letais que as demais atividades da construção civil, por estarem em menor frequência em relação ao número de trabalhadores e mesmo assim apresentarem muitas ocorrências de fatalidades.

Normalmente, as atividades da construção civil acontecem de segunda à sexta, nos horários das 7 às 17 horas, com a exceção da sexta-feira que geralmente possui expediente reduzido, até às 16. Algumas atividades também podem se estender durante a semana, com a realização de horas extras, ou nos finais de semana, com jornadas reduzidas. Dessa forma, a análise exploratória também identificou outras relações importantes relacionadas aos dias de ocorrência dos acidentes. Durante a semana, pôde-se observar em geral que a segunda-feira foi o dia com maior número de registros de acidentes, enquanto o final de semana, especialmente o domingo, apresentou os menores índices. Para os acidentes típicos, os percentuais durante a semana variaram pouco e apresentaram decaimento ao longo dos dias, no entanto, para os de trajeto, apesar da pequena variação entre os dias, observou-se maiores valores nas quartas e sextas-feira; este último pode estar relacionado à urgência dos profissionais de chegar em suas residências, fazendo com que se envolvam em acidentes de trânsito. Nos acidentes fatais, foi possível observar outra distribuição, onde o início da semana, relativo à segunda, e os dias finais da semana, relativos à quinta e sexta, apresentaram maiores índices comparados aos outros dias. As fatalidades ao final da semana, por sua vez, podem estar relacionadas ao cansaço dos funcionários após jornadas exaustivas de trabalho, fazendo com que adotem comportamentos menos seguros. Além disso, na análise dos meses de ocorrência dos acidentes, observou-se que o mês de outubro foi relacionado a maiores números de acidentes e que o mês de dezembro normalmente apresentou menores índices, este último, no entanto, pode estar

relacionado aos feriados e férias coletivas que normalmente ocorrem neste período. A distribuição dos meses também evidenciou o período de pandemia no primeiro semestre de 2020, mostrando que a paralisação das atividades da construção afetou de forma significativa o número de acidentes de trabalho no setor, principalmente os relacionados ao trajeto.

Após analisar os casos de forma ampla, foram utilizados os algoritmos de *clustering K-Modes* e *K-Prototype* para agrupar os registros de acidentes da construção de edifícios, atividade mais representativa observada, com o objetivo de criar grupos prioritários para direcionamento de ações de segurança. Devido ao fato das colunas das CAT possuírem, em sua maioria, valores categóricos, os algoritmos escolhidos utilizaram a moda para criar os *clusters*, pela definição de semelhanças e diferenças entre os dados. No entanto, os resultados encontrados por essas abordagens não foram satisfatórios e revelaram a dificuldade dos algoritmos em criar *clusters* distintos, fato que por sua vez rejeitou a hipótese levantada neste trabalho de que o método de *clustering* poderia ser útil para a criação dos grupos de acidentes utilizando a base da CAT. Os motivos para esta conclusão podem estar associados à ausência de uma distribuição adequada dos valores na base de dados, determinada pela própria natureza complexa dos acidentes de trabalho, além da forma de preenchimento do documento.

Apesar das dificuldades observadas, a CAT é a forma de registro oficial e, portanto, mais representativa dos acidentes de trabalho no Brasil. Assim, entende-se que a extração das informações desta base de dados possibilita a compreender mais facilmente os fatores gerais responsáveis pela ocorrência dos acidentes e doenças, algo considerado de grande importância para o gerenciamento de SST no setor da construção.

Assim, os resultados deste trabalho têm o potencial de contribuir com a formulação de planos de ação de caráter preventivo e/ou corretivo, tanto na esfera governamental quanto na empresarial, aplicados em âmbito nacional ou diretamente nos canteiros de obra. Como sugestão para políticas públicas, poderiam ser realizadas ou elaboradas, por exemplo: campanhas nacionais de prevenção de acidentes no mês de outubro, devido à alta incidência de acidentes nesse mês; campanhas nacionais de conscientização sobre o uso adequado de ferramentas manuais, visando a proteção das partes do corpo mais atingidas, como mãos e dedos; e diretrizes para sinalização de segurança

em serviços de terraplenagem, construção de rodovias ou serviços de preparação, consideradas atividades mais letais. Como sugestão para aplicação dos resultados deste trabalho de forma direta nos canteiros de obras, poderiam ser realizados ou adotados, por exemplo: treinamentos semanais de segurança nas empresas nas segundas-feiras, dia da semana mais propenso a ocorrer um acidente; EPIs mais adequados para proteção de partes do corpo como mãos, pés e cabeça; cronogramas de treinamentos periódicos para os serventes de obras; procedimentos de trabalho específicos para a realização de serviços em altura e com eletricidade; entre outros.

Por fim, como sugestão para trabalhos futuros, entende-se que seria interessante:

- Utilizar outros tipos de algoritmo com diferentes abordagens, como o aprendizado profundo, para encontrar outras relações importantes entre os dados e obter resultados mais interessantes;
- Utilizar ferramentas e técnicas no pré-processamento de dados para facilitar a utilização dos algoritmos de ML;
- Cruzar as informações de CAT com outras bases de dados governamentais para identificar o perfil do trabalhador da construção (idades, funções, etc.), valores de benefícios acidentários concedidos, períodos de afastamento do trabalho, entre outras informações relevantes;
- Incorporar registros de acidentes de outros anos para compreender de uma melhor forma a distribuição dos eventos ao longo do tempo;
- Investigar hipóteses levantadas na discussão sobre o motivo de certos dias e meses serem mais propensos a ocorrência de um acidente, bem como determinadas funções e atividades;
- Utilizar e mesclar outros dados, seja em relação às abordagens proativas (buscando levar em consideração as falhas de segurança que ocorrem antes que um evento) ou aos tipos de dados (buscando mesclar imagens, vídeos, áudios) para gerar modelos de predição de acidentes;
- Utilizar tecnologias integradas, como Big Data e IoT, para obter e analisar uma maior quantidade de dados, alcançando resultados mais precisos e detalhados dos acidentes, inclusive em tempo real, antecipando ocorrências futuras.

6 CONCLUSÕES

Os acidentes de trabalho ocorridos na Construção Civil são responsáveis por milhares de mortes todos os anos no mundo. Com o objetivo de diminuir o alto número de ocorrências e com isso contribuir com a manutenção da saúde e segurança dos trabalhadores, pesquisadores têm buscado ferramentas e tecnologias que possam aperfeiçoar a gestão de segurança nos canteiros de obra. Uma das tecnologias disponíveis atualmente é o aprendizado de máquinas, onde sistemas computacionais são estimulados a solucionar problemas gerais de forma autônoma por meio do uso interativo de dados.

Inicialmente, esta pesquisa teve o objetivo de encontrar publicações relevantes sobre o tema por meio de uma RSL, identificando os principais objetivos e características dos modelos de ML desenvolvidos no mundo. A partir das informações obtidas, foi possível mapear as pesquisas, identificando os principais países, autores, instituições e periódicos com maior número de publicações. Os resultados mostraram ainda a ausência do Brasil no desenvolvimento de pesquisas relacionadas ao tema, confirmando a primeira hipótese deste trabalho.

A segunda etapa da pesquisa consistiu na análise de dados de acidentes de trabalho brasileiros com o auxílio de técnicas de ML. O modelo analisou 50.380 dados de CAT, relacionados a acidentes e doenças ocupacionais do setor da construção civil ocorridos entre os anos 2018 e 2021. A partir da análise inicial, foram visualizadas informações gerais sobre os casos, como os tipos de acidentes, atividades econômicas, natureza das lesões, partes do corpo atingidas e o perfil do profissional acidentado, com informações sobre a função, idade e gênero, entre outros.

O entendimento inicial dos acidentes e doenças possibilitou a seleção de atributos específicos para aplicação de dois algoritmos de *clustering*, o *K-Modes* e o *K-Prototype*, utilizados com o objetivo de obter grupos prioritários de acidentes relacionados à atividade de construção de edifícios. No entanto, a utilização da técnica de *clustering* não se mostrou válida para extrair grupos de acidentes da base da CAT, devido à dificuldade dos algoritmos em encontrar *clusters* distintos, rejeitando a segunda hipótese deste trabalho.

Por fim, foram relatadas as dificuldades da pesquisa, sugeridas propostas de aplicações dos resultados e sugeridos trabalhos futuros para otimizar o modelo, contribuindo cada vez mais com a incorporação de tecnologias para o gerenciamento da saúde e segurança do trabalho do setor da construção e, conseqüentemente, com a redução de acidentes e doenças ocupacionais no país.

REFERÊNCIAS

- ABDAT, F. *et al.* Extracting recurrent scenarios from narrative texts using a Bayesian network: Application to serious occupational accidents with movement disturbance. **Accident Analysis and Prevention**, v. 70, p. 155–166, 2014. DOI: 10.1016/j.aap.2014.04.004.
- AKOBENG, A. K. Principles of evidence based medicine. **Archives of Disease in Childhood**, v. 90, n. 8, p. 837–840, 2005. DOI: 10.1136/adc.2005.071761.
- ALLI, B. **Fundamental principles of Occupational Health and Safety**. OIT: Geneva, 2001. ISBN 92-2-110869-4.
- ALSHARIF *et al.* Machine Learning Algorithms for Smart Data Analysis in Internet of Things Environment: Taxonomies and Research Trends. **Symmetry**, v. 12, n. 1, p. 1-17, 2020. DOI: 10.3390/SYM12010088
- ALWASEL, A. *et al.* Identifying poses of safe and productive masons using machine learning. **Automation in Construction**, v. 84, p. 345–355, 2017. DOI: 10.1016/j.autcon.2017.09.022.
- ARCISZEWSKI, T.; MICHALSKI, R.; DYBALA, T. STAR methodology-based learning about construction accidents and their prevention. **Automation in Construction**, v. 4, n. 1, p. 75–85, 1995. DOI: 10.1016/0926-5805(94)00035-L.
- ARYAL, A.; GHAHRAMANI, A.; BECERIK-GERBER, B. Monitoring fatigue in construction workers using physiological measurements. **Automation in Construction**, v. 82, p. 154–165, 2017. DOI: 10.1016/j.autcon.2017.03.003.
- ASSAAD, R.; EL-ADAWAY, I. Determining Critical Combinations of Safety Fatality Causes Using Spectral Clustering and Computational Data Mining Algorithms. **Journal of Construction Engineering and Management**, v. 147, n. 5, 04021035, 2021. DOI: 10.1061/(asce)co.1943-7862.0002040
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS – ABNT. **ISO 45001**: Sistema de Gestão de Saúde e Segurança Ocupacional. ABNT, 2018. ISBN 978-85-07-07514-1.
- AYHAN, B.; TOKDEMIR, O. Accident Analysis for Construction Safety Using Latent Class Clustering and Artificial Neural Networks. **Journal of Construction Engineering and Management**, v. 146, n. 3, 04019114, 2020. DOI: 10.1061/(asce)co.1943-7862.0001762
- AYHAN, B.; TOKDEMIR, O. Predicting the outcome of construction incidents. **Safety Science**, v. 113, n. maio 2018, p. 91–104, 2019. DOI: 10.1016/j.ssci.2018.11.001.
- BADRI, A. *et al.* Occupational health and safety in the industry 4.0 era: A cause for major concern? **Safety Science**, v. 109, n. 2018, p. 403-411, 2018. DOI: 10.1016/j.ssci.2018.06.012.

BAKER, H.; HALLOWELL, M.; TIXIER, A. Automatically learning construction injury precursors from text. **Automation in Construction**, v. 118, n. agosto 2019, p. 103145, 2020. DOI: 10.1016/j.autcon.2020.103145.

BONIFAZI, G. *et al.* A machine learning based sentient multimedia framework to increase safety at work. **Multimedia Tools and Applications**, v. 81, n. 1, p. 141–169, 2022. DOI: 10.1007/s11042-021-10984-z

BRASIL. **Lei nº 8.213, de 24 de julho de 1991 da Constituição Federal**. Estabelece Benefícios da Previdência Social. Brasília, 1991.

BRASIL. Ministério da Saúde. **Notificação de acidentes do trabalho fatais, graves e com crianças e adolescentes**. Brasília: Ministério da Saúde, 2006.

BRASIL. Ministério do Trabalho e Emprego. **Classificação Brasileira de Ocupações – CBO**. 2017. Disponível em: <http://www.mtecho.gov.br/cbsite/pages/home.jsf>. Acesso em: 18 de abril de 2022.

BRASIL. Ministério do Trabalho e Previdência Social. **Norma Regulamentadora 01 – Disposições Gerais e Gerenciamento de Riscos Ocupacionais**. Disponível em: <https://www.gov.br/trabalho-e-previdencia/pt-br/composicao/orgaos-especificos/secretaria-de-trabalho/inspecao/seguranca-e-saude-no-trabalho/normas-regulamentadoras/nr-01-atualizada-2020.pdf>. Acesso em: 22 de agosto de 2022.

BRASIL. Ministério do Trabalho e Previdência Social. **Normas Regulamentadoras – NR**. 2022. Disponível em: <https://www.gov.br/trabalho-e-previdencia/pt-br/composicao/orgaos-especificos/secretaria-de-trabalho/inspecao/seguranca-e-saude-no-trabalho/ctpp-nrs/normas-regulamentadoras-nrs>. Acesso em: 22 de agosto de 2022.

BRASIL. **Portaria n.º 25, de 29 de dezembro de 1994**. Brasília, 1994.

BRASIL. Receita Federal. **Classificação Nacional de Atividades Econômicas – CNAE**. 2020. Disponível em: <https://www.gov.br/receitafederal/pt-br/assuntos/orientacao-tributaria/cadastros/cnpj/classificacao-nacional-de-atividades-economicas-2013-cnae/apresentacao>. Acesso em: 18 de abril de 2022.

BRILAKIS, I.; PARK, M.; JOG, G. Automated vision tracking of project related entities. **Advanced Engineering Informatics**, v. 25, n. 4, p. 713–724, 2011. DOI: 10.1016/j.aei.2011.01.003.

CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO. Banco de dados CBIC. **RAIS - Ministério do Trabalho**. 2023. Disponível em: <http://www.cbicdados.com.br/menu/emprego/rais-ministerio-do-trabalho>. Acesso em: 20 de janeiro de 2023.

CAO, H.; GOH, Y. Analyzing construction safety through time series methods. **Frontiers of Engineering Management**, v. 6, n. 2, p. 262–274, 2019. DOI: 10.1007/s42524-019-0015-6

CHEN, S.; DEMACHI, K. Towards on-site hazards identification of improper use of personal protective equipment using deep learning-based geometric relationships and hierarchical scene graph. **Automation in Construction**, v. 125, n. março 2020, p. 103619, 2021. DOI: 10.1016/j.autcon.2021.103619.

CHEN, H.; LUO, X. Severity Prediction Models of Falling Risk for Workers at Height. **Procedia Engineering**, v. 164, n. junho, p. 439–445, 2016. DOI: 10.1016/j.proeng.2016.11.642.

CHI, N.; LIN, K.; HSIEH, S. Using ontology-based text classification to assist Job Hazard Analysis. **Advanced Engineering Informatics**, v. 28, n. 4, p. 381–394, 2014. DOI: 10.1016/j.aei.2014.05.001.

CHOI, J. *et al.* Machine learning predictive model based on national data for fatal accidents of construction workers. **Automation in Construction**, v. 110, n. setembro 2019, 102974, 2020. DOI: 10.1016/j.autcon.2019.102974.

CHOKOR, A. *et al.* Analyzing Arizona OSHA Injury Reports Using Unsupervised Machine Learning. **Procedia Engineering**, v. 145, p. 1588–1593, 2016. DOI: 10.1016/j.proeng.2016.04.200.

CIARAPICA, F. E.; GIACCHETTA, G. Classification and prediction of occupational injury risk using soft computing techniques: An Italian study. **Safety Science**, v. 47, n. 1, p. 36–49, 2009. DOI: 10.1016/j.ssci.2008.01.006.

DEBNATH, J. *et al.* Fuzzy inference model for assessing occupational risks in construction sites. **International Journal of Industrial Ergonomics**, v. 55, p. 114–128, 2016. DOI: 10.1016/j.ergon.2016.08.004.

DING, L. *et al.* A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. **Automation in Construction**, v. 86, n. março 2017, p. 118–124, 2018. DOI: 10.1016/j.autcon.2017.11.002.

DUARTE, M. **Análise do acidentes de trabalho causados por meio de transporte motocicleta em uma capital brasileira**. Dissertação (Mestrado em Enfermagem) – Universidade Federal de Minas Gerais, Minas Gerais, 2011.

DUTTA, A. *et al.* Fusing imperfect experimental data for risk assessment of musculoskeletal disorders in construction using canonical polyadic decomposition. **Automation in Construction**, v. 119, n. February, 2020. DOI: 10.1016/j.autcon.2020.103322

FAN, C. Defect Risk Assessment Using a Hybrid Machine Learning Method. **Journal of Construction Engineering and Management**, v. 146, n. 9, 04020102, 2020. DOI: 10.1061/(asce)co.1943-7862.0001897.

FANG, Q. *et al.* A semantic and prior-knowledge-aided monocular localization method for construction-related entities. **Computer-Aided Civil and Infrastructure Engineering**, v. 35, n. 9, p. 979–996, 2020. DOI: 10.1111/mice.12541

- FANG, W. *et al.* Automated text classification of near-misses from safety reports: An improved deep learning approach. **Advanced Engineering Informatics**, v. 44, n. fevereiro, 101060, 2020. DOI: 10.1016/j.aei.2020.101060.
- FANG, Q. *et al.* Computer vision aided inspection on falling prevention measures for steeplejacks in an aerial environment. **Automation in Construction**, v. 93, n. dezembro 2017, p. 148–164, 2018. DOI: 10.1016/j.autcon.2018.05.022.
- FENG, Q.; SHI, X.; ZHANG, J. Influence of rent-seeking on safety supervision in Chinese construction: Based on a simulation technology. **Technological Forecasting and Social Change**, v. 138, n. junho 2018, p. 1–9, 2019. DOI: 10.1016/j.techfore.2018.10.016
- FERRO, A. **Estatísticas de Acidentes de Trabalho na Indústria da Construção**. Câmara Brasileira da Indústria da Construção. Brasília: CBIC, 2019.
- GIACOMELLO, D. **Técnicas de Agrupamento aplicadas aos dados de Acidente de Trabalho**. 2020. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Uberlândia, Minas Gerais, 2020.
- GOH, Y.; UBEYNARAYANA, C. Construction accident narrative classification: An evaluation of text mining techniques. **Accident Analysis and Prevention**, v. 108, n. agosto, p. 122–130, 2017. DOI: 10.1016/j.aap.2017.08.026.
- HALE, A.; HOVDEN, J. **Management and culture: the third age of safety**. Occupational Injury: Risk Prevention and Intervention. Taylor & Francis: Londres, 1998.
- HAN, S.; LEE, S.; PEÑA-MORA, F. Comparative Study of Motion Features for Similarity-Based Modeling and Classification of Unsafe Actions in Construction. **Journal of Computing in Civil Engineering**, v. 28, n. 5, p. 1–11, 2014. DOI: 10.1061/(asce)cp.1943-5487.0000339
- HUANG, L. *et al.* Detection algorithm of safety helmet wearing based on deep learning. **Concurrency and Computation: Practice and Experience**, v. 33, n. 13, p. 1–14, 2021. DOI: 10.1002/cpe.6234
- HUANG, Z. Clustering large data sets with mixed numeric and categorical values. In: PROCEEDINGS OF THE FIRST PACIFIC ASIA KNOWLEDGE DISCOVERY AND DATA MINING CONFERENCE, 1997, Singapore. **Anais [...]** Singapore: 1997, pp. 21–34.
- HUANG, Z. **Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values**. Data Mining and Knowledge Discovery, v. 2, p. 283-304, 1998.
- HURWITZ, J.; KIRSCH, D. **Approaches to machine learning**. Hoboken: John Wiley & Sons Inc., 2018, v. 35. ISSN: 10974571.
- INSTITUTO NACIONAL DO SEGURO SOCIAL – INSS. **Comunicação de Acidente de Trabalho – CAT**. 2022. Disponível em: <https://www.gov.br/inss/pt-br/saiba-mais/auxilios/comunicacao-de-acidente-de-trabalho-cat>. Acesso em: 18 de abril de 2022.

JAHANGIRI, M.; SOLUKLOEI, H.; KAMALINIA, M. A neuro-fuzzy risk prediction methodology for falling from scaffold. **Safety Science**, v. 117, n. julho 2018, p. 88–99, 2019. DOI: 10.1016/j.ssci.2019.04.009.

KENNY, C. **Construction, Corruption, and Developing Countries**. The World Bank: Washington, DC, USA, 2007.

KHAN, N. *et al.* Utilizing safety rule correlation for mobile scaffolds monitoring leveraging deep convolution neural networks. **Computers in Industry**, v. 129, 103448, 2021. DOI: 10.1016/j.compind.2021.103448.

KIM, T.; CHI, S. Accident Case Retrieval and Analyses: Using Natural Language Processing in the Construction Industry. **Journal of Construction Engineering and Management**, v. 145, n. 3, 04019004, 2019. DOI: 10.1061/(asce)co.1943-7862.0001625

KINES, S. Prioritizing occupational injury prevention in the construction industry: Injury severity or absence? **Journal of Safety Research**, v. 38, n. 1, 2007, p. 53-58. DOI: 10.1016/j.jsr.2006.09.002.

KOLAR, Z.; CHEN, H.; LUO, X. Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. **Automation in Construction**, v. 89, n. maio 2017, p. 58–70, 2018. DOI: 10.1016/j.autcon.2018.01.003.

LEE, J. *et al.* A study on data pre-processing and accident prediction modelling for occupational accident analysis in the construction industry. **Applied Sciences (Switzerland)**, v. 10, n. 21, p. 1–23, 2020. DOI: 10.3390/app10217949

LEE, Y. *et al.* Evidence-driven sound detection for prenotification and identification of construction safety hazards and accidents. **Automation in Construction**, v. 113, n. fevereiro, p. 103127, 2020. DOI: 10.1016/j.autcon.2020.103127.

LI, H.; LUO, X.; SKITMORE, M. Intelligent Hoisting with Car-Like Mobile Robots. **Journal of Construction Engineering and Management**, v. 146, n. 12, 04020136, 2020. DOI: 10.1061/(asce)co.1943-7862.0001931

LI, X. *et al.* MetaInjury: Meta-learning framework for reusing the risk knowledge of different construction accidents. **Safety Science**, v. 140, n. outubro 2020, p. 105315, 2021. DOI: 10.1016/j.ssci.2021.105315.

LIAO, C.; PERNG, Y.; CHIANG, T. Discovery of unapparent association rules based on extracted probability. **Decision Support Systems**, v. 47, n. 4, p. 354–363, 2009. DOI: 10.1016/j.dss.2009.04.006.

LIM, T. *et al.* Artificial Neural Network–Based Slip-Trip Classifier Using Smart Sensor for Construction Workplace. **Journal of Construction Engineering and Management**, v. 142, n. 2, 04015065, 2016. DOI: 10.1061/(asce)co.1943-7862.0001049

LIMA JÚNIOR, J. **Gestão de Segurança e Saúde no Trabalho na Indústria da Construção**. Câmara Brasileira da Indústria da Construção. Brasília: CBIC, 2019.

LINGARD, H. Occupational health and safety in the construction industry. **Construction Management and Economics**, v. 31, n. 6, p. 505-514, 2013.
DOI: 10.1080/01446193.2013.816435.

LIU, S. *et al.* Multi-scale personnel deep feature detection algorithm based on Extended-YOLOv3. **Journal of Intelligent and Fuzzy Systems**, v. 40, n. 1, p. 773–786, 2021.
DOI: 10.3233/JIFS-200778

MARISCAL, M. *et al.* The influence of employee training and information on the probability of accident rates. **International Journal of Industrial Ergonomics**, v. 72, n. junho, p. 311–319, 2019. DOI: 10.1016/j.ergon.2019.06.002.

MARTÍN, J. E. *et al.* A Bayesian network analysis of workplace accidents caused by falls from a height. **Safety Science**, v. 47, n. 2, p. 206–214, 2009. DOI: 10.1016/j.ssci.2008.03.004.

MATÍAS, J. *et al.* A machine learning methodology for the analysis of workplace accidents. **International Journal of Computer Mathematics**, v. 85, n. 3–4, p. 559–578, 2008.
DOI: 10.1080/00207160701297346

MISTIKOGLU, G. *et al.* Decision tree analysis of construction fall accidents involving roofers. **Expert Systems with Applications**, v. 42, n. 4, p. 2256–2263, 2015.
DOI: 10.1016/j.eswa.2014.10.009.

MOHAMMADFAM, I. *et al.* Constructing a Bayesian network model for improving safety behavior of employees at workplaces. **Applied Ergonomics**, v. 58, p. 35–47, 2017.
DOI: 10.1016/j.apergo.2016.05.006.

MUTLU, N.; ALTUNTAS, S. Assessment of occupational risks In Turkish manufacturing systems with data-driven models. **Journal of Manufacturing Systems**, v. 53, n. novembro 2018, p. 169–182, 2019. DOI: 10.1016/j.jmsy.2019.09.008.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS - ONU. **The Sustainable Development Agenda**. UN, 2015. Disponível em: <https://www.un.org/sustainabledevelopment/development-agenda/>
Acesso em: 15 de março de 2022.

ORGANIZAÇÃO INTERNACIONAL DO TRABALHO - OIT. **Join in building a culture of prevention on OSH**. OIT, 2015. Disponível em: https://www.ilo.org/beijing/information-resources/public-information/speeches/WCMS_376164/lang--en/index.htm.
Acesso em: 31 de março de 2022.

ORGANIZAÇÃO INTERNACIONAL DO TRABALHO. **Safety and Health in the Construction Sector – Overcoming the Challenges**. 2014. Disponível em: https://www.ilo.org/empent/Eventsandmeetings/WCMS_310993/lang--en/index.htm.
Acesso em: 22 de agosto de 2022.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Healthy Workplace Framework and Model: Background and Supporting Literature and Practices**. OMS: Geneva, 2010. ISBN 978-92-4-150024-1.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. **International Statistical Classification of Diseases and Related Health Problems (ICD)**. 2022.
Disponível em: <https://www.who.int/classifications/classification-of-diseases>.
Acesso em: 18 de abril de 2022.

ORGANIZAÇÃO MUNDIAL DA SAÚDE; ORGANIZAÇÃO INTERNACIONAL DO TRABALHO. **WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury, 2000–2016: Global Monitoring Report**. Geneva: OMS/OIT, 2021. ISBN: 978-92-4-003494-5.

PAGE, M. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews Systematic reviews and Meta-Analyses. **BMJ**, v. 372, n. 71, 2021.
DOI: 10.1136/bmj.n71

PARK, J. *et al.* Online Learning Control of Hydraulic Excavators Based on Echo-State Networks. **IEEE Transactions on Automation Science and Engineering**, v. 14, n. 1, p. 249–259, 2017.
DOI: 10.1109/TASE.2016.2582213

POH, C.; UBEYNARAYANA, C.; GOH, Y.. Safety leading indicators for construction sites: A machine learning approach. **Automation in Construction**, v. 93, n. outubro 2017, p. 375–386, 2018. DOI: 10.1016/j.autcon.2018.03.022.

POUYANFAR *et al.* A Survey on Deep Learning: Algorithms, Techniques, and Applications. **ACM Computing Surveys**, v. 51, n. 5, 2018. DOI: 10.1145/3234150.

RIVAS, T. *et al.* Explaining and predicting workplace accidents using data-mining techniques. **Reliability Engineering and System Safety**, v. 96, n. 7, p. 739–747, 2011.
DOI: 10.1016/j.res.2011.03.006

ROBINSON, K.; AKINS, R. Machine learning in epigenetic diseases. **Medical Epigenetics**, v. 29, p. 513-525, 2021. DOI: 10.1016/B978-0-12-823928-5.00038-4.

RYU, J. *et al.* Analysis of Relationships between Body Load and Training, Work Methods, and Work Rate: Overcoming the Novice Mason’s Risk Hump. **Journal of Construction Engineering and Management**, v. 146, n. 8, 04020097, 2020.
DOI: 10.1061/(asce)co.1943-7862.0001889

SAKHAKARMI, S.; PARK, J.; CHO, C. Enhanced Machine Learning Classification Accuracy for Scaffolding Safety Using Increased Features. **Journal of Construction Engineering and Management**, v. 145, n. 2, 04018133, 2019. DOI: 10.1061/(asce)co.1943-7862.0001601

SALAMA, D.; EL-GOHARY, N. Semantic Text Classification for Supporting Automated Compliance Checking in Construction. **Journal of Computing in Civil Engineering**, v. 30, n. 1, 04014106, 2013. DOI: 10.1061/(asce)cp.1943-5487.0000301

SARKAR, S. *et al.* Application of optimized machine learning techniques for prediction of occupational accidents. **Computers and Operations Research**, v. 106, p. 210–224, 2019. DOI: 10.1016/j.cor.2018.02.021.

SARKAR, S.; MAITI, J. Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis. **Safety Science**, v. 131, n. junho, p. 104900, 2020. DOI: 10.1016/j.ssci.2020.104900.

SARKER, I. Machine Learning: Algorithms, Real-World Applications and Research Directions. **SN Computer Science**, v. 2, n. 160, 2021. DOI: 10.1007/s42979-021-00592-x.

SERVIÇO NACIONAL DE APRENDIZAGEM INDUSTRIAL – SENAI; CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO – CBIC. **Informativo Econômico**. 2021. Disponível em: <https://cbic.org.br/wp-content/uploads/2021/07/informativo-economico-importancia-construcao-civil-final-julho-2021.pdf>. Acesso em: 15 de março de 2022.

SHEN, J. *et al.* Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning. **Computer-Aided Civil and Infrastructure Engineering**, v. 36, n. 2, p. 180–196, 2021. DOI: 10.1111/mice.12579

SIDDULA, M. *et al.* Unsupervised Feature Learning for Objects of Interest Detection in Cluttered Construction Roof Site Images. **Procedia Engineering**, v. 145, p. 428–435, 2016. DOI: 10.1016/j.proeng.2016.04.010.

SILVA, D. *et al.* Condições de trabalho e riscos no trânsito urbano na ótica de trabalhadores motociclistas. **Physis**, v. 18, n. 2, 2008.

SINGH, A.; MISRA, S. A Dominance based Rough Set analysis for investigating employee perception of safety at workplace and safety compliance. **Safety Science**, v. 127, n. fevereiro, 104702, 2020. DOI: 10.1016/j.ssci.2020.104702.

SMARTLAB. **Observatório de Segurança e Saúde do Trabalho**. 2022. Disponível em: <https://smartlabbr.org/sst>. Acesso em: 15 de março de 2022.

SOMAVIA, J. **Facts on Safety at Work**. OIT, 2005.

SON, H.; KIM, C. Integrated worker detection and tracking for the safe operation of construction machinery. **Automation in Construction**, v. 126, n. março, 103670, 2021. DOI: 10.1016/j.autcon.2021.103670.

TANG, S. *et al.* Video-Based Motion Trajectory Forecasting Method for Proactive Construction Safety Monitoring Systems. **Journal of Computing in Civil Engineering**, v. 34, n. 6, 04020041, 2020. DOI: 10.1061/(asce)cp.1943-5487.0000923

- TANG, S.; ROBERTS, D.; GOLPARVAR-FARD, M. Human-object interaction recognition for automatic construction site safety inspection. **Automation in Construction**, v. 120, n. outubro 2019, 103356, 2020. DOI: 10.1016/j.autcon.2020.103356.
- TEIXEIRA, J.; REIS, A.; MENDES, F. **Uso de tecnologias para gerenciamento de áreas de risco na construção**. Câmara Brasileira da Indústria da Construção. Brasília: CBIC, 2019.
- TEIZER, J. 3D range imaging camera sensing for active safety in construction. **Electronic Journal of Information Technology in Construction**, v. 13, n. junho 2007, p. 103–117, 2008.
- TIXIER, A. *et al.* Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports. **Automation in Construction**, v. 62, n. 2016, p. 45–56, 2016. DOI: 10.1016/j.autcon.2015.11.001.
- TIXIER, A. *et al.* Application of machine learning to construction injury prediction. **Automation in Construction**, v. 69, n. 2016, p. 102–114, 2016. DOI: 10.1016/j.autcon.2016.05.016.
- TIXIER, A. *et al.* Construction Safety Clash Detection: Identifying Safety Incompatibilities among Fundamental Attributes using Data Mining. **Automation in Construction**, v. 74, p. 39–54, 2017. DOI: 10.1016/j.autcon.2016.11.001.
- TOLEDO, J.; TIMOTEO, R.; BARBOSA, E. **Inteligência Artificial para predição de acidentes laborais no Brasil e sua aplicação pela Inspeção do Trabalho**. Revista da Escola Nacional da Inspeção do Trabalho. Rev. ENIT: Brasília, 2020.
- WACHTER, J.; YORIO, P. A system of safety management practices and worker engagement for reducing and preventing accidents: An empirical and theoretical investigation. **Accident Analysis and Prevention**, v. 68, p. 117-130, 2014. DOI: 10.1016/j.aap.2013.07.029
- WÄNNSTRÖM, T.; LARSSON, T. **Using information properly makes for effective prevention**. Swedish Council for Working Life: Estocolmo, 1996.
- WU, H.; ZHAO, J. An intelligent vision-based approach for helmet identification for work safety. **Computers in Industry**, v. 100, n. agosto 2017, p. 267–277, 2018. DOI: 10.1016/j.compind.2018.03.037.
- XIAO, B.; KANG, S. Development of an Image Data Set of Construction Machines for Deep Learning Object Detection. **Journal of Computing in Civil Engineering**, v. 35, n. 2, 05020005, 2021. DOI: 10.1061/(asce)cp.1943-5487.0000945.
- YANG, K. *et al.* Semi-supervised near-miss fall detection for ironworkers with a wearable inertial measurement unit. **Automation in Construction**, v. 68, p. 194–202, 2016. DOI: 10.1016/j.autcon.2016.04.007.

YU, Y. *et al.* Joint-Level Vision-Based Ergonomic Assessment Tool for Construction Workers. **Journal of Construction Engineering and Management**, v. 145, n. 5, 04019025, 2019. DOI: 10.1061/(asce)co.1943-7862.0001647.

ZHANG, F. *et al.* Construction site accident analysis using text mining and natural language processing techniques. **Automation in Construction**, v. 99, n. dezembro 2018, p. 238–248, 2019. DOI: 10.1016/j.autcon.2018.12.016.

ZHANG, J. *et al.* A C-BiLSTM approach to classify construction accident reports. **Applied Sciences (Switzerland)**, v. 10, n. 17, 2020. DOI: 10.3390/APP10175754.

ZHANG, M. *et al.* Utilizing Computer Vision and Fuzzy Inference to Evaluate Level of Collision Safety for Workers and Equipment in a Dynamic Environment. **Journal of Construction Engineering and Management**, v. 146, n. 6, 04020051, 2020. DOI: 10.1061/(asce)co.1943-7862.0001802.

ZHANG, M.; CAO, T.; ZHAO, X. Using Smartphones to Detect and Identify Construction Workers' Near-Miss Falls Based on ANN. **Journal of Construction Engineering and Management**, v. 145, n. 1, 04018120, 2019. DOI: 10.1061/(asce)co.1943-7862.0001582.

ZHAO, J.; OBONYO, E. Convolutional long short-term memory model for recognizing construction workers' postures from wearable inertial measurement units. **Advanced Engineering Informatics**, v. 46, n. agosto, 101177, 2020. DOI: 10.1016/j.aei.2020.101177.

ZHONG, B. *et al.* Hazard analysis: A deep learning and text mining framework for accident prevention. **Advanced Engineering Informatics**, v. 46, n. abril, p. 101152, 2020. DOI: 10.1016/j.aei.2020.101152.

ZHU, R. *et al.* Application of machine learning techniques for predicting the consequences of construction accidents in China. **Process Safety and Environmental Protection**, v. 145, p. 293–302, 2021. DOI: 10.1016/j.psep.2020.08.006.

ANEXO A – Modelo de Comunicação de Acidente de Trabalho



PREVIDÊNCIA SOCIAL

2.Tipo de CAT

1- Emitente 1 - Empregador	2 - Sindicato	3 -Médico
4 - Segurado ou dependente	5 - Autoridade Pública	

2 - Tipo de CAT 1 - Início	2 - Reabertura	/	/
3 - Comunicação de óbito em		/	/

COMUNICAÇÃO DE ACIDENTE DO TRABALHO - CAT
(LER ATENTAMENTE AS ORIENTAÇÕES ANTES DO PREENCHIMENTO)

Empregado	3 - Razão Social/Nome			4 - Tipo			1-CGC 2-CEI 3-CPF 4-NIT		5 - CNAE	
	6 - Endereço		Rua/Av/Nº/comp.	Bairro	CEP	7-Município	8 - UF	9 - Telefone		

Acidentado	10 - Nome				11 - Nome da mãe									
	12-Data de Nasc.		13 - Sexo		14 - Estado civil		15-CTPS/SerieData de emissao				16-UF			
	17 - Carteira de identidade		Data da emissão		Orgão Exp.		18-UF		19-PIS/PASEP		20-Remuneração Mensal			
	21-Endereço		Rua/Av/Nº/comp.		Bairro		CEP		22-Município		23 - UF		24 - Telefone	
	25-Nome da ocupação		26 - CBO		27-Filiação à previdência Social				28-Aposentado?		29-Area			

Acidente ou Doença	30-Data do acidente		31-Hora do acidente		32- Após quantas horas de Trabalho?				Houve afastamento?		34-Ultimo dia de Trab.	
	35-Local do Acidente		36-CGC		37-Município do local do Acidente				38-UF		39-Especif. do local do acidente	
	40-Parte(s) do corpo atingida(s)				41-Agente causador							
	42-Descrição da situação do acidente ou doença								43-Houve registro policial?		1-Sim 2-Não	
									44 -Houve morte?		1-Sim 2-Não	

Testemunha	47 -Nome: Não houve testemunhas do ocorrido.												
	46 - Endereço		Rua/Av/Nº/comp.		Bairro		CEP		47-Município		48 - UF		Telefone
	49-Nome												
	50- Endereço		Rua/Av/Nº/comp.		Bairro		CEP		51-Município		52 - UF		Telefone

Atendimento	53 -Unidade de atendimento médico				54 - Data				55 - Hora			
	56-Houve internação?		57 - Duração provavel do tratamento				58 - Deverá o acidentado afastar-se do trabalho durante o tratamento ?					
1-SIM 2- NÃO		15 dias				1- SIM 2- NÃO						

Diagnóstico com Lesão	59 - Descrição e natureza da lesão											
	60 - Diagnóstico provável								61 - CID - 10			
	62 - Observações											

Local e data				Assinatura do emitente							
63 - Recebida Em / /		64 - Código de unidade		65 - Número de acidente				Notas:			
66 - E reconhecido o direito do segurado à habilitação de benefício acidentário?				67 - Tipo		1 - A inexistência das declarações desta comunicação implicará nas sanções previstas nos Art. 171 e 299 do código penal. 2 - A comunicação de acidente do trabalho deverá ser feita até o 1º dia útil após o acidente, sob pena de multa. 3 - A comunicação do acidente do trabalho reger-se pelos Art. 134 do Dec. 2.172/97. 4 - Os conceitos de acientes do trabalho e doença ocupacional estão definidos nos arts. 131 a 133 do Dec 2172/ 97 5 - A caracterização do acidente reger-se-á pelo art. 135 do Dec 2172/ 97.					
1 - Sim 2 - Não		3 - Tajueto									
68 - Matrícula do servidor				Matrícula		Assinatura do servidor					

A COMUNICAÇÃO DE ACIDENTE É OBRIGATORIA, MESMO NO CASO EM QUE NÃO HAJA AFASTAMENTO DO TRABALHO.

APÊNDICE A – Modelo de ML

Link para acesso:

https://github.com/mariliabezerrac/Dados_CAT.git