

Aprendizado de Máquina para Predição de Diagnósticos de Doenças Cardiovasculares

Francisco Romes da Silva Filho¹, Emanuel F. Coutinho¹

¹Programa de Pós-Graduação em Computação (PCOMP)
Universidade Federal do Ceará (UFC) – Quixadá – CE – Brasil

romesfilho.cc@gmail.com, emanuel.coutinho@ufc.br

Resumo. *Doenças Cardiovasculares (DCV) são adversidades que afetam o coração e vasos sanguíneos, sendo estimado em 2019 a causa de óbito de 17,9 milhões de pessoas. Devido a dados disponíveis em diversas formas, a análise de dados em informática médica ganhou importância, gerando interesse na geração de modelos analíticos orientados em Aprendizado de Máquina (ML). A predição de DCV é um desafio complexo na área de análise de dados clínicos, e a classificação com ML desempenharia um papel significativo na previsão de doenças cardíacas e investigação, para diminuir os impactos no coração e evitar uma morte prematura. O objetivo desse trabalho é treinar e avaliar modelos de aprendizado de máquina e aprendizado profundo para predição de diagnósticos de doenças cardiovasculares.*

Abstract. *Cardiovascular diseases (CVD) are adversities that affect the heart and blood vessels, being estimated in 2019 the cause of death of 17.9 million people. Due to data available in various forms, data analysis in medical informatics has gained importance, generating interest in producing analytical models oriented in Machine Learning (ML). CVD prediction is a complex challenge in the area of clinical data analysis, and classification with ML would play a significant role in heart disease prediction and research, to lessen impacts on the heart and prevent premature death. The objective of this work is to train and evaluate machine learning and deep learning models for predicting cardiovascular disease diagnoses.*

1. Introdução

As Doenças Cardiovasculares (DCV) são adversidades que afetam o coração e os vasos sanguíneos. Conhecidas também como doenças cardíacas, a incidência dessas patologias se dá principalmente por um bloqueio que estreita as artérias impedindo que o sangue flua para o coração ou para o cérebro [Heart 2021]. Estão inclusas várias enfermidades, como a doença isquêmica do coração, doenças cerebrovasculares, doença arterial periférica, doenças reumáticas e entre outras condições [PAHO 2021]. Em 2019, estima-se que as DCV tenham sido a causa de óbito de 17,9 milhões de pessoas, ou seja, 32% das mortes daquele ano. Ataques cardíacos e acidentes vasculares cerebrais geralmente são os eventos agudos mais comuns do conjunto, mais especificamente representaram 85% dos acontecimentos fatais [WHO 2021].

Os fatores de risco para DCV compreendem fatores comportamentais como tabagismo, alimentação não saudável, uso nocivo de álcool e atividade física insuficiente, e

fatores fisiológicos (metabólicos), entre os quais estão a hipertensão arterial e os níveis sanguíneos elevados de colesterol e glicose [WHO 2018]. Com a influência maciça de dados disponíveis em diversas formas, o papel da análise de dados em informática médica cresceu rapidamente na última década [Ravi et al. 2017]. Isso também gerou um interesse crescente na geração de modelos analíticos orientados por dados com base em Aprendizado de Máquina (ML - *Machine Learning*) em informática médica.

O ML é um campo de estudo da ciência da computação, especificamente um sub-campo da inteligência artificial. Envolve algoritmos de autoaprendizagem que extraem conhecimento de dados para fazer previsões, dando aos computadores a capacidade de aprender com os dados e não exigir que os humanos extraiam manualmente regras e construam modelos a partir de análise de grandes quantidades de dados. O ML oferece uma alternativa eficiente para capturar o conhecimento em dados para melhorar gradualmente o desempenho dos modelos preditivos e tomar decisões baseadas em dados [Raschka and Mirjalili 2019].

A predição de DCV é um dos desafios mais complicados na área de análise de dados clínicos. Contudo, a classificação usando ML desempenha um papel significativo na previsão de doenças cardíacas e na investigação de dados, para diminuir os impactos no coração e evitar uma possível morte prematura [Mohan et al. 2019]. As DCV podem vir a ser amenizadas por diagnósticos precoces, a fim de reduzir a taxa de mortalidade. Dessa forma, a identificação de fatores de riscos com o uso de ML é uma abordagem de grande potencial para mitigar a doença [Ghosh et al. 2021]. Diversas pesquisas desenvolvem modelos preditivos com o objetivo de realizar a predição de DCV. Por exemplo, em Al-Absi et al. (2021) os autores procuraram desenvolver um modelo para identificação de pessoas saudáveis e de pessoas com DCV, na base de dados do *Qatar Biobank* (QBB), possibilitando revelar e analisar os principais fatores de risco associados as doenças, no Catar. Outro exemplo é o estudo desenvolvido em Smigiel et al. (2021) que realizou classificações multiclasse de diversas classes de diagnósticos de DCV após a observação de eletrocardiogramas.

Dado o contexto, o presente trabalho tem como principal objetivo treinar e avaliar modelos de aprendizado de máquina e aprendizado profundo para predição de diagnósticos de doenças cardiovasculares. Como contribuições o trabalho realça o uso de redes neurais LSTM para séries temporais ao avaliar e comparar com outras abordagens ao realizar classificação multiclasse de diagnósticos de eletrocardiogramas (ECG).

Este artigo está organizado da seguinte forma: A Seção 2 apresenta alguns trabalhos relacionados à proposta deste trabalho. A Seção 3 apresenta a metodologia e suas etapas que foram definidas para guiar o presente trabalho. Os resultados e análises obtidos são apresentados, discutidos e comentados na Seção 4. Por fim, a Seção 5 comenta as considerações finais identificadas neste trabalho e apresenta os trabalhos futuros.

2. Trabalhos Relacionados

O diagnóstico precoce é uma etapa chave para alcançar o objetivo de reduzir o impacto e as consequências da DCV. Ghosh et al. (2021), similar a diversos estudos, procuraram classificar DCV com uso de ML. Bons resultados foram obtidos com a concepção de um modelo de acurácia 99,05% quando usado *Relief* para extrair os melhores atributos e *Random Forest Bagging Method* (RFBM) na classificação de doenças cardíacas.

Al-Absi et al. (2021) desenvolveram um modelo de ML para distinguir indivíduos saudáveis de pessoas com DCV, para revelar a lista de fatores de riscos em potencial associados à doença. O estudo considerou uma coleção de medidas biomédicas que representam as diferentes medidas biomédicas, inclusive comportamentais, do grupo DCV do *Qatar Biobank* (QBB). Como resultados, os autores apresentaram o *CatBoost* como o melhor modelo, alcançando uma acurácia de 93%. Outros resultados foram a identificação de uma nova lista de fatores de risco, além dos já conhecidos, como distúrbio renal, função hepática, aterosclerose, etc.

Cui et al. (2020) destacaram que altos valores de lipídios no sangue estão entre os principais fatores de risco de DCV. Portanto, com a previsão antecipada de anormalidade nos valores de lipídios no sangue, é possível uma intervenção precoce e diminuição de riscos de Dislipidemia que é o colesterol anormalmente elevado ou lipídeos no sangue [Cui et al. 2020]. Dessa forma, o trabalho teve o objetivo de prever o risco de Dislipidemia em trabalhadores de uma empresa siderúrgica. Como resultados, foi apresentada a rede neural LSTM como o melhor modelo, obtendo acurácia superior a 95%.

Strodthoff et al. (2020) introduziram um *benchmarking* para o conjunto de dados de ECG clínicos, o PTB-XL. Como contribuições, o trabalho além de apresentar uma base de dados para amenizar os problemas com a análise automática de ECG, implementa e aplica diferentes modelos de *Deep Learning* (DL) do estado da arte. Dentre todos os modelos avaliados, aqueles que alcançaram melhores desempenhos nas tarefas propostas foram as redes neurais convolucionais, com as arquiteturas *ResNet* e *Inception*.

O principal objetivo em Smigiel et al. (2021) foi realizar a classificação binária e multiclasse automática de diagnósticos de DCV utilizando modelos de DL. Para isso, o conjunto de dados de ECG clínicos PTB-XL foi escolhido. Esse trabalho propôs três abordagens na construção dos modelos de redes neurais, o uso de uma rede convulacional, uma segunda rede convulacional do tipo *SincNet* e uma terceira abordagem com uma rede convulacional, mas adicionando uma análise de atributos utilizando 13 atributos não lineares, conhecida como entropia. O trabalho tem como melhor classificador a rede convulacional que utiliza a entropia como estratégia adicional, obtendo acurácias de 89%, 77% e 70%, respectivamente nas classificações de 2, 5 e 20 classes.

Os trabalhos apresentados nesta seção estão alinhados com o presente trabalho ao realizarem tarefas de ML para prever eventos de saúde, especificamente DCV. Assim, como o presente trabalho, todos apresentam a tarefa de classificação. Porém, diferentemente desse trabalho, Ghosh et al. (2021) e Cui et al. (2020) realizam classificação do tipo binária, já o presente trabalho apresenta a classificação multiclasse, como os trabalhos Strodthoff et al. (2020) e Smigiel et al. (2021) que também compartilham da mesma base de dados que o presente trabalho, a PTB-XL. A Tabela 1 faz uma análise comparativa. Portanto, o presente trabalho apresenta como uma nova contribuição estudar diferentes modelos de ML e DL para classificação com a base PTB-XL.

3. Metodologia

3.1. Obtenção de base de dados

A primeira etapa do desenvolvimento da pesquisa consiste na busca e seleção de uma base de dados no domínio da saúde para doenças cardiovasculares, estruturada para um estudo

Tabela 1. Tabela comparativa de trabalhos

| Trabalho | Tarefa | Base de dados | Técnicas de predição |
|--------------------------|--|--|-------------------------------------|
| [Ghosh et al. 2021] | Classificação binária | Cleveland, Hungary, Switzerland, VA Long Beach e Statlog, do repositório UCI | DTBM, RFBM, KNNBM, ABBM e GBBM |
| [Al-Absi et al. 2021] | Classificação binária | Qatar Biobank | DT, ANN, RF, XGBoost, CatBoost e LR |
| [Cui et al. 2020] | Classificação binária | Base de dados contruída pela pesquisa | RNN tradicional e LSTM |
| [Strodthoff et al. 2020] | Classificação binária, multiclasse e regressão | PTB-XL | Redes convolucionais e LSTM |
| [Śmigiel et al. 2021] | Classificação binária e multiclasse | PTB-XL | Redes convolucionais |
| Este trabalho | Classificação multiclasse | PTB-XL | DT, RF, XGBoost e LSTM |

de aprendizagem supervisionada com o objetivo de realizar a classificação multiclasse das amostras.

3.2. Pré-processamento e análise de dados

O processo de um estudo em ML requer atividades que incluem pré-processamento e preparação dos dados, após obtenção do conjunto de dados. Objetivo dessa etapa é utilizar síntese estatística e técnicas de visualização para entender melhor os dados e identificar compreensões sobre tendências e a qualidade dos dados, bem como para formular hipóteses e fazer suposições nas análises. Na preparação dos dados, será realizada uma limpeza nos dados de forma a otimizar qualidade dos dados, procurando reduzir assim amostras com valores faltantes, dados com ruídos, dados discrepantes, etc. Além disso, pode-se transformar os dados originais em formatos mais apropriados e adequados para o processo. Outro passo relevante antes dos dados serem conhecidos pelos modelos de predição é a seleção de atributos. Nesse passo, os atributos mais úteis importantes são identificadas para a resolução do problema. E ainda sobre os atributos, é possível realizar engenharia de atributos também, com remoção, transformação e descoberta de novos através dos já fornecidas pelo *dataset*.

3.3. Treinamento e otimização dos modelos

O presente trabalho utilizará de algoritmos que se enquadrem como ML tradicional – possui um único modelo preditivo –, *ensemble learning* (EL) – combine múltiplos modelos preditivos – e redes neurais de DL. Para atender ao uso de ML tradicional, assim como foi apresentado anteriormente os algoritmos previamente escolhidos e que serão utilizados serão regressão logística e árvore de decisão ambos da biblioteca *Scikit-learn*¹, como os modelos mais simples. Os modelos mais robustos propostos, que usam EL, serão a floresta aleatória, da biblioteca *Scikit-learn*, como representante da técnica de *Bagging*, e *XGBoost*, da biblioteca *XGBoost*² como representantes da técnica de *Boosting*, especificamente *Gradient Boosting*. Por fim, para atender ao uso de DL para classificação

¹Scikit-learn - <https://scikit-learn.org/stable/>

²XGBoost - <https://xgboost.readthedocs.io/en/latest/>

no presente trabalho, será implementada a rede neural do tipo LSTM utilizando a biblioteca *TensorFlow*³. A biblioteca *Tensorflow* tem bons padrões de redes neurais e o funcionamento é eficiente sem que tenha que gastar muito tempo ajustando parâmetros manualmente [Chollet 2017]. Na fase de treino, os modelos conhecerão apenas os dados de treino. Utilizando a técnica de validação cruzada, *k-fold cross-validation*, os modelos serão submetidos a *k* diferentes partes do *dataset* de treino, para a realização do treino. Parte do conjunto de dados de treino também será separada servindo como um conjunto de validação. Aplicando assim o conceito de validação cruzada com um valor de *k* a ser escolhido durante os experimentos. Ainda nessa fase, será possível realizar o estudo dos hiperparâmetros dos modelos propostos, com a validação cruzada, portanto, as melhores configurações serão escolhidas para a construção dos modelos finais. O estudo com diferentes hiperparâmetros é a fim de encontrar o melhor ajuste que permita um bom desempenho. A otimização de hiperparâmetros é feita o informando a lista de hiperparâmetros e a sua lista de valores que serão testados, o valor de *k* para a validação cruzada e a definição de uma métrica, para avaliar durante o treino. Os melhores preditores posteriormente serão submetidos ao conjunto de teste.

3.4. Avaliação dos modelos

Os dados de testes serão submetidos aos melhores modelos para a predição. Para analisar de forma quantitativa os modelos finais a fim de realizar uma análise comparativa as métricas de classificação serão aplicadas para analisar os resultados da predição dos modelos. Os modelos serão avaliados em relação a Acurácia, Precisão, Revocação e *F1 Score*, assim como a maioria dos trabalhos citados na Seção 2. Em que os melhores modelos serão aqueles que apresentaram valores médios mais altos nas métricas que variam entre 0 e 1.

4. Resultados e Análises

4.1. Obtenção da base de dados

Neste trabalho, são utilizados dados da base de dados PTB-XL⁴, base de dados disponibilizada por um projeto do Physikalisch-Technische Bundesanstalt (PTB) que é o instituto nacional de metrologia da Alemanha. O banco de dados PTB-XL é um conjunto de dados de ECG clínico de tamanho sem precedentes, com alterações aplicadas para avaliar algoritmos de aprendizado de máquina. O conjunto de dados contém 21.837 ECGs clínicos de 12 derivações de 18.885 pacientes de 10s de duração, amostrados em 500 Hz e 100 Hz com resolução de 16 bits [Wagner et al. 2020]. O conjunto de dados é balanceado com relação ao sexo (52% masculino e 48% feminino) e cobre toda a faixa de idades de 0 a 95 anos (mediana de 62 e intervalo interquartil de 22). Os registros de ECG foram anotados por até dois cardiologistas com declarações de ECG potencialmente múltiplas de um conjunto de 71 declarações diferentes em conformidade com o padrão SCP-ECG. As declarações cobrem as declarações de forma, ritmo e diagnóstico de uma forma unificada e legível por máquina [Wagner et al. 2020]. Para os rótulos de diagnóstico, é fornecida uma organização hierárquica, em termos de 5 superclasses e 24 subclasses para os rótulos de diagnóstico [Wagner et al. 2020].

³TensorFlow - <https://www.tensorflow.org/>

⁴PTB-XL - <https://physionet.org/content/ptb-xl/1.0.1/>

4.2. Pré-processamentos dos dados

Na etapa de pré-processamento dos dados, é previsto na Seção de 3 a realização de procedimentos para o processamento de dados, como: eliminação de amostras com dados ausentes, remoção de dados discrepantes, normalização dos dados, seleção de atributos. Todos esses procedimentos tem o objetivo de maximizar a qualidade dos dados e a performance dos modelos de predição. Dessa forma, os procedimentos empregados por este trabalho foram os seguintes: (i) **Limpeza de dados**: Os dados originais do *dataset* podem conter muitas partes irrelevantes ou ausentes. Para lidar com esta situação, a limpeza de dados é essencial; e (ii) **Transformação de dados**: Essa etapa é executada para transformar os dados originais em formatos mais apropriados e adequados para o processo de treinamento dos modelos preditivos.

Para a realização desses procedimentos, utilizou-se o código exemplo oferecido pela *PhysioNet*⁵. O código de pré-processamento implementa um exemplo simples de como carregar os dados em forma de onda separando os dados (treino e teste) e os rótulos (treino e teste), propondo já uma divisão treino e teste.

Sobre os arquivos processados, há duas pastas com exames e dois arquivos com dados dos pacientes e os diagnósticos. A pasta *records100* detém os valores dos eletrocardiogramas com taxa de amostragem de 100Hz. De forma análoga, a pasta *records500* detém os valores dos eletrocardiogramas com taxa de amostragem de 500Hz.

O pré-processamento, inicialmente, é feito em dois arquivos do tipo csv. Inicialmente, é carregado o *ptbxl_database.csv* que possui um índice de todos os exames realizados, com informações demográficas do paciente, anotação do diagnóstico, forma e ritmo, dispositivo que foi realizado, enfermeira que o acompanha, entre outras informações médicas. A coluna *scp_codes* é essencial, pois ela apresenta as declarações informadas pelos médicos e suas probabilidades em formato de dicionário. A limpeza dos dados com a filtragem das amostras com diagnósticos informados por médicos reduziu as amostras de 21837 para 21430. O segundo arquivo é o *scp_statements.csv* que possui a descrição de todas as anotações de diagnóstico utilizadas pelos médicos, assim como anotações de forma e ritmo, e outros dados relacionados às anotações. As anotações estão organizadas em uma superclasse e uma subclasse, como forma de detalhar a hierarquia das anotações. Neste trabalho são observados os dados do tipo diagnóstico, superclasse e subclasse informadas. Para a classificação há 44 diagnósticos, 23 subclasses e 5 superclasses. Tendo em vista isso, o objetivo é avaliar os modelos preditivos nessas 3 situações.

4.2.1. Transformação de dados

Os dados utilizados nesse trabalho são os dados dos exames ECG armazenados nas pastas *records100* e *records500* que possui os dados brutos de ritmo e forma. O carregamento desses dados é feito associando aos ids dos pacientes do arquivo *ptbxl_database.csv*. Após o carregamento os dados têm a forma (21430, 1000, 12), ou seja, possuem 3 dimensões.

Para a realização do treinamento foi feita uma mudança na forma dos dados, alguns algoritmos e abordagens escolhidas, neste trabalho, exigem uma entrada na forma

⁵PhysioNet - <https://physionet.org/>

de matriz e não em 3 dimensões, como a série temporal, portanto, um tratamento adicional foi realizado. Isso inclui converter o conjunto de treinamento em uma matriz, e remodelar os dados de entrada para obtê-los na forma que o modelo espera receber posteriormente. Portanto, para aplicar aos modelos de ML clássicos foram feitas mudanças da forma (21430, 1000, 12) através da multiplicação da dimensão Y e Z , obtendo assim uma matriz com o formato de (21430, 12000).

Algumas redes neurais permitem trabalhar com dados em 3 dimensões. É o caso da rede LSTM que espera que os dados de entrada X sejam fornecidos com uma estrutura específica na forma de: (amostras, passos de tempo, características). Portanto, a forma original é mantida para os experimentos com LSTM [Chollet 2017].

As amostras são desbalanceadas referentes a quantidade de amostras por superclasses. Observando as subclasses e os diagnósticos, esse desbalanceamento é maior ainda. Como solução para esse problema foi aplicada a estratégia de *undersampling* utilizando a biblioteca *Imbalanced learn*⁶. Porém, essa estratégia não foi aplicada a rede neural, pois para usar essa solução é necessária informar uma entrada X na forma de matriz $M \times N$ e a LSTM espera receber uma entrada em um formato diferente como citado anteriormente.

Outra técnica utilizada para transformação dos dados foi a codificação *Multi-hot encoding* dos rótulos para a tarefa de predição dos diagnósticos. Tendo em vista que amostras poderiam ter múltiplos diagnósticos anotados, assim, o *One-hot encoding* não seria uma opção ideal [Li et al. 2022]. Por fim, baseado no atributo *scp_codes*, os diagnósticos com probabilidade informada como 0 foram retirados das amostras.

4.3. Treinamento e otimização dos modelos

Os algoritmos e abordagens procuram englobar algoritmos que utilizam um único preditor (árvore de decisão), algoritmos que combinam múltiplos preditores (floresta aleatória e o *XGBoost*) e, por fim, redes neurais (LSTM). A seguir é apresentada e discutida a arquitetura de rede neural proposta e a otimização hiperparâmetros dos modelos, para finalmente, realizar o treinamento, validação e teste dos modelos.

4.3.1. Arquitetura da rede neural LSTM

A rede possui uma camada visível com 1 entrada, uma camada oculta com a quantidade de blocos ou neurônios LSTM igual a quantidade de valores a serem previstos e uma camada de saída que faz a predição dos múltiplos valores de classes. A função de ativação sigmóide padrão é usada para os blocos LSTM, por conta da modelagem *Multi-hot encoding*, pois essa função de ativação tem o objetivo de prever a probabilidade de que um certo exemplo pertença a uma classe particular [Raschka and Mirjalili 2019].

4.3.2. Otimização de hiperparâmetros

O estudo dos hiperparâmetros é feito utilizando a validação cruzada informando a lista de hiperparâmetros para cada algoritmo (listados nos quadros a seguir), o valor de k definido

⁶Imbalanced learn - <https://imbalanced-learn.org/stable/>

como 5 e a métrica escolhida foi o *f1 score* para avaliação dos melhores parâmetros para determinado algoritmo na fase de treinamento.

Os hiperparâmetros e o conjunto de valores analisados para a árvore de decisão estão descritos na Tabela 2. O primeiro é o *criterion*, ou seja, o critério ou função para medir a qualidade de uma divisão da árvore, e, o segundo, é o *max_depth* que representa a profundidade máxima da árvore. Os valores são ‘*gini*’ que mede o grau de heterogeneidade dos dados e ‘*entropy*’ que calcula o ganho de informação. Após o processo de validação cruzada a melhor configuração de valores para a árvore de decisão foi ‘*entropy*’ para *criterion* e 10 para *max_depth*.

Tabela 2. Hiperparâmetros testados para árvore de decisão

| Hiperparâmetro | Conjunto de valores |
|------------------|--------------------------------------|
| <i>criterion</i> | ‘ <i>gini</i> ’ e ‘ <i>entropy</i> ’ |
| <i>max_depth</i> | 2, 5, 7, 10, 12 e 15 |

O parâmetro *max_depth* também foi analisado para o algoritmo floresta aleatória e com os mesmos valores como apresentado na Tabela 3. Além dele, foi analisado também o *n_estimators* que representa a quantidade de estimadores ou preditores, neste caso, árvores de decisão para a predição. Os resultados do processo foram o valor 100 para *n_estimators* e 10 para o *max_depth* como a melhor configuração.

Tabela 3. Hiperparâmetros testados para floresta aleatória

| Hiperparâmetro | Conjunto de valores |
|---------------------|----------------------|
| <i>n_estimators</i> | 100, 200 e 300 |
| <i>max_depth</i> | 2, 5, 7, 10, 12 e 15 |

No algoritmo de XGBoost é adicionado ao processo o hiperparâmetro *learning_rate* que significa a taxa de aprendizagem que o modelo usará para controlar o tamanho da etapa em cada iteração, por consequência, será responsável em dizer qual a distância que ele deverá percorrer em cada iteração. Os hiperparâmetros e o conjunto de valores para o XGBoost podem ser verificados na Tabela 4. Os experimentos indicaram o valor 10 para o *max_depth*, 100 para *n_estimators* e 0.1 para *learning_rate* como melhores.

Tabela 4. Hiperparâmetros testados para o XGBoost

| Hiperparâmetro | Conjunto de valores |
|----------------------|---------------------|
| <i>max_depth</i> | 3, 5, 7 e 10 |
| <i>n_estimators</i> | 100, 200 e 300 |
| <i>learning_rate</i> | 0.1, 0.01 e 0.05 |

A Tabela 5 apresenta os hiperparâmetros analisados na execução da LSTM. O *batch_size* é o tamanho do lote que define o número de amostras que serão propagadas pela rede e *epochs* descreve o número de vezes que o algoritmo vê o conjunto de dados. A melhor configuração apresenta o valor de 30 *epochs* e *batch_size* de 32.

Após essa etapa, todos os modelos construídos com as suas respectivas configurações apresentadas por essa subseção são expostos aos dados de treinamento e validação, para então, serem avaliados quanto a predição ao encontrar novos dados como o conjunto de teste.

Tabela 5. Hiperparâmetros testados para a rede neural LSTM

| Hiperparâmetro | Conjunto de valores |
|-------------------|---------------------|
| <i>epochs</i> | 10, 20 e 30 |
| <i>batch_size</i> | 16, 32 e 64 |

4.4. Avaliação dos modelos

Nessa etapa, os modelos são avaliadas pelas métricas informadas anteriormente, quando expostos ao conjunto de teste. Levando em conta o objetivo de realizar a classificação multiclasse, no cálculo das métricas precisão, revocação e *f1 score* para avaliação foram observados de forma que as métricas são calculadas para cada rótulo, obtendo a média ponderada pelo número de instâncias verdadeiras para cada rótulo.

4.4.1. Classificação de 5 classes

O objetivo dessa tarefa era prever a superclasse (normal, distúrbio de condução, STTC, infarto do miocárdio e hipertrofia) para cada amostra da base de dados. Como apresentado na Seção 4.2.1 os modelos são árvore de decisão, floresta aleatória e XGBoost foram expostos a amostras com a quantidade de rótulos balanceados e não balanceados. Dessa forma, esses modelos são analisados nessas duas situações.

Os modelos de aprendizado de máquina obtiveram rendimentos similares. O pior dentre os três foi a árvore de decisão como pode ser visto na Tabela 6. A floresta aleatória e o XGBoost foram levemente superiores como apresentado nas Tabelas 7 e 8. Com isso, podemos observar a superioridade de ter múltiplos preditores como é caso de floresta aleatória e o XGBoost ao aplicar o uso de *ensemble learning* em comparação a um único preditor como é o caso da árvore de decisão.

Tabela 6. Desempenho da árvore de decisão na classificação das 5 superclasses

| | Acurácia | Precisão | Revocação | F1 score |
|-------------------|----------|----------|-----------|----------|
| Sem balanceamento | 45% | 36% | 45% | 36% |
| Com balanceamento | 30% | 29% | 30% | 27% |

Tabela 7. Desempenho da floresta aleatória na classificação das 5 superclasses

| | Acurácia | Precisão | Revocação | F1 score |
|-------------------|----------|----------|-----------|----------|
| Sem balanceamento | 50% | 38% | 50% | 38% |
| Com balanceamento | 46% | 44% | 46% | 44% |

Tabela 8. Desempenho do XGBoost na classificação das 5 superclasses

| | Acurácia | Precisão | Revocação | F1 score |
|-------------------|----------|----------|-----------|----------|
| Sem balanceamento | 49% | 32% | 49% | 40% |
| Com balanceamento | 50% | 48% | 50% | 48% |

Os resultados apresentados nas métricas não são excelentes, mas é possível observar o impacto de diferentes estratégias para maximizar a qualidade dos resultados, como

utilizar *ensemble learning* e o balanceamento. A floresta aleatória e o XGBoost apresentaram valores superiores de precisão, revocação e *f1 score*, mostrando que o modelo melhorou sua capacidade de acertar diferentes classes e não apenas a classe majoritária presente na base de dados. Em contrapartida, ao balancear os dados utilizando *undersampling*, diminui-se a quantidade de amostras apresentadas a um modelo e juntamente a essa estratégia foi realizada a remodelação dos dados o que, possivelmente, ocasiona a perda de informações, o que afetou o modelo de árvore de decisão.

O melhor modelo para a classificação de 5 classes foi a rede neural LSTM. O modelo se apresenta bem superior em todas as métricas a todos os modelos, como observado na Tabela 9. Recorde que este modelo não conheceu dados balanceados por conta da sua natureza e a da estratégia proposta. Esses resultados apoiam que o uso de redes neurais do tipo LSTM são indicadas para séries temporais e para dados que tenham uma forma de 3 dimensões. Dessa forma, a LSTM se apresenta uma alternativa eficiente para a predição de séries temporais como a deste presente trabalho, necessitando pouco trabalho de pré-processamento.

Tabela 9. Desempenho da rede neural LSTM na classificação das 5 superclasses

| Acurácia | Precisão | Revocação | F1 score |
|----------|----------|-----------|----------|
| 77% | 43% | 68% | 52% |

Tendo em vista isso, apenas a rede neural foi utilizada na tarefa de predizer as 23 subclasses de diagnósticos e os 44 diagnósticos, levando em conta sua superioridade.

4.4.2. Classificação de 23 e 44 classes

Essa subseção apresenta os resultados para a classificação das 23 subclasses de diagnósticos e 44 diagnósticos. Os resultados apresentados pelas Tabelas 10 e 11 apresentam uma alta acurácia na predição de ambos os casos. O modelo obteve acurácia igual superior a 90% nos dois caso, o que intuitiva surpreende, observando a acurácia ao predizer menos classes.

Tabela 10. Desempenho da rede neural LSTM na classificação das 23 subclasses

| Acurácia | Precisão | Revocação | F1 score |
|----------|----------|-----------|----------|
| 94% | 21% | 70% | 30% |

A explicação para esses resultados se deve ao fato da coluna *scp_codes* presente nos metadados da base de dados apresentar os diagnósticos informados e as probabilidades da ocorrência de fato para a amostra em questão. Por consequência, todos os diagnósticos que apresentavam 0 como probabilidade foram retirados da amostra em questão, no pré-processamento dos dados. Entretanto, esse atributo dos metadados não inclui algumas superclasses o que ocasiona mais erros do que quando para prever subclasses de diagnósticos e os diagnósticos.

Em comparação com os trabalhos de Strodthoff et al. (2020) e Smigiel et al. (2021), o modelo proposto aqui é superior em acurácia. Strodthoff et al. (2020) utiliza outras métricas diferentes como auxiliares na avaliação não utilizadas por este trabalho, o

Tabela 11. Desempenho da rede neural LSTM na classificação dos 44 diagnósticos

| Acurácia | Precisão | Revocação | F1 score |
|----------|----------|-----------|----------|
| 95% | 35% | 51% | 39% |

que dificulta uma explícita comparação com ele. Entretanto, essa comparação é possível ser feita com Smigiel et al. (2021), conforme Tabela 12. Observa-se, então, que o presente trabalho é inferior na classificação de 5 classes em precisão, revocação e *f1 score*, porém, apresenta-se como ligeiramente melhor na classificação a partir de 20 classes.

Tabela 12. Comparação do desempenho dos melhores modelos de [Smigiel et al. 2021] e o presente trabalho

| Trabalho | Classes | Acurácia | Precisão | Revocação | <i>F1 Score</i> |
|-----------------------|---------|----------|----------|-----------|-----------------|
| [Smigiel et al. 2021] | 5 | 77% | 71% | 66% | 68% |
| | 20 | 70% | 35% | 33% | 33% |
| Este trabalho | 5 | 77% | 43% | 68% | 52% |
| | 23 | 94% | 21% | 70% | 30% |

Observando os resultados de precisão, revocação e *f1 score* a rede neural teve dificuldade para prever corretamente as classes minoritárias corretamente, o que pode representar um problema, no contexto desse trabalho, já que é o domínio dos dados é da área da saúde, especificamente doenças cardiovasculares, ou seja, o modelo erra ao classificar diversos diagnósticos corretamente. Principalmente, quando recordado que os dados possuem como classe majoritária o diagnóstico de ECG normal, ou seja, paciente sem doença.

5. Considerações Finais

Doenças cardiovasculares são a principal causa de mortalidade no mundo. Neste trabalho, foram propostos modelos de aprendizado de máquina para a classificação multiclasse de diagnósticos de doenças cardiovasculares, como forma de apoio a tomada de decisão no sentido de contribuir com a investigação dos dados e oferecer um direcionamento para a identificação de diagnósticos e, por consequência, ajudar na indicação de tratamentos. Foram selecionadas 3 abordagens na construção dos modelos de *machine learning* de classificação, sendo estas: uso de algoritmos tradicionais com um único preditor com a utilização da árvore de decisão; uso de *ensemble learning* para combinar múltiplos preditores para a classificação com a escolha da floresta aleatória e o XGBoost; e, por fim, redes neurais profundas, categorizando o uso de *deep learning* com a rede neural recorrente LSTM. Os experimentos revelaram superioridade do uso de aprendizado profundo com a rede LSTM como solução para o problema resultando em uma acurácia de 77% ao prever 5 superclasses de diagnósticos, e 94% e 95% ao classificar 23 subclasses e 44 diagnósticos. Como contribuições o trabalho realça o uso de redes neurais LSTM para séries temporais ao avaliar e comparar com outras abordagens.

Como trabalhos futuros, pretende-se avaliar diferentes estratégias de balanceamento, estudar a otimização de mais hiperparâmetros para os modelos e realizar um pré-processamento que especifique melhor as superclasses das amostras com base na probabilidade dos diagnósticos, como referido nos metadados dos dados. Além disso, estudar o uso de diferentes arquiteturas de redes neurais como as redes convolucionais que

são normalmente utilizadas para trabalhar com imagens e redes com arquiteturas do tipo *transformers* que tem desempenhado bem em diversos trabalhos do estado da arte.

Referências

- Al-Absi, H. R. H., Refaee, M. A., Rehman, A. U., Islam, M. T., Belhaouari, S. B., and Alam, T. (2021). Risk factors and comorbidities associated to cardiovascular disease in qatar: A machine learning based case-control study. *IEEE Access*, 9:29929–29941.
- Chollet, F. (2017). *Deep Learning with Python*. Manning.
- Cui, S., Li, C., Chen, Z., Wang, J., and Yuan, J. (2020). Research on risk prediction of dyslipidemia in steel workers based on recurrent neural network and lstm neural network. *IEEE Access*, 8:34153–34161.
- Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F. M. J. M., Ignatious, E., Shultana, S., Beeravolu, A. R., and De Boer, F. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques. *IEEE Access*, 9:19304–19326.
- Heart, A. H. A. (2021). What is cardiovascular disease? Disponível em: <https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-diseases>. Acesso em: 13 jun. 2021.
- Li, B., Tang, X., Qi, X., Chen, Y., Li, C.-G., and Xiao, R. (2022). Effective multi-hot encoding and classifier for lightweight scene text recognition with a large character set. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1.
- Mohan, S., Thirumalai, C., and Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 7:81542–81554.
- PAHO, P. A. H. O. (2021). Doenças cardiovasculares. <https://www.paho.org/pt/topicos/doencas-cardiovasculares>. Acesso em: 13 jun. 2021.
- Raschka, S. and Mirjalili, V. (2019). *Python Machine Learning*. Packt Publishing, Birmingham, UK, 3 edition.
- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., and Yang, G.-Z. (2017). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1):4–21.
- Strodthoff, N., Wagner, P., Schaeffter, T., and Samek, W. (2020). Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. *CoRR*, abs/2004.13701.
- Wagner, P., Strodthoff, N., Bousseljot, R., Kreiseler, D., Lunze, F. I., Samek, W., and Schaeffter, T. (2020). Ptb-xl, a large publicly available electrocardiography dataset. *Scientific Data*, 7.
- WHO, W. H. O. (2018). Technical package for cardiovascular disease management in primary health care: healthy-lifestyle counselling. Technical documents.
- WHO, W. H. O. (2021). Cardiovascular diseases (cvds). Disponível em: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). Acesso em: 13 jun. 2021.
- Śmigiel, S., Pałczyński, K., and Ledziński, D. (2021). Ecg signal classification using deep learning techniques based on the ptb-xl dataset. *Entropy*, 23(9).