Um Modelo de Predição de Mortalidade em Unidades de Terapia Intensiva Baseado em Deep Learning

Diogo Schmidt, Denise Bandeira da Silva, Cristiano André da Costa, Rodrigo da Rosa Righi
SOFTWARELAB – Software Innovation Laboratory,
Programa de Pós-Graduação em Computação Aplicada (PPGCA),
Universidade do Vale do Rio dos Sinos (Unisinos)
São Leopoldo, RS, Brazil

Abstract. *The usage of Deep Learning techniques has become even more frequent in medical research due to the possibilities that it offers to improve the quality of Clinical Decision Support. Several conventional prognostic models have been used in Intensive Care Units (ICU) to evaluate the risk of death. However, those models still cannot accurately predict that risk. For this reason, the aim of this article is to offer a model based on Deep Learning to predict the risk of mortality, especially in the fields of intensive care medicine, in order to make healthcare more efficient. The model consists of a Convolutional Neural Network that is divided into five stages, which contain nine hidden layers. In the proposed model we use quantitative methods in its processes. An experimental approach is given when comparing the predictive power of the proposed model to one of the most used models for predictions in ICU, the APACHE II. The data used were extracted from medical records available in the Multiparameter Intelligent Monitoring in Intensive Care III (MIMIC III) database. In order to evaluate the performance of the models, measures of accuracy, sensitivity, specificity and Area Under the ROC Curve (AUC) were used. After comparing the performance of the proposed model to the APACHE II, the proposed model presented positive results as it reached an AUC of more than 0,80, whereas the APACHE II reached an AUC of 0,71.*

Resumo. *Cada vez mais, pesquisas na área da medicina vêm utilizando a técnica de Deep Learning na tentativa de melhorar a qualidade dos sistemas de Apoio à Decisão Clínica. Diversos índices prognósticos convencionais são utilizados em Unidades de Terapia Intensiva (UTIs) para avaliar o risco de morte. Entretanto, esses índices ainda não conseguem prever com precisão o risco de morte. O objetivo deste trabalho é apresentar um modelo baseado em Deep Learning, na área de medicina intensiva, para a predição do risco de morte, a fim de tornar a decisão terapêutica mais eficiente. O modelo de predição proposto é formado por uma Rede Neural Convolucional (RNC) dividida em cinco estágios e contendo nove camadas ocultas. Neste trabalho, de abordagem quantitativa, é dado um enfoque experimental através da comparação do poder preditivo do modelo proposto com um dos modelos de predição mais utilizados em UTIs, o Acute Physiology and Chronic Health Evaluation II (APACHE II). Os dados utilizados foram obtidos através da extração de informações contidas em registros médicos da base de dados Multiparameter Intelligent Monitoring in Intensive Care III (MIMIC III). Para avaliar o desempenho dos modelos foram utilizadas as medidas de acurácia,*

sensibilidade, especificidade e área sob a curva ROC (AUC). Comparando o desempenho do modelo proposto com o APACHE II, observou-se que ele apresentou bons resultados ao atingir AUC acima de 0,80, enquanto o APACHE II atingiu AUC de 0,71.

1. Introdução

De acordo com LeCun et al. (2015), diversas áreas como negócios, ciências e medicina têm apresentado sucesso na utilização dessa vasta disponibilidade de dados ao empregarem a tecnologia de Machine Learning (Aprendizado de Máquina) para análise de Big Data. Os autores relatam que Machine Learning tem aprimorado pesquisas realizadas na internet, recomendações de amizades nas redes sociais e predição de texto em smartphones. Sistemas utilizando Machine Learning também estão sendo usados, com alto grau de precisão, para identificação automatizada de objetos em imagens, processamento de linguagem natural e reconhecimento de padrões.

Dentre as técnicas de aprendizado de máquina, os algoritmos de Deep Learning (Aprendizagem Profunda) demonstram grandes avanços na extração automatizada de representações com vários níveis de abstração a partir de dados complexos e são, portanto, aplicáveis a diferentes domínios da ciência. Tais algoritmos são capazes de detectar correlações existentes entre variáveis que não estão aparentemente visíveis, possibilitando indicar um padrão não observado quando as variáveis são analisadas isoladamente [LeCun et al. 2015].

Nesse contexto, o campo de Health Informatics (Informática em Saúde) tem alcançado perspectivas importantes com a utilização de Big Data. Novas possibilidades de obter e registrar dados como Registros Eletrônicos de Saúde (RES), sensores eletrônicos e dispositivos móveis, em conjunto com técnicas de Aprendizado de Máquina, estão tornando possível o aprimoramento de diagnósticos e predições, qualificando o tratamento e reduzindo custos [Herland et al. 2014].

O aumento de doenças crônicas e o envelhecimento populacional em escala global vêm contribuindo significativamente para a crescente demanda por medicina intensiva. Em diversos países do mundo, Unidades de Terapia Intensiva (UTIs) são as unidades hospitalares onde a alocação de recursos material, humano e tecnológico é fundamental para a preservação da vida. Nesse cenário, sistemas de Apoio à Decisão Clínica (Clinical Decision Support - CDS) ou Sistemas de Informação em Saúde (SIS) têm sido desenvolvidos, em nível nacional e mundial, para auxiliar no processo de investigação médica, diagnóstico, terapêutica e auxiliar no processo de tomada de decisão. [Marin 2010]. Especificamente nas UTIs, com o objetivo de melhorar a alocação de recursos e na avaliação dos serviços prestados, foram desenvolvidos índices prognósticos a fim de predizer a mortalidade e estimar a gravidade de doenças, visando aumentar a eficácia na área de medicina intensiva [Neto et al. 2015].

Segundo Hissa et al. (2013), diversos índices prognósticos vêm sendo utilizados na área de cuidados intensivos para predizer a mortalidade e estimar a gravidade da doença: Acute Physiology and Chronic Health Evaluation (APACHE II), UNICAMP II (Modelo da Universidade de Campinas II), Simplified Acute Physiology Score II (SAPS II), Logistic Organ Dysfunction System (LODS) e Sepsis Related Organ Failure Assessment (SOFA). Conforme Moritz et al. (2005), no Brasil, o índice APACHE II é usado desde 1985 para

calcular probabilidade de óbito hospitalar. De acordo com os autores, esse índice prognóstico é um dos escores mais utilizados no mundo como preditor de mortalidade, sendo de fácil e rápida execução. O escore APACHE II requer a utilização de variáveis fisiológicas, laboratoriais, idade, status neurológico e presença de comorbidades.

Entretanto, Pirracchio et al. (2015) relatam que apesar de vários índices prognósticos, incluindo o APACHE II, terem sido desenvolvidos e aperfeiçoados ao longo dos anos, estes escores ainda não conseguem prever com precisão o risco de morte. Diante dessa limitação, técnicas de Aprendizado de Máquina aplicadas à Big Data na área da saúde, em especial a técnica de Deep Learning, oportunizam o desenvolvimento de novos algoritmos preditores, visando obter melhores resultados na predição do risco de morte em UTIs [Deo 2015, Lecun et al. 2015].

A implementação de um projeto para a análise de Big Data (Big Data analytics) na área da saúde envolve a adoção de uma sequência de passos que constituem um pipeline de processamento de informações de saúde (Figura 1) [Fang et al. 2016]. O objetivo principal desse pipeline é explorar padrões significativos que possam ser encontrados ao analisar grandes volumes de dados, seguindo os seguintes passos: capturar grandes volumes de dados, armazenar os dados brutos capturados, compartilhar adequadamente os dados, analisar os dados (pré-processamento, seleção de variáveis, aplicação de técnicas de Machine Learning para gerar conhecimento), procurar por padrões de informações e dar suporte a decisões clínicas.

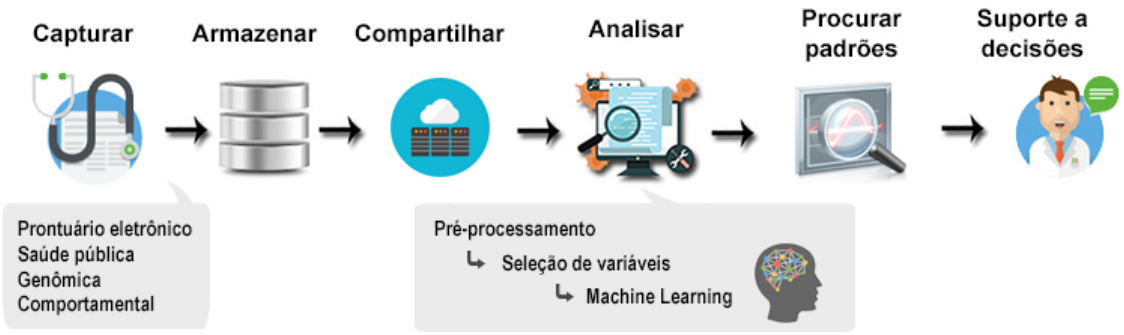

Figura 1 - Pipeline de processamento de informações de saúde

Nesse âmbito, o presente trabalho busca apresentar um modelo para aplicação da técnica de Deep Learning na predição de mortalidade em UTIs, oportunizando uma decisão terapêutica mais eficiente e assim reduzindo as chances de erros na área de medicina intensiva. Este trabalho apresenta um enfoque experimental através da comparação do poder preditivo de um dos índices prognósticos convencionais mais utilizados em UTIs, o escore APACHE II, com um modelo preditivo utilizando Deep Learning.

O artigo está organizado em 8 seções. A seção 2 apresenta o referencial teórico contendo a revisão bibliográfica que traz embasamento para o presente artigo. Na seção 3 são apresentados alguns trabalhos relacionados. As seções 4 e 5 descrevem, respectivamente, o modelo proposto e sua implementação. Na seção 6 são apresentados os resultados obtidos que são discutidos na seção 7. Por fim, a seção 8 traz as conclusões acerca do modelo proposto.

2. Fundamentação Teórica

O termo Deep Learning (Aprendizagem Profunda) é um conceito emergente que vem apresentando grandes avanços na solução de problemas há muitos anos tratados pela comunidade de inteligência artificial. Algoritmos de Deep Learning permitem ao computador aprender conceitos complexos a partir de conceitos mais simples. Essa abordagem já demonstrou sucesso empírico em uma série de aplicações como visão computacional e processamento de linguagem natural [Bengio et al. 2013].

Embora seja, muitas vezes, desafiador treinar de forma eficiente arquiteturas profundas, elas têm sido objeto de diversas pesquisas, pois apresentam duas vantagens importantes: promovem a reutilização de recursos e podem levar a compressões progressivamente mais abstratas nas camadas mais altas. A capacidade de construir múltiplos níveis de representação está relacionada com a profundidade da arquitetura. Em circuitos profundos, o número de caminhos, ou a forma de reutilizar diferentes partes dele, cresce exponencialmente com a profundidade, permitindo que, mudando a definição do que cada nodo pode calcular, a profundidade do circuito seja alterada [Bengio et al. 2013].

Uma Rede Neural Convolucional (RNC) é um tipo de rede neural de múltiplas camadas com arquitetura profunda, inspirada no funcionamento biológico de processamento de dados visuais, que utilizam a operação de convolução ao invés da multiplicação de matrizes em pelo menos uma das camadas. RNCs vêm sendo amplamente utilizadas pela comunidade da computação, alcançando resultados práticos muito positivos (veículos e robôs autônomos), e apresentando o maior sucesso no tratamento de problemas envolvendo detecção, segmentação e reconhecimento de objetos em imagens [Lecun et al. 2015].

Convolução é um tipo especializado de operação linear, que combina operações para extrair (filtrar) determinadas informações de um conjunto de dados. Essa operação possui características importantes que permitem melhorar o desempenho do Aprendizado de Máquina: conectividade esparsa e compartilhamento de parâmetros [Goodfellow et al. 2016].

3. Trabalhos relacionados

As técnicas de Machine Learning têm sido aplicadas na área da saúde com o objetivo de antever a evolução do estado de saúde dos pacientes. Os modelos desenvolvidos têm por objetivo aprimorar os resultados obtidos pelos índices prognósticos comumente empregados com o mesmo propósito, como os escores APACHE II, SOFA, EWS (Early Warning Score), NEWS (National Early Warning Score) e MEWS (Modified Early Warning Score) [Hissa et al. 2013]. Nesta seção são apresentados alguns trabalhos que propõem soluções algorítmicas, baseadas em análises de dados por meio de técnicas de Machine Learning, para esse problema.

Na tentativa de buscar preditores de mortalidade hospitalar mais precisos, aplicado à pacientes internados em UTIs, Pirracchio et al. (2015) apresentam uma nova proposta utilizando Machine Learning. O projeto denominado Super ICU Learner Algorithm (SICULA), visa melhorar a predição em relação aos escores convencionais como o APACHE II, SOFA e SAPS II. A técnica empregada chamada de Super Learner (SL) é uma técnica do tipo Ensemble Machine Learning que utiliza diversos algoritmos para

obter um melhor desempenho de predição dentro de um conjunto de algoritmos determinados. De acordo com os autores, o estudo demonstrou uma melhor performance ao analisar a Área sob a curva Característica de Operação do Receptor (AUROC) do algoritmo SL (0,85 versão 1 e 0,88 versão 2) em comparações aos escores SOFA (0,71), SAPS II (0,78), SAPS II-adaptado (0,83) e APACHE II-adaptado (0,82).

Com o objetivo de identificar precocemente o risco iminente de morte de um paciente, Ghosh et al. (2017) desenvolveram um preditor chamado de Early Deterioration Indicator (EDI), que busca detectar sinais de deterioração aguda do paciente através de mudanças sutis nos sinais vitais. Segundo os autores, muitas vezes essas pequenas variações passam desapercebidos pelas escalas de alerta baseadas em atribuição ponderada de pontos Early Warning Score (EWS). O EDI utiliza uma abordagem baseada em dados (data-driven approach) para obter um maior desempenho preditivo em relação aos escores EWS mais amplamente utilizados, o MEWS e o NEWS. Utilizando o algoritmo de classificação probabilística Naive Bayes, o EDI apresentou melhores resultados na previsão de deterioração (EDI AUROC: 0,6504/ 0,8418, NEWS AUROC: 0,572/0,7649, MEWS AUROC 0,5552/0,7565).

Em Eskofier et al. (2015), é apresentado um estudo combinando o grande volume de dados fornecidos por sensores vestíveis, chamados de wearables (sistemas que combinam giroscópios e acelerômetros), e RNCs para a classificação automatizada de distúrbios do movimento, como a Doença de Parkinson. De acordo com os autores, ao comparar com algoritmos de Machine Learning padrão, o algoritmo Deep Learning foi o que apresentou melhor desempenho, atingindo acurácia de 90,9%. Dentre os ou outros algoritmos, o AdaBoost.M1 apresentou maior acurácia, 86,3%, e o pior algoritmo foi o kNN, atingindo 67,1%.

4. Modelo Proposto

Conforme apresentado no referencial teórico, RNCs são formadas por uma série de camadas empilhadas, agrupadas em diferentes funções (estágios). Os primeiros estágios são formados por dois tipos de camadas: a camada de convolução e a camada de subamostragem (pooling layer). As camadas de convolução têm a função de extrair características locais para a formação do mapa de características (feature maps), através de um filtro bidimensional (kernel) que é deslizado por toda a entrada (Figura 2) e seus parâmetros (pesos) são ajustados automaticamente ao longo do aprendizado. As operações de convolução são acompanhadas de uma função de ativação, como a Rectified Linear Units (ReLU). A ReLU é um tipo de função de não-linearidade, utilizada em arquiteturas Deep Learning por proporcionar uma aprendizagem mais rápida [Nair and Hinton 2010].

A camada de subamostragem é responsável por mesclar, semanticamente, características semelhantes, com o objetivo de reduzir a dimensionalidade (tamanho da entrada) na camada seguinte. Outra função da camada de subamostragem é tornar a rede invariante a pequenas distorções na entrada. Uma camada típica de subamostragem é a Max-pooling (Figura 2), que consiste em capturar e transmitir o valor mais significativo (máximo de ativação) de determinada região da entrada. Após as camadas de convolução e subamostragem, são acrescentadas duas ou três camadas totalmente conectadas (RNAs com ReLU) e, por fim, uma camada de saída onde é realizada a classificação final, que é

ativada pela Função Sigmóide no caso de classificações binárias (0 ou 1) [Jarrett et al. 2009].

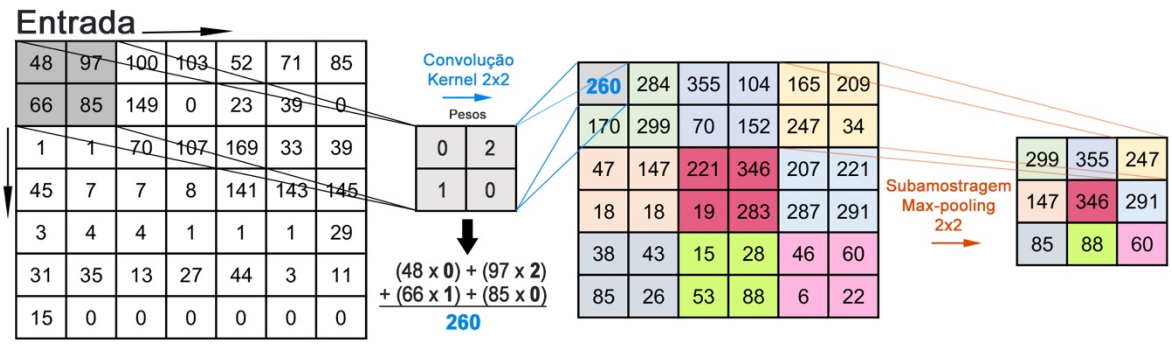

Figura 2 – Convolução e Max-pooling: processo de subamostragem para determinar a máxima ativação das regiões de entrada

Diante do exposto, o modelo de Deep Learning proposto neste projeto (Figura 3) é uma RNC, tendo como base a arquitetura utilizada por Eskofier et al. (2015), em que os autores aplicaram RNC para detecção da Doença de Parkinson a partir de dados capturados por sensores de movimento, obtendo resultados promissores. A arquitetura do modelo proposto é formada pelos seguintes estágios: 1 camada de entrada; 2 estágios formados por camadas de convolução, ReLU e Max-Pooling; 2 camadas totalmente conectadas (RNA com ReLU); 1 camada de saída utilizando a função Sigmóide.

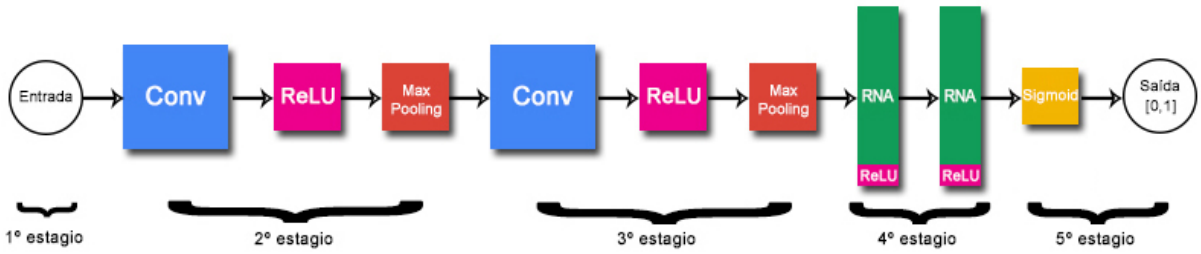

Figura 3 – Modelo proposto para predição de mortalidade baseado em Deep Learning

Diversos frameworks (conjunto de bibliotecas) para Deep Learning têm sido desenvolvidos nos últimos anos. Dentre os mais populares está o TensorFlow, um dos frameworks mais recentes para desenvolvimento de aplicações Deep Learning. Desenvolvido pelo Google, é uma ferramenta de código aberto que foi construído para ser flexível, eficiente, extensível e portável. Possui uma interface completa em Python e conta com uma comunidade crescente de usuários e contribuintes [Angermueller et al. 2016].

5. Implementação

Os experimentos foram realizados utilizando o framework TensorFlow (versão 1.0.1), a interface Keras API (versão 2.0.2), a linguagem de programação Python (versão 2.7.10) e a aplicação web Jupyter Notebook (versão 4.3.0), em um computador com processador Intel Core i5 3.2GHz e 8GB de memória RAM com sistema operacional MacOS (versão 10.12.3).

Os dados utilizados neste artigo foram obtidos através da extração de informações contidas em registros médicos de pacientes adultos (com tempo de internação maior que 24 horas), não identificados, existentes na base de dados Multiparameter Intelligent Monitoring in Intensive Care III (MIMIC III), versão v1.4 [Goldberger et al. 2000, Johnson et al. 2016]. Foram selecionados dois conjuntos de dados. O primeiro (D1) é formado pelas variáveis consideradas na versão padrão do escore APACHE II: temperatura, pressão arterial média, frequência cardíaca, frequência respiratória, oxigenação (FiO2, PaO2, PaCO2), pH arterial, sódio sérico, potássio sérico, creatinina sérica, hematócrito, leucócitos, pontos na escala de coma de Glasgow e idade, exceto problemas crônicos de saúde. O segundo (D2) é formado pelas variáveis presentes no primeiro conjunto mais 11 outras variáveis: peso na admissão, glicose, SpO2, plaquetas, cloreto, hemoglobina, magnésio, pressão arterial sistólica, pressão arterial diastólica, CO2, pontos na escala de Braden. Em ambos os conjuntos a variável dependente (rótulo) considerada foi a mortalidade durante a internação. Para cada variável, exceto idade, foram considerados os valores mínimo, médio e máximo após a admissão na UTI, dos registros de internações que continham valores não nulos em todas variáveis (D1 com 43 variáveis e D2 com 64 variáveis).

A implementação ocorreu em 4 passos. Primeiro, foi implementado e testado o modelo de predição com base no escore APACHE II, utilizando o conjunto de dados D1. Segundo, foi implementada e treinada a versão 1 (V1) do modelo proposto utilizando o conjunto de dados D1. Para o treinamento da V1, cada entrada do conjunto de dados (43 variáveis) foi transformada em uma matriz 7x6 (Figura 4a), para tanto foi necessário descartar uma variável. No terceiro passo, com o objetivo de considerar todas as variáveis, a versão 2 (V2) foi treinada utilizando o mesmo conjunto de dados D1, onde cada entrada foi transformada em uma matriz 9x9, tendo as bordas da matriz sido preenchidas com valor “0” (Figura 4b). No quarto passo, com o objetivo de aumentar o desempenho do modelo, a versão 3 (V3) foi implementada e treinada utilizando o conjunto de dados D2 (64 variáveis), onde cada entrada foi transformada em uma matriz 10x10 e as bordas da matriz também foram preenchidas com valor “0” (Figura 4c).

De acordo com Pinheiro e Collobert (2014), um método comum para evitar a perda de dados é adicionar zeros no contorno da matriz de entrada, pois não altera a estrutura espacial da entrada e preserva o tamanho original da entrada (é como adicionar um fundo preto à uma imagem). Segundo os autores, aplicar a convolução a uma imagem grande é mais rápido do que aplicar a mesma convolução em pequenos blocos.

A fim de assegurar a capacidade de generalização, o modelo foi treinado e testado utilizando o método de validação Holdout [Angermueller et al. 2016]. O conjunto de dados D1, contendo 12.919 entradas, foi dividido em 7.751 registros para treino, 1.292 para validação e 3.876 para avaliação. Já o conjunto de dados D2, contendo 11.191 entradas, foi dividido em 6.715 registros para treino, 1.119 para validação e 3.858 para avaliação.

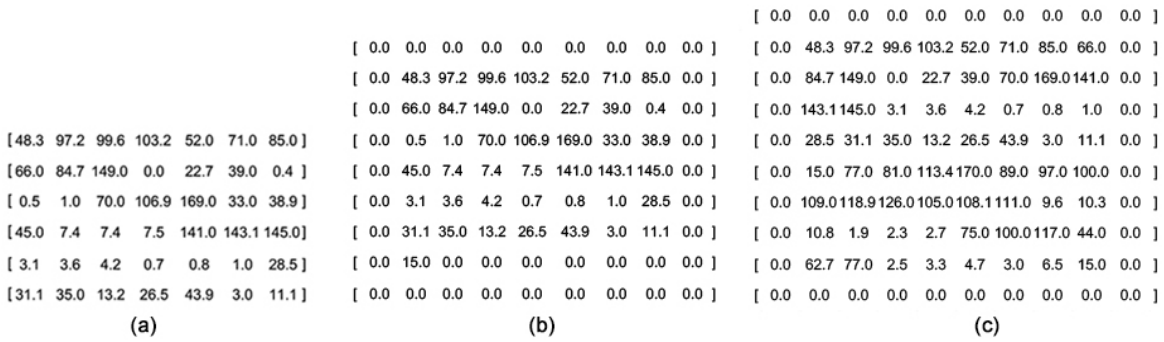

Figura 4 - Modelo de entrada dos dados em versão V1 (a), V2 (b) e V3 (c)

6. Resultados

Após a implementação dos algoritmos necessários para a construção e aprendizado do modelo proposto e do escore APACHE II, foram realizados os testes de validação e avaliação a fim de avaliar o poder preditivo do modelo quanto à mortalidade hospitalar. A seguir são apresentados os resultados obtidos em relação à acurácia, sensibilidade, especificidade e a Área sob a curva ROC (AUC) de cada versão (Tabela 1), assim como o gráfico da curva de Característica de Operação do Receptor (curva ROC) (Figura 5).

	APACHE II	V1	V2	V3
Conjunto de dados	**D1**	**D1**	**D1**	**D2**
Acurácia	68,80%	90,40%	**91,60%**	91,80%
Sensibilidade	**73,80%**	68,40%	69,70%	75,00%
Especificidade	67,90%	94,70%	**95,90%**	95,10%
AUC	0,71	0,82	**0,83**	0,85

Tabela 1 - Resultados obtidos

Utilizando o conjunto de dados D1, as versões V1 e V2 do modelo proposto apresentaram maior acurácia, especificidade e AUC (V1 = 90.4%, 94.7%, 0,82 e V2 = 91.6%, 95.9%, 0,83) em relação ao APACHE II (68.8%, 67.9%, 0,71), considerando o ponto de corte de 34 pontos (melhor resultado para o escore). Já a sensibilidade foi maior no APACHE II (73.8%). Utilizado o conjunto de dados D2, na versão V3, observou-se maior acurácia (91,8%), sensibilidade (75,0%) e AUC (0,85) em relação às versões V1 e V2.

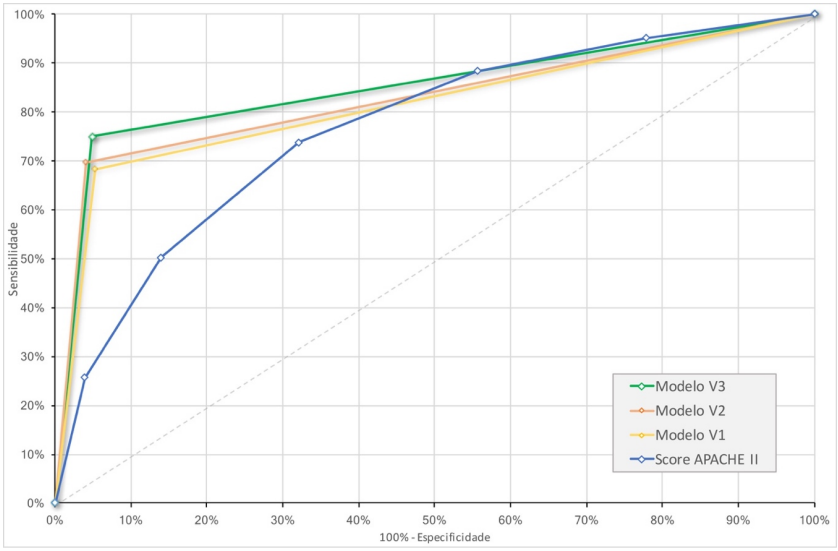

Figura 5 - Curva ROC demonstrando o desempenho dos modelos V1, V2 e V3 em comparação com o APACHE II

7. Discussão

Os resultados mostraram que, utilizando a mesma base de dados D1, o modelo proposto, versão V2, apresentou desempenho 17% maior que escore APACHE II, ao analisar a área sob a curva ROC, e acurácia 33% maior. Entretanto a sensibilidade foi 6% menor. Comparando as versões V1 e V2, foi possível observar desempenhos semelhantes, melhora de 1% na versão V2 em função da variável excluída em V1. Os resultados ilustram que o poder do modelo proposto (V1 e V2) de identificar casos positivos (sensibilidade) foi semelhante do escore APACHE II, mas apresentou capacidade superior de não predizer casos positivos equivocadamente (especificidade). Quando analisado o desempenho do modelo proposto, utilizando o conjunto de dados D2 (contendo 49% mais variáveis para cada entrada e 13% menos entradas, representando 29% mais dados) a versão V3 obteve desempenho superior em relação às demais versões, alcançando AUC de 0,85 e 75% de sensibilidade. Segundo Cardoso e Chiavone (2013), quando analisados, os índices prognósticos devem ter valores acima de 0,70 na área sob a curva ROC, sendo considerados bons, valores acima de 0,80 e excelentes acima de 0,90. Os autores também atribuem diferenças entre valores encontrados em diferentes estudos, ao perfil da base de dados e as características particulares de atendimento e protocolos de cada hospital.

Também foi possível observar a capacidade da RNC aprender com dados brutos, isto é, não houve a necessidade de realizar nenhum tratamento para correção de dados discrepantes ou incorretos. RNCs exploram o conhecimento da estrutura espacial da entrada, sendo tolerantes a pequenas distorções. Todavia, os autores apontam o tamanho e a qualidade do conjunto de treino como os principais fatores que alteram a qualidade de um sistema aprendido [Simard et al. 2003].

Como desvantagem do modelo proposto, é importante destacar que, segundo Eskofier et al. (2015) e Fang et al. (2016), na área da medicina, redes neurais são vistas como uma caixa-preta, isto é, a estrutura da rede, resultante do treinamento, não possui uma interpretação direta. De acordo com os autores, a falta de transparência do modelo

dificulta a obtenção de conhecimento a partir da estrutura da rede aprendida, em geral sendo avaliados somente os resultados das predições.

8. Conclusão

Técnicas de Deep Learning, combinadas com o fenômeno conhecido como Big Data, têm obtido importantes avanços na busca por padrões ocultos em grandes quantidades de dados. Nesse sentido, cada vez mais, pesquisas na área médica vêm utilizando Deep Learning para melhorar a qualidade dos sistemas de informação em saúde. Atualmente, diversos índices prognósticos convencionais são utilizados em UTIs para avaliar o risco de morte.

Nesse contexto, foram realizados estudos, pesquisas e implementações a fim de propor um modelo para aplicação da técnica de Deep Learning na predição de mortalidade em UTIs. Comparando o desempenho do modelo proposto, empregando RNCs, com o escore APACHE II, foi possível concluir que, apesar do estudo apresentar algumas limitações, o modelo apresentou bons resultados ao superar o APACHE II, com AUC acima de 0,80 em todas versões, demonstrando ser uma alternativa promissora na prática clínica.

Ainda que os resultados sejam animadores, é necessário abordar as limitações deste trabalho. Primeiro, a ausência da variável problemas crônicos de saúde pode ter afetado o poder preditivo do modelo. Isso ocorreu por essa informação não estar diretamente disponível na base de dados MIMIC-III. Em segundo lugar, o fato da base de dados provir de um único hospital poderia impactar na capacidade de generalização do modelo, embora os dados tenham sido coletados de diferentes UTIs. Por fim, as limitações dos recursos computacionais impediram a implementação de um modelo mais profundo e complexo, como a GoogLeNet que é composta por mais de cem camadas e contendo nove estágios de convolução [Szegedy et al. 2015].

Com base nos resultados e nas limitações expostas, é possível fazer algumas proposições para trabalhos futuros com o intuito de melhorar o desempenho da técnica de Deep Learning: aumentar a base de dados, expandir o conjunto de variáveis, validar externamente o modelo em diferentes populações, utilizar uma arquitetura mais profunda como a GoogLeNet, implementar a visualização gráfica dos dados, interpretar a estrutura da rede aprendida e a parceria com pesquisadores da área médica. Também é possível propor o emprego do modelo para a predições de doenças como diabetes, sepse, infarto e Alzheimer.

Agradecimentos

Os autores gostariam de agradecer ao CNPq pelo apoio a essa pesquisa.

Referências

Angermueller, C. et al. (2016). Deep learning for computational biology. Molecular systems biology, v. 12, n. 7, p. 878.

Bengio, Y., Courville, A., and Vicent, P. (2013). Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, v. 35, n. 8, p. 1798-1828.

Cardoso, L. G. dos S., and Chiavone, P. A. (2013). APACHE II medido na saída dos pacientes da Unidade de Terapia Intensiva na previsão da mortalidade. Revista Latino-Americana de Enfermagem, v. 21, n. 3, p. 811-819.

Deo, R. C. (2015). Machine learning in medicine. Circulation, v. 132, n. 20, p. 1920-1930.

Eskofier, B. M. et al. (2016). Recent Machine Learning Advancements in Sensor-Based Mobility Analysis: Deep Learning for Parkinson's Disease Assessment. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 38., Orlando. Conference Proceedings. p. 655-658.

Fang, R. et al. (2016). Computational health informatics in the big data age: a survey. ACM Computing Surveys (CSUR), v. 49, n. 1, p. 12.

Ghosh, E., Eshelman, L., Yang, L., Carlson, E., & Lord, B. (2017). Early Deterioration Indicator: Data-driven approach to detecting deterioration in general ward. Resuscitation, 122, 99-105.

Goldberger, A. L. et al. (2000). Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. Circulation, v. 101, n. 23, p. e215-e220.

Goodfellow, I. J., Benfio, Y., and Courville, A. (2016). Deep learning. Cambridge: MIT Press.

Herland, M., Khoshgoftaar, T. M., and Wald, R. (2014). A review of data mining using big data in health informatics. Journal of Big Data, v. 1, n. 1, p. 1.

Hissa, P., Hissa, M., and Araújo, P. (2013). Análise comparativa entre dois escores na previsão de mortalidade em unidade terapia intensiva. Revista Brasileira de Clínica Médica, v. 11, n. 1, p. 21-6.

Jarrett, K. et al. (2009). What is the best multi-stage architecture for object recognition? In: IEEE International Conference on Computer Vision, 12., Kyoto.

Johnson, A. E. W. et al. (2016). MIMIC-III, a freely accessible critical care database. Scientific data, v. 3.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature, v. 521, n. 7553, p. 436-444.

Marin, H. de F. (2010). Sistemas de informação em saúde: considerações gerais. Journal of Health Informatics, v. 2, n. 1.

Moritz, R. D., Schwingel, R. F., and Machado, F. O. (2005). Critérios prognósticos de pacientes graves: comparação entre a percepção dos médicos e o índice APACHE II. Revista Brasileira de Terapia Intensiva, v. 17, n. 3, p. 176-80.

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10). p. 807-814.

Neto, A. S. et al. (2015). Feasibility of transitioning from APACHE II to SAPS III as prognostic model in a Brazilian general intensive care unit. A retrospective study. Sao Paulo Medical Journal, São Paulo, v. 133, n. 3, p. 199-205.

Pinheiro, P. and Collobert, R. (2014). Recurrent convolutional neural networks for scene labeling. In: International Conference on Machine Learning.

Pirracchio, R. et al. (2015). Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. The Lancet Respiratory Medicine, v. 3, n. 1, p. 42-52.

Simard, P. Y., Steinkraus, D., and Platt, J. C. (2003). Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. In: International Conference on Document Analysis and Recognition (ICDAR), IEEE Computer Society, Los Alamitos, pp. 958-962.

Szegedy, C. et al. (2015). Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.