UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Faculdade de Engenharia e Ciências de Guaratinguetá

TAYNARA DE OLIVEIRA CASTELLÕES

Parto prematuro e poluentes do ar em São José dos Campos-SP:

uma abordagem *fuzzy* e neuro-*fuzzy*

Guaratinguetá
2023

Taynara de Oliveira Castelllões

Parto prematuro e poluentes do ar em São José dos Campos-SP:

uma abordagem *fuzzy* e neuro-*fuzzy*

Dissertação apresentada à Faculdade de Engenharia e Ciências do Campus de Guaratinguetá, Universidade Estadual Paulista, para obtenção do título de Mestre em Engenharia em Simulação, otimização e tomada de decisão

Orientadora: Prof. Dr. Paloma Maria Silva Rocha Rizol
Coorientador: Prof. Dr. Luiz Fernando Costa Nascimento

Guaratinguetá
2023

Castellões, Taynara de Oliveira
C348p Parto prematuro e poluentes do ar em São José dos Campos-SP: uma abordagem *fuzzy* e *neuro-fuzzy* / Taynara de Oliveira Castellões - Guaratinguetá, 2023.
84 f : il.
Bibliografia: f. 78-84

Dissertação (Mestrado) – Universidade Estadual Paulista, Faculdade de Engenharia e Ciências de Guaratinguetá, 2023.

Orientadora: Profª. Drª. Paloma M. Silva Rocha Rizol
Coorientador: Prof. Dr. Luiz F. Costa Nascimento

1. Prematuros. 2. Ar - Poluição.
3. Mortalidade infantil. 4. Lógica difusa. I. Título.

CDU 614 (043)

Luciana Máximo
Bibliotecária/CRB-8 3595

IMPACTO POTENCIAL DESTA PESQUISA

Essa pesquisa possui impacto significativo para toda a sociedade, pois aborda questões importantes relacionadas ao parto prematuro. A divulgação desse trabalho pode aumentar a conscientização da população em relação a melhoria de cuidados pré-natal, como a realização de pelo menos seis consultas durante a gravidez. Os resultados obtidos podem auxiliar os profissionais de saúde na tomada de decisões e influenciar a criação de políticas públicas de redução de gases poluentes. Além disso, a minimização da incidência do parto prematuro pode levar à redução de custos associados ao tratamento de complicações geradas pela prematuridade, e por fim, à melhoria na saúde de gerações futuras.

POTENTIAL IMPACT OF THIS RESEARCH

This research has a significant impact on society as a whole, as it addresses important issues related to preterm birth. The dissemination of this work can increase the population's awareness regarding the improvement of prenatal care, such as the performance of at least six consultations during pregnancy. The results obtained can assist health professionals in decision-making and influence the creation of public policies to reduce polluting gases. In addition, minimizing the incidence of preterm birth can lead to a reduction in costs associated with the treatment of complications generated by prematurity, and ultimately, to an improvement in the health of future generations.

unesp

UNIVERSIDADE ESTADUAL PAULISTA
CAMPUS DE GUARATINGUETÁ

Taynara de Oliveira Castellões

ESTA DISSERTAÇÃO FOI JULGADA ADEQUADA PARA A OBTENÇÃO DO TÍTULO DE
"MESTRE EM ENGENHARIA"

PROGRAMA: ENGENHARIA
CURSO: MESTRADO

APROVADA EM SUA FORMA FINAL PELO PROGRAMA DE PÓS-GRADUAÇÃO

gov.br Documento assinado digitalmente
MANOEL CLEBER DE SAMPAIO ALVES
Data: 04/12/2023 10:05:05-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Manoel Cléber de Sampaio Alves
Coordenador

B A N C A E X A M I N A D O R A:

gov.br Documento assinado digitalmente
PALOMA MARIA SILVA ROCHA RIZOL
Data: 08/11/2023 11:09:36-0300
Verifique em https://validar.iti.gov.br

Profa Dra Paloma Maria Silva Rocha Rizol
Orientadora / UNESP/FEG
participou por videoconferência

gov.br Documento assinado digitalmente
MARCELA APARECIDA GUERREIRO MACHADO DE
Data: 17/11/2023 10:45:21-0300
Verifique em https://validar.iti.gov.br

Profa Dra. Marcela Aparecida Guerreiro Machado de Freitas
UNESP/FEG
participou por videoconferência

gov.br Documento assinado digitalmente
SANDRA REGINA MONTEIRO MASALSKIENE ROVE
Data: 13/11/2023 13:04:16-0300
Verifique em https://validar.iti.gov.br

Profa Dra Sandra Regina Monteiro Masalskiene Roveda
Unesp ICT Sorocaba
participou por videoconferência

Outubro de 2023

DADOS CURRICULARES

TAYNARA DE OLIVEIRA CASTELLÕES

NASCIMENTO 15/05/1994 – Resende / RJ

FILIAÇÃO Calebe Gonçalves da Silva
Verônica Aparecida de Oliveira

2013/2018 Curso de Graduação em Engenharia de Produção - ênfase em mecânica na Universidade do Estado do Rio de Janeiro - UERJ

2018/2020 Curso de Pós-Graduação em Gerenciamento de Projetos na Pontifícia Universidade Católica de Minas Gerais – PUC/MG

2021/2023 Curso de Pós- Graduação, nível Mestrado em Engenharia com área de concentração em Pesquisa Operacional (Simulação, Otimização e Tomada de Decisão), na Universidade Estadual Paulista - UNESP

Dedico esse trabalho a todas as pessoas que, de alguma, forma fizeram esse sonho ser possível.

AGRADECIMENTOS

Agradeço à Prof. Dra. Paloma Maria Silva Rocha Rizol, por ter me convidado e auxiliado durante toda a etapa do processo de construção dessa tese;

ao Prof. Dr. Luiz Fernando Costa Nascimento, por todos os ensinamentos e contribuições passadas a mim;

a todos os funcionários da Pós-Graduação da FEG/UNESP pelas informações prestadas, apoio e contribuições;

a todos os colegas da FEG/UNESP que fiz durante meu período de estudos;

a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- Brasil) por me apoiarem durante o levantamento desse estudo;

em suma, a Pós Graduação da FEG/UNESP por me permitirem aderir ao programa;

e a todos os amigos e familiares que contribuíram de forma direta ou não para a realização desse trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- Brasil (CAPES) - código de financiamento 001.

RESUMO

Os efeitos do parto prematuro são diversos e não se limitam apenas a infância, podendo gerar dificuldades de aprendizagem e complicações até a fase adulta. Estudos relacionam a prematuridade a inúmeros fatores como falta de acompanhamento pré-natal, idade e escolaridade materna, alcoolismo, tabagismo até a exposição a poluentes do ar. Esse trabalho tem por objetivo investigar a associação entre parto prematuro, idade da mãe, número de consultas pré-natal e exposição da gestante ao Material particulado com diâmetro menor que 10 μm (PM_{10}) e Dióxido de Nitrogênio (NO_2). Foram desenvolvidos modelos de predição utilizando técnicas de *inteligência artificial*: *fuzzy* e neuro-*fuzzy*. Os modelos foram validados por meio das métricas de avaliação como correlação (r), coeficiente de determinação (R^2), erro médio absoluto (MAE), raiz do erro quadrático médio (RMSE), erro percentual absoluto médio (MAPE) e erro médio quadrático (MSE) entre os dados reais e preditos. Para aquisição dos dados, foram utilizadas informações da cidade de São José dos Campos-SP, entre os anos de 2016 a 2018. Os relatórios diários de poluentes utilizados foram extraídos do portal da CETESB (Companhia Ambiental do Estado de São Paulo), e os dados da Declaração de Nascidos Vivos (DNV), extraídos do portal de livre acesso do DATASUS (Departamento de Informática do Sistema Único de Saúde). Dentre todos os modelos de predição desenvolvidos, apresentou melhor desempenho o modelo neuro-*fuzzy* com validação simples 50:50, estrutura do tipo *grid partition,* método de otimização híbrido, com 20 épocas, entradas gaussianas, sendo as entradas idade da mãe, número de consultas pré-natal e concentração de PM_{10} com quatro funções de pertinência e a concentração de NO_2 com três funções, e variável de saída constante, com métricas r=0,59, R^2=0,35, MAE=1,70, RMSE=2,83, MAPE=5,35% e MSE=8,00. A abordagem apresentada desse estudo possui potencial para beneficiar a sociedade como um todo, auxiliando gestores na implementação de medidas preventivas voltadas à redução da emissão de gases poluentes. Visando não apenas melhoria na saúde da gestante e recém-nascido, mas também promoção de qualidade de vida para gerações futuras.

PALAVRAS-CHAVE: Prematuridade; Mortalidade Infantil; Poluentes do Ar; Lógica *Fuzzy*; Saúde Pública.

ABSTRACT

The effects of premature birth are diverse and are not limited to childhood, but can cause learning difficulties and complications until adulthood. Studies link prematurity to numerous factors such as lack of prenatal care, maternal age and education, alcoholism, smoking and exposure to air pollutants. This work aims to investigate the association between premature birth, mother's age, number of prenatal consultations and pregnant woman's exposure to Particulate Matter with a diameter smaller than 10 μm (PM_{10}) and Nitrogen Dioxide (NO_2). Prediction models were developed using artificial intelligence techniques: fuzzy and neuro-fuzzy. The models were validated using evaluation metrics such as correlation (r), coefficient of determination (R_2), mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) and mean square error (MSE) between actual and predicted data. To acquire the data, information from the city of São José dos Campos-SP was used, between the years 2016 and 2018. The daily pollutant reports used were extracted from the CETESB (*Companhia Ambiental do Estado de São Paulo*) portal, and the data from *Declaração de Nascidos Vivos* (DNV), extracted from the free access portal of DATASUS (*Departamento de Informática do Sistema Único de Saúde*). Among all the prediction models developed, the neuro-fuzzy model with simple 50:50 validation, grid partition type structure, hybrid optimization method, with 20 epochs, Gaussian inputs, the inputs being mother's age, number of prenatal consultations and PM_{10} concentration with four membership functions and NO_2 concentration with three functions, and constant output variable, with metrics r=0.59, R_2=0.35, MAE=1.70, RMSE= 2.83, MAPE=5.35% and MSE=8.00. The approach presented in this study has the potential to benefit society as a whole, helping managers implement preventive measures aimed at reducing the emission of polluting gases. Aiming not only to improve the health of pregnant women and newborns, but also to promote quality of life for future generations.

KEYWORDS: Prematurity Neonatal; Mortality; Fuzzy Logic; Air Pollution; Public Health.

LISTA DE FIGURAS

Figura 1 - Classificação da pesquisa científica 20
Figura 2 – Trajetória de publicações no SCOPUS de 2000 a 2020 22
Figura 3 - Trajetória de publicações "Neuro-*fuzzy* e poluição do ar" no SCOPUS, de 2000 a 2020 23
Figura 4- Etapas básicas do Sistema *Fuzzy* 28
Figura 5 - Representação da arquitetura de uma rede neural artificial 30
Figura 6 – Representação do neurônio 31
Figura 7 - Arquitetura do sistema neuro-*fuzzy* ANFIS 32
Figura 8 - Mortalidade Infantil em São José dos Campos-SP de 2006 a 2020 37
Figura 9 – Tela de extração dos arquivos de dados do portal DATASUS 38
Figura 10 - Dados dos nascidos vivos de 2016 a 2018, após descompactação pelo TABWIN 39
Figura 11 - Localização das estações de medição de poluentes de 2021 em São José dos Campos-SP 40
Figura 12 – Modelo de relatório diário de NO_2 de 2016 pelo portal CETESB 41
Figura 13- Etapas de tratamento de dados DATASUS e CETESB 41
Figura 14 - Conjunto *fuzzy* para a variável de entrada: idade da mãe 42
Figura 15 - Conjunto *fuzzy* para a variável de entrada: número de consultas pré-natal 43
Figura 16 - Conjunto *fuzzy* para a variável de entrada: concentração de NO_2 44
Figura 17 - Conjunto *fuzzy* para a variável de entrada: concentração de PM_{10} 45
Figura 18 - Conjunto *fuzzy* para a variável de saída: prematuridade 46
Figura 19 - Editor de Regras do sistema *fuzzy* no MATLAB 48
Figura 20 – Treino de dados do sistema neuro-*fuzzy* no MATLAB 49
Figura 21 - Verificação de dados do sistema neuro-*fuzzy* no MATLAB 50
Figura 22 – Adição de funções de pertinência no sistema neuro-*fuzzy* do MATLAB 51
Figura 23 - Treino da função FIS no sistema neuro-*fuzzy* do MATLAB 51
Figura 24 - Treino do sistema neuro-*fuzzy* no MATLAB 52
Figura 25 - Teste do sistema neuro-*fuzzy* do MATLAB 52
Figura 26 - Estrutura do sistema neuro-*fuzzy* do MATLAB 53
Figura 27 - Conjunto neuro-*fuzzy* para a variável de entrada: idade da mãe 54
Figura 28 - Conjunto neuro-*fuzzy* para a variável de entrada: número de consultas 54
Figura 29 - Conjunto neuro-*fuzzy* para a variável de entrada: concentração NO_2 55

Figura 30 - Conjunto neuro-*fuzzy* para a variável de entrada: concentração PM_{10} 56
Figura 31 - Parâmetros para variável de saída "prematuridade" 56
Figura 32 - Editor de Regras do sistema neuro-*fuzzy* no MATLAB 57
Figura 33 - Valores diários de NO_2 nos anos de (a) 2016 (b) 2017 e (c) 2018 2018 na cidade de São José dos Campos-SP 58
Figura 34 - Valores diários de PM_{10} nos anos de (a) 2016 (b) 2017 e (c) 2018 2018 na cidade de São José dos Campos-SP 59
Figura 35 - Relação Mensal entre parto total e prematuros nos anos de (a) 2016 (b) 2017 e (c) 2018 2018 na cidade de São José dos Campos-SP 60
Figura 36 - Estrutura MATLAB do modelo *fuzzy* 64
Figura 37 - Primeiras regras do modelo *fuzzy* 65
Figura 38 - Regra 18 do sistema *fuzzy* 65
Figura 39 - Superfície do sistema *fuzzy* no MATLAB: Idade da mãe x Prematuridade 66
Figura 40 - Superfície do Sistema de Inferência *Fuzzy* no MATLAB: Idade da mãe x Concentração PM_{10} 66
Figura 41 - Dispersão dados reais x preditos *fuzzy* 67
Figura 42 - Parâmetros padrões *sub. clustering* 70
Figura 43 - Sequência de testes do modelo neuro-*fuzzy* 71
Figura 44 - Estrutura MATLAB do modelo neuro-*fuzzy* 71
Figura 45 - *Rule viewer* do sistema neuro-*fuzzy* 72
Figura 46 - *Rule viewer* da regra 77 72
Figura 47 - Validação cruzada 5-*fold* 73
Figura 48 - Dispersão dados reais x preditos neuro-*fuzzy* 74
Figura 49 - Métricas de desempenho: RMSE, MSE, MAPE e MAE para o modelo *fuzzy* e neuro-*fuzzy* 75
Figura 50 - Métricas de desempenho: R^2 e correlação para o modelo *fuzzy* e neuro-*fuzzy* .. 75

LISTA DE TABELAS

Tabela 1 – Classificação dos poluentes entre primários e secundários 25
Tabela 2 - Parâmetros do sistema *fuzzy* para classificação da idade da mãe 42
Tabela 3 - Parâmetros do sistema *fuzzy* para classificação do número de consultas pré-natal 43
Tabela 4 - Parâmetros do sistema *fuzzy* para classificação Dióxido de Nitrogênio 44
Tabela 5 - Parâmetros do sistema *fuzzy* para classificação Material Particulado..... 44
Tabela 6 - Parâmetros do sistema *fuzzy* para classificação da prematuridade 45
Tabela 7 - Base de regras inicial 46
Tabela 8 - Parâmetros do sistema neuro-*fuzzy* para classificação da idade da mãe 53
Tabela 9 - Parâmetros do sistema neuro-*fuzzy* para classificação do número de consulta pré-natal 54
Tabela 10 – Parâmetros do sistema neuro-*fuzzy* para classificação da concentração de NO_2 55
Tabela 11 - Parâmetros do sistema neuro-*fuzzy* para classificação da concentração de PM_{10} 55
Tabela 12 - Valores mínimos, máximos, média e desvio padrão para variáveis de entrada (idade da mãe, número de consultas, concentração de NO_2 e PM_{10}) e saída (prematuridade) para dados de nascidos vivos entre 2016 a 2018 em São José dos Campos-SP 61
Tabela 13 - Modelos *fuzzy* realizados 63
Tabela 14 - Testes neuro-*fuzzy* realizados..... 68
Tabela 15 - Melhores cinco testes neuro-*fuzzy* realizados 69
Tabela 16 - Melhor teste neuro-*fuzzy* realizado 70
Tabela 17 - Sintaxe para métricas de avaliação 74

LISTA DE ABREVIATURAS E SIGLAS

ANFIS	*Adaptive Neuro-fuzzy Inferece System*
CETESB	Companhia Ambiental do Estado de São Paulo
DNV	Declaração de Nascido Vivo
DATASUS	Departamento de Informática do Sistema Único de Saúde
MAE	Erro médio absoluto
MSE	Erro médio quadrático
MAPE	Erro percentual absoluto médio
IDH	Índice de Desenvolvimento Humano
IDHM	Índice de Desenvolvimento Humano Municipal
IBGE	Instituto Brasileiro de Geografia e Estatística
IA	Inteligência Artificial
IAM	Inteligência Artificial Médica
IC	Intervalo de Confiança
MATLAB	*Matrix Laboratory*
PNUD	Programa das Nações Unidas para o Desenvolvimento
QUALAR	Qualidade do Ar
RMSE	Raiz do erro quadrático médio
RNA	Rede Neural Artificial
SINASC	Sistema de Informações sobre Nascidos Vivos
MacOS	Sistema Operacional MAC
SUS	Sistema Único de Saúde

LISTA DE SÍMBOLOS

μ	micro
m	metro
m^3	metro cúbico
g	grama
ppm	partes por milhão
hv	radiação solar
km	quilômetro
km^2	quilômetro quadrado
ºC	graus Celsius
pdf	portable document format
dbc	data base CAN
dbf	data base file
xls	extensible style language
fis	file extension

SUMÁRIO

1 INTRODUÇÃO 17
1.1 QUESTÃO DA PESQUISA E OBJETIVOS 19
1.1.1 Questão da Pesquisa 19
1.1.2 Objetivo Geral 19
1.1.3 Objetivos Específicos 19
1.2 CLASSIFICAÇÃO DA PESQUISA CIENTÍFICA 19
1.3 ESTRUTURA DA TESE 20
2 DESENVOLVIMENTO 22
2.1 REVISÃO DA LITERATURA 22
2.1.1 Parto prematuro, poluentes do ar e *fuzzy* 22
2.1.2 Parto prematuro, poluentes do ar e neuro-*fuzzy* 23
2.2 REFERENCIAL TEÓRICO 24
2.2.1 Poluentes do Ar 24
2.2.1.1 Material Particulado 25
2.2.1.2 Dióxido de Nitrogênio 26
2.2.2 Inteligência Artificial 26
2.2.3 Inteligência Computacional 27
2.2.3.1 Computação Evolutiva 27
2.2.3.2 Lógica *Fuzzy* 28
2.2.3.3 Rede Neural Artificial 29
2.2.3.4 Sistema Neuro-*Fuzzy* 31
2.2.4 Métricas de Avaliação 34
2.2.4.1 Mean Absolute Error (MAE) 34
2.2.4.2 Root Mean Squared Error (RMSE) 35
2.2.4.3 Mean Absolute Percent Error (MAPE) 35
2.2.4.4 Mean Square Error (MSE) 35
2.2.4.5 Correlação (r) 36
2.2.4.5 Coeficiente de Determinação (R^2) 36
3 MATERIAIS E MÉTODOS 37
3.1 LOCAL DE ESTUDO 37
3.2 NASCIDOS VIVOS 37
3.3 RELATÓRIOS DE POLUENTES DO AR 39

3.4 MODELO ELABORADO POR *FUZZY* 41
3.4.1 Idade da mãe 42
3.4.2 Número de consultas pré-natal 43
3.4.3 Concentração de poluentes NO_2 e PM_{10} 43
3.4.4 Prematuridade 45
3.5 MODELO ELABORADO POR NEURO-*FUZZY* ANFIS 48
3.5.1 Idade da mãe 53
3.5.2 Número de consultas pré-natal 54
3.5.3 Concentração de poluentes NO_2 e PM_{10} 55
3.5.4 Prematuridade 56
4 RESULTADOS E DISCUSSÕES 57
4.1 RESULTADO DO MODELO ELABORADO POR *FUZZY* 52
4.1.1 Avaliação do Modelo Fuzzy 66
4.2 RESULTADO MODELO ELABORADO POR NEURO-*FUZZY* ANFIS 67
4.2.1 Validação Simples 67
4.2.2 Validação Cruzada k-fold 72
4.2.3 Avaliação do Modelo Neuro-Fuzzy 73
4.3 MÉTRICAS DE AVALIAÇÃO 74
5 CONCLUSÃO 76
REFERÊNCIAS 78
ANEXO 1 84

1 INTRODUÇÃO

Parto prematuro é definido como o parto ocorrido antes de 37 semanas completas de gestação. Estima-se que 15 milhões de nascimentos prematuros ocorrem todos os anos, e um milhão de crianças morrem em decorrência de complicações do parto (BICK, 2012).

Esse tipo de parto é causa de 75% a 95% dos óbitos neonatais em todo mundo, quando excluídas as malformações congênitas, e 50% dos comprometimentos neurológicos (PASSINI *et al.*, 2010).

A ocorrência desse parto é considerada expressiva em países subdesenvolvidos e em desenvolvimento (ALMEIDA *et al.*, 2012). No Brasil, em 2000, ocorreram 29 óbitos a cada mil nascidos. Projeções do IBGE apresentam queda na taxa de mortalidade infantil ao longo dos anos. Em números mais recentes divulgados, em 2021 ocorreram 11,2 óbitos a cada mil nascimentos realizados no país (IBGEa, 2023).

Além de ser a principal causa de mortes neonatais no mundo, a prematuridade ocasiona enfermidades que podem perdurar até a fase adulta (HARRISON e GOLDENBERG, 2016). Sendo assim, o nascimento prematuro pode comprometer o desenvolvimento humano, aumentando o risco de doenças graves e causando dificuldade de aprendizagem, problemas visuais e auditivos durante toda a vida do indivíduo (WHO, 2006).

Fatores como o nível socioeconômico, idade e peso da mãe, presença ou ausência de companheiros, hipertensão, tabagismo, uso de drogas ilícitas, escolaridade e fatores médicos, como prematuridade anterior, gravidez múltipla são agentes responsáveis pelos partos prematuros (BEHRMAN e BUTLER, 2007; STEER e FLINT, 1999).

Em estudos mais recentes, Nascimento, Machin e Santos (2017) encontraram associação entre a exposição materna a poluentes atmosféricos, como dióxido de nitrogênio (NO_2), Ozônio (O_3) e Material Particulado com diâmetro aerodinâmico menor de 10 μm (PM_{10}), principalmente para nascidos do sexo feminino. Foi utilizado o método de regressão logística para quantificar os efeitos desses poluentes nas gestantes.

Também há estudos que relacionam o efeito agudo no parto prematuro a exposição de Tolueno e Benzeno. Santos e Nascimento (2019) utilizaram um modelo de regressão logística com dados maternos e concentração dos poluentes, e constataram uma maior associação na janela de 5 a 10 dias anteriores ao parto.

Carvalho Nascimento e Nakazato (2014) apontam que, o parto prematuro tem relação com a exposição materna ao PM_{10} sete dias antes do parto, a significância estatística foi calculada por meio da regressão logística com modelo multipoluente (PM_{10}, O_3,NO e NO_2). O estudo foi realizado com dados de gestantes de idade entre 18 e 34 anos, residentes da cidade de Ribeiro Preto -SP, com parto normal e gestação única, no período de 2009 a 2010.

Os poluentes atmosféricos também estão associados ao peso insuficiente ao nascer, fator de risco para a morbidade e mortalidade infantil. O estudo de Santos *et al.* (2016), com dados de 2006 a 2010, utilizou a regressão logística univariada e múltipla para confirmar que a exposição materna entre o primeiro e terceiro semestre da gestação a PM_{10}, SO_2 e O_3 pode ser um fator determinante para o baixo peso ao nascer.

Machine learning ou aprendizado de máquina é um ramo da inteligência artificial baseado na ciência da computação e estatística, capaz de construir algoritmos por meio da aprendizagem de dados de uma maneira eficaz (MITCHELL, 2006). Um dos paradigmas da inteligência artificial é a lógica *fuzzy,* criada para lidar com problemas de imprecisão e incerteza, no qual é permitido que um elemento pertença a mais de um conjunto, a determinar pelo seu grau de pertinência, diferentemente dos conjuntos clássicos (IYODA, 2000; NASCIMENTO, RIZOL, ABIUZI, 2009). A união do sistema *fuzzy* e a capacidade de aprendizagem da rede neural permitiu a criação do sistema neuro-*fuzzy* ANFIS (*Adaptative neuro fuzzy inference system)*, também conhecido como sistema híbrido, podendo ser utilizado em aplicações diversas (JANG, 1993).

O estudo realizado por David, Rizol e Nascimento (2018) utilizou a lógica *fuzzy*, para estimar os efeitos da poluição do ar causados pelo NO_2 e PM_{10}, temperatura e velocidade do vento, em internações de crianças de até 10 anos por doenças respiratórias, na cidade de São José do Rio Preto - SP durante 2011 a 2013.

Chaves (2013), em seu estudo, utiliza a técnica *fuzzy* e neuro-*fuzzy* para verificar a associação entre a exposição ao material particulado (PM_{10}), dióxido de enxofre (SO_2), ozônio (O_3) e temperatura aparente nos casos de internações hospitalares por asma e pneumonia. Para esse estudo, foram utilizados dados da cidade de São José dos Campos-SP, no ano de 2007. Os dois modelos apresentaram bons resultados, o modelo neuro-*fuzzy* gerou maior precisão enquanto o modelo *fuzzy* apresentou melhor acurácia.

Esse trabalho aborda as técnicas *fuzzy* e neuro-*fuzzy*, para estimar a prematuridade, tendo como ponto de partida a exposição da grávida a poluentes de ar (NO_2 e PM_{10}) e a

outros fatores, como idade da mãe e número de consultas pré-natal realizadas. A utilização dessas técnicas nesse domínio é totalmente inédita na literatura.

1.1 QUESTÃO DA PESQUISA E OBJETIVOS

1.1.1 Questão da pesquisa

Existe associação entre os poluentes do ar, PM_{10} e NO_2, idade da mãe, número de consultas pré-natal realizadas e parto prematuro?

1.1.2 Objetivo Geral

Identificar se existe associação entre a exposição da grávida a poluentes de ar (PM_{10} e NO_2) e a outros fatores, como idade da mãe e número de consultas pré-natal realizadas, por meio da utilização das técnicas *fuzzy* e neuro-*fuzzy*.

1.1.3 Objetivos Específicos

- Identificar dados dos recém-nascidos da cidade de São José dos Campos-SP, nos anos de 2016 a 2018
- Identificar relatórios diários dos poluentes do ar, PM_{10} e NO_2, para o mesmo período
- Desenvolver um modelo baseado na lógica *fuzzy e* neuro-*fuzzy*, relacionando exposição das gestantes aos poluentes do ar com parto prematuro
- Realizar a comparação entre o modelo *fuzzy* e o sistema neuro-*fuzzy ANFIS* utilizados

1.2 CLASSIFICAÇÃO DA PESQUISA CIENTÍFICA

Para Fleury *et al.* (2012), uma pesquisa científica pode ser classificada quanto a sua natureza, básica ou aplicada, objetivos, como exploratório, descritivo, explicativo e empírico normativo, abordagem, de forma quantitativa, qualitativa ou combinada e método, podendo ser experimento, modelagem e simulação, levantamento/*survey*, estudo de caso, pesquisa-ação ou *soft system methodology*.

A classificação dessa pesquisa é classificada conforme apresentado na Figura 1, com natureza aplicada, objetivo empírico normativo, abordagem quantitativa e método de

modelagem e simulação. A seguir, serão abordados os conceitos de cada uma das classificações.

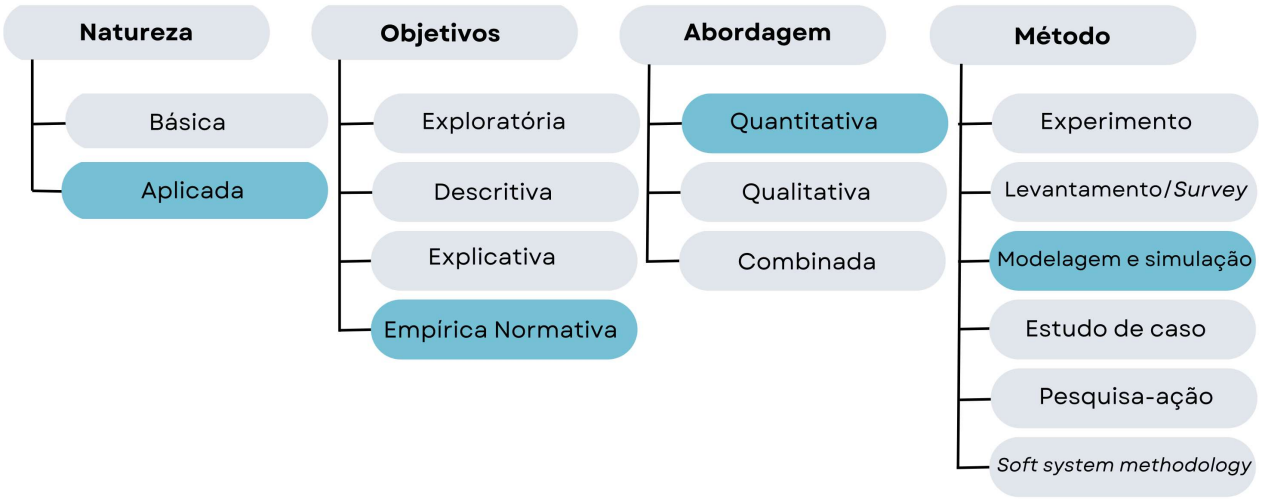
Natureza aplicada: segundo Silva e Menezes (2001), a pesquisa com tipo de natureza aplicada gera conhecimentos voltados para a solução de problemas específicos, envolvendo verdades e interesses locais.

Objetivo empírico normativo: para Bertrand e Fransoo (2002) esse objetivo está ligado ao desenvolvimento de possíveis políticas e ações para a solução de um problema específico.

Abordagem quantitativa: nessa abordagem, os resultados da pesquisa podem ser quantificados, a partir da utilização de linguagem matemática e técnicas estatísticas, sendo possível realizar classificações e análises dos dados (GIL, 1996).

Método de modelagem e simulação: utilizado quando há desenvolvimento de um modelo capaz de simular um sistema real (FLEURY *et al.* 2012).

Figura 1 - Classificação da pesquisa científica



Fonte: Adaptado de Fleury *et al.* (2012).

1.3 ESTRUTURA DA TESE

Esse trabalho está organizado em cinco capítulos, o primeiro deles contém a introdução com conceitos de parto prematuro, além do objetivo geral e específicos.

O segundo capítulo, Desenvolvimento, aborda a revisão da literatura, assim como o referencial teórico, a respeito dos poluentes do ar (Material Particulado, Dióxido de Nitrogênio) e a inteligência artificial, passando pela inteligência computacional e seus principais paradigmas, computação evolutiva, lógica *fuzzy* e rede neural artificial, chegando ao sistema neuro-*fuzzy* e métricas de avaliação.

O terceiro capítulo, Metodologia, apresenta o local e período de estudo, assim como foram coletados e utilizados os dados de nascidos vivos e poluentes do ar. Além disso, há explicação da forma de utilização dos sistemas *fuzzy* e neuro-*fuzzy*, a partir de cada variável de entrada e saída para os dois tipos de sistema.

No quarto capítulo, são apresentados os resultados e discussões do estudo.

Por fim, o quinto e último capítulo, cita as limitações encontradas e contribuições do estudo para a sociedade.

2 DESENVOLVIMENTO

O capítulo de Desenvolvimento apresenta a revisão na literatura sobre o tema, assim como o referencial teórico para todos os tópicos utilizados para a construção desse trabalho.

2.1 REVISÃO DE LITERATURA

2.1.1 Parto prematuro, poluentes do ar e *fuzzy*

O estudo bibliométrico sobre as palavras chaves "*fuzzy*", "*air pollution*", "*premature birth*" ou "*preterm birth*" foi realizado no artigo "*Bibliometric analysis: premature birth, air pollution and fuzzy logic*" por Castellões, T. de O.; David, T. M.; Nascimento, L. F. C.; Rizol, P. M. S. R., publicado pela revista *Research, Society and Development* em 2022.

O artigo citado acima trata das publicações de artigos completos entre os anos de 2000 a 2020. Foi realizada a busca singular das palavras chaves, assim como suas combinações. Para a "*fuzzy*" foram encontrados 156.204 artigos completos, "*premature birth*" ou "*preterm birth*" 24.769, enquanto para "*air pollution*" resultaram em 87.385 publicações. Para as combinações, "*preterm birth*" ou "*premature birth*" e "*air pollution*" resultaram em 299 resultados não duplicados, "*fuzzy*" e "*air pollution*", 437 publicações, "*preterm birth*" ou "*premature birth*" e "*fuzzy*" resultaram em 7 artigos não publicados, enquanto para a combinação completa ("*fuzzy*", "*air pollution*" e "*preterm birth*" ou "*premature birth*") não foram encontrados artigos. A trajetória anual das publicações combinadas é representada pela Figura 2.

Figura 2 – Trajetória de publicações no SCOPUS de 2000 a 2020



Fonte: Castellões *et al*. (2022).

Além do número de publicações, o artigo também cita os principais países de publicação, os Estados Unidos ocupam a primeira posição nas publicações, singulares e combinadas, seguidos de China, Canadá e Reino Unido. Ao final, o artigo destaca uma lacuna existente na literatura, com oportunidades de pesquisa nesse campo com o acréscimo da lógica *fuzzy*.

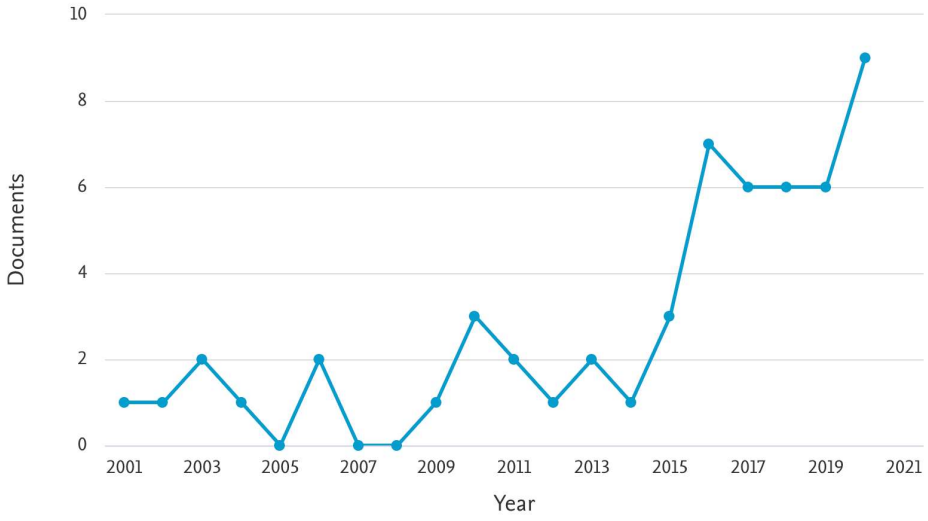
2.1.2 Parto prematuro, poluentes do ar e neuro-*fuzzy*

Para a verificação da literatura existente a respeito do parto prematuro, poluentes do ar e o sistema neuro-*fuzzy*, foi realizada uma pesquisa na base de dados SCOPUS nas palavras chave, resumos e títulos sobre artigos completos e revisões, durante os anos de 2000 a 2020. As análises bibliométricas das palavras "*air pollution*", "*preterm birth*", "*premature birth*" em sua singularidade e combinações foram realizadas por Castellões *et al.* (2022) e contempladas nesse estudo.

Já para o singular "neuro-*fuzzy*", 8.526 publicações foram encontradas, entre artigos completos e revisões. O artigo mais citado foi o "*Direct torque control of PWM inverter-fed AC motors - A survey*" de Buja, G.S., Kazmierkowski, M.P. de 2004, citado 1.154 vezes. O país com maior número de publicações relacionadas a neuro-*fuzzy* foi o Irã, com 1.840 publicações, seguido por Índia, China e Turquia, com 1.688, 715 e 680 publicações, respectivamente.

No mesmo período, para as palavras chaves "neuro-*fuzzy*" e "*air pollution*" foram encontradas 54 publicações, com trajetória anual representada na Figura 3.

Figura 3 - Trajetória de publicações "Neuro-*fuzzy* e poluição do ar" no SCOPUS, de 2000 a 2020



Fonte: SCOPUS (2023).

O país com o maior número de publicações dessa combinação (neuro-*fuzzy* e *air pollution*) foi o Irã, totalizando 10, em segundo e terceiro lugar, Índia e Turquia, com oito publicações cada, seguido dos Estados Unidos com cinco publicações. O artigo mais citado, com um total de 181 citações, foi "*Using data-mining approaches for wind turbine power curve monitoring: A comparative study*" de Schlechtingen, M., Santos, I.F., Achiche, S., publicado em 2013 pelo *IEEE Transactions on Sustainable Energy.*

Utilizando os mesmos parâmetros, tipo e ano de publicação, para as palavras chaves "neuro *fuzzy*" e "*preterm birth*" ou "*premature birth*", foi encontrado apenas um artigo. O artigo intitulado "*Green computing methodology on screening of retinopathy of prematurity using vital statistical features of retcam images*" foi publicado por Samson Isaac, J., Prabakar, S., Porkumaran, K. na Índia, em 2020, pelo *Journal of Green Engineering*. Nenhuma citação foi encontrada para essa pesquisa.

Quando todas as palavras chaves foram pesquisadas na base SCOPUS, "neuro *fuzzy*", "*air pollution*", "*preterm birth*" ou "*premature birth*", não foi encontrada nenhuma publicação.

2.2 REFERENCIAL TEÓRICO

2.2.1 Poluentes do Ar

> A poluição atmosférica pode ser definida como a presença de substâncias estranhas na atmosfera, resultantes da atividade humana ou de processos naturais, em concentrações suficientes para interferir direta ou indiretamente na saúde, segurança e bem estar dos seres vivos. (DEREK, 1992)

Conforme o manual da CETESB (2023), considera-se poluente qualquer substância presente no ar e que, pela sua concentração, possa torná-lo impróprio, nocivo ou ofensivo à saúde, causando inconveniente ao bem estar público, danos aos materiais, à fauna e à flora ou prejudicial à segurança, ao uso e gozo da propriedade e às atividades normais da comunidade. O nível de poluição atmosférica é medido pela quantidade de substâncias poluentes presentes no ar. A variedade das substâncias que podem ser encontradas na atmosfera é muito grande, o que torna difícil a tarefa de estabelecer uma classificação. Para facilitar essa classificação, os poluentes são divididos em duas categorias, conforme apresentado na Tabela 1.

Tabela 1 – Classificação dos poluentes entre primários e secundários

Poluentes Primários	Poluentes Secundários
Aqueles emitidos diretamente pelas fontes de emissão	Aqueles formados na atmosfera através da reação química entre poluentes primários e componentes naturais da atmosfera

Fonte: CETESB (2023)

Requia e Roig (2016) apontam que nove em cada dez pessoas no mundo respiram ar contaminado, e tal contaminação pode levar a morte e a doenças como derrame, doenças cardíacas, câncer de pulmão, doenças pulmonares obstrutivas crônicas e infecções respiratórias incluindo pneumonia. As principais substâncias que auxiliam na poluição atmosférica são os particulados, óxidos de nitrogênio, dióxido de enxofre, monóxido de carbono e ozônio.

A seguir, são apresentados os poluentes escolhidos como principais percursores do parto prematuro, Material Particulado e Dióxido de Nitrogênio.

2.2.1.1 Material Particulado

Para WHO (2006), são denominados material particulado (MP) as partículas sólidas ou líquidas presentes na atmosfera. A maioria delas são visíveis a olho nu, e se manifestam a partir da poeira, cinzas e fumaça. Partículas menores que 10 μm ficam retidas no trato respiratório superior, ao passo que partículas menores que 2,5 μm atingem os pulmões, ficando retidas e podendo causar danos à saúde.

Araújo *et al*. (2014) apontam que as emissões do PM_{10} têm origens variadas, podendo ser provenientes da queima de combustíveis fósseis em veículos a chaminés industriais, queimadas e até vias sem pavimentação. Esse poluente está relacionado a doenças respiratórias, como asma brônquica e outras doenças alérgicas, além de doenças cardiovasculares.

Para Nascimento e Medeiros (2012), o PM_{10} inalável apresenta uma característica importante que é o transporte dos gases retidos na superfície até as vias aéreas mais internas, onde acontecem as trocas de gasosas no pulmão. Conforme os gases poluentes entram no trato respiratório, o corpo ativa alguns mecanismos de defesa. Um dos mecanismos iniciais é o espirro, como as partículas são grandes, não conseguem passar das narinas e se depositam, ocasionando o espirro. Outro mecanismo é a tosse, que ocorre de

forma similar ao espirro, quando há invasão além da laringe. Quando as partículas se depositam em células com cílios e células secretoras de muco, o aparelho mucociliar entra em ação. Os cílios empurram o muco do pulmão para a boca, fora do trato respiratório, carregando também as partículas que se depositam sobre os mucos, dessa forma, são removidas via sistema linfático ou aparelho mucociliar.

2.2.1.2 Dióxido de Nitrogênio

Conforme WHO (2006), o óxido nítrico (NO) e dióxido de nitrogênio (NO_2) são os dois principais associados com os óxidos de nitrogênio em fontes de combustão. O óxido nítrico é oxidado no ar para formar dióxido de nitrogênio. Na sua forma líquida, coloração do dióxido de nitrogênio varia de incolor a marrom. Na forma gasosa ele é volátil, marrom-avermelhado e mais pesados do que o ar. Tem um odor característico intenso, perceptível a partir de uma concentração de 188 μg / m^3 (0,1 ppm). Ele reage com água e é solúvel em ácidos sulfúrico e nítrico.

Esse poluente é apontado como o óxido de nitrogênio mais tóxico, sendo capaz de formar potenciais agentes cancerígenos e infecções respiratórias por bactérias (ZAKRZEWSK, 1997).

2.2.2 Inteligência artificial

“A inteligência artificial é definida como uma habilidade de um sistema para interpretar corretamente dados externos para aprender por tal data e usar aqueles aprendizados para atingir objetivos específicos e tarefas por meio de adaptações flexíveis” (KAPLAN; HAENLEIN, 2019, p.22).

Para Moura, Andrade e Kikuchi (2005), a IA possui características bem definidas, como aplicar o conhecimento por meio do aprendizado, tanto com situações simples como complexas, podendo ser aplicada em resoluções de problemas, mesmo se as informações se apresentarem de forma incompleta, pois ela é capaz de interpretar dados a partir da experiencia. Segundo Rodrigues *et al*. (2021), o uso desse tipo de inteligência se aplica em diversas áreas, como na saúde, por conseguir, através do *machine learning*, identificar padrões e prever resultados através de técnicas computacionais.

Chaves (2013) aponta que cada vez mais, a Inteligência Artificial Médica (IAM) ou Inteligência Artificial em Medicina, utiliza recursos avançados para auxiliar os profissionais nas tomadas de decisão e em atividades como diagnósticos e análises médicas.

2.2.3 Inteligência Computacional

"Inteligência computacional significa conceitos, paradigmas, algoritmos e implementações de sistemas que em teoria exibem comportamentos inteligentes em ambientes complexos" (HILLER; LEIBERMAN, 2017).

Para Iyoda (2000), a inteligência computacional busca desenvolver uma inteligência baseada nos sistemas biológicos. Algumas inspirações estão nas redes neurais e computação evolutiva, enquanto outros utilizam sistemas naturais como base para o desenvolvimento. Os principais paradigmas da inteligência computacional são: redes neurais artificiais, sistema *fuzzy* e computação evolutiva.

A seguir, será abordada uma breve explicação a respeito dos principais paradigmas da inteligência computacional.

2.2.3.1 Computação Evolutiva

Para Eiben e Smith (2003), a computação evolutiva é o termo utilizado para designar o conjunto de técnicas de resolução de problemas, como o algoritmo genético, que utiliza como base os conceitos da evolução biológica, seleção natural e herança genética. Dessa forma, a computação evolutiva pode ser capaz de obter soluções para grandes problemas, com aplicações em diversos locais, como por exemplo, na indústria e comércio.

Para Nagashima (2006), os algoritmos utilizados na computação evolutiva (ou evolucionária) é chamado de algoritmo evolucionário, entre eles os mais conhecidos são algoritmos genéticos, programação evolucionária e estratégias evolucionárias.

Goldberg (1989) afirma que os algoritmos Genéticos constituem uma técnica de busca e otimização, altamente paralela, inspirada no princípio Darwiniano de seleção natural e reprodução genética. Destacam-se as principais vantagens do algoritmo genético:

1) Não requer informações sobre o gradiente da superfície de resposta;
2) Eventuais descontinuidades da superfície de resposta não afetam a performance da otimização;
3) A presença de mínimos locais não reduzem a eficiência do algoritmo;
4) A performance do algoritmo tem apresentado excelentes resultados para problemas de otimização de grande escala (GOLDBERG, 1989).

2.2.3.2 Lógica *Fuzzy*

A chamada lógica *fuzzy* ou difusa foi implementada em meados dos anos 60, por Lotfi Zadeh, após perceber deficiências no desenvolvimento de problemas complexos. O desenvolvimento da teoria representou uma grande mudança na forma da resolução de problemas anteriormente não sanados pela lógica clássica (KRYKHTINE, 2018).

Lógica *fuzzy* de um ponto de vista simplório pode ser considerada um sistema lógico com uma extensão da lógica multi-valor. Entretanto, de uma perspectiva mais ampla e conceitual, lógica *fuzzy* reflete a teoria de conjuntos *fuzzy* que trata de classificações (em conjuntos) com fronteiras suaves determinadas por graus de pertinência e, portanto, abarca reflexões teóricas mais profundas (MATHWORKS, 2012).

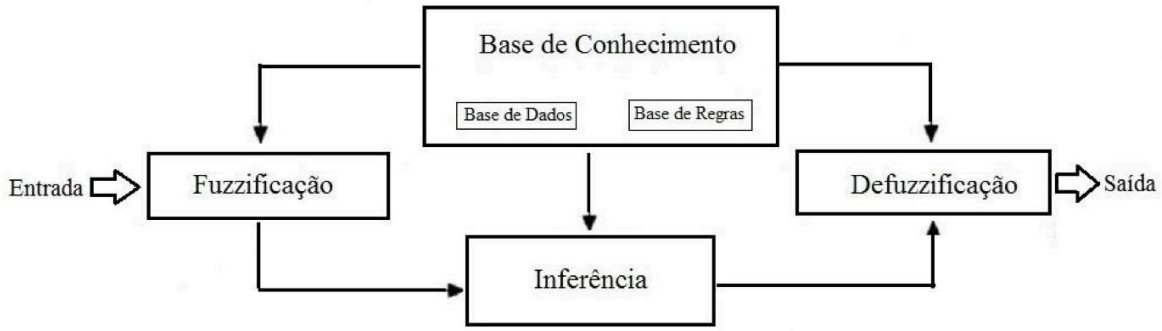
De acordo com Zadeh (1965), a lógica *fuzzy* é um sistema conceitual preciso de raciocínio, dedução e computação em que os objetos do discurso e da análise são ou podem ser associados a informações imperfeitas, as quais podem ser imprecisas, incertas, incompletas, não confiáveis, vagas e/ou parcialmente verdadeiras. A capacidade da lógica *fuzzy* de expressar transições graduais da pertinência a não pertinência e vice-versa tem uma excelente utilidade, provendo não somente uma significante e poderosa representação da medida de incerteza, mas também uma significante representação de conceitos vagos expressados na linguagem natural (MORAES, 2008).

Segundo Cox (1994), para casos complexos em que há incidência de imprecisão e incerteza, a lógica *fuzzy* é amplamente recomendada pois apresenta um modelo em que há a combinação da imprecisão e o poder computacional, capazes de produzir um sistema com boas respostas.

Existem cinco etapas básicas em um sistema de inferência *fuzzy,* ilustrado na Figura 4.

1) Interface de fuzzificação que transforma as informações de entrada em conjuntos *fuzzy*;
2) A base de regras que contém as regras do tipo “se ... então” elaboradas pelos especialistas;
3) A base de dados que contém as funções de pertinência dos conjuntos *fuzzy* usados nas regras. Normalmente a base de dados e a base de regras são unificadas formando a base de conhecimento.
4) O método de inferência para saída *fuzzy*;
5) Uma interface de defuzzificação, cuja função é converter os valores *fuzzy* em valores precisos. (JANG, 1993)

Figura 4- Etapas básicas do Sistema *Fuzzy*



Fonte: Jang (1993).

No sistema *fuzzy*, os métodos de inferência mais utilizados são Takagi-Sugeno e Mamdani. Segundo Coelho e Mattos (2014), o método de inferência normalmente utilizado é o Método Mamdani, também conhecido como Mínimo-Máximo, cujas regras são criadas por especialistas, para o operador de junção das regras "e" é utilizado o mínimo para a combinação de graus de pertinência, enquanto no operador disjunção "ou" é utilizado o máximo. Os dois métodos (Takagi-Sugeno e Mamdani) se diferem no modo de representação de cada regra e resposta do sistema. Mamdani permite que o conhecimento de especialistas seja aplicado na base de regras, enquanto Tagaki-Sugeno utiliza as funções de entrada como respostas das regras e saídas (SIVANANDAM, 2007).

O MATLAB (*Matrix Laboratory*) é uma linguagem de computação técnica, amplamente utilizada para cálculos matemáticos, modelagens, simulações e desenvolvimentos de algoritmos. O software possui um pacote padrão com ferramentas de diversas áreas de conhecimento, porém é possível adicionar novas ferramentas de acordo com a área e problema e ser solucionado (GILAT, 2009).

Uma das ferramentas adicionais ao MATLAB é o *Fuzzy Logic Toolbox*, com funções capazes de analisar, projetar e simular sistemas baseados na lógica *fuzzy*. A combinação entre o software MATLAB e *Fuzzy Logic Toolbox* gera uma poderosa ferramenta de resolução de problemas utilizando a lógica difusa (MATLAB, 2012).

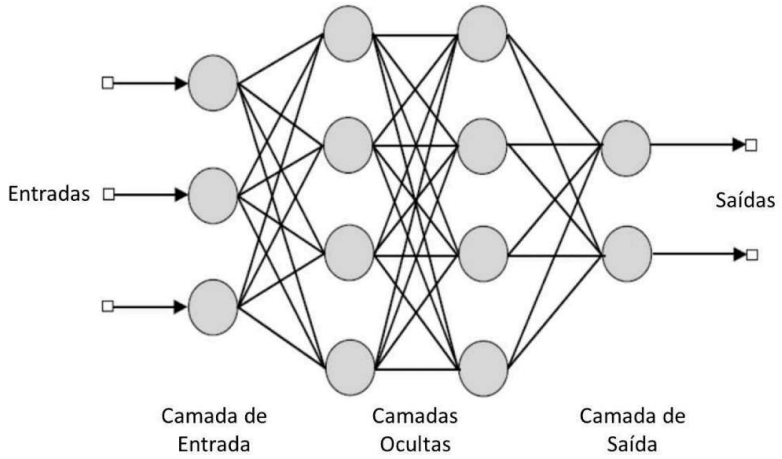
2.2.3.3 Rede Neural Artificial

"Redes Neurais podem ser definidas como uma rede massivamente paralela de elementos interconectados e suas organizações hierárquicas que estão preparadas para iterar com objetos do mundo real do mesmo modo que um sistema nervoso biológico faz" (TAFNER, 1996).

Em sua forma mais geral, para Haykin (2001), uma rede neural é um sistema confeccionado para realizar suas tarefas de forma similar ao cérebro humano, porém com estímulos eletrônicos e simulado por um computador digital. Braga, Carvalho e Ludemir (2000) afirmam que as redes neurais artificiais fazem parte de um sistema capaz de adquirir aprendizado e generalização, assim como o cérebro.

Para possuir um bom desempenho, as redes neurais se interligam em células computacionais, assim organizadas por camadas, com um número finito de unidades individualizadas, também designadas por neurônios, cuja arquitetura proposta por Silva (2006) está representada na Figura 5.

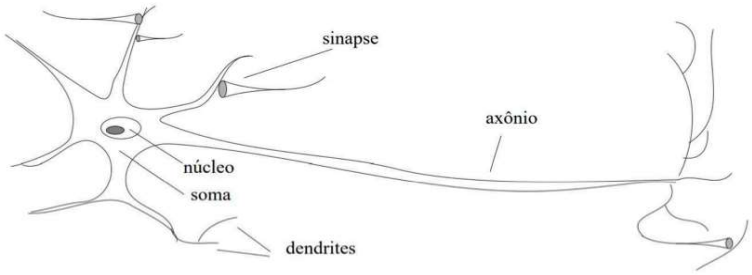
Figura 5 - Representação da arquitetura de uma rede neural artificial



Fonte: Silva (2006).

Conforme Rauber (2001), o neurônio é uma célula com núcleo e corpo (ou soma), todos os processamentos de informações ocorrem por reações químicas e elétricas. As saídas são propagadas a partir dos axônios, e suas ramificações são capazes de distribuir as informações para outros neurônios. Cada ligação de neurônio é realizada através da chamada sinapse, conectada por dendrites. A representação do modelo do neurônio é apresentada na Figura 6.

Figura 6 – Representação do neurônio



Fonte: Rauber (2001).

Para Sporl *et al*. (2011), uma das principais característica das RNAs é a sua eficiência em solucionar problema complexos, por meio da aprendizagem de exemplos e informações recebidas. Além dessa característica, as RNAs apresentam:

1) Bom desempenho em tarefas pouco ou mal definidas, onde falta o conhecimento explicito de como resolvê-las, o aprendizado se dá através de exemplos;
2) Robustez à presença de informações falsas ou ausentes, escolha dos elementos no próprio conjunto de treinamento (integridade do conjunto de treinamento);
3) No contexto de classificação de padrões, uma rede neural pode fornecer informações sobre quais padrões selecionar em função do grau de confiança apresentado (confiabilidade do conjunto de treinamento);
4) Tolerância à falha. (BARONE, 1999)

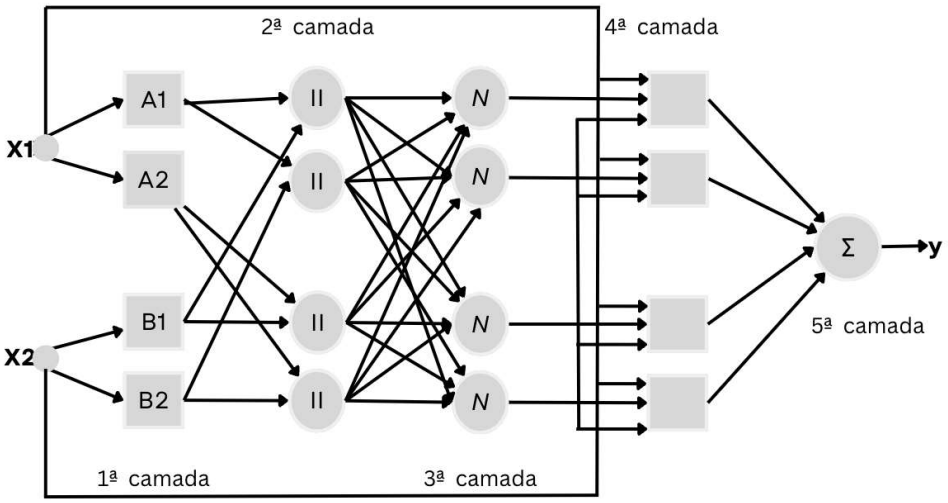
2.2.3.4 Sistema Neuro-*fuzzy*

Silva *et al*. (2014) apontam que com a integração entre os modelos *fuzzy* e redes neurais, um novo sistema formado é capaz de responder a comportamentos mais complexos e não lineares, formando assim o sistema neuro-*fuzzy*, também conhecido como sistema hibrido.

O objetivo da integração entre os sistemas, para Fuller (1995) é minimizar suas desvantagens isoladas, evidenciando e promovendo as qualidades da unificação dos sistemas. Como, por exemplo, a utilização do sistema de inferência *fuzzy* para o aprendizado usando algum tipo de algoritmo de treinamento baseado nos modelos de redes neurais artificiais, criando assim uma arquitetura homogênea.

Para Freitas (2021), o *adaptative neuro fuzzy inference system*, mais conhecido como ANFIS, é uma rede neuro-*fuzzy* composta por cinco camadas, cada uma responsável por uma etapa do processo, como representado na Figura 7.

Figura 7 - Arquitetura do sistema neuro-*fuzzy* ANFIS



Fonte: Adaptado de Freitas (2021).

Jang (1993) considerou um modelo representativo para o sistema neuro *fuzzy* com método de inferência Takagi-Sugeno, duas variáveis de entrada, funções de pertinência do tipo gaussiana e uma única saída para explicar as cinco camadas existentes na estrutura ANFIS.

Dessa forma, x_1 e x_2 são as entradas do sistema, enquanto a primeira camada da arquitetura do ANFIS representa os nós responsáveis por calcular o grau de pertinência referente a cada variável de entrada. Essa camada pode ser chamada de *fuzzificação*. As funções da 1ª camada são representadas em (1):

$$y_i^{(1)} = \mu A_{1\,(x1)} \text{ para i = 1,2}$$

$$y_i^{(1)} = \mu B_{i-2\,(x1)} \text{ para i = 3,4} \qquad (1)$$

A_i e B_{i-2} são os termos linguísticos relacionados as entradas x_1 e x_2, respectivamente.

Na segunda camada, são estabelecidas as relações a parir da criação da parte "SE" da base de regras, e o operador "E" é definido por uma t-norma. Suas funções representadas em (2):

$$y_1^{(2)} = w1 = \mu A_{1\,(x1)} . \mu B_{1\,(x2)}$$

$$y_2^{(2)} = w2 = \mu A_{1\,(x1)} . \mu B_{2\,(x2)}$$

$$y_3^{(2)} = w3 = \mu A_{2\,(x1)} . \mu B_{1\,(x2)}$$

$$y_4^{(2)} = w4 = \mu A_{2\,(x1)} . \mu B_{2\,(x2)} \qquad (2)$$

Em que $y_1^{(2)}$ é a saída do w_i obtida pela t-norma do produto algébrico.

A terceira camada, normalização, é composta por nós fixos que realizam a normalização das regras criadas na camada anterior. As funções dessa camada são representadas abaixo (3):

$$y_i^{(3)} = \overline{wi} = \frac{wi}{w2 + w3 + w4} \tag{3}$$

Na penúltima camada, *defuzzificação*, há criação da parte "ENTÃO" das regras, por meio de parâmetros consequentes $\{k_{i0}, k_{i1}, k_{i2}\}$. Nessa camada, o valor de cada média ponderada de cada regra é calculado a partir das funções representadas a seguir (4):

$$y_i^{(4)} = \overline{wi}\, fi = \overline{wi}\, (k_{i0}{}^{x1}{}_{+}\, k_{i1}{}^{x2}{}_{+} k_{i2}) \tag{4}$$

A última camada, saída, é composta por um único nó, responsável por calcular a saída do sistema neuro *fuzzy*. O cálculo se dá através do somatório da média ponderada das saídas de cada regra escrita. A função final do sistema é representada na equação (5):

$$y_i^{(5)} = \sum_{i=1}^{4} \overline{wi}\, fi \tag{5}$$

"O ANFIS é utilizado em aplicações de previsão e aproximação de funções e emprega como sistema de inferência um modelo *fuzzy* Takagi-Sugeno" (JANG & SUN, 1995).

Assim como na lógica *fuzzy*, a utilização do software MATLAB R2022b®, acrescida do *Fuzzy Logic Toolbox* também é indicada na resolução de problemas com a técnica neuro-*fuzzy*. Dessa vez, esse toolbox calcula os parâmetros das funções de pertinência conforme dados de entrada e saída fornecidos e a função ANFIS realiza os ajustes dos parâmetros das funções de pertinências (MATHWORKS, 2012).

A geração do sistema neuro-*fuzzy* pode ser do tipo *grid partition* ou s*ubstractive clustering*. O tipo *grid partition* gera um *fis* de saída única do tipo Sugeno, dividindo o espaço de dados em subespaços retangulares, tendo como base as funções de pertinência e dimensões definidas. Já o tipo *subtractive clustering*, aplica o primeiro agrupamento

subtrativo de dados, e é geralmente utilizado quando não há clareza a respeito da quantidade de agrupamentos a serem determinados para o conjunto de dados. (WEI *et al.*, 2007; MATHWORKS, 2012).

Para otimização e treinamento do sistema neuro-*fuzzy* ANFIS, dois métodos são os mais utilizados: *backpropagation* e *hybrid.* O método *backpropagation* (retropropagação) é um algoritmo de aprendizado e treinamento capaz de reduzir a taxa de erro por meio de ajuste de peso na diferença da saída e destino desejado, logo, o método *hybrid*, é a combinação de mínimos quadrados (*least-squares*) e o *backpropagation* (MATHWORKS, 2012; WYTHOFF, 1993).

Para James *et al.* (2017), a melhor forma para se avaliar um modelo de predição é dividir a amostra em duas partes, uma para treino e outra para avaliação, método conhecido como validação simples, podendo ser nas proporções 70:30, 80:20 ou 50:50. Castanho (2007) estuda o prognóstico de câncer de próstata utilizando a técnica de neuro-*fuzzy* com validação simples 50:50, onde metade dos dados são destinados a treino e restante para teste. Porém, outros estudos apontam o método de validação cruzada, *k-fold* como técnica mais precisa de avaliação. Essa técnica consiste em separar a amostra total em k subconjuntos, onde cada bloco é testado e os k-1 restantes são utilizados como dados para treinamento do modelo (GÉRON, 2017).

2.2.4 Métricas de Avaliação

Paliwal e Kumar (2009) afirmaram que existem algumas métricas de regressão linear, como MAPE (*Mean Absolute Percent Error*), MSE (*Mean Square Error*), MAE (*Mean Absolute Error*) e RMSE (*Root Mean Square Error*) capazes de medir o desempenho de modelos em diversos estudos. O conceito de cada métrica é apresentado a seguir.

2.2.4.1 *Mean Absolute Error* (MAE)

Para Moura (2021), ao ser calculado o erro (diferença entre valor real e previsto), o resultando retornaria com valores positivos ou negativos, não representando o valor total do erro. Por esse motivo, o cálculo do erro médio absoluto é indicado, pois os erros estarão em forma de módulo, com o mesmo sinal para todos os dados. A equação que representa a MAE é ilustrada em (6):

$$\mathrm{MAE} = \frac{1}{n}\sum_{1}^{n} |(y_i - \hat{y}_i)|$$

(6)

Onde y_i é o valor real do período i, $\hat{y}_i$ é a previsão para o período i e n é o número de observações.

Essa métrica é indicada por Hyndman e Koehler (2006) quando todos as séries estiverem na mesma escala.

2.2.4.2 *Root Mean Squared Error* (RMSE)

Conforme Mingoti (2005), a raiz quadrada do erro quadrático é uma métrica que avalia a adequabilidade do modelo, apresentando um índice de variância entre os valores reais e os encontrados pelo modelo proposto que varia entre zero e um, sendo que quanto mais próximo de zero, melhor o ajuste do modelo. A equação é apresentada em (7):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}((yi - \hat{y}_i)^2)}$$

(7)

2.2.4.3 *Mean Absolute Percent Error* (MAPE)

O erro percentual absoluto médio, segundo Wyatt (2012), é o valor absoluto entre o real e o proposto pelo modelo, comparado ao real, expresso em porcentagem. Dessa forma, quanto menor o valor MAPE, melhor a previsão. A equação desse métrica é representada em(8):

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} |(y_i - \hat{y}_i)/y_i \text{ | x } 100$$

(8)

2.2.4.4 *Mean Square Error* (MSE)

Para Devore (2012), o erro médio quadrático utiliza a potência para elevar todos os valores ao mesmo sinal. Com essa métrica, os maiores erros ficam mais evidenciados, visto que a potenciação aumenta a diferença entre os valores.

A equação do MSE é representada em (9):

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

(9)

2.2.4.5 Correlação (r)

A função correlação é um parâmetro de validação do sistema pois determina a relação existente entre os dados reais e os propostos. A equação da correlação é apresentada em (10):

$$r = \frac{\sum_{i=1}^{n}(x - \overline{x})(y_i - \overline{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \overline{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \overline{y})^2\right]}} \quad (10)$$

Onde n é a quantidade da amostra, x_i e y_i representam as variáveis, $\overline{x}$ e $\overline{x}$ as médias das amostras.

Um valor ótimo para correlação seria próximo de 1,00. Porém para a área biológica este valor é dificilmente alcançado e considera-se bom um valor próximo de 0,60 (FARAWAY, apud TADANO, 2012).

2.2.4.6 Coeficiente de Determinação (R^2)

Conforme Devore (2012), o coeficiente de determinação é utilizado em modelos lineares, para melhor ajuste ao conjunto real de dados. Essa métrica possui a capacidade de determinar a porcentagem da variação do modelo. A fórmula para o R^2 é representada na equação (11):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - \hat{y}_i)^2}{\sum_{i=1}^{n}(yi - \overline{y})^2} \quad (11)$$

Onde $\overline{y}$ é o valor médio das amostras.

Com o encerramento desse capítulo, será realizada a transição para as próximas etapas da pesquisa, como o estabelecimento do local de estudo, extração de dados e modelagem, tendo como base o referencial teórico apresentado.

3 MATERIAIS E MÉTODOS

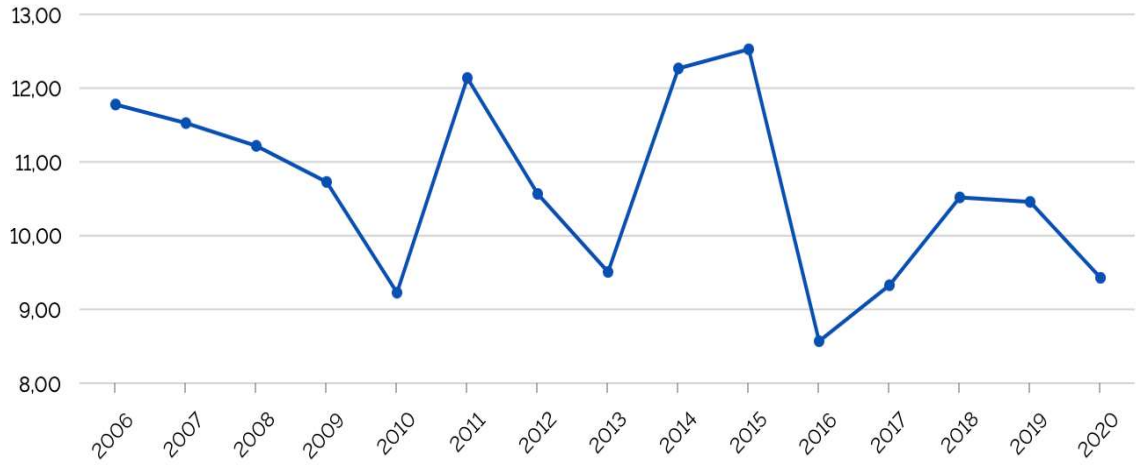
3.1 LOCAL DE ESTUDO

A cidade de São José dos Campos está localizada no interior do Estado de São Paulo, na região metropolitana do Vale do Paraíba. Segundo o Instituto Brasileiro de Geografia e Estatística (IBGEb, 2023), o município possui área territorial de 1.099,409 km² , população estimada em 2021 de 737.310 pessoas, com PIB per capita de 53.646,74 e Índice de desenvolvimento humano municipal (IDHM) de 0,807.

Esse município é um importante polo industrial, respondendo por 51,96% da economia do município, com destaque para as áreas aeroespacial, petroquímica, farmacêutica e automobilística, contando com empresas como a Embraer, Petrobrás, General Motors, entre outras.

Em 2020, ocorreram 9,44 óbitos por mil nascidos vivos, a taxa anual pode ser observada na Figura 8, onde o eixo Y representa a quantidade de óbitos por mil nascidos vivos e o eixo X representa os anos.

Figura 8 - Mortalidade Infantil em São José dos Campos-SP de 2006 a 2020



Fonte: IBGEb (2023).

3.2 NASCIDOS VIVOS

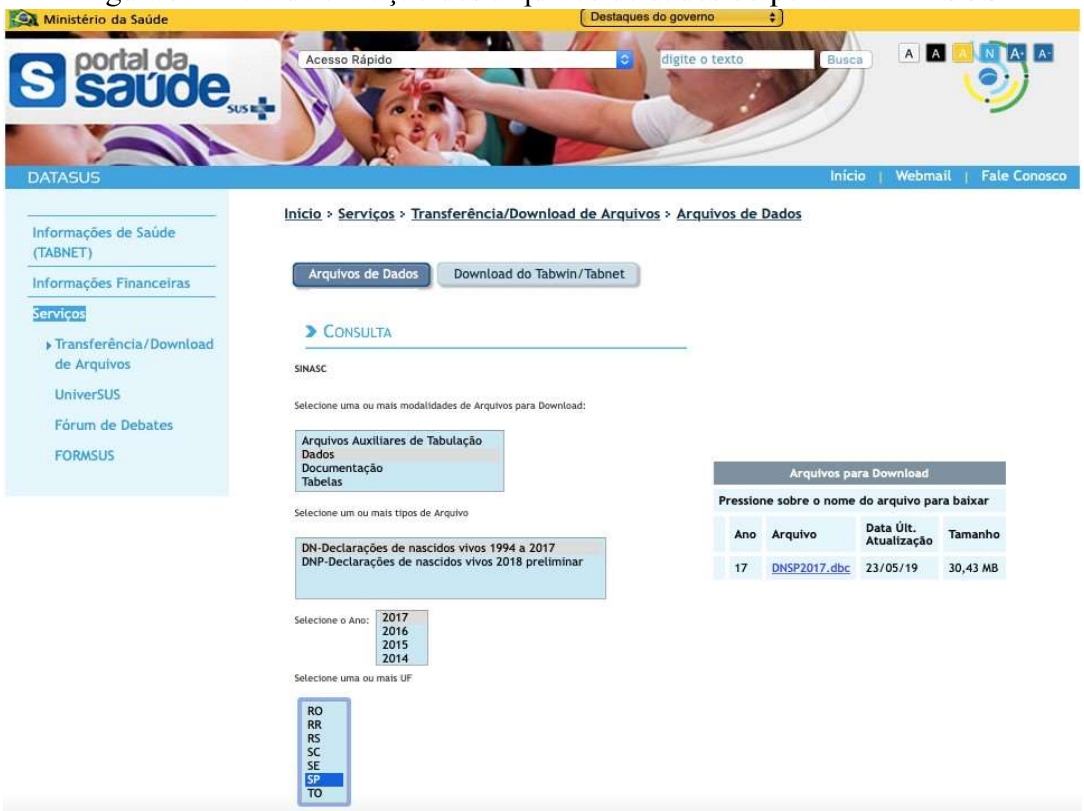
O Departamento de Informática do Sistema Único de Saúde (DATASUS) surgiu em 1992 e tem como missão “promover modernização por meio da tecnologia da informação para apoiar o Sistema Único de Saúde (SUS)” (DATASUS, 2023). Esse departamento é responsável por prover sistemas de informação e suporte de informática, necessários ao

processo de planejamento, operação e controle, além de coletar, processar e disseminar informações a respeito da saúde no Brasil.

Em seu portal de livre acesso, auxiliado pelo programa TABWIN, é possível visualizar um banco de dados com informações coletadas de todo o país. Um dos dados disponibilizados é a Declaração de Nascido Vivo (DNV) através do Sistema de Informações sobre Nascidos Vivos (SINASC).

Na Figura 9 abaixo, tela capturada do portal do DATASUS, está o local de onde foram obtidas as informações dos nascidos vivos. O arquivo extraído contém os dados de nascidos em todos os mais de 600 municípios de São Paulo, com cerca de 600 mil informações.

Figura 9 – Tela de extração dos arquivos de dados do portal DATASUS



Fonte: DATASUS (2021).

Após extraída, o arquivo em .dbc é descompactada para extensão .dbf pelo programa de acesso público TABWIN, apresentado na Figura 10.

Figura 10 - Dados dos nascidos vivos de 2016 a 2018, após descompactação pelo TABWIN

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CODMUNRES	DTNASC	SEMAGESTAC	PESO	PARTO	ESCMAEAGR1	IDADEMAE	ESTCIVMAE	QTDFILMORT	GRAVIDEZ	PARTO	SEXO	CONSPRENAT	MESPRENAT	IDANOMAL
2	SAO JOSE DOS CAMPOS	22062017	40	3580	1	02	23	1	00	1	1	1	09	02	2
3	SAO JOSE DOS CAMPOS	21042017	38	2785	2	08	32	2	00	1	2	2	08	01	2
4	SAO JOSE DOS CAMPOS	17032017	38	2980	2	08	32	2	01	1	2	1	12	02	2
5	SAO JOSE DOS CAMPOS	10052017	39	3025	2	06	28	2	00	1	2	1	11	01	2
6	SAO JOSE DOS CAMPOS	20032017	36	2390	2	08	33	1	00	1	2	2	08	03	2
7	SAO JOSE DOS CAMPOS	04012017	38	3025	2	06	23	2	00	1	2	1	06	07	2
8	SAO JOSE DOS CAMPOS	16012017	39	3550	2	06	37	2	01	1	2	1	08	03	2
9	SAO JOSE DOS CAMPOS	02012017	38	3400	2	05	25	1	00	1	2	2	10	01	2
10	SAO JOSE DOS CAMPOS	15022017	38	2825	2	05	29	2	00	1	2	1	08	02	2
11	SAO JOSE DOS CAMPOS	20032017	41	3385	2	08	30	2	00	1	2	1	06	05	2
12	SAO JOSE DOS CAMPOS	03042017	40	2915	2	06	26	1	00	1	2	1	05	04	2
13	SAO JOSE DOS CAMPOS	04052017	37	2845	1	08	23	5	02	1	1	1	09	01	2
14	SAO JOSE DOS CAMPOS	29052017	36	2830	2	06	18	1	00	1	2	1	07	04	2
15	SAO JOSE DOS CAMPOS	10012017	27	1080	2	06	29	2	00	1	2	2	01	04	2
16	SAO JOSE DOS CAMPOS	04032017	39	3140	2	03	20	1	00	1	2	2	01	06	2
17	SAO JOSE DOS CAMPOS	13052017	34	2860	2	06	24	1	00	1	2	1	10	01	2
18	SAO JOSE DOS CAMPOS	01052017	30	1450	1	02	16	5	00	1	1	1	04	02	
19	SAO JOSE DOS CAMPOS	20022017	38	3760	2	08	35	2	00	1	2	2	08	07	2
20	SAO JOSE DOS CAMPOS	01012017	36	3290	2	08	26	2	00	1	2	1	05	04	2
21	SAO JOSE DOS CAMPOS	02012017	40	3405	1	12	35	2	00	1	1	2	12	01	2
22	SAO JOSE DOS CAMPOS	12012017	38	2920	2	08	29	2	00	2	2	2	09	03	2
23	SAO JOSE DOS CAMPOS	12012017	38	1625	2	08	29	2	00	2	2	1	09	03	2
24	SAO JOSE DOS CAMPOS	15012017	45	3770	1	08	29	2	00	1	1	1	19	02	2
25	SAO JOSE DOS CAMPOS	10012017	37	2710	2	08	43	2	02	1	2	2	11	03	2
26	SAO JOSE DOS CAMPOS	14012017	38	3175	2	08	34	4	01	1	2	1	08	02	2
27	SAO JOSE DOS CAMPOS	25012017	36	2835	1	12	20	1	00	1	1	1	06	99	2
28	SAO JOSE DOS CAMPOS	27012017	36	3060	1	12	38	2	00	1	1	1	08	03	2
29	SAO JOSE DOS CAMPOS	29012017	41	3600	2	08	38	2	00	1	2	2	08	02	2
30	SAO JOSE DOS CAMPOS	02022017	41	3325	2	08	26	2	01	1	2	2	12	03	2
31	SAO JOSE DOS CAMPOS	07022017	39	3600	1	08	28	2		1	1	1	06	06	2
32	SAO JOSE DOS CAMPOS	10022017	37	3775	2	05	35	2	00	1	2	1	08	03	2
33	SAO JOSE DOS CAMPOS	11022017	41	2860	2	08	33	9	00	1	2	2	03	03	2
34	SAO JOSE DOS CAMPOS	11022017	36	2890	2	06	25	1	00	1	2	2	06	02	2
35	SAO JOSE DOS CAMPOS	15022017	39	3460	1	08	35	2	00	1	1	2	10	02	2
36	SAO JOSE DOS CAMPOS	22022017	38	3220	1	08	24	1	00	1	1	1	11	01	2

Fonte: Elaborado pelo autor (2021).

Foram obtidas informações de todos os nascidos vivos de gestantes residentes nos municípios de São José dos Campos (código IBGE 354990) entre 01/01/2016 e 31/12/2018. E excluídos da análise os recém-nascidos abaixo de 22 semanas completas de gestação e os com peso menor de 500 gramas, por serem considerados abortamento, os partos gemelares e tri gemelares, bem como os fetos com má-formação congênita. Também foram excluídos os partos ocorridos por cirurgia cesariana, para descartar possíveis iatrogenias.

Para esse estudo, os campos de dados referentes a idade da mãe, escolaridade, estado civil, quantidade de consultas pré-natal e semanas de gestação foram mantidos para refinamento da pesquisa, os demais campos foram ocultados conforme auxílio de um especialista e consulta em artigos existentes. A partir dos dados consolidados foi possível a criação de uma tabela final para análise das informações.

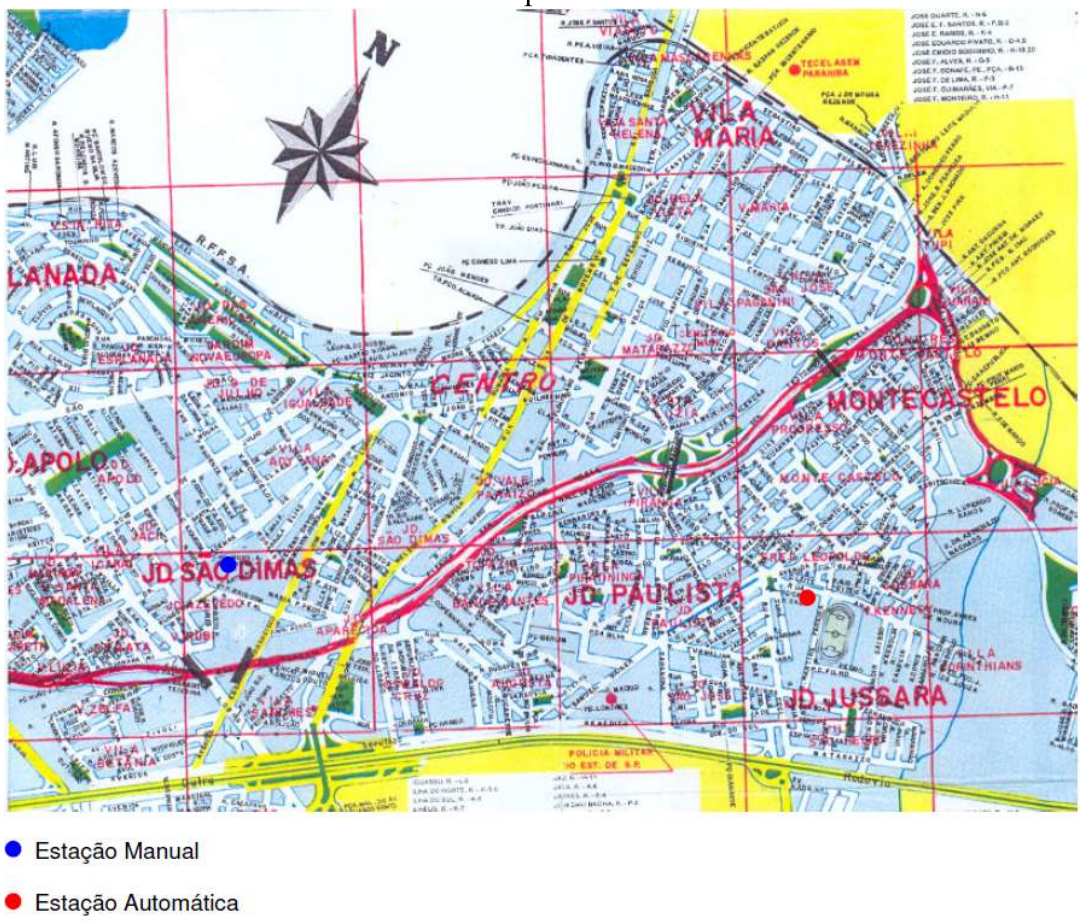
3.3 RELATÓRIOS DE POLUENTES DO AR

A Companhia Ambiental do Estado de São Paulo – CETESB é a agência do Governo do Estado responsável pelo controle, fiscalização, monitoramento e licenciamento de atividades geradoras de poluição, com a preocupação fundamental de preservar e recuperar a qualidade das águas, do ar e do solo (CETESB, 2023).

No portal CETESB, consultado por meio do QUALAR (Qualidade do Ar), é possível visualizar a qualidade do ar em tempo real, localização das estações, relatórios diários de poluentes como PM_{10} e NO_2.

O município de São José dos Campos conta com duas estações de medição para os poluentes em estudo, localizadas conforme Figura 11. Para fins de análise foi utilizada a média dos valores das duas estações da cidade.

Figura 11 - Localização das estações de medição de poluentes de 2021 em São José dos Campos-SP



Fonte: CETESB (2021).

No portal, na aba de relatórios de valores diários, foi possível selecionar o período de análise e o poluente. O *template* da extração dos dados capturados é demonstrado na Figura 12. Inicialmente em versão pdf, os dados foram posteriormente convertidos para xls afim de compilar todas as informações coletadas em uma base única.

Figura 12 – Modelo de relatório diário de NO_2 de 2016 pelo portal CETESB

CETESB

Sistema de Informações de Qualidade do Ar

RELATÓRIO DE VALORES DIÁRIOS

Tipo de Rede: Automático Data Inicial: 01/09/2016 Data Final: 31/12/2017

Realizar Representatividade: N

Utilizar maior média móvel do dia: N

UGRHI: 2 - PARAIBA DO SUL
Estação: S.José Campos
Unidade de µg/m3
PQAR: 260,0
Parâmetro: 15 - NO2 (Dióxido de Nitrogênio)

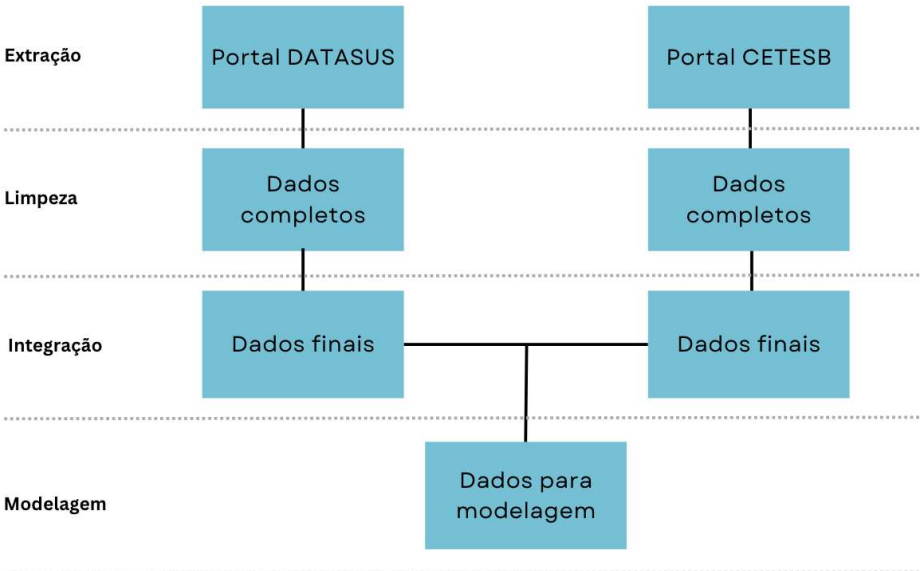
Data	Valor Diário	Horário de Ocorrência	Número de Amostras	Qualidade do Ar
01/09/2016	47	09:00	23	N1 - BOA
02/09/2016	34	07:00	23	N1 - BOA
03/09/2016	27	07:00	23	N1 - BOA
04/09/2016	40	19:00	23	N1 - BOA
05/09/2016	34	11:00	23	N1 - BOA
06/09/2016	63	22:00	23	N1 - BOA
07/09/2016	38	21:00	23	N1 - BOA
08/09/2016	36	07:00	23	N1 - BOA
09/09/2016	26	08:00	23	N1 - BOA
10/09/2016	29	24:00	23	N1 - BOA
11/09/2016	29	02:00	23	N1 - BOA
12/09/2016	71	21:00	23	N1 - BOA
13/09/2016	89	20:00	23	N1 - BOA
14/09/2016	68	09:00	23	N1 - BOA

Fonte: CETESB (2021).

3.4 MODELO ELABORADO POR *FUZZY*

Após o tratamento dos dados dos nascidos vivos e poluentes, provenientes dos portais DATASUS e CETESB, conforme esquema apresentado na Figura 13, é possível iniciar a etapa da modelagem do sistema *fuzzy*.

Figura 13- Etapas de tratamento de dados DATASUS e CETESB



Fonte: Elaborado pelo autor (2023).

O software utilizado para a aplicação do sistema *fuzzy* foi o MATLAB R2022b®, acrescido da ferramenta *Fuzzy Logic Toolbox*, com aplicação do sistema de inferência Mamdani.

As variáveis de entrada e saída foram selecionadas a partir das informações extraídas e compiladas, tanto pelo portal DATASUS, quanto pelo CETESB, descritos anteriormente. Para a criação do primeiro modelo, as variáveis de entrada escolhidas foram: idade da mãe, número de consultas pré-natal, concentração de NO_2 e contração de PM_{10}. Retornando com uma única variável de saída, a prematuridade.

3.4.1 Idade da mãe

A idade da mãe é classificada como baixa, média e alta, conforme parâmetros apresentados na Tabela 2. No período de estudo, a idade mínima encontrada foi de 13 anos, enquanto a máximo foi de 45 anos.

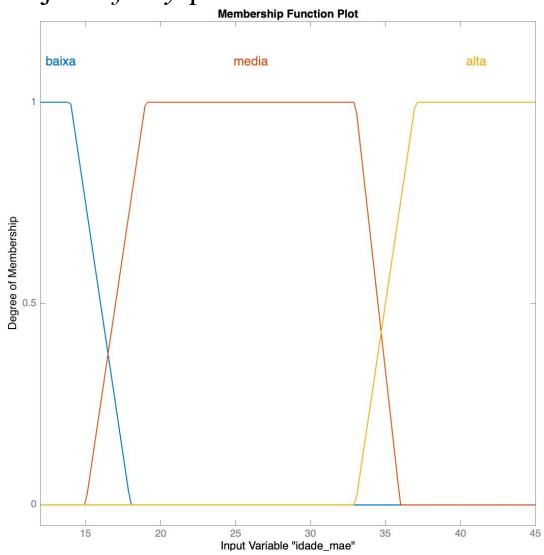
Tabela 2 - Parâmetros do sistema *fuzzy* para classificação da idade da mãe

Classificação	Parâmetros
baixa	[-10 0 14 18]
média	[15 19 33 36]
alta	[33 37 45 50]

Fonte: Elaborado pelo autor (2023).

Para essa variável de entrada, é utilizada a função de pertinência do tipo trapezoidal, demonstrado na Figura 14.

Figura 14 - Conjunto *fuzzy* para a variável de entrada: idade da mãe



Fonte: Elaborado pelo autor (2023).

3.4.2 Número de consultas pré-natal

Conforme previsto pelo SUS, seis consultas pré-natal é a quantidade ideal mínima para o acompanhamento da gestação. No modelo proposto, a quantidade de consultas é classificada como inadequada, regular e adequada. O valor máximo de consultas no período do estudo é de quinze durante toda a gravidez, enquanto aproximadamente 4% das gestantes não realizaram nenhuma consulta pré-natal.

Os parâmetros foram classificados conforme Tabela 3.

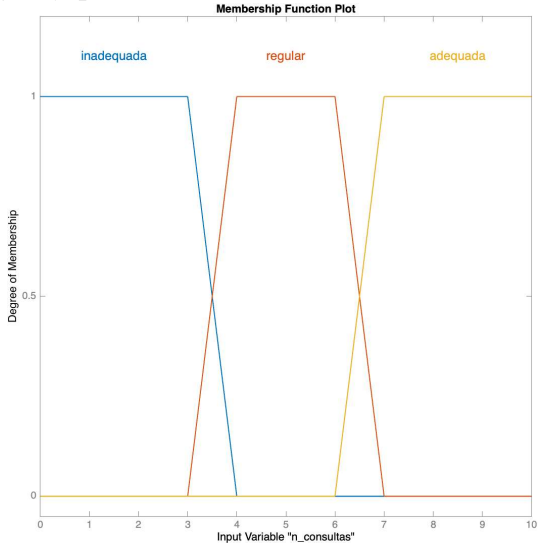
Tabela 3 - Parâmetros do sistema *fuzzy* para classificação do número de consultas pré-natal

Classificação	Parâmetros
inadequada	[-3 0 3 3 4]
regular	[3 4 6 7]
adequada	[6 7 10 15]

Fonte: Elaborado pelo autor (2023).

Há utilização de função de pertinência do tipo trapezoidal para essa variável de entrada, conforme ilustrado na Figura 15.

Figura 15 - Conjunto *fuzzy* para a variável de entrada: número de consultas pré-natal



Fonte: Elaborado pelo autor (2023).

3.4.3 Concentração de poluentes NO_2 e PM_{10}

Como base para a concentração de poluentes, é utilizado o método do primeiro quartil para estabelecer os limites de concentração. Para o NO_2, taxas de até 38 $\mu g/m^3$ são

consideradas aceitáveis, e superiores a esse valor são consideradas inaceitáveis. Já para a concentração de PM_{10}, o limite aceitável estabelecido é de 22 $\mu g/m^3$, sendo considerados inaceitáveis todos os valores superiores. Ajustes foram realizados nos valores para que o modelo obtivesse melhor resultado. Os parâmetros para NO_2 e PM_{10} são apresentados na Tabela 4 e 5, respectivamente.

Tabela 4 - Parâmetros do sistema *fuzzy* para classificação Dióxido de Nitrogênio

Classificação	**Parâmetros**
aceitável	[-36 0 36 39]
inaceitável	[36 40 50 60]

Fonte: Elaborado pelo autor (2023).

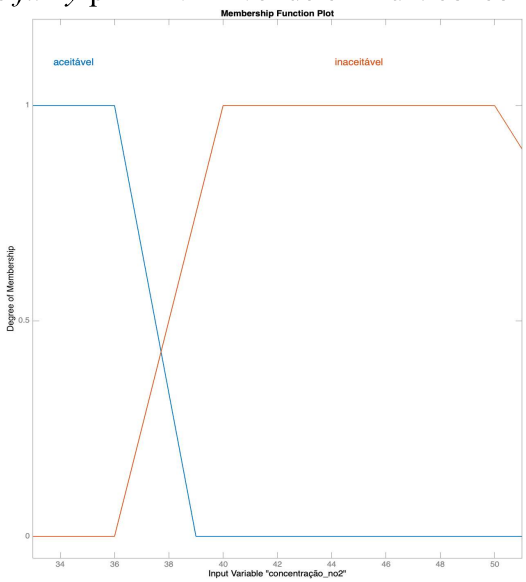
Tabela 5 - Parâmetros do sistema *fuzzy* para classificação Material Particulado

Classificação	**Parâmetros**
aceitável	[-20 0 20 24]
inaceitável	[20 24 30 40]

Fonte: Elaborado pelo autor (2023).

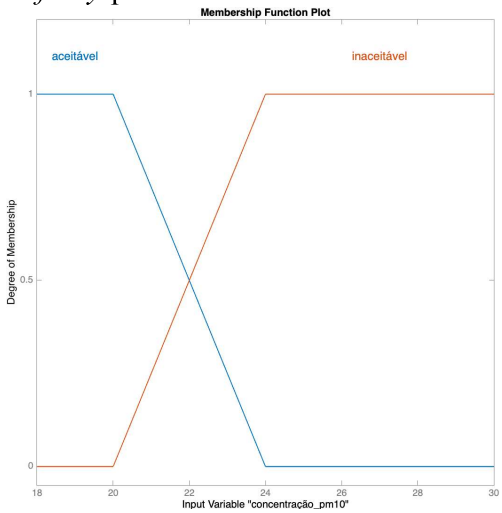
Para ambas as concentrações de poluentes, a função de pertinência do tipo trapezoidal foi escolhida, conforme Figura 16 para NO_2 e Figura 17 para PM_{10}.

Figura 16 - Conjunto *fuzzy* para a variável de entrada: concentração de NO_2



Fonte: Elaborado pelo autor (2023).

Figura 17 - Conjunto *fuzzy* para a variável de entrada: concentração de PM_{10}



Fonte: Elaborado pelo autor (2023).

3.4.4 Prematuridade

O estudo realizado por Dias *et al.* (2022) considera como prematuro precoce, todos os nascidos com idade gestacional menor ou igual a 33 semanas e prematuro tardio aqueles nascidos entre 34 e 36 semanas.

Nesse estudo, a variável de saída, prematuridade, é dividida entre quatro classificações, de acordo com as semanas de gestação: Prematuro extremo, muito prematuro, prematuro moderado, e prematuro tardio, os parâmetros são apresentados na Tabela 6.

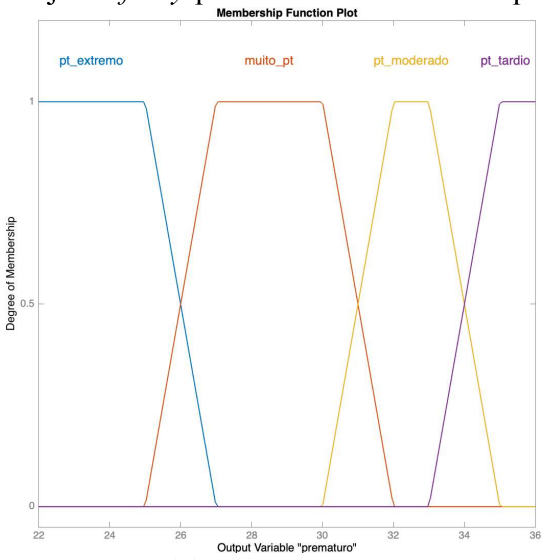
Tabela 6 - Parâmetros do sistema *fuzzy* para classificação da prematuridade

Classificação	Parâmetros
pt extremo	[-10 0 25 27]
muito pt	[25 27 30 32]
pt moderado	[30 32 33 35]
pt tardio	[33 34 36 40]

Fonte: Elaborado pelo autor (2023).

Para a variável de saída, é utilizada a função de pertinência do tipo trapezoidal, demonstrada na Figura 18.

Figura 18 - Conjunto *fuzzy* para a variável de saída: prematuridade



Fonte: Elaborado pelo autor (2023).

Após estabelecer as classificações das variáveis de entrada e saída, foi realizada a criação da base de regras manualmente, validada com um especialista, abrangendo todas as possibilidades de combinações e totalizando 36 regras, conforme esquema apresentado na Tabela 7.

Tendo como exemplo da Regra 36, uma mãe com idade alta, realizou um número de consultas pré-natal considerado regular, exposta a níveis aceitáveis de PM_{10} e NO_2 durante toda a sua gravidez, tem a tendência de ter um prematuro tardio.

Tabela 7 - Base de regras inicial

(continua)

	SE				ENTÃO
N DE REGRAS	**IDADE DA MÃE**	**CONSULTA PRÉ-NATAL**	**CONC. NO2**	**CONC. PM10**	**PREMATURIDADE**
REGRA 1	baixa	inadequada	inaceitável	inaceitável	tardio
REGRA 2	baixa	inadequada	inaceitável	aceitável	moderado
REGRA 3	baixa	inadequada	aceitável	inaceitável	moderado
REGRA 4	baixa	inadequada	aceitável	aceitável	moderado
REGRA 5	baixa	adequada	inaceitável	inaceitável	tardio
REGRA 6	baixa	adequada	inaceitável	aceitável	tardio
REGRA 7	baixa	adequada	aceitável	inaceitável	tardio
REGRA 8	baixa	adequada	aceitável	aceitável	tardio
REGRA 9	baixa	regular	inaceitável	inaceitável	moderado
REGRA 10	baixa	regular	inaceitável	aceitável	muito prematuro
REGRA 11	baixa	regular	aceitável	inaceitável	muito prematuro
REGRA 12	baixa	regular	aceitável	aceitável	tardio
REGRA 13	média	inadequada	inaceitável	inaceitável	tardio
REGRA 14	média	inadequada	inaceitável	aceitável	extremo
REGRA 15	média	inadequada	aceitável	inaceitável	muito prematuro

Tabela 7 – Base de regras inicial

(conclusão)

REGRA 16	média	inadequada	aceitável	aceitável	Extremo
REGRA 17	média	adequada	inaceitável	inaceitável	tardio
REGRA 18	média	adequada	inaceitável	aceitável	tardio
REGRA 19	média	adequada	aceitável	inaceitável	tardio
REGRA 20	média	adequada	aceitável	aceitável	tardio
REGRA 21	média	regular	inaceitável	inaceitável	tardio
REGRA 22	média	regular	inaceitável	aceitável	tardio
REGRA 23	média	regular	aceitável	inaceitável	tardio
REGRA 24	média	regular	aceitável	aceitável	tardio
REGRA 25	alta	inadequada	inaceitável	inaceitável	tardio
REGRA 26	alta	inadequada	inaceitável	aceitável	tardio
REGRA 27	alta	inadequada	aceitável	inaceitável	tardio
REGRA 28	alta	inadequada	aceitável	aceitável	tardio
REGRA 29	alta	adequada	inaceitável	inaceitável	tardio
REGRA 30	alta	adequada	inaceitável	aceitável	tardio
REGRA 31	alta	adequada	aceitável	inaceitável	tardio
REGRA 32	alta	adequada	aceitável	aceitável	tardio
REGRA 33	alta	regular	inaceitável	inaceitável	tardio
REGRA 34	alta	regular	inaceitável	aceitável	tardio
REGRA 35	alta	regular	aceitável	inaceitável	tardio
REGRA 36	alta	regular	aceitável	aceitável	tardio

Fonte: Elaborado pelo autor (2022).

No sistema, a base de regras se deu no formato 3x3x2x2, podendo ser explicada da seguinte maneira: 3 funções para a entrada idade da mãe (baixa, média e alta), 3 para número de consultas pré-natal (inadequada, regular e adequada), 2 para concentração de NO_2 (inaceitável e aceitável) e 2 para a concentração de PM_{10} (inaceitável e aceitável). A Figura 19 ilustra o Editor de Regras do MATLAB R2022b® para as primeiras dezesseis regras do sistema.

Figura 19 - Editor de Regras do sistema *fuzzy* no MATLAB

	Rule	Weight	Name
1	If idade_mae is baixa and n_consultas is inadequada and concentração_no2 is inac…	1	rule1
2	If idade_mae is baixa and n_consultas is inadequada and concentração_no2 is inac…	1	rule2
3	If idade_mae is baixa and n_consultas is inadequada and concentração_no2 is aceit…	1	rule3
4	If idade_mae is baixa and n_consultas is inadequada and concentração_no2 is aceit…	1	rule4
5	If idade_mae is baixa and n_consultas is adequada and concentração_no2 is inaceit…	1	rule5
6	If idade_mae is baixa and n_consultas is adequada and concentração_no2 is inaceit…	1	rule6
7	If idade_mae is baixa and n_consultas is adequada and concentração_no2 is aceitá…	1	rule7
8	If idade_mae is baixa and n_consultas is adequada and concentração_no2 is aceitá…	1	rule8
9	If idade_mae is baixa and n_consultas is regular and concentração_no2 is inaceitáv…	1	rule9
10	If idade_mae is baixa and n_consultas is regular and concentração_no2 is inaceitáv…	1	rule10
11	If idade_mae is baixa and n_consultas is regular and concentração_no2 is aceitável …	1	rule11
12	If idade_mae is baixa and n_consultas is regular and concentração_no2 is aceitável …	1	rule12
13	If idade_mae is media and n_consultas is inadequada and concentração_no2 is inac…	1	rule13
14	If idade_mae is media and n_consultas is inadequada and concentração_no2 is inac…	1	rule14
15	If idade_mae is media and n_consultas is inadequada and concentração_no2 is acei…	1	rule15
16	If idade_mae is media and n_consultas is inadequada and concentração_no2 is acei…	1	rule16

Fonte: Elaborado pelo autor (2023).

Uma forma de validar o sistema é por meio da rotina “ddeinit” no MATLAB R2022b®, realizando os comandos representados a seguir, a partir de um sistema operacional macOS:

```
>> mymatrix = xlsread('V1.xlsx', 1, 'A2:D1097')
>> fismat= readfis ('teste1')
>> output=evalfis(fismat,mymatrix)
```

Onde ‘V1’ é o nome da planilha padrão dos dados, ‘A2:D1097’ é o intervalo de dados e ‘teste1’ o nome do arquivo *fis* contendo o sistema realizado.

3.5 MODELO ELABORADO POR NEURO-*FUZZY* ANFIS

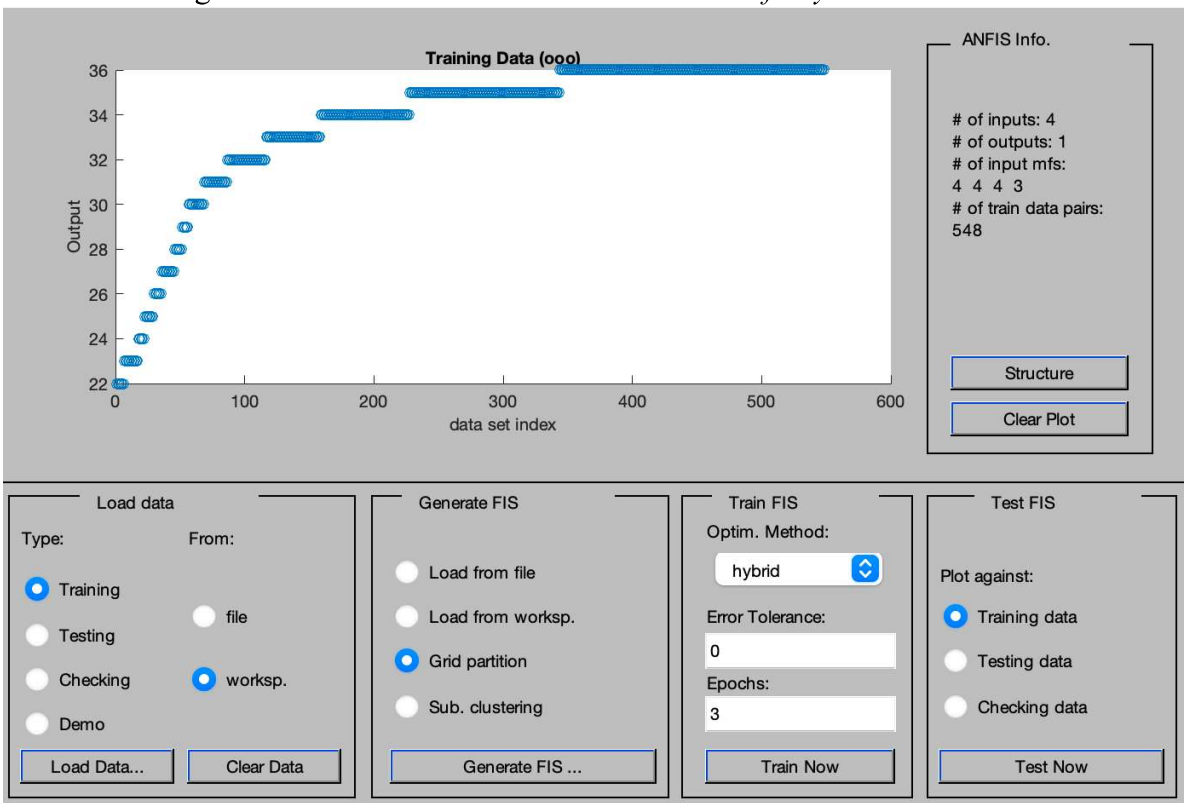
Para a comparação de resultados, a construção de um modelo utilizando, dessa vez, a técnica neuro-*fuzzy* é realizada. Igualmente ao modelo anterior, foi utilizado o software MATLAB R2022b®, para a modelagem do sistema, porém com o acréscimo do *Toolbox* ANFIS, delimitando a lógica do neuro-*fuzzy*, a partir do comando *anfisedit.*

No sistema neuro-*fuzzy* é possível adicionar os dados de entrada, o número de funções para cada variável e o próprio sistema retornará com os parâmetros correspondentes, tornando essa outra característica que difere o método *fuzzy* ao neuro-*fuzzy* na hora da modelagem. Para meios de comparação, foram utilizadas as mesmas variáveis em relação ao modelo *fuzzy* realizado anteriormente: idade da mãe, número de consultas pré-natal, concentração de NO_2 e contração de PM_{10}, como entradas e a prematuridade como saída do sistema.

Esse sistema utiliza o método de inferência Takagi-Sugeno, diferentemente do *fuzzy*, no qual foi utilizado o método de inferência Mamdani.

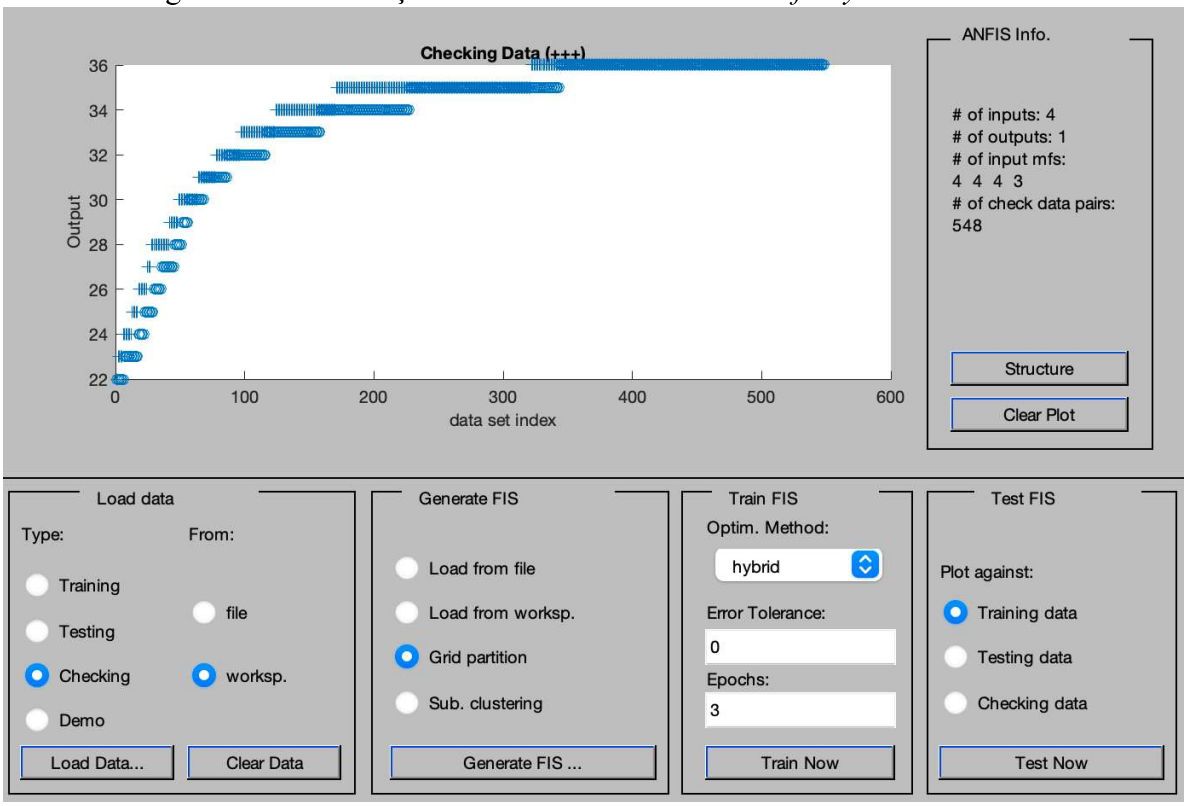
As Figuras 20 e 21 representam, no MATLAB, os dados de entrada para o modelo neuro-*fuzzy*, nesse exemplo, divididos em duas partes, 50% dos dados para treino (Figura 18) e 50% para verificação (Figura 20). Vale ressaltar que os dados foram selecionados de forma coerente para que não gerassem intervenções ao sistema final.

Figura 20 – Treino de dados do sistema neuro-*fuzzy* no MATLAB



Fonte: MATLAB (2023).

Figura 21 - Verificação de dados do sistema neuro-*fuzzy* no MATLAB



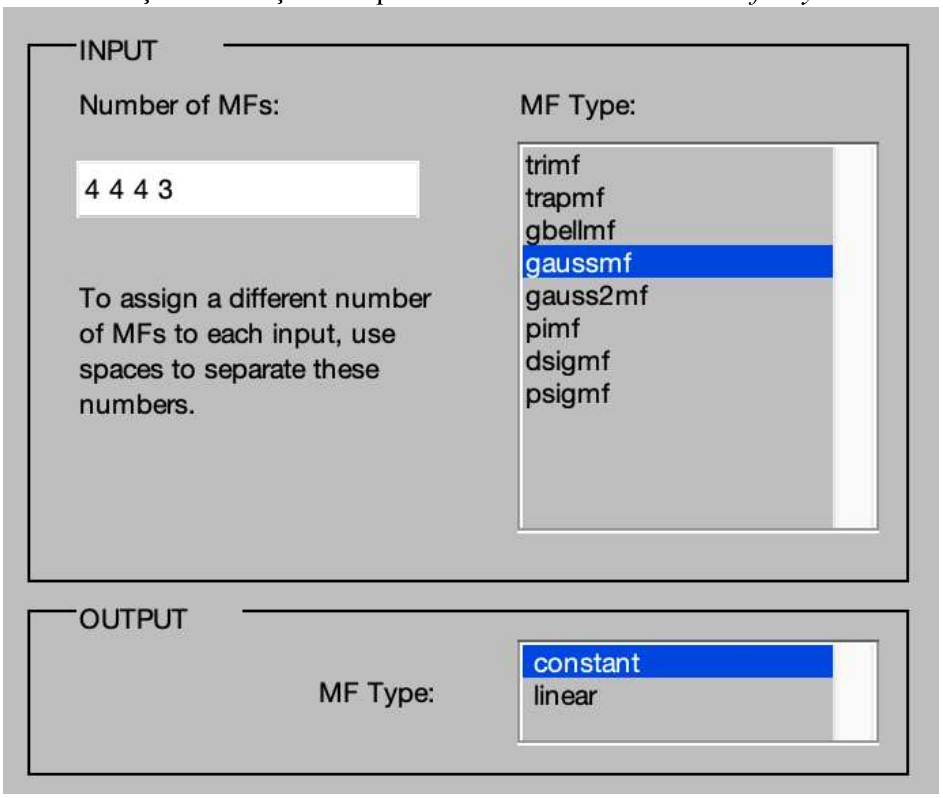
Fonte: MATLAB (2023).

A partir do “Generate FIS”, é possível adicionar o número de funções para cada uma das variáveis de entrada, assim como o tipo, triangular, trapezoidal, gaussiana ou outras, podendo também a variável de saída ser escolhida como constante ou linear, conforme exemplo apresentado na Figura 22. Além disso, é possível selecionar a geração do modelo *.fis* inicial entre as técnicas de *grid partition* ou *subtractive clustering.*

Após essa etapa, o método de otimização é definido, podendo ser *backpropagation* ou *hybrid*. Assim como a tolerância de erro e quantidade de épocas, conforme exemplo demonstrado na Figura 23.

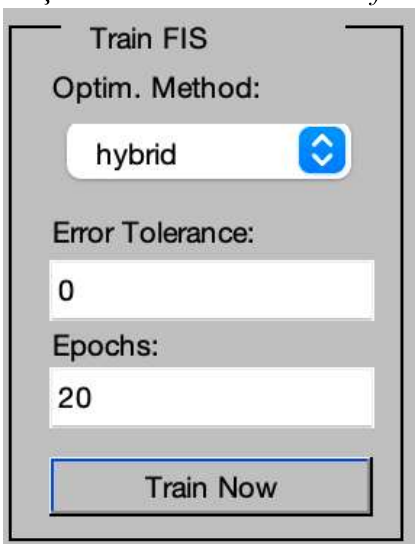
Com todos os parâmetros ajustados, o treino do sistema é realizado, representado na Figura 24.

Figura 22 – Adição de funções de pertinência no sistema neuro-*fuzzy* do MATLAB



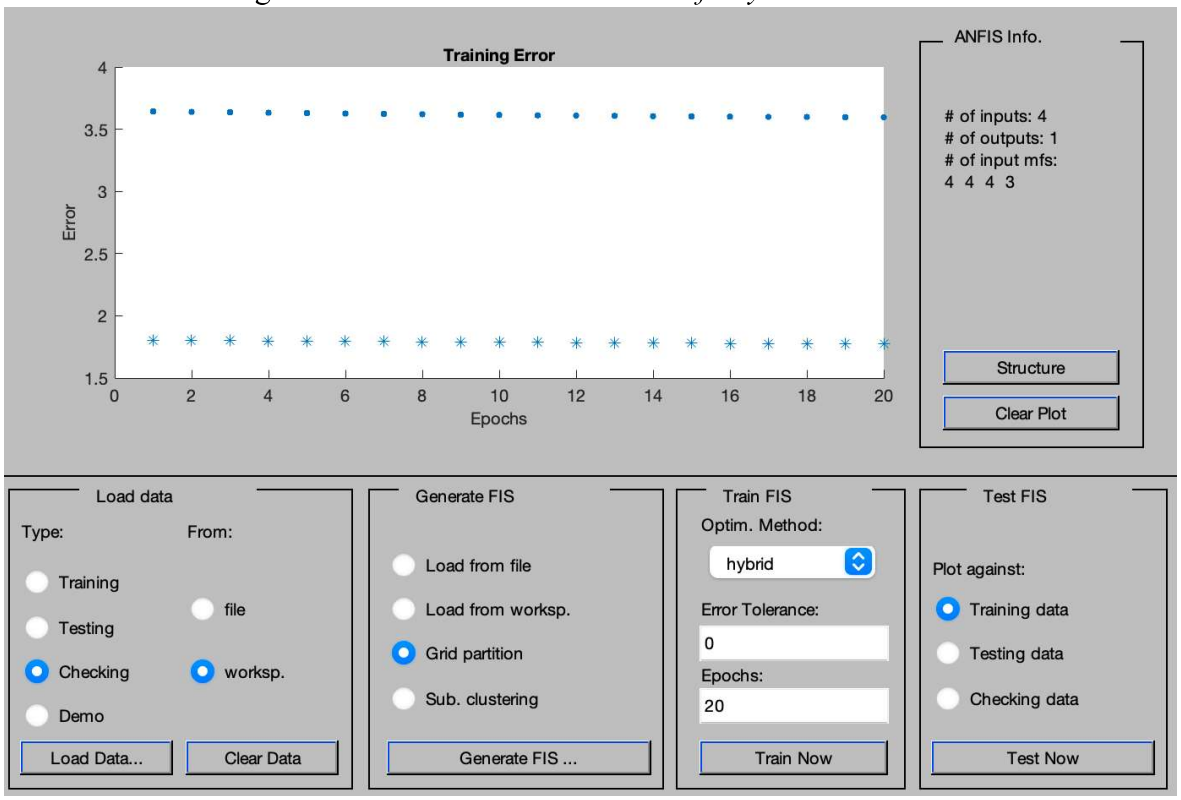
Fonte: MATLAB (2023).

Figura 23 - Treino da função FIS no sistema neuro-*fuzzy* do MATLAB



Fonte: MATLAB (2023).

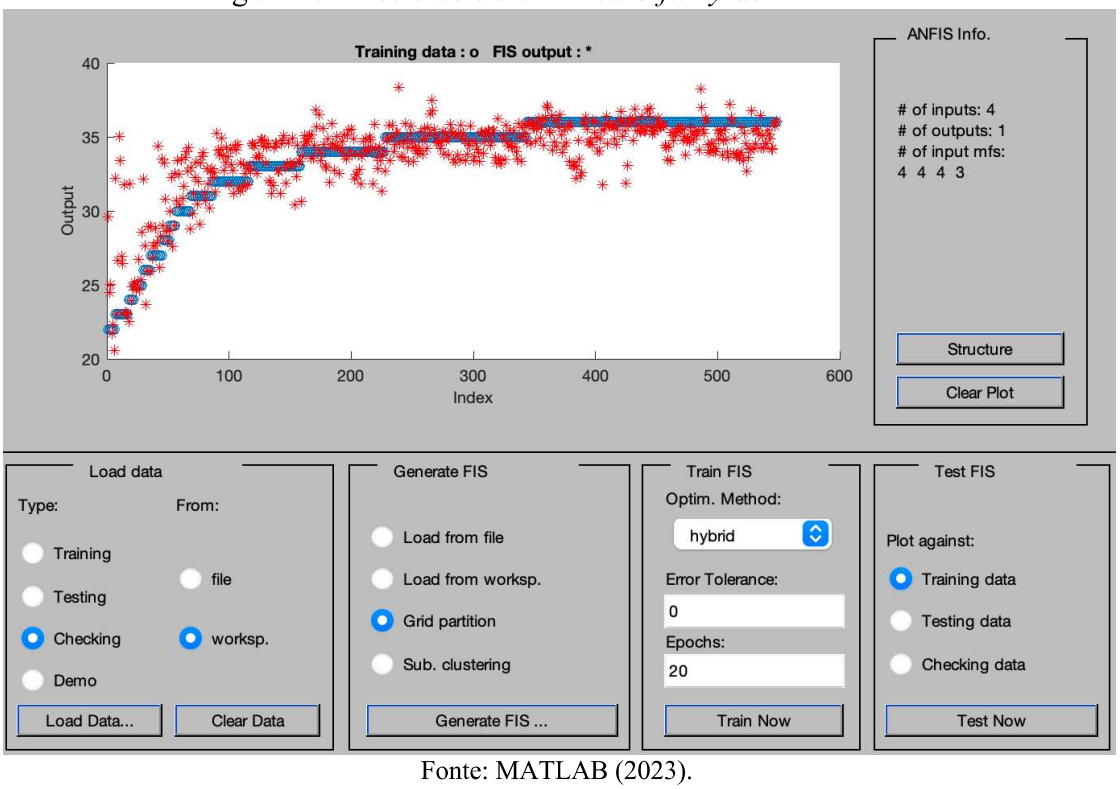
Figura 24 - Treino do sistema neuro-*fuzzy* no MATLAB



Fonte: MATLAB (2023).

Após o treino, o teste será realizado sobre os dados do sistema, representado pela Figura 25.

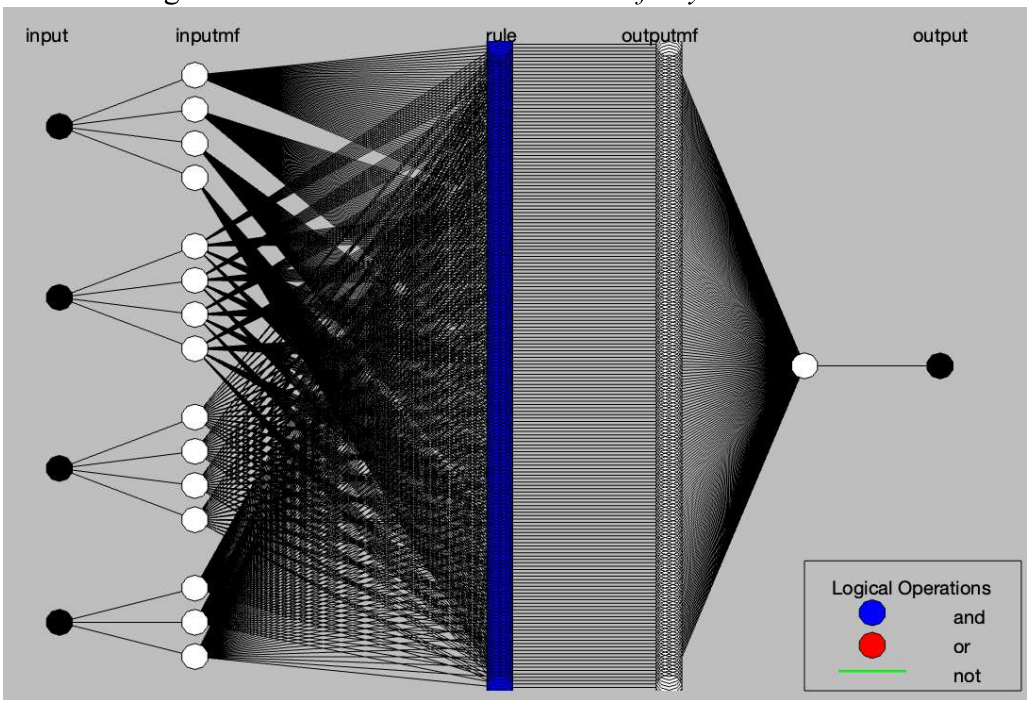
Figura 25 - Teste do sistema neuro-*fuzzy* do MATLAB



Fonte: MATLAB (2023).

Com o sistema neuro-*fuzzy* é possível visualizar a estrutura do modelo, exemplificada na Figura 26.

Figura 26 - Estrutura do sistema neuro-*fuzzy* do MATLAB



Fonte: MATLAB (2023).

No modelo utilizando o neuro-*fuzzy*, todas as variáveis obtiveram parâmetros estabelecidos pelo próprio sistema, a partir dos dados de treino e verificação disponibilizados.

3.5.1 Idade da mãe

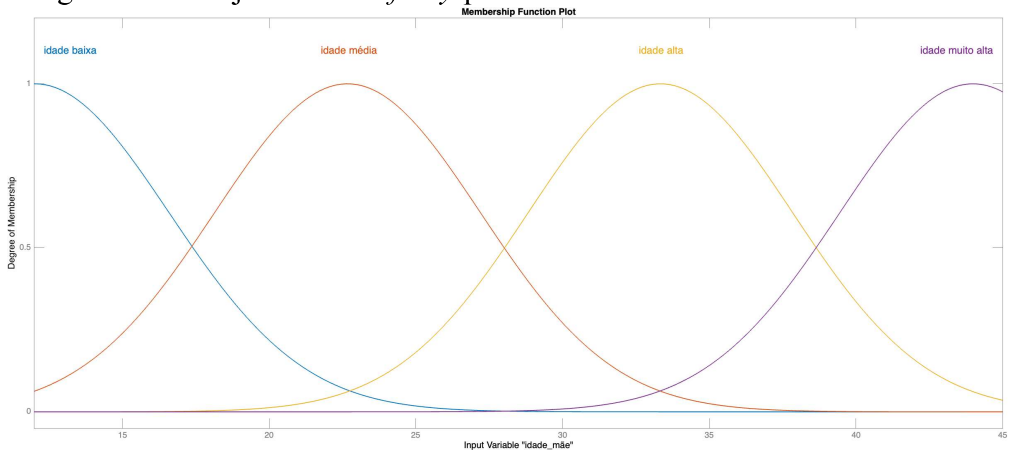
Para a idade da mãe, os parâmetros são classificados pelo sistema e dadas as denominações idade baixa, média, alta e muito alta. Cada linha, conforme Tabela 8, representa uma classificação realizada pelo sistema e seus parâmetros correspondentes.
A função de pertinência para essa entrada é representada na Figura 27.

Tabela 8 - Parâmetros do sistema neuro-*fuzzy* para classificação da idade da mãe

Classificação	Parâmetros
Idade baixa	[4.5734 12.0236]
Idade média	[4.53528 22.6738]
Idade alta	[4.51123 33.3379]
Idade muito alta	[4.54194 43.99]

Fonte: Elaborado pelo autor (2023).

Figura 27 - Conjunto neuro-*fuzzy* para a variável de entrada: idade da mãe



Fonte: Elaborado pelo autor (2023).

3.5.2 Número de consultas pré-natal

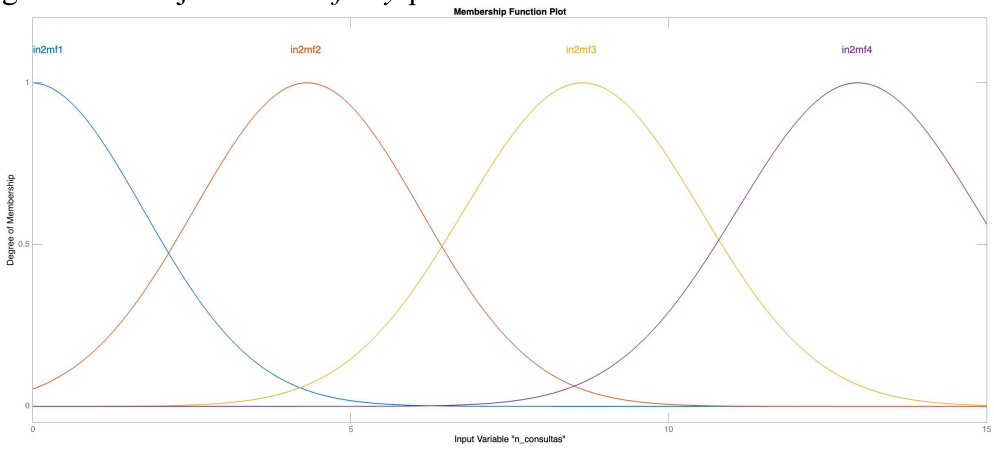
Da mesma forma realizada com a primeira variável de entrada, a segunda variável, número de consultas, também teve seus parâmetros e classificações implementados pelo sistema, apresentado na Tabela 9. Suas funções de pertinência foram estabelecidas conforme Figura 28.

Tabela 9 - Parâmetros do sistema neuro-*fuzzy* para classificação do número de consulta pré-natal

Classificação	Parâmetros
Inadequada	[1.74993 -0.0376237]
Regular	[1.7257 4.27426]
Adequada	[1.87331 8.62223]
Extrapolada	[1.92148 12.9622]

Fonte: Elaborado pelo autor (2023).

Figura 28 - Conjunto neuro-*fuzzy* para a variável de entrada: número de consultas



Fonte: Elaborado pelo autor (2023).

3.5.3 Concentração de poluentes NO_2 e PM_{10}

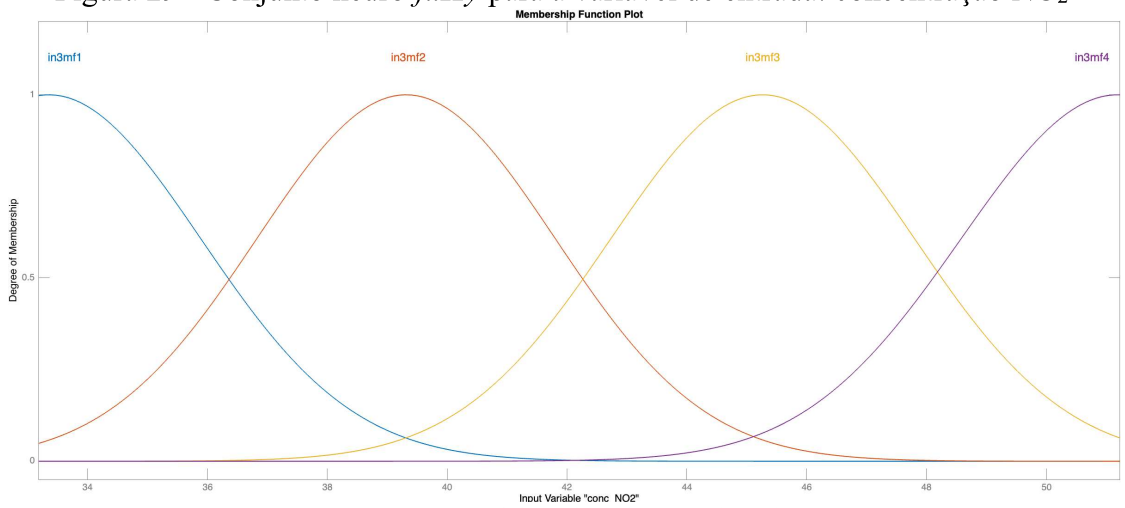
Os poluentes também tiveram denominações e parâmetros estabelecidos pelo sistema neuro-*fuzzy*. As classificações e parâmetros do NO_2 são representadas na Tabela 10, com suas funções de pertinência ilustradas na Figura 29. Já os parâmetros para o PM_{10} foram descritos conforme Tabela 11, com funções de pertinência apresentadas na Figura 30.

Tabela 10 – Parâmetros do sistema neuro-*fuzzy* para classificação da concentração de NO_2

Classificação	Parâmetros
baixo risco	[2.53771 33.3539]
aceitável	[2.49449 39.3139]
risco	[2.53756 45.2658]
alto risco	[2.61562 51.1897]

Fonte: Elaborado pelo autor (2023).

Figura 29 - Conjunto neuro-*fuzzy* para a variável de entrada: concentração NO_2



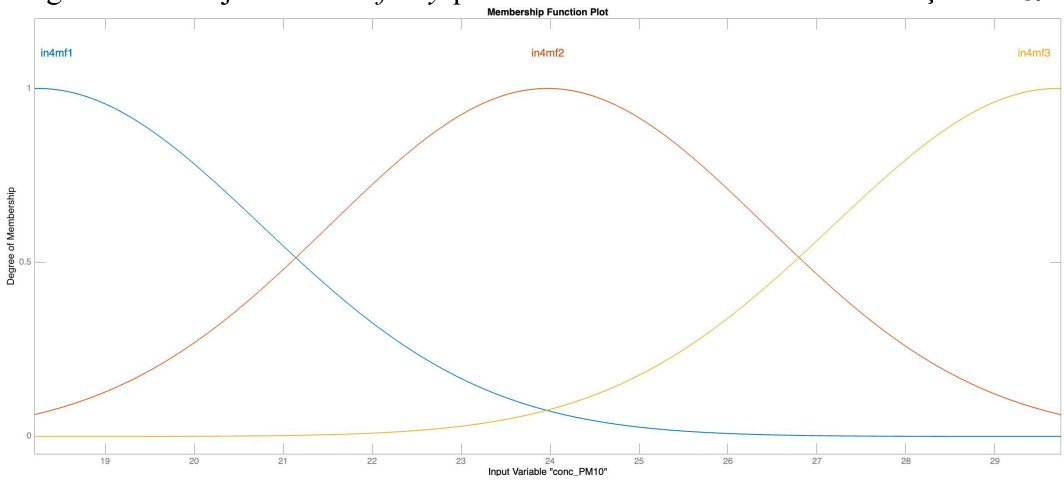
Fonte: Elaborado pelo autor (2023).

Tabela 11 - Parâmetros do sistema neuro-*fuzzy* para classificação da concentração de PM_{10}

Classificação	Parâmetros
baixo risco	[2.51273 18.2429]
aceitável	[2.45095 23.9717]
risco	[2.52553 29.7087]

Fonte: Elaborado pelo autor (2023).

Figura 30 - Conjunto neuro-*fuzzy* para a variável de entrada: concentração PM_{10}



Fonte: Elaborado pelo autor (2023).

3.5.4 Prematuridade

Para a variável de saída, foi escolhido o tipo constante e parâmetros criados pelo próprio sistema. A Figura 31 ilustra as doze primeiras saídas do sistema, assim como os parâmetros de cada variável.

Figura 31 - Parâmetros para variável de saída "prematuridade"

Name	Type	Parameters
out1mf1	Constant	-29.5403
out1mf2	Constant	-90.7523
out1mf3	Constant	-315.0742
out1mf4	Constant	165.7715
out1mf5	Constant	31.3927
out1mf6	Constant	-373.1467
out1mf7	Constant	-482.8886
out1mf8	Constant	56.3598
out1mf9	Constant	169.9063
out1mf10	Constant	1803.1758
out1mf11	Constant	-133.7553
out1mf12	Constant	-415.7707

Fonte: Elaborado pelo autor (2023).

As regras são representadas automaticamente pelo sistema, através do editor de regras. As dezenove primeiras regras do sistema são representadas na Figura 32. Para todas as regras foi adotado peso 1.

A primeira regra, por exemplo, demonstra que se a mãe tiver a idade baixa, com número de consultas inadequado, exposta a níveis de baixo risco de PM_{10} e NO_2, então a saída será um prematuro constante 1.

Figura 32 - Editor de Regras do sistema neuro-*fuzzy* no MATLAB

	Rule	Weight	Name
1	If idade_mãe is idade baixa and n_consultas is inadequado and conc_NO2 is baixo risc…	1	rule1
2	If idade_mãe is idade baixa and n_consultas is inadequado and conc_NO2 is baixo risc…	1	rule2
3	If idade_mãe is idade baixa and n_consultas is inadequado and conc_NO2 is baixo risc…	1	rule3
4	If idade_mãe is idade baixa and n_consultas is inadequado and conc_NO2 is aceitável …	1	rule4
5	If idade_mãe is idade baixa and n_consultas is inadequado and conc_NO2 is aceitável …	1	rule5
6	If idade_mãe is idade baixa and n_consultas is inadequado and conc_NO2 is aceitável …	1	rule6
7	If idade_mãe is idade baixa and n_consultas is inadequado and conc_NO2 is risco and …	1	rule7
8	If idade_mãe is idade baixa and n_consultas is inadequado and conc_NO2 is risco and …	1	rule8
9	If idade_mãe is idade baixa and n_consultas is inadequado and conc_NO2 is risco and …	1	rule9
10	If idade_mãe is idade baixa and n_consultas is inadequado and conc_NO2 is alto risco …	1	rule10
11	If idade_mãe is idade baixa and n_consultas is inadequado and conc_NO2 is alto risco …	1	rule11
12	If idade_mãe is idade baixa and n_consultas is inadequado and conc_NO2 is alto risco …	1	rule12
13	If idade_mãe is idade baixa and n_consultas is regular and conc_NO2 is baixo risco an…	1	rule13
14	If idade_mãe is idade baixa and n_consultas is regular and conc_NO2 is baixo risco an…	1	rule14
15	If idade_mãe is idade baixa and n_consultas is regular and conc_NO2 is baixo risco an…	1	rule15
16	If idade_mãe is idade baixa and n_consultas is regular and conc_NO2 is aceitável and …	1	rule16
17	If idade_mãe is idade baixa and n_consultas is regular and conc_NO2 is aceitável and …	1	rule17
18	If idade_mãe is idade baixa and n_consultas is regular and conc_NO2 is aceitável and …	1	rule18
19	If idade_mãe is idade baixa and n_consultas is regular and conc_NO2 is risco and conc…	1	rule19

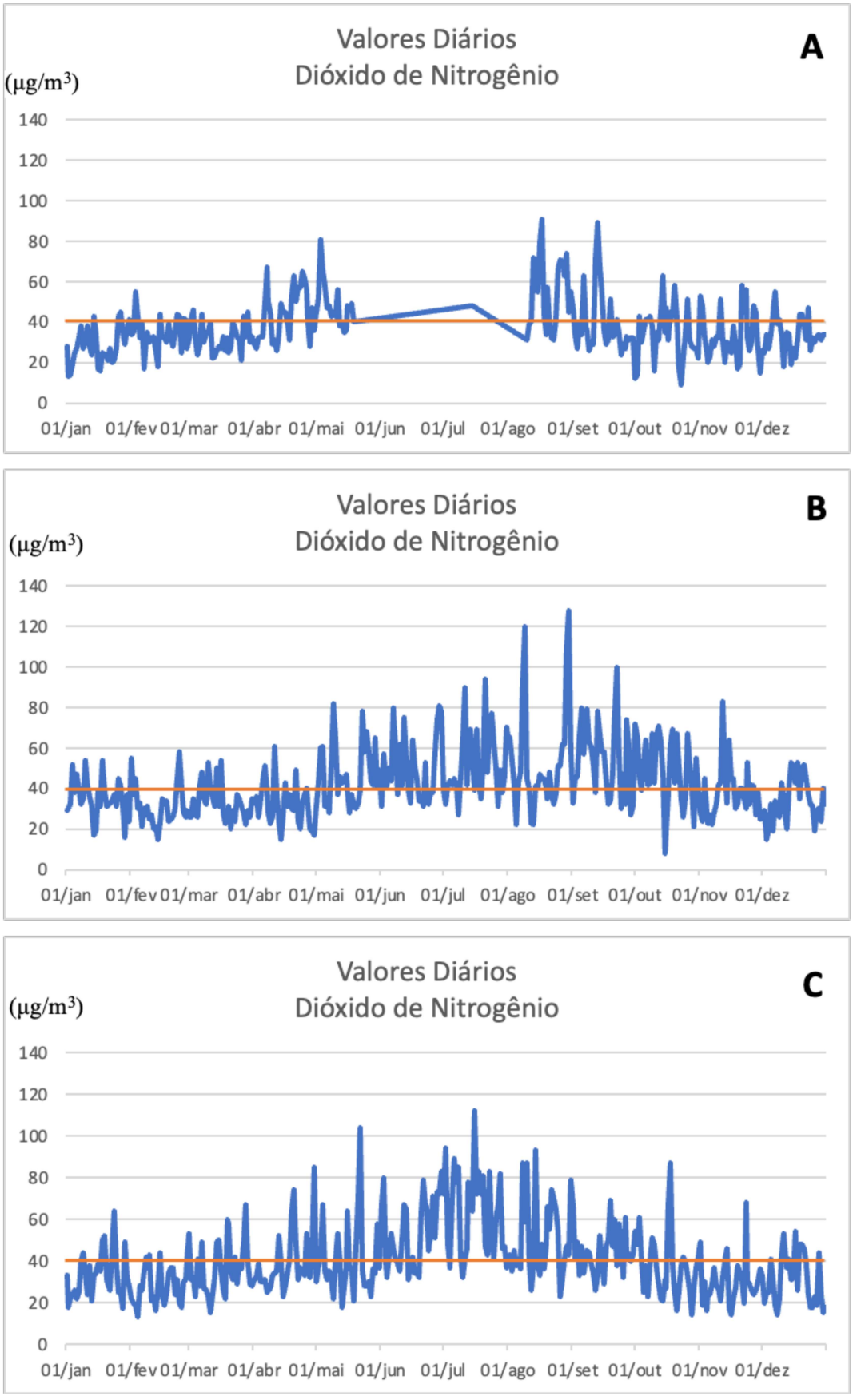
Fonte: Elaborado pelo autor (2023).

4 RESULTADOS E DISCUSSÕES

Com os dados compilados, concentrações de poluentes (PM_{10} e NO_2) e informações de nascidos vivos dos anos de 2016 a 2018, em uma base única, é possível realizar algumas considerações sobre a taxa de prematuridade em relação as informações coletadas.

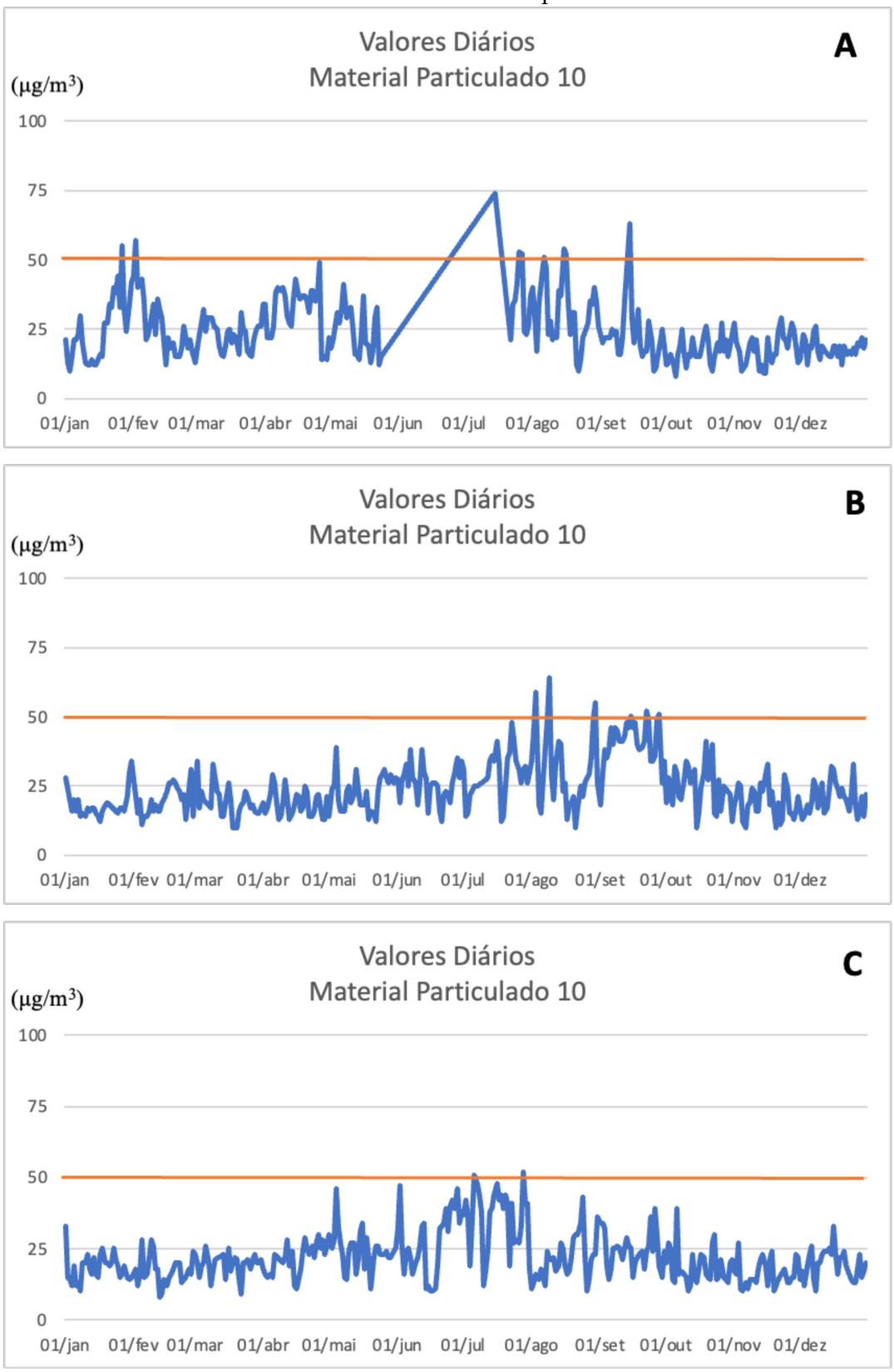
A partir da análise dos dados de poluentes atmosféricos, é possível visualizar os valores diários durante o período de estudo. Nas Figuras 33 e 34 são apresentados os valores reais obtidos por dia para NO_2 e PM_{10}, respectivamente. A linha laranja no gráfico evidencia o limite aceitável de cada poluente, conforme padrão final estabelecido pela QUALAR.

Figura 33 - Valores diários de NO_2 nos anos de (a) 2016 (b) 2017 e (c) 2018 2018 na cidade de São José dos Campos-SP



Fonte: Elaborado pelo autor (2022).

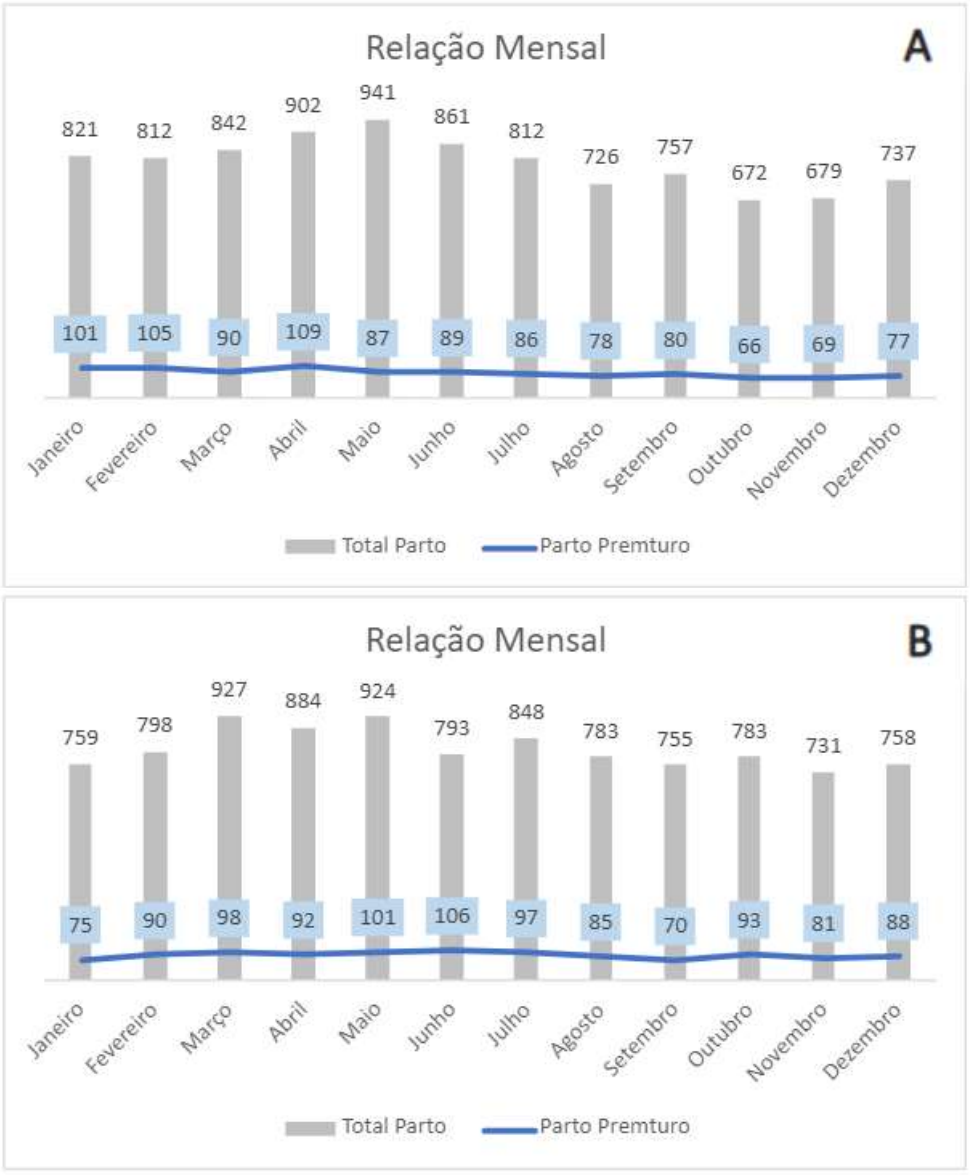
Figura 34 - Valores diários de PM_{10} nos anos de (a) 2016 (b) 2017 e (c) 2018 2018 na cidade de São José dos Campos-SP

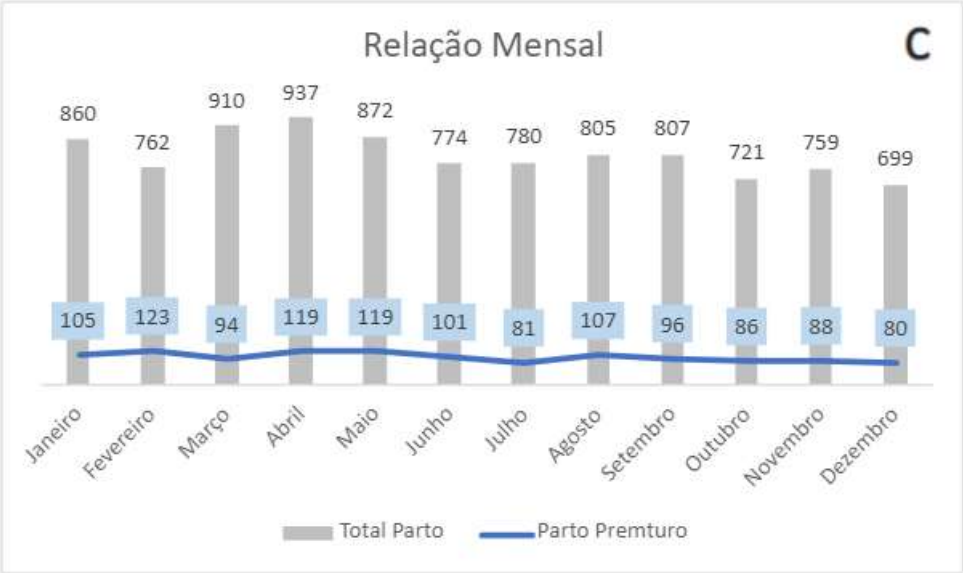


Fonte: Elaborado pelo autor (2022).

Em 2016, a taxa prematuridade de São José dos Campos a partir dos dados coletados, foi de 9,78%, 9,95% no ano de 2017 e 11,02% em 2018, números próximos à taxa média de prematuridade brasileira nos períodos. A relação entre partos totais e nascimentos prematuros mensais no período de estudo é ilustrada na Figura 35.

Figura 35 - Relação Mensal entre parto total e prematuros nos anos de (a) 2016 (b) 2017 e (c) 2018 2018 na cidade de São José dos Campos-SP





Fonte: Elaborado pelo autor (2021).

Após a análise gráfica dos dados, foram calculados os valores mínimos, máximos, média e desvio padrão para cada variável escolhida, conforme Tabela 12. As concentrações dos poluentes são medidas em μg/m^3, idade da mãe em anos, e prematuridade em semanas de gestação.

A idade da mãe mais jovem encontrada na base de dados foi de 12 anos, e 45 anos para a mais velha, com média de 26,4 anos e desvio padrão de 6,9. O número de consultas variou entre nenhuma e quinze, com média de 6,4 consultas realizadas durante toda a gravidez e desvio padrão de 2,7. A mínima concentração média de NO_2 no período foi de 33,2 μg/m^3, enquanto a máxima foi de 51,2 μg/m^3, com média de 41 μg/m^3 e desvio padrão de 3,9. Já a concentração de PM_{10} variou entre 18,2 e 29,7 μg/m^3, com média de 23,5 μg/m^3 e desvio padrão de 2,3. O nascido vivo mais novo, teve seu parto realizado com 22 semanas de gestação, enquanto o mais velho, 36 semanas de gestação, a média foi de 33,9 semanas, com desvio padrão de 3,1.

Tabela 12 - Valores mínimos, máximos, média e desvio padrão para variáveis de entrada (idade da mãe, número de consultas, concentração de NO_2 e PM_{10}) e saída (prematuridade) para dados de nascidos vivos entre 2016 a 2018 em São José dos Campos-SP

Variável	Mínimo	Máximo	Média	Desvio Padrão
Idade da mãe	12,0	45,0	26,4	6,9
Número de consultas	0,0	15,0	6,4	2,7
Concentração Dióxido de Nitrogênio	33,2	51,2	41,0	3,9
Concentração Material Particulado 10	18,2	29,7	23,5	2,3
Prematuridade	22,0	36,0	33,9	3,1

Fonte: A Elaborado pelo autor (2023).

4.1 RESULTADO DO MODELO ELABORADO POR *FUZZY*

Após a inserção das variáveis de entradas, idade da mãe, número de consultas, concentração de NO_2 e PM_{10}, o sistema *fuzzy* retornou com os outputs, através do software MATLAB R2022b®. Os dados extraídos pelo "ddeinit" foram inseridos na tabela base do excel para ser realizada a correlação, função utilizada para determinar a relação entre dois intervalos.

O primeiro teste *fuzzy* utilizado, teve com função de pertinência do tipo triangular com três funções para as variáveis número de consultas e quatro para prematuridade, mantendo as outras variáveis com funções do tipo trapezoidal, três funções para idade da mãe e duas para cada concentração de poluente, para esse teste, a correlação encontrada foi de 0,19. Outros testes foram realizados, trocando os tipos de funções para as variáveis ou realizando leves ajustes de parâmetros. Em um dos testes, por exemplo, todas as variáveis foram consideradas trapezoidais (três para idade da mãe e número de consultas e duas para as concentrações de poluentes), exceto a saída, com quatro funções do tipo triangular, para esse caso a correlação foi de 0,18.

Em outro teste, a função de pertinência da saída foi do tipo triangular, com quatro funções, enquanto as variáveis de entrada foram consideradas trapezoidais (quatro para idade da mãe, três para número de consultas, e duas para as concentrações de NO_2 e PM_{10}), a correlação resultante foi de 0,11 para esse caso.

A pior correlação encontrada foi de 0,04, onde foram inseridas quatro funções de pertinência para a idade da mãe, quatro para o número de consultas, duas para cada concentração de poluente e saída com quatro funções de pertinência, ambas do tipo trapezoidal.

A seguir, é apresentada a Tabela 13 com os testes *fuzzy* realizados, indicando a quantidade e tipo de função de pertinência para cada variável, assim como a correlação, desvio padrão e erro médio entre os valores reais e preditos. A palavra TRAP significa função de pertinência do tipo trapezoidal, enquanto TRIAN simboliza a função triangular. Os leves ajustes de parâmetros não apresentaram grandes modificações ao valor da correlação e não foram considerados nessa tabela.

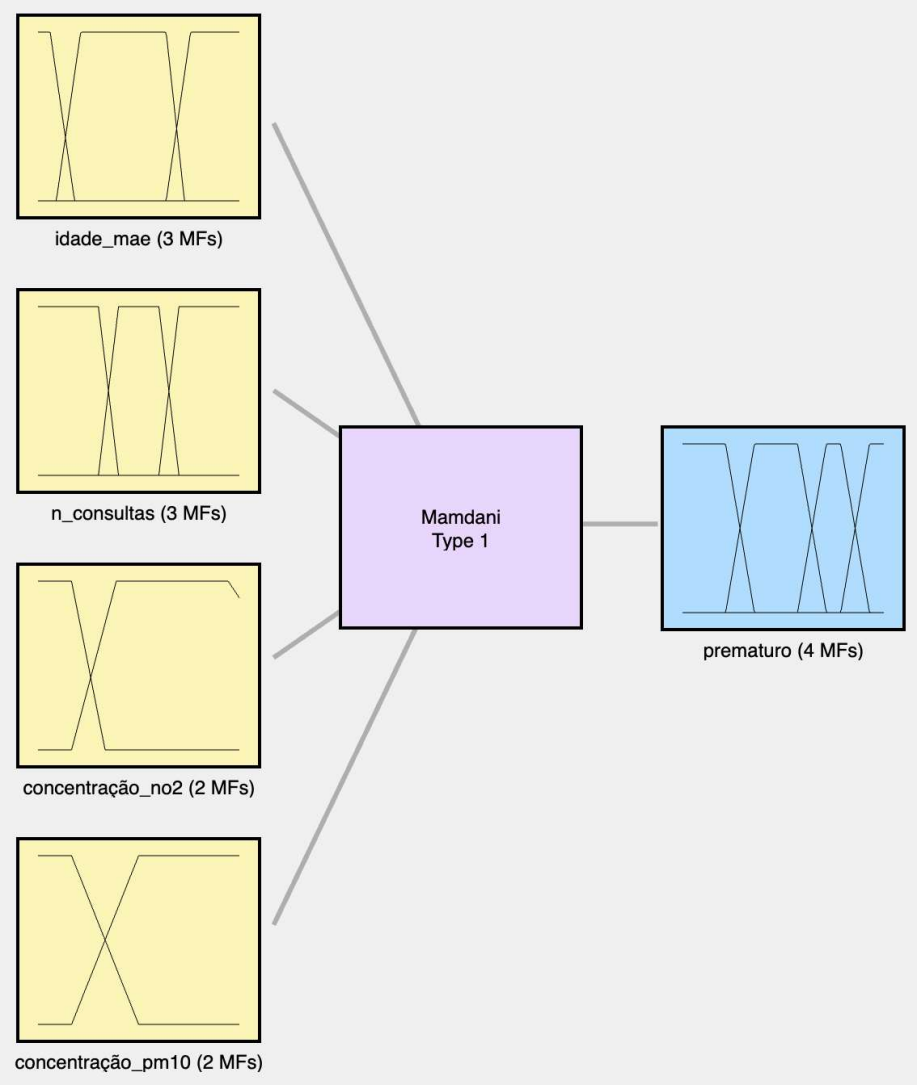
Tabela 13 - Modelos *fuzzy* realizados

	Idade da mãe	**N. Consultas**	**Conc. NO_2**	**Conc. PM_{10}**	**Prematur.**	**Correl.**	**Desvio Padrão**	**Erro médio**
Teste 1	3 TRAP	3 TRIAN	2 TRAP	2 TRAP	4 TRIAN	0,19	1,9	2,03
Teste 2	3 TRAP	3 TRAP	2 TRAP	2 TRAP	4 TRIAN	0,18	1,92	2,67
Teste 3	4 TRAP	3 TRAP	2 TRAP	2 TRAP	4 TRAP	0,1	2,64	2,65
Teste 4	4 TRAP	3 TRAP	2 TRAP	2 TRAP	4 TRIAN	0,11	2,37	2,61
Teste 5	4 TRAP	4 TRAP	2 TRAP	2 TRAP	4 TRAP	0,04	2,96	3,7
Teste 6	4 TRAP	3 TRAP	2 TRAP	2 TRAP	4 TRAP	0,11	2,8	2,87
Teste 7	3 TRAP	3 TRAP	2 TRAP	2 TRAP	4 TRAP	0,2	1,87	2,01
Teste 8	3 TRIAN	3 TRAP	2 TRAP	2 TRAP	4 TRAP	0,19	1,47	2,66

Fonte: Elaborado pelo autor (2023).

Após a realização dos testes, as melhores respostas de correlação e erro médio encontradas foram para três funções de pertinência para idade da mãe e número de consultas, duas para as concentrações de poluentes e quatro funções para o parto prematuro, todas do tipo trapezoidal (Teste 7). A estrutura completa do melhor modelo *fuzzy* é apresentada na Figura 36.

Figura 36 - Estrutura MATLAB do modelo *fuzzy*



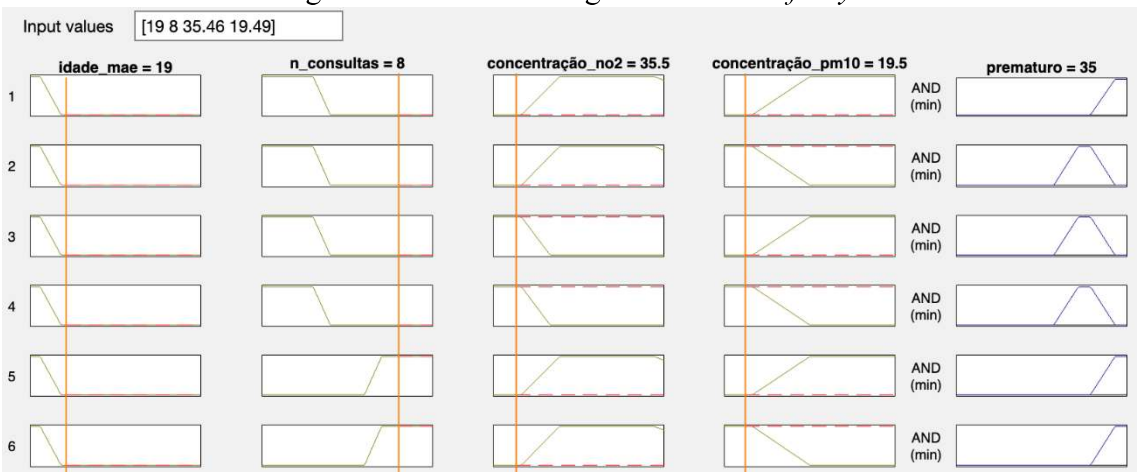
Fonte: Elaborado pelo autor (2022).

O detalhamento das funções de pertinência para cada variável de entrada e saída para esse modelo estão ilustradas nas Figuras 14 a 18.

A primeira validação realizada foi por meio da visualização de regras do MATLAB R2022b®. Para uma mãe de 19 anos, com 8 consultas pré-natais realizadas durante a gestação, exposta a 35,46 $\mu g/m^3$ de NO_2 e 19,49 $\mu g/m^3$ de PM_{10}, o resultado real é um parto prematuro de 35 semanas. Pelo modelo, o valor proposto foi idêntico ao real, 35 semanas.

Na Figura 37 são ilustradas as seis primeiras regras do sistema, assim como a visualização dos dados testados e seu resultado.

Figura 37 - Primeiras regras do modelo *fuzzy*



Fonte: Elaborado pelo autor (2022).

Na Figura 38 é possível observar a regra de número 18, demonstrando a ativação de todas as regras desse sistema *fuzzy*.

Figura 38 - Regra 18 do sistema *fuzzy*

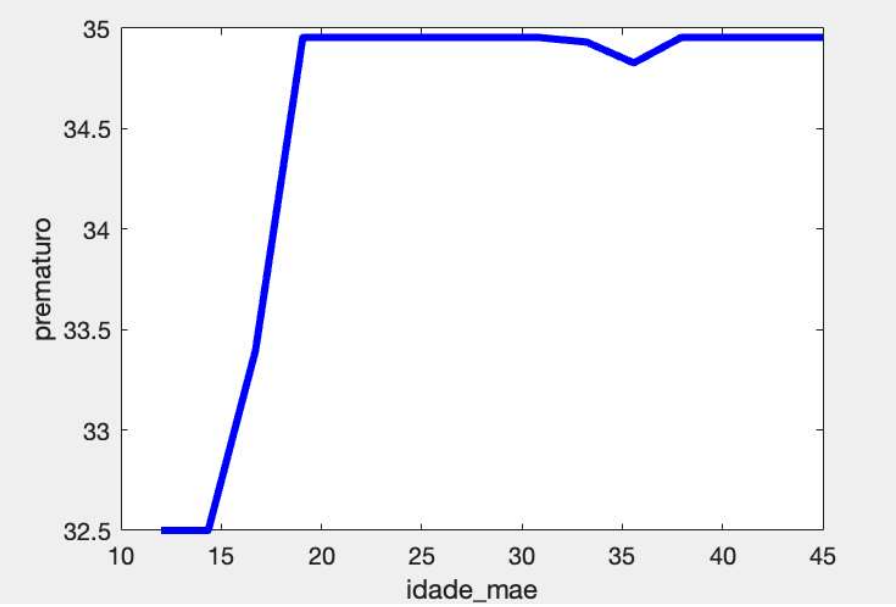


Fonte: Elaborado pelo autor (2022).

Pelo MATLAB R2022b® também é possível visualizar a superfície do sistema de forma gráfica. Apresentada no exemplo da Figura 39, em uma superfície bidimensional, o eixo x representa a idade da mãe e o eixo y marca a prematuridade em semanas de gestação. A partir dessa representação é possível visualizar que mães com idade baixa tendem a ter gestações mais curtas, porém conforme a idade avança, as semanas de gestação aumentam até permanecerem constantes.

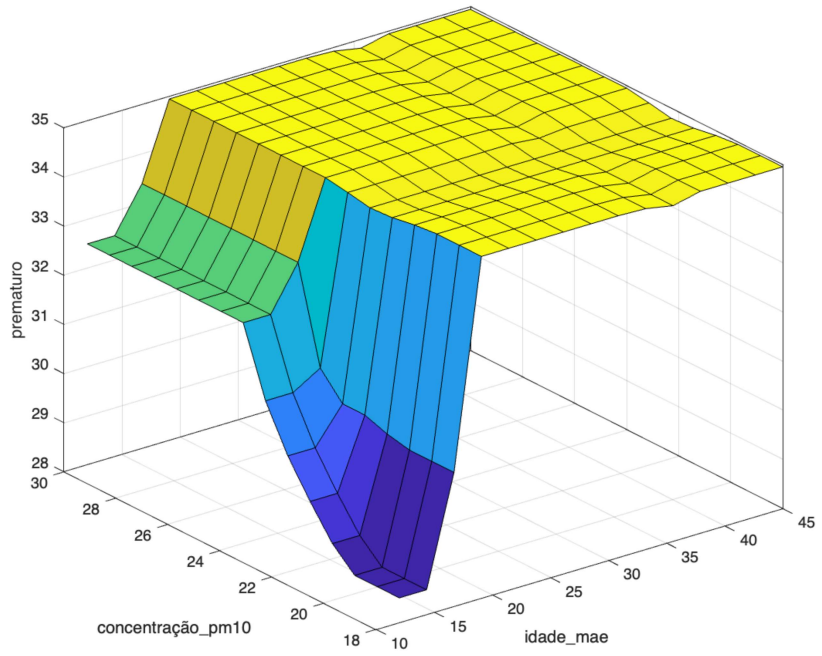
Já a Figura 40, ilustra a superfície de outra forma, tridimensional, onde o eixo x mantém representado a idade de mãe, porém o eixo y representa a concentração de Material Particulado (PM_{10}), e o eixo z marca a prematuridade.

Figura 39 - Superfície do sistema *fuzzy* no MATLAB: Idade da mãe x Prematuridade



Fonte: Elaborado pelo autor (2023).

Figura 40 - Superfície do Sistema de Inferência *Fuzzy* no MATLAB: Idade da mãe x Concentração PM_{10}



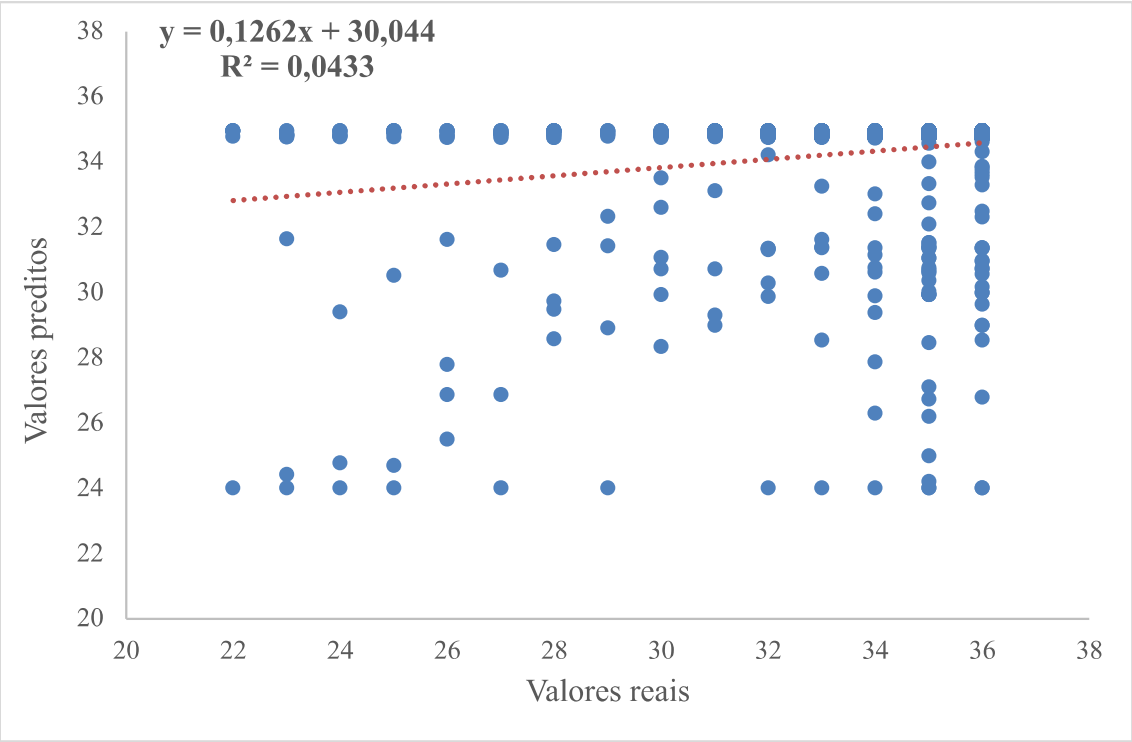
Fonte: Elaborado pelo autor (2023).

4.1.1 Avaliação do Modelo *Fuzzy*

Nesse modelo desenvolvido, a correlação retornou com o valor de 0,20, abaixo de um resultado considerado bom.

O desvio padrão entre os dados preditos pelo modelo *fuzzy* é de 1,87. A seguir, a Figura 41 apresenta um gráfico de dispersão entre os valores reais (eixo x) e preditos pelo sistema *fuzzy* (eixo y). A linha laranja simboliza uma tendência linear entre os dados, o valor do coeficiente de determinação e equação da linha é calculado automaticamente pelo excel, a partir dos dados informados.

Figura 41 - Dispersão dados reais x preditos *fuzzy*



Fonte: Elaborado pelo autor (2023).

4.2 RESULTADO DO MODELO ELABORADO POR NEURO-*FUZZY* ANFIS

4.2.1 Validação Simples

Para a técnica de neuro-*fuzzy*, testes foram realizados a partir da modificação do número de funções de pertinência e tipo de função para cada entrada, e verificação dos erros, conforme apresentado na Tabela 14. NF é o número de funções para cada variável, TIPO é o tipo de função de pertinência e ERRO é o erro médio do teste calculado pelo próprio sistema, *average testing error*. Os testes realizados possuem a estrutura do tipo *grid partition,* método de otimização híbrido, com 20 épocas e saída constante, com validação simples, sendo metade dos dados destinados a teste e metade restante para treino (50:50).

Tabela 14 - Testes neuro-*fuzzy* realizados

	NF	TIPO	ERRO		NF	TIPO	ERRO
Teste 1	3/4/3/3	TRIANG	2,30	Teste 25	3/4/3/3	TRAP	2,42
Teste 2	3/4/4/3	TRIANG	2,09	Teste 26	3/4/4/3	TRAP	2,36
Teste 3	2/2/2/2	TRIANG	2,82	Teste 27	2/2/2/2	TRAP	2,90
Teste 4	3/2/2/2	TRIANG	2,71	Teste 28	3/2/2/2	TRAP	2,89
Teste 5	3/3/2/2	TRIANG	2,69	Teste 29	3/3/2/2	TRAP	2,94
Teste 7	3/3/3/3	TRIANG	2,40	Teste 30	3/3/3/2	TRAP	2,78
Teste 8	2/3/2/2	TRIANG	2,80	Teste 31	3/3/3/3	TRAP	2,54
Teste 9	2/2/3/2	TRIANG	2,79	Teste 32	2/3/2/2	TRAP	2,93
Teste 10	2/2/2/3	TRIANG	2,78	Teste 33	2/2/3/2	TRAP	2,75
Teste 11	2/2/3/2	TRIANG	2,79	Teste 34	2/2/2/3	TRAP	2,66
Teste 12	2/2/2/3	TRIANG	2,78	Teste 35	2/2/3/2	TRAP	2,75
Teste 13	3/2/2/3	TRIANG	2,62	Teste 36	2/2/2/3	TRAP	2,66
Teste 14	2/3/3/2	TRIANG	2,76	Teste 37	3/2/2/3	TRAP	2,72
Teste 15	2/3/2/3	TRIANG	2,74	Teste 38	2/3/3/2	TRAP	2,80
Teste 16	2/2/3/3	TRIANG	2,60	Teste 39	2/3/2/3	TRAP	2,69
Teste 17	2/3/3/3	TRIANG	2,54	Teste 40	2/2/3/3	TRAP	2,60
Teste 18	3/4/3/3	GAUSS	2,27	Teste 41	2/3/3/3	TRAP	2,62
Teste 19	3/4/4/3	GAUSS	2,00	Teste 42	3/4/2/2	TRAP	2,85
Teste 20	4/4/4/3	GAUSS	1,77	Teste 43	3/4/3/3	BELL	2,26
Teste 21	4/4/4/4	GAUSS	1,68	Teste 44	3/4/4/3	BELL	2,08
Teste 22	3/3/4/3	GAUSS	2,07	Teste 44	4/4/3/4	GAUSS	1,82
Teste 23	3/3/3/4	GAUSS	2,18	Teste 45	4/4/3/3	GAUSS	1,95

Fonte: Elaborado pelo autor (2023).

Também foram testados modelos com configuração similar ao citado acima, dessa vez com 70% dos dados destinados a treino e 30% para teste (70:30). Não houve alteração significativa nos valores de *average testing* error, logo, não foram considerados nesse ambiente de estudo.

Com a realização de 44 testes, os cinco melhores foram selecionados e estão representados na Tabela 15, onde além do número, tipo de função de pertinência e *average testing error*, foram calculados correlação, desvio padrão, erro médio e realizado novo teste do número de épocas. Nota-se que os melhores resultados foram com a função de pertinência do tipo gaussiana e houve pouca diferença entre os valores para 15 e 20 épocas.

Tabela 15 - Melhores cinco testes neuro-*fuzzy* realizados

	NF	Tipo	N. épocas	*Average testing error*	Correlação	Desvio Padrão	Erro médio
Teste 21	4/4/4/4	GAUSS	20	1,68	0,55	3,80	1,82
			15	1,69	0,55	3,77	1,81
Teste 20	4/4/4/3	GAUSS	20	1,77	0,59	3,15	1,70
			15	1,77	0,59	3,16	1,70
Teste 44	4/4/3/4	GAUSS	20	1,82	0,41	5,31	1,92
			15	1,83	0,43	5,08	1,89
Teste 45	4/4/3/3	GAUSS	20	1,95	0,56	3,34	1,80
			15	1,95	0,56	3,30	1,79
Teste 19	3/4/4/3	GAUSS	20	2,00	0,44	3,15	1,89
			15	2,00	0,45	3,17	1,88

Fonte: Elaborado pelo autor (2023).

Verificou-se que os resultados de *average testing error*, correlação, desvio padrão e erro médio entre os testes com 20 e 15 épocas não apresentaram grandes diferenças.

Os melhores resultados encontrados foram para um sistema com idade da mãe, número de consultas, concentração de NO_2 e PM_{10} com 4, 4, 4 e 3 funções de pertinência, respectivamente, todas do formato gaussiana, com 20 épocas. O *average testing error* de 1,77, correlação de 0,59, desvio padrão de 3,15 e erro médio de 1,70.

Para o teste 20, com 20 épocas, a estrutura *subtractive clustering* também foi configurada, com parâmetros padrões de raio de inferência (0,5), fator multiplicador (1,25), raio de aceitação (0,5) e rejeição do agrupamento (0,15), conforme demonstrados na Figura 42.

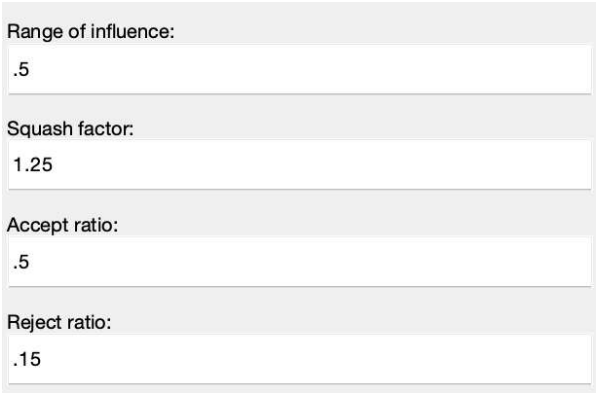
De forma simular à tabela anterior, foram calculados a correlação, desvio padrão e erro médio para as estruturas *grid partition* e *sub. clustering*. Os valores são representados na Tabela 16.

Tabela 16 - Melhor teste neuro-*fuzzy* realizado

	Estrutura	***Average testing error***	**Correlação**	**Desvio Padrão**	**Erro médio**
Teste 20	G*rid partition*	1,77	0,59	3,15	1,70
	Sub. clustering	2,75	0,47	1,66	2,02

Fonte: Elaborado pelo autor (2023).

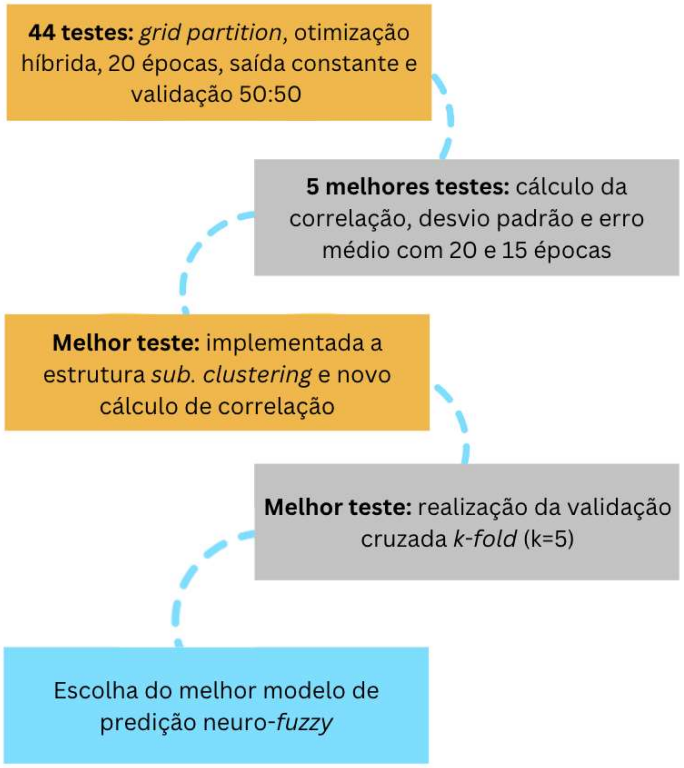
Figura 42 - Parâmetros padrões *sub. clustering*



Fonte: Elaborado pelo autor (2023).

A Figura 43, ilustra a sequência e critérios de testes realizados até a escolha do melhor modelo de predição. Após a sua realização, o modelo com melhor erro encontrado (*average testing error)* manteve-se para o treino em 20 épocas na estrutura *grid partition.* Nessa configuração, as variáveis de entrada foram escolhidas com quatro funções de pertinência, exceto a concentração de Material Particulado, com três funções de pertinência (4/4/4/3). Todas as variáveis de entrada apresentaram melhor performance para o tipo gaussiano, com saída constante.

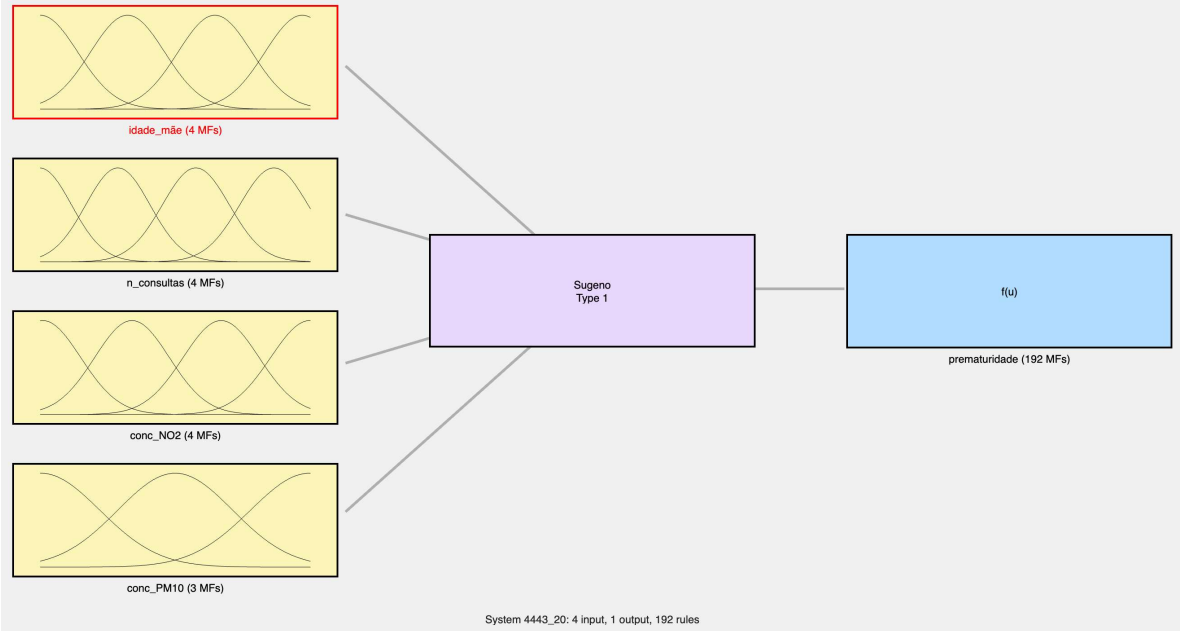
Figura 43 - Sequência de testes do modelo neuro-*fuzzy*



Fonte: Elaborado pelo autor (2023).

Os parâmetros, conjuntos e regras desse sistema são representados nas Tabelas 8 a 11, Figuras 27 a 31 e Figura 32, respectivamente. Esse modelo obteve ao total 192 regras (formato 4x4x4x3). A sua estrutura geral é representada conforme Figura 44.

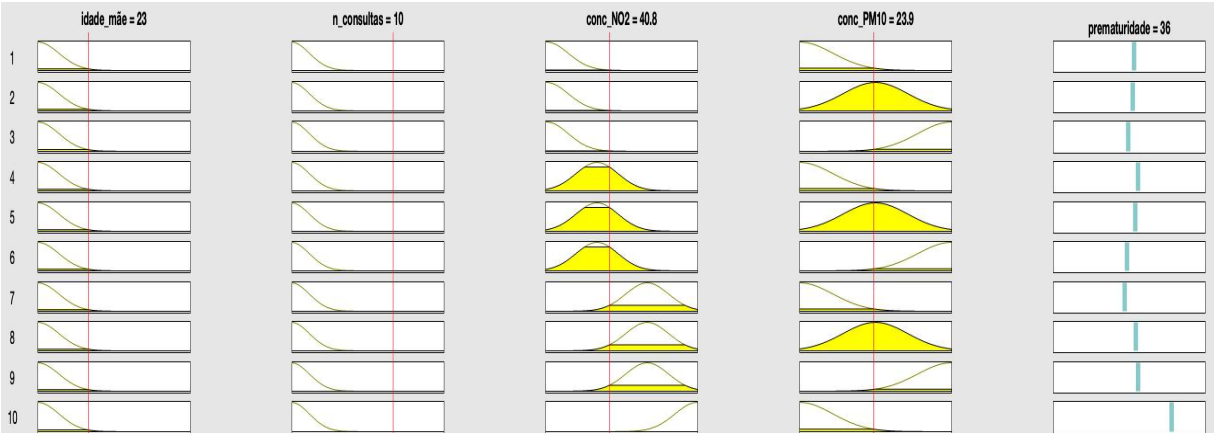
Figura 44 - Estrutura MATLAB do modelo neuro-*fuzzy*



Fonte: Elaborado pelo autor (2023).

Para esse sistema, utilizou-se um exemplo para verificação, com uma mãe com idade de 23 anos, dez consultas pré-natal realizadas, concentração média de NO_2 e PM_{10} durante a gravidez de 40,8 e 23,9 respectivamente. O resultado real, segundo as informações compiladas, é um parto prematuro com 36 semanas, e o resultado encontrado pelo sistema foi idêntico ao real, mostrando significativa confiabilidade do modelo produzido, ilustrado no *rule viewer* na Figura 45 a partir das dez primeiras regras.

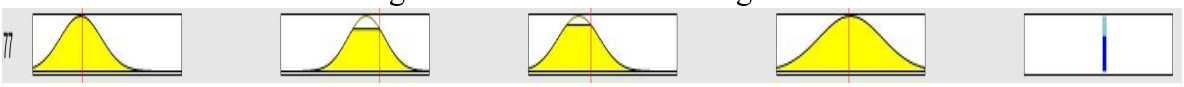
Figura 45 - *Rule viewer* do sistema neuro-*fuzzy*



Fonte: Elaborado pelo autor (2023).

A Figura 46 ilustra a regra 77, onde é possível visualizar a ativação de todas as regras desse sistema neuro-*fuzzy*.

Figura 46 - *Rule viewer* da regra 77



Fonte: Elaborado pelo autor (2023).

Após a realização de "ddeinit" e correlação, similarmente à logica *fuzzy*, os resultados neuro-*fuzzy* foram registrados. A correlação encontrada foi de 0,59 para o sistema neuro-*fuzzy* com validação simples.

4.2.2 Validação Cruzada *k-fold*

A partir da melhor estrutura neuro-*fuzzy* encontrada, de formato 4/4/4/3, com funções de pertinência gaussianas, método de otimização *grid partition,* 20 épocas e saída constante, foi também utilizado o método de validação cruzada *k-fold* como alternativa para

avaliar o modelo. A Figura 47 simboliza a divisão de dados utilizada para essa avaliação, com k=5, tendo em verde o conjunto aplicado para teste e em azul para treino.

Para cálculo da validação cruzada 5-*fold*, foi utilizado um script no MATLAB R2022b®, representado no Anexo 1. Nesse código criado, foi possível colocar as variáveis de entrada e saída, assim como o número de *folds*, método de otimização, números e tipos de cada função de pertinência. Como resultado, o software retornou com valores de correlação para cada *fold*, sendo o resultado final a média das correlações. Com a correlação (r), r *fold* 1=0,486, r *fold* 2=0,355, r *fold* 3=0,580, r fold 4=0,501 e r *fold* 5=0,521, a correlação final foi igual a 0,49.

Comparando as correlações realizadas por meio da validação simples 50:50 (r=0,59) e validação cruzada 5-*fold* (r=0,49), é possível afirmar que o modelo mais fiel aos dados reais foi o desenvolvido por meio de técnica neuro-*fuzzy* construído pela validação simples 50:50.

Figura 47 - Validação cruzada 5-*fold*



Fonte: Adaptado de Soares, Golliant, Santis e Aguiar (2017).

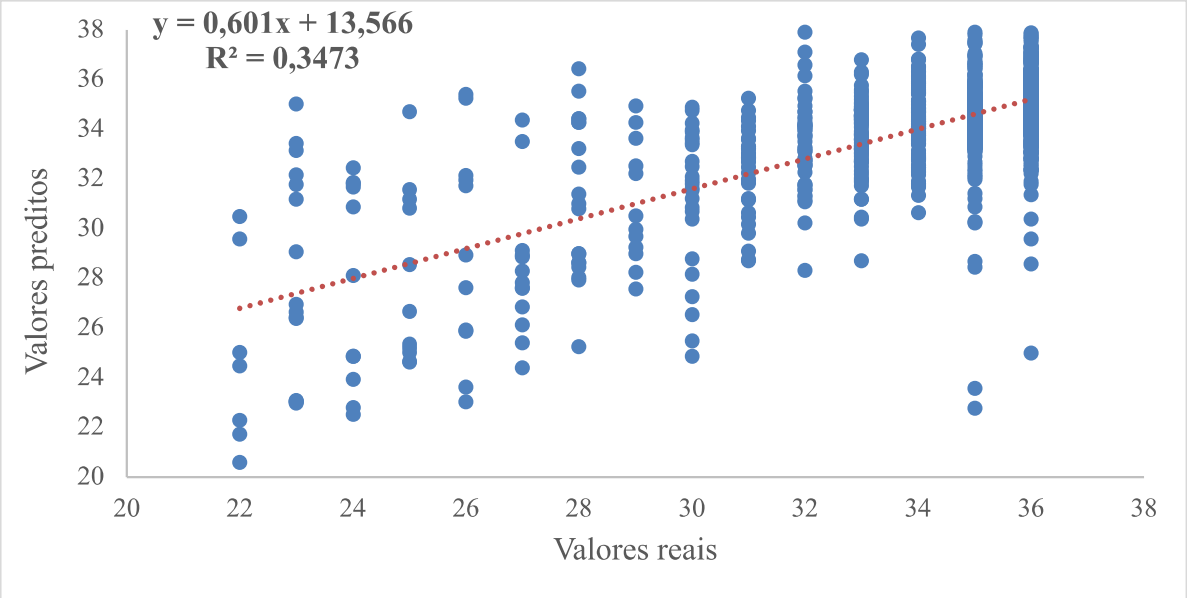
4.2.3 Avaliação do Modelo Neuro-*Fuzzy*

Além da correlação, outras métricas como desvio padrão e erro médio do melhor modelo também foram calculados.

O desvio padrão entre os dados modelados pela técnica neuro-*fuzzy* é de 3,15. A Figura 48 representa a correlação entre os dados reais (eixo x) e preditos (eixo y) através do gráfico de dispersão. A linha laranja simboliza a linha de tendência entre os dados, também

é demonstrado o valor do coeficiente de determinação, R^2, assim como a equação da linha de tendência.

Figura 48 - Dispersão dados reais x preditos neuro-*fuzzy*



Fonte: Elaborado pelo autor (2023).

4.3 MÉTRICAS DE AVALIAÇÃO

Métricas como MAE, RMSE, MSE, MAPE, R^2 e correlação foram utilizadas como método de validação e comparação entre os modelos *fuzzy* e neuro-*fuzzy* realizados.

A coeficiente de determinação (R^2) e correlação foram calculados pelo excel. Para o cálculo das métricas RMSE, MSE, MAPE e MAE, foi realizado o download do *Deep Learning Toolbox*, desenvolvido pelo *MathWorks* para a criação, treino e simulação de redes neurais. O *toolbox* foi utilizado no MATLAB R2022b®, a partir de um sistema operacional macOS. A sintaxe de cada uma das métricas utilizadas é representada na Tabela 17 abaixo.

Tabela 17 - Sintaxe para métricas de avaliação

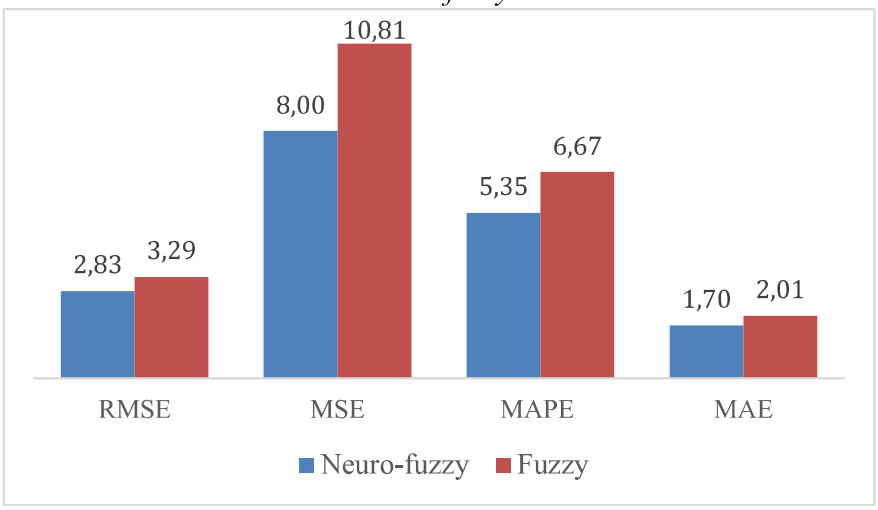
Métrica	**Sintaxe MATLAB**
RMSE	>> mse (F,A)
MSE	>> mse (F,A)
MAPE	>> mape (F,A)
MAE	>> mae (F,A)

Fonte: Elaborado pelo autor (2023).

Onde F é o conjunto de dados preditos e A é o conjunto observado.

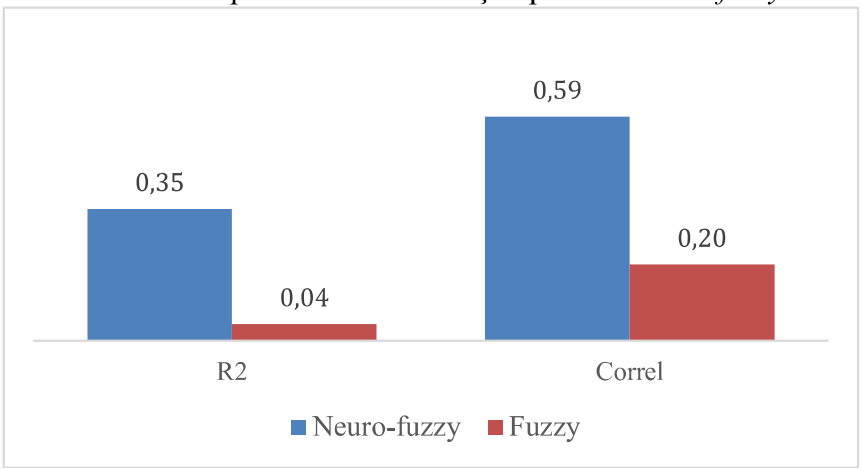
A Figura 49 e 50 ilustram os resultados de cada métrica calculada. A Figura 49 representa os erros RMSE, MSE, MAPE, MAE, enquanto a Figura 50 ilustra o coeficiente de determinação e correlação, para o melhor modelo *fuzzy,* representado pela barra vermelha e neuro-*fuzzy*, representado pela barra azul.

Figura 49 - Métricas de desempenho: RMSE, MSE, MAPE e MAE para o modelo *fuzzy* e neuro-*fuzzy*



Fonte: Elaborado pelo autor (2023).

Figura 50 - Métricas de desempenho: R^2 e correlação para o modelo *fuzzy* e neuro-*fuzzy*



Fonte: Elaborado pelo autor (2023).

O modelo neuro-*fuzzy* apresentou melhores resultados, em relação ao *fuzzy*, para todas as métricas de avaliação calculadas. Para os casos da raiz do erro quadrático médio (RMSE), erro médio quadrático (MSE), erro médio absoluto (MAE) e erro percentual absoluto médio (MAPE), quanto menor o valor, melhor o resultado, já para o coeficiente de determinação (R^2) e correlação, quanto mais próximo do valor 1, melhor o desempenho.

5 CONCLUSÃO

A partir desse estudo, foi possível coletar informações referentes aos nascidos vivos da cidade de São José dos Campos (SP), e relatórios dos poluentes como PM_{10} e NO_2, durante o período de 2016 a 2018.

Esse estudo abordou, de forma inédita, a utilização da lógica *fuzzy* e do sistema neuro-*fuzzy ANFIS*, para a confirmação da correlação entre os nascimentos prematuros e a poluição do ar na qual a gestante foi exposta durante toda a gravidez.

Foram desenvolvidos modelos baseados na lógica *fuzzy* e neuro-*fuzzy*, relacionando as variáveis de entrada: idade da mãe, número de consultas, concentração de PM_{10} e NO_2 com a variável de saída, prematuridade. Entre todos os modelos desenvolvidos, o sistema neuro-*fuzzy* utilizado com validação simples 50:50, estrutura do tipo *grid partition,* método de otimização híbrido, com 20 épocas, entradas gaussianas, sendo as três primeiras com quatro funções de pertinência e a última com três funções, e variável de saída constante, apresentou melhores valores de correlação (r=0,59), coeficiente de determinação (R^2=0,35), erro médio absoluto (MAE=1,70), raiz do erro quadrático médio (RMSE=2,83), erro percentual absoluto médio (MAPE=5,35%) e erro médio quadrático (MSE=8,00).

Ao longo do estudo, foram encontradas algumas limitações que se dizem respeito, principalmente, a base de dados. Foi apontada a possibilidade de erros na base de informações das gestantes do DATASUS, uma vez que esses dados foram preenchidos de forma manual por pessoas que poderiam não ser especializadas.

Das limitações de informações dos gases, a concentração de poluentes foi considerada homogênea em todo o território de São José dos Campos, visto que no município existiam apenas duas estações de medição, além disso, para esse estudo, foi considerado que a gestante esteve na cidade durante toda a gravidez.

Para trabalhos futuros, destaca-se a oportunidade de exploração da temperatura como variável de entrada do sistema, visto que o aquecimento global e as mudanças climáticas apresentam impactos na vida humana. Além disso, poderão ser utilizados dados de cidades com características semelhantes ao município de São José dos Campos-SP, para verificação de similaridade do modelo e resultados de predição.

A criação dos modelos desenvolvidos em função da idade da mãe, número de consultas pré-natal e exposição a NO_2 e PM_{10} é inovador na literatura e de grande valia para a sociedade, podendo auxiliar na criação e implementação de políticas de redução de gases

poluentes, melhorando, consequentemente, a saúde das gestantes e recém-nascidos. Com a melhora da qualidade do ar e redução das taxas de prematuridade, a saúde das futuras gerações poderá ser preservada.

REFERÊNCIAS

ALMEIDA, A. C.; JESUS, A. C. P.; LIMA, P. F. T.; ARAÚJO, M. F. M.; ARAÚJO, T. M. Fatores de risco maternos para prematuridade em uma maternidade pública de Imperatriz- MA. **Revista Gaúcha de Enfermagem**, Porto Alegre, v. 33, n. 2, p. 86-94, 2012. https://doi.org/10.1590/S1983-14472012000200013.

ARAÚJO, S. L. L.; SALDIVA, P. H. N.; SCHWARTZ, J.; ZANOBETTI, A**.** Efeitos isolados e sinérgicos do MP_{10} e da temperatura média na mortalidade por doenças cardiovasculares e respiratórias. **Revista Saúde Pública,** São Paulo, v. 48, n. 6, p. 881-888, 2014. https://doi.org/10.1590/S0034-8910.2014048005218.

BARONE, D. A. C. **Projeto Revox**. [s. l.], 1999.

BEHRMAN, R. E.; BUTLER, A. S. **Preterm birth:** causes, consequences, and prevention. Washington: The National Academies Press, 2007.

BERTRAND, J. W. M.; FRANSOO, J. C. Modeling and simulation: operations management research methodologies using quantitative modeling. **Internation Journal of Operations & Production Managemet**, v. 22, n. 2, p. 241-264, 2002. https://doi.org/10.1108/01443570210414338.

BICK, D. Born too soon: the global issue of preterm birth. **Midwifery**, New York, v. 28, n. 4, p. 401-402, 2012. https://doi.org/10.1016/j.midw.2012.06.010.

BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDEMIR, T. B. **Redes Neurais Artificiais**: Teoria e aplicações. Rio de Janeiro: LTC, 2000. 250p.

BUJA, G. S.; KAZMIERKOWSKI, M. P. Direct torque control of PWM inverter-fed AC motors - A survey. **IEEE Transactions on Industrial Electronics**, New Jersey, v. 51, n. 4, p. 744-757, 2004. DOI: 10.1109/TIE.2004.831717.

CARVALHO, P. C.; NASCIMENTO, L. F. C.; NAKAZATO, L. F. Associação entre exposição materna a poluentes do ar e parto prematuro em Ribeirão Preto. **Revista Biociências**, Taubaté, v. 20, n. 2, p. 107-114, 2014.

CASTANHO, M. J. P. Sistema neuro-*fuzzy* para prognóstico de câncer de próstata. **Revista Biomatemática**, Campinas, v. 17, n. 2, p. 11-20, 2007.

CASTELLÕES, T. de O.; DAVID, T. M.; NASCIMENTO, L. F. C.; RIZOL, P. M. S. R. Bibliometric analysis: premature birth, air pollution and fuzzy logic. **Research, Society and Development**, Itajubá, [S. l.], v. 11, n. 12, 2022. DOI: 10.33448/rsd-v11i12.34233.

SÃO PAULO (Estado). CETESB. COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO. **Qualidade do ar, Informações básicas:** poluentes. Disponível em: https://cetesb.sp.gov.br/ar/poluentes/. Acesso em: 04 abr. 2023.

CHAVES, L. E. **Modelos computacionais Fuzzy e Neuro-Fuzzy para avaliarem os efeitos da poluição do ar.** Tese (Doutorado em Engenharia Mecânica) - Universidade Estadual Paulista, Guaratinguetá, 2013.

COELHO, L.; MATTOS, V. Um experimento para determinar algumas propriedades de controladores fuzzy. 2014.In: MOSTRA DA PRODUÇÃO UNIVERSITÁRIA DA UNIVERSIDADE FEDERAL DO RIO GRANDE, 13, 2014, Porto Alegre. Anais da MPU. Rio Grande: Editora da FURG, 2014. Disponível em: https://propesp.furg.br/anaismpu/cd2014/trabalhos.html. Acesso em: 19 fev. 2022.

COX, E. **The fuzzy systems handbook**: a practitioner's guide to building, using, anda maintainning fuzzy systems. New York: AP Professional, 1994.

DAVID, G.; RIZOL, P. M. S. R.; NASCIMENTO, L. F. C. Modelos computacionais fuzzy para avaliar efeitos da poluição do ar em crianças. **Revista Paul. Pediatria**, São Paulo, v. 36, n. 1, 2018. http://dx.doi.org/10.1590/1984-0462/;2018;36;1;00013

BRASIL. MINISTÉRIO DA SAÚDE. DEPARTAMENTO DE INFORMÁTICA DO SUS. **O DATASUS** Disponível em: https://datasus.saude.gov.br/sobre-o-datasus/. Acesso em: 04 abr. 2023.

DEREK, E. **Atmospheric pollution:** a global problem. 2nd ed. Oxford: Blackwell, 1992.

DEVORE, J. **Probability and statistics for engineering and the sciences**. 8th ed. Boston: Cengage Learning, 2012.

DIAS, B. A. S.; LEAL, M. C.; MARTINELLI, K. G.; NAKAMURA, M. P.; PEREIRA, A. P. P.; NETO, E. T. Prematuridade recorrente: dados do estudo "Nascer no Brasil". **Rev Saude Publica**, São Paulo, v. 56, n. 7, 2022. https://doi.org/10.11606/s1518-8787.2022056003527

EIBEN, A. E.; SMITH, J. E. **Introduction to evolutionary computing**. 2nd ed. Berlin: Springer, 2003.

FREITAS, F. A. **Aplicação da teoria dos conjuntos Fuzzy no estudo da impedância eletro-mecânica.** Dissertação (Mestrado em Matemática) - Universidade Federal de Uberlândia, Uberlândia, 2021.

FULLER, R**. Neural Fuzzy Systems**. Turku: Abdo Akademi University, 1995.

GÉRON, A. **Hands-on machine learning with scikit-learn and tersonFlow**: Concepts, Tools and Techniques to build intelligent systems. 1st ed. Sebastopol: Editions O'Reilly, 2017.

GIL, A. **Como elaborar projetos de pesquisa**. 3. ed. São Paulo: Editora Atlas, 1996. 159p.

GILAT, A. **MATLAB com aplicações em engenharia**. 4. ed. Porto Alegre: Bookman Editora, 2009, p.1. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=n1cdVJRs3qQC&oi=fnd&pg=PR5&dq=MATLAB&ots=HZvnc2Lvqc&sig=

pUmiQXrLm8PqiZfvYZT-2f68zqM#v=onepage&q=MATLAB&f=false. Acesso em 07 jun. 2023.

GOLDBERG, D. **Genetic algorithms in search, optimization and machine learning**. Glenview: Addison Wesley, 1989.

HARRISON, M. S.; GOLDENBERG, R. L. Global burden of prematurity. **Seminars In Fetal And Neonatal Medicine**, Michigan, v. 21, n. 2, p.74-79, abr. 2016. http://dx.doi.org/10.1016/j.siny.2015.12.007.

HAYKIN, S. **Redes neurais:** princípios e práticas. 2. ed. São Paulo: Bookman, 2001. 900p.

HILLER, F. S.; LEIBERMAN, G. J. **Introdução à pesquisa operacional**. 8. ed. São Paulo: Mc Graw Hill, 2017.

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International Journal of Forecasting**, Amsterdam, v.22, n. 4, p. 679-688, 2006. DOI: 10.1016/j.ijforecast.2006.03.001

BRASIL. IBGEa. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Painel de Indicadores**: Brasil. Disponível em: https://www.ibge.gov.br/indicadores.html. Acesso em: 04 abr. 2023.

BRASIL. IBGEb. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Cidades e Estados**: São José dos Campos. Disponível em: https://www.ibge.gov.br/cidades-e-estados/sp/sao-jose-dos-campos.html. Acesso em: 05 abr. de 2023.

IYODA, E. M.; **Inteligência computacional no projeto automático de redes neurais híbridas e redes neuro-fuzzy heterogêneas.** Dissertação (Mestrado em Engenharia Elétrica e da Computação) – Universidade Estadual de Campinas, Campinas, 2023.

JAMES, G. *et al.* **An introduction to statistical learning with applications in R**. 8th ed. New York: Springer, 2017.

JANG, J. S. R. ANFIS: Adaptive-network-based Fuzzy inference system. **IEEE Transactions on Systems, Man, and Cybernetics**, New Jersey, v. 23, n. 3, p. 665–685, 1993. DOI: 10.1109/21.256541.

JANG, J. S. R.; SUN, C. T. Neuro-fuzzy modeling and control. **Proceedings of the IEEE**, New York, v. 83, n. 3, p. 378-406, 1995. DOI: 10.1109/5.364486.

KAPLAN, A.; HAENLEIN, M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. **Business Horizons**, Blomington, v. 62, n. 1, p. 15-25, 2019. DOI: 10.1016/j.bushor.2018.08.004.

KRYKHTIME, F. L. P. **Abordagem fuzzy para otimização de planos de vôo**. Dissertação (Doutorado em Engenharia de produção) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.

MATHWORKS**. Fuzzy logic toolbox**: User's Guide. The Mathworks, Inc, 2012.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada:** uma abordagem aplicada. Belo Horizonte: Editora UFMG, 2005. 295p.

MITCHELL, T. M. **The discipline of machine learning**. Pittsburgh: Mach Learning Department; 2006.

MOURA, A. O.; ANDRADE, D. S.; KIKUCHI, W. K. Inteligência artificial aplicada no auxílio a tomadas de Decisões na área de Saúde. 2005. In: SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA (SEGeT), 2, 2005, Resende, Site SEGeT (AEDB). Rio de Janeiro: 2005. Disponível em: https://www.aedb.br/seget/arquivos/artigos05/293_Inteligencia%20Artificial%20Aplicada%20no%20Auxilio%20a%20Tomadas%20de%20Decisoes%20na%20Area%20de%20Saude.pdf. Acesso em: 03 de maio 2022.

MORAES, O. **Método de análise de dados para avaliação de áreas urbanas recuperadas -** uma abordagem utilizando a lógica fuzzy. Tese (Doutorado em Engenharia de Construção Civil e Urbana) - Universidade de São Paulo, São Paulo, 2008.

MOURA, M. **Proposta de rede neural para previsão de fatura de energia elétrica**. Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica) - Universidade Federal Rural do Semi-Àrido, Mossoró, 2021.

NAGASHIMA, R. **Treinamento de uma rede neural de base radial usando computação evolutiva:** implementação e aplicações. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Estadual Paulista, Ilha Solteira, 2006.

NASCIMENTO, L. F. C.; MACHIN, A. B.; SANTOS, D. A. A. Are there differences in birth weight according to sex and associations with maternal exposure to air pollutants? A cohort study. **Sao Paulo Medical Journal**, São Paulo, v. 135, n. 4, p.347-354, 2017. http://dx.doi.org/10.1590/15163180.2016.0262100317.

NASCIMENTO, L. F. C.; MEDEIROS, A. P. P. Internações por pneumonias e queimadas: uma abordagem espacial. **Jornal de Pediatria,** Porto Alegre, v. 88, n. 2, p. 177-183, 2012. DOI: 10.2223/JPED.2161

NASCIMENTO, L. F. C.; RIZOL, P. M. S. R.; ABIUZI, L. B. **Cad. Saúde Pública**, Rio de Janeiro, v. 25, n. 9, 2009. DOI: 10.1590/s0102-311x2009000900018.

PALIWAL, M.; KUMAR, U. A. Neural networks and statistical techniques: A review of applications. **Expert Systems with Applications,** New York, v. 36, p. 2-17, 2009. https://doi.org/10.1016/j.eswa.2007.10.005.

PASSINI, R. J.; TEDESCO, R. P.; MARBA, S. T.; CECATTI, J. G.; GUINSBURG, R.; MARTINES, F. E.; NOMURA, M. L. Brazilian multicenter study on prevalence of

preterm birth and associated factors.**BMC Pregnancy Childbirth**, v. 10, n. 22, 2010. DOI: 10.1186/1471-2393-10-22.

RAUBER, T. W. **Redes neurais artificiais -** Universidade Federal do Espírito Santo, Vitória, 2001. Disponível em: https://www.researchgate.net/publication/228686464_Redes_neurais_artificiais. Acesso em 10 fev. 2023.

REQUIA, J. W.; ROIG, H. L. Avaliação espacial entre poluição do ar e saúde em áreas com limitação de dados. **Bol. Ciênc. Geod.** Curitiba, v.22, n.4, p.807-821,2016. https://doi.org/10.1590/S1982-21702016000400046

RODRIGUES, J. P.; SOUZA, M.; FERREIRA, P.; MARQUES, B. A. Aplicações da IA na saúde humana, 2021. In: ANAIS DO CONGRESSO NACIONAL UNIVERSIDADE, EAD E SOFTWARE LIVRE, v. 2, n. 12, 2021, *[s.l.]*. Anais UEADSL. 2021. Disponível em: https://ueadsl.anais.nasnuv.com.br/index.php/UEADSL/article/view/631. Acesso em: 05 jun. 2023.

SANTOS, D. A. A.; NASCIMENTO, L. F. C. Maternal exposure to benzene and toluene and preterm birth. A longitudinal study. **Sao Paulo Med. J.**,São Paulo , v. 137, n. 6, p. 486-490, 2019. https://doi.org/10.1590/1516-3180.2019.0224170919.

SANTOS, V. P.; MEDEIROS, A. P. P. LIMA, T. A. C.; NASCIMENTO, L. F. C. Poluentes atmosféricos associados ao peso insuficiente ao nascimento. **Revista Brasileira de Epidemiologia**, São Paulo, v. 19, n. 1, p. 89-99, 2016. https://doi.org/10.1590/1980-5497201600010008.

SAMSON, I. J.; PRABAKAR, S.; PORKUMARAM, K. Green computing methodology on screening of retinopathy of prematurity using vital statistical features of retcam images. **Journal of Green Engineering,** Denmark, v. 10, n. 9, p. 6709-6721, 2020.

SCHLECHTINGEN, M.; SANTOS, I. F.; ACHICHE, S. Using data-mining approaches for wind turbine power curve monitoring: A comparative study. **IEEE Transactions on Sustainable Energy**, New Jersey, v. 4, n. 3, p. 671-679, 2013. DOI: 10.1109/TSTE.2013.2241797.

SILVA, A. V. da; SILVA, I. A. F.; TEIXEIRA FILHO, M. C. M.; BUZETTI, S.; TEIXEIRA, M. C. Estimativa da produtividade de trigo em função da adubação nitrogenada utilizando modelagem neuro fuzzy. **Rev. Bras. Eng. Agric. ambient**, Campo Grande, v.18, n.2, 2014. https://doi.org/10.1590/S1415-43662014000200008

SILVA, E.; MENEZES, E. **Metodologia da pesquisa e elaboração de dissertação.** 3. ed. Florianópolis: Laboratório de Ensino a Distância da UFSC, 2001. 20p.

SILVA, G. V. **Controle não linear.** 3. ed. Lisboa: Edição digital, 2006. Disponível em: http://www.microsistemascol.com/doc/ControleNaoLinearLisboa.pdf. Acesso em 25 jul. 2022.

SIVANADAM, S. N.; SUMATHI, S.; DEEPA, S. N. **Introduction to fuzzy logic using MATLAB**. Berlim: Springer, 2007.

SOARES, N.; GOLLIANT, L.; SANTIS, R.; AGUIAR, E. Máquinas de aprendizado extremo para a predição do volume de clones de eucaliptos, 2017. In: CONFERENCE IBERO LATIN AMERICAN CONGRESS ON COMPUTACIONAL METHODS IN ENGINEERING (CILAMCE), 38, 2017, Florianópolis. Researchgate. Santa Catarina: 2017. Disponível em: https://www.researchgate.net/publication/321027902_Maquinas_de_aprendizado_extremo_para_a_predicao_do_volume_de_clones_de_eucaliptos Acesso em: 02 mar 2023.

SPORL, C.; CASTRO, E. G.; LUCHIARI, A. Aplicação de redes Neurais artificiais na construção de modelos de fragilidade ambiental. **Revista do Departamento de Geografia**, São Paulo, v. 21, n.1, p. 113-135, 2011. https://doi.org/10.7154/RDG.2011.0021.0006.

STEER, P.; FLINT, C. ABC of labour care: preterm labour and premature rupture of membranes. **BMJ**, London, v. 318, n. 7190, 1999 Abr 17, p. 1059-1062. DOI: 10.1136/bmj.318.7190.1059

TADANO, Y. S. **Simulação da dispersão dos poluentes atmosféricos para aplicação em análise de impacto**. Tese (Doutorado em Engenharia Mecânica) - Universidade Estadual de Campinas, Campinas, 2012.

TAFNER, M. A. **Reconhecimento de palavras isoladas usando redes neurais artificiais**. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal de Santa Catarina, Florianópolis, 1996.

FLEURY, A. *et al.* **Metodologia de pesquisa em engenharia de produção** e gestão de operadores. 2. ed. Rio de Janeiro: Elsevier, 2012.

WEI, M.; BAI, B.; SUNG, A. H.; LIU, Q.; WANG, J.; CATHER, M. E. Predicting injection profiles using ANFIS. **Information Sciences**, Edmonton, v. 177, n. 20, p. 4445-4461, 2007. https://doi.org/10.1016/j.ins.2007.03.021.

WORLD HEALTH ORGANIZATION. **Air quality guidelines**: Global Update 2005. Genebra: WHO, 2006

WYATT, N. **The financial times essential guide to budgeting and forecasting:** how to deliver accurate numbers. Londres: Pearson Education, 2012.

WYTHOFF, B. J. Backpropagation neural networks. A tutorial. **Chemometrics and Intelligent Laboratory Systems,** v. 18, n. 2, p. 115–155, 1993. https://doi.org/10.1016/0169-7439(93)80052-J.

ZADEH, L. A. Fuzzy sets. **Information and Control**, v. 8, n. 3, p. 338-353, 1965. https://doi.org/10.1016/S0019-9958(65)90241-X.

ZAKRZEWSK, S. P. **Principles of Environmental Toxicology.** 2nd ed. New York: Oxford University Press, 1997. 133p.

ANEXO 1

```
data = readtable('matlab.xlsx');
x = data(:,["IDADEMAE","CONSULTAS","MEDIANO2","MEDIAMP10"]);
x = table2array(x);
y = data(:,"SEMANAGESTACAO");
y = table2array(y);
%
cv = cvpartition(height(x),'KFold',5);
for i=1:5


    fprintf('Modelo gerado, fold %d\n',i);
    TrainInd = training(cv,i);
    TestInd = test(cv,i);
    x_train = x(TrainInd,:);
    y_train = y(TrainInd,:);
    x_test = x(TestInd,:);
    y_test = y(TestInd,:);
    % Modelo FIS
    genOpt = genfisOptions('GridPartition');
    genOpt.NumMembershipFunctions = [4 4 4 3];
    genOpt.InputMembershipFunctionType = ["gaussmf", "gaussmf", "gaussmf",
"gaussmf"];
    genOpt.OutputMembershipFunctionType = ['constant'];
    inFIS = genfis(x_train, y_train,genOpt);
    % ANFIS
    %'OptimizationMethod'
    %1 - hybrid
    %0 - backpropagation
    trainOpt = anfisOptions('InitialFIS',inFIS, ...
                    'EpochNumber',20, ...
                    'OptimizationMethod',1);
    xtrain = [x_train, y_train];
    [outFIS,trainError,stepSize] = anfis(xtrain,trainOpt);
    % Avaliação
    y_pred = evalfis(outFIS,x_test);
    fprintf('Correlação: %d\n', corr2(y_test, y_pred));
end
```