Aplicação de Mineração de Dados para Predição de Mortalidade em UTI: balanceamento, dados ausentes e classificadores

Jorge S.A.M. Barreto¹, Angelo C. Loula¹

¹Programa de Pós-Graduação em Computação Aplicada (PGCA)
Universidade Estadual de Feira de Santana (UEFS)
Feira de Santana – BA – Brazil

Abstract. *Severity scores provide a consolidated index of the patient's health status in the ICU. These scores are based on linear models and analysis of each single variable in isolation. Data mining techniques have been applied to generate more complex prediction models for prediction, but did not deepen in the analysis of class imbalance and treatment of missing data. This work analyzed techniques for class balancing and imputation of missing values, together with Random Forest (RF), Artificial Neural Networks (RNA) and Logistic Regression (RL) classification models. As a result the RF obtained the best performance with the mean AUC of 0.784 ± 0.006 , sensitivity of 0.738 ± 0.002 and specificity of 0.700 ± 0.003 with missing values replaced by default values and trained with the base with NCL undersampling.*

Resumo. *Escores de severidade fornecem um índice consolidado do estado de saúde do paciente na UTI. Estes escores são baseados em modelos lineares e em análises das variáveis isoladamente. Trabalhos prévios aplicaram mineração de dados para indução de modelos mais complexos de predição, mas não se aprofundaram na análise do desbalanceamento de classes e tratamento dos dados ausentes. Este trabalho analisou técnicas de balanceamento e imputação de valores, em conjunto com modelos de classificação de Random Forest (RF), Redes Neurais Artificiais (RNA) e Regressão Logística (RL). Como resultado a RF obteve o melhor desempenho, com a AUC média de 0.784 ± 0.006 , sensibilidade de 0.738 ± 0.002 e especificidade de 0.700 ± 0.003 com valores ausentes substituídos por valores padrões e treinada com a base com sub-amostragem NCL.*

1. Introdução

A Unidade de Terapia Intensiva (UTI) é um ambiente hospitalar em que recursos médicos são utilizados para tratamento de pacientes de alto risco de mortalidade. Escores preditivos de severidade são um importante índice de avaliação da condição de gravidade destes pacientes para predição do resultado do tratamento, tipicamente predição de mortalidade. Esses escores, utilizados em hospitais em todo o mundo, auxiliam no direcionamento de recursos, na avaliação de qualidade das unidades e na estratificação de casos clínicos.

Os escores de severidade são compostos pelos atributos fisiológicos dos pacientes, resultados de exames laboratoriais e informações sociais e demográficas, transformados

em um único valor que pode ser interpretado pela equipe médica para determinar a severidade do estado de saúde do paciente e por em prática ações correspondentes. Um dos escores de severidade de pacientes em UTI mais conhecidos e mais utilizados é o APACHE - *Acute Physiology and Chronic Health disease Classification System* [Knaus et al. 1991]. O APACHE foi proposto em sua primeira versão em 1981 e utilizou 34 variáveis para cálculo do escore [Knaus et al. 1981], utilizando dados de 583 admissões de um hospital. Em sua terceira revisão, APACHE-III, foram utilizados 20 variáveis e dados de 17.000 admissões em 40 UTIs [Knaus et al. 1991].

De forma geral, seja para o APACHE ou outros escores de severidade, a análise das variáveis é feita para cada atributo isolado e discretizado com posterior atribuição de pesos e composição em um modelo linear, gerando o valor final do escore e aplicando regressão logística somente para este valor. Apesar da facilidade de análise e aplicação, este modelo apresenta limitações por não considerar relações não-lineares ou relações multidimensionais para predição de severidade.

Técnicas de mineração de dados foram aplicadas em trabalhos anteriores buscando induzir modelos mais complexos para predição de mortalidade em UTI a partir de dados disponíveis em grandes bases. As metodologias utilizadas foram diversas, demonstrando uma diversidade de modelos e técnicas. Embora a preparação e pré-processamento dos dados tenham grande impacto nos resultados obtidos na mineração de dados, poucos estudos abordaram procedimentos para tratamento de balanceamento de classe e substituição de dados ausentes, principalmente combinando as duas soluções. Estes são dois aspectos relevantes em bases de dados de UTI, a quantidade de sobreviventes e não-sobreviventes são desiguais e há grande quantidade de dados ausentes nas variáveis consideradas.

Este trabalho propõe comparar diferentes procedimentos de balanceamento de classes e para tratamento de dados ausentes, em conjunto com utilização de modelos de classificação de *Random Forest* (RF), Redes Neurais Artificiais (RNA) e Regressão Logística (RL), avaliando o desempenho na predição da mortalidade em UTI e comparando com a predição obtida pelo escore APACHE III, buscando estabelecer uma metodologia de mineração de dados que possa ser seguida para esta tarefa de predição.

2. Trabalhos relacionados

Trabalhos anteriores já buscaram aplicar técnicas de mineração de dados para indicação de modelos de predição de mortalidade em UTI. Nesta seção apresentamos estudos representativos que compararam modelos obtidos por mineração de dados com escores de severidade.

Em [Kim et al. 2011], foi desenvolvido um modelo de predição de mortalidade a partir dos dados de 38.474 admissões em UTI de um hospital, com base nas variáveis do APACHE III. Foi comparada a acurácia de várias técnicas de mineração de dados, dentre elas RNA e RL. A metodologia de tratamento de dados ausentes utilizada foi a substituição pela mediana dos valores não ausentes, porém não foi mencionado tratamento para balanceamento de classes. A RNA utilizada foi a *Multi-Layer Perceptron*, com 2 camadas intermediárias. Para avaliação da performance dos modelos de predição utilizou-se a AUC sendo que a RNA obteve o valor de 0.874, enquanto a RL obteve um AUC de 0.871.

Em [Xia et al. 2012] foi criada uma RNA com 2 camadas intermediárias com 15

neurônios cada para predição da mortalidade em UTI. Os dados dos pacientes foram obtidos do *PhysioNET Challenge 2012* [Silva et al. 2012], com 12.000 registros, que foram divididos treinamento, teste e validação com 4000 registros cada. Dados ausentes foram substituídos por valores padrão. Para tratar o desbalanceamento de classes, foi realizada uma super-amostragem de registros de não-sobreviventes, embora não seja especificado o procedimento de super-amostragem. O estudo mostrou como resultado a sensibilidade obtida de 0.5088.

Em [Schmidt et al. 2018], foi utilizado Deep Learning, avaliando o resultado de uma Rede Neural Convolutacional - RNC em comparação com o APACHE II, utilizando dados da base MIMIC III. Não foi mencionada a aplicação de procedimentos de balanceamento dos conjuntos de dados ou tratamento de valores ausentes. A base de dados foi dividida em dois conjuntos de dados: um conjunto com 12.919 registros contendo as variáveis do APACHE II e outro com 11.191 registros contendo as variáveis anteriores mais 11 outras variáveis obtidas de exames laboratoriais. O melhor valor de AUC obtido foi de 0,85 com uma sensibilidade de 0,75 e especificidade de 0,95.

Embora a preparação dos dados pelo tratamento aplicado para desbalanceamento e para dados ausentes afete as etapas posteriores de construção de modelos de dados, trabalhos anteriores dedicaram pouca atenção a esta questão. Neste presente estudo, propomos avaliar e comparar diferentes procedimentos de balanceamento de classes e imputação de dados ausentes, em conjunto com diferentes modelos de classificação, avaliando o desempenho na predição de mortalidade em UTI.

3. Metodologia

Para realização deste estudo, os dados utilizados foram obtidos da base MIMIC-III (*Medical Information Mart for Intensive Care*), uma iniciativa do MIT - Massachusetts Institute of Technology, em parceria com o *Medic Center Beth Israel Deaconess* [Johnson et al. 2016]. A base possui 65.132 registros de entradas de pacientes em UTI. Foram utilizadas 35 variáveis a partir das 20 variáveis do APACHE III (neste score, algumas variáveis são agregações de outras): Batimentos Cardíacos (mínimo e máximo), Pressão sanguínea (mínimo e máximo), Temperatura (mínimo e máximo), Taxa respiratória (mínimo e máximo), Pressão arterial de oxigênio, Tensão arterial alveolar, PH, Pressão de CO₂, Hematócrito (mínimo e máximo), Contagem de glóbulos brancos (mínimo e máximo), Creatinina (mínimo e máximo), Nitrogênio ureico no sangue (mínimo e máximo), Sódio (mínimo e máximo), Albumina (mínimo e máximo), Bilirrubina (mínimo e máximo), Glucose (mínimo e máximo), Ventilação extracorpórea, Volume de urina, Escore Glasgow, mínimo, Escore Glasgow motor, Escore Glasgow verbal, Escore Glasgow visual.

Nesta base de dados, somente 7% dos registros possuíam todos os atributos com dados presentes. Variáveis fisiológicas como batimentos cardíacos, temperatura, taxa respiratória temos poucas ausências, próximo a 1% de dados ausentes. Já exames como tensão arterial alveolar chegam a 26% de ausência. Assim, para tratamento de dados ausentes, foram considerados métodos de imputação de valores ausentes, uma vez que abordagens de remoção de registros ou variáveis com dados ausentes não seria viável.

Para tratamento de dados ausentes, foram aplicada quatro abordagens. Primeiro foi avaliada a imputação pelo valor da mediana do atributo. Depois foi avaliada a

imputação do valor padrão dos atributos, encontrado na literatura médica e documentação do APACHE como valor esperado para um paciente em condições normais, associado ou não a um novo atributo binário de *flag* indicando se o atributo estava ausente. Por fim foi aplicada uma técnica de regressões em sequência para imputação dos dados ausentes através do modelo de predição *Multivariate Imputation by Chained Equations* (MICE)[Azur et al. 2011].

O número de pacientes sobreviventes na base de dados é de 45.318, enquanto o número de não sobreviventes é de 16.214. Dessa forma a classe sobrevivente representa cerca de 64% da base, indicando desbalanceamento e demanda por obtenção de balanceamento para treinamento dos modelos de classificação. Para tratamento do desbalanceamento de classe, primeiro foram avaliadas técnicas de super-amostragem e sub-amostragem aleatória. Em seguida, utilizamos as técnicas de amostragem seletiva, a super-amostragem *Synthetic Minority Oversampling Technique* SMOTE [Chawla et al. 2002] e super-amostragem aleatória, e a sub-amostragem *Neighborhood Cleaning Rule* - NCL [Laurikkala 2001].

Para modelos de predição, foram escolhidos 3 classificadores: Regressão Logística (RL), Random Forest (RF) e Redes Neurais Artificiais (RNA). Cada classificador tem características distintas de linearidade, RL é um modelo linear, RF é um modelo não-linear mas linear por partes, e RNA é não-linear com capacidade de aproximação universal. RL utilizou o otimizador SAGA. Para determinação dos parâmetros dos classificadores RF e RNA foi utilizada uma metodologia gulosa, avaliando um parâmetro por vez e mantendo o valor melhor do parâmetro. A RNA utilizada foi a Multi-Layer Perceptron com otimizador ADAM com 300 neurônios na camada intermediária. A RF foi definida com 100 estimadores e profundidade máxima de 400 nós.

Para avaliação dos modelos foram utilizadas 2 repetições da validação cruzada com 10 sub-conjuntos (2x10 CV). Os métodos de balanceamento foram realizados somente nos sub-conjuntos de treinamento e não nos sub-conjuntos de teste, para aproximar o conjunto de teste da situação real de aplicação. As medidas de avaliação utilizadas nos modelos foram a Área sob a Curva - AUC, Sensibilidade (Sens) e Especificidade (Esp). Para comparação de desempenho dos classificadores foi utilizado o teste não-paramétrico *Wilcoxon signed-ranks* com correção de *Bonferroni* com $\alpha=0.05$.

4. Resultados

4.1. APACHE III

Como base inicial de referência, o escore APACHE III foi avaliado em cada conjunto de teste da base do MIMIC-III. O escore obteve os valores médios de AUC de 0.660 ± 0.0006 , Sensibilidade de 0.531 ± 0.001 e Especificidade de 0.788 ± 0.0001 . Enquanto a Especificidade obtida indica uma razoável capacidade de identificar sobreviventes, a sensibilidade obteve um desempenho deteriorado, indicando baixa identificação dos não-sobreviventes, evidenciando fragilidade do modelo. O modelo mesmo assim demonstra alta estabilidade com baixo desvio-padrão pela alternância dos sub-conjuntos.

4.2. Resultados dos classificadores por estratégia de balanceamento de classe

Para avaliar o impacto do desbalanceamento de classes e a aplicação de técnicas de balanceamento, foi necessário um tratamento inicial de dados ausentes, uma vez que a

imputação de dados é necessária para treinamento dos classificadores. Foi selecionada a técnica substituição de valores ausentes pela mediana, por ser uma das técnicas mais simples e que obtém valores a partir dos próprios dados. Foram avaliadas as estratégias de balanceamento de classe: super-amostragem aleatória, super-amostragem aleatória SMOTE, sub-amostragem aleatória, sub-amostragem NCL, e SMOTE/NCL combinados. Ressalta-se que o balanceamento de classe só é aplicado nos conjuntos de treinamento de validação cruzada.

Foram realizadas variações dos 3 modelos de classificadores propostos, buscando avaliar o comportamento de cada um nestas condições. A tabela 1 mostra os valores médios de AUC, sensibilidade e especificidade encontrados para cada procedimento de balanceamento e para cada classificador.

Tabela 1. Resultados dos classificadores por estratégia de balanceamento de classe com substituição de dados ausentes por mediana dos atributos.

	Sem balanceamento			Super-amostragem SMOTE		
	AUC	Sens	Esp	AUC	Sens	Esp
RL	0.635±0.002	0.006±0.004	0.998±0.001	0.653 ± 0.002	0.725 ± 0.009	0.463 ± 0.008
RF	0.784±0.002	0.310±0.010	0.929±0.003	0.776±0.002	0.545±0.006	0.810±0.005
RNA	0.766±0.004	0.226±0.040	0.951±0.008	0.773±0.001	0.752±0.035	0.649±0.030
	Sub-amostragem aleatória			Super-amostragem aleatória		
	AUC	Sens	Esp	AUC	Sens	Esp
RL	0.622±0.002	0.707±0.010	0.448±0.008	0.656±0.001	0.721±0.007	0.473±0.009
RF	0.781±0.001	0.737±0.004	0.684±0.002	0.782±0.002	0.420±0.004	0.880±0.004
RNA	0.759±0.004	0.678±0.056	0.703±0.054	0.772±0.002	0.754±0.048	0.644±0.052
	Sub-amostragem NCL			SMOTE/NCL		
	AUC	Sens	Esp	AUC	Sens	Esp
RL	0.637±0.002	0.042±0.001	0.985±0.001	0.650±0.005	0.773±0.170	0.364±0.234
RF	0.784±0.006	0.719±0.014	0.703±0.009	0.773±0.002	0.626±0.139	0.743±0.098
RNA	0.762±0.002	0.659±0.057	0.717±0.048	0.771±0.003	0.773±0.070	0.617±0.082

O treinamento dos classificadores para um base sem balanceamento gerou modelos com alta especificidade mas baixa sensibilidade, evidenciado que o classificador gerado obteve forte viés para classe majoritária de sobreviventes. Os procedimentos de balanceamento, por outro lado, foram capazes de melhorar os baixos valores de sensibilidade, porém com comportamento final variado entre sensibilidade e especificidade.

Os valores de AUC não tiveram grande variação para a base desbalanceada e as variações com balanceamento. Isto de certa forma era esperado uma vez que AUC não é uma medida que captura tal condição. Avaliando os resultados de AUC por modelo de classificador, nota-se que em geral o modelo RL tem AUC inferior aos modelos de RF e de RNA. Isto possivelmente está associado a limitação do modelo linear da RL. Já a RF e RNA, ambas não-lineares, obtiveram valores superiores de AUC mas próximos entre si.

Na super-amostragem SMOTE, o maior AUC (0.776±0.002) ($p = 0.004$) foi observado pela RF, mas o melhor equilíbrio entre sensibilidade e especificidade foi obtido pela RNA (Sens=0.752±0.035, Esp=0.649±0.030). Situação análoga foi observada também para a super-amostragem aleatória.

Na sub-amostragem aleatória, novamente temos a RF com maior AUC (0.781±0.001) ($p = 0.004$), mas com esse balanceamento, tanto a RF quanto a RNA obtiveram equilíbrios semelhantes de sensibilidade e especificidade (RF: Sens=0.737±0.00,

Esp=0.684±0.002, RNA: Sens=0.678±0.056, Esp=0.703±0.054). Na sub-amostragem com NCL, observa-se também maior AUC para a RF e equilíbrio entre sensibilidade e especificidade RF e RNA.

Na combinação de super-amostragem SMOTE e sub-amostragem NCL, quanto comparadas com estes procedimentos isolados, não trouxeram resultados diferentes no desempenho dos classificadores. A RF permanece com maior AUC enquanto RF e RNA tem relativo equilíbrio de sensibilidade e especificidade.

Comparando o desempenho dos classificadores em cada procedimento de desbalanceamento, nota-se que a RL não manteve seu desempenho consistente quando aplicado as técnicas de balanceamento de classe, mostrando tendência de desequilíbrio entre a sensibilidade e especificidade conforme valores obtidos nos modelos de balanceamento. Seu maior AUC foi obtido na aplicação a base com balanceamento por super-amostragem aleatória (0.656±0.001) ($p = 0.005$).

A RF teve seu maior AUC na base balanceada por sub-amostragem NCL (0.784±0.006) ($p = 0.005$), assim como obteve valores de sensibilidade e especificidade equilibrados (Sens=0.719±0.014 Esp=0.703±0.009). Já a RNA teve seu maior AUC na base balanceada por super-amostragem SMOTE (0.776±0.002) ($p = 0.005$) e conforme tabela 1 obteve valores equilibrados de sensibilidade e especificidade, mas obteve valores mais altos de desvio padrão.

A RF demonstrou uma consistência maior nos resultados obtidos, com maiores valores médios de AUC ($p = 0.005$) em todos os procedimentos de balanceamento. Com base na RF como referência, vemos que os valores mais equilibrados de sensibilidade e especificidade e maiores AUC foram obtidos pela base de dados balanceada pelo procedimento de NCL.

4.3. Resultados dos classificadores por estratégia de substituição de valores ausentes

Para avaliar as alternativas de imputação de dados ausentes, selecionamos para balanceamento de classe a técnica de NCL, pelo indicativo de melhores resultados para a RF. Além da substituição dos valores ausentes pela mediana de cada variável, foram aplicados os procedimentos de substituição de valores por valores padrões, substituição de valores por valores padrões com inclusão de atributo indicando ausência ou presença do mesmo, e substituição dos valores pelo MICE. Foi realizada novamente a variação dos classificadores em cada condição e obtidos os valores médios de AUC, sensibilidade e especificidade, descritos na tabela 2.

As variações dos procedimentos de imputação de valores, comparado com a imputação pelo valor da mediana, em geral, indicam melhoria dos valores de AUC. A RF obteve seu maior valor de AUC para a imputação por mediana e por valor padrão. Já a RNA e o RL obtiveram valores de AUC maiores na imputação por valor padrão ($p = 0.005$). Mas a RNA teve os maiores desvios padrões de AUC comparando-se aos demais modelos, demonstrando uma instabilidade maior que os outros classificadores.

Embora a RL tenha obtido melhoria de AUC, o desequilíbrio entre sensibilidade e especificidade ainda está presente. Por outro lado, a RF demonstra maior estabilidade em sua capacidade de identificar tanto a classe de sobreviventes e quanto a classe de

Tabela 2. Resultados dos classificadores por estratégia de substituição de valores ausentes com balanceamento NCL

	Mediana			Valor padrão		
	AUC	Sens	Esp	AUC	Sens	Esp
RL	0.637±0.002	0.042±0.001	0.985±0.001	0.723±0.009	0.786±0.013	0.525±0.013
RF	0.784±0.006	0.719±0.014	0.703±0.009	0.784±0.006	0.738±0.002	0.700±0.003
RNA	0.762±0.002	0.659±0.057	0.717±0.048	0.772±0.057	0.680±0.058	0.712±0.043
	Valor padrão + flag			MICE		
	AUC	Sens	Esp	AUC	Sens	Esp
RL	0.717±0.009	0.787±0.021	0.506±0.005	0.707±0.005	0.778±0.010	0.478±0.010
RF	0.776±0.004	0.710±0.017	0.697±0.004	0.781±0.005	0.706±0.006	0.705±0.005
RNA	0.765±0.041	0.658±0.057	0.723±0.045	0.766±0.046	0.692±0.061	0.694±0.055

não-sobreviventes.

Apesar da imputação por MICE possuir um procedimento sofisticado que busca gerar modelos sequências de regressão para aproximar valores ausentes, isso não se traduz em ganhos de AUC, mas sim em valores mais equilibrados de sensibilidade e especificidade.

5. Considerações Finais

Neste artigo foram comparados diferentes procedimentos para balanceamento de classe e tratamento de dados ausentes de modo a avaliar o desempenho da predição de mortalidade em UTI de modelos de classificação como *Random Forest* (RF), Redes Neurais Artificiais (RNA) e Regressão Logística (RL). Em comparação com o escore do APACHE III, os resultados obtidos dos modelos de classificadores treinados com procedimento adequados de pré-processamento apresentaram desempenho superior para predição de mortalidade, com valores superiores de AUC e melhor equilíbrio entre Sensibilidade e Especificidade, particularmente os modelos de *Random Forest* e Redes Neurais Artificiais. O modelo de Regressão Logística, no entanto, demonstra limitações pelo seu modelo linear.

Dentre todos os classificadores, a *Random Forest* obteve valores superiores de AUC ao compararmos com os demais. Em alguns casos, a RF apresentou desequilíbrio de Especificidade e Sensibilidade, enquanto a Rede Neural Artificial era capaz de equilibrar melhor em todos os casos. Com a escolha adequada de procedimentos de balanceamento e imputação, RF e RNA foram superiores ao APACHE III em AUC e no equilíbrio de Sensibilidade e Especificidade.

A imputação de valor padrão em substituição aos valores ausentes teve o AUC mais alto dentre os procedimentos de tratamento de valores ausentes, mantendo também o equilíbrio entre Sensibilidade e Especificidade, dessa forma a aplicação desse procedimento se apresenta como uma escolha para tratamento desta base de dados. A utilização da técnica de balanceamento classe por NCL, manteve o equilíbrio de Sensibilidade e Especificidade da *Random Forest* e Rede Neural Artificial, sendo que o AUC obtido pela *Random Forest* foi significativamente maior ($p = 0.005$).

Enquanto o escore do APACHE III é um modelo rígido, definido *a priori* para um conjunto de dados estudado para sua concepção, modelos como o *Random Forest* e Redes Neurais Artificiais são induzidos automaticamente a partir dos dados. Além disso, a aplicação da metodologia utilizada neste trabalho pode permitir que cada hospital

gere seu próprio modelo de predição de mortalidade em UTI com base em seu próprio conjunto de dados e portanto mais adequado à sua realidade.

Referências

- Azur, M. J., Stuart, E. A., Frangakis, C., and Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research*, 20(1):40–49.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.
- Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L.-w. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., and Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific data*, 3.
- Kim, S., Kim, W., and Park, R. W. (2011). A comparison of intensive care unit mortality prediction models through the use of data mining techniques. *Healthcare informatics research*, 17(4):232–243.
- Knaus, W. A., Wagner, D. P., Draper, E. A., Zimmerman, J. E., Bergner, M., Bastos, P. G., Sirio, C. A., Murphy, D. J., Lotring, T., Damiano, A., et al. (1991). The apache iii prognostic system: risk prediction of hospital mortality for critically iii hospitalized adults. *Chest*, 100(6):1619–1636.
- Knaus, W. A., Zimmerman, J. E., Wagner, D. P., Draper, E. A., and Lawrence, D. E. (1981). Apache-acute physiology and chronic health evaluation: a physiologically based classification system. *Critical care medicine*, 9(8):591–597.
- Laurikkala, J. (2001). Improving identification of difficult small classes by balancing class distribution. In *Conference on Artificial Intelligence in Medicine in Europe*, pages 63–66. Springer.
- Schmidt, D., da Silva, D. B., da Costa, C. A., and da Rosa Righi, R. (2018). Um modelo de predição de mortalidade em unidades de terapia intensiva baseado em deep learning. In *18º Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2018)*, volume 18, Porto Alegre, RS, Brasil. SBC.
- Silva, I., Moody, G., Scott, D. J., Celi, L. A., and Mark, R. G. (2012). Predicting in-hospital mortality of ICU patients: The physionet/computing in cardiology challenge 2012. In *Computing in Cardiology (CinC), 2012*, pages 245–248. IEEE.
- Xia, H., Daley, B. J., Petrie, A., and Zhao, X. (2012). A neural network model for mortality prediction in ICU. In *2012 Computing in Cardiology*, pages 261–264. IEEE.