

Um Modelo para Predição Individualizada da Deterioração da Saúde de Pacientes no âmbito da Internet das Coisas

Naira Kaieski¹, Diogo Schmidt¹, Carolina Rosa Kelsch¹,
Denise Bandeira da Silva¹, Cristiano André da Costa¹, Rodrigo da Rosa Righi¹

¹ Software Innovation Laboratory (SOFTWARELAB)
Programa de Pós-graduação em Computação Aplicada (PPGCA)
Universidade do Vale do Rio dos Sinos (UNISINOS)
São Leopoldo — RS — Brasil

{naira.kaieski, diogoschmidt, carolinarkelsch}@gmail.com
{bandeira, cac, rrrighi}@unisinoss.br

Abstract. *This article presents the development of a model based on internet of health things and artificial intelligence to predict the deterioration of patients' health. We construct the model using the Deep Learning technique of Convolutional Neural Networks to process the patients' data that can be collected by ubiquitous sensors. We used two distinct strategies to select the data to train the model, one based on a traditional population approach, and other, more individualized with data just from patients with a similar age to the monitored individual. The results show an improvement in the accuracy of mortality prediction of the individualized model with an average AUC of 0.73 for all age ranges, compared to AUC of 0.70 from the population approach.*

Resumo. *Este artigo apresenta o desenvolvimento de um modelo baseado em internet das coisas de saúde e inteligência artificial para prever a deterioração do estado de saúde de pacientes. Foi construído um modelo de Deep Learning baseado em Redes Neurais Convolucionais para processar os dados que podem ser coletados por sensores ubíquos. Foram utilizadas duas estratégias diferentes para treinar o modelo, uma baseada na abordagem populacional tradicional e outra mais individualizada onde os dados de treinamento são oriundos de pacientes similares em termos de idade ao indivíduo monitorado. Os resultados mostraram uma sensível melhora na acurácia da predição de mortalidade do modelo individualizado com AUC média de 0,73 para todas as faixas de idade enquanto o modelo populacional apresentou AUC de 0,70.*

1. Introdução

Técnicas de inteligência artificial (IA) têm sido utilizadas com expressivo sucesso e crescente interesse na área da saúde [He et al. 2019], especialmente para diagnosticar doenças, identificar e prever condições agudas dos pacientes. No entanto, desenvolver sistemas de apoio à decisão clínica baseados em IA constitui uma tarefa repleta de desafios, visto que características específicas de cada paciente (idade, gênero, etnia e comorbidades) podem influenciar as respostas fisiológicas às doenças e tratamentos [Royal College of Physicians 2017].

As abordagens médicas modernas visam personalizar diagnósticos e terapias para cada paciente [Barbour 2019] visando fornecer o "tratamento certo para o paciente certo na hora certa"[Mills 2017]. Emerge a chamada medicina de precisão (personalizada ou individualizada) que é fortemente norteadas por dados e tecnologia em áreas que incluem informação médica e engenharia da computação [Kim et al. 2020]. Assim o cuidado com a saúde baseado unicamente na experiência está mudando para a terapia de precisão baseada em *big data* para pacientes individuais [Shen 2020]. Quanto à disponibilidade de dados, as tecnologias relacionadas à Internet das Coisas de Saúde (*Internet of Health Things* - IoHT) [da Costa et al. 2018] possibilitam a coleta remota e constante de dados do paciente de forma ativa ou passiva. E, estima-se que já em 2021 existirão aproximadamente 930 milhões de dispositivos vestíveis conectados em todo o mundo [Caulfield et al. 2019] coletando dados que podem ser armazenados em sistemas de Registro Eletrônico de Saúde (RES) para uso posterior.

Apesar da disponibilidade de dispositivos de IoHT, são raras ou inexistentes as bases de dados (BD) que contemplem um monitoramento mais extensivo do paciente para além do ambiente hospitalar. Tal fato inviabiliza o desenvolvimento de um modelo totalmente individualizado e treinado unicamente com dados de um paciente. Assim, os estudos de IA aplicada à saúde são normalmente desenvolvidos com BD de pacientes hospitalizados. Devido ao período mais restrito de dados de um único indivíduo, são utilizados dados de outros pacientes para viabilizar o treinamento de modelos de predição. Tais modelos podem ser treinados com os dados de todos os pacientes do RES (abordagem tradicional populacional) ou com dados de pacientes similares caracterizando uma estratégia mais individualizada dentro da viabilidade técnica. Os resultados apresentados pelos modelos de IA dependem muito da qualidade do conjunto de dados disponível, ficando a IA sujeita ao viés dos dados utilizados para treinar os algoritmos [Gianfrancesco et al. 2018].

Neste âmbito, o presente artigo visa explorar a hipótese de que um modelo de IA para prever a mortalidade de pacientes com uma abordagem mais individualizada pode gerar resultados mais precisos minimizando o viés dos dados de treinamento. O conceito de individualização deve ser entendido aqui como a seleção de pacientes similares baseado na idade. Onde, a utilização da idade é respaldada pela análise das variáveis dos escores de risco heurísticos como APACHE¹ e SAPS² que foram desenvolvidos por profissionais da saúde e são amplamente utilizados em Unidade de Terapia Intensiva (UTI) para avaliar o risco de morte dos pacientes [Krishnan and S. 2019]. Tais escores atribuem uma pontuação específica para faixas de idade dos pacientes. Além disso, os dados apontam uma maior longevidade da população mundial sendo que com o envelhecimento, cresce a prevalência de doenças crônicas aumentando a mortalidade [Martha et al. 2020].

Para testar a hipótese apresentada, foram desenvolvidos modelos de *Deep Learning* (DL) baseados em Redes Neurais Convolucionais (RNC). Onde, os dados de treinamento foram selecionados com duas abordagens diferentes (populacional e individualizada) para comparação dos resultados. A definição dos sinais vitais a serem utilizados como entrada do modelo foi fundamentada nos critérios utilizados pelos escores APACHE e SAPS e que pudessem ser coletadas por dispositivos de IoHT de maneira não invasiva. A principal contribuição científica deste artigo é analisar os efeitos em termos

¹ Acute Physiology and Chronic Health Evaluation [Zimmennan et al. 2006]

² Simplified Acute Physiology Score [Metnitz et al. 2005]

de acurácia da utilização de uma abordagem mais individualizada em um modelo de DL treinado com um conjunto de variáveis de entrada mais restritas.

Dentro do conceito de IA e aprendizado de máquina (AM) encontram-se as Redes Neurais Artificiais (RNA), que buscam assemelhar-se aos conceitos biológicos do funcionamento do cérebro humano tendo como menor unidade de processamento o neurônio, logo, as RNA's são constituídas por camadas de neurônios [Haykin 2007]. Já o DL é uma subárea da AM, que emprega RNA's mais complexas, com um maior número de camadas de neurônios, para resolver problemas classificados em aprendizado semi-supervisionado [Goodfellow et al. 2016]. Uma RNC é um modelo de DL que utiliza uma RNA de múltiplas camadas com uma arquitetura profunda e aplica a convolução de matrizes em ao menos uma das suas camadas. Essas redes foram designadas para o processamento de dados em forma de matrizes e são muito utilizadas em algoritmos de reconhecimento de objetos em imagens [LeCun et al. 2015].

Este artigo está organizado em seis seções. Na Seção seguinte são apresentados os trabalhos relacionados; a Seção 3 expõe o modelo proposto neste trabalho; a metodologia utilizada para a construção, validação e avaliação do modelo proposto é descrita na Seção 4; a Seção 5 apresenta os resultados e discussão; a Seção 6 versa sobre as considerações finais e aponta as próximas etapas de pesquisa.

2. Trabalhos relacionados

O desenvolvimento de estratégias personalizadas de cuidado com a saúde em diferentes populações e ambientes é um dos componentes da medicina de precisão [Martha et al. 2020]. No entanto, quando o AM é utilizado para treinar modelos de predição em saúde, uma estratégia comumente aplicada é a populacional ou *one-size-fits-all* onde o diagnóstico e o tratamento são os mesmos para todos os pacientes sem distinção entre aspectos individuais do paciente como idade, gênero e etnia [Boness et al. 2020].

Na abordagem populacional é selecionado um conjunto de variáveis de todos os pacientes do RES para ser utilizado como entrada do modelo de ML com diferentes objetivos como a predição de mortalidade [Thorsen-Meyer et al. 2020, Caicedo-Torres and Gutierrez 2019, Todd et al. 2019]. Em outros casos são utilizados dados de grupos de pacientes com características comuns como comorbidades [dos Santos et al. 2020, Ershoff et al. 2020, Lin et al. 2019], gênero [Qi et al. 2018, Mansoor et al. 2017] e idade [Brajer et al. 2020, Cooper et al. 2018, van Loon et al. 2017]. Mesmo com subgrupos de interesse os modelos ainda são muito suscetíveis ao viés dos dados devido a grande heterogeneidade da população de pacientes. Além disso, ainda existe o problema clássico das classes desbalanceadas que afeta o resultado do modelo. É relativamente fácil apontar possíveis fontes de viés nos dados, no entanto, é desafiador encontrar formas de selecionar variáveis de treinamento para os algoritmos a fim de atenuar tal efeito [Obermeyer et al. 2019].

Considerando o objetivo do presente artigo são elencadas na Tabela 1 características importantes de alguns trabalhos considerados relacionados.

O estudo apresentado por [Ma et al. 2020] visa prever o tempo de internação de um paciente em uma UTI e utiliza este tempo em conjunto com a idade para estimar o risco de morte. Pacientes similares são definidos utilizando JITL baseado nas distâncias

Tabela 1. Comparativo de modelos de predição de mortalidade individualizados

Estudo	Técnicas	Variáveis entrada	Critério de similaridade
[Ma et al. 2020]	JITL ¹ ; one-class ELM ¹	15 - Fisiológicas	Distância euclidiana e angular
[Shamout et al. 2019] [Krishnan and S. 2019]	Distribuição estatística Algoritmo genético; ELM	7 - Sinais vitais 578 - Exames; Demográficas	Idade Algoritmo genético
[Ding et al. 2018]	JITL; PCA ³ ; ELM	26 - Sinais vitais; Exames; Demográficas	Distância sintética

¹ JITL - *Just-in-time Learning*.

² ELM - *Extreme Learning Machine*.

³ PCA - *Principal Component Analysis*.

euclidianas e angulares entre os valores das 15 variáveis de entrada do modelo. Com os dados dos pacientes similares é treinado um modelo ELM para prever se o paciente pode receber alta em até de 10 dias. [Shamout et al. 2019] desenvolveram um sistema paramétrico de alerta precoce (*Early Warning Score* - EWS) baseado em idade onde os pacientes foram separados em dois subgrupos (abaixo e acima de 45 anos). Os valores dos 7 sinais vitais que compõe o EWS foram divididos em intervalos utilizando distribuições estatísticas. Para cada intervalo de um sinal vital é atribuído um escore sendo a soma dos escores o valor final do EWS ao qual uma probabilidade de mortalidade é associada.

O modelo de predição de mortalidade proposto por [Krishnan and S. 2019] analisa inicialmente um conjunto complexo de 578 variáveis entre resultados de exames e dados demográficos. Algoritmos genéticos são utilizados para selecionar pacientes afins e as 10 variáveis mais significativas do conjunto inicial são utilizadas como entrada do modelo baseado em ELM. Já o modelo dos autores [Ding et al. 2018] baseado em JITL utiliza a distância sintética entre os dados das 26 variáveis de entrada para selecionar pacientes similares. A PCA é utilizada para reduzir a dimensionalidade dos dados de entrada do modelo de predição construído com ELM.

Conforme destacado através dos critérios de comparação estabelecidos na Tabela 1 os modelos de predição de mortalidade podem variar muito em termos de estratégias de construção. Não existe um consenso sobre uma técnica de ML ou heurística mais apropriada para determinado problema na área da saúde humana. Desta forma, a proposta deste artigo extrapola as soluções apresentadas na literatura propondo a avaliação do impacto da adoção de um critério de individualização em um modelo de predição de mortalidade baseado em *Deep Learning* utilizando como entrada um conjunto reduzido de variáveis que se considera possível serem coletadas pelos dispositivos de IoHT disponíveis atualmente.

3. Modelo proposto

O modelo de DL para predição de mortalidade utilizado neste artigo é baseado em uma RNC dividida em 5 estágios e contendo 9 camadas ocultas conforme proposto por [Schmidt et al. 2018] e apresentado na Figura 1.

A fim de avaliar o impacto da adição do critério de individualização foram treinados modelos diferentes, porém utilizando a mesma arquitetura de rede RNC. Para permitir a melhor compreensão das ideias apresentadas neste estudo, o modelo com a abordagem



Figura 1. Arquitetura da RNC [Schmidt et al. 2018]

populacional será referido como **Modelo A** ao longo do texto, e o modelo individualizado proposto neste artigo como **Modelo B**.

O **Modelo A** é considerado neste artigo como uma abordagem tradicional, onde os dados de todos os pacientes armazenados no RES são utilizados para treinar uma única RNC com o objetivo de prever o risco de morte. Uma ilustração desta abordagem é apresentada na Figura 2 onde é possível observar que esta estratégia não considera especificidades relacionadas a quaisquer características do conjunto de pacientes na construção do modelo. Algumas pesquisas têm mostrado que a inclusão de elementos de similaridade no modelo podem produzir resultados mais assertivos [Qi et al. 2018, Innocenti et al. 2018, Alaa et al. 2018].

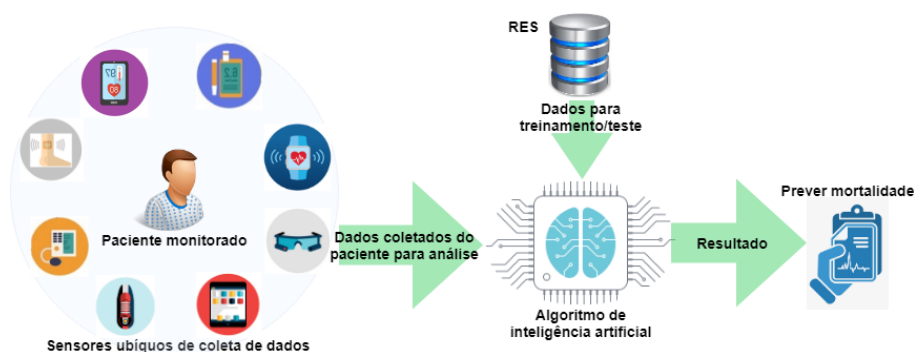


Figura 2. Representação de uma abordagem populacional (Modelo A)

A proposta apresentada neste artigo, **Modelo B** (Figura 3), sugere uma abordagem mais individualizada, onde várias RNCs serão treinadas, cada uma com uma parcela dos dados armazenados no RES segundo o critério de similaridade entre pacientes. O objetivo é que o indivíduo monitorado tenha seus dados usados como entrada para um modelo especificamente treinado com dados similares aos seus. No caso da implementação proposta para o este modelo, os dados dos pacientes serão segmentados por faixa de idade como critério de similaridade.

Evidências geradas a partir de dados de grupos de pacientes semelhantes capturados em RES podem ajudar na tomada de decisões da equipe de saúde [Gombar et al. 2019]. Assim, a hipótese que se deseja investigar é o efeito da utilização da estratégia individualizada (baseada em idade) aplicada ao modelo de DL utilizado no contexto de IoHT. Conjectura-se que a abordagem individualizada possa influenciar o processo de aprendizagem do algoritmo e, conseqüentemente, beneficiar o **Modelo B** em relação a sua assertividade.

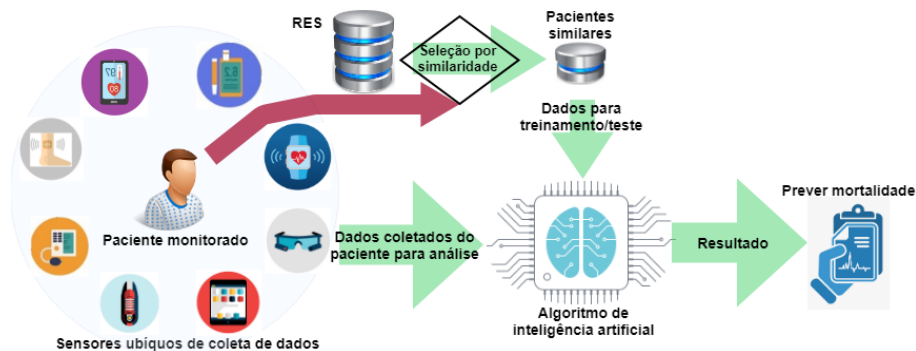


Figura 3. Representação de uma abordagem individualizada (Modelo B)

4. Metodologia

Esta seção apresenta a metodologia empregada no trabalho detalhando a origem do conjunto de dados utilizado, o critério de individualização adotado, as estratégias de treinamento, validação e teste das RNCs e as métricas de desempenho para avaliar os modelos.

4.1. Fonte de dados

Neste artigo foi considerado um conjunto de variáveis de entrada para a RNC que se presume poderem ser coletadas por dispositivos de IoHT tanto em ambientes hospitalares como de internação domiciliar. Devido à inexistência de uma base de dados apenas com dados coletados de um ambiente de internação domiciliar foram utilizados os dados do MIMIC III (*Multiparameter Intelligent Monitoring in Intensive Care III*). O MIMIC é um banco de dados composto por informações de pacientes internados em UTI, desenvolvido pelo Laboratório de Fisiologia Computacional do Instituto de Tecnologia de Massachusetts. Os dados incluem sinais vitais, exames laboratoriais, procedimentos clínicos, tempo de internação hospitalar, dados de sobrevivência, entre outros [Johnson et al. 2016].

Foram 8 as variáveis selecionadas como entrada para a RNC: idade, gênero, frequência cardíaca, frequência respiratória, glicose, pressão arterial sistólica, pressão arterial diastólica e temperatura corporal.

4.2. Critério de individualização

A individualização utilizada é baseada na idade dos pacientes no momento da internação, assim a seleção de dados armazenados no RES de pacientes similares para treinamento da RNC é efetuada com base em faixas de idade apresentadas na Tabela 2. Foram utilizadas as mesmas 5 faixas de idade estabelecidas no escore APACHE II [Knaus et al. 1985].

4.3. Pré-processamento

Foram extraídos dados da base MIMIC III de 25.451 internações de pacientes com idade ≥ 18 anos e com tempo de internação mínimo de 1 hora, sendo que 13% destas internações apresentaram óbito do paciente. As internações consideradas neste ensaio não apresentavam dados faltantes e não foi efetuado nenhum filtro ou tratamento para medidas possivelmente incorretas. Para cada uma das 8 variáveis foi calculado o valor mínimo, médio e máximo do período de internação para ser utilizado como entrada da RNC.

Os dados utilizados são detalhados na Tabela 2 já separados nas cinco faixas de idade estabelecidas no escore APACHE II. Baseado nestas informações é possível perceber que os dados são desbalanceados pois apenas 13% das internações evoluíram para óbito do paciente. Um possível ponto de viés para uma abordagem populacional é a prevalência de pessoas com idade ≥ 65 anos que correspondem a mais de 51% dos pacientes. Estas características podem afetar de forma negativa o desempenho da técnica de aprendizado de máquina utilizada para prever a mortalidade.

Tabela 2. Detalhes dos dados extraídos da base MIMIC III

Faixa	Idade (anos)	Internações	Mortalidade (%)
1	≤ 44	3801	248 (6,53%)
2	45-54	3690	342 (9,27%)
3	55-64	4882	524 (10,73%)
4	65-74	5083	643 (12,66%)
5	≥ 75	7995	1721 (21,53%)

4.4. Treinamento, validação e teste do modelo de *deep learning*

Os algoritmos de aprendizado de máquina precisam ser treinados, validados e testados em conjuntos de dados independentes para assegurar a capacidade de generalização, isto é, produzir previsões corretas ao analisar novos dados não rotulados. Segundo [Angermueller et al. 2016] o método chamado *holdout* é o padrão para divisão do conjunto de dados para técnicas de DL e consiste em dividir os dados, aleatoriamente, em três subconjuntos independentes (mutuamente exclusivos) tipicamente com a seguinte proporção:

- Treinamento (e validação): compreende 70% do total de dados onde 60% é utilizado para aprender modelos com diferentes hiper-parâmetros. Estes modelos são então avaliados com o conjunto de dados de validação que corresponde a 10% do conjunto de dados;
- Teste: o modelo com melhor desempenho na fase de treinamento, baseado na acurácia, é selecionado e avaliado com o conjunto de dados de teste (30% restantes do conjunto de dados) para quantificar o desempenho em dados não submetidos ao modelo.

Acurácia é uma métrica para medir o desempenho de classificadores, verificando a proporção entre as previsões corretamente feitas pelo modelo e o total de dados analisados. Previsões corretas são aquelas em que a saída dada pelo modelo corresponde, de fato, ao rótulo esperado, por exemplo, saídas classificadas como “positivo” e que possuíam o rótulo “positivo” (verdadeiros positivos - VP) e saídas classificadas como “negativo”, que possuíam o rótulo “negativo” (verdadeiros negativos - VN) [Nicolas 2017].

A curva de Característica de Operação do Receptor (curva ROC) é um gráfico de sensibilidade versus especificidade. Sensibilidade é a capacidade de o modelo prever casos positivos corretamente (taxa de VP) e especificidade é a capacidade de prever casos negativos corretamente (taxa de VN) [Ho et al. 2007]. A comparação entre os Modelos A e B foi feita por meio da área sob a curva ROC (AUC), onde o valor pode variar de 0,5 (predição aleatória) a 1,0 (predição perfeita).

A Tabela 3 apresenta a divisão do conjunto de dados extraídos do MIMIC III entre treino (70% que corresponde a 17.815 internações) e teste (30% que equivale a 7.636

Tabela 3. Divisão do número de interações para treino e teste

Faixa de idade	Modelo A		Modelo B	
	Treinamento	Teste	Treinamento	Teste
1	17815	1142	2659	1142
2	17815	1148	2542	1148
3	17815	1446	3436	1446
4	17815	1537	3546	1537
5	17815	2363	5632	2363

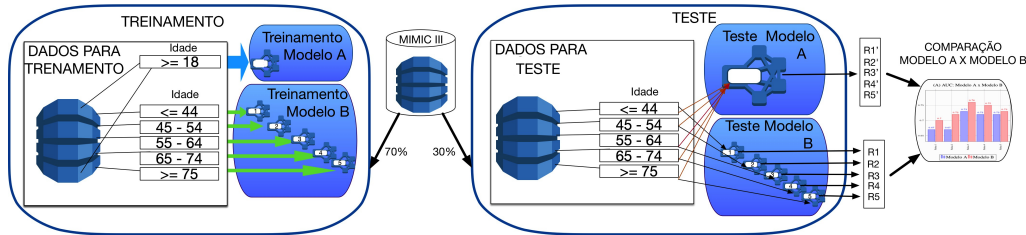


Figura 4. Esquema de distribuição de dados para treinamento e teste das RNC

interações). Estes conjuntos de dados passaram por um novo processo de divisão que consiste na segmentação por faixa de idade.

No Modelo A, abordagem populacional, todo o conjunto treino (17.815 interações) é utilizado para treinar um único modelo. Já no Modelo B, com a abordagem individualizada, o conjunto de treino é segmentado em 5 faixas de idade (conforme Tabela 2) e são treinados vários modelos diferentes, um para cada faixa de idade.

Durante os testes o conjunto de dados de teste foi segmentado nas mesmas 5 faixas de idade já mencionadas para ambos os Modelos A e B. Isso foi necessário para viabilizar a comparação dos resultados gerados. A Figura 4 ilustra esta segmentação dos dados de treino e teste por faixa de idade nas duas abordagens analisadas (Modelo A e Modelo B).

5. Resultados e discussão

Os dois modelos, A e B, foram treinados e testados de acordo com o delineamento apresentado na Figura 4 e seus valores de AUC comparados a fim de avaliar o impacto da individualização no modelo de DL apresentado. O resultado da comparação é exposto na Figura 5(a). É possível observar que o AUC para o **Modelo B** superou o resultado obtido pelo **Modelo A** em todas as faixas de idade. Isso confirma a hipótese de que modelos de DL treinados com dados de pacientes que tenham alguma similaridade são capazes de realizar melhores previsões. Assim, esse resultado corrobora os indicativos de que aspectos específicos de cada paciente influenciam os desfechos clínicos [Royal College of Physicians 2017] ressaltando a relevância em aprofundar estudos a fim de viabilizar tecnicamente o desenvolvimento de modelos de predição clínica mais individualizados e menos suscetíveis ao viés dos dados.

A Figura 5(b) retrata o percentual de ganho no AUC em relação à taxa de mortalidade presente em cada faixa de dados, onde os valores do eixo x representam a taxa de mortalidade presente no conjunto de dados e os valores do eixo y representam o percentual de ganho no AUC do Modelo B em relação ao Modelo A. É possível observar que o ganho em termos de AUC é menor na faixa de idade 5 (≥ 75 anos) que corresponde ao

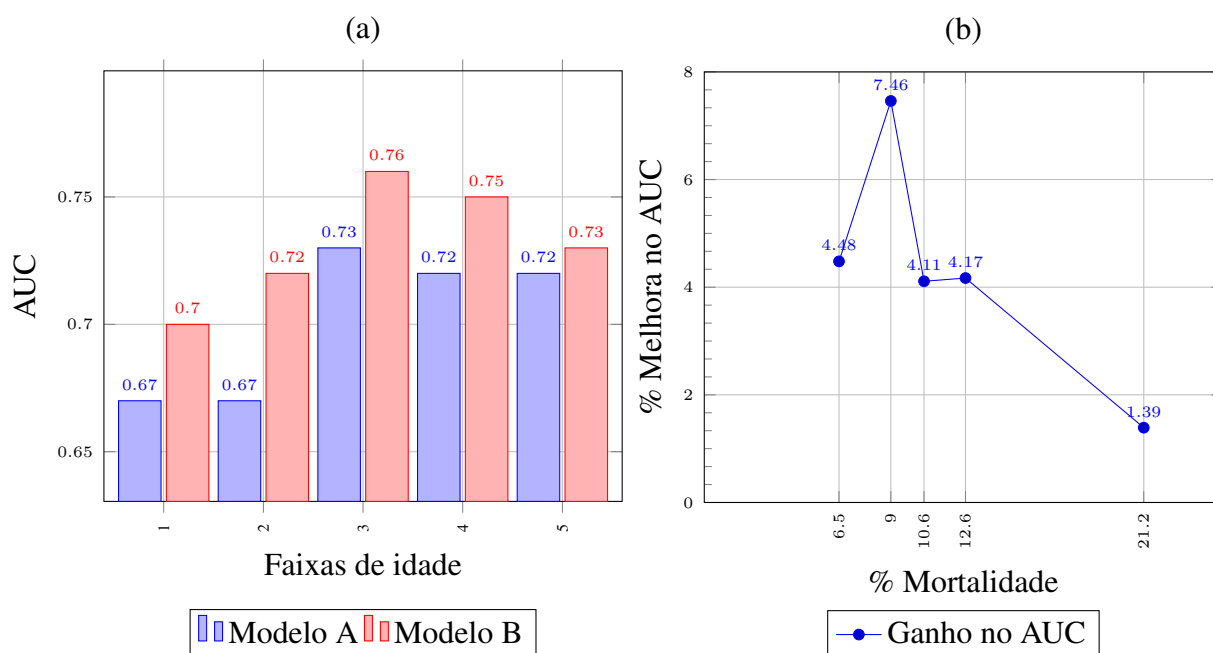


Figura 5. (a) AUC: Modelo A X Modelo B (b) % de ganho no AUC por faixa etária

conjunto de dados com maior número de internações e também maior percentual de mortalidade (21,53%). Este resultado ilustra como a população com maior representatividade influencia no treinamento do modelo de DL principalmente na abordagem populacional.

O Modelo B com a abordagem individualizada mostra expressivo ganho em termos de acurácia da predição em faixas de idade onde tanto o número de pacientes é menor quanto a incidência de mortalidade em relação ao grupo predominante. Entre estas faixas o ganho no AUC é similar, em média 4.2%, à exceção da faixa 2, que obteve 7.46% de ganho no AUC, possivelmente por conter o menor volume de internações dentre todas as faixas de idade. Assim o impacto da abordagem individualizada é mais proeminente por justamente melhorar os resultados de predição neste grupo menos numeroso, mas, não menos importantes em termos de objetivo final de predição de mortalidade.

Essa análise dos resultados demonstra como o conjunto de dados impacta no desempenho do modelo de DL principalmente em termos de viés. É possível afirmar que, quanto mais individualizado o modelo de DL, no sentido da medicina de precisão, melhores resultados serão produzidos em termos de predição. Sem mensurar os possíveis impactos em termos de custos computacionais para uma implementação em um ambiente de UTI de modelos de DL individualizados pode-se afirmar que os resultados apresentados são promissores e alinhados a outros estudos na área.

6. Conclusão

Neste artigo foi retratada uma proposta de uma abordagem individualizada de predição de mortalidade de pacientes que, em parte, está alinhada com o conceito emergente de medicina de precisão. A individualização foi baseada no treinamento de modelos de DL usando apenas dados de pacientes similares em termos de idade. Os dados dos pacientes armazenados no sistema RES foram segmentados em faixas de idade seguindo o mesmo critério estabelecido no escore APACHE II. Os resultados denotam uma sensível melhora

na acurácia da predição de mortalidade de um modelo de DL treinado com abordagem individualizada proposta em relação à estratégia de treinamento tradicional (populacional).

Em trabalhos futuros pretende-se avaliar outros critérios de individualização e algoritmos de aprendizado de máquina como *Random Forest*. Outro ponto importante a ser considerado em pesquisas futuras é analisar a capacidade de generalização do modelo individualizado proposto testando outras bases de dados.

Agradecimentos

Esse estudo foi parcialmente financiado pela CAPES (código de financiamento 001) e pelo CNPq (processos 405354/2016-9 e 303640/2017-0).

Referências

- Alaa, A. M., Yoon, J., Hu, S., and van der Schaar, M. (2018). Personalized risk scoring for critical care prognosis using mixtures of gaussian processes. *IEEE Transactions on Biomedical Engineering*, 65(1):207–218.
- Angermueller, C., Pärnamaa, T., Parts, L., and Stegle, O. (2016). Deep learning for computational biology. *Molecular systems biology*, 12(7):878.
- Barbour, D. L. (2019). Precision medicine and the cursed dimensions. *npj Digital Medicine*, 2(1).
- Boness, C. L., Loeffelman, J. E., Steinley, D., Trull, T., and Sher, K. J. (2020). Using complete enumeration to derive “one-size-fits-all” versus “subgroup-specific” diagnostic rules for substance use disorder. *Assessment*.
- Brajer, N., Cozzi, B., Gao, M., Nichols, M., Revoir, M., Balu, S., Futoma, J., Bae, J., Setji, N., Hernandez, A., and Sendak, M. (2020). Prospective and external evaluation of a machine learning model to predict in-hospital mortality of adults at time of admission. *JAMA Network Open*, 3(2).
- Caicedo-Torres, W. and Gutierrez, J. (2019). Iseeu: Visually interpretable deep learning for mortality prediction inside the icu. *Journal of Biomedical Informatics*, 98(103269).
- Caulfield, B., Reginatto, B., and Slevin, P. (2019). Not all sensors are created equal: a framework for evaluating human performance measurement technologies. *npj Digital Medicine*, 2.
- Cooper, J. N., Minneci, P. C., and Deans, K. J. (2018). Postoperative neonatal mortality prediction using superlearning. *Journal of Surgical Research*, 221:311–319.
- da Costa, C. A., Pasluosta, C. F., Eskofier, B., da Silva, D. B., and da Rosa Righi, R. (2018). Internet of health things: Toward intelligent vital signs monitoring in hospital wards. *Artificial Intelligence in Medicine*, 89:61 – 69.
- Ding, Y., Wang, Y., and Zhou, D. (2018). Mortality prediction for icu patients combining just-in-time learning and extreme learning machine. *Neurocomputing*, 281:12–19.
- dos Santos, H. G., Zampieri, F. G., Normilio-Silva, K., da Silva, G. T., de Lima, A. C. P., Cavalcanti, A. B., and Filho, A. D. P. C. (2020). Machine learning to predict 30-day quality-adjusted survival in critically ill patients with cancer. *Journal of Critical Care*, 55:73–78.

- Ershoff, B. D., Lee, C. K., Wray, C. L., Agopian, V. G., Urban, G., Baldi, P., and Cannesson, M. (2020). Training and validation of deep neural networks for the prediction of 90-day post-liver transplant mortality using unos registry data. *Transplantation Proceedings*, 52(1):246–258.
- Gianfrancesco, M. A., Tamang, S., Yazdany, J., and Schmajuk, G. (2018). Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Internal Medicine*, 178(11):1544–1547.
- Gombar, S., Callahan, A., Califf, R., Harrington, R., and Shah, N. H. (2019). It is time to learn from patients like mine. *npj Digital Medicine*, 2.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.
- Haykin, S. (2007). *Redes neurais: princípios e prática*. Bookman, Porto Alegre.
- He, J., Baxter, S. L., Xu, J., Xu, J., and Zhou, X. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, 25(1).
- Ho, K., Lee, K., Williams, T., Finn, J., Knuiman, M., and Webb, S. (2007). Comparison of acute physiology and chronic health evaluation (apache) ii score with organ failure scores to predict hospital mortality. *Anaesthesia*, 62(5):466–473.
- Innocenti, F., Tozzi, C., Donnini, C., Villa, E. D., Conti, A., Zanobetti, M., and Pini, R. (2018). Sofa score in septic patients: incremental prognostic value over age, comorbidities, and parameters of sepsis severity. *Internal and Emergency Medicine*, 13.
- Johnson, A. E., Pollard, T. J., Shen, L., wei H. Lehman, L., 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. Article number: 160035.
- Kim, H. J., Kim, H. J., Park, Y., Lee, W. S., Lim, Y., and Kim, J. H. (2020). Clinical genome data model (cgdm) provides interactive clinical decision support for precision medicine. *Scientific Reports*, 10(1):1414.
- Knaus, W. A., Draper, E. A., Wagner, D. P., and Zimmerman, J. E. (1985). APACHE II: a severity of disease classification system. *Critical care medicine*, 13(10):818–829.
- Krishnan, G. S. and S., S. K. (2019). A novel ga-elm model for patient-specific mortality prediction over large-scale lab event data. *Applied Soft Computing*, 80:525–533.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553).
- Lin, K., Hu, Y., and Kong, G. (2019). Predicting in-hospital mortality of patients with acute kidney injury in the icu using random forest model. *International Journal of Medical Informatics*, 125:55–61.
- Ma, X., Si, Y., Wang, Z., and Wang, Y. (2020). Length of stay prediction for icu patients using individualized single classification algorithm. *Computer Methods and Programs in Biomedicine*, 186.
- Mansoor, H., Elgendy, I. Y., Segal, R., Bavry, A. A., and Bian, J. (2017). Risk prediction model for in-hospital mortality in women with st-elevation myocardial infarction: A machine learning approach. *Heart & Lung*, 46(6):405–411.
- Martha, S. R., Auld, J. P., Hash, J. B., and Hong, H. (2020). Precision health in aging and nursing practice. *Journal of Gerontological Nursing*, 46(3):3–6.

- Metnitz, P. G. H., Moreno, R. P., Almeida, E., Jordan, B., Bauer, P., Campos, R. A., Iapichino, G., Edbrooke, D., Capuzzo, M., and Gall, J.-R. L. (2005). Saps 3—from evaluation of the patient to evaluation of the intensive care unit. part 2: Development of a prognostic model for hospital mortality at icu admission. *Intensive care medicine*, 31(10):1345–1355.
- Mills, J. R. (2017). Precision medicine—right treatment, right patient, right time, wrong approach? *Clinical Chemistry*, 63(4).
- Nicolas, P. R. (2017). *Scala for Machine Learning: Data processing, ML algorithms, smart analytics, and more*. Packt Publishing Ltd, Birmingham, UK, second edition.
- Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453.
- Qi, Y., Wang, W., Zhang, K., An, S., Wang, S., Zheng, J., and Yi-DaTang (2018). Development and validation of women acute myocardial infarction in-hospital mortality score (wami score). *International Journal of Cardiology*, 259:31–39.
- Royal College of Physicians (2017). National early warning score (news) 2: Standardising the assessment of acute-illness severity in the NHS. Technical report, Royal College of Physicians, London.
- 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 *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2018)*. SBC.
- Shamout, F., Zhu, T., Clifton, L., Briggs, J., Prytherch, D., Meredith, P., Tarassenko, L., Watkinson, P. J., and Clifton, D. A. (2019). Early warning score adjusted for age to predict the composite outcome of mortality, cardiac arrest or unplanned intensive care unit admission using observational vital-sign data: a multicentre development and validation. *BMJ open*, 9(11).
- Shen, X. (2020). An imprecise path to precision medicine. *Nature Medicine*, 26(1):14–14.
- Thorsen-Meyer, H.-C., Nielsen, A. B., Nielsen, A. P., Kaas-Hansen, B. S., Toft, P., Schierbeck, J., Strøm, T., Chmura, P. J., Heimann, M., Dybdahl, L., Spangsege, L., Hulsén, P., Belling, K., Brunak, S., and Perner, A. (2020). Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. *The Lancet Digital Health*.
- Todd, J., Gepp, A., Richards, B., and Vanstone, B. J. (2019). Improving mortality models in the icu with high-frequency data. *International Journal of Medical Informatics*, 129:318–323.
- van Loon, I. N., Bots, M. L., Boereboom, F. T. J., Grooteman, M. P. C., Blankestijn, P. J., van den Dorpel, M. A., Nubé, M. J., Ter Wee, P. M., Verhaar, M. C., and Hamaker, M. E. (2017). Quality of life as indicator of poor outcome in hemodialysis: relation with mortality in different age groups. *BMC nephrology*, 18(1):217–217.
- Zimmennan, J. E., Kramer, A. A., McNair, D. S., and Malila, F. M. (2006). Acute physiology and chronic health evaluation (apache) iv: hospital mortality assessment for today’s critically ill patients. *Critical care medicine*, 34(5):1297–1310.