

Instituto Federal de Santa Catarina

Desenvolvimento de redes Naive Bayes e sistema de inferência *fuzzy* para suporte à tomada de decisão acerca da interrupção de ventilação mecânica assistida

Especialização em Computação Científica para a Indústria

Pós-Graduação Lato Sensu

Unidade Curricular: Sistemas probabilísticos e difusos.

Campus Florianópolis

Professor: Cesar Alberto Penz

Alunos: Deivity Rosa de Andrade – deivity.a14@aluno.ifsc.edu.br

Lucas Gabriel Coliado Bandeira – lucas.qcb@aluno.ifsc.edu.br

Florianópolis, maio de 2023.

Este trabalho foi desenvolvido segundo a proposta do professor Cesar Alberto Penz, para a disciplina de Sistemas probabilísticos e difusos do Curso de Especialização em Computação Científica para a Indústria.

O trabalho consiste em (1) desenvolvimento de redes Naive Bayes para suporte à tomada de decisão acerca da interrupção de ventilação mecânica assistida e (2) desenvolvimento de um sistema de inferência fuzzy para suporte à tomada de decisão acerca da interrupção, para finalmente (3) comparar os resultados obtidos com a rede Naive Bayes, sistema fuzzy e mapa de Kohonen de ventilação mecânica assistida.

1- DESENVOLVIMENTO DE REDES NAIVE BAYES PARA SUPORTE À TOMADA DE DECISÃO ACERCA DA INTERRUPTÃO DE VENTILAÇÃO MECÂNICA ASSISTIDA

Para o desenvolvimento de uma rede Naive Bayes para suporte à tomada de decisão acerca da interrupção/desmame da ventilação mecânica, segundo a proposição, foram consideradas as variáveis de entrada:

- a) NIF - pico negativo da pressão inspiratória (cm.H₂O);
- b) RR - taxa respiratória (respirações/minuto);
- c) VT – volume tidal (ml); e,
- d) RSBI - (Rapid Shallow Breathing Index).

Foram coletadas informações no documento disponibilizado, denominado “Base de Conhecimento Weaning”. Deste documento foi obtido as seguintes observações em relação às funções de pertinência:

- Volume Tidal (VT): sucesso acima de 315 ml, fracasso abaixo de 315 ml;
- RSBI: sucesso ≤ 80 BM/L, fracasso ≥ 120 BM/L;
- Taxa respiratória (RR) : sucesso < 30 R/M, fracasso ≥ 35 R/M;
- Pico negativo da pressão inspiratória (NIF): sucesso < -26 CM H₂O, fracasso > -20 CM H₂O.

Como a rede Naive Bayes não aceita intervalos que se interseccionam e como no trabalho proposto entendemos que a pior situação seria retirar a ventilação forçada quando não deveria, ocasionando no risco de morte do paciente, foi adotado a premissa da opção mais abrangente ao fracasso, evitando assim falsos positivos (retirar quando não deveria), conforme segue:

Tabela 1 - Critérios utilizados na rede bayes

CASOS NIF		CASOS RR		CASOS VT		CASOS RSBI	
< -26	≥ -26	< 30	≥ 30	≥ 315	< 315	≤ 80	> 80
Sucesso	Fracasso	Sucesso	Fracasso	Sucesso	Fracasso	Sucesso	Fracasso

1.1 Implementação

Para desenvolvimento do sistema foi utilizado a linguagem de programação Python em conjunto uma biblioteca específica para rede Naive Bayes, denominada “sklearn.naive_bayes”¹ e utilizando o módulo “GaussianNB” que implementa o algoritmo Gaussian Naive Bayes para classificação. A probabilidade dos recursos é assumida como Gaussiana.

Foram utilizadas bibliotecas auxiliares “sklearn.preprocessing” com o módulo “LabelEncoder”² e “pandas”³, todo código, no qual serão abordados os resultados logo abaixo, estão disponibilizados em arquivo formato “py” e , disponibilizados de modo permanente no seguinte link: <https://github.com/lucasgcb/fuzzy-bayes>

Para a aplicação deste caso em Python foi necessário o tratamento dos dados para ficarem aptos à utilização das bibliotecas citadas, vamos discorrer um pouco sobre este tratamento antes de apresentar os resultados.

Os dados originais tiveram que ser transformados, inicialmente, segundo a caracterização aderente, conforme abaixo:

Tabela 2 - De-Para dos dados originais

De:	NIF	VT	RR	RSBI	Label	Para:	NIF	VT	RR	RSBI	Label
	-30	500	17	34	F		<_-26	>=_315	<_30	<=_80	F
	-25	412	30	73	F		>=_-26	>=_315	>=_30	<=_80	F
	-22	350	30	86	F		>=_-26	>=_315	>=_30	>_80	F
	-18	370	32	86	F		>=_-26	>=_315	>=_30	>_80	F
	-10	320	29	91	F		>=_-26	>=_315	<_30	>_80	F
	-28	400	38	95	F		<_-26	>=_315	>=_30	>_80	F
	-24	276	38	138	F		>=_-26	<_315	>=_30	>_80	F
	-12	220	32	145	F		>=_-26	<_315	>=_30	>_80	F
	-15	270	40	148	F		>=_-26	<_315	>=_30	>_80	F
	-26	265	44	166	F		>=_-26	<_315	>=_30	>_80	F
	-20	225	46	204	F		>=_-26	<_315	>=_30	>_80	F
	-20	124	33	266	F		>=_-26	<_315	>=_30	>_80	F
	-42	950	13	14	S		<_-26	>=_315	<_30	<=_80	S
	-40	916	22	24	S		<_-26	>=_315	<_30	<=_80	S
	-60	650	16	25	S		<_-26	>=_315	<_30	<=_80	S
	-42	750	24	32	S		<_-26	>=_315	<_30	<=_80	S
	-30	500	19	38	S		<_-26	>=_315	<_30	<=_80	S
	-32	530	21	40	S		<_-26	>=_315	<_30	<=_80	S
	-24	300	24	80	S		>=_-26	<_315	<_30	<=_80	S
	-23	300	29	97	S		>=_-26	<_315	<_30	>_80	S
	-30	250	30	120	S		<_-26	<_315	>=_30	>_80	S

Considerando que os dados estão em formato categórico, foi necessário transformá-los em dados numéricos, com o módulo da biblioteca específica citada, conforme segue:

¹ Os métodos Naive Bayes são um conjunto de algoritmos de aprendizado supervisionados baseados na aplicação do teorema de Bayes com a suposição “naive” de independência condicional entre cada par de recursos dado o valor da variável de classe.

² Utilizados para transformação dos dados categóricos em numéricos.

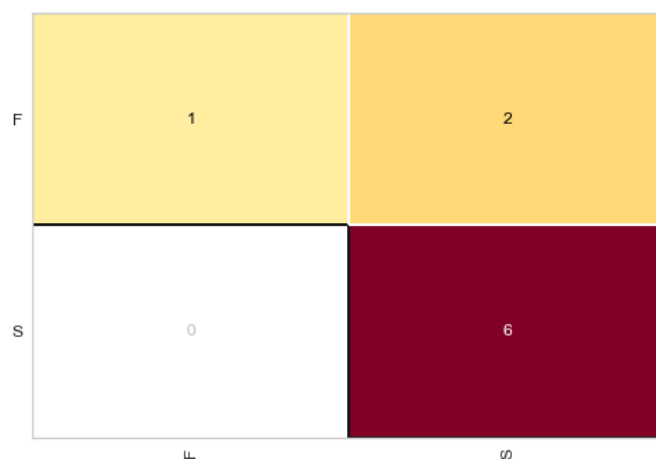
³ Biblioteca que pode ser utilizada para várias atividades e processos, entre eles: limpeza e tratamento de dados, análise exploratória de dados (EDA).

Tabela 3 - Dados preparados para treino da rede bayes

NIF	VT	RR	RSBI	Label
0	1	0	0	F
1	1	1	0	F
1	1	1	1	F
1	1	1	1	F
1	1	0	1	F
0	1	1	1	F
1	0	1	1	F
1	0	1	1	F
1	0	1	1	F
1	0	1	1	F
1	0	1	1	F
1	0	1	1	F
0	1	0	0	S
0	1	0	0	S
0	1	0	0	S
0	1	0	0	S
0	1	0	0	S
0	1	0	0	S
1	0	0	0	S
1	0	0	1	S
0	0	1	1	S

Após o treino da rede Naive Bayes (Anexo A- “rede_bayesiana.py”), segundo os dados de treino (Anexo B – “treino_bayseana.txt”) a previsão em relação aos dados de teste (Anexo C – “teste_bayseana.txt”), a rede apresentou os seguintes resultados, aqui representado no figura da matriz de confusão:

Figura 1- Gráfico com a Matriz Confusão da Previsão x Real



Desta figura podemos extrair as seguintes informações, das previsões para não retirada da ventilação (F), houve um verdadeiro positivo (TP), representado pelo quadro superior esquerdo. No quadro superior direito contam os falsos positivos (FP), ou seja,

indicaram que o paciente poderia ter a ventilação retirada quando não deveria. Desta parte podemos retirar a especificidade, de 1 para 3 casos possíveis, ou 33,3%.

No canto inferior esquerdo temos os casos falsos negativos (FN), onde não houve nenhum caso, já no canto inferior direito temos os casos de verdadeiro positivo (TP), desta parte inferior podemos extrair a sensibilidade de 6 para 6 casos possíveis, ou 100%.

Embora a rede Naive Bayes tenha obtido uma excelente sensibilidade, o caso em tela necessita que a especificidade fosse mais expressiva, já que é menos arriscado manter um paciente com ventilação quando não deveria do que retirar quando não devido.

Complementarmente para avaliar a sistemática da biblioteca em Python, foi analisado o mesmo caso, efetuando manualmente uma rede Naive Bayes em Microsoft Excel (Anexo D-“Rede NB Excel.xlsx”), onde foi obtido os mesmos resultados, comprovando a aderência da biblioteca em Python.

2 - DESENVOLVIMENTO DE UM SISTEMA DE INFERÊNCIA FUZZY PARA SUPORTE À TOMADA DE DECISÃO ACERCA DA INTERRUPÇÃO

Para o desenvolvimento de um sistema de inferência fuzzy para suporte à tomada de decisão acerca da interrupção/desmame da ventilação mecânica, segundo a proposição, foram consideradas as variáveis de entrada:

- e) NIF - pico negativo da pressão inspiratória (cm.H₂O);
- f) RR - taxa respiratória (respirações/minuto);
- g) VT – volume tidal (ml); e,
- h) RSBI - (Rapid Shallow Breathing Index).

A implementação considerou a aplicação do método de inferência do tipo Mamdani e a definição das funções de pertinência foi baseada nas informações coletadas no documento disponibilizado, denominado “Base de Conhecimento Weaning”.

Deste documento foi obtido as seguintes observações em relação às funções de pertinência:

- Volume Tidal (VT): sucesso acima de 315 ml, fracasso abaixo de 315 ml;
- RSBI: sucesso ≤ 80 BM/L, fracasso ≥ 120 BM/L;
- Taxa respiratória (RR) : sucesso < 30 R/M, fracasso ≥ 35 R/M;
- Pico negativo da pressão inspiratória (NIF): sucesso < -26 CM H₂O, fracasso > -20 CM H₂O.

2.1 Implementação

Para desenvolvimento do sistema foi utilizado a linguagem de programação Python em conjunto com a biblioteca específica para sistema de inferência fuzzy, denominada “skfuzzy”⁴ e as bibliotecas auxiliares “numpy”⁵ e “pandas”⁶, todo código, no qual serão

⁴ Biblioteca com coleção de algoritmos de lógica difusa.

⁵ Biblioteca de código aberto destinada a realizar operações em arrays multidimensionais.

⁶ Biblioteca que pode ser utilizada para várias atividades e processos, entre eles: limpeza e tratamento de dados, análise exploratória de dados (EDA).

abordados os resultados logo abaixo, estão disponibilizados em arquivo formato “py” e , disponibilizados de modo permanente no seguinte link: <https://github.com/lucasgcb/fuzzy-bayes>

Inicialmente não foi utilizado o RSBI para análise, sendo o mesmo adicionado posteriormente para comparação do resultado.

Os dados utilizados para definição da amplitude dos antecedentes ou entradas foram o conjunto de dados (com 30 elementos, sendo a soma do dataset de treino e teste disponibilizado para o trabalho do mapa de Kohonen) contendo os parâmetros das variáveis NIF, RR e VT e as saídas reais obtidas (S-Sucesso e F- Fracasso), conforme a tabela 4 abaixo:

Tabela 4 - Dataset respiradores

NIF	VT	RR	Label	RSBI
-24	300	24	S	80,00
-23	300	29	S	96,67
-60	650	16	S	24,62
-30	250	30	S	120,00
-42	950	13	S	13,68
-42	750	24	S	32,00
-32	530	21	S	39,62
-40	916	22	S	24,02
-30	500	19	S	38,00
-30	500	17	F	34,00
-26	265	44	F	166,04
-18	370	32	F	86,49
-22	350	30	F	85,71
-10	320	29	F	90,63
-20	225	46	F	204,44
-24	276	38	F	137,68
-15	270	40	F	148,15
-12	220	32	F	145,45
-20	124	33	F	266,13
-25	412	30	F	72,82
-28	400	38	F	95,00
-40	465	23	S	49,46
-15	450	25	F	55,56
-22	400	41	F	102,50
-28	310	24	F	77,42
-48	380	24	S	63,16
-34	530	28	S	52,83
-40	740	19	S	25,68
-42	550	27	S	49,09
-55	480	19	S	39,58

Destes dados (Anexo 1) extraímos e utilizaremos como referência:

- NIF: Mínimo = -60 (cm.H₂O), Máximo = -10 (cm.H₂O);
- RR: Mínimo = 13 (respirações/minuto), Máximo = 46 (respirações/minuto);
- VT: Mínimo = 124 (ml), Máximo = 950 (ml); e,
- RSBI: Mínimo = 13 (respirações/minuto /litro), Máximo = 267 (respirações/minuto /litro);

Com base nestes limites e com a pertinência baseada nas informações coletadas no documento disponibilizado, denominado “Base de Conhecimento Weaning”, foram criados os antecedentes ou entradas utilizando a função triangular e a saída ou consequente no qual apresentamos os gráficos nas Figura 2 e 2a seguir:

Figura 2- Função triangular aplicada às entradas

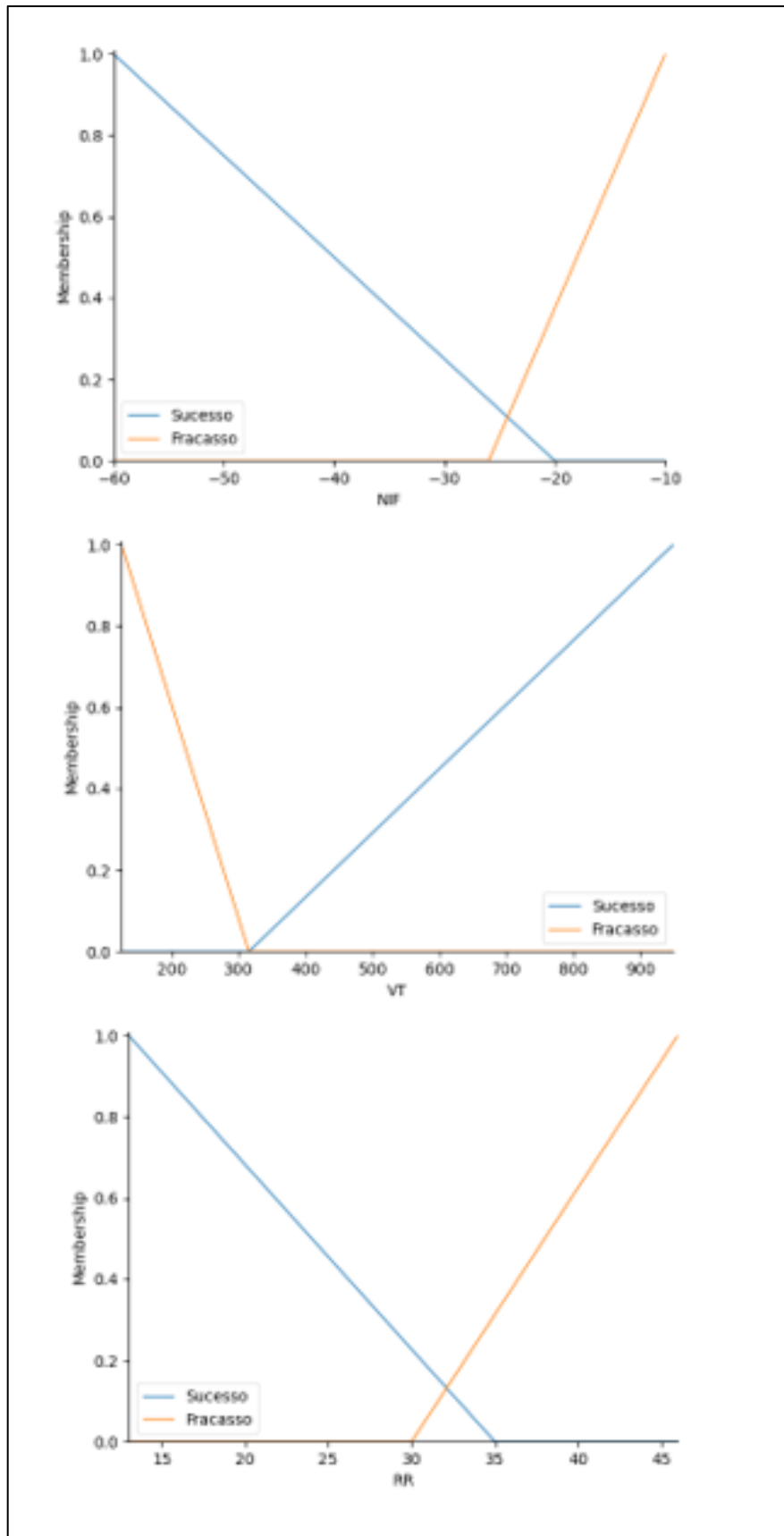
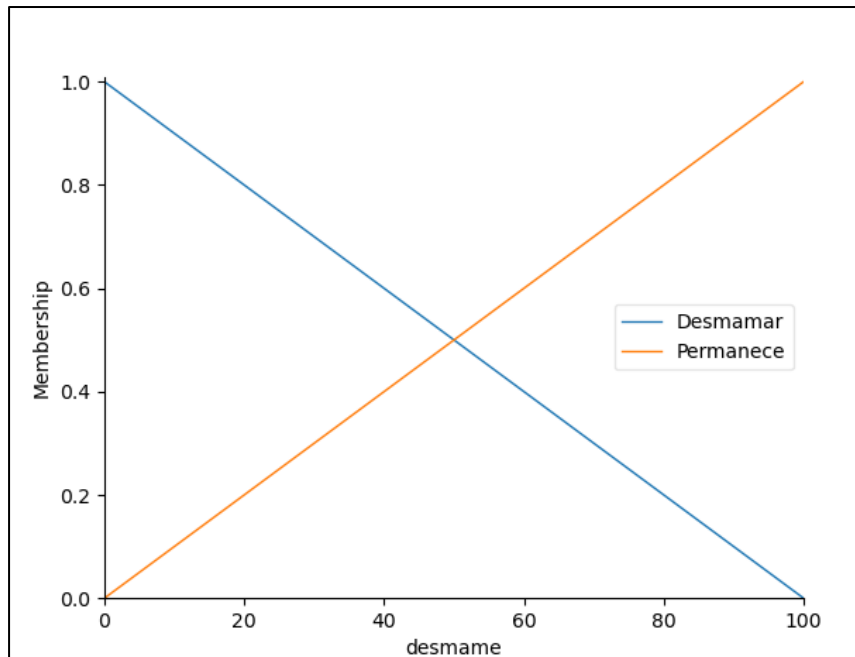


Figura 3- Função triangular aplicada à saída



Baseado nestas pertinências foram adotadas, inicialmente, as seguintes regras:

“regra1 = NIF = Fracasso e VT = Fracasso e RR = Fracasso então desmame = Permanece
regra2 = NIF = Fracasso e VT = Fracasso e RR = Sucesso então desmame = Permanece
regra3 = NIF = Fracasso e VT = Sucesso e RR = Fracasso então desmame = Permanece
regra4 = NIF = Sucesso e VT = Fracasso e RR = Fracasso então desmame = Permanece
regra5 = NIF = Sucesso e VT = Sucesso e RR = Sucesso então desmame = Desmamar
regra6 = NIF = Fracasso e VT = Sucesso e RR = Sucesso então desmame = Desmamar
regra7 = NIF = Sucesso e VT = Sucesso e RR = Fracasso então desmame = Desmamar
regra8 = NIF = Sucesso e VT = Fracasso e RR = Sucesso então desmame = Desmamar”

Com base nestes parâmetros e regras o modelo baseado em centroide obteve os seguintes resultados (Anexo 2 - “Fuzzy_trinf_inicial.py”), conforme tabela abaixo:

Tabela 5 - Resultados obtidos

Verdadeiros Positivos (TP) :	13	Verdadeiros Negativos (TN):	8
Total de positivos base (UP):	15	Total Negativos (UN):	15
Sensibilidade (TP/UP):	86,67%	Especificidade (TN/UN)	53%

Neste caso, entendemos que o resultado não foi satisfatório em razão de a especificidade ter sido baixa, como é um caso em que a retirada da ventilação forçada envolve risco de o paciente perder a vida, o ideal é que os Verdadeiros Negativos (TN) sejam o maior possível e consequentemente os Falsos Positivos sejam o menor possível, os Falsos Positivos (FP) são os casos em que indicam que o paciente deva ter a ventilação forçada retirada (desmame) quando na realidade não deveriam.

Com uma simples alteração nas regras (Anexo 3 - “Fuzzy_trinf_alt_regra.py”), fazendo que os casos de desmame ocorram apenas quando as 3 entradas indiquem sucesso (regras 6, 7 e 8 igual a “Permaneça”), resultou no aumento a especificidade, conforme tabela abaixo:

Tabela 6 - Resultados obtidos com alteração da regra

Verdadeiros Positivos (TP):	12	Verdadeiros Negativos (TN):	13
Total de positivos base (UP):	15	Total Negativos (UN):	15
Sensibilidade (TP/UP):	80%	Especificidade (TN/UN)	86,67%

Foi testado uma rede fuzzy considerando o RSBI” (Anexo 4 - “Geral_fuzzy_RSBI.py”) a compor mais uma entrada ou antecedente, com os dados ajustados (Anexo 5 – “Respira_DB_Tds_RSBI.txt”), e como consequência as seguintes regras:

```

regra1 = se NIF['Fracasso'] e VT['Fracasso'] e RR['Fracasso'] e RSBI['Fracasso'] então
desmame['Permaneça']
regra2 = se NIF['Fracasso'] e VT['Fracasso'] e RR['Fracasso'] e RSBI['Sucesso'] então
desmame['Permaneça']
regra3 = se NIF['Fracasso'] e VT['Fracasso'] e RR['Sucesso'] e RSBI['Fracasso'] então
desmame['Permaneça']
regra4 = se NIF['Fracasso'] e VT['Sucesso'] e RR['Fracasso'] e RSBI['Fracasso'] então
desmame['Permaneça']
regra5 = se NIF['Sucesso'] e VT['Fracasso'] e RR['Fracasso'] e RSBI['Fracasso'] então
desmame['Permaneça']
regra6 = se NIF['Fracasso'] e VT['Fracasso'] e RR['Sucesso'] e RSBI['Sucesso'] então
desmame['Permaneça']
regra7 = se NIF['Fracasso'] e VT['Sucesso'] e RR['Fracasso'] e RSBI['Sucesso'] então
desmame['Permaneça']
regra8 = se NIF['Sucesso'] e VT['Fracasso'] e RR['Fracasso'] e RSBI['Sucesso'] então
desmame['Permaneça']
regra9 = se NIF['Sucesso'] e VT['Sucesso'] e RR['Fracasso'] e RSBI['Fracasso'] então
desmame['Permaneça']
regra10 = se NIF['Sucesso'] e VT['Fracasso'] e RR['Sucesso'] e RSBI['Fracasso'] então
desmame['Permaneça']
regra11 = se NIF['Fracasso'] e VT['Sucesso'] e RR['Sucesso'] e RSBI['Fracasso'] então
desmame['Permaneça']
regra12 = se NIF['Fracasso'] e VT['Sucesso'] e RR['Sucesso'] e RSBI['Sucesso'] então
desmame['Desmamar']
regra13 = se NIF['Sucesso'] e VT['Sucesso'] e RR['Fracasso'] e RSBI['Sucesso'] então
desmame['Desmamar']
regra14 = se NIF['Sucesso'] e VT['Fracasso'] e RR['Sucesso'] e RSBI['Sucesso'] então
desmame['Desmamar']
regra15 = se NIF['Sucesso'] e VT['Sucesso'] e RR['Sucesso'] e RSBI['Fracasso'] então
desmame['Desmamar']
regra16 = se NIF['Sucesso'] e VT['Sucesso'] e RR['Sucesso'] e RSBI['Sucesso'] então
desmame['Desmamar']

```

Neste caso houve uma pequena melhora na especificidade em relação ao modelo anterior sem ajuste de regra:

Tabela 7 - Resultados obtidos com RSBI

Verdadeiros Positivos (TP) :	12	Verdadeiros Negativos (TN):	9
Total de positivos base (UP):	15	Total Negativos (UN):	15
Sensibilidade (TP/UP):	80%	Especificidade (TN/UN)	60%

Já com ajuste de regra, deixando apenas a regra 16 como “Desmamar” (Anexo 6 - “Geral_fuzzy_RSBI_ajustado.py”), o resultado foi o seguinte:

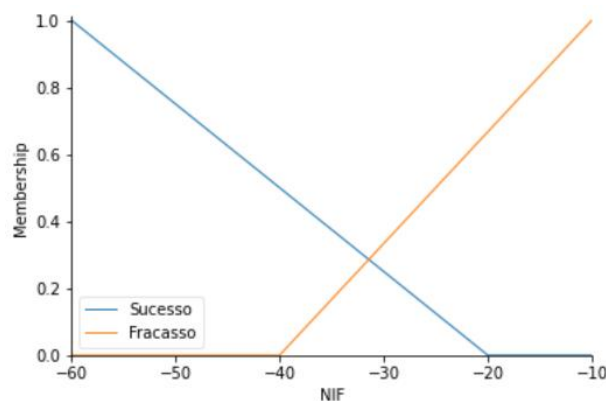
Tabela 8 - Resultados obtidos com RSBI com ajuste de regra

Verdadeiros Positivos (TP):	12	Verdadeiros Negativos (TN):	13
Total de positivos base (UP):	15	Total Negativos (UN):	15
Sensibilidade (TP/UP):	80%	Especificidade (TN/UN)	86,67%

Ao analisar os 2 casos de falso positivos, observamos que são casos em que todos os indicadores ou entradas estão dentro dos parâmetros indicados pela literatura para que seja efetuada a retirada da ventilação forçada, contudo naqueles casos ocorreu que não houve sucesso com a retirada.

Percebemos que alterar os parâmetros das entradas ou antecedentes poderia aumentar a especificidade, por exemplo, alteramos os parâmetros do NIF para fracasso quando maior que -20CM H₂O para maior que -40CM H₂O:

Figura 4- Alteração do NIF para aumento da área de Fracasso



Neste caso entendemos que o resultado foi positivo, pois obteve-se um ganho de dois elementos da especificidade, atendendo o objetivo de 100% ao custo de apenas 1 elemento da sensibilidade, conforme segue:

Tabela 9 - Resultados obtidos com alteração de parâmetro de entrada NIF

Verdadeiros Positivos (TP):	11	Verdadeiros Negativos (TN):	15
Total de positivos base (UP):	15	Total Negativos (UN):	15
Sensibilidade (TP/UP):	73,3%	Especificidade (TN/UN)	100%

3 - COMPARAR OS RESULTADOS OBTIDOS COM A REDE NAIVE BAYES, SISTEMA FUZZY E MAPA DE KOHONEN DE VENTILAÇÃO MECÂNICA ASSISTIDA

O resultado do trabalho relativo ao mapa de Kohonen, como havia conjunto treino (com 21 elementos) e conjunto teste com 9 (nove) elementos para a validação do modelo, neste cenário o resultado do modelo para o mapa de Kohonen foi, para o conjunto teste:

Tabela 10 - Resultado obtidos do conjunto teste do mapa de kohonen treinado

Verdadeiros Positivos (TP):	3	Verdadeiros Negativos (TN):	3
Total de positivos base (UP):	6	Total Negativos (UN):	3
Sensibilidade (TP/UP):	50%	Especificidade (TN/UN)	100%

A rede Naive Bayes por sua vez apresentou um resultado quase inverso no que diz respeito as Sensibilidade e Especificidade, ou seja, foi eficiente na apuração de verdadeiros positivos:

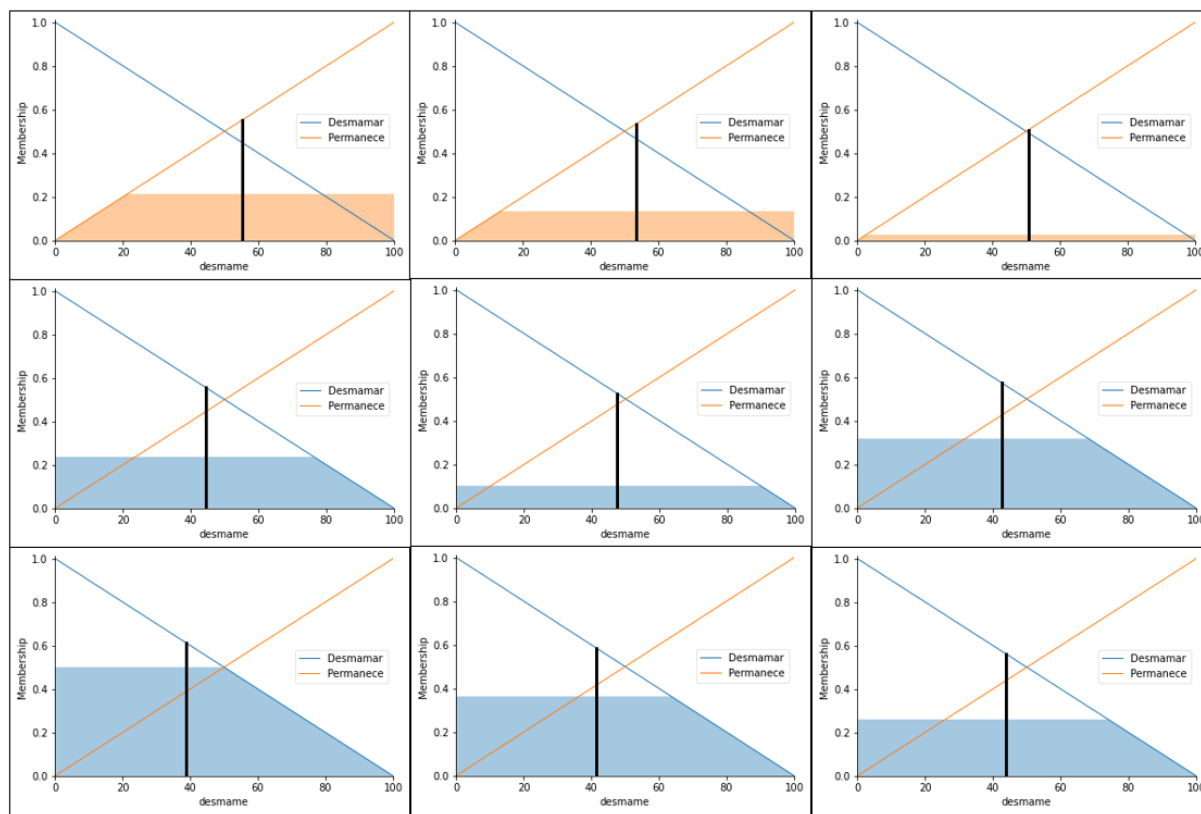
Verdadeiros Positivos (TP):	6	Verdadeiros Negativos (TN):	1
Total de positivos base (UP):	6	Total Negativos (UN):	3
Sensibilidade (TP/UP):	100%	Especificidade (TN/UN)	33%

Para termos a mesma base de comparação, já que para a implementação do sistema fuzzy demonstrado acima foi utilizado informações exógenas ao conjunto de dados de treino e teste para definição dos antecedentes, foi utilizado apenas o conjunto de dados teste (Anexo 7 – “Respira_DB_Tds_RSBI_teste.txt”) aplicado aos parâmetros de antecedentes e regras ajustadas contendo o RSBI (Anexo 8 - “Geral_fuzzy_RSBI_ajustado_dados_teste.py”), neste cenário a lógica fuzzy obteve o seguinte resultado:

Tabela 11 – Resultado da lógica fuzzy sobre os dados teste

Verdadeiros Positivos (TP):	6	Verdadeiros Negativos (TN):	3
Total de positivos base (UP):	6	Total Negativos (UN):	3
Sensibilidade (TP/UP):	100%	Especificidade (TN/UN)	100%

Figura 5- Saídas da Rede Fuzzy para os 9 casos de teste, laranjas devem permanecer e os azuis devem executar o procedimento



Deste modo, observamos que o sistema de inferência Fuzzy se sobressai sobre os demais segundo as métricas aqui utilizadas, destacamos que os parâmetros para classificação das entradas podem ser ajustados, porém quando efetuados existe uma migração de casos entre sensibilidade e especificidade,

Concluimos que cabe ao tomador de decisão a melhor escolha sobre os parâmetros para classificação das entradas do problema em questão e que cada tipo de técnica escolhida, neste caso Mapa de Kohonen, Inferência Fuzzy e Naive Bayes possui resultados distintos que podem ser utilizados como parâmetros em forma de comitê para a tomada de decisão acerca da retirada da ventilação mecânica ou não.

Referências Bibliográficas

[Scikit-learn: aprendizado de máquina em Python](#) , Pedregosa *et al.* , JMLR 12, pp. 2825-2830, 2011.

ASHUTOSH, K., L. et all. (1992). Prediction criteria for successful weaning from respiratory support: statistical and connectionist analyses. Critical care medicine, 1992, 1295–1301.

<https://doi.org/10.1097/00003246-199209000-00017>