

UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

KELLEN ADRIELY ALVARENGA GUIMARÃES

PREDIÇÃO DE MORTALIDADE DE PACIENTES COM TRAUMATISMO
CRANIOENCEFÁLICO NO BRASIL USANDO TÉCNICAS DE APRENDIZAGEM DE
MÁQUINA

MANAUS

2022

KELLEN ADRIELY ALVARENGA GUIMARÃES

PREDIÇÃO DE MORTALIDADE DE PACIENTES COM TRAUMATISMO
CRANIOENCEFÁLICO NO BRASIL USANDO TÉCNICAS DE APRENDIZAGEM DE
MÁQUINA

Dissertação apresentada ao Curso de Mestrado em Engenharia Elétrica, área de concentração Controle e Automação de Sistemas e linha de pesquisa Reconhecimento de Padrões e Otimização do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas.

Orientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho
Coorientadora: Prof.^a Dra. Marly Guimarães Fernandes Costa

MANAUS

2022

Ficha Catalográfica

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

G963p Guimarães, Kellen Adriely Alvarenga
Predição de mortalidade de pacientes com traumatismo cranioencefálico no Brasil usando técnicas de aprendizagem de máquina / Kellen Adriely Alvarenga Guimarães . 2022
89 f.: il. color; 31 cm.

Orientador: Cícero Ferreira Fernandes Costa Filho
Coorientadora: Marly Guimarães Fernandes Costa
Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Amazonas.

1. Traumatismo cranioencefálico. 2. Aprendizagem de máquina.
3. Redes neurais. 4. Redes neurais convolucionais. I. Costa Filho, Cícero Ferreira Fernandes. II. Universidade Federal do Amazonas
III. Título

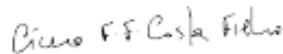
KELLEN ADRIELY ALVARENGA GUIMARÃES

**PREDIÇÃO DE MORTALIDADE DE PACIENTES COM
TRAUMATISMO CRANIOENCEFÁLICO NO BRASIL USANDO
TÉCNICAS DE APRENDIZAGEM DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Aprovada em 18 de fevereiro de 2022.

BANCA EXAMINADORA



Prof. Dr. Cícero Ferreira Fernandes Costa Filho, Presidente
Universidade Federal do Amazonas



Prof. Dr. Robson Luís Oliveira de Amorim, Membro
Universidade Federal do Amazonas



Prof. Dr. José Raimundo Gomes Pereira, Membro
Universidade Federal do Amazonas

AGRADECIMENTOS

Agradeço a mim, por não desistir quando tudo pareceu tão difícil, por acreditar que a educação tem o poder de transformação no mundo em que vivemos.

Agradeço aos meus orientadores, Prof. Dr. Cícero Ferreira Fernandes Costa Filho e Prof.^a Dra. Marly Guimarães Fernandes Costa, pela compreensão, apoio e por terem me dado a honra de ser sua orientanda e pelos inúmeros aprendizados nessa jornada.

À Universidade Federal do Amazonas e, em especial, ao Centro de Tecnologia Eletrônica e da Informação – CETELI – pela concessão de toda infraestrutura para realização deste trabalho. Esta pesquisa, conforme previsto no Art. 48 do decreto nº 6.008/2006, foi financiada pela Samsung Eletrônica da Amazônia Ltda, nos termos da Lei Federal nº 8.387/1991, através de convênio nº 004, firmado com o Centro de P&D em Eletrônica e Tecnologia da Informação da Universidade Federal do Amazonas - CETELI / UFAM.

RESUMO

Esse trabalho propõe, de forma original, a utilização de redes neurais convolutivas para a predição da mortalidade em até 14 dias de pacientes com trauma crânio encefálico. O desempenho das redes neurais é comparado com o desempenho de outras ferramentas clássicas de aprendizado de máquina, como regressor logístico, perceptron multicamadas, máquina de vetores de suporte, árvores de decisão e floresta randômica. Na simulação dos modelos utilizando redes neurais, diversos métodos de otimização foram utilizados, como RMSProp, Adam, Adamax e SGDM. A base de dados utilizadas é constituída de 529 registros e 16 variáveis preditoras, tendo sido obtida no Hospital das Clínicas (São Paulo, Brasil). Devido a presença de muitos valores ausentes nas variáveis preditoras propõe-se e avaliou-se dois procedimentos para o preenchimento dos valores ausentes das mesmas, utilizando métodos como árvore de decisão, floresta randômica, k-vizinho mais próximo e regressão linear. Os melhores resultados obtidos para a taxa de predição foram precisão de 0,845 e área sob a curva ROC de 0,911.

Palavras-chave: Traumatismo cranioencefálico, aprendizagem de máquina, redes neurais, redes neurais convolucionais.

ABSTRACT

This work proposes, in an original way, the use of convolutional neural networks for the prediction day mortality until the 14th day in patients with traumatic brain injury. The performance of neural networks is compared with the performance of other classic machine learning tools such as logistic regressor, multilayer perceptron, support vector machine, decision trees and random forest. In the simulation of models using neural networks, several optimization methods were used, such as RMSProp, Adam, Adamax and SGDM. The database used consists of 529 records and 16 predictor variables, and was obtained from Hospital das Clínicas (São Paulo, Brazil). Due to the presence of many missing values in the predictor variables, two procedures were proposed and evaluated for filling in the missing values, using several methods, such as decision tree, random forest, k-nearest neighbor and linear regression. The best results obtained for the prediction rate were an accuracy of 0.845 and an area under the ROC curve of 0.911.

Keywords: Traumatic brain injury, machine learning, neural networks, convolutional neural networks.

LISTA DE FIGURAS

Figura 3.1 – Estrutura de uma árvore de decisão	36
Figura 3.2 – Processo de Classificação Floresta Randômica	37
Figura 3.3 – Gráfico k-vizinhos mais próximos	38
Figura 3.4 – Gráfico Máquina de Vetores de Suporte	38
Figura 3.5 – Estrutura básica de um neurônio artificial	40
Figura 3.6 – Operação de <i>pooling</i>	43
Figura 4.1 – Recorte das 5 primeiras observações da base de dados utilizada.....	48
Figura 4.2 – (a) Fluxograma das Etapas da Metodologia; (b) Pré-processamento; (c) Treinamento e Avaliação dos Modelos Preditivos	50
Figura 4.3 – Arquitetura da Rede MLP1	55
Figura 4.4 – Arquitetura da Rede MLP2	55
Figura 4.5 – Arquitetura da Rede MLP3	56
Figura 4.6 – Arquitetura da Rede MLP4	56
Figura 4.7 – Arquitetura da Rede CNN1	57
Figura 4.8 – Arquitetura da Rede CNN2.....	58
Figura 4.9 – Gráfico mostrando o desbalanceamento das classes.....	59
Figura 5.1 – Recorte das 5 primeiras observações do conjunto de dados após preenchimento dos valores ausentes	66
Figura 5.2 – Gráfica de AUC – ROC de MLP4	72

LISTA DE TABELAS

Tabela 4.1 – Características das 17 variáveis utilizadas nos métodos de predição	50
Tabela 4.2 – Range de valores das variáveis selecionadas	53
Tabela 4.3 – Parâmetros utilizados no treinamento das redes neurais	58
Tabela 4.4 – Modelos utilizados para cada tipo de predição	61
Tabela 5.1 – Acurácia dos modelos de classificação para as variáveis categóricas, utilizando o Procedimento 1	64
Tabela 5.2 – MSE dos modelos de regressão para as variáveis numéricas, utilizando o Procedimento 1	65
Tabela 5.3 – Acurácia dos modelos de classificação para as variáveis categóricas, utilizando o Procedimento 2	65
Tabela 5.4 – MSE dos modelos de regressão para as variáveis numéricas, utilizando o Procedimento 2	66
Tabela 5.5 – Área sob a curva ROC (AUC) obtida para o modelo MLP1 para diferentes técnicas de normalização dos dados	67
Tabela 5.6 – Comparação inicial dos modelos sem o balanceamento de classes	67
Tabela 5.7 – Avaliação inicial dos modelos utilizando o balanceamento de classes para MLP1 .	68
Tabela 5.8 – Resultados para a predição da mortalidade obtidos com o modelo MLP1 usando diversos métodos de otimização	68
Tabela 5.9 – Resultados para a predição da mortalidade obtidos com o modelo MLP2 usando diversos métodos de otimização	69
Tabela 5.10 – Resultados para a predição da mortalidade obtidos com o modelo MLP3 usando diversos métodos de otimização	69
Tabela 5.11 – Resultados para a predição da mortalidade obtidos com o modelo MLP4 usando diversos métodos de otimização	69
Tabela 5.12 – Resultados para a predição da mortalidade obtidos com o modelo CNN1 usando diversos métodos de otimização	69
Tabela 5.13 – Resultados para a predição da mortalidade obtidos com o modelo CNN2 usando diversos métodos de otimização	70

LISTA DE SIGLAS

AUC	Área sob a curva (do inglês, <i>Area Under the Curve</i>)
ROC	Curva Característica de Operação do Receptor
ECG	Escala de coma de Glasgow
TCE	Trauma cranioencefálico
PBMR	Países de baixa e média renda

SUMÁRIO

1	INTRODUÇÃO.....	15
1.1	Objetivo Geral.....	17
1.2	Objetivos Específicos.....	17
1.3	Organização do Trabalho.....	18
2	REVISÃO BIBLIOGRÁFICA.....	19
2.1	<i>Application of machine learning to predict the outcome of pediatric traumatic brain injury (TUNTHANNATHIP e OEARSAKUL, 2021)</i>	19
2.2	<i>Prediction of Early TBI Mortality Using a Machine Learning Approach in a LMIC Population (AMORIM et al., 2020)</i>	20
2.3	<i>Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries (FENG et al., 2019)</i>	21
2.4	<i>Machine learning-based dynamic mortality prediction after traumatic brain injury (RAJ et al., 2019)</i>	22
2.5	<i>Using and artificial neural network to predict traumatic brain injury (HALE et al., 2018)</i> 22	
2.6	<i>Predicting outcome after traumatic brain injury: practical prognostic models based on a large cohort of international patients (COLLABORATORS et al., 2008)</i>	23
2.7	<i>Predicting Outcome after Traumatic Brain Injury: Development and International Validation of a Prognostic Score Based on Admission Characteristics (STEYERBERG et al., 2005)</i> 24	
2.8	<i>Neural Network Modeling for surgical decisions on traumatic brain injury patients (LI et al., 2000)</i>	25
2.9	<i>Predicting survival using simple clinical variables: a case study in traumatic brain injury (SIGNORINI et al., 1999)</i>	25
3	REFERENCIAL TEÓRICO.....	31
3.1	Traumatismo Cranioencefálico.....	31

3.2	Pré-processamento dos Dados	32
3.2.1	Normalização de dados.....	33
3.3	Aprendizado Supervisionado.....	36
3.3.1	Árvore de Decisão	36
3.3.2	Floresta Randômica	37
3.3.3	K-Vizinhos Mais Próximos	37
3.3.4	Máquina de Vetores de Suporte.....	38
3.3.5	Regressão Logística.....	39
3.3.6	Regressão Linear	39
3.4	Redes Neurais Artificiais	40
3.5	Redes Neurais Convolucionais	41
3.5.1	Camada Convolutiva	42
3.5.2	Camada <i>Pooling</i> (subamostragem).....	42
3.5.3	Camada <i>Dropout</i>	43
3.5.5	Regularização L2.....	44
3.6	Métodos de Otimização	44
3.6.1	Estimativa Dinâmica Adaptativa (Adam).....	44
3.6.2	AdaMax	45
3.6.3	Propagação da Raiz Média Quadrática (RMSProp).....	45
3.6.4	Gradiente Descendente Estocástico com Momento (SGDM)	46
4	MATERIAIS E MÉTODOS.....	47
4.1	Materiais	47
4.1.1	Base de Dados	47
4.1.2	Divisão da Base de Dados	48
4.2	Métodos.....	49

4.2.1	Pré-processamento.....	50
4.2.2	Avaliação de Modelos	54
4.2.3	Métricas de Avaliação	61
4.2.4	Ambiente de Desenvolvimento	63
5	RESULTADOS E DISCUSSÃO.....	64
5.1	Resultados do Preenchimento dos Valores Ausentes	64
5.2	Resultados da Normalização de Variáveis.....	66
5.3	Resultados dos Métodos de Otimização para as Redes Neurais.....	67
5.4	Discussão	70
5.5	Comparação dos Resultados obtidos com a Literatura.....	71
6.	CONCLUSÃO.....	72
5	REFERÊNCIAS	74
	APÊNDICE	80
	INTRODUCTION	81
	MATERIALS AND METHODS.....	83
	A. <i>Materials</i>	83
	B. <i>Methods</i>	83
	C. <i>Filling in missing values</i>	83
	D. <i>Data normalization</i>	84
	D. <i>Definition of machine learning models</i>	85
	RESULTS	87
	A. <i>Results of filling in missing values of variables</i>	87
	B. <i>Results of the normalization process</i>	87
	C. <i>Results of 14-day mortality prediction</i>	87
	DISCUSSION.....	88

CONCLUSION 89

REFERENCES 89

1 INTRODUÇÃO

Atualmente, existem 4,4 milhões de mortes relacionadas a lesões em todo o mundo. Dessas mortes, 3,16 milhões estão relacionadas a lesões não intencionais. No geral, 1 em cada 3 dessas mortes estão relacionadas a acidentes de trânsito. Segundo a Organização Mundial da Saúde (OMS), as mortes decorrentes de lesões que ocorrem nas estradas são a 7ª causa de morte no mundo, totalizando 189.000 óbitos (Organização Mundial da Saúde, 2021b). As lesões no trânsito são agora a principal causa de morte de pessoas com idade entre 5 e 29 anos (Organização Mundial da Saúde, 2018).

A pobreza aumenta o risco de lesão de uma pessoa. Cerca de 90% dos casos de lesões ocorrem em países de baixa e média renda (PBMR), onde vive 85% da população (Organização Mundial da Saúde, 2021a).

A distribuição desigual das lesões, com maior prevalência nos países mais pobres, está relacionada a uma série de fatores de risco, como: morar, trabalhar, viajar e ir à escola em condições mais precárias e menor acesso a serviços de atendimento emergencial de trauma e reabilitação de qualidade (Organização Mundial da Saúde, 2021a).

Entre as lesões, o trauma cranioencefálico (TCE) é responsável por um terço a metade de todas as mortes por lesão e é a principal causa de incapacidade em pessoas com menos de quarenta anos (WHITENECK et al., 2006). O TCE é responsável por uma grande carga de morbidade, mortalidade, incapacidade, perdas socioeconômicas, redução da expectativa de vida e qualidade de vida, principalmente em PBMR (UMERANI et al., 2014) (KAMAL et al., 2016) (HOFMAN et al., 2005).

Os médicos normalmente tomam decisões terapêuticas com base no prognóstico do paciente. De acordo com Perel et al. (2007), 80% dos médicos acreditam que é importante ter um prognóstico confiável ao tomar decisões pertinentes sobre o uso específico de algumas terapias, como hiperventilação, barbitúricos ou manitol. A utilização de prognósticos fornecidos por ferramentas computacionais em pacientes com TCE favorece o uso de algumas terapias intervencionistas naqueles pacientes com bom prognóstico e reduz seu uso naqueles pacientes com prognóstico ruim (MURRAY et al., 1993).

Com relação a TCE, as classes prognósticas mais utilizadas são: 1. Mortalidade em 14 dias (AMORIM et al., 2020); 2. Mortalidade em 30 dias (RAJ et al., 2019); 3. sobrevida do paciente um ano após a alta (HUKKELHOVEN et al., 2000); 3. mortalidade, sobrevida (estado vegetativo ou incapacidade grave), desfecho favorável (incapacidade média ou boa recuperação); 4. Decisão de realizar cirurgia aberta de crânio (STEYERBERG et al., 2008); 5. TCE Clinicamente Relevante - CRTBI: a) procedimento neurocirúrgico, b) intubação > 24 horas como resultado direto de traumatismo cranioencefálico, c) internação maior que 48 horas e evidência de traumatismo cranioencefálico através de tomografia computadorizada, e d) óbito por trauma cranioencefálico (HALE et al., 2019); 6. Resultado KOSCHI após 6 meses da alta hospitalar, compreendendo as seguintes categorias: a) óbito, b) estado vegetativo, c) incapacidade moderada, d) boa recuperação (TUNTANATHIP & OEARSAKUL, 2021).

As técnicas de aprendizado de máquina mais usadas para previsão são máquinas de vetor de suporte (TUNTHANATHIP & OEARSAKUL, 2021), redes neurais perceptron multicamadas (HALE et al., 2019; LI et al., 2000), função de base radial (RBF) (LI et al., 2000), regressão logística (HUKKELHOVEN et al., 2000; LI et al., 2000; STEYERBERG et al., 2008; FENG et al., 2019; RAJ et al., 2019), Naive-Bayes (AMORIM et al., 2020), modelo linear Bayesiano generalizado (AMORIM et al., 2020), análise discriminante penalizada (AMORIM et al., 2020) e floresta randômica (AMORIM et al., 2020).

Os trabalhos publicados utilizaram diversas bases de dados proprietárias e diferentes variáveis preditoras.

No trabalho de Amorim et al. (2020), os autores utilizaram um banco de dados com 517 pacientes do Hospital das Clínicas da Universidade de São Paulo, Brasil. Para a constituição do banco de dados foram utilizados alguns critérios: inclusão de pacientes com idade a partir de 14 anos, com anormalidade intracraniana em tomografia computadorizada inicial e exclusão de pacientes com trauma penetrante, bem como aqueles com escala de coma de Glasgow 15 e sem associação de lesões intracranianas quando realizada tomografia. Foram selecionadas as seguintes variáveis preditoras: sexo, idade, nível de reatividade pupilar na admissão, ECG (Escala de Coma de Glasgow) no local do trauma (ECG pré-hospitalar), ECG na admissão, escore do componente motor da ECG e presença de hipóxia e hipotensão. Também foram incluídos desvio da linha média maior que 5 mm, herniação cerebral detectada na TC (definida como apagamento do terceiro

ventrículo ou das cisternas basais), hemorragia subaracnóidea, hemorragia epidural, hemorragia subdural, hemorragia intracerebral, gravidade do trauma, tempo de protrombina e tromboplastina parcial Tempo. Os autores utilizaram os seguintes preditores: Naive-Bayes, modelo Linear Bayesiano generalizado, análise discriminante penalizada e floresta aleatória. A avaliação dos preditores foi realizada por meio de métricas como área sob a curva ROC, sensibilidade e especificidade. Os melhores resultados alcançados pelos autores para a predição da mortalidade em 14 dias foram obtidos pela técnica de Naive Bayes, com área sob a curva ROC, $AUC = 0,906$, seguida do modelo Linear Bayesiano generalizado, com $AUC = 0,881$.

Até onde sabemos, nenhum trabalho utilizou redes neurais profundas para o prognóstico de pacientes com TCE. CNNs são normalmente redes 2D usadas para processamento de imagens, principalmente para classificação de imagens e segmentação de objetos (AGGARWAL, 2018; JIANG, 2019). No entanto, CNNs 1D também foram usadas com sucesso em muitas aplicações médicas, como detecção de doença de Parkinson (PRINCE e DE VOS, 2018; ZHANG et al., 2020) e detecção de arritmias (PLAWIAK, 2018; ZHENG et al., 2020). Neste trabalho, utilizando a mesma base de dados e as mesmas variáveis preditoras utilizadas por Amorim et al. (2020), comparamos o desempenho uma rede neural profunda, mais precisamente uma rede neural convolucional (CNN), para a previsão de mortalidade em 14 dias em pacientes com TCE, com outros modelos de aprendizado de máquina, como perceptron multicamadas (*Multilayer Perceptron* – MLP), árvores de decisão, regressão logística, máquina de vetores de suporte e floresta randômica.

1.1 Objetivo Geral

Comparar o desempenho de redes neurais profundas, mais especificamente de redes convolutivas, com técnicas clássicas de aprendizado de máquina, na revisão de mortalidade em 14 dias em pacientes com TCE.

1.2 Objetivos Específicos

1. Avaliar o desempenho de diferentes técnicas de preenchimento de dados ausentes na predição da mortalidade de pacientes com trauma cranioencefálico.

2. Avaliar o desempenho de diferentes técnicas de normalização de dados na predição da mortalidade de pacientes com trauma cranioencefálico.
3. Realizar benchmarking com o trabalho de Amorim et al. (2020).

1.3 Organização do Trabalho

Este trabalho está organizado conforme a divisão descrita a seguir:

- Capítulo 1: Introdução;
- Capítulo 2: Revisão Bibliográfica;
- Capítulo 3: Referencial Teórico;
- Capítulo 4: Materiais e Métodos;
- Capítulo 5: Resultados e discussão.

O Capítulo 1 apresenta o contexto do trabalho, citando a importância da predição de mortalidade de pacientes com trauma cranioencefálico para a área médica, uma breve revisão da literatura sobre tema, e os objetivos geral e específicos deste trabalho.

O Capítulo 2 apresenta trabalhos que foram publicados na literatura na área de trauma cranioencefálico por meio de técnicas de aprendizagem de máquina, usando algoritmos como regressão logística e redes neurais.

O Capítulo 3 apresenta os principais fundamentos teóricos utilizados no desenvolvimento deste trabalho: trauma cranioencefálico, pré-processamento de dados, redes neurais artificiais, redes neurais convolucionais e os métodos de otimização avaliados.

O Capítulo 4 apresenta os materiais utilizados nessa proposta de projeto, as características do banco de dados, o pré-processamento dos dados, as arquiteturas propostas e o fluxo do treinamento, validação e teste.

O Capítulo 5 apresenta os resultados preliminares alcançados.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo são apresentadas as técnicas descritas na literatura para a predição de mortalidade de pacientes que tiveram traumatismo cranioencefálico. Para fins de comparação com os materiais e a metodologia do presente trabalho, os seguintes aspectos foram analisados em cada trabalho: o banco de dados, variáveis preditoras, modelo de aprendizagem de máquina e métricas de desempenho empregadas. Além disso, dedicou-se especial atenção às técnicas empregadas na predição.

2.1 *Application of machine learning to predict the outcome of pediatric traumatic brain injury* (TUNTHANNATHIP e OEARSAKUL, 2021)

Nesse estudo, os autores utilizaram uma base de dados com registros de crianças com idade menor do que 15 anos, que foram admitidas no centro de trauma do sudeste da Tailândia, entre o período de Janeiro de 2009 a Julho de 2020. Foram excluídos do estudo pacientes que não fizeram a tomografia do cérebro, morreram nas primeiras 24 horas, que não tiveram todos os registros coletados durante a admissão ou que não foi possível fornecer resultados atualizados. Um total de 828 registros de pacientes foram incluídos no estudo. As variáveis preditoras incluídas no estudo foram 14: idade, gênero, mecanismo da lesão, pontuação média de gravidade de lesão, lesões relacionadas, comorbidades, ECG, reflexo da pupila à luz, fratura do crânio, lesões intracranianas, cisterna basal, deslocamento da linha média, tratamento cirúrgico, resultado *KOSCHI* (do inglês, *King's Outcome Scale for Childhood Head Injury*) na alta hospitalar. A variável predita considerada foi o parâmetro *KOSCHI* após 6 meses da alta hospitalar, compreendendo as seguintes categorias: 1) morte, 2) estado vegetativo, 3) incapacidade moderada, 4) boa recuperação. Os modelos de máquina de vetor de suporte, rede neural, *Naive Bayes*, regressão logística e vizinhos mais próximos foram construídos, utilizando validação cruzada com 10-*pastas*. O desempenho de cada algoritmo foi acessado por meio das seguintes métricas: sensibilidade, especificidade, valor positivo predito, valor negativo predito, acurácia, F1-score, precisão média. O modelo máquina de vetores de suporte obteve o melhor desempenho: sensibilidade = 0,95, especificidade = 0,60, valor positivo predito = 0.99, valor negativo predito = 1,0, acurácia 0.94 e AUC = 0,78.

2.2 *Prediction of Early TBI Mortality Using a Machine Learning Approach in a LMIC Population* (AMORIM et al., 2020)

Neste trabalho, os autores efetuaram uma coleta de dados de pacientes do hospital das Clínicas, da Universidade de São Paulo. Para inclusão de um paciente no estudo, os seguintes critérios foram utilizados: inclusão de pacientes com idade a partir de 14 anos, com anormalidade intracraniana em uma tomografia inicial computadorizada, exclusão de pacientes com trauma penetrante, assim como aqueles com escala de coma de Glasgow de 15 e sem associação de lesões intracranianas quando realizada a tomografia. Um total de 517 pacientes foram incluídos. Os autores, então selecionaram 15 variáveis preditoras: gênero, idade, nível de reação da pupila na admissão, ECG na cena onde o trauma ocorreu, ECG na admissão, o escore do componente motor do ECG, presença de hipóxia, hipotensão, mudança da linha média maior que 5mm, herniação cerebral identificado na tomografia, hemorragia subaracnóidea, hemorragia epidural, severidade do trauma, tempo de protrombina e tempo parcial de tromboplastina. Em seguida, uma análise de dados foi realizada para avaliar a frequência, porcentagem, variação próxima de zero para todas as variáveis categóricas, distribuição para as variáveis numéricas, valores ausentes e padrões para todas as variáveis. Para treinar e testar os modelos, o método de validação cruzada de 5 pastas foi utilizado. Os modelos de aprendizado de máquina utilizados para predição de saídas numéricas foram *random forest*, redes neurais, árvore de decisão, *boosting*, modelo linear generalizado, mínimos quadrados parciais (do inglês, *partial least squares*) e *spline* de regressão adaptativa multivariada. Por sua vez, os modelos usados para classificação foram: *random forest*, análise discriminante, métodos bayesianos, redes neurais, mínimos quadrados parciais (do inglês, *partial least squares*) e *spline* de regressão adaptativa multivariada. A comparação entre os modelos foi realizada utilizando métricas como área sob a curva ROC, sensibilidade e especificidade, entre outros. Os melhores resultados alcançados pelos autores para predição de mortalidade em 14 dias foram obtidos utilizando a técnica *naive Bayes*, com uma área sob a curva ROC, $AUC = 0,906$, seguido do modelo linear bayesiano generalizado, com $AUC = 0,881$.

2.3 Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries (FENG et al., 2019)

Neste trabalho, os autores coletaram dados de pacientes que tiveram traumatismo cranioencefálico severo admitidos no hospital Sichuan Provincial People no período de Dezembro de 2009 e Novembro de 2011, sendo considerados três critérios de inclusão de pacientes no estudo: idade entre 18 e 86 anos, histórico de lesão na cabeça e histórico de craniotomia. Foram considerados também três critérios de exclusão: gravidez, craniotomia em outros hospitais e hospitalização por um período menor do que 24 horas. A base de dados compreende um total de 117 registros de pacientes.

Um total de 40 variáveis preditoras foram utilizadas nesse estudo: idade, ECG (hospitalização), pontuação de gravidade da lesão temperatura, pressão sistólica, pressão diastólica lesão cerebral aberta ou não, presença de concussão, presença de contusão cerebral presença de lesão do tronco encefálico, presença de grupo de controle, presença de hematoma epidural, presença de hematoma subdural, volume de hematoma, presença de hematoma intracerebral, presença de hérnia cerebral, saturação de oxigênio, presença de complicação com infecção, presença de outras complicações, número de cirurgias, tempo de permanência, tempo de permanência na uti, presença de múltiplos traumas, presença de traqueostomia, período de ventilação mecânica, presença de aspiração, ECG (alta), ocorrência anterior de TCE, presença de hipotermia, presença de acidose, presença de pneumonia adquirida em hospital, contagem de leucócitos, dose de glicose, uso de glicocorticoide, uso de tubo nasogástrico alteração de coagulação, uso de nutrição parenteral, uso de emulsão lipídica, tempo de nutrição enteral, sequelas ou não.

A variável predita foi a sobrevivência do paciente após o trauma cranioencefálico severo. Para verificação da sobrevivência dos pacientes, 22 modelos de aprendizado de máquina foram testados: Logistic regression, Cubic SVM, Cubic KNN, Complex tree, Fine Gaussian SVM, Weighted KNN, Medium tree, Medium Gaussian SVM, Boosted trees, Simple tree, Coarse Gaussian SVM, Bagged trees, Linear discriminant, Fine KNN, Subspace discriminant, Quadratic discriminant, Medium KNN, Subspace KNN, Linear SVM, Coarse KNN, RUSBoosted trees, Quadratic SVM, Cosine KNN.

Para avaliação do desempenho, 5 métricas foram utilizadas: AUC, acurácia, precisão, sensibilidade e F1-score. Para a simplicidade do resumo, serão abordados os resultados de desempenho gerais da área sob a curva ROC, AUC. Todos os modelos de aprendizagem de máquina testados obtiveram AUC entre 0,83 a 0,94, sendo que os modelos que obtiveram AUC igual a 0,94 foram: Linear discriminant, Cosine KNN e Cubic SVM. O modelo de regressão logística obteve AUC = 0,83 .

2.4 *Machine learning-based dynamic mortality prediction after traumatic brain injury (RAJ et al., 2019)*

Neste trabalho, os autores realizaram a predição de mortalidade em tempo real durante a estadia do paciente na unidade intensiva de tratamento. Os pacientes incluídos no estudo foram pacientes adultos, com idade maior ou igual a 16 anos, admitidos em um período de até 24 horas após a ocorrência do trauma. Foram incluídos pacientes que ficaram mais de 24 horas sob monitoramento para verificação da pressão intracranial (ICP, do inglês, *intracranial pressure*). Um total de 472 pacientes foi incluído no estudo. As seguintes variáveis preditoras foram utilizadas: idade, ECG na admissão, pontuação motora, reatividade da pupila, hipóxia, hipotensão, classificação de Marshall para a tomografia computadorizado (CT), tSAH no CT, massa epidural no CT, glicose, Hb, tempo de estadia na unidade intensiva de tratamento e procedimentos neurocirúrgicos. A variável predita é a mortalidade em 30 dias. Dois modelos foram desenvolvidos utilizando a técnica de regressão logística, sendo aplicadas as bibliotecas do *scikit-learn* para criação dos algoritmos. O desempenho dos algoritmos foi avaliado através da área sob a curva ROC. Os resultados discriminaram entre sobreviventes e não sobreviventes com AUC de 0,81 a 0,84.

2.5 *Using and artificial neural network to predict traumatic brain injury (HALE et al., 2018)*

Neste trabalho, os autores utilizaram a base de dados de PECARN (Pediatric Emergency Care Applied Research Network). Foram selecionados os pacientes com idade menor do que 18 anos e com escaneamento computadorizado completo da cabeça. Um total de 12 902 pacientes foram incluídos no estudo, sendo apenas incluídos registros em que não havia variáveis ausentes. Para as variáveis preditoras, foram incluídas 17 variáveis selecionadas por neurologistas:

hemorragia cerebelar, contusão cerebral, edema cerebral, hemorragia cerebral ou hematoma intracerebral, diástase do crânio, hematoma epidural, hematoma extra axial, hemorragia intraventricular, deslocamento da linha média das estruturas do cérebro, pneumocefalia, fratura do crânio, hemorragia subaracnóide, hematoma subdural, infarto traumático, lesão axonal difusa, herniação e lesão por cisalhamento. Além disso, foram incluídas variáveis com importância estatística, que foram: mecanismo da lesão (e.g colisão com veículo), severidade do mecanismo de lesão, perda de consciência, ECG na apresentação, idade e gênero. As variáveis preditas foram o risco para um trauma cranioencefálico clinicamente relevante (do inglês, *Clinically Relevant TBI - CRTBI*): 1) procedimento neurocirúrgico, 2) intubação > 24 horas como resultado direto do trauma na cabeça, 3) hospitalização maior que 48 horas e evidência do trauma cranioencefálico por meio de escaneamento computadorizado e 4) morte devido ao trauma cranioencefálico. O modelo de aprendizagem de máquina utilizado foi um MLP com 11 neurônios, com função de ativação sigmoide na camada escondida e função *softmax* para a camada de saída. Como método de otimização foi utilizado o método *backpropagation* com gradiente conjugado escalonado. Os resultados obtidos foram acurácia de 97,98%, precisão 98,19%, sensibilidade de 99,73%, e AUC = 0,9906.

2.6 *Predicting outcome after traumatic brain injury: practical prognostic models based on a large cohort of international patients (COLLABORATORS et al., 2008)*

Neste trabalho os autores utilizaram uma base de dados MRC CRASH, contendo 10.008 pacientes adultos com ECG ≤ 14 , que entraram no hospital com até 8 horas após o momento do trauma. As variáveis preditoras utilizadas foram: gênero, idade, causa do trauma, tempo do trauma até a randomização, ECG na randomização, reatividade da pupila, resultados da tomografia computadorizada, se o paciente teve maior trauma extracranial, nível de receita no país (países desenvolvidos ou países em desenvolvimento). As variáveis preditas foram morte em 14 dias e morte ou incapacidade severa seis meses após o trauma. Dois modelos de regressão foram desenvolvidos: (1) modelo básico, no qual as variáveis utilizadas foram variáveis clínicas e demográficas (idade, ECG, reatividade da pupila, presença de trauma extra cranial) e (2) modelo CT (tomografia computadorizada), que também incluiu os resultados da tomografia

computadorizada (hemorragias petequiais, obliteração do terceiro ventrículo, desvio da linha média, hematoma não evacuado). Duas métricas para verificação do desempenho do modelo foram utilizadas: teste Hosmer-Lemeshow para calibração do modelo e *C statistics*, um conceito equivalente a área sob a curva ROC. Os dois modelos apresentaram discriminação *C statistics* acima de 0,80. Todos os modelos apresentaram boa calibração, exceto o modelo CT quando aplicados a dados de pacientes de países em desenvolvimento.

2.7 *Predicting Outcome after Traumatic Brain Injury: Development and International Validation of a Prognostic Score Based on Admission Characteristics (STEYERBERG et al., 2005)*

Neste trabalho os autores utilizaram uma base de dados denominada IMPACT, a qual inclui pacientes com trauma cranioencefálico moderado e severo e ECG menor ou igual a 12. Os dados foram coletados a partir de oito estudos clínicos randômicos e 3 estudos clínicos controlados, conduzidos entre 1984 e 1997, o que resultou em um grupo de 8059 pacientes com idade maior ou igual a 14 anos. Para a validação externa foram utilizados dados de pacientes que participaram da pesquisa médica denominada MRC CRASH, que participaram da pesquisa durante o período de 1999 a 2004, sendo incluídos 10.008 pacientes adultos com $ECG \leq 14$, que entraram no hospital com até 8 horas após o momento do trauma. Para esse trabalho, os autores selecionaram pacientes com $ECG \leq 12$ e com ECG completo após 6 meses. As variáveis preditas foram mortalidade, sobrevivência (estado vegetativo ou incapacidade severa), saída favorável (incapacidade média ou boa recuperação). Para o modelo principal, foram incluídas as variáveis preditoras idade, pontuação motora do ECG e reação da pupila. Para o modelo estendido, foram incluídas as três variáveis preditoras utilizadas no modelo principal e as variáveis hipóxia, hipotensão e características tomografia computadorizada. O modelo denominado modelo de laboratório incluiu todas as variáveis do modelo estendido e informações de glicose e hemoglobina. Para preenchimento das variáveis ausentes, foi utilizado o algoritmo MICE. Os modelos são baseados na regressão logística e a métrica de desempenho do modelo selecionada é a área sob a curva ROC. Os resultados obtidos para todos os modelos foram AUC acima de 0,80. Com a validação externa foi possível confirmar o desempenho do modelo principal com a métrica de AUC de 0,80.

2.8 *Neural Network Modeling for surgical decisions on traumatic brain injury patients (LI et al., 2000)*

Neste trabalho os autores utilizaram uma base de dados resultante da coleta de 116 hospitais de ensino de Taiwan, sendo selecionados pacientes com ferimento no crânio ou nos ossos do rosto, nos vasos sanguíneos, nervos e pacientes com contusão, concussão com perda de consciência e hematoma intracraniano. Embora tenham sido coletadas 132 variáveis a partir dos registros dos pacientes, após a remoção das variáveis ausentes e seleção das variáveis mais significantes ($P < 0,05$), o conjunto de dados utilizado tem um total de 12.640 registros e apenas 10 variáveis preditoras: gênero, idade do grupo, uso de capacete, tempo de perda de consciência, tempo de amnésia, presença de amnésia, episódio de convulsão, presença de déficit neurológico, presença de complicações e presença de fraturas cranianas. Como variável predita foi utilizada a decisão de realizar cirurgia do crânio aberto. Três modelos foram desenvolvidos: modelo utilizando a técnica de regressão linear e dois modelos utilizando MLP. O primeiro MLP tinha a seguinte arquitetura: 11 nós na camada de entrada, 7 neurônios na camada escondida e a camada de saída. A função sigmóide foi utilizada como função de ativação das camadas da rede. O segundo modelo de rede neural é uma rede de funções de base radial (RBF) (do inglês, *radial-basis-function*), com uma camada de 5-centros e função de ativação não-linear *spline*. A métrica para verificação dos resultados utilizada foi a área sob a curva ROC. Os resultados obtidos pelos autores mostraram que o modelo de regressão logística obteve $AUC = 0,761$. Já o modelo com MLP obteve uma $AUC = 0,897$ e o modelo com RBF obteve uma $AUC = 0,880$.

2.9 *Predicting survival using simple clinical variables: a case study in traumatic brain injury (SIGNORINI et al., 1999)*

Neste trabalho, os autores utilizaram dados dos pacientes admitidos no centro de trauma regional *Royal Infirmary*, em Edimburgo. Foram selecionados para o estudo pacientes com idade a partir de 14 anos que, durante a admissão, apresentassem a última medição de ECG menor do que 12 ou entre 13 e 15, com lesões concomitantes e para uma pontuação da severidade do trauma

maior do que 16. Um total de 372 pacientes foram incluídos no estudo. As variáveis preditoras selecionadas foram: idade, gênero, ECG, causa do trauma, pontuação da severidade do trauma, resposta da pupila, resultado do escaneamento do cérebro e consumo anterior de álcool. A variável predita foi a sobrevivência do paciente um ano após ele ter recebido a alta. O modelo de regressão logística e o modelo aditivo generalizado foram utilizados para verificar a significância das variáveis. Para predição da sobrevivência dos pacientes, o método de seleção direta (do inglês, *forward selection*) foi utilizado. Nesse método, o modelo inicia sem variáveis. A cada iteração novas variáveis são adicionadas, até quando não se tenham mais melhorias no modelo. Os valores ausentes para qualquer variável do modelo não foram considerados no estudo. Os autores utilizaram algumas métricas para validação da calibração do modelo, como por exemplo, Hosmer e Lemeshow e pontuação de Brier. Para verificar a acurácia do modelo, a métrica utilizada foi a área sob a curva ROC. Foram obtidos os seguintes resultados: 1. Estatística de Hosmer-Lemeshow, que compara as taxas observadas com as taxas esperadas, igual a 12,5, com 8 graus de liberdade, com valores p de 0,129 para a rejeição da hipótese de que o modelo é inadequado. 2. A pontuação de Brier igual a 0,0797, a taxa de erro de 10,1% e a área sob a curva ROC foi de 0,901.

Quadro 1 - Resumo da análise realizada nos trabalhos relacionados.

Referência	Base de dados	Variáveis Predictoras	Variáveis Preditas	Resultados Obtidos em termos de AUC
Application of machine learning to predict the outcome of pediatric traumatic brain injury (TUNTHANNATHIP e OEARSAKUL, 2021)	828 registros de crianças com idade menor do que 15 anos, admitidas no centro de trauma do sudeste da Tailândia, no período de Janeiro de 2009 a Julho de 2020.	14 variáveis predictoras: idade, gênero, mecanismo da lesão, pontuação média de gravidade de lesão, lesões relacionadas, comorbidades, ECG, reflexo da pupila à luz, fratura do crânio, lesões intracranianas, cisterna basal, deslocamento da linha média, tratamento cirúrgico. 1 variável predita: resultado KOSCHI da alta hospitalar.	Resultado KOSCHI após 6 meses da alta hospitalar, compreendendo as seguintes categorias: 1) morte, 2) estado vegetativo, 3) incapacidade moderada, 4) boa recuperação	Modelo máquina de vetor de suporte obteve o melhor desempenho: sensibilidade = 0,95, especificidade = 0,60, valor positivo predito = 0,99, valor negativo predito = 1,0, acurácia 0,94 e AUC-ROC = 0,78
Prediction of Early TBI Mortality Using a Machine Learning Approach in a LMIC Population (AMORIM et al., 2020)	517 registros de pacientes do hospital das Clínicas, da Universidade de São Paulo. Foram incluídos pacientes com idade a partir de 14 anos, com anormalidade intracraniana em uma tomografia inicial computadorizada, exclusão de pacientes com trauma penetrante, assim como aqueles com ECG de 15 e sem associação de lesões intracranianas quando realizada a tomografia.	16 variáveis predictoras: Gênero, idade, nível de reação da pupila na admissão, ECG na cena em que o trauma ocorreu, ECG na admissão, o escore do componente motor do ECG, presença de hipóxia, hipotensão, mudança da linha média maior que 5mm, herniação cerebral identificado na tomografia, hemorragia subaracnóide, hemorragia epidural, severidade do trauma, tempo de protrombina e tempo parcial de tromboplastina.	Mortalidade em 14 dias.	Técnica naive Bayes: AUC= 0,906, seguido do modelo linear bayesiano generalizado, com AUC = 0,881

Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries (GRAVESTIJN, Benjamin Y. et al., 2020)	117 registros de pacientes admitidos no hospital Sichuan Provincial People no período de dezembro de 2009 e Novembro de 2011.	Um total de 40 variáveis foram utilizadas nesse estudo, dentre as quais destaca-se: idade, tempo de estadia no hospital, tempo de estadia na unidade intensiva de tratamento, entre outras.	Sobrevivência do paciente que teve trauma cranioencefálico severo.	AUC entre 0,83 a 0,94. O modelo de regressão logística obteve AUC-ROC = 0,83.
Machine learning-based dynamic mortality prediction after traumatic brain injury (RAJ et al., 2019)	472 pacientes adultos, com idade maior ou igual a 16 anos, admitidos em até 24 horas após a ocorrência do trauma.	13 variáveis preditoras: Idade, ECG na admissão, pontuação motora, reatividade da pupila, hipóxia, hipotensão, classificação de Marshall para a tomografia computadorizado (CT), tSAH no CT, massa epidural no CT, glicose, Hb, tempo de estadia na unidade intensiva de tratamento e procedimentos neurocirúrgicos	Mortalidade em 30 dias	Modelo de regressão logística obteve AUC de 0,81 a 0,84.
Using and artificial neural network to predict traumatic brain injury (HALE et al., 2019)	Base de dados de PECARN (Pediatric Emergency Care Applied Research Network), Um total de 12 902 pacientes foram incluídos no estudo.	23 variáveis preditoras: hemorragia cerebelar, contusão cerebral, edema cerebral, hemorragia cerebral ou hematoma intracerebral, diástase do crânio, hematoma epidural, hematoma extra axial, hemorragia intraventricular, deslocamento da linha média das estruturas do cérebro, pneumocefalia, fratura do crânio, hemorragia	Risco para um trauma cranioencefálico clinicamente relevante (do inglês, <i>Clinically Relevant TBI</i> - CRTBI): 1) procedimento neurocirúrgico, 2) intubação > 24 horas como resultado direto do trauma na cabeça, 3) hospitalização maior que 48 horas e evidência do trauma cranioencefálico por meio de escaneamento computadorizado e 4) morte	MLP, acurácia de 97,98%, precisão 98,19%, sensibilidade de 99,73%, especificidade, AUC ROC = 0,9906.

		subaracnóide, hematoma subdural, infarto traumático, lesão axonal difusa, herniação e lesão por cisalhamento, mecanismo da lesão (e.g colisão com veículo), severidade do mecanismo de lesão, perda de consciência, ECG na apresentação, idade e gênero	devido ao trauma cranioencefálico.	
Predicting outcome after traumatic brain injury: practical prognostic models based on a large cohort of international patients (COLLABORATORS et al., 2008)	Base de dados MRC CRASH, contendo 10.008 pacientes adultos com ECG 14, que entraram no hospital com até 8 horas após o momento do trauma	8 variáveis preditoras: Gênero, idade, causa do trauma, tempo do trauma até a randomização, ECG na randomização, reatividade da pupila, resultados da tomografia computadorizada, se o paciente teve maior trauma extracranial, nível de receita no país (países desenvolvidos ou países em desenvolvimento).	Prognóstico para morte em 14 dias e para morte ou incapacidade severa seis meses após o trauma.	Os modelos apresentaram discriminação C statistics acima de 0,80
Predicting Outcome after Traumatic Brain Injury: Development and International Validation of a Prognostic Score Based on Admission Characteristics (STEYERBERG et al., 2005)	Base de dados IMPACT, constituída por 8059 pacientes, que inclui pacientes com trauma cranioencefálico moderado e severo e ECG menor ou igual a 12 e com ECG completo após 6 meses.	8 variáveis preditoras: Idade, pontuação motora do ECG, reação da pupila, hipóxia, hipotensão, características tomografia computadorizada (do inglês, CT), informações de glicose e hemoglobina.	Mortalidade, sobrevivência (estado vegetativo ou incapacidade severa), saída favorável (incapacidade média ou boa recuperação)	Os resultados obtidos para todos os modelos obtiveram AUC > 0,80.

Neural Network Modeling for surgical decisions on traumatic brain injury patients (LI et al., 2000)	Base de dados com 12.640 registros de pacientes de 116 hospitais de ensino de Taiwan.	10 variáveis preditoras: gênero, idade do grupo, uso de capacete, tempo de perda de consciência, tempo de amnésia, presença de amnésia, episódio de convulsão, presença de déficit neurológico, presença de complicações, presença de fraturas cranianas.	Decisão de realizar cirurgia do crânio aberto.	regressão logística obteve AUC = 0,761, MLP obteve AUC = 0,897, RBF obteve AUC = 0,880.
Predicting survival using simple clinical variables: a case study in traumatic brain injury (SIGNORINI et al., 1999)	372 pacientes admitidos no centro de trauma regional Royal Infirmary, em Edimburgo. Critérios de inclusão: idade a partir de 14 anos que, durante a admissão, apresentassem a última medição de ECG menor que 12 ou entre 13 e 15, com lesões concomitantes e para uma pontuação da severidade do trauma maior do que 16.	8 variáveis preditoras: Idade, gênero, ECG, causa do trauma, pontuação da severidade do trauma, resposta da pupila, resultado do escaneamento do cérebro e consumo anterior de álcool.	Sobrevivência do paciente um ano após ele ter recebido a alta	Área sob a curva ROC foi de 0,901

3 REFERENCIAL TEÓRICO

Neste capítulo serão abordados os conceitos necessários para o entendimento da metodologia desse trabalho. Inicialmente, apresenta-se os conceitos sobre trauma cranioencefálico (seção 3.1). Em seguida, introduz-se às técnicas de pré-processamento de dados. Na sequência, são apresentados os conceitos de modelos de classificação e regressão (seção 3.2), e conceitos de redes neurais (seção 3.3). Por fim, são apresentados os métodos de otimização das arquiteturas que serão implementadas (seção 3.4).

3.1 Traumatismo Cranioencefálico

Traumatismo cranioencefálico é um tipo de lesão física ao tecido cerebral que pode incapacitar a função cerebral. Temporariamente ou de forma permanente. Tem sido considerado um dos maiores problemas de saúde em todo o mundo, resultando em deficiências motoras e psíquicas, e mortes, principalmente de pessoas jovens e economicamente ativas. Em sua maioria, resultam de acidentes no trânsito (60%), quedas (30%) e violência (10%), sendo interessante notar que o grupo mais afetado é o de pessoas com idade entre 21 e 30 anos (THORNHILL *et al.*, 2000). Um dos motivos que nos traz ao estudo desse problema na saúde é a preocupação com um atendimento adequado a essas pessoas, bem como o estudo da predição de mortalidade desses pacientes.

Diferentemente de países desenvolvidos, os países em desenvolvimento apresentam um número muito maior de acidentes no trânsito, o que induz em uma maior probabilidade na ocorrência de acidentes com traumas cranioencefálicos. De acordo com Schmucker *et al.* (2010), países em desenvolvimento possuem 48% dos veículos motorizados, no entanto, as mortes fatais por acidentes de trânsito nesses países é 91% de 1,3 milhões de mortes fatais em todo o mundo. Além disso, mais de 90% dos Anos de Vida Ajustados por Incapacidade (do inglês DALY – *Disability Adjusted Life Years*) em todo o mundo devido a acidentes de trânsito ocorrem em países em desenvolvimento (MURRAY e LOPEZ, 1996).

Em países desenvolvidos, a primeira base de dados trauma computadorizada foi estabelecida em Chicago, no Hospital Cook County, no ano de 1969 (NWOMEH *et al.*, 2006). Atualmente, muitos dos estados dos EUA mantêm registros históricos dos pacientes e tratamentos

de trauma realizados, de forma a contribuir não somente com os estudos como também com os provedores de planos de saúde, legislação e outras partes interessadas.

Portanto, modelos preditivos de mortalidade, utilizando aprendizagem de máquina, puderam ser aplicados há mais tempo em países desenvolvidos, mormente por conta de dados históricos existentes. A existência dessa base de dados, possibilitou o desenvolvimento de um modelo baseado em pontuação do trauma, em um estudo realizado em 1982, que tinha como objetivo agregar dados de pacientes acidentados para desenvolver e testar normas de probabilidade de sobrevivência, com base nas pontuações de gravidade de lesões (NWOMEH *et al.*, 2006).

De acordo com Amorim et al. (2020), a maioria das tentativas de predição de resultado de um TCE é baseada em cálculos manuais. Um dos exemplos é a Escala de Coma de Glasgow (ECG), que associa os pacientes com pontuações entre 13-14 à amnésia pós-traumática mais longa, além de maior anormalidade nas imagens do cérebro em até seis meses após o trauma inicial. A ECG associada a outras variáveis pode trazer um impacto positivo para a predição da mortalidade dos pacientes. Outras variáveis podem ser associadas a ela, como por exemplo, idade, reflexos do cérebro, entre outras.

Um dos motivos para aplicação de modelos de aprendizagem de máquina no contexto de países em desenvolvimento está relacionado a contribuição na tomada de decisões médicas relacionadas a pacientes desse tipo de país. Muitos dos modelos de aprendizagem de máquinas criados são treinados com dados de pacientes de países desenvolvidos, não sendo muitas vezes, aplicáveis para prognóstico de pacientes de países em desenvolvimento.

3.2 Pré-processamento dos Dados

O pré-processamento dos dados é uma fase importante para alcançar um bom desempenho de classificação, sendo aplicado antes da avaliação dos dados pelo algoritmo de aprendizagem de máquina (SINGH e SINGH, 2020). Existem várias técnicas que podem ser aplicadas nessa fase, como por exemplo, remoção de *outliers*, integração de dados de diversas fontes, limpeza dos dados e normalização. Nesta seção, são apresentadas algumas técnicas que podem ser utilizadas para o pré-processamento de dados.

3.2.1 Normalização de dados

Normalização de dados é uma etapa essencial no pré-processamento de dados, que envolve a transformação de variáveis para um intervalo comum, evitando que variáveis com valores mais altos não ofusquem variáveis com valores menores, no método de classificação. O principal, então, é reduzir o viés daquelas variáveis em que a contribuição numérica é mais alta no momento de discriminar as classes (GARCIA *et al.*, 2015).

A normalização é realizada nos dados não tratados, visando escalá-los novamente ou transformá-los. Alguns métodos são considerados para esse estudo: Normalização Z-score (NZS), Normalização Mínimo-Máximo (NMM) e Raiz Cúbica (NRC) (SINGH e SINGH, 2020).

Na Normalização Z-score (NZN), as medidas de média e desvio padrão são utilizadas para reescalar os dados, o que faz com que as variáveis resultantes tenham média zero e variância unitária. Cada observação do conjunto de dados é transformada como segue:

$$x'_i = \frac{x_i - \mu}{\sigma}$$

Em que:

x'_i é o i-ésimo valor normalizado

x_i é o i-ésimo valor não normalizado

μ é uma média da variável x_i

σ o desvio padrão da variável x_i

O método de normalização Mínimo Máximo (NMM) escala os dados não normalizados de acordo com os limites inferiores e superiores pré-definidos, sendo que normalmente os dados são escalados para um intervalo entre 0 e 1 ou mesmo entre -1 e 1. A equação é dada como segue:

$$x'_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \cdot (nMax - nMin) + nMin$$

Em que:

x'_i é o i-ésimo valor normalizado

x_i é o i-ésimo valor não normalizado

$\min(x_i)$ – valor mínimo da variável x_i

$\max(x_i)$ – valor máximo da variável x_i

$nMax$ – Máximo valor desejado para x_i

$nMin$ – Mínimo valor desejado para x_i

A normalização por Raiz Cúbica (NRC), proposto por Mulenga *et al.* (2021), multiplica cada observação no conjunto de dados pelo desvio padrão dos dados não tratados. Em seguida, para reescalar os dados, calcula a raiz cúbica do produto. Diferentemente dos métodos NZN e NMM, que são aplicados em variáveis individuais, o método NRC é aplicado no conjunto de dados completo, como mostrado na equação que segue:

$$x'_i = \sqrt[3]{x_i \cdot \sigma}$$

Em que:

x'_i é o i -ésimo valor normalizado

x_i é o i -ésimo valor não normalizado

σ – desvio padrão da variável x_i

3.2.2 Métodos para Abordagem do Problema de Variáveis Ausentes

Dados ausentes ou dados incompletos ocorrem com frequência em estudos médicos, normalmente devido a formulários respondidos de forma parcial, descumprimento de testes clínicos, entre outros motivos. Por representar uma perda de informação, os dados ausentes complicam a manipulação e a análise dos dados, gerando um viés nas conclusões obtidas a partir dos dados presentes. (BARNARD e MENG, 1999).

De acordo com Han, Pei e Jian (2011), existem várias abordagens para realizar o preenchimento de valores ausentes, como por exemplo: (a) ignorar registros contendo variáveis ausentes, (b) preencher os valores ausentes manualmente, (c) utilizar uma constante global para preencher todos os valores ausentes, (d) usar uma medida de tendência central para preencher o valor ausente, (e) usar a média ou mediana para as observações de todas as variáveis que pertencem à mesma classe, (f) usar o valor mais provável por meio de algoritmos de aprendizagem de máquina para preencher o valor ausente. A seguir faremos alguns comentários sobre cada uma dessas abordagens.

Abordagem (a) - Essa abordagem é normalmente usada quando a informação referente à definição da classe do registro está faltando (assumindo a predição será de classificação) ou mesmo vários atributos de um registro estão faltando. No entanto, isso pode fazer com o que o modelo tenha baixo desempenho se alguns registros forem ignorados ou removidas, pois informações importantes podem estar sendo removidas.

Abordagem (b) - Essa abordagem pode requerer muito tempo para ser completada, dependendo do tamanho do conjunto de dados.

Abordagem (c) - Nessa abordagem, todos os valores ausentes de cada atributo são substituídos pela mesma constante. Portanto, antes de realizar os cálculos, os valores ausentes são substituídos por valores entre zero e um, ou qualquer valor, dependendo da magnitude de cada atributo.

Abordagem (d) Nessa abordagem, calcula-se uma medida de tendência central. Se a distribuição da variável é normal, a média pode ser usada, caso contrário, a mediana deve ser empregada para substituição dos valores ausentes. Portanto, os valores ausentes são removidos, e com os valores restantes é realizado o cálculo da medida de tendência central. O valor encontrado é substituído onde antes se encontrava o valor ausente. Exemplos de medidas de tendência central podem ser média, moda e mediana.

(e) Nessa abordagem, são substituídos os valores ausentes pela média ou mediana das observações que pertencem à mesma classe.

(f) Nessa abordagem, são aplicadas regressão, ferramentas baseadas em inferência usando formalismo Bayesiano ou mesmo árvores de decisão para estimar valores que substituirão os valores ausentes para cada variável.

A abordagem (f) é uma estratégia popular, sendo bastante utilizado por fazer uso da maioria das informações presentes no conjunto de dados. Além disso, é importante observar que, em alguns casos, os valores ausentes podem significar que os verdadeiros valores não precisavam ser preenchidos para aquela observação, o que não deveria resultar em um erro no conjunto de dados (HAN; PEI; KAMBER, 2011).

3.3 Aprendizado Supervisionado

Esse tipo de aprendizado é caracterizado pela existência de pares entrada-saída para treinamento do classificador. Para compreensão sobre tipos de aprendizado supervisionado, nesta seção serão abordados os seguintes tipos de aprendizado supervisionado mais utilizados: Árvores de Decisão, Regressão Linear, Regressão Logística, Máquina de Vetores de Suporte, K-Vizinhos Mais Próximos e Redes Neurais Artificiais. Vale ressaltar que os modelos apresentados nessa seção podem ser utilizados tanto na fase de Pré-Processamento de Dados, onde serão empregados para predição de valores ausentes, quanto na fase de predição de mortalidade dos pacientes que sofreram traumatismo cranioencefálico.

3.3.1 Árvore de Decisão

Árvores de decisão são modelos estatísticos que utilizam um treinamento supervisionado para a classificação e previsão de dados (DA SILVA, 2005). Podem ser utilizadas tanto para classificação como para regressão, sendo possível a utilização de variáveis categóricas e contínuas. A estrutura de uma árvore de decisão (Figura 3.1) é apresentada abaixo.

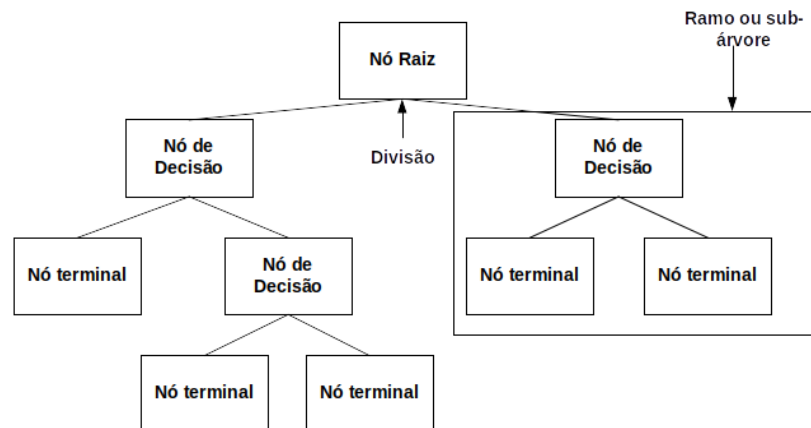


Figura 3.1 – Estrutura de uma árvore de decisão

De modo geral, podemos dizer que uma árvore é uma estrutura de dados formada por um conjunto de elementos que armazenam informações, esses elementos são chamados de nós. Toda árvore de decisão possui um nó, chamado nó raiz, que é o ponto hierárquico mais alto e de onde derivam as ligações nós filhos. Um nó que não possui filho é chamado de nó folha ou terminal, pois a partir dele não há mais divisões. Doravante, cada subseção a partir de um nó é denominada

de ramo ou sub-árvore. Em uma árvore de decisão, no nó raiz e em cada nó de decisão da árvore, uma decisão é tomada, só não ocorrendo decisão em um nó terminal.

3.3.2 Floresta Randômica

O método é Floresta Randômica é um classificador que consiste em uma coleção de classificadores com estrutura de árvore, em que cada árvore é construída a partir de uma coleção randômica de variáveis (ZHANG e MA, 2012). Para a tomada de decisão sobre a que classe um padrão de teste pertence, utiliza-se um processo de votação, em que cada árvore contribui com um voto. O voto de cada árvore representa a classe em que o padrão de teste foi classificado (LEITE; DE MORAES; LOPES, 2021). O fato do algoritmo Floresta Randômica utilizar um processo de votação para a tomada de decisão faz com que ele seja mais robusto. A Figura 3.2 ilustra o processo de classificação de uma Floresta Randômica.

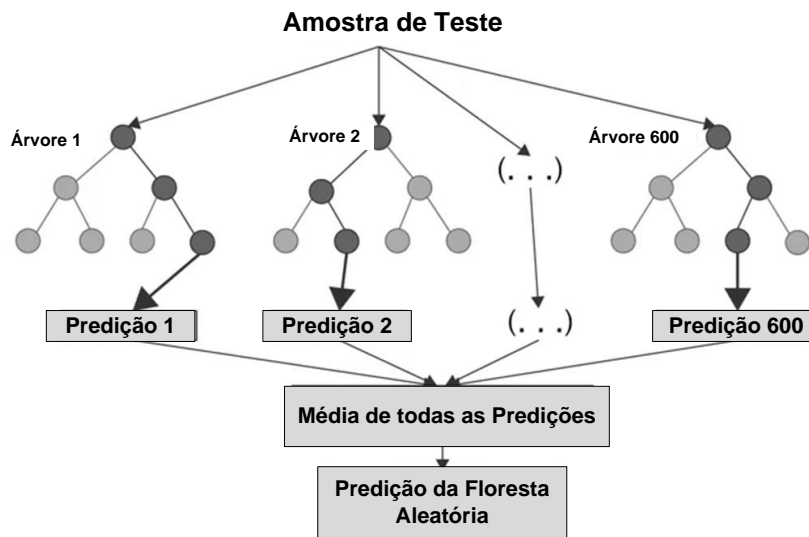


Figura 3.2 – Processo de Classificação Floresta Randômica

Fonte: [Corporate Finance Institute](#)

3.3.3 K-Vizinhos Mais Próximos

O K-Vizinhos Mais Próximos (do inglês, *K-Nearest Neighbors*) é um classificador onde o aprendizado é baseado em analogia, sendo o conjunto de treinamento formado por vetores n-dimensionais. Cada elemento do conjunto de treinamento representa um ponto no espaço n-dimensional. Neste algoritmo, para classificar uma amostra de teste, o classificador procura k elementos no conjunto de treinamento que estejam mais próximos da amostra de teste (DA SILVA,

2005). Esses k elementos são chamados de k -vizinhos mais próximos. Obtida a classe dos k -vizinhos mais próximos, sendo atribuída à classe da amostra de teste a classe mais frequente entre os vizinhos mais próximos (DA SILVA, 2005). Na Figura 3.3 temos uma amostra de teste que precisa ser classificada com base nos vizinhos mais próximos. Se $k=1$, conforme mostrado, a amostra pertence à classe “estrela”.

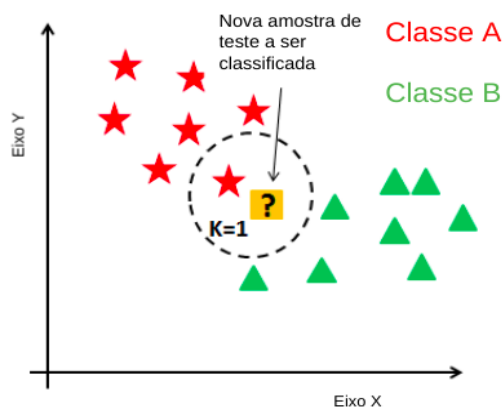


Figura 3.3 – Gráfico k -vizinhos mais próximos
 Fonte: [Infer \(2019\)](#)

3.3.4 Máquina de Vetores de Suporte

Máquina de Vetores de Suporte (do inglês, *Support Vector Machine*) é um algoritmo que pode ser aplicado para classificação e regressão, possuindo boa capacidade de generalização. Baseia-se no conceito de planos de decisão que são definidos por pontos extremos das classes, conforme mostrado na Figura 3.4, também chamados de “vetores de suporte”. Esses hiperplanos dividem os dados em diferentes categorias ou classes (LEITE; DE MORAES; LOPES, 2021), como exemplificado na Figura 3.4.

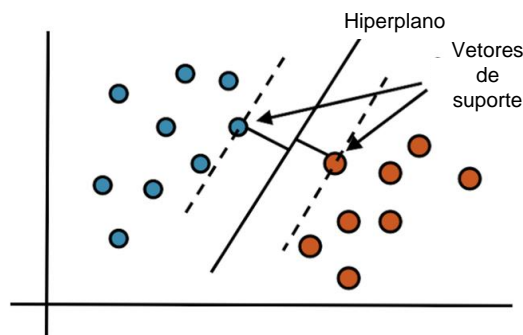


Figura 3.4 – Gráfico Máquina de Vetores de Suporte
 Fonte: Ichi (2021)

3.3.5 Regressão Logística

Regressão Logística é uma técnica estatística que visa produzir um modelo para predição de valores assumidos por uma variável categórica, normalmente binária, em função de uma ou mais variáveis independentes contínuas e/ou binárias (GONZALEZ, 2018). Trata-se, assim, de um recurso que permite estimar a probabilidade associada à ocorrência de um evento, em função de um conjunto de variáveis.

Na regressão logística, a probabilidade da ocorrência do evento ou probabilidade de um valor pertencer a uma classe, pode ser determinada por meio da fórmula que segue:

$$P(Y = 1) = \frac{1}{1 + e^{-g(x)}}$$

sendo que Y assume apenas dois estados (1 ou 0). Assumindo-se que temos um conjunto de p variáveis independentes $x = [x_1, x_2, \dots, x_p]$. Temos que $g(x) = b_0 + b_1 x_1 + \dots + b_p x_p$.

3.3.6 Regressão Linear

A regressão linear é um algoritmo que pode ser usado para predição de um valor baseado em uma série de eventos históricos. Esse algoritmo tem como principal objetivo verificar se determinadas variáveis preditoras influenciam em um valor de saída Y, e assim, determinar um modelo para prever o valor de Y em função dos valores dessas variáveis preditoras.

Basicamente, existem dois tipos de regressão: simples e múltipla. Regressão linear simples ocorre quando há apenas uma variável preditora X. Por sua vez, a regressão linear múltipla ocorre quando há várias variáveis preditoras (REHBEIN, 2020). A função básica da regressão linear simples é dada a seguir:

$$f(Y_i) = w_0 + w_1 * x_i$$

sendo que w_0 representa o ponto em que a reta cruza o eixo x, w_1 representa a inclinação da reta e x_i representa a variável preditora.

3.4 Redes Neurais Artificiais

Rede Neural é uma ferramenta de aprendizado de máquina composta por uma série de elementos, denominados de neurônios, dispostos em camadas, que se conectam umas com as outras. A inspiração das redes neurais vem do funcionamento do cérebro, no qual uma rede neural é composta por vários neurônios dispostos em camadas (REHBEIN, 2020).

A estrutura de um neurônio artificial contém, basicamente, os seguintes elementos: entradas (X), pesos (W), polarização (b), função de ativação (θ). Como mostrado na Figura 3.5, as entradas são conectadas ao neurônio através dos pesos que simulam a estrutura do dendrito (extensões do corpo celular que atuam na recepção dos estímulos vindos de outros neurônios). O somatório, a polarização e a função de ativação, mostrados nessa figura, representam o análogo do processamento realizado pelo corpo da célula de um neurônio. A variável Y representa a saída do axônio da célula nervosa (AWAD; KHANNA, 2015).

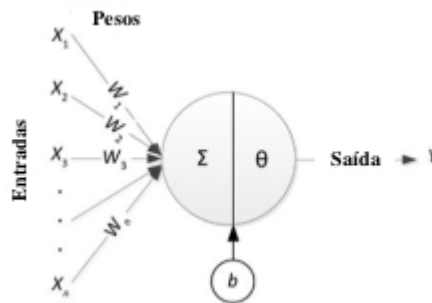


Figura 3.5 – Estrutura básica de um neurônio artificial
Fonte: AWAD; KHANNA (2015).

A representação matemática de um neurônio artificial é dada por:

$$Y = \theta\left(\sum_{i=1}^n W_i X_i + b\right)$$

Sendo representado de forma matricial por:

$$Y = \theta(W \cdot X + b)$$

onde, $W = [W_1 \ W_2 \ \dots \ W_n]$ e $X = [X_1 \ X_2 \ \dots \ X_n]^t$

Assim, é importante notar que X representa um conjunto de características que influenciam a saída Y , sendo essa influência intermediada pelos valores dos pesos. Portanto, para que a rede seja treinada para realizar uma dada função, os pesos são alterados durante o processo de

treinamento. A polarização, por sua vez, permite um melhor ajuste da função de ativação. Diversas funções de ativação já foram propostas na literatura (SERRÃO, 2020). As funções de ativação mais utilizadas são:

1. Sigmóide: $\theta(a) = \frac{1}{1+e^a}$
2. Tangente Hiperbólica: $\theta(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$
3. Degrau: $\theta(a) = \{0 \text{ se } a < 0; 1 \text{ se } a > 0$
4. Linear: $\theta(a) = \{0 \text{ se } a < 0; a \text{ se } 0 \leq a \leq 1; 1 \text{ se } a > 1$

A estrutura de uma rede neural artificial é formada por várias camadas lineares de neurônios. Cada camada é definida pelo número de neurônios que a mesma possui. Os neurônios de uma camada conectam-se a todos os neurônios da camada anterior, mas não se conectam a outros neurônios da mesma camada. Por isso, essas redes são também denominadas de redes inteiramente conectadas. (AWAD; KHANNA, 2015). Não se costuma ter redes neurais com mais de três ou 4 camadas. O número de neurônios e de camadas é uma função da complexidade do problema de aprendizado de máquina que se tem na mão para resolver (SERRÃO, 2020).

3.5 Redes Neurais Convolucionais

Redes Neurais Convolucionais (do inglês, *Convolutional Neural Networks*) são normalmente utilizadas para processamento de imagens, principalmente, para classificação de imagem e segmentação de objetos. Na arquitetura de uma CNN, tem-se camadas tridimensionais. Na camada de entrada, a terceira dimensão diz respeito à quantidade de planos do sinal de entrada. Nas outras camadas, a terceira dimensão refere-se ao número de filtros convolutivos que são usados na camada. Cada filtro convolutivo é responsável pela geração de um mapa de características resultante do processo de convolução do filtro pelos valores de ativação da camada anterior (AGGARWAL, 2018). Diferentemente das redes neurais simples, abordadas na seção anterior, os neurônios de uma camada convolutiva não se conectam a todos os neurônios da camada anterior, mas tão somente a um pequeno grupo de neurônios.

Diferentemente das redes simples abordadas na seção anterior, a arquitetura de uma rede convolutiva é formada por diversos tipos de camadas. Dentre essas camadas, citamos: (1) camada

convolutiva, (2) camada de amostragem (*pooling*), (3) camada de *dropout*, (3) camada *ReLU*, (4) camada de regularização L2.

3.5.1 Camada Convolutiva

A camada convolutiva ou convolucional é utilizada para extrair características representativas das entradas, sendo composta de vários kernels de convolução, que são usados para calcular diferentes mapas de características (UTSCH, 2018). De forma geral, o mapa de características final é determinado em duas etapas: primeiramente, a entrada é convoluída com um filtro. Em seguida, aplica-se uma função não linear aos resultados da convolução. A não linearidade é de suma importância no processo de aprendizado. Por meio dela ocorre a detecção de características não lineares presentes nos dados de entrada (GOODFELLOW; BENGIO; COURVILLE, 2016).

3.5.2 Camada *Pooling* (subamostragem)

Uma camada de rede convolucional consiste em três fases. Na primeira fase, várias convoluções são executadas pela camada para que seja produzido um conjunto de ativações lineares. Na segunda fase, cada ativação linear é executada por meio de uma ativação linear, como por exemplo, a ativação ReLu. E, finalmente, na terceira fase, uma função de *pooling* é usada para redução da dimensão da saída (GOODFELLOW; BENGIO; COURVILLE, 2016).

A camada de *pooling* opera em cada mapa de características da entrada para produzir outro mapa de características reduzido (SERRÃO, 2020). Contudo essas camadas não realizam alterações no número de mapas de características, com isso a profundidade da imagem de saída da camada de *pooling* é igual a imagem de entrada na camada. Na Figura 3.6, temos um exemplo de *max pooling*, onde uma região de tamanho 2x2 com passo de 2, tem-se a saída reduzida em um fator 2.

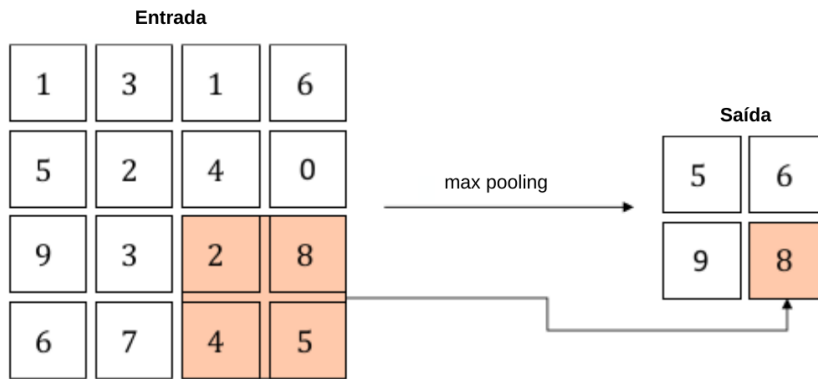


Figura 3.6 – Operação de *pooling*
 Fonte: SERRÃO (2020).

3.5.3 Camada *Dropout*

Dropout é um método para melhoria da generalização de uma rede, que consiste na eliminação de neurônios de uma camada, criando modelos de redes neurais diferentes. Se um conjunto de neurônios de uma camada for eliminado, então todas as conexões de entrada e saída desse neurônio também serão descartadas (AGGARWAL, 2018). Nos simuladores, quando, na arquitetura da rede, tem-se uma camada de *dropout* à frente de uma camada convolutiva ou inteiramente conectada, o simulador gera um conjunto de modelos diferentes de redes neurais, alterando o número de neurônios da camada na frente da qual a mesma é colocada.

A principal razão para se utilizar uma camada *dropout* em uma CNN é reduzir os problemas de *overfitting*, ou seja, reduz-se a chance da rede apresentar excelente desempenho no conjunto de treinamento e baixo desempenho nos conjuntos de validação e teste.

3.5.4 Camada ReLU

Nos últimos anos, o uso da função de ativação ReLU substituiu as funções de ativação sigmóide e tangente hiperbólica no projeto de uma CNN, pois a ReLU tem grandes vantagens sobre essas funções de ativação, tanto em termos de velocidade quanto de precisão (AGGARWAL, 2018). A função é expressa pela equação abaixo:

$$f(x) = \max(0, x)$$

Em que x representa o valor de ativação de uma camada anterior.

3.5.5 Regularização L2

A regularização é um método de generalização que consiste na adição de um termo à função de custo. O termo que é adicionado à função de custo é a soma do quadrado de todos os pesos w da rede escalonado por um fator $\frac{\lambda}{2n}$, onde λ é o parâmetro de regularização e n o tamanho do conjunto de treinamento (SERRÃO, 2020). Assim, a rede tende a minimizar a função de custo e a aprender pesos pequenos. A função de custo regularizada é dada por:

$$E_R = E + \frac{\lambda}{2n} \sum_w w^2$$

3.6 Métodos de Otimização

Métodos de otimização são técnicas de treinamento que ajudam a minimizar uma função de custo das redes neurais simples e das redes convolucionais. O emprego desses métodos, baseados no algoritmo de propagação reversa, possibilitam o treinamento desses classificadores para realizarem determinadas tarefas de classificação e regressão. Neste trabalho, para o treinamento de redes simples e de redes convolucionais, foram utilizados diversos métodos de otimização. A seguir listamos esses métodos: Estimativa de Dinâmica Adaptativa (ADAM, do inglês *Adaptive Moment Estimation*), AdaMax, Propagação da Raiz Média Quadrática (RMSProp, do inglês *Root Mean Square Propagation*) e Gradiente Descendente Estocástico com Momento (SGDM, do inglês *Stochastic Gradient Descent with Momentum*).

3.6.1 Estimativa Dinâmica Adaptativa (Adam)

O método Adam atua sobre a componente do gradiente, utilizando a média exponencial ponderada dos gradientes passados (V_t). Além disso, atua na taxa de aprendizagem (α), dividindo-a pela raiz da média móvel (S_t) do gradiente ao quadrado. A atualização dos pesos nesse método é feita através da equação a seguir:

$$w_{t+1} = w_t - \frac{\alpha}{\sqrt{\hat{S}_t + \epsilon}} \cdot \hat{V}_t$$

\hat{V}_t e \hat{S}_t sofrem correções de vieses dadas por:

$$\hat{V}_t = \frac{V_t}{1 - \beta_1^t}$$

$$\hat{S}_t = \frac{S_t}{1-\beta_2^t}$$

V_t e S_t são inicializados em zero, β_1^t e β_2^t são hiper-parâmetros a serem otimizados e ϵ um valor muito pequeno para evitar divisão por zero (KINGMA; BA, 2015). Os valores de V_t e S_t são dados por:

$$V_t = \beta_1 V_{t-1} + (1 - \beta_1) \frac{dL}{dw_t}$$

$$S_t = \beta_2 S_{t-1} + (1 - \beta_2) \left[\frac{dL}{dw_t} \right]^2$$

3.6.2 AdaMax

AdaMax é uma adaptação do método de otimização Adam, que utiliza normas infinitas. Essa adaptação foi criada pelos mesmos autores do método Adam. Sendo, V_t a média móvel exponencial dos gradientes e S_t a média móvel exponencial da norma-p dos gradientes passados (KINGMA; BA, 2015), a atualização dos pesos nesse método é dado pela equação a seguir

$$w_{t+1} = w_t - \frac{\alpha}{\sqrt{\hat{S}_t + \epsilon}} \cdot \hat{V}_t$$

em que:

$$\hat{V}_t = \frac{V_t}{1-\beta_1^t}$$

é a equação que realiza a correção do viés em V_t e

$$V_t = \beta_1 V_{t-1} + (1 - \beta_1) \frac{dL}{dw_t}$$

$$S_t = \max(\beta_2 S_t, \left| \frac{dL}{dw_t} \right|)$$

V_t e S_t inicializados em zero.

3.6.3 Propagação da Raiz Média Quadrática (RMSProp)

RMSProp é um método de otimização que possui taxa de aprendizagem adaptativa, sendo que esse método armazena a média ponderada exponencial dos quadrados anteriores e divide a taxa de aprendizagem por essa média para acelerar a convergência (RUDER, 2016).

$$w_{t+1} = w_t - \frac{\alpha}{\sqrt{\hat{S}_t + \epsilon}} \cdot \frac{dL}{dw}$$

em que:

$$\hat{S}_t = \frac{S_t}{1 - \beta^t}$$

e

$$S_t = \beta S_t + (1 - \beta) \left[\frac{dL}{dw} \right]^2$$

S_t é inicializado em zero.

3.6.4 Gradiente Descendente Estocástico com Momento (SGDM)

O SGD e suas variantes são provavelmente os métodos de otimização mais utilizados. No entanto, muitas vezes podem apresentar lentidão no treinamento. Por isso, uma das estratégias utilizadas para acelerar a convergência é a redução da taxa de aprendizagem após cada iteração. Outra estratégia utilizada é utilizar a variação com momento, na qual é adicionada uma variável que controla a velocidade na qual os parâmetros se movem através do espaço de parâmetros (GOODFELLOW; BENGIO; COURVILLE, 2016).

Esse método acelera a taxa de aprendizagem em regiões onde o gradiente é constante e evita oscilações desnecessárias, resultando em uma convergência mais rápida. Portanto, seja a função de custo, L , taxa de aprendizagem, α e a variável que controla a velocidade, V , que é suavizada de forma exponencial, controlada por um parâmetro de suavização β . As equações abaixo representam como a atualização do vetor \underline{W} é realizada.

$$\underline{V} = \beta \underline{V} - \alpha \frac{dL}{dw}$$

$$\underline{W} = \underline{W} + \underline{V}$$

V é inicializado em zero.

4 MATERIAIS E MÉTODOS

A metodologia proposta para este trabalho consiste na avaliação de arquiteturas de redes profundas e de modelos tradicionais de aprendizado de máquina, tendo em vista a contribuir para o estado da arte da predição de mortalidade de pacientes com trauma cranioencefálico, levando em consideração um contexto de pequeno volume e baixa qualidade de dados. Para isso, serão utilizados dados de treinamento, validação e teste, selecionados a partir da base de dados disponibilizada para a pesquisa.

No Capítulo 3, foram abordados os conceitos iniciais sobre os modelos de aprendizagem de máquina que serão utilizados neste trabalho. Nesta seção serão descritos os materiais utilizados e a metodologia aplicada na pesquisa, que consistiu das seguintes etapas: pré-processamento dos dados, avaliação de modelos preditivos, validação e teste do melhor modelo. Os modelos preditivos propostos envolveram árvores de decisão, perceptron multicamadas, redes convolutivas, máquinas de vetores de suporte e regressão logística.

4.1 Materiais

4.1.1 Base de Dados

A base de dados utilizada neste trabalho foi fornecida pelo artigo de Amorim et al. (2020). Essa base de dados foi gerada a partir de dados coletados de pacientes que foram encaminhados para o Hospital das Clínicas (São Paulo, Brasil). A coleta dos dados foi realizada no período de março de 2012 a janeiro de 2015. O acompanhamento final dos pacientes encerrou em junho de 2015. Essa base de dados contém 529 registros com 71 colunas. Cada registro corresponde aos dados coletados de um paciente com traumatismo cranioencefálico e as colunas referem-se as variáveis que foram coletadas de cada paciente. A base de dados fornecida tem aprovação do Comitê de Ética em Pesquisa da Universidade de São Paulo (São Paulo, Brasil), CAAE 46831315.3.0000.0068.

De acordo com Amorim et. al (2020), os pacientes que foram admitidos no estudo foram aqueles referenciados pela equipe de neurocirurgia que precisavam de admissão a unidade de tratamento intensiva (UTI). Para esse estudo, os seguintes critérios de seleção foram realizados: excluídos pacientes com menos de 14 anos de idade, pacientes com traumatismo craniano

penetrante, pacientes com ECG igual a 15 e pacientes que não apresentaram anormalidades intracranianas no escaneamento computadorizado.

A partir da base de dados original, utilizando os critérios de exclusão acima citados, uma nova base de dados foi gerada, com 470 registros. Com respeito às informações de cada paciente, foram selecionadas apenas 17 colunas de dados que contêm as variáveis utilizadas nesse estudo. As 17 variáveis utilizadas neste trabalho correspondem às variáveis utilizadas no trabalho de Amorim et al. (2020). Com isso, será possível uma comparação dos resultados obtidos nesse trabalho com os resultados obtidos em Amorim et al. (2020). Na Seção 4.2.1 detalharemos essas variáveis. Na Figura 4.1 exibe-se um recorte das 5 primeiras observações da base de dados. As células que possuem valores NaN (não é um número - Not a Number) referem-se aos valores ausentes na base de dados. Como veremos adiante, na etapa de pré-processamento, diversas técnicas foram avaliadas para preenchimento de valores ausentes.

	Genero	Idade	Dilatacao_Pupilar	ECG_Local_Trauma	ECG_Admissao	Pontuacao_Motora	Hipoxia	Hipotensao_Admissao	TC_Desvio_Cranio	I
0	1	44.0	2.0	NaN	7.0	5.0	0.0	0.0	0	0
1	1	57.0	2.0	NaN	14.0	6.0	0.0	0.0	0	0
2	1	75.0	NaN	NaN	NaN	NaN	NaN	NaN	0	0
3	1	21.0	NaN	3.0	3.0	1.0	0.0	1.0	0	0
4	1	37.0	2.0	NaN	5.0	3.0	0.0	NaN	1	1

Figura 4.1 – Recorte das 5 primeiras observações da base de dados utilizada.

Portanto, as variáveis utilizadas neste estudo serão: Hipóxia, Idade, ECG Local do Trauma, Severidade do Trauma, Hipotensão na Admissão, Gênero, ECG na Admissão, TC Desvio do Crânio > 5mm, Tempo Tromboplastina Parcial, Herniação Cerebral, Pontuação Motora, TC Hemorragia Subaracnoidea, Tempo de Protrombina, TC Hemorragia Subdural, Dilatação Pupilar, TC Hemorragia Intracerebral e morte em até 14 dias.

4.1.2 Divisão da Base de Dados

Em aprendizado de máquina, a obtenção de um modelo ótimo para a solução de um problema envolve três etapas distintas, denominadas de treinamento, validação e teste. Cada uma dessas etapas exige um conjunto próprio, denominados de conjunto de treinamento, conjunto de validação e conjunto de teste.

O conjunto de treinamento é utilizado para obtenção dos parâmetros do modelo. Por exemplo, numa rede neural, os parâmetros são os pesos e polarizações da rede. O conjunto de validação é utilizado para definição dos hiperparâmetros do modelo. Por exemplo, numa rede neural, a taxa de aprendizado ótima, a melhor arquitetura, o método para melhoria de generalização com o melhor desempenho etc. O conjunto de teste é utilizado para avaliar o melhor modelo definido através do processo de validação.

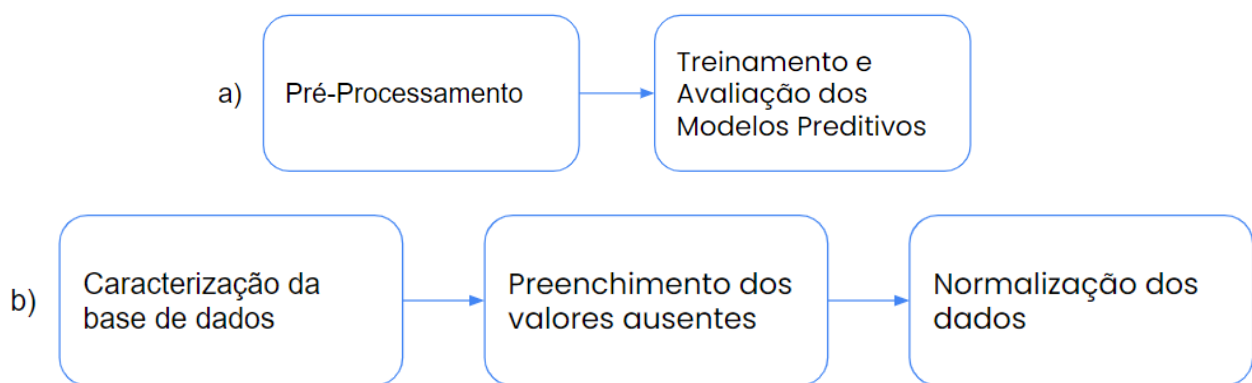
Neste trabalho, as proporções do conjunto inicial utilizadas para treinamento e teste foram aproximadamente 70%, 30%, respectivamente. Assim, o número de observações utilizadas em cada um dos conjuntos foram as seguintes: conjunto de treinamento – 329 registros; conjunto de teste – 141 registros.

4.2 Métodos

A metodologia utilizada neste trabalho, mostrada nos diagramas em blocos da Figura 4.2, é composta de três etapas: pré-processamento, avaliação de modelos de predição e teste do melhor modelo.

No pré-processamento é efetuada uma caracterização da base de dados, o preenchimento de valores ausentes e a normalização dos dados. Os seguintes modelos são utilizados na predição ou no preenchimento de valores ausentes: perceptron multicamadas, redes convolutivas, árvores de decisão, floresta randômica, máquina de vetores de suporte, regressão logística, regressão linear e K-vizinhos mais próximos. O desempenho do melhor modelo é então avaliado utilizando o conjunto de testes.

A seguir são descritas cada uma dessas etapas e as sub-etapas delas.



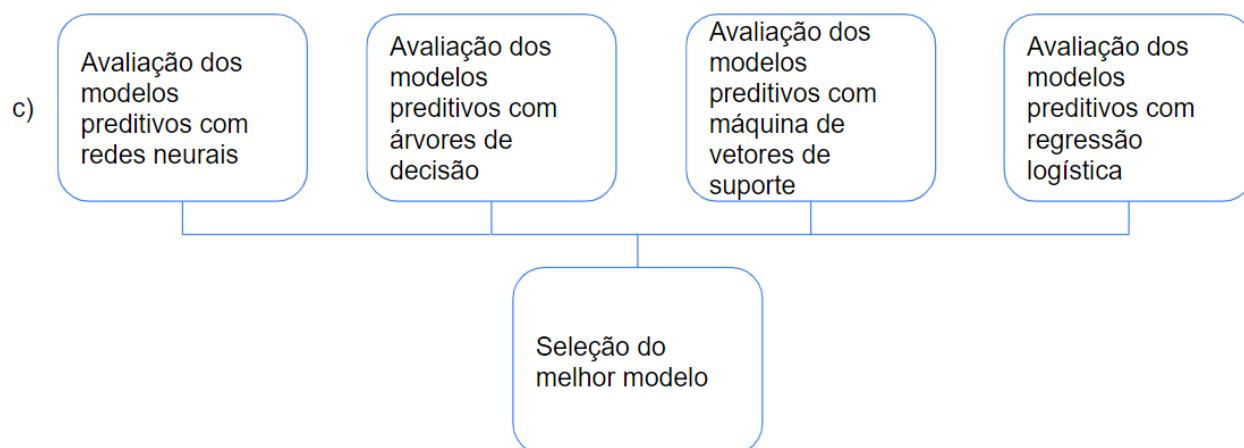


Figura 4.2 – (a) Fluxograma das Etapas da Metodologia; (b) Pré-processamento; (c) Treinamento e Avaliação dos Modelos Preditivos

4.2.1 Pré-processamento

4.2.1.1 Caracterização da base de dados

Ao inspecionar a base de dados, para entender como os dados estavam organizados, os nomes das colunas foram renomeados para simplificar a leitura das variáveis preditoras. Também foram verificados o tipo de dado de cada variável e a quantidade de valores ausentes, conforme mostrado na Tabela 4.1.

Tabela 4.1 – Características das 17 variáveis utilizadas nos métodos de predição

Nome Anterior da Coluna	Nome Atual da Coluna	Tipo de Variável	Quantidade de Valores Ausentes	% Valores Ausentes
Hipoxia	Hipoxia	Categórica	254	54%
Trauma_LocalGSC	ECG_Local_Trauma	Categórica	127	27%
hipot adm	Hipotensao_Admissao	Categórica	77	16%
Trauma_PSGCS	ECG_Admissao	Categórica	47	10%
Exames_TTPAR	Tempo_Trombo_Parcial	Numérica	39	8%
motor score	Pontuacao_Motora	Categórica	37	8%
Exames_TPINR	Tempo_Protombina	Numérica	35	7%
EF_PupilasReagente	Dilatacao_Pupilar	Categórica	33	7%
Idade	Idade	Numérica	11	2%
SAPS3	Severidade_Trauma	Numérica	3	1%

Sexo	Genero	Catagórica	0	0%
EC_TCcrânio_Desvio>5	TC_Desvio_Cranio	Catagórica	0	0%
EC_TCcrânio_Herniacao	Herniacao_Cerebral	Catagórica	0	0%
EC_TCcrânio_HSA	TC_Hemo_Subara	Catagórica	0	0%
EC_TCcrânio_Petequias	TC_Hemo_Subdural	Catagórica	0	0%
EC_TCcrânio_Hematoma	TC_Hemo_Intracerebral	Catagórica	0	0%
morte 14 dias	morte_em_14_dias	Catagórica	0	0%

A partir desta verificação, constatou-se que a quantidade de valores ausentes em algumas variáveis teria um forte impacto no momento de realizar as predições, inviabilizando a realização delas. Sendo assim, procedeu-se ao preenchimento desses valores ausentes.

4.2.1.2 Preenchimento de Dados Ausentes

O preenchimento de dados para substituição dos valores ausentes foi realizado utilizando a combinação de técnicas apresentadas na seção 3.5.2, do capítulo 3 - Referencial Teórico. As técnicas utilizadas foram utilização do valor mais provável para predição do valor ausente e utilização de uma medida central para predição do valor ausente.

Para a imputação dos dados, dois procedimentos usando as técnicas citadas foram propostos. Ao final, avaliou-se qual deles produziu melhores resultados, através de uma comparação com observações sem valores ausentes. A seguir, esses procedimentos são descritos.

Procedimento 1:

1. Ordenação das variáveis preditoras pelo critério de maior taxa de dados ausentes.
2. Seleção de uma variável para predição dos valores ausentes.
3. Preenchimento dos valores ausentes com uma medida de tendência central.
4. Realização da predição da variável selecionada utilizando os modelos escolhidos de acordo com o tipo de dado.
5. Substituição dos valores ausentes pelos valores preditos, sendo utilizado como parâmetro os valores provenientes do modelo com melhor desempenho.
6. Repetição dos passos de 2 a 5 para as demais variáveis.

7. Em variáveis que apresentam menor taxa de valores ausentes em relação à variável que está sendo predita, substituir os valores ausentes por uma medida de tendência central, pela moda se a variável for categórica ou pela média caso contrário.
8. Geração de uma nova base de dados com os valores ausentes preenchidos por meio de predição.

Exemplo: Seja T a lista de variáveis listadas em ordem decrescente em relação a porcentagem de dados ausentes: $T = [A, B, C, D, E, F, G]$

Para realizar a predição da variável B utilizaram-se os valores preditos de A, e os valores mais frequentes de C, D, F e G. E assim, sucessivamente até o término da predição de todos os valores ausentes da base de dados.

Procedimento 2:

1. Ordenação das variáveis preditoras pelo critério de maior taxa de valores ausentes.
2. Seleção de uma variável para predição dos valores ausentes.
3. Preenchimento dos valores ausentes com uma medida de tendência central.
4. Realização da predição da variável selecionada utilizando os modelos escolhidos de acordo com o tipo de dado.
5. Substituição dos valores ausentes pelos valores preditos, sendo utilizado como parâmetro os valores provenientes do modelo com melhor desempenho.
6. Após realizar a predição da variável com maior taxa de dados ausentes, realizar a remoção da mesma quando realizar a predição das variáveis seguintes.
7. Repetição dos passos de 2 a 6 para as demais variáveis.
8. Em variáveis que apresentaram menor taxa de dados ausentes, substituir os valores ausentes por uma medida de tendência central, moda se a variável for categórica ou pela média, caso contrário.
9. Geração de uma nova base de dados utilizando os valores ausentes preenchidos por meio da predição.

Exemplo: Seja T a lista de variáveis ordenadas em ordem decrescente de acordo com os dados ausentes, temos que: $T = [A, B, C, D, E, F, G]$

Portanto, para predição da variável E, são removidas as variáveis A, B, C e D, utilizando os valores mais frequentes de F e G.

Para realização do preenchimento das variáveis ausentes, foi removida a variável morte em 14 dias, que representa a variável a ser predita, para que não seja criado um viés em relação a essa variável. Para as 16 variáveis restantes, foram utilizados modelos em conformidade com o tipo de dado. Para variáveis categóricas foram utilizados modelos de classificação. Para variáveis numéricas foram utilizados modelos de regressão. Esses modelos são descritos na seção 4.2.2. Assim, modelos de classificação foram utilizados para o preenchimento de valores ausentes das seguintes variáveis: Hipóxia, ECG no Local do Trauma, Hipotensão na Admissão, ECG na Admissão, Pontuação Motora e Dilatação Pupilar. Por sua vez, modelos regressivos foram utilizados para o preenchimento de valores ausentes das seguintes variáveis: Tempo Tromboplastina Parcial e Tempo de Protrombina. Para obtenção de uma melhor confiabilidade estatística da capacidade de generalização dos modelos, foi aplicada o método de validação cruzada de k pastas, sendo o número de pastas $k=4$.

4.2.1.3 Normalização dos Dados

A normalização dos dados é aplicada nas variáveis Idade e Severidade de Trauma, por estas variáveis estarem em uma faixa de valores muito diferente das outras variáveis numéricas, conforme mostrado na Tabela 4.3. As outras variáveis numéricas são: Tempo de Protrombina e Tempo Tromboplastina parcial. Estas apresentam valores na faixa entre 0 e 10, o que está em uma faixa de valores próxima das variáveis categóricas. Para Idade e Severidade de Trauma, então, foram validados os seguintes métodos de normalização: raiz cúbica, mínimo máximo e pontuação z (z-score). Para essas variáveis, a Tabela 4.2 apresenta uma descrição dos valores máximos e mínimos:

Tabela 4.2 – Range de valores das variáveis selecionadas

Variável	Valor Mínimo	Valor Máximo
Idade	16	99
Severidade do Trauma	8	101

Após a normalização das variáveis, utilizando cada um dos métodos citados, as variáveis são inseridas novamente no conjunto de dados de treinamento, substituindo os valores anteriores delas. Para verificar qual método de normalização seria mais eficiente, um MLP foi utilizado para validação. A arquitetura utilizada (RNR1) é apresentada na seção 4.2.2.1.

É importante ressaltar que para as arquiteturas de Redes Neurais não precisaríamos realizar a normalização dessas variáveis, por ser um algoritmo que já faz o ajuste dos pesos em cada neurônio. No entanto, por estar sendo realizada a comparação com algoritmos mais tradicionais, decidimos por adotar a mesma metodologia para todos os métodos de predição.

4.2.2 Avaliação de Modelos

Conforme mencionado anteriormente, para obtenção de métricas de avaliação mais confiáveis, do ponto de vista estatístico, os modelos de predição utilizados nesse trabalho foram avaliados utilizando a técnica de validação cruzada. As métricas utilizadas para avaliação são descritas mais adiante.

4.2.2.1 Modelos Preditivos Neurais

Um total de 6 (seis) modelos preditivos utilizando redes neurais foram avaliados nesse trabalho na tarefa de predição de mortalidade de pacientes que sofreram traumas cranioencefálicos: 4 arquiteturas rasas: perceptron multicamadas 1 (MLP1), perceptron multicamadas 2 (MLP2), perceptron multicamadas 3 (MLP3), perceptron multicamadas 4 (MLP4), e duas arquiteturas profundas: rede neural convolutiva 1 (CNN1) e rede neural convolutiva 2 (CNN2).

Para o treinamento dessas arquiteturas, 4 (quatro) métodos de otimização foram utilizados e seus desempenhos comparados: gradiente descendente com momento (Stochastic Gradient Descent with Momentum - SGDM), propagação da raiz média quadrática (Root Mean Square Propagation – RMSprop), otimizador da estimativa do momento adaptativo (Adaptive Moment Estimation Optimizer – Adam) e Adamax.

Para melhoria da generalização das arquiteturas propostas, os seguintes métodos de regularização foram avaliados: sem nenhum método de regularização, *dropout*, regularização L_2 + *dropout*.

Em todas as arquiteturas que possuem camada de *dropout*, a taxa de *dropout* foi ajustada para $p = 0,2$.

A primeira arquitetura (MLP1), mostrada na Figura 4.3, tem a camada de entrada, duas camadas escondidas com 8 neurônios em cada camada, sendo a ativação das camadas escondidas realizada por meio da função ReLU. Em seguida é adicionada uma camada linear (*flatten*). Por último, tem-se a camada de classificação com ativação por meio da função sigmóide.

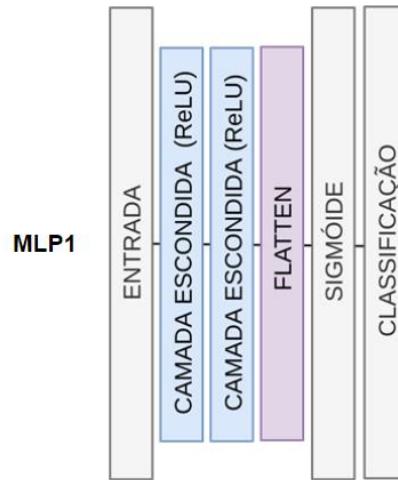


Figura 4.3 – Arquitetura da Rede MLP1

A segunda arquitetura (MLP2), mostrada na Figura 4.4, também tem a camada de entrada, com duas camadas escondidas com 8 neurônios em cada camada e ativação com a função ReLU, camada *dropout* e camada linear (*flatten*). Em seguida é adicionada a camada de classificação com ativação sigmóide.

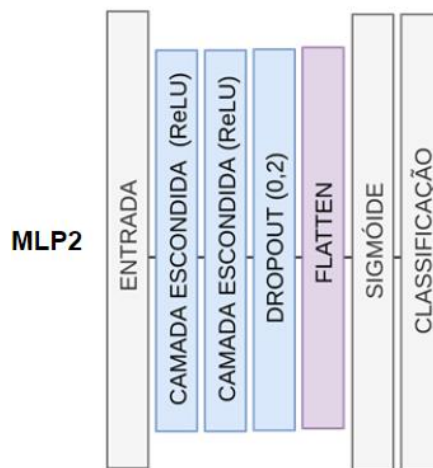


Figura 4.4 – Arquitetura da Rede MLP2

A terceira arquitetura (MLP3), mostrada na Figura 4.5, tem a camada de entrada, camada escondida com 8 neurônios e ativação com a função ReLU, camada de *dropout*, camada escondida novamente com 8 neurônios e ativação com a função ReLU e camada *dropout*. Em seguida temos uma camada linear (*flatten*) e camada de classificação com ativação por meio da função sigmóide.

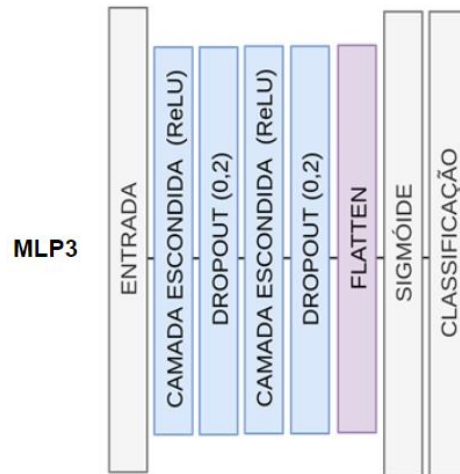


Figura 4.5 – Arquitetura da Rede MLP3

A quarta rede proposta (MLP4), mostrada na Figura 4.6, representa uma arquitetura rasa com camada de entrada, camada linear (*flatten*), duas sequenciais de camada escondida com 16 neurônios e função de ativação ReLU, camada de *dropout* após a camada escondida. Em seguida uma camada linear (*flatten*) e uma camada de classificação com ativação sigmóide.

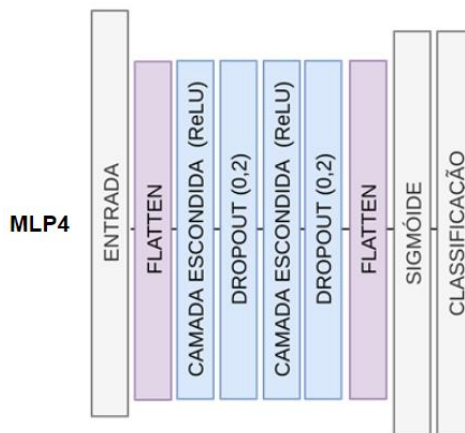


Figura 4.6 – Arquitetura da Rede MLP4

A quinta arquitetura (CNN1), mostrada na Figura 4.7, é uma arquitetura paralela, com camada de entrada, três sequências de duas camadas convolutivas e camada linear (*flatten*), com função de ativação ReLU, camada de concatenação na saída das camadas *flatten*, uma camada de *dropout* e uma camada escondida com 8 neurônios com a função de ativação pela função ReLU. Cada camada convolutiva tem 64 kernels de tamanho 3x1. Por fim, uma camada de classificação com ativação sigmóide.

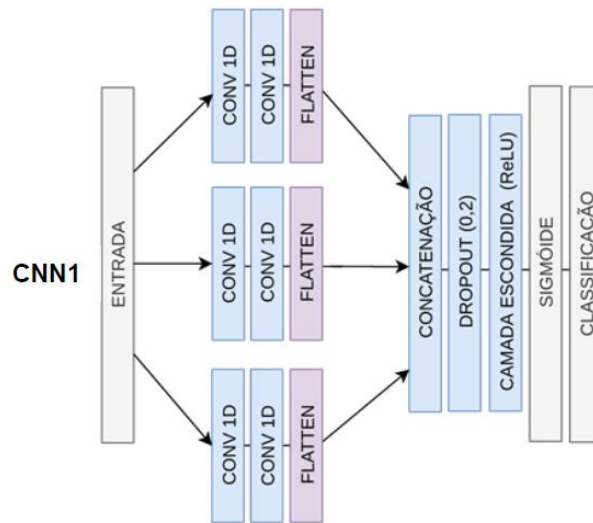


Figura 4.7 – Arquitetura da Rede CNN1

A sexta arquitetura proposta (CNN2), mostrada na Figura 4.8, é uma arquitetura paralela, com camada de entrada, três sequências de duas camadas convolutivas e camada linear (*flatten*), com função de ativação ReLU, camada de concatenação, camada escondida com 8 neurônios com função de ativação ReLU e camada de *dropout* após a camada escondida. Cada camada convolutiva tem 64 kernels de tamanho 3x1. Por último, uma camada de classificação com ativação sigmóide.

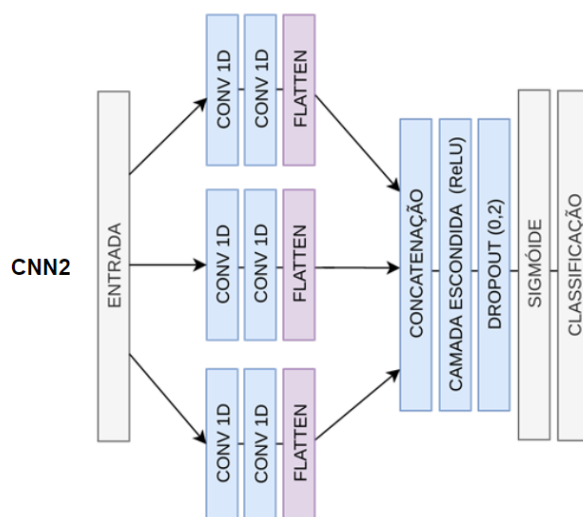


Figura 4.8 – Arquitetura da Rede CNN2

Para realizar o processo de treinamento das arquiteturas de redes neurais descritas anteriormente, foi necessário o ajuste de alguns parâmetros. Na Tabela 4.3 são mostrados os valores dos parâmetros de treinamento utilizados para cada um dos métodos de otimização propostos.

Tabela 4.3 – Parâmetros utilizados no treinamento das redes neurais

Taxa de aprendizagem	Métodos de Otimização	Método para melhoria da generalização	Métodos de Normalização	Função de Perda	Épocas de Treinamento
0,001	Adamax	sem regularização	z-score	Entropia binária cruzada	4000
0,003	RMSprop	dropout	Max-Min	Entropia binária cruzada	4000
0,007	Adam	dropout + regularização L2	Raiz cúbica	Entropia binária cruzada	4000
0,1	SGDM			Entropia binária cruzada	4000

Ao verificar os dados de treinamento, observou-se, conforme mostrado na Figura 4.9, que os dados estão desbalanceados, com a maior quantidade de observações sem morte em 14 dias (imagem abaixo). Para que seja possível realizar o treinamento de maneira adequada, sem que o modelo se especialize nos casos em que não há morte, é necessário realizar o balanceamento das classes.

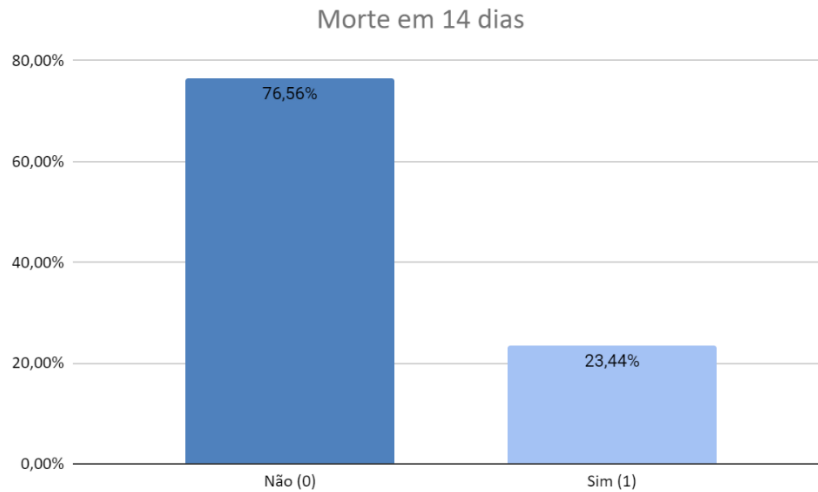


Figura 4.9 – Gráfico mostrando o desbalanceamento das classes

Portanto, fez-se necessária a correção do problema de desbalanceamento das classes de treinamento. Para realizar a correção, utilizou-se um método de balanceamento das classes durante o treinamento. Para tanto, multiplica-se a taxa de aprendizado por um peso inversamente proporcional à frequência de ocorrência da classe. Assim, ao apresentar-se na entrada da rede uma observação pertencente a uma classe com uma menor frequência de ocorrência, a taxa de aprendizado utilizada para atualização dos pesos e das polarizações é multiplicada por um fator multiplicativo maior, como mostrado na equação 4.1.

$$p_i = \frac{n}{(n_{classes} * n_{classe\ i})} \quad (4.1)$$

Em que: p_i é o peso da classe i , n o número total de observações, $n_{classes}$ sendo o número de classes presente e $n_{classe\ i}$ o número de observações da classe i .

4.2.2.2 Modelos Preditivos árvore de decisão, floresta randômica, k-vizinhos mais próximos, máquina de vetores de suporte, regressão logística e regressão linear

Os modelos árvore de decisão, floresta randômica, k-vizinhos mais próximos, máquina de vetores de suporte, regressão logística e regressão linear também foram utilizados ou para a

predição de mortalidade dos pacientes, ou para preenchimento de valores ausentes, ou para ambos. Nesta subsecção são descritos os parâmetros utilizados nesses modelos. A ferramenta de simulação utilizada foi o software *PyCharm* (JETBRAINS, 2021) com a biblioteca *scikit-learn* 0.23.2 (SCIKIT-LEARN, 2021) para todas as simulações deste trabalho.

A função de erro utilizada no modelo árvore de decisão, quando ele foi utilizado para tarefa de classificação, foi baseada no índice de Gini. Para a tarefa preditiva de regressão, foi utilizado o critério do erro médio quadrático. Ao utilizar-se a ferramenta de simulação *PyCharm* (JETBRAINS, 2021), os seguintes parâmetros comuns foram ajustados: `splitter='best'`, `max_depth=None`, `min_samples_split=2`, `min_samples_leaf=1`, `min_weight_fraction_leaf=0.0`, `max_features=None`, `random_state=None`, `max_leaf_nodes=None`, `min_impurity_decrease=0.0`, `min_impurity_split=None`, `class_weight=None`, `ccp_alpha=0.0`. Para a predição de mortalidade, é utilizado o balanceamento de classes através da variável *class_weight*.

De forma semelhante, no modelo floresta randômica, quando ele foi utilizado na classificação, utilizou-se uma função de erro baseada no índice de Gini, e, quando ele foi utilizado em tarefas de regressão, utilizou-se o erro quadrático médio. Ao utilizar-se a ferramenta de simulação *PyCharm* (JETBRAINS, 2021), os seguintes parâmetros comuns foram ajustados: `n_estimators=100`, `max_depth=None`, `min_samples_split=2`, `min_samples_leaf=1`, `min_weight_fraction_leaf=0.0`, `max_features='auto'`, `max_leaf_nodes=None`, `min_impurity_decrease=0.0`, `min_impurity_split=None`, `bootstrap=True`, `oob_score=False`, `n_jobs=None`, `random_state=None`, `verbose=0`, `warm_start=False`, `class_weight=None`, `ccp_alpha=0.0`, `max_samples=None`. Para a predição de mortalidade, é utilizado o balanceamento de classes.

Para o modelo k-vizinhos mais próximos, ao utilizar-se a ferramenta de simulação *PyCharm* (JETBRAINS, 2021), os seguintes parâmetros comuns foram ajustados: `weights='uniform'`, `algorithm='auto'`, `leaf_size=30`, `p=2`, `metric='minkowski'`, `metric_params=None`, `n_jobs=None`. Para o preenchimento de variáveis ausentes foram utilizados 2 vizinhos mais próximos.

Para o modelo máquina de vetores de suporte, ao utilizar-se a ferramenta de simulação *PyCharm* (JETBRAINS, 2021), os seguintes parâmetros comuns foram ajustados: `C=0.2`, `kernel='linear'`, `class_weight=class_weights`, `degree=3`, `gamma='scale'`, `coef0=0.0`,

shrinking=True, probability=False, tol=0.001, cache_size=200, verbose=False, max_iter=- 1, decision_function_shape='ovr', break_ties=False, random_state=None).

Para o modelo regressão logística, ao utilizar-se a ferramenta de simulação *PyCharm* (JETBRAINS, 2021), os seguintes parâmetros comuns foram ajustados: penalty='l2', dual=False, tol=0.0001, C=1.0, fit_intercept=True, intercept_scaling=1, class_weight=class_weights, random_state=None, solver='lbfgs', max_iter=4000, multi_class='auto', verbose=0, warm_start=False, n_jobs=None, l1_ratio=None

Para o modelo de regressão linear, ao utilizar-se a ferramenta de simulação *PyCharm* (JETBRAINS, 2021), os seguintes parâmetros comuns foram ajustados: fit_intercept=True, normalize=False, copy_X=True, n_jobs=None, positive=False.

Na Tabela 4.4 discrimina-se quais dos modelos anteriormente citados foram utilizados para predição de mortalidade e para preenchimento de valores ausentes.

Tabela 4.4 – Modelos utilizados para cada tipo de predição

Modelo	Predição de Valores Ausentes	Predição de Mortalidade	Singularidade
Árvore de Decisão	x	x	Parâmetro de balanceamento de classes para predição de mortalidade
Floresta Randômica	x	x	Parâmetro de balanceamento de classes para predição de mortalidade
K-Vizinhos Mais Próximos	x		
Máquina de Vetores de Suporte		x	
Regressão Logística		x	
Regressão Linear	x		

4.2.3 Métricas de Avaliação

Os desempenhos dos modelos apresentados nesse capítulo foram avaliados por meio de métricas das seguintes métricas de desempenho: acurácia, precisão, sensibilidade, especificidade e F1-Score. No cálculo dessas métricas, os seguintes conceitos foram empregados: Verdadeiro Positivo (VP), ocorre quando o modelo classifica um dado de entrada como positivo corretamente;

Verdadeiro Negativo (VN), ocorre quando o modelo classifica um dado de entrada como negativo, corretamente; Falso Positivo (FP), ocorre quando o modelo classifica um dado de entrada como positivo, erroneamente; Falso Negativo (FN), ocorre quando o modelo classifica um dado de entrada como negativo, erroneamente. Nesse trabalho, um dado ser classificado como positivo significou a mortalidade ser menor do que 14 dias e negativo, a mortalidade ser maior do que 14 dias. Com base nesses conceitos, definiremos as métricas anteriormente citadas:

- Acurácia: determina a proporção de acertos do modelo:

$$acurácia = \frac{VP + VN}{VP + FP + VN + FN}$$

- Precisão: determina a proporção de observações classificadas como positivas e que eram positivas.

$$precisão = \frac{VP}{VP + FP}$$

- Sensibilidade ou *recall*: determina a capacidade de um modelo prever corretamente que um dado de entrada é positivo. Quanto maior a sensibilidade menos falso positivos são detectados pelo modelo.

$$sensibilidade = \frac{VP}{VP + FN}$$

- Especificidade: determina a capacidade de um modelo prever corretamente que um dado de entrada é negativo. Quanto maior o valor da especificidade menos são os falsos positivos detectados pelo modelo.

$$especificidade = \frac{VN}{VN + FP}$$

- F1-score: resultado da combinação da precisão e da sensibilidade, demonstrando a qualidade geral do modelo e estabelecendo um equilíbrio entre precisão e sensibilidade.

$$f1 - score = 2 * \frac{precisão * sensibilidade}{precisão + sensibilidade}$$

Além das métricas anteriormente citadas, o poder de discriminação dos modelos de classificação foi avaliado através da área sob a curva “Característica de Operação do Receptor” (*Receiver Operating Characteristic* – ROC). Por outro lado, a precisão de predição dos modelos regressivos foi avaliada através do erro médio quadrático. A seguir, detalhamos melhor essas medidas.

- Área sob a curva (AUC) ROC: demonstra a variação entre a sensibilidade e a especificidade para diferentes pontos de corte na probabilidade estimada. Quanto mais alto o valor de AUC, maior o poder de discriminação do modelo.
- Erro Quadrático Médio (EQM): mostra o quanto a linha de regressão está próxima dos pontos de observação fornecidos. Assim, quanto menor o valor de MSE, mais precisa é a predição.

$$EQM = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$

Sendo y_i o valor atual e \tilde{y}_i o valor predito e n o número de observações.

4.2.4 Ambiente de Desenvolvimento

Para desenvolvimento das arquiteturas de algoritmos auxiliares foi utilizado o software *Pycharm* (JETBRAINS, 2021), na versão estudante universitário. Os treinamentos, validações e testes das arquiteturas propostas foram executados em um computador com as seguintes especificações: processador Intel® Core™ i7-7500U CPU @ 2.70GHz × 4, memória RAM 16GB.

As bibliotecas utilizadas para auxiliar no desenvolvimento foram: *tensorflow* na versão 2.2.1, *Keras* na versão 2.4.3, *scikit-learn* na versão 0.23.2. A linguagem de programação escolhida foi Python 3.6. A biblioteca *pandas* na versão 1.0.3 foi utilizada para trabalharmos com as tabelas dos conjuntos de dados, a biblioteca *matplotlib* na versão 3.2.2 foi utilizada para visualização.

5 RESULTADOS E DISCUSSÃO

Neste capítulo serão abordados os resultados obtidos, tanto no que diz respeito aos métodos de preenchimento de valores ausentes, como de avaliação de desempenho dos modelos preditivos apresentados no capítulo anterior. Por fim, será efetuada uma comparação com o resultado de outros trabalhos publicados na literatura.

5.1 Resultados do Preenchimento dos Valores Ausentes

Nas Tabelas 5.1 e 5.2 são apresentados resultados para avaliação do “Procedimento 1”, visando o preenchimento de valores ausentes das variáveis categóricas e numéricas, respectivamente, apresentado na Seção 4.2.1.2 – “Preenchimento dos valores ausentes”. Em função do tipo da variável preditora, categórica ou numérica, são apresentados resultados de regressão ou de classificação, respectivamente. A apresenta resultados da acurácia de classificação para as variáveis categóricas, enquanto a Tabela 5.2 apresenta resultados do erro médio quadrático da regressão para as variáveis numéricas. Em negrito são marcados os melhores resultados de preenchimento para cada variável preditora.

Tabela 5.1 – Acurácia dos modelos de classificação para as variáveis categóricas, utilizando o Procedimento 1

Variáveis	Modelos		
	Árvore de Decisão	Floresta Randômica	KNN
Hipoxia	0,7373	0,8093	0,7754
ECG Local Trauma	0,5084	0,4321	0,2432
Hipotensão Admissão	0,7985	0,8707	0,8209
ECG Admissao	0,4543	0,4471	0,3341
Pontuação Motora	0,9012	0,7977	0,5242
Dilatação Pupilar	0,6811	0,7768	0,5946

* Em negrito são marcados os melhores resultados para cada variável

Tabela 5.2 – MSE dos modelos de regressão para as variáveis numéricas, utilizando o Procedimento 1

Variáveis	Modelos		
	Árvore de Decisão	Floresta Randômica	Regressão Linear
Tempo Tromboplastina Parcial	0,2613	0,1539	0,1182
Tempo Protrombina	0,2613	0,1661	0,1121

* Em negrito são marcados os melhores resultados para cada variável

Nas Tabelas 5.3 e 5.4 são apresentados os resultados para avaliação do “Procedimento 2”, visando o preenchimento de valores ausentes das variáveis categóricas e numéricas, respectivamente, apresentado na Seção 4.2.1.2 – “Preenchimento dos valores ausentes”. Em função do tipo da variável preditora, categórica ou numérica, são apresentados resultados de regressão ou de classificação, respectivamente. A Tabela 5.3 apresenta resultados da acurácia de classificação para as variáveis categóricas, enquanto a Tabela 5.4 apresenta resultados do erro médio quadrático da regressão para as variáveis numéricas. Em negrito são marcados os melhores resultados de preenchimento para cada variável preditora.

Tabela 5.3 – Acurácia dos modelos de classificação para as variáveis categóricas, utilizando o Procedimento 2

Variáveis	Modelos		
	Árvore de Decisão	Floresta Randômica	KNN
Hipóxia	0,7373	0,8051	0,7754
ECG Local Trauma	0,5027	0,4689	0,246
Hipotensão Admissão	0,7662	0,8407	0,8309
ECG Admissão	0,3646	0,4164	0,3012
Pontuação Motora	0,3288	0,3862	0,3103
Dilatação Pupilar	0,6741	0,7745	0,581

* Em negrito são marcados os melhores resultados para cada variável

Tabela 5.4 – MSE dos modelos de regressão para as variáveis numéricas, utilizando o Procedimento 2

Variáveis	Modelos		
	Árvore de Decisão	Floresta Randômica	Regressão Linear
Tempo Tromboplastina Parcial	0,2615	0,1539	0,1137
Tempo Protrombina	0,2738	0,2262	0,1813

* Em negrito são marcados os melhores resultados para cada variável

Comparando os resultados apresentados nas tabelas 5.1, 5.2, 5.3 e 5.4, conclui-se que o procedimento 1 obteve melhor acurácia na maioria das variáveis analisadas. Assim temos que: 1) para predição das variáveis hipóxia, ecg local trauma, hipotensão admissão e dilatação pupilar, será utilizado o método floresta randômica; 2) para as variáveis ECG local trauma, ecg admissão e pontuação motora será utilizado o método árvore de decisão; 3) para as variáveis tempo tromboplastina parcial e tempo protrombina será utilizado o método de regressão linear.

As variáveis idade e severidade do trauma, por apresentarem uma taxa de variáveis ausentes menor que 3% tiveram o preenchimento dos valores ausentes realizados pela média dos valores. Assim, o conjunto de dados não apresenta mais valores ausentes, sendo possível realizar previsões com todos os valores das variáveis preditoras. Na Figura 16 é possível observar as 5 primeiras linhas do novo conjunto de dados.

	Genero	Idade	Dilatacao_Pupilar	ECG_Local_Trauma	ECG_Admissao	Pontuacao_Motora	Hipoxia	Hipotensao_Admissao	TC_Desvio_Cranio
0	1	19.0	0.0	6.0	6.0	4.0	1.0	0.0	0
1	2	40.0	2.0	11.0	11.0	5.0	0.0	0.0	0
2	1	61.0	1.0	8.0	3.0	5.0	0.0	0.0	0
3	1	80.0	2.0	10.0	6.0	5.0	1.0	1.0	0
4	2	60.0	2.0	10.0	11.0	5.0	0.0	0.0	1

Figura 5.1 – Recorte das 5 primeiras observações do conjunto de dados após preenchimento dos valores ausentes

5.2 Resultados da Normalização de Variáveis

Ao analisar os dados, observou-se que as variáveis numéricas idade e severidade do trauma não estão normalizadas, contendo valores em uma faixa bem diferente das demais variáveis.

Portanto, realizou-se a normalização dessas variáveis através das seguintes técnicas de normalização: mínimo-máximo (min-max), z-score e raiz cúbica. Na Tabela 5.5 mostra-se o desempenho das três técnicas de normalização, utilizando-se modelo de previsão MLP1, e a métrica AUC, apresentada no capítulo anterior.

Tabela 5.5 – Área sob a curva ROC (AUC) obtida para o modelo MLP1 para diferentes técnicas de normalização dos dados

Modelo	Técnica de normalização		
	Raiz Cúbica	Z-score	Min-Max
MLP1	0,8694	0,8669	0,8739

Conforme observa-se na Tabela 5.5, o melhor desempenho do modelo MLP1 foi obtido com a técnica de normalização min-max. Sendo assim, doravante, nos resultados a serem apresentados posteriormente, será utilizada essa técnica de normalização dos dados.

5.3 Resultados dos Métodos de Otimização para as Redes Neurais

Esta seção apresenta o desempenho dos modelos propostos para predição da morte em 14 dias: quatro arquiteturas de redes MLP, duas arquiteturas de redes CNN, apresentados na Seção 4.2.2.1, e dos modelos preditivos árvores de decisão, floresta randômica, regressão logística e máquina de vetores de suporte, apresentados na Seção 4.2.2.2. As métricas utilizadas foram descritas na Seção 4.2.3.

Inicialmente, fez-se uma avaliação dos seguintes modelos preditivos: MLP1, regressão logística, máquinas de vetores de suporte, árvore de decisão e floresta randômica. Os resultados para a predição de mortalidade, sem o balanceamento das classes, são mostrados na Tabela 5.6.

Tabela 5.6 – Comparação inicial dos modelos sem o balanceamento de classes

Modelo	AUC	Acurácia
MLP1	0,8719	0,8451
Regressão Logística	0,7535	0,8591
Máquina de Vetor de Suporte	0,7272	0,845
Árvore de Decisão	0,6553	0,7887
Floresta Randômica	0,6842	0,8309

Usando o balanceamento de classes, os resultados para a predição da mortalidade, são apresentados na Tabela 5.7:

Tabela 5.7 – Avaliação inicial dos modelos utilizando o balanceamento de classes para MLP1

Modelo	AUC	Acurácia
MLP1	0,8613	0,8309
Regressão Logística	0,7844	0,8309
Máquina de Vetor de Suporte	0,8011	0,8309
Árvore de Decisão	0,6766	0,7464
Floresta Randômica	0,6219	0,7887

A partir das Tabelas 5.6 e 5.7, verifica-se que o modelo MLP1 apresentou melhor desempenho quando comparado aos outros modelos com a métrica AUC-ROC. Para alguns modelos, o desempenho com o balanceamento de classes é superior. Para outros, ocorre o contrário. Doravante, serão avaliados diversos métodos de otimização utilizando-se apenas redes neurais, tanto MLPs como CNNs. Para isso, foram selecionados os seguintes métodos de otimização: Adamax, RMSprop, SGDM, Adam. As Tabelas 5.8, 5.9, 5.10, 5.11, 5.12 e 5.13 mostram os resultados de predição da mortalidade obtidos com as arquiteturas MLP1, MLP2, MLP3, MLP4, CNN1 e CNN2, respectivamente, com os dados balanceados.

Tabela 5.8 – Resultados para a predição da mortalidade obtidos com o modelo MLP1 usando diversos métodos de otimização

Método de otimização	AUC	F1-score	Acurácia
Adamax	0,8694	0,6809	0,7887
RMSprop	0,876	0,6667	0,7606
Adam	0,8739	0,6667	0,7746
SGDM (taxa de aprendizagem = 0.1, momento = 0.01)	0,8785	0,6667	0,8169
SGDM (taxa de aprendizagem = 0.3, momento = 0.001)	0,8097	0,4928	0,507
SGDM (taxa de aprendizagem = 0.007, momento = 0.001)	0,8694	0,7143	0,831

Tabela 5.9 – Resultados para a predição da mortalidade obtidos com o modelo MLP2 usando diversos métodos de otimização

Método de otimização	AUC	F1-score	Acurácia
RMSprop	0,8719	0,6667	0,7606
Adam	0,8709	0,68	0,7746
SGDM (taxa de aprendizagem = 0.1, momento = 0.01)	0,8775	0,5862	0,662

Tabela 5.10 – Resultados para a predição da mortalidade obtidos com o modelo MLP3 usando diversos métodos de otimização

Método de otimização	AUC	F1-score	Acurácia
RMSprop	0,8866	0,4578	0,3662
Adam	0,8765	0,6415	0,7324
SGDM (taxa de aprendizagem = 0.1, momento = 0.01)	0,8755	0,6383	0,7606

Tabela 5.11 – Resultados para a predição da mortalidade obtidos com o modelo MLP4 usando diversos métodos de otimização

Método de otimização	AUC	F1-score	Acurácia
RMSprop	0,8618	0,5957	0,7324
Adam	0,8618	0,6512	0,7887
SGDM (taxa de aprendizagem = 0.1, momento = 0.01)	0,9091	0,6857	0,8451
SGDM (taxa de aprendizagem = 0.01, momento = 0.01)	0,8831	0,5806	0,6338
SGDM (taxa de aprendizagem = 0.0009, momento = 0.9)	0,9114	0,6296	0,7183

Tabela 5.12 – Resultados para a predição da mortalidade obtidos com o modelo CNN1 usando diversos métodos de otimização

Método de otimização	AUC	F1-score	Acurácia
Adam	0,8441	0,5714	0,7465
RMSprop	0,8345	0,6061	0,8169
Adamax	0,8522	0,6383	0,7606
SGDM (taxa de aprendizagem = 0.007, momento = 0.01)	0,8502	0,6977	0,8169
SGDM (taxa de aprendizagem = 0.1, momento = 0.01)	0,8831	0,625	0,831
SGDM (taxa de aprendizagem = 0.3, momento = 0.001)	0,8826	0,667	0,8169

Tabela 5.13 – Resultados para a predição da mortalidade obtidos com o modelo CNN2 usando diversos métodos de otimização

Método de otimização	AUC	F1-score	Acurácia
Adam	0,8441	0,6154	0,7183
RMSprop	0,8395	0,5714	0,662
Adamax	0,8537	0,5161	0,7465
SGDM (taxa de aprendizagem = 0.007, momento = 0.01)	0,8694	0,5424	0,6197
SGDM (taxa de aprendizagem = 0.1, momento = 0.01)	0,8927	0,6977	0,8169
SGDM (taxa de aprendizagem = 0.3, momento = 0.001)	0,5	0,2676	0,4222

5.4 Discussão

Inicia-se a discussão comentando-se sobre a metodologia utilizada para realizar a imputação dos valores ausentes das variáveis preditoras selecionadas para predição de mortalidade de pacientes. Dois procedimentos foram propostos para imputação dos valores ausentes, os quais tem as variáveis ordenadas de forma decrescente, com base na quantidade de valores ausentes. No Procedimento 1, à medida que os valores são preditos, as variáveis com os valores completos são utilizadas na predição dos valores ausentes da variável seguinte; no Procedimento 2, após a predição dos valores ausentes de uma variável preditora, a mesma não é mais utilizada para predição das variáveis seguintes. O Procedimento 1, que apresentou os melhores desempenhos de modelos, foi utilizado para a predição das variáveis ausentes.

Para realizar a predição das variáveis ausentes, alguns modelos supervisionados foram utilizados, sendo eles: árvore de decisão, floresta randômica, k-vizinhos mais próximos e regressão linear. Na predição da mortalidade, utiliza-se os valores preditos com o modelo que apresentou melhor desempenho na predição de cada variável. De forma geral, para as variáveis categóricas os modelos de floresta randômica e árvore de decisão apresentaram o melhor desempenho, e, para as variáveis numéricas, o melhor modelo foi a regressão linear.

Ainda na etapa de pré-processamento foram aplicadas técnicas de normalização nas variáveis numéricas idade e severidade do trauma, de modo que as faixa dinâmica de valores dessas variáveis sejam semelhantes. Avaliando o desempenho do modelo MLP1, verificamos que a técnica de Min-Max sobressaiu-se sobre as outras técnicas. Corrigiu-se também o desbalanceamento das classes: a classe “Não” compreendia aproximadamente 75% dos valores.

Conforme mostrado anteriormente, o balanceamento das classes foi feito através do ajuste da taxa de aprendizado para os modelos de redes neurais. Efetuando-se o balanceamento de classes, observa-se há uma pequena diferença no desempenho do modelo MLP1. Todos os modelos foram avaliados utilizando-se o balanceamento das classes.

Na etapa de classificação da mortalidade dos pacientes com trauma cranioencefálico, temos a construção de seis arquiteturas de redes neurais, sendo 4 arquiteturas de redes neurais tipo perceptron multicamadas e 2 arquiteturas de redes neurais convolucionais. Além dessas arquiteturas, os modelos de floresta randômica, árvore de decisão, máquina de vetores de suporte e regressão logística foram utilizados para a predição de mortalidade.

Ao comparar-se os resultados obtidos com redes neurais, mostrados nas Tabelas 5.8 a 5.12, com os resultados obtidos com métodos convencionais de aprendizado de máquina, como árvores de decisão, floresta randômica, regressão logística e máquina de vetores de suporte, mostrados na Tabela 5.7, observa-se que o desempenho das redes neurais foi superior ao desempenho dos métodos convencionais quando comprando-se a métrica de AUC-ROC.

No que tange aos modelos que utilizam redes neurais, no treinamento dos mesmos utilizou-se diversos métodos de otimização: RMSProp, ADAM, Adamax e SGDM. De uma forma geral, o método SGDM se sobrepôs aos outros métodos de otimização, destacando-se nas arquiteturas MLP1, MLP2, MLP4, CNN1 e CNN2.

Avaliando-se os resultados apresentados nas Tabelas 5.8 a 5.12, observa-se que a maioria dos modelos neurais apresentou uma área sob a curva ROC superior a 0,85. Os melhores valores para área sob a curva ROC e para a acurácia, 0,911 e 0,845, respectivamente, foram obtidos com o modelo MLP4, método de otimização SGDM e método de generalização *dropout*. O modelo MLP4 é o mais complexo dentre os modelos de redes rasas. Já o melhor valor para a métrica F1-score, 0,714, foi obtido com o modelo MLP1, o modelo MLP menos complexo.

5.5 Comparação dos Resultados obtidos com a Literatura

Ao observar os trabalhos da literatura apresentados no Quadro 1, observa-se que o número de variáveis preditoras varia entre 8 e 40. Dois trabalhos, o de Amorim et al. (2020) e o de Tunthannathip e Oearsaku (2021), utilizaram 14 variáveis, o mesmo número de variáveis utilizadas nesse trabalho. Ressalta-se, no entanto, que o trabalho de Amorim et al. (2020) utilizou a mesma

base de dados e as mesmas variáveis preditoras, enquanto o trabalho de Tunthannathip e Oearsaku (2021) utilizou outra base de dados e outras variáveis preditoras.

Nos trabalhos que utilizaram 14 variáveis preditoras, o de Amorim et al. (2020) obteve uma área sob a curva ROC de 0,906, enquanto o trabalho de Tunthannathip e Oearsaku (2021) obteve uma área sob a curva ROC de 0,78. No trabalho ora apresentado, uma área sob a curva ROC de 0,911 foi obtida com o modelo MLP4, como mostrado na Figura 5.2.

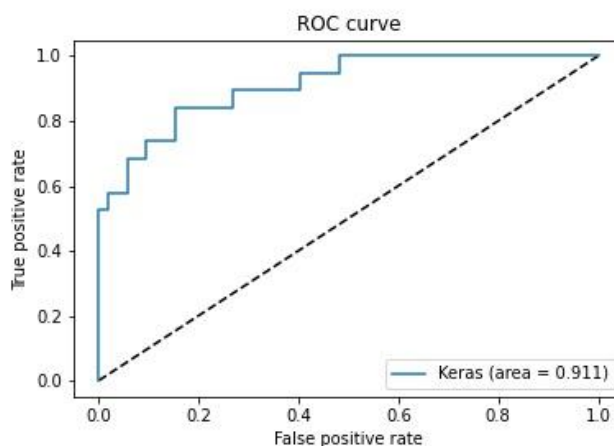


Figura 5.2 – Gráfica de AUC – ROC de MLP4

Registramos, por último que o trabalho ora apresentado foi o primeiro trabalho da literatura a avaliar o desempenho de redes profundas, mais especificamente, o de redes convolutivas no prognóstico de TCE.

6. CONCLUSÃO

As principais contribuições desse trabalho foram as seguintes: 1) a avaliação de redes neurais profundas, mas especificamente, redes neurais convolucionais, para realizar a tarefa de predição de mortalidade de pacientes que tiveram trauma cranioencefálico. Embora os resultados obtidos apontem para um melhor desempenho das redes neurais rasas (MLPs) frente as redes neurais profundas (CNNs), acreditamos que um trabalho mais profundo, avaliando outras arquiteturas de redes convolutivas, seja necessário, para uma conclusão definitiva sobre o tema, pois as redes neurais se beneficiam mais quando há grande volume de dados para treinamento; 2) a avaliação de dois procedimentos para o preenchimento de valores ausentes, mostrando a superioridade do Procedimento1, que utiliza variáveis preditoras previamente preenchidas para o

preenchimento das outras. Essa contribuição é de suma importância, uma vez que os dados médicos, mormente em países de baixa e média renda, a qualidade dos dados deixa a desejar, seja pela urgência de se lidar com o paciente, seja pela qualidade da mão de obra.

Como trabalhos futuros sugere-se: 1) exploração de novas arquiteturas de redes profundas; 2) utilização de métodos de análise de variáveis, para remoção de variáveis com informações redundantes, como o método de seleção escalar de características; 3) avaliação de metodologia proposta com outras bases de dados mais robustas; 4) desenvolvimento de uma plataforma web para utilização dos modelos de forma pública.

5 REFERÊNCIAS

AGGARWAL, C. C. *Neural Networks and Deep Learning*. New York: Springer, 2018.

AMORIM, Robson Luis, et al. Prediction of early TBI mortality using a machine learning approach in a LMIC population. *Frontiers in neurology* 10, p. 1366, 2020. <https://doi.org/10.3389/fneur.2019.01366>. Acesso em: 28 jan. 2022.

AWAD, M.; KHANNA, R. *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*. New York: Apress Open, 2015.

BARNARD, John; MENG, Xiao-Li. *Applications of multiple imputation in medical studies: from AIDS to NHANES*, 1999.

CHOU DHURY, Ambika; GUPTA, Deepak. A survey on medical diagnosis of diabetes using machine learning techniques. In: *Recent developments in machine learning and data analytics*. Springer, Singapore, p. 67-78, 2019.

COLLABORATORS, M. C. T. et al. "Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients." *Bmj* 336, no. 7641, 2008.

DA SILVA, L. M. O. *Uma Aplicação de Árvores de Decisão, Redes Neurais e KNN para a Identificação de Modelos ARMA Não-Sazonais e Sazonais*. Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro, p. 145, 2005.

ESFANDIARI, Nura et al. Knowledge discovery in medicine: Current issue and future trend. *Expert Systems with Applications*, v. 41, n. 9, p. 4434-4463, 2014.

GARCÍA, S., Luengo, J., & Herrera, F. *Data Preprocessing in Data Mining*. Intelligent Systems Reference Library, 2015.

GONZALEZ, L. D. A. *Regressão logística e suas aplicações*, 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [s.l.] The MIT Press, 2016.

FENG, Jin-zhou et al. Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries. *Journal of critical care*, v. 54, p. 110-116, 2019.

HALE, Andrew T. et al. Using an artificial neural network to predict traumatic brain injury. *Journal of Neurosurgery: Pediatrics*, v. 23, n. 2, p. 219-226, 2018.

HAN, Jiawei; PEI, Jian; KAMBER, Micheline. *Data mining: concepts and techniques*. Elsevier, 2011.

HOFMAN, K. et al. (2005). Addressing the growing burden of trauma and injury in low- and middle-income countries. *American Journal of Public Health*, 95(1), 13–17, 2005. <https://doi.org/10.2105/AJPH.2004.039354>. Acesso em: 28 jan. 2022.

HUKKELHOVEN, C. W. P. M. et al. Predicting survival using simple clinical variables: A case study in traumatic brain injury (multiple letters). *Journal of Neurology Neurosurgery and Psychiatry*, 68(3), 396–397, 2000. <https://doi.org/10.1136/jnnp.68.3.396>. Acesso em: 28 jan. 2022.

JAYATILAKE, Senerath Mudalige Don Alexis Chinthaka; GANEGODA, Gamage Upeksha. Involvement of Machine Learning Tools in Healthcare Decision Making. *Journal of Healthcare Engineering*, v. 2021, 2021.

JETBRAINS: The Python IDE for Professional Developers. Disponível em: <https://www.jetbrains.com/pycharm/>. 2021. Acesso em: 28 jan. 2022.

JIANG, Xiaoyue, et al. *Deep Learning in object detection and recognition*. Singapore: Springer, 2019.

JIAWEI, Han. *Data Mining: Concepts and Techniques*, 3rd ed. 2012.

KAMAL, V., AGRAWAL, D., & PANDEY, R. (2016). Epidemiology, clinical characteristics and outcomes of traumatic brain injury: Evidences from integrated level 1 trauma center in India. *Journal of Neurosciences in Rural Practice*, 7(4), 515–525. <https://doi.org/10.4103/0976-3147.188637>. Acesso em: 28 jan. 2022.

KINGMA, Diederik P.; BA, Jimmy. A Method for Stochastic Optimization. 3rd International Conference for Learning Representations (ICLR). San Diego, 2015.

LEITE, Danilo Rangel Arruda; DE MORAES, Ronei Marcos; LOPES, Leonardo Wanderley. Método de Aprendizagem de Máquina para Classificação da intensidade do desvio vocal utilizando Random Forest. *Journal of Health Informatics*, v. 12, 2021.

LI, Yu-Chuan et al. Neural network modeling for surgical decisions on traumatic brain injury patients. *International journal of medical informatics*, v. 57, n. 1, p. 1-9, 2000. [https://doi.org/10.1016/S1386-5056\(99\)00054-4](https://doi.org/10.1016/S1386-5056(99)00054-4). Acesso em: 28 jan. 2022.

MULENGA, Mwenge et al. Feature Extension of Gut Microbiome Data for Deep Neural Network-Based Colorectal Cancer Classification. *IEEE*, 2021. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9319639>. Acesso em: 28 jan. 2022.

MURRAY, L. S. et al. Does prediction of outcome alter patient management? *The Lancet*, 341(8859), 1487–1491. 1993. Disponível em: [https://doi.org/10.1016/0140-6736\(93\)90631-P](https://doi.org/10.1016/0140-6736(93)90631-P).

MURRAY, C.; LOPEZ, A. *The global burden of disease*. Vol. 1. Cambridge, MA: Harvard University Press; 1996.

NWOMEH, B. et al. History and development of trauma registry: lessons from developed to developing countries. *World Journal of Emergency Surgery* I:32, 2006.

OMS: Organização Mundial da Saúde. Injuries and violence (Issue March 2021). 2021a. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/injuries-and-violence>

OMS: Organização Mundial da Saúde. World health statistics 2021: monitoring health for the SDGs, sustainable development goals. 2021b. Disponível em: <https://emea.mitsubishielectric.com/ar/products-solutions/factory-automation/index.html>

OMS: Organização Mundial da Saúde. Global status report on road safety 2018. 2018. Disponível em: <https://www.who.int/publications/i/item/9789241565684>

PEREL, P. et al. Prognosis following head injury: A survey of doctors from developing and developed countries. *Journal of Evaluation in Clinical Practice*, 13(3), 464–465. 2007. Disponível em: <https://doi.org/10.1111/j.1365-2753.2006.00713.x>. Acesso em: 28 jan. 2022.

PLAWIAK, P. “Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals,” *Swarm Evol. Comput.*, vol. 39, no. April, pp. 192–208, 2018. doi: 10.1016/j.swevo.2017.10.002.

PRINCE, John; DE VOS, Maarten. "A deep learning framework for the remote detection of Parkinson's disease using smart-phone sensor data." 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018.

RAJ, Rahul et al. Machine learning-based dynamic mortality prediction after traumatic brain injury. *Scientific reports*, v. 9, n. 1, p. 1-13, 2019.

REHBEIN, Matheus Henrique. Estudo comparativo de Deep Learning e regressão linear na predição de mensagens processadas pela plataforma de integração guaraná. Trabalho de Conclusão de Curso (Graduação em e Engenharias da Universidade Regional do Noroeste do Estado do Rio Grande do Sul. Rio Grande do Sul, 2020.Ciência da Computação) - Departamento de Ciências Exatas

RUDER, Sebastian. An overview of gradient descent optimization algorithms. arXiv:1609.04747, 2016.

SAMANAMALEE, Samitha et al. Traumatic brain injury (TBI) outcomes in an LMIC tertiary care centre and performance of trauma scores. *Traumatic brain injury (TBI) outcomes in an LMIC tertiary care centre and performance of trauma scores. BMC anesthesiology*, 18(1), 1-7. 2018.

SCHMUCKER, U. et al. Road traffic crashes in developing countries. *Unfallchirurg*. 2010 May;113(5):373-7. German. doi: 10.1007/s00113-010-1777-9. PMID: 20376616.

SCIKIT-LEARN: Machine Learning in Python. Disponível em: <https://scikitlearn.org/>. 2021. Acesso em: 28 jan. 2022.

SERRÃO, Mikaela Kalline Maciel. Detecção automática de bacilos em baciloscopia de campo claro usando aprendizado profundo e técnica de imagem mosaico. Dissertação (Mestrado em Engenharia Elétrica) - Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, Universidade Federal do Amazonas. Amazonas, 2020.

SIGNORINI, David F., et al. Predicting survival using simple clinical variables: a case study in traumatic brain injury. *Journal of Neurology, Neurosurgery & Psychiatry* 66.1. p. 20-25, 1999.

SINGH, D., & SINGH, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97, 105524. 2020.

SOMASUNDARAM, R.S.; NEDUNCHEZHIAN, R. Evaluation of three Simple Imputation Methods for Enhancing Preprocessing of Data with Missing Values. 2011.

STEYERBERG, Ewout W. et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med.* 2008 Aug 5;5(8): e165; discussion e165. doi: 10.1371/journal.pmed.0050165. PMID: 18684008; PMCID: PMC2494563.

THORNHILL, S. et al. Disability in young people and adults one year after head injury: prospective cohort study. *Bmj*, 320(7250), p. 1631-1635, 2000.

TUNTHANATHIP, Thara; OEARSAKUL, Thakul. Application of machine learning to predict the outcome of pediatric traumatic brain injury. *Chinese Journal of Traumatology*, 2021.

UMERANI, M. S.; ABAS, A.; SHARIF, S. Traumatic brain injuries: Experience from a tertiary care center in Pakistan. *Turkish Neurosurgery*, 24(1), 19–24, 2014. Disponível em: <https://doi.org/10.5137/1019-5149.JTN.7080-12.1>. Acesso em: 28 jan. 2022.

UTSCH, Kaio Giurizzato. Uso de Redes Neurais Convolucionais para Classificação de Imagens Digitais de Lesões de Pele. Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica) - Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo. Espírito Santo, 2018.

WHITENECK, Gale et al. Long-term outcomes after traumatic brain injury. *The Journal of Head Trauma Rehabilitation*, v. 21, n. 5, p. 435, 2006. Disponível em: <https://doi.org/10.1097/00001199-200609000-00060>. Acesso em: 28 jan. 2022.

WORLD HEALTH ORGANIZATION. Global status report on road safety 2018. 2018. Disponível em: <https://www.who.int/publications/i/item/9789241565684>. Acesso em: 28 jan. 2022.

WORLD HEALTH ORGANIZATION. World health statistics 2021: monitoring health for the SDGs, sustainable development goals. 2021b. Disponível em: <https://emea.mitsubishielectric.com/ar/products-solutions/factory-automation/index.html>. Acesso em: 28 jan. 2022.

WORLD HEALTH ORGANIZATION. Injuries and violence (Issue March 2021). 2021a. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/injuries-and-violence>. Acesso em: 28 jan. 2022.

ZHANG, Cha; MA, Yunqian. Ensemble machine learning: methods and applications. Springer Science & Business Media, 2012.

ZHANG, Hanrui, et al. "Deep learning identifies digital biomarkers for self-reported Parkinson's disease." *Patterns* 1.3, p. 100042, 2020. doi: 10.1016/j.patter.2020.100042.

ZHENG, Zhenyu, et al. "An automatic diagnosis of arrhythmias using a combination of CNN and LSTM technology." *Electron.*, vol. 9, no. 1, pp. 1–15, 2020, doi: 10.3390/electronics9010121.

APÊNDICE

Artigo submetido à 44TH ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE & BIOLOGY SOCIETY (EMBC).

Predicting early TBI mortality with convolutional neural networks and conventional machine learning techniques

K. A. A. Guimarães, M. G. F. Costa, Member, IEEE, R. L. Amorim and C. F. F. Costa Filho, Member, IEEE

Abstract— In this work, we compare the performance of a convolutional network with other conventional machine learning techniques, such as a multilayer perceptron neural network, decision tree, random forest, logistic regression and support vector machine, for predicting 14-day mortality in patients with TBI, using a database obtained in a low-and middle-income country, with 529 records and 16 predictor variables. The missing values of predictor variables were filled in with techniques such as decision tree, random forest, k-nearest-neighbor and linear regression. Two methods for filling in predictor variable were proposed. In the first method, previous filled in variables were used for filling in other variables. In the second one, previous filled in variables were not used for filling in other variables. In the simulation of neural networks, several optimization methods were used, such as RMSProp, Adam, Adamax and SGDM. The best prediction results were obtained with a multilayer perceptron network, with an accuracy of 0.845 and an area under the ROC curve of 0.911.

Clinical Relevance—Our findings support using a multilayer perceptron network model to predict early mortality in patients with TBI, with an area under ROC curve of 0.911.

INTRODUCTION

Currently, there are 4.4 million injury-related deaths worldwide. Of these deaths, 3.16 million are related to unintentional injuries. Overall, 1 in 3 of these deaths are related to traffic accidents. According to the World Health Organization (WHO), deaths resulting from injuries occurring on roads are the 7th cause of death worldwide, totaling 189,000 deaths [1]. Road traffic injuries are now the leading killer of people aged 5 to 29 years [2].

Poverty is associated with an increased risk of injury. Around 90% of injury cases occur in low- and middle-income countries (LMIC), where 85% of the world's population live [3].

The uneven distribution of injuries, with a higher prevalence in poorer countries, is related to a series of risk factors, such as living, working, traveling and going to school in more precarious conditions and poorer access to quality emergency trauma care and rehabilitation services [3].

Millions of people suffer nonfatal injuries each year that lead to emergency and critical care visits, hospitalizations or treatment by general practitioners and often result in temporary or permanent disability and the need for health care and physical and mental rehabilitation [3].

Among the injuries, traumatic brain injury (TBI) accounts for one-third to one-half of all injury deaths and is the leading cause of disability in people under forty [4]. TBI accounts for a major burden regarding morbidity, mortality, disability, socioeconomic losses, reduced life expectancy, and the quality of life, mainly in LMIC [5][6][7].

Physicians normally make therapeutic decisions based on the patient's prognosis. According to Perel et al. [8], 80% of physicians believe that it is important to have a reliable prognosis when making pertinent decisions about the specific use of some therapies such as hyperventilation, barbiturates, or mannitol. The use of prognostics provided by computational tools in patients with TBI favors the use of some interventional therapies in these patients with a good prognosis and reduces their use in those patients with a poor prognosis [9].

In recent years, several studies have been published aiming at the prognosis of patients with TBI, using machine learning techniques such as support vector machines [10][11], radial-basis-function (RBF) [12], multilayer perceptron neural networks [11][12][13], naive-Bayes [14], logistic regression [12][15][16][17][11][12][17][18], generalized Bayesian linear model [14], random forest [11][14], penalized discriminant analysis [14] and gradient boosting machines [11]. To the best of our knowledge, no work has used deep neural networks for the prognosis of patients with TBI

Concerning TBI, the prognostic classes most used are: 1. 30-day mortality [15]; 2. 14-day mortality [14]; 3. mortality, survival (vegetative state or severe disability) and favorable outcome (medium disability or good recovery) [18]; 4. patient survival one year after being discharged [16]; 5. 14-day death and death and disability at six months in patients with traumatic brain injury [17]; 6. KOSCHI results after 6 months of hospital discharge, comprising the following categories: a) death, b) vegetative state, c) moderate disability, d) good recovery [10]; 7. Clinically Relevant TBI - CRTBI: a) neurosurgical procedure, b) intubation > 24 hours as a direct result of head trauma, c) hospitalization greater than 48 hours and evidence of traumatic brain injury through computerized scanning, and d) death due to trauma cranioencephalic [13]; 8. Decision to perform open skull surgery [12].

Table 1 shows a summary of some works published in recent years concerning TBI prognosis. Each work is described in terms of the following variables: data base used, predicting variables, predictor variable and results.

Reference	Data base	Predictor variables	Predicted variable	Results
[10]	828 children under the age of 15 years, admitted to the trauma center of Southeast Thailand, from January 2009 to July 2020.	14 predictor variables: age, gender, mechanism of injury, mean injury severity score, related injuries, comorbidities, ECG, pupillary light reflex, skull fracture, intracranial injuries, basal cistern, midline displacement, surgical treatment.	KOSCHI result after 6 months of hospital discharge, comprising the following categories: a) death, b) vegetative state, c) moderate disability, d) good recovery	Support vector machine: sensitivity = 0.95, specificity = 0.60, accuracy = 0.94 and AUC = 0.78
[14]	517 patient records from the Hospital das Clínicas, University of São Paulo, Brazil. Patients aged 14 years and older were included, with intracranial abnormality on an initial CT scan, exclusion of patients with penetrating trauma, as well as those with an ECG of 15 and without association of intracranial lesions when CT was performed.	16 predictor variables: gender, age, level of pupil reactivity at admission, GCS at the scene where the trauma occurred (prehospital GCS), GCS at admission, the motor component score of the GCS, and presence of hypoxia and hypotension. Also included were midline shift higher than 5 mm, brain herniation detected on CT (defined as effacement of the third ventricle or the basal cisterns), subarachnoid hemorrhage, epidural hemorrhage, subdural hemorrhage, intracerebral hemorrhage, trauma severity, prothrombin time, and partial thromboplastin time.	Mortality over 14 days	Naive Bayes: AUC= 0.906; generalized Bayesian linear model: AUC = 0.881
[11]	117 records of patients admitted to the Sichuan Provincial People Hospital from December 2009 to November 2011.	40 variables were used in this study, among which the following stand out: age, length of stay in the hospital, length of stay in the intensive care unit, among others.	Survival of the patient who had severe traumatic brain injury.	Logistic regression: AUC = 0.83
[15]	472 adult patients, aged 16 years or older, admitted within 24 hours of the trauma.	13 predictor variables: Age, admission ECG, motor score, pupil reactivity, hypoxia, hypotension, Marshall computed tomography (CT) score, CT tSAH, CT epidural mass, glucose, Hb, length of stay in the intensive care unit, and neurosurgical procedures	Mortality over 30 days	Logistic regression: AUC between 0.81 and 0.84
[13]	PECARN (Pediatric Emergency Care Applied Research Network) database. 902 patients were included in the study.	23 predictor variables: Cerebellar hemorrhage, cerebral contusion, cerebral edema, cerebral hemorrhage or intracerebral hematoma, skull diastasis, epidural hematoma, extra-axial hematoma, intraventricular hemorrhage, midline displacement of brain structures, pneumocephaly, skull fracture, subarachnoid hemorrhage, subdural hematoma, traumatic infarction, diffuse axonal injury, herniation and shear injury, mechanism of injury (eg vehicle collision), severity of injury mechanism, loss of consciousness, ECG at presentation, age and gender	Clinically Relevant TBI - CRTBI: a) neurosurgical procedure, b) intubation > 24 hours as a direct result of head trauma, c) hospitalization greater than 48 hours and evidence of traumatic brain injury through computerized scanning, and d) death due to trauma cranioencephalic	Multilayer perceptron: accuracy = 0.98; precision = 0.98; sensitivity = 0.99; specificity = 0.99; AUC = 0.99.
[18]	MRC CRASH database, containing 10,008 adult patients with ECG 14, who entered the hospital within 8 hours of the moment of trauma	8 predictor variables: Gender, age, cause of trauma, time from trauma to randomization, ECG at randomization, pupil reactivity, CT scan results, whether the patient had major extracranial trauma, country income level (developed or developing countries).	Mortality, survival (vegetative state or severe disability), favorable outcome (medium disability or good recovery).	Logistic regression: C statistic above 0.80
[17]	IMPACT database, with 8059 patients, including patients with moderate and severe traumatic brain injury, ECG less than or equal to 12 and complete ECG after 6 months.	8 predictor variables: Age, ECG motor score, pupil reaction, hypoxia, hypotension, computed tomography (CT) features, glucose and hemoglobin information.	death at 14 days and death and disability at six months in patients with traumatic brain injury	logistic regression: AUC > 0.80.
[12]	Database of 12,640 patient records from 116 teaching hospitals in Taiwan.	10 predictor variables: gender, age of the group, helmet use, time of loss of consciousness, time of amnesia, presence of amnesia, episode of convulsion, presence of neurological deficit, presence of complications, presence of skull fractures.	Decision to perform open skull surgery	logistic regression: AUC = 0.761; multi layer perceptron: AUC = 0.897; radial base function network: AUC = 0.880.
[16]	372 patients admitted to the Royal Infirmary regional trauma center in Edinburgh. Inclusion criteria: age 14 years and older who, at admission, had a final ECG measurement of less than 12 or between 13 and 15, with concomitant injuries, and had a trauma severity score greater than 16.	8 predictor variables: Age, gender, ECG, cause of trauma, trauma severity score, pupil response, brain scan result, and previous alcohol consumption.	Patient survival one year after discharge	logistic regression: AUC = 0.901

Amorim et al. [14], used a database with 517 patients from the Hospital das Clínicas, University of São Paulo, Brazil. For the constitution of the database, some criteria were used: inclusion of patients aged from 14 years, with intracranial abnormality in an initial computed tomography and exclusion of patients with penetrating trauma, as well as those with a Glasgow Coma scale of 15 and without association of intracranial lesions shown by tomography. The following predictor variables were selected: gender, age, level of pupil reactivity at admission, GCS at the scene where the trauma occurred (prehospital GCS), GCS at admission, the motor component score of the GCS (Glasgow motor score), and presence of hypoxia and hypotension. Also included were midline shifts greater than 5 mm, brain herniation detected on CT (defined as effacement of the third ventricle or the basal cisterns), subarachnoid hemorrhage, epidural hemorrhage, subdural hemorrhage, intracerebral hemorrhage, trauma severity, prothrombin time, and partial thromboplastin time. The authors used the following predictors: naive-Bayes, generalized Bayesian linear model, penalized discriminant analysis and random forest. The evaluation of predictors was performed using metrics such as area under the ROC curve, sensitivity and specificity. The best results achieved by the authors for predicting 14-day mortality were obtained using the naive Bayes technique, with an area under the ROC curve, $AUC = 0.906$, followed by the generalized Bayesian linear model, with $AUC = 0.881$.

Amorim et al. [14] carried out their study in an LMIC. According to Perel et al. [18] models drawn up with samples from developed countries performed worse when applied in LMICs, likely due to the mismatch in case mixes as well as overall health care infrastructure. In addition, a major challenge in LMICs is the quality of data recording, done locally, with a high incidence of variables with missing values.

To the best of our knowledge, no work has used a convolutional neural network (CNN) for the prognosis of patients with TBI. CNNs are normally 2D networks used for image processing, mainly for image classification and object segmentation [19][20]. Nevertheless, 1D CNNs have also been used successfully in many medical applications such as detection of Parkinson disease [21][22] and detection of arrhythmias [23][24].

In this work, we propose the use of 1D CNN to predict 14-day mortality in patients with TBI. This work, using the same database and the same predictor variables used by Amorim et al. [14], presents the following contributions:

1. Propose and evaluate CNN to predict 14-day mortality in patients with TBI.

2. Compare the performance of a CNN with conventional machine learning techniques, such as multilayer perceptron neural network (MLP), support vector machines (SVM), decision tree and random forest.
3. Propose and evaluate two methods for filling in values of missing variables.

MATERIALS AND METHODS

A. Materials

The database used in this work was provided by Amorim et al. [14]. This database was generated from data collected from patients referred to Hospital das Clínicas (São Paulo, Brazil). Data collection was carried out from March 2012 to January 2015. The final follow-up of patients ended in June 2015. This database contains 529 records with 71 variables. Each record corresponds to data collected from a patient with TBI.

The patients admitted to the study were those referred by the neurosurgical team needing to be admitted to the intensive care unit (ICU). For this study, the following selection criteria were performed: patients under 14 years of age, patients with penetrating head trauma, patients with an ECG equal to 15 and patients not presenting intracranial abnormalities on computerized scanning were excluded.

From the original database, using the exclusion criteria mentioned above, a new database was generated, with 470 records. Regarding the information of each patient, only 17 columns of data were selected containing the variables used in this study. The 17 variables (16 predictors variables and 1 predicted variable), cited previously in the introduction, correspond to the variables used in the work by Amorim et al. [14]. With this, it will be possible to compare the results obtained in this work with the results previously obtained by Amorim et al. [14].

The proportions of the initial set used for training and testing the machine learning models for prediction of mortality in 14 days were 80% and 20%, respectively. Thus, the number of observations used in each of the sets were as follows: training set – 376 observations; test set – 94 observations.

B. Methods

Figure 1 shows a block diagram of the adopted methodology. The following presents each block shown in Fig. 1.

C. Filling in missing values

Table II shows the number of patients with missing values for each variable.

To fill in these missing variables, the variable *death up to 14 days* was removed, which represents the prognostic class, so that a bias is not created in relation

to this variable. For the remaining 16 variables, different methods were evaluated in accordance with the data type. For the categorical variables, the methods decision tree, random forest, and k-nearest-neighbor were evaluated. For numerical variables, the methods decision tree, random forest, and linear regression models were evaluated. The variables age and trauma severity, as they present a rate of absent variables lower than 3%, had the missing values filled in by the mean value.

To fill in the missing values of a variable a , the training and testing sets are defined as follows: Suppose there are X records with missing values of variable a and Y ($470-X$) records with no missing values. The training set is defined with 80% of the records of Y , while the testing set, with 20% of the remaining records of Y .

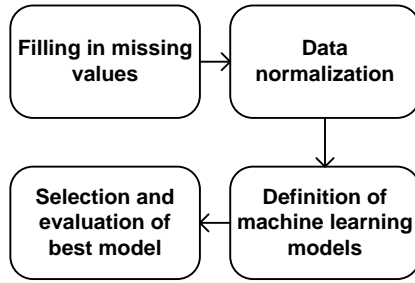


Figure 1. Block diagram of the adopted methodology.

NUMBER OF PATIENTS WITH MISSING VALUES			
Predictor	Variable type	Number of records with missing values	% missing values
Hypoxia	Categorical	254	54%
Prehospital GCS	Categorical	127	27%
hypotension at admission	Categorical	77	16%
GCS at admission	Categorical	47	10%
Partial thromboplastin time	Numeric	39	8%
Glasgow motor score	Categorical	37	8%
Prothrombin time	Numeric	35	7%
level of pupil reactivity at admission	Categorical	33	7%
Age	Numeric	11	2%
Trauma severity	Numeric	3	1%
Gender	Categorical	0	0%
Midline brain shift > 5 mm	Categorical	0	0%
Brain herniation detected on CT	Categorical	0	0%
Subarachnoid hemorrhage	Categorical	0	0%
Subdural hemorrhage	Categorical	0	0%
Intracerebral hemorrhage	Categorical	0	0%
Death up to 14 days	Categorical	0	0%

D. Data normalization

When analyzing the data, it was observed that the numerical variables *age* and *trauma severity* contain values in a very different range from the other variables. The values of the variable *age* vary between 16 and 99, while the values of the variable *trauma severity* vary between 8 and 101. The other numerical variables, *prothrombin time* and *partial thromboplastin time*, have a dynamic range between 0 and 10, which is in a range of values close to the categorical variables. To normalize the variables *age* and *severity of trauma*, three methods tailored to datasets characterized by outliers and dominant features were evaluated: min-max, cubic root and z score [25]. The min-max normalization technique is described in (1). The cubic technique, is shown in (2) and the cubic root technique, in (3).

$$x'_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \cdot (nMax - nMin) + nMin \quad (1)$$

$$x'_i = \sqrt[3]{x_i \cdot \sigma} \quad (2)$$

$$x'_i = \frac{x_i - \mu}{\sigma} \quad (3)$$

Where:

x'_i - i -th normalized value

x_i - i -th non normalized value

$\min(x_i)$ - minimum value of variable x_i

$\max(x_i)$ - maximum value of variable x_i

$nMax$ - maximum desired value for x_i ($= 1$)

$nMin$ - minimum desired value for x_i ($= -1$)

σ - standard deviation of variable x_i

μ - mean value of variable x_i

Two procedures, shown below, were evaluated for filling in missing values.

Procedure 1:

9. Sort the variables in descending order, according to the number of missing values.: $\{v_1, v_2, v_3, \dots, v_n\}$, where variable v_i has more missing values than variable v_{i+1} .
10. To predict missing values of v_1 choose a prediction method and use all other variables. In the prediction, for $i > 1$, fill in all missing values of variables v_i with a measure of the central tendency, the mode if the variable is categorical, or the mean otherwise.
11. To predict missing values of $v_i, i > 1$, choose a prediction method and use

all other variables. In the prediction, for $j < i$, use variable v_j with predicted values, and, for $k > i$, fill in all missing values of variables v_k with a measure of central tendency, the mode if the variable is categorical, or the mean otherwise.

Example 1: Consider the following variables: v_1, v_2, v_3, v_4 and v_5 , with 4, 3, 2, 1 and no missing values, respectively. To predict missing values of variable v_2 use all other variables. Missing values of v_1 are filled in with predicted values; Missing values of variables v_3 and v_4 are filled in with a measure of central tendency and use variable v_5 .

Procedure 2:

1. Sort the variables in descending order, according to the number of missing values.: $\{v_1, v_2, v_3, \dots, v_n\}$, where variable v_i has more missing values than variable v_{i+1} .
2. To predict missing values of v_1 choose a prediction method and use all other variables. In the prediction, for $i > 1$, fill in all missing values of variables v_i with a measure of central tendency, the mode if the variable is categorical, or the mean otherwise.
3. To predict missing values of v_i , $i > 1$, choose a prediction method and use only the variables v_j , where $j > i$. In prediction, for $j > i$, fill in all missing values of variables v_j with a measure of central tendency, the mode if the variable is categorical, or the mean otherwise.

Example 2: Consider the following variables: v_1, v_2, v_3, v_4 and v_5 , with 4, 3, 2, 1 and no missing values, respectively. To predict missing values of variable v_2 use the following variables: v_3 and v_4 with missing values filled in with a measure of central tendency and variable v_5 .

D. Definition of machine learning models

Two groups of machine learning models are used in this work. The first group is used to fill in missing values of variables. The second group is used to predict 14-day mortality in patients with TBI. Table III shows these models and their use.

MACHINE LEARNING MODELS FOR FILLING IN MISSING VARIABLES AND FOR PREDICTION OF 14-DAY MORTALITY

Machine learning model	Application	
	Fill in missing values	14-day mortality prediction
Decision tree	x	x
Random Forest	x	x
K-nearest neighbor	x	
Support vector machines		x
Logistic regression		x
Linear regression	x	
MLP		x
CNN		x

When the decision tree and random forest models were used for the mortality prediction (classification task), the error function was based on the Gini index. For filling in missing values of variables (regression task), the mean-square-error (MSE) criterion was used. The minimum number of leaves per node was fixed in one.

When the k-nearest neighbors' model was used for filling in missing values, 2 nearest neighbors were used.

When the support vector machines model was used for 14-day mortality prediction, a linear kernel was employed and parameter C was fixed in 0.2.

When the logistic regression model was used for 14-day mortality prediction, the training stop criterion was fixed in 4000 epochs and the L_2 regularization method was used.

When the MLP was used for 14-day mortality prediction, the 4 architectures shown in Figure 2 were evaluated.

The first MLP architecture (MLP1), shown in Figure 2(a), has an input layer and three hidden layers with 8 neurons each and a ReLU activation function. At the end, there is a classification layer with a sigmoid activation function. The second MLP architecture (MLP2) differs from the MLP1 architecture by introducing a dropout layer before the last hidden layer. The dropout factor used is 0.2. The introduction of the dropout layer aimed to improve the generalization of the network. The third MLP architecture (MLP3) differs from the MLP2 architecture by introducing a dropout layer after the first hidden layer. The dropout factor used is 0.2. The fourth MLP architecture (MLP4) differs from the MLP3 architecture in two ways: 1) by introducing a second hidden layer before the first dropout layer; 2) by fixing the number of neurons in hidden layers in 16.

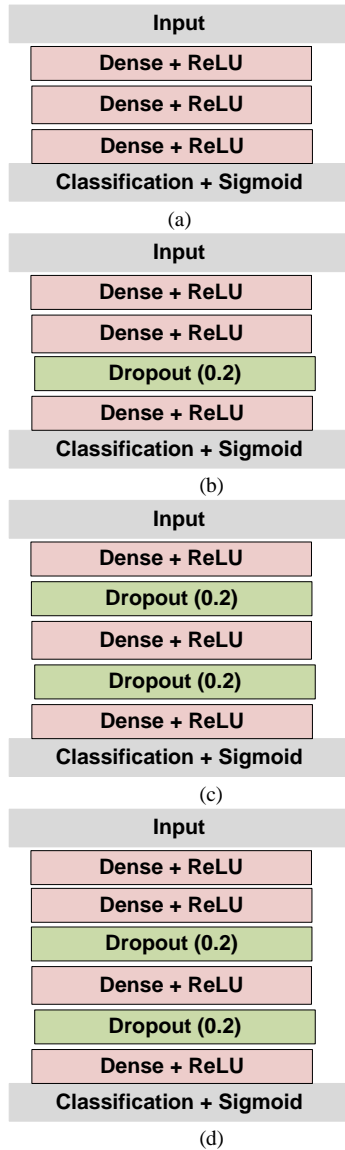


Fig. 2 Multilayer perceptron architectures proposed for predicting 14-day mortality. (a) architecture MLP1; (b) architecture MLP2; (c) architecture MLP3 and (d) architecture MLP3

When the CNN was used for 14-day mortality prediction, the architecture shown in Figure 3 was used. The CNN network has a parallel architecture with three blocks. Each block is formed by two Conv1D layers: one of them with 64-kernels size 3x1, another one with kernel size of 4x1 and the third block with kernels size of 5x1, followed by a flatten layer. All the hidden layers of the CNN use the ReLU activation function. Also, a dropout layer was used with a dropout factor of 0.2.

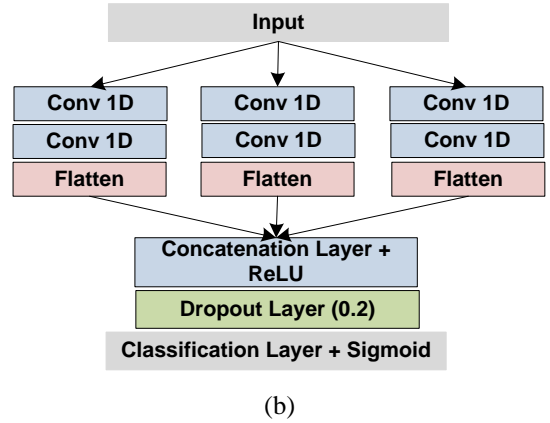


Figure 3. CNN architectures proposed for the prediction of 14-day mortality in patients with TBI.

Both neural networks, MLP and CNN, have the sigmoid function at output layer. The output of 1 corresponds to death >14 days and the output 0 corresponds to death <14 days.

For training the networks, the following parameters were evaluated: learning rate: 0.001, 0.003, 0.007 and 0.1; optimization algorithm: RMSProp, Adamax, Adam and SGDM; generalization improvement: dropout; number of training epochs: 4000.

The dataset used for training is not balanced. Fig. 3 shows the percentage of patients in each class.



Figure 3. Percentage of patients in each training class.

To correct the imbalance during training, the learning rate was multiplied by a weight inversely proportional to the frequency of occurrence of the class. Thus, when an observation belonging to a class i is presented at the input of the network, the learning rate used to update the weights and polarizations is multiplied by a multiplicative factor f_i given by (2).

$$f_i = \frac{n}{(n_{classes} * n_{class\ i})} \quad (2)$$

Where:

n – number of patients

$n_{classes}$ – number of classes

$n_{class\ i}$ – number of patients on class i

The models were implemented in the PyCharm environment (JETBRAINS, 2021), in the university student version, using a computer with the following specifications: Intel® Core™ i7-7500U CPU @ 2.70GHz × 4, 16GB RAM.

The libraries used to help development were tensor flow in version 2.2.1, *Keras* in version 2.4.3, *scikit-learn* in version 0.23.2. The programming language chosen was *Python* 3.6. The *panda* library in version 1.0.3 was used to work with dataset tables. The *matplotlib* library in version 3.2.2 was used for visualization.

RESULTS

Three types of results are shown. The first one concerns filling in missing values of variables. The second one concerns the normalization process. The third one concerns the prediction of mortality over 14 days.

A. Results of filling in missing values of variables

Table IV shows the accuracy obtained by filling in missing values of categorical variables, with the two procedures, 1 and 2. The bold values show the best accuracies obtained for each variable.

ACCURACY OBTAINED IN FILL IN MISSING VALUES OF CATEGORICAL VARIABLES

Variable	Procedure 1			Procedure 2		
	Model			Model		
	Decision tree	Random Forest	KNN	Decision tree	Random Forest	KNN
Hypoxia	0.737	0.809	0.775	0.737	0.805	0.775
Prehospital GCS	0.508	0.432	0.243	0.503	0.469	0.246
Hypotension at admission	0.798	0.871	0.821	0.766	0.841	0.831
GCS at admission	0.454	0.447	0.334	0.365	0.416	0.301
Glasgow motor score	0.901	0.798	0.524	0.329	0.386	0.310
Pupil reactivity at admission	0.681	0.777	0.595	0.674	0.774	0.581

For all categorical variables, the best accuracy values were obtained with procedure 1. For the categorical variables *hypoxia*, *hypotension at admission* and *level of pupil reactivity at admission*, the best results for filling in missing values were obtained with the random forest method (accuracy = 0.809, 0.871, and 0.777, respectively). For the categorical variables *prehospital GCS*, *GCS at admission* and *Glasgow motor score*, using the accuracy metric, the best results for filling in the missing values were obtained with the decision tree

method (accuracy = 0.508, 0.454 and 0.901, respectively).

Table V shows the MSE obtained when filling in the missing values of numeric variables, with the two procedures. The bold values show the best accuracies obtained for each variable.

For numeric variable *prothrombin time*, the best MSE value, 0.112, was obtained with procedure 1 and with linear regression method. For numeric variable *partial prothrombin time*, the best MSE value, 0.114, was obtained with procedure 1 and with linear regression.

MSE VALUES OBTAINED WHEN FILLING IN THE MISSING VALUES OF A NUMERIC VARIABLE

Variables	Procedure 1			Procedure 2		
	Models			Models		
	Decision tree	Random forest	Linear regression	Decision tree	Random forest	Linear regression
Partial prothrombin time	0.261	0.154	0.118	0.261	0.154	0.114
Prothrombin time	0.261	0.166	0.112	0.274	0.226	0.181

B. Results of the normalization process

To evaluate the performance of the three normalization techniques, min-max, cubic root and z score, when applied to the variables *age* and *trauma severity*, we use the MLP1 network to predict the mortality in 14-days. Table VI shows the performance of the MLP1 network when using all the variables, with age and trauma severity normalized by one of the normalization methods. As observed in Table VI, the best performance of the MLP1 model was obtained with the min-max normalization technique. Therefore, from now on, in the results to be presented later, this data normalization technique is used to normalize the variables age and severity of trauma.

AREA UNDER ROC CURVE FOR MLP1 MODEL, WHEN USING THE THREE NORMALIZATION METHODS

Model	Normalization technique		
	Cubic Root	Z-score	Min-Max
MLP1	0.8694	0.8669	0.8739

C. Results of 14-day mortality prediction

Table VII shows values of accuracy and area under the ROC curve (AUC) of models' logistic regression, decision tree, random forest and support vector machines, for prediction of mortality over 14 days. As shown in Table VII, the best result for AUC,

0.811, is obtained with SVM. The best value for accuracy, 0.831, is obtained with SVM and logistic regression.

The tables below show the results of the neural networks, both MLP and CNN, exploring different hyperparameters, like optimization method, learning rate and momentum.

Tables VIII, IX, X, XI and XII show the values of AUC, F1-score and accuracy of MLP1, MLP2, MLP3, MLP4 and CNN, for predicting mortality over 14 days.

AUC AND ACCURACY OF MODELS LOGISTIC REGRESSION, DECISION TREE, RANDOM FOREST AND SUPPORT VECTOR MACHINES, FOR PREDICTING MORTALITY OVER 14 DAYS

Model	AUC	Accuracy
Logistic regression	0.784	0.831
Support vector machines	0.801	0.831
Decision tree	0.677	0.746
Random forest	0.622	0.789

AUC, F1-SCORE AND ACCURACY) OF MODEL MLP1 FOR PREDICTING MORTALITY OVER 14 DAYS

Hyperparameters	AUC	F1-score	Accuracy
RMSProp	0.876	0.667	0.761
Adam	0.874	0.667	0.775
SGDM (learning rate = 0.1, momentum = 0.01)	0.878	0.667	0.817
SGDM (learning rate = 0.3, momentum = 0.001)	0.810	0.493	0.507
SGDM (learning rate = 0.007, momentum = 0.001)	0.869	0.714	0.831

AUC, F1-SCORE AND ACCURACY) OF MODEL MLP2 FOR PREDICTING MORTALITY OVER 14 DAYS

Hyperparameters	AUC	F1-score	Accuracy
RMSProp	0.872	0.668	0.761
Adam	0.871	0.68	0.775
SGDM (learning rate = 0.1, momentum = 0.01)	0.878	0.586	0.662

AUC, F1-SCORE AND ACCURACY) OF MODEL MLP3 FOR PREDICTING MORTALITY OVER 14 DAYS

Hyperparameters	AUC	F1-score	Accuracy
RMSProp	0.887	0.458	0.366
Adam	0.876	0.641	0.732
SGDM (learning rate = 0.1, momentum = 0.01)	0.875	0.638	0.761

AUC, F1-SCORE AND ACCURACY) OF MODEL MLP4 FOR PREDICTING MORTALITY OVER 14 DAYS

Hyperparameters	AUC	F1-score	Accuracy
RMSProp	0.862	0.596	0.732
Adam	0.862	0.651	0.789
SGDM (learning rate = 0.1, momentum = 0.01)	0.909	0.686	0.845
SGDM (learning rate = 0.3, momentum = 0.001)	0.883	0.581	0.634
SGDM (learning rate = 0.007, momentum = 0.001)	0.911	0.630	0.718

AUC, F1-SCORE AND ACCURACY) OF MODEL CNN FOR PREDICTING MORTALITY OVER 14 DAYS

Hyperparameters	AUC	F1-score	Accuracy
RMSProp	0.844	0.615	0.718
Adam	0.840	0.571	0.662
SGDM (learning rate = 0.1, momentum = 0.01)	0.878	0.652	0.775
SGDM (learning rate = 0.3, momentum = 0.001)	0.660	0.560	0.732
SGDM (learning rate = 0.007, momentum = 0.001)	0.5	0.268	0.422

DISCUSSION

Two procedures were proposed for filling in the variables missing values, which have the variables ordered in descending order, based on the number of missing values. In Procedure 1, as values are predicted, variables with full values are used in predicting the missing values of the next variable; in Procedure 2, after predicting the missing values of a predictor variable, they are no longer used to predict the following variables.

For the categorical variables' hypoxia, prehospital GCS, hypotension at admission, GCS at admission, Glasgow motor score and pupil reactivity at admission, procedure 1, as shown in Table IV, presented the best performance. For the numeric variable partial prothrombin time and prothrombin time the best performances, as shown in Table V, were obtained with procedures II and I, respectively.

To predict the missing variables, some supervised models were used, namely: decision tree, random forest, k-nearest neighbors and linear regression. In general, for the categorical variables, as shown in Table IV, the random forest and decision tree models presented the best performances, while, for the numerical variables, as shown in Table V, the best performances were obtained with the linear regression model.

Still in the preprocessing stage, normalization techniques were applied to the numerical variables *age and severity of trauma*, so that the dynamic range of values of these variables are similar. Evaluating the performance of the MLP1 model with several normalization techniques, we found, as shown in Table VI, that the min-max technique excelled over the other techniques.

Regarding the classification of mortality over 14 days of patients with TBI, we used five architectures of neural networks, being 4 architectures of MLP networks and 1 CNN architecture. In addition to these architectures, conventional machine learning methods such as random forest, decision tree, support vector machine and logistic regression models, were also used.

As shown in Table VII, when comparing the results obtained with models' random forest, decision tree, support vector machines and logistic regression, the best AUC and accuracy, 0.801 and 0.831, respectively, were obtained with the SVM method.

Concerning MLP networks, as shown in Tables VIII to XII, the best results for AUC, and accuracy, 0.911 and 0.845, respectively, were obtained with MLP4, the most complex MLP architecture, while the best value of F1-score, 0.714, was obtained with the MLP1 network, the simplest MLP architecture.

The proposed CNN architecture, as shown in Table XII, presented an AUC, F1-score and accuracy of 0.878, 0.652 and 0.775, respectively. This performance is slightly inferior to the one obtained with the MLP4 network. The AUC value, nevertheless, is higher than the AUC values obtained with the conventional machine learning techniques, random forest, decision tree, support vector machines and logistic regression.

Considering data from the literature presented in Table I, the number of predictor variables varies between 8 and 40. Amorim et al. [14] and Tunthanathip and Oearsaku [10] used 14 variables, the same number of variables used in this work. It is noteworthy, however, that Amorim et al. [14] used the same database and the same predictor variables, while Tunthanathip and Oearsaku [10] used another database and other predictor variables.

Amorim et al. [14] obtained an area under the ROC curve of 0.906, while Tunthanathip and Oearsaku [10] obtained an area under the ROC curve of 0.78. In the work presented here, an area under the ROC curve of 0.911 was obtained with the MLP4 model.

Finally, we note that the work presented here was the first study reported in the literature to evaluate the performance of deep networks, more specifically, that of convolutional networks in the prognosis of TBI.

Although the results obtained point to a better performance by shallow neural networks (MLPs)

compared to deep neural networks (CNN), we believe that a deeper study, evaluating other architectures of convolutional networks, is necessary, for a definitive conclusion on the subject. Another point to be highlighted is that the small database used does not favor the use of deep networks, which need more data to adjust the parameters, avoiding overfitting.

CONCLUSION

The main contributions of this work were the following: 1) the evaluation of deep neural networks, more specifically, 1D convolutional neural networks, to perform the task of predicting 14-day mortality in patients who had TBI.; 2) the evaluation of two procedures for filling in missing values, showing the superiority of procedure 1, which uses predictor variables previously filled in to fill in the others. This contribution is of paramount importance, especially in low- and middle-income countries, since the quality of medical data leaves something to be desired, either due to the urgency of dealing with the patient or the quality of the workforce.

For future studies, we suggest: 1) explore new deep network architectures and work with large and robust databases; 2) use of variable analysis methods to remove variables with redundant information, such as the scalar feature selection method; 3) development of a web platform to use the models publicly.

REFERENCES

- [1] World Health Organization, *World health statistics 2021: monitoring health for the SDGs, sustainable development goals*. 2021.
- [2] World Health Organization, "Global status report on road safety 2018," 2018. [Online]. Available: <https://www.who.int/publications/i/item/9789241565684>.
- [3] World Health Organization, "Injuries and violence," 2021. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/injuries-and-violence>.
- [4] G. Whiteneck, M. Sendroy-Terrill, J. Coll, and C. A. Brooks, "Long-term Outcomes After Traumatic Brain Injury," *J. Head Trauma Rehabil.*, vol. 21, no. 5, p. 435, 2006, doi: 10.1097/00001199-200609000-00060.
- [5] M. S. Umerani, A. Abas, and S. Sharif, "Traumatic brain injuries: Experience from a tertiary care center in Pakistan," *Turk. Neurosurg.*, vol. 24, no. 1, pp. 19–24, 2014, doi: 10.5137/1019-5149.JTN.7080-12.1.
- [6] V. Kamal, D. Agrawal, and R. Pandey, "Epidemiology, clinical characteristics and outcomes of traumatic brain injury: Evidences from integrated level 1 trauma center in India," *J. Neurosci. Rural Pract.*, vol. 7, no. 4, pp. 515–525, 2016, doi: 10.4103/0976-3147.188637.
- [7] K. Hofman, A. Primack, G. Keusch, and S. Hrynkow, "Addressing the growing burden of trauma and injury in low- and middle-income countries," *Am. J. Public Health*, vol. 95, no. 1, pp. 13–17, 2005, doi: 10.2105/AJPH.2004.039354.
- [8] P. Perel, J. Wasserberg, R. R. Ravi, H. Shakur, P. Edwards, and I. Roberts, "Prognosis following head injury: A survey of doctors from developing and

- developed countries,” *J. Eval. Clin. Pract.*, vol. 13, no. 3, pp. 464–465, 2007, doi: 10.1111/j.1365-2753.2006.00713.x.
- [9] L. S. Murray *et al.*, “Does prediction of outcome alter patient management?,” *Lancet*, vol. 341, no. 8859, pp. 1487–1491, 1993, doi: 10.1016/0140-6736(93)90631-P.
- [10] T. Tunthanathip and T. Oearsakul, “Application of machine learning to predict the outcome of pediatric traumatic brain injury,” *Chinese J. Traumatol. - English Ed.*, vol. 24, no. 6, pp. 350–355, 2021, doi: 10.1016/j.cjtee.2021.06.003.
- [11] B. Y. Gravesteijn *et al.*, “Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury,” *J. Clin. Epidemiol.*, vol. 122, pp. 95–107, 2020, doi: 10.1016/j.jclinepi.2020.03.005.
- [12] Y. C. Li, L. Liu, W. T. Chiu, and W. S. Jian, “Neural network modeling for surgical decisions on traumatic brain injury patients,” *Int. J. Med. Inform.*, vol. 57, no. 1, pp. 1–9, 2000, doi: 10.1016/S1386-5056(99)00054-4.
- [13] A. T. Hale, D. P. Stonko, J. Lim, O. D. Guillaumondegui, C. N. Shannon, and M. B. Patel, “Using an artificial neural network to predict traumatic brain injury,” *J. Neurosurg. Pediatr.*, vol. 23, no. 2, pp. 219–226, 2019, doi: 10.3171/2018.8.PEDS18370.
- [14] R. L. Amorim *et al.*, “Prediction of Early TBI Mortality Using a Machine Learning Approach in a LMIC Population,” *Front. Neurol.*, vol. 10, no. January, pp. 1–9, 2020, doi: 10.3389/fneur.2019.01366.
- [15] R. Raj *et al.*, “Machine learning-based dynamic mortality prediction after traumatic brain injury,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–13, 2019, doi: 10.1038/s41598-019-53889-6.
- [16] C. W. P. M. Hukkelhoven *et al.*, “Predicting survival using simple clinical variables: A case study in traumatic brain injury (multiple letters),” *J. Neurol. Neurosurg. Psychiatry*, vol. 68, no. 3, pp. 396–397, 2000, doi: 10.1136/jnnp.68.3.396.
- [17] E. W. Steyerberg *et al.*, “Predicting outcome after traumatic brain injury: Development and international validation of prognostic scores based on admission characteristics,” *PLoS Med.*, vol. 5, no. 8, pp. 1251–1261, 2008, doi: 10.1371/journal.pmed.0050165.
- [18] P. A. Perel *et al.*, “Predicting outcome after traumatic brain injury: Practical prognostic models based on large cohort of international patients,” *Bmj*, vol. 336, no. 7641, pp. 425–429, 2008, doi: 10.1136/bmj.39461.643438.25.
- [19] C. C. Aggarwal, *Neural Networks and Deep Learning*. Gewerbestrasse: Springer International Publishing, 2018.
- [20] X. Jiang, A. Hadid, Y. Pang, E. Granger, and X. Feng, *Deep Learning in Object Detection and Recognition*. Singapore: Springer Nature Singapore Pte Ltd, 2019.
- [21] J. Prince and M. De Vos, “A Deep Learning Framework for the Remote Detection of Parkinson’s Disease Using Smart-Phone Sensor Data,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2018-July, pp. 3144–3147, 2018, doi: 10.1109/EMBC.2018.8512972.
- [22] H. Zhang, K. Deng, H. Li, R. L. Albin, and Y. Guan, “Deep Learning Identifies Digital Biomarkers for Self-Reported Parkinson’s Disease,” *Patterns*, vol. 1, no. 3, p. 100042, 2020, doi: 10.1016/j.patter.2020.100042.
- [23] Z. Zheng, Z. Chen, F. Hu, J. Zhu, Q. Tang, and Y. Liang, “An automatic diagnosis of arrhythmias using a combination of CNN and LSTM technology,” *Electron.*, vol. 9, no. 1, pp. 1–15, 2020, doi: 10.3390/electronics9010121.
- [24] P. Pławiak, “Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals,” *Swarm Evol. Comput.*, vol. 39, no. April, pp. 192–208, 2018, doi: 10.1016/j.swevo.2017.10.002.
- [25] M. Mulenga *et al.*, “Feature Extension of Gut Microbiome Data for Deep Neural Network-Based Colorectal Cancer Classification,” *IEEE Access*, vol. 9, pp. 23565–23578, 2021, doi: 10.1109/ACCESS.2021.3050838.