

POLITECHNIKA WARSZAWSKA

DYSCYPLINA NAUKOWA INŻYNIERIA BIOMEDYCZNA  
DZIEDZINA NAUK INŻYNIERYJNO-TECHNICZNYCH

# **Rozprawa doktorska**

mgr inż. Maciej Rosoł

**Zastosowanie analizy sygnałów krążeniowo-oddechowych  
z rozszerzeniem o domenę przyczynową i informacyjną  
w celu poprawy oceny stanu zdrowia**

Promotor

dr hab. inż. Marcel Młyńczak, prof. uczelni

WARSZAWA 2025



## **Podziękowania**

*Pragnę złożyć serdeczne podziękowania:*

*dr hab. inż. Marcelowi Młynczakowi prof. uczelni – promotorowi tej rozprawy, jak również mojej pracy inżynierskiej i magisterskiej, koledze i mentorowi, za poświęcony czas i wsparcie w całym procesie powstawania niniejszej pracy. Dziękuję za inspirację oraz za zaproszenie na tę naukową ścieżkę.*

*dr n. o zdr. Jakubowi S. Gąsiorowi – koledze z Warszawskiego Uniwersytetu Medycznego, za całą okazaną pomoc w przeprowadzeniu prezentowanych w tej pracy badań i opublikowaniu ich wyników.*

*Wszystkim współautorom – za ich wkład w powstanie opisywanego w tej rozprawie cyklu publikacji.*

*Szczególne podziękowania dla mojej cudownej żony Kasi, za jej nieocenione wsparcie, wyrozumiałość i cierpliwość oraz motywację do osiągnięcia zamierzonych celów.*

*Podziękowania kieruję także do moich rodziców Lidii i Pawła, którzy od najmłodszych lat wspierali mnie na mojej drodze edukacji. Dziękuję za Waszą cierpliwość, zaangażowanie i płynący od Was przykład.*

*Dziękuję również dziadkowi Zygmuntowi oraz śp. babci Basi, za każde słowo i gest wsparcia w moim rozwoju naukowym.*

*Wszystkim przyjaciółom, którzy przez cały ten czas kibicowali mi w mojej karierze naukowej i zawodowej oraz pomagali mi znaleźć równowagę zapewniając towarzystwo i rozrywkę poza nią.*





## Streszczenie

W ramach niniejszej rozprawy doktorskiej zaprezentowany został cykl pięciu artykułów poświęconych weryfikacji możliwości zastosowania analizy sygnałów krążeniowo-oddechowych z rozszerzeniem o domenę przyczynową i informacyjną w celu poprawy oceny stanu zdrowia. W pierwszym kroku, w analizie przeprowadzonej na podstawie 369 zapisów z prób wysiłkowych, dokonano predykcji wartości  $VO_{2peak}$  dla różnych zestawów parametrów krążeniowo-oddechowych. Uzyskane wyniki wskazały na statystycznie istotną poprawę jakości predykcji przy uwzględnieniu parametrów oddechowych poza parametrami kardiologicznymi przy najniższej wartości średniego procentowego błędu bezwzględnego równej 10,51%. Następnie zaprezentowano utworzony w języku Python pakiet *nonlincausality*, służący do analizy przyczynowości z zastosowaniem modeli uczenia maszynowego. Dzięki temu narzędziu możliwe było sparametryzowanie nieliniowych zależności przyczynowych, które nie były wykrywane przy użyciu klasycznego podejścia metodą Grangera. W dalszej kolejności sprawdzono możliwość zastosowania opracowanego pakietu oraz innych dostępnych technik analizy przyczynowości na sygnałach krążeniowo-oddechowych w grupie 20 pediatrycznych pacjentów kardiologicznych, z czego każda metoda była w stanie wykryć oraz określić liczbowo współzależności między aktywnością układu krążenia i układu oddechowego. Kolejna praca przedstawia pilotażowe wykorzystanie szerokiego zestawu cech opisujących aktywność układu krążenia, parametrów oddechowych oraz z domeny przyczynowej i informacyjnej w połączeniu z metodami uczenia maszynowego do predykcji czasu trwania próby wysiłkowej na podstawie 5-minutowych statycznych pomiarów w pozycji leżącej w grupie 36 młodych piłkarzy. Uzyskane wyniki charakteryzowały się umiarkowaną dokładnością ze średnim procentowym błędem bezwzględnym wynoszącym 17,1% oraz średnim błędem bezwzględnym równym 129 sekund. Techniki wyjaśnialnej sztucznej inteligencji dostarczyły wglądu w poziom wpływu poszczególnych cech na wyniki modelu, wskazując na kluczowe znaczenie parametrów kardiologicznych, ale jednocześnie podkreślając potrzebę uwzględnienia informacji z sygnałów oddechowych oraz współzależności krążeniowo-oddechowych. Ostatnia z publikacji opisuje wykorzystanie metod uczenia maszynowego do klasyfikacji stanu zdrowia badanego na podstawie parametrów krążeniowo-oddechowych, w tym tych z domeny przyczynowej i informacyjnej. W tym badaniu wzięło udział 135 osób w wieku od 6 do 17 lat, przy czym każda z nich została przydzielona do jednej z trzech grup w zależności od stanu zdrowia. Dla zestawu danych składającego się z 35 najbardziej istotnych parametrów osiągnięto dokładność równą 89,1%. Uzyskane wyniki

wykazały statystycznie istotnie lepszą jakość klasyfikacji dla zbiorów danych zawierających parametry z domeny przyczynowej i informacyjnej, co wskazuje na zawartą w nich potencjalnie cenną informację diagnostyczną. Przeprowadzone analizy demonstrują praktyczny potencjał wykorzystania przedstawionych metod i narzędzi w diagnostyce medycznej. Zastosowanie parametrów krążeniowo-oddechowych, w tym cech z domeny przyczynowej i informacyjnej, może umożliwić bardziej precyzyjne monitorowanie stanu zdrowia, ocenę postępów treningowych oraz personalizację opieki zdrowotnej.

Słowa kluczowe: Sprężenie krążeniowo-oddechowe, analiza przyczynowości, uczenie maszynowe, parametry krążeniowo-oddechowe, pneumografia impedancyjna

## Abstract

This doctoral dissertation presents a series of five articles focused on the verification of the potential application of cardiorespiratory signal analysis, extended with causal and informational domains, to improve health status assessment. In the first step, an analysis based on 369 cardiopulmonary exercise test recordings was conducted to predict  $\text{VO}_{2\text{peak}}$  values with various sets of cardiorespiratory parameters. The obtained results indicated a statistically significant improvement in prediction quality when respiratory parameters were included alongside cardiac parameters, achieving the lowest mean absolute percentage error of 10.51%. Subsequently, the *nonlincausality* package, developed in Python for causality analysis using machine learning models, was introduced. This tool enabled the parameterization of nonlinear causal relationships that were undetectable using the classical Granger method. Next, the possibility of applying the developed package and other available causality analysis techniques to cardiorespiratory signals in a group of 20 pediatric cardiac patients was tested, with each method successfully identifying and numerically quantifying the interdependence between cardiovascular and respiratory system activity. Another study presented the pilot use of an extensive set of features describing cardiovascular activity, respiratory parameters, and features from causal and informational domains combined with machine learning methods to predict the duration of a cardiopulmonary exercise test based on 5-minute static measurements in a lying position in a group of 36 young soccer players. The obtained results were characterized by moderate accuracy, with a mean absolute percentage error of 17.1% and a mean absolute error of 129 seconds. Explainable artificial intelligence techniques provided insight into the influence of individual features on the model's outcomes, highlighting the critical importance of cardiac parameters while emphasizing the need to include information from respiratory signals and cardiorespiratory interdependencies. The final publication described the use of machine learning techniques to classify an individual's health status based on cardiorespiratory parameters, including those from causal and informational domains. This study involved 135 participants aged between 6 and 17 years, each assigned to one of three groups depending on their health status. For a dataset comprising the 35 most significant parameters, an accuracy of 89.1% was achieved. The results showed a statistically significant improvement in classification quality for datasets containing parameters from causal and informational domains, indicating their possibly valuable diagnostic information. The conducted analyses demonstrate the practical potential of utilizing the presented methods and tools in medical diagnostics. The application of cardiorespiratory parameters, including features from causal and

informational domains, could enable more precise health monitoring, training progress evaluation, and the personalization of healthcare.

Keywords: Cardiorespiratory coupling, causal analysis, machine learning, cardiorespiratory parameters, impedance pneumography

## Wykaz stosowanych skrótów

| <b>Skrót</b> | <b>Rozwinięcie</b>                                        | <b>Angielskie tłumaczenie</b>           |
|--------------|-----------------------------------------------------------|-----------------------------------------|
| AI           | Sztuczna inteligencja                                     | Artificial intelligence                 |
| ApEn         | Aproksymowana entropia                                    | Approximate entropy                     |
| AR           | Autoregresja liniowa                                      | Linear regression                       |
| AUC          | Pole pod krzywą charakterystyki operacyjnej odbiornika    | Area under curve                        |
| BMI          | Wskaźnik masy ciała                                       | Body mass index                         |
| CPET         | Test wysiłkowy                                            | Cardiopulmonary exercise test           |
| CRC          | Sprzężenie krążeniowo-oddechowe                           | Cardiorespiratory coupling              |
| CRF          | Wydolność krążeniowo-oddechowa                            | Cardiorespiratory fitness               |
| DFA          | Beztrendowa analiza fluktuacji                            | Detrended fluctuation analysis          |
| EKG          | Elektrokardiogram                                         | Electrocardiogram                       |
| GC           | Przyczynowość Grangera                                    | Granger causality                       |
| GRBF         | Uogólnione radialne funkcje bazowe                        | Generalized radial basis functions      |
| GRU          | Bramkowana jednostka rekurencyjna                         | Gated recurrent unit                    |
| HF           | Wysoka częstotliwość                                      | High frequency                          |
| HR           | Częstość akcji serca                                      | Heart rate                              |
| HRmax        | Maksymalna wartość częstości akcji serca                  | Maximal heart rate                      |
| HRV          | Zmienność rytmu zatokowego                                | Heart rate variability                  |
| IP           | Pneumografia impedancyjna                                 | Impedance pneumography                  |
| KGC          | Jądrowa przyczynowość Grangera                            | Kernel Granger causality                |
| LF           | Niska częstotliwość                                       | Low frequency                           |
| lsNGC        | Metoda wielkoskalowej nieliniowej przyczynowości Grangera | Large-scale nonlinear Granger causality |

|                     |                                                   |                                           |
|---------------------|---------------------------------------------------|-------------------------------------------|
| LLM                 | Duży model językowy                               | Large language model                      |
| LSTM                | Komórka długiej pamięci krótkotrwałej             | Long-short term memory                    |
| MAE                 | Średni błąd bezwzględny                           | Mean absolute error                       |
| MAPE                | Średni procentowy błąd bezwzględny                | Mean absolute percentage error            |
| MCC                 | Współczynnik korelacji Matthews                   | Matthews correlation coefficient          |
| MI                  | Informacja wzajemna                               | Mutual information                        |
| ML                  | Uczenie maszynowe                                 | Machine learning                          |
| MLP                 | Perceptron wielowarstwowy                         | Multilayer perceptron                     |
| RespRate            | Częstość oddechowa                                | Respiratory rate                          |
| RMSE                | Pierwiastek błędu średniokwadratowego             | Root mean square error                    |
| ROC                 | Krzywa charakterystyki operacyjnej odbiornika     | Receiver operating characteristic         |
| RRi                 | Interwały RR                                      | RR intervals                              |
| RSA                 | Niemiarowość oddechowa                            | Respiratory sinus arrhythmia              |
| SamEn               | Entropia próbek                                   | Sample entropy                            |
| SMOTE               | Technika nadpróbkowania mniejszości syntetycznych | Synthetic minority oversampling technique |
| TV                  | Objętość oddechowa                                | Tidal volume                              |
| VE                  | Minutowa wentylacja płuc                          | Ventilation                               |
| VLf                 | Bardzo niska częstotliwość                        | Very low frequency                        |
| VO <sub>2max</sub>  | Maksymalna wartość pochłaniania tlenu             | Maximal oxygen consumption                |
| VO <sub>2peak</sub> | Szczytowa wartość pochłaniania tlenu              | Peak oxygen consumption                   |
| XAI                 | Wyjaśnialna sztuczna inteligencja                 | Explainable AI                            |

## **SPIS TREŚCI**

|                                                                                                                                                                                       |           |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>1. Publikacje stanowiące cykl prac.....</b>                                                                                                                                        | <b>13</b> |
| <b>2. Wstęp.....</b>                                                                                                                                                                  | <b>14</b> |
| <b>3. Prediction of peak oxygen consumption using cardiorespiratory parameters from warmup and submaximal stage of treadmill cardiopulmonary exercise test.....</b>                   | <b>26</b> |
| 3.1. Materiały i metody.....                                                                                                                                                          | 26        |
| 3.2. Wyniki.....                                                                                                                                                                      | 28        |
| 3.3. Dyskusja .....                                                                                                                                                                   | 34        |
| <b>4. Granger causality test with nonlinear neural-network-based methods: Python package and simulation study .....</b>                                                               | <b>37</b> |
| 4.1. Materiały i metody.....                                                                                                                                                          | 37        |
| 4.2. Wyniki.....                                                                                                                                                                      | 39        |
| 4.3. Dyskusja .....                                                                                                                                                                   | 41        |
| <b>5. Causality in cardiorespiratory signals in pediatric cardiac patients.....</b>                                                                                                   | <b>43</b> |
| 5.1. Materiały i metody.....                                                                                                                                                          | 43        |
| 5.2. Wyniki.....                                                                                                                                                                      | 44        |
| 5.3. Dyskusja .....                                                                                                                                                                   | 47        |
| <b>6. Prediction of the duration of maximal exercise test in professional adolescent football players based on the cardiorespiratory signals – a pilot study .....</b>                | <b>49</b> |
| 6.1. Materiały i metody.....                                                                                                                                                          | 49        |
| 6.2. Wyniki.....                                                                                                                                                                      | 50        |
| 6.3. Dyskusja .....                                                                                                                                                                   | 52        |
| <b>7. Machine Learning Classification of Pediatric Health Status Based on Cardiorespiratory Signals with Causal and Information Domain Features Applied—An Exploratory Study.....</b> | <b>54</b> |
| 7.1. Materiały i metody.....                                                                                                                                                          | 54        |
| 7.2. Wyniki.....                                                                                                                                                                      | 57        |
| 7.3. Dyskusja .....                                                                                                                                                                   | 61        |
| <b>8. Podsumowanie.....</b>                                                                                                                                                           | <b>65</b> |
| <b>9. Bibliografia.....</b>                                                                                                                                                           | <b>69</b> |
| <b>10. Opinia Komisji Bioetycznej.....</b>                                                                                                                                            | <b>80</b> |

|                                                      |           |
|------------------------------------------------------|-----------|
| <i>11. Oświadczenia współautorów publikacji.....</i> | <i>81</i> |
| <i>12. Pozostałe publikacje.....</i>                 | <i>86</i> |
| <i>Załączniki.....</i>                               | <i>87</i> |



## 1. Publikacje stanowiące cykl prac

- **M. Rosol**, M. Petelczyc, J. S. Gąsior, and M. Młyńczak, “Prediction of peak oxygen consumption using cardiorespiratory parameters from warmup and submaximal stage of treadmill cardiopulmonary exercise test,” PLoS One, vol. 19, no. 1, p. e0291706, Jan. 2024, doi: 10.1371/journal.pone.0291706. (100 MNiSW, 2,9 IF)
- **M. Rosol**, M. Młyńczak, and G. Cybulski, “Granger causality test with nonlinear neural-network-based methods: Python package and simulation study.,” Comput Methods Programs Biomed, vol. 216, 2022, doi: 10.1016/j.cmpb.2022.106669. (100 MNiSW, 4,9 IF)
- **M. Rosol**, J. S. Gąsior, I. Walecka, B. Werner, G. Cybulski, and M. Młyńczak, “Causality in cardiorespiratory signals in pediatric cardiac patients,” in 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2022, pp. 355–358. doi: 10.1109/EMBC48229.2022.9871750. (20 MNiSW)
- **M. Rosol**, J. S. Gąsior, K. Korzeniewski, J. Łaba, R. Makuch, and M. Młyńczak, “Prediction of the duration of maximal exercise test in professional adolescent football players based on the cardiorespiratory signals – a pilot study,” in 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2024, pp. 1–4. doi: 10.1109/EMBC53108.2024.10782255. (20 MNiSW)
- **M. Rosol**, J. S. Gąsior, K. Korzeniewski, J. Łaba, R. Makuch, B. Werner and M. Młyńczak, “Machine Learning Classification of Pediatric Health Status Based on Cardiorespiratory Signals with Causal and Information Domain Features Applied—An Exploratory Study,” J Clin Med, vol. 13, no. 23, 2024, doi: 10.3390/jcm13237353. (140 MNiSW, 3,0 IF)

Pozostałe publikacje zostały wyszczególnione w Rozdziale 12.

## 2. Wstęp

Jednym z wyzwań stawianych współczesnej medycynie jest optymalizacja obciążeń fizycznych, którym pacjent lub sportowiec jest poddawany podczas rehabilitacji lub treningu w taki sposób, aby były one spersonalizowane i dostosowane do możliwości danej osoby [1], [2], [3]. Indywidualny dobór obciążeń zależy między innymi od wydolności krążeniowo-oddechowej (CRF, z ang. cardiorespiratory fitness), występowania przemęczenia w związku z uprzednim wysiłkiem fizycznym i występującymi problemami zdrowotnymi. W tym kontekście istotne jest monitorowanie parametrów fizjologicznych, które dostarczają cennych informacji diagnostycznych o obecnym stanie zdrowia danej osoby i możliwości adaptacji organizmu do obciążenia fizycznego. Kluczowe znaczenie ma tu zwłaszcza monitorowanie pracy układu krążenia i oddechowego, które ze względu na swoją dostępność (na przykład z wykorzystaniem urządzeń ubieralnych) i wartość diagnostyczną, stanowią podstawę oceny reakcji organizmu na obciążenia fizyczne. Ocena, jak układ sercowo-naczyniowy i oddechowy przystosowują się do wysiłku fizycznego, jest niezwykle istotna nie tylko dla pacjentów w trakcie rehabilitacji, ale także dla osób uprawiających sport, zarówno w czasie okresu treningowego, jak i zawodów. Wydaje się to niezbędne dla opracowania, dostosowania i personalizacji planów treningowych oraz czasu potrzebnego na regenerację.

Obecnie ocena funkcji krążeniowo-oddechowych jest możliwa między innymi na podstawie: parametrów kardiologicznych, takich jak częstość akcji serca (HR, z ang. heart rate) lub parametrów zmienności rytmu zatokowego (HRV, z ang. heart rate variability) [4], wyników testu wysiłkowego (CPET, z ang. cardiopulmonary exercise test) [5], badań laboratoryjnych [6] oraz technik obrazowania medycznego [7]. Jednakże metody te są albo kosztowne i mało dostępne, jak wykonanie pełnego testu wysiłkowego, gdzie konieczne jest zaangażowanie wykwalifikowanego personelu, sprzętu, na którym test mógłby być wykonany (jak bieżnia lub cykloergometr) oraz między innymi zaawansowanej aparatury do analizy gazów oddechowych, albo nie przedstawiają pełnego obrazu funkcji krążeniowo-oddechowych, jak w przypadku parametrów HRV, gdzie informacja o funkcji oddechowej jest uzyskiwana wyłącznie pośrednio. W związku z tym wciąż trwają badania nad nowymi podejściami, które mogłyby dostarczyć dodatkowych informacji na temat stanu zdrowia pacjenta, co pozwoliłoby na bardziej indywidualne podejście. Jedną z możliwości poprawy jakości oceny adaptacji do wysiłku fizycznego zdaje się być jednocześnie uwzględnienie informacji o układzie sercowo-naczyniowym oraz o układzie oddechowym, jak również uwzględnienie informacji

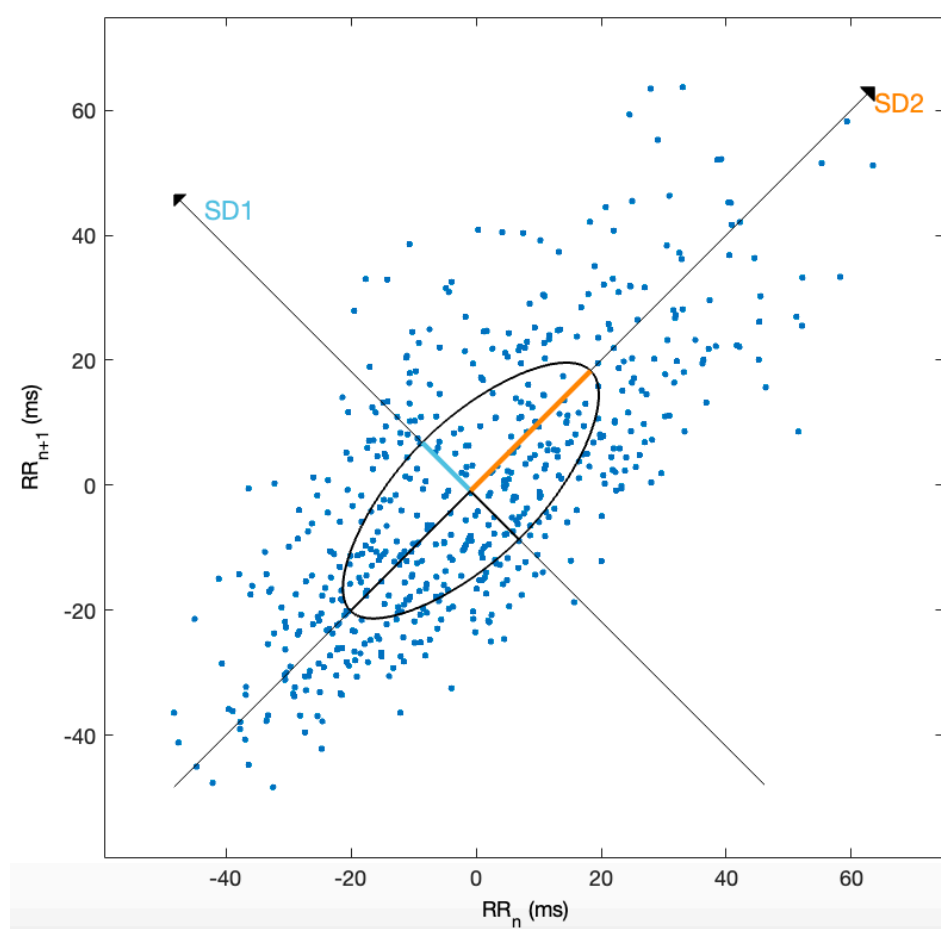
o współzależnościach krążeniowo-oddechowych opisywanych w domenie przyczynowej i informacyjnej.

W dzisiejszych czasach monitorowanie aktywności pracy serca jest możliwe dzięki popularyzacji inteligentnych zegarków, monitorów pracy serca oraz innych urządzeń ubieralnych. Urządzenia te pozwalają na wyznaczenie parametrów HRV opisujących pracę serca zarówno w domenie czasu, częstotliwości, jak i cech nieliniowych. Parametry te mogą zostać wyliczone przy użyciu takich narzędzi jak KubiosHRV [8] lub pakietu NeuroKit2 [9]. Do najpopularniejszych parametrów HRV w dziedzinie czasu należą: średnia wartość interwałów RR lub odpowiadająca jej średnia częstość rytmu serca, odchylenie standardowe interwałów RR, które odzwierciedla ogólną zmienność rytmu serca zarówno w krótkim, jak i długim okresie, pierwiastek kwadratowy ze średnich kwadratów różnic kolejnych interwałów RR, służący do oceny krótkoterminowej zmienności rytmu serca, a także NN50, określająca liczbę par sąsiadujących normalnych interwałów RR, których różnica przekracza 50 milisekund. Często wykorzystuje się również jej wariant procentowy, pNN50, który wyraża tę samą informację w formie względnej.

Dla parametrów w dziedzinie częstotliwości wyróżnia się trzy odrębne pasma: bardzo niskie częstotliwości (VLF, z ang. very low frequency), niskie częstotliwości (LF, z ang. low frequency) oraz wysokie częstotliwości (HF, z ang. high frequency). W przypadku krótkoterminowych pomiarów HRV u zdrowych osób, pasma te zazwyczaj definiuje się w następujących zakresach częstotliwości: 0–0,04 Hz dla VLF, 0,04–0,15 Hz dla LF oraz 0,15–0,4 Hz dla HF. Kluczowe parametry HRV wyprowadzane z tych pasm obejmują częstotliwość szczytową (częstotliwość, przy której występuje maksymalna moc w każdym z pasm), wartości mocy absolutnej i względnej (dla VLF, LF i HF), moce znormalizowane (szczególnie dla LF i HF), stosunek mocy LF/HF oraz całkowitą moc spektralną. Moce dla poszczególnych pasm oblicza się przez całkowanie oszacowań widma w granicach danego pasma częstotliwości, natomiast całkowitą moc uzyskuje się przez całkowanie w całym spektrum.

Nieliniowe metody parametryzacji HRV obejmują między innymi: wykres Poincaré, aproksymowaną entropię (ApEn, z ang. approximate entropy), entropię próbki (SampEn, z ang. sample entropy), wymiar korelacyjny  $D_2$  oraz beztrendową analizę fluktuacji (DFA, z ang. detrended fluctuation analysis). Wykres Poincaré przedstawia graficznie zależność między kolejnymi interwałami RR, wykreślając  $RR_{n+1}$  względem  $RR_n$ , gdzie  $n$  oznacza numer interwału w analizowanym zapisie. Z wykresu tego wyznaczane są dwa parametry: SD1, który

oznacza odchylenie standardowe wykresu Poincaré prostopadle do linii tożsamości, podczas gdy SD2 reprezentuje odchylenie standardowe wykresu Poincaré wzdłuż linii tożsamości [10]. SD1 odzwierciedla głównie krótkoterminową zmienność, natomiast SD2 opisuje zmienność długoterminową. Przykładowy wykres Poincaré z wizualną prezentacją parametrów SD1 i SD2 został zaprezentowany na Rycinie 1. Aby ocenić złożoność lub nieregularność HRV, stosuje się miary bazujące na teorii informacji takie jak ApEn i SampEn. Główną różnicą między tymi dwiema miarami jest to, że ApEn oblicza regularność na podstawie krótkich sekwencji czasowych, podczas gdy SampEn uwzględnia cały szereg czasowy [11]. DFA to technika służąca do oceny korelacji między interwałami RR na różnych skalach czasowych. W tej metodzie korelacje krótkoterminowe  $\alpha_1$  są oceniane na podstawie danych z segmentów obejmujących 4 do 16 uderzeń serca, natomiast  $\alpha_2$  opisuje korelacje długoterminowe w przedziale od 16 do 64 uderzeń [12]. Korelacje krótkoterminowe są związane z odruchem z baroreceptorów, natomiast korelacje długoterminowe odzwierciedlają szersze mechanizmy regulacyjne w układzie krążenia [13].



Rycina 1. Przykładowy wykres Poincaré wygenerowany przy użyciu programu KubiosHRV.

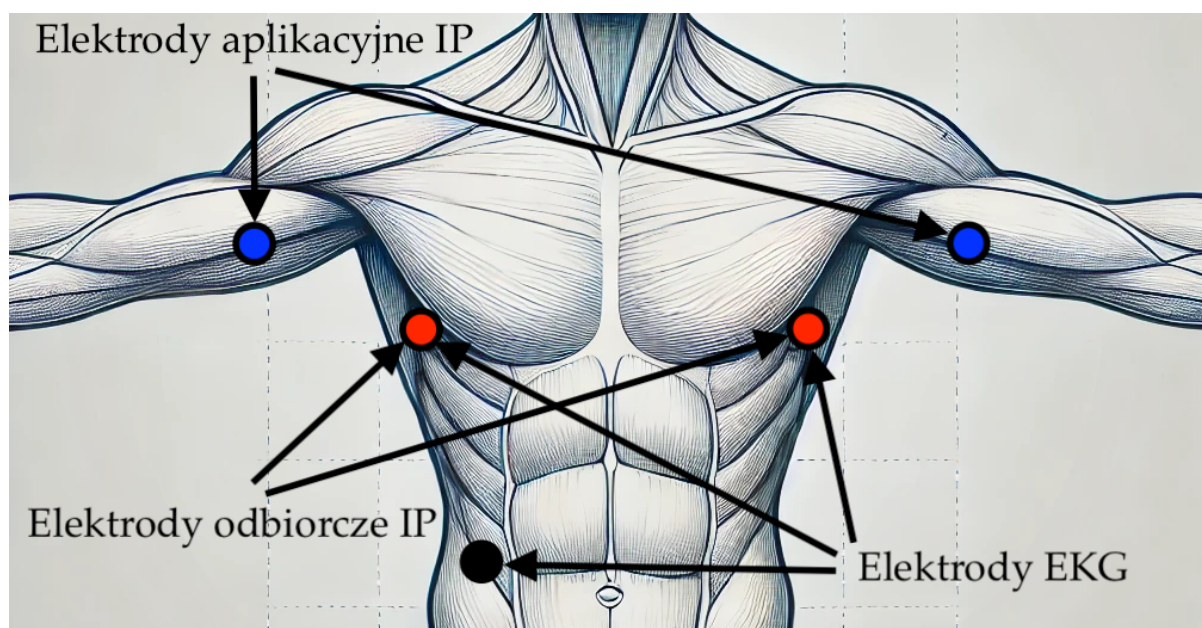
Parametry HRV są cennymi wskaźnikami śmiertelności u osób z niewydolnością serca [14], stanu zdrowia psychicznego [15], skutków zachorowania na COVID-19 w populacji osób starszych [16], przebiegu rehabilitacji po przebytym udarze [17], [18], wyników sportowych [19], a także stanu przetrenowania i stopnia regeneracji [20]. Jednakże poza parametrami krążeniowymi istotną rolę dla oceny szeroko pojętego stanu zdrowia odgrywają również parametry oddechowe, takie jak częstość oddechowa (RespRate, z ang. respiratory rate), objętość oddechowa (TV, z ang. tidal volume) i minutowa wentylacja płuc (VE, z ang. ventilaton), które są istotne w wykrywaniu i monitorowaniu takich schorzeń, jak przewlekła obturacyjna choroba płuc [21], [22] czy zapalenie płuc [23]. Monitorowanie aktywności oddechowej może być przeprowadzone przy użyciu różnych metod i urządzeń, jak spirometry, paski oddechowe oraz czujniki i urządzenia ubieralne. Każda z tych metod może być stosowana w zależności od potrzeb pacjenta, warunków i wymagań specyficznych dla monitorowania stanu zdrowia lub aktywności fizycznej. Istnieją również algorytmy pozwalające na wyznaczenie krzywej oddechowej na podstawie sygnału elektrokardiograficznego (EKG) [24], jednakże niektóre prace wskazują na ich niską użyteczność w wybranych zastosowaniach [25].

Jedną z metod, która pozwala na ciągłe monitorowanie aktywności oddechowej, a przy tym nie wymaga specjalistycznej aparatury do analizy gazów wydechowych oraz przylegającej maski, która wpływa na naturalny przepływ wdychanego i wydychanego powietrza, jest pneumografia impedancyjna (IP, z ang. impedance pneumography) [26], [27], [28], [29]. W tej technice, za pomocą elektrod aplikacyjnych, do tkanek osoby badanej aplikowany jest prąd zmienny o niskim natężeniu (poniżej progu pobudliwości komórek organizmu), natomiast przy użyciu elektrod odbiorczych rejestrowane jest napięcie będące odpowiedzią na zadany prąd. Na podstawie wartości zaaplikowanego prądu i zarejestrowanego napięcia wyznaczana jest wartość impedancji. Zastosowanie tetrapolarnej metody pomiaru i rozdzielenie elektrod aplikacyjnych od odbiorczych jest wskazane ze względu na zniwelowanie wpływu impedancji elektroda-skóra w tej metodzie. Zgodnie z [30], elektrody odbiorcze powinny być umieszczone w linii pachowej, w okolicy piątego i szóstego żebra, a elektrody aplikacyjne przymocowane do wewnętrznej strony ramienia na wysokości elektrod odbiorczych. Przy takim umiejscowieniu elektrod regresja liniowa zapewnia najlepszą zgodność pomiędzy IP a pneumotachometrią, dzięki czemu sygnał IP może być traktowany jako sygnał względnej objętości oddechowej [30]. Uzyskanie bezwzględnych wartości objętości oddechowej i przepływu powietrza z sygnału IP jest możliwe, jednak wymagana jest do tego odpowiednia kalibracja [31], [32], [33].

Jednym z urządzeń pozwalających na rejestrację sygnału IP jest opracowany na Politechnice Warszawskiej Pneumonitor [34], [35], [36]. Pneumonitor pozwala na jednoczesną rejestrację IP oraz EKG ze wspólną podstawą czasu obu sygnałów. Do pomiaru IP wykorzystywana jest konfiguracja tetrapolarna, w której aplikowany jest prąd sinusoidalny o częstotliwości 100 kHz i maksymalnej amplitudzie do 1 mA. Mierzona jest natomiast wartość napięcia będącego odpowiedzią na zadany sygnał prądowy, na podstawie której wyliczana jest wartość impedancji. Elektrody pomiarowe umieszczone są w linii pachowej na wysokości piątego i szóstego żebra, a elektrody aplikacyjne są przytwierdzone na ramieniu, na poziomie odpowiadającym elektrod pomiarowym. Do rejestracji używane są elektrody typu Holter Ag/AgCl, a sygnały próbkowane są z częstotliwością 250 Hz. Taka konfiguracja umożliwia precyzyjną ocenę zmian impedancji klatki piersiowej związanych z oddychaniem, co czyni Pneumonitor skutecznym narzędziem do monitorowania parametrów oddechowych. Urządzenie Pneumonitor przedstawiono na Rycinie 2, natomiast schemat podłączenia elektrod służących do pomiaru IP oraz EKG zaprezentowano na Rycinie 3.

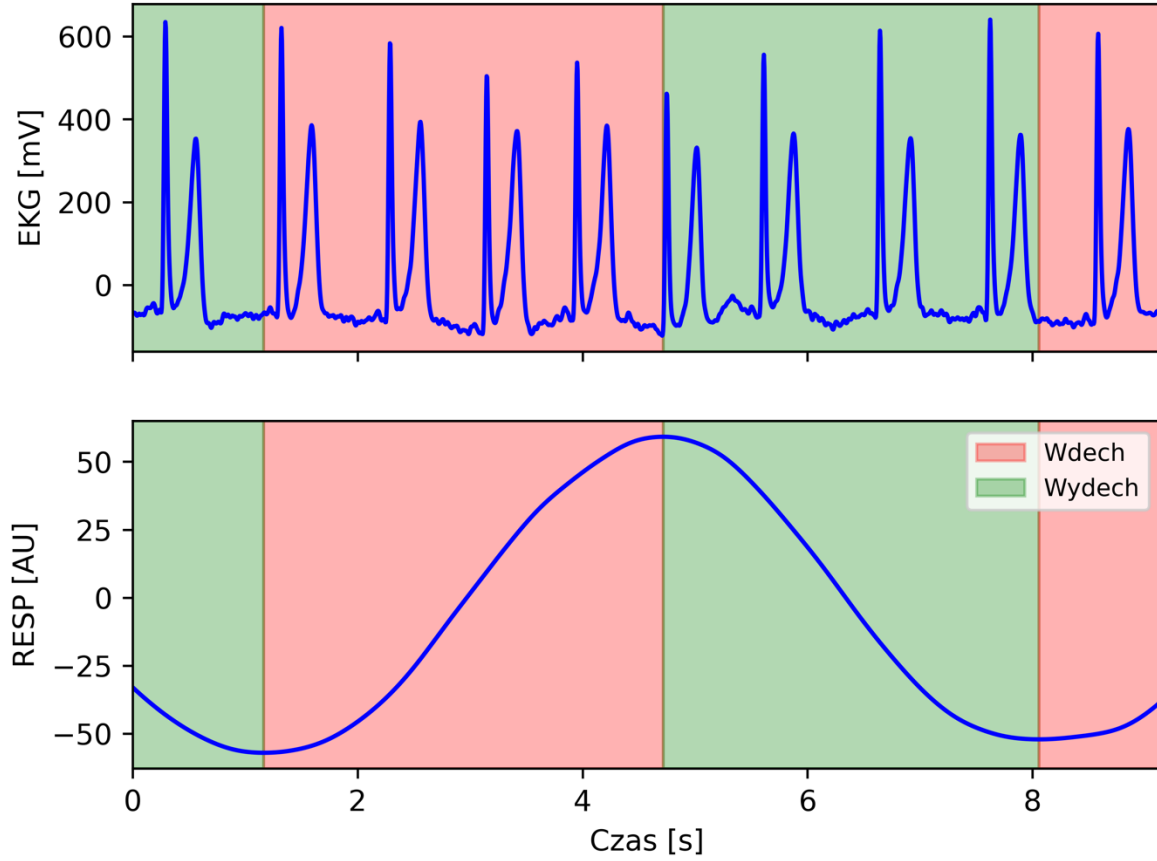


Rycina 2. Pneumonitor wersja 2.



Rycina 3. Schemat umieszczenia elektrod wykorzystywanych w urządzeniu Pneumonitor do pomiaru IP oraz EKG.

Coraz więcej prac wskazuje na nierozzerwalny charakter współzależności między układem sercowo-naczyniowym i układem oddechowym oraz na korzyści płynące z uwzględnienia parametrów oddechowych wraz z parametrami krążeniowymi [15], [37]. Związek ten określany jest mianem sprzężenia krążeniowo-oddechowego (CRC, z ang. cardiorespiratory coupling), u podstaw którego leży między innymi zjawisko niemiaryowości oddechowej (RSA, z ang. respiratory sinus arrhythmia). RSA przejawia się w skróceniu interwałów RR w trakcie fazy wdechu oraz ich wydłużeniu w trakcie wydechu, co przekłada się odpowiednio na chwilowy wzrost i spadek tętna. Zjawisko to jest związane z pracą autonomicznego układu nerwowego, przy czym zwiększenie aktywności układu współczulnego odpowiada za skrócenie średniego cyklu akcji serca (wzrost tętna), a zwiększenie aktywności układu przywspółczulnego wpływa na jego wydłużenie [38]. Przykład omawianego sprzężenia został zaprezentowany na Rycinie 4.



Rycina 4. Sygnał EKG oraz sygnał relatywnej zmiany objętości oddechowej z widocznym skróceniem interwałów RR podczas wdechu oraz ich wydłużeniem podczas wydechu.

Informację o CRC można uzyskać na podstawie parametrów HRV, w szczególności z domeny częstotliwościowej [13], jednakże pełen obraz tej zależności możliwy jest przy wykorzystaniu zarówno sygnału opisującego czynność układu krążenia, jak i oddechowego. Parametryzacja CRC może być wykonana na wiele sposobów [39], między innymi poprzez podejścia bazujące na przyczynowości Grangera (GC, z ang. Granger causality). Podejście Grangera do oceny zależności przyczynowych między szeregami czasowymi zostało zaprezentowane w 1969 roku [40]. Zakłada ono zbadanie przyczynowości od szeregu czasowego  $Y$  do szeregu czasowego  $X$  (oznaczonej jako  $Y \rightarrow X$ ) przy użyciu dwóch modeli autoregresji liniowej, zaprezentowanych w Równaniach (1-2).

$$X(t) = \sum_{l=1}^L a_l X(t-l) + E_X(t) \quad (1)$$

$$X(t) = \sum_{l=1}^L a_l X(t-l) + \sum_{l=1}^L b_l Y(t-l) + E_{X,Y}(t) \quad (2)$$



Gdzie  $L$  jest maksymalną wartością przeszłych obserwacji uwzględnionych w modelowaniu,  $a_l$  i  $b_l$  to współczynniki modeli (udział przeszłych wartości w predykcji obecnej wartości szeregu czasowego  $X$ ), a  $E_1$  oraz  $E_2$  to residua modeli. Aby ocenić, czy zmienna  $Y$  jest przyczyną zmiennej  $X$  w sensie Grangera, można wykorzystać test F [41], lub test chi-kwadrat [42]. W obu przypadkach należy wyliczyć sumę kwadratów residuów (RSS, z ang. residual sum of squares) obliczonych dla obu modeli autoregresji liniowej zgodnie z Równaniami (3-4), gdzie  $T$  jest liczbą przewidzianych wartości szeregu czasowego  $X$ .

$$RSS_X = \sum_{t=1}^T E_X(t)^2 \quad (3)$$

$$RSS_{X,Y} = \sum_{t=1}^T E_{X,Y}(t)^2 \quad (4)$$

Statystyka testowa  $S$  dla testu F przyjmuje postać zaprezentowaną w Równaniu (5), gdzie  $S$  pochodzi z rozkładu Fishera z  $L$  i  $T - 2L - 1$  stopniami swobody.

$$S = \frac{(RSS_X - RSS_{X,Y})/L}{RSS_{X,Y}/(T - 2L - 1)} \sim F_{L,T-2L-1} \quad (5)$$

Natomiast statystyka testowa  $S$  dla testu chi-kwadrat wyliczana jest zgodnie z Równaniem (6), gdzie  $S$  pochodzi z rozkładu chi-kwadrat z  $L$  stopniami swobody.

$$S = \frac{T(RSS_X - RSS_{X,Y})}{RSS_{X,Y}} \sim \chi_L^2 \quad (6)$$

Oprócz stwierdzenia występowania przyczynowości Grangera, taką współzależność między szeregami czasowymi można również określić liczbowo zgodnie z Równaniem (7), gdzie  $\sigma^2 E_X$  oraz  $\sigma^2 E_{X,Y}$  oznaczają odpowiednio wariancje residuów modelu opartego na przeszłych wartościach szeregu czasowego  $X$  oraz modelu opartego na przeszłych wartościach szeregów czasowych  $X$  i  $Y$ .

$$F_{Y \rightarrow X} = \ln \frac{\sigma^2 E_X}{\sigma^2 E_{X,Y}} \quad (7)$$

Podjęcie to ma jednak swoje ograniczenia ze względu na założenie o stacjonarności analizowanych szeregów czasowych oraz na brak możliwości wykrywania zależności nieliniowych. Dlatego też opracowywane są nowe metody analizy przyczynowości, które pozwalają na przezwycięzenie tych ograniczeń. Inną możliwością liczbowego określenia CRC jest zastosowanie metod, określanych jako nieliniowe. Na ten zbiór składają się metody oparte na technikach z teorii informacji [43], [44], [45], symbolizacji [46], [47], synchronizacji faz [48] lub innych metodach [39]. Przykładowo zależność między dwoma szeregami czasowymi może zostać sparametryzowana za pomocą informacji wzajemnej (MI, z ang. mutual information). Taka zależność dla dwóch dyskretnych zmiennych  $X$  i  $Y$  może być wyliczona na podstawie Równania (8) [39]:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log_2 \left( \frac{p(x,y)}{p(x)p(y)} \right) \quad (8)$$

gdzie  $p(x)$  i  $p(y)$  są rozkładami prawdopodobieństwa zmiennych  $X$  i  $Y$ , a  $p(x,y)$  jest rozkładem prawdopodobieństwa łącznego obu szeregów czasowych. Obecnie wciąż trwają prace nad poszukiwaniem nowych metod liczbowego określenia CRC, które pozwalałyby na bardziej precyzyjne zbadanie złożonych i dynamicznych interakcji między tymi układami, a także byłyby zoptymalizowane pod konkretne cele diagnostyczne [49], [50].

W związku z rozwojem narzędzi sztucznej inteligencji (AI, z ang. artificial intelligence) oraz algorytmów uczenia maszynowego (ML, z ang. machine learning), techniki te są coraz częściej i coraz chętniej wykorzystywane w zastosowaniach medycznych [51]. Rozwój ten obejmuje nie tylko systemy wspomagania decyzji klinicznych [52], [53] czy analizy sygnałów i obrazów medycznych [54], [55], [56], [57], ale również coraz bardziej zaawansowane narzędzia do przetwarzania języka naturalnego, które usprawniają pracę lekarzy [58], [59]. W ostatnim czasie nastąpił duży wzrost popularności dużych modeli językowych, które są w stanie przetwarzać język naturalny na poziomie pozwalającym na zaliczenie lekarskiego egzaminu końcowego [60]. Szybki postęp w szeroko pojętej dziedzinie AI wpływa na rozwój precyzyjnej i spersonalizowanej medycyny [61], [62], [63]. Pomimo tego, że modele uczenia maszynowego osiągają w różnych zadaniach poziom wydajności porównywalny z ludzkim, ich postrzeganie jako "czarnych skrzynek" znacznie utrudnia zrozumienie podstaw zwracanych przez nie wyników, co w rezultacie ogranicza ich szersze przyjęcie i zastosowanie w medycynie [64]. Aby rozwiązać ten problem, coraz większą popularność zyskują techniki tzw. „wyjaśnialnej” sztucznej inteligencji (XAI, z ang. explainable AI), które odgrywają kluczową rolę

w zwiększaniu interpretowalności i wiarygodności modeli uczenia maszynowego. Ich wykorzystanie pozwala na ograniczenie ryzyka związanego z implementacją narzędzi AI, umożliwiając integrację tych narzędzi w różnych dziedzinach w sposób zarówno efektywny, jak i etycznie odpowiedzialny [65]. Techniki ML znajdują zastosowanie również w analizie danych krążeniowo-oddechowych, np. w kardiologii, pulmonologii i medycynie sportowej. Użycie algorytmów uczenia maszynowego okazało się przydatne w przewidywaniu ryzyka choroby niedokrwiennej serca [66], klasyfikacji stopnia ograniczenia możliwego wysiłku [67], identyfikacji profili na podstawie testów wysiłkowych do oceny ryzyka wystąpienia zdarzeń sercowo-naczyniowych [68], wykrywaniu bezdechu centralnego u wcześniaków [69] oraz w wielu innych zastosowaniach [70].

Biorąc pod uwagę rosnący trend, polegający na bardziej holistycznym podejściu do ludzkiego organizmu, w tym do współpracy między systemem krążeniowym a oddechowym oraz rosnące możliwości analityczne dostarczane dzięki metodom uczenia maszynowego, na potrzeby niniejszej rozprawy zdefiniowano problem badawczy jako poszukiwanie sposobu zastosowania analizy sygnałów krążeniowo-oddechowych z rozszerzeniem o domenę przyczynową i informacyjną w celu poprawy oceny stanu zdrowia.

W ramach tej pracy postawiono hipotezę: **uwzględnienie informacji o aktywności oddechowej i parametrach z domeny przyczynowej lub informacyjnej dla sygnałów krążeniowo-oddechowych pozwala uzyskać dodatkową wartość diagnostyczną w ocenie stanu zdrowia.**

Celem cyklu prac było zbadanie istotności aktywności oddechowej i parametrów z domeny przyczynowej lub informacyjnej dla sygnałów krążeniowo-oddechowych w kontekście stanu zdrowia osoby badanej.

Powyższy cel został podjęty w publikacjach stanowiących cykl artykułów składający się z następujących pozycji:

1. M. Rosoł, M. Petelczyc, J. S. Gąsior, and M. Młyńczak, "Prediction of peak oxygen consumption using cardiorespiratory parameters from warmup and submaximal stage of treadmill cardiopulmonary exercise test," PLoS One, vol. 19, no. 1, p. e0291706, Jan. 2024, doi: 10.1371/journal.pone.0291706.

2. M. Rosoł, M. Młyńczak, and G. Cybulski, “Granger causality test with nonlinear neural-network-based methods: Python package and simulation study.,” *Comput Methods Programs Biomed*, vol. 216, 2022, doi: 10.1016/j.cmpb.2022.106669.
3. M. Rosoł, J. S. Gąsior, I. Walecka, B. Werner, G. Cybulski, and M. Młyńczak, “Causality in cardiorespiratory signals in pediatric cardiac patients,” in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022, pp. 355–358. doi: 10.1109/EMBC48229.2022.9871750.
4. M. Rosoł, J. S. Gąsior, K. Korzeniewski, J. Łaba, R. Makuch, and M. Młyńczak, “Prediction of the duration of maximal exercise test in professional adolescent football players based on the cardiorespiratory signals – a pilot study,” in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024, pp. 1–4. doi: 10.1109/EMBC53108.2024.10782255.
5. M. Rosoł, J. S. Gąsior, K. Korzeniewski, J. Łaba, R. Makuch, B. Werner, M. Młyńczak, “Machine Learning Classification of Pediatric Health Status Based on Cardiorespiratory Signals with Causal and Information Domain Features Applied—An Exploratory Study,” *J Clin Med*, vol. 13, no. 23, 2024, doi: 10.3390/jcm13237353.

W pierwszej z wymienionych prac [71] zaprezentowano predykcję wartości szczytowego pochłaniania tlenu ( $VO_{2peak}$ ) na podstawie różnych zestawów parametrów krążeniowo-oddechowych. Przedstawione w tej pracy wyniki wskazują na potrzebę uwzględniania informacji oddechowej w celu lepszego profilowania sportowców ze względu na statystycznie istotnie mniejsze błędy predykcji w sytuacji, gdy oprócz parametrów krążeniowych zostały uwzględnione parametry oddechowe. Zaproponowano również możliwość rozszerzenia zestawu parametrów o cechy z domeny przyczynowej i informacyjnej, w celu dalszej poprawy dokładności otrzymywanych predykcji. W pracy [72] opisano pakiet *nonlincausality* opracowany w języku Python, służący do parametryzacji związków przyczynowych przy użyciu nieliniowych technik modelowania opartych o uczenie maszynowe. W tej publikacji przedstawiono analizy symulacyjne, pokazujące możliwość wykrywania związków przyczynowych o złożoności większej niż liniowa w paradygmacie Grangera, za pomocą opracowanego pakietu używającego sieci neuronowych. Wskazano również na możliwość zastosowania zaproponowanych narzędzi do analizy współzależności krążeniowo-oddechowych. Kolejna praca [73] prezentuje zastosowanie opracowanego pakietu do sygnałów krążeniowo-oddechowych w grupie pacjentów z problemami kardiologicznymi. Tachogramy oraz krzywe oddechowe uzyskane z pięciominutowych rejestracji w spoczynku zostały

poddane analizie przyczynowości w grupie 20 pediatrycznych pacjentów kardiologicznych. Dokonano również porównania uzyskanych wyników parametryzacji CRC z użyciem różnych metod analizy przyczynowej. W publikacji [74] przedstawione zostały wyniki predykcji czasu trwania próby wysiłkowej wykonywanej przez młodych piłkarzy przy użyciu technik ML, na podstawie parametrów kardiologicznych, oddechowych oraz z domeny przyczynowej i informacyjnej, wyliczonych z 5-minutowych pomiarów pracy serca i układu oddechowego podczas spoczynku. Ostatnia praca z cyklu [75], ukazuje analizę klasyfikacji osób badanych ze względu na stan zdrowia, na podstawie parametrów krążeniowych, oddechowych oraz z domeny przyczynowej i informacyjnej. Wydzielone zostały 3 grupy osób badanych: dzieci z problemami kardiologicznymi, dzieci zdrowe oraz wytrenowani młodzi sportowcy. Sygnały krążeniowo-oddechowe były rejestrowane u każdej osoby przez przynajmniej 5 minut podczas spoczynku, a następnie dokonano ich parametryzacji. Analizie poddano 6 zestawów danych składających się z różnych kombinacji wyliczonych parametrów. Uzyskane wyniki wskazały na użyteczność parametrów kwantyfikujących CRC ze względu na znaczącą poprawę jakości klasyfikacji w sytuacji uwzględnienia tej grupy cech w analizie.

Podsumowując, prace stanowiące zaprezentowany cykl prezentują badania przeprowadzone nad istotnością aktywności oddechowej i parametrów z domeny przyczynowej lub informacyjnej dla sygnałów krążeniowo-oddechowych w kontekście stanu zdrowia osoby badanej, dążąc do opracowania nowych narzędzi i metod analitycznych, które umożliwią bardziej precyzyjną ocenę stanu zdrowia, wspomogą proces diagnostyki oraz przyczynią się do personalizacji opieki zdrowotnej.

### 3. Prediction of peak oxygen consumption using cardiorespiratory parameters from warmup and submaximal stage of treadmill cardiopulmonary exercise test

#### 3.1. Materiały i metody

W pracy poświęconej predykcji  $VO_{2peak}$  wykorzystano otwarty zbiór danych, opublikowany przez Mongina i innych [76], [77], który zawiera rejestracje parametrów krążeniowo-oddechowych wykonane między 2008 a 2018 rokiem w grupie amatorskich i profesjonalnych sportowców, podczas 992 maksymalnych prób wysiłkowych na bieżni. Badania te obejmowały dwa rodzaje protokołów: ciągły oraz skokowy wzrost prędkości bieżni. W analizie postanowiono ograniczyć się wyłącznie do eksperymentów z ciągłym wzrostem prędkości. Protokół badania obejmował: rozgrzewkę trwającą 8–10 minut przy prędkości 5 km/h, bieg na bieżni z prędkością wzrastającą o 1 km/h na minutę do momentu maksymalnego zmęczenia oraz fazę regeneracji po wysiłku przy prędkości 5 km/h. Do analizy wybrano uczestników w wieku od 18 do 40 lat, co zmniejszyło liczbę zapisów do 692. Wybrano tylko testy z ciągłym wzrostem prędkości, aby uzyskać spójne warunki badania, w wyniku czego pozostało 485 nagrań. Następnie wykluczono rejestracje osób, które zostały określone jako obserwacje odstające na pod względem wagi, wzrostu i  $VO_{2peak}$  w odniesieniu do danej płci, ograniczając próbkę do 462 nagrań. Ponadto dane zostały poddane wizualnej ocenie w celu odrzucenia pomiarów, podczas których wystąpiły widoczne artefakty w akwizycji HR (np. nagły spadek o ponad 30 uderzeń na minutę lub brak ciągłości szeregów czasowych HR podczas próby wysiłkowej, prawdopodobnie z powodu odłączenia elektrody). Ostatecznie do analizy zostało wykorzystanych 369 nagrań. Końcowe nagrania należały do 327 sportowców (42 osoby przeszły więcej niż jeden test), w tym 275 mężczyzn i 52 kobiet. Pełna charakterystyka grup jest przedstawiona w Tabeli 1.

Tabela 1. Charakterystyka grupy badanej z podziałem na płeć, przedstawiona jako wartość średnia  $\pm$  odchylenie standardowe oraz zakres w nawiasie.

| Płeć      | Wiek [lata]                   | Wzrost [cm]                      | Waga [kg]                     | BMI [kg/m <sup>2</sup> ]      | VO <sub>2peak</sub> [ml/min/kg] |
|-----------|-------------------------------|----------------------------------|-------------------------------|-------------------------------|---------------------------------|
| Mężczyźni | 27,3 $\pm$ 5,8<br>(18,0–39,8) | 177,4 $\pm$ 6,3<br>(160,5–193,0) | 76,6 $\pm$ 8,3<br>(55,3–97,0) | 24,3 $\pm$ 2,2<br>(17,9–31,4) | 47,7 $\pm$ 7,5<br>(28,9–67,3)   |
| Kobiety   | 26,9 $\pm$ 6,3<br>(18,0–40,0) | 165,2 $\pm$ 6,1<br>(154,0–178,0) | 62,2 $\pm$ 8,2<br>(46,0–83,0) | 22,8 $\pm$ 2,3<br>(18,0–29,6) | 38,1 $\pm$ 6,3<br>(24,8–53,8)   |
| Wszyscy   | 27,3 $\pm$ 5,9<br>(18–40)     | 175,5 $\pm$ 7,6<br>(154,0–193,0) | 74,5 $\pm$ 9,7<br>(46,0–97,0) | 24,1 $\pm$ 2,3<br>(17,9–31,3) | 46,3 $\pm$ 8,1<br>(24,8–67,3)   |

W oparciu o uzyskany zbiór danych, na podstawie parametrów krążeniowo-oddechowych z różnych etapów CPET, przy użyciu metod uczenia maszynowego zbadana została jakość predykcji wartości VO<sub>2peak</sub> oraz istotność cech opartych na danych oddechowych dla modelowania VO<sub>2peak</sub>. W tym celu wykorzystano zarejestrowane szeregi czasowe tętna (HR), częstotliwości oddechów (RespRate) oraz wentylacji minutowej (VE). Wartość VO<sub>2peak</sub> została określona jako maksymalna wartość poboru tlenu uzyskana po zastosowaniu okna średniej ruchomej na podstawie 15 oddechów, zgodnie z zaleceniami przedstawionymi przez Robergsa i innych [78].

Jako cechy dla modeli ML obliczono parametry takie jak: średnia, odchylenie standardowe, wartość maksymalna i minimalna, mediana, 25. i 75. kwantyl, skośność, kurtoza, współczynnik z regresji liniowej, współczynniki impulsu i kształtu dla HR, RespRate i VE dla różnych etapów podjętego wysiłku. Łącznie utworzono 11 zestawów danych (oznaczonych jako D1-D11) na podstawie różnych kombinacji parametrów i etapów CPET. Badania skupiły się na fazie submaksymalnym z testu wysiłkowego, trwającym od momentu zakończenia rozgrzewki do momentu uzyskania 85% maksymalnego tętna (HR<sub>max</sub>). Wykorzystano zarówno faktyczne HR<sub>max</sub> uzyskane podczas testu, jak i HR<sub>max</sub> wyliczone na podstawie wieku (220-wiek), aby dostarczyć informacji o użyteczności prognozowania VO<sub>2peak</sub> w testach submaksymalnych bez uprzedniej wiedzy o wartości HR<sub>max</sub> dla danego uczestnika.

Dokładność prognozy oceniano za pomocą 10-krotnej walidacji krzyżowej. W każdej iteracji, na podstawie średniej i odchylenia standardowego z zestawu danych treningowych, dokonywano standaryzacji cech niekategorycznych. Wykorzystano różne algorytmy ML powszechnie stosowane w problemach regresji. Dostrojenie hiperparametrów dla każdego

algorytmu przeprowadzono za pomocą techniki przeszukiwania siatki. W każdej iteracji walidacji obliczano metryki takie jak: średni procentowy błąd bezwzględny (MAPE, z ang. mean absolute percentage error), współczynnik determinacji  $R^2$ , średni błąd bezwzględny (MAE, z ang. mean absolute error), pierwiastek błędu średniokwadratowego (RMSE, z ang. root mean square error) oraz  $f^2$  Cohena do oceny wielkości efektu. Najlepszy model dla każdego zestawu danych określano na podstawie najniższego wyniku MAPE, uzyskanego z walidacji krzyżowej. Dla najlepszego modelu obliczano współczynnik korelacji zgodności  $L_{ina}$  oraz zwizualizowano zależność między przewidywanymi i rzeczywistymi wartościami  $VO_{2peak}$ , jak również utworzono wykres Blanda-Altmana. Ponadto sprawdzono różnice w wartościach uzyskanych metryk dla mężczyzn i kobiet.

Wyniki uzyskane dla wszystkich zestawów danych porównywano parami za pomocą testu Wilcoxona. Poziom istotności ustalono na 0,05. Do obliczeń użyto Pythona 3.9.13. Aby zbadać znaczenie poszczególnych cech wykorzystanych w modelowaniu ML, zastosowano narzędzia do wyjaśnialnej sztucznej inteligencji.

### 3.2. Wyniki

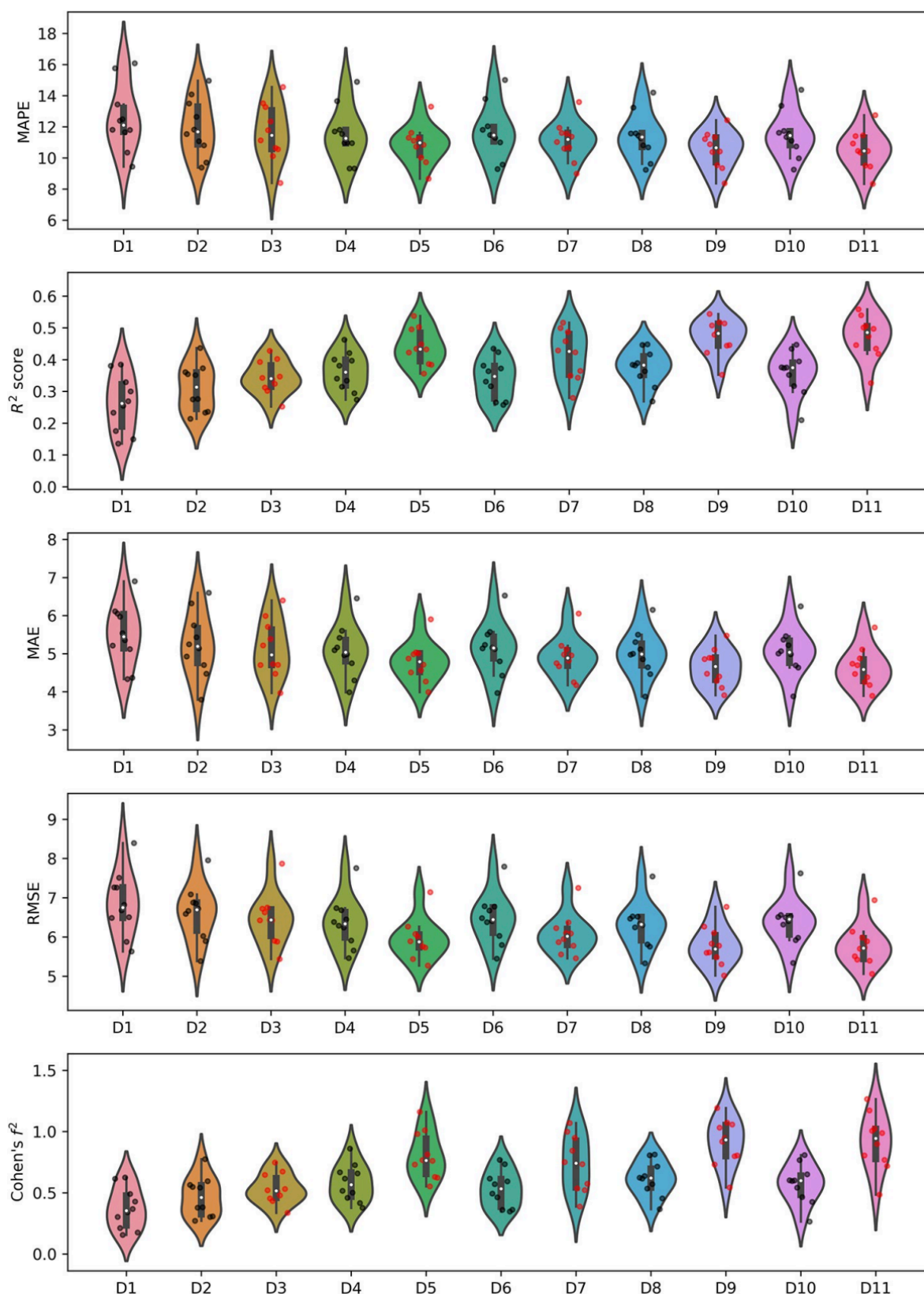
W ramach przeprowadzonej analizy statystycznej, oceniono skuteczność różnych algorytmów na podstawie najniższej wartości MAPE uzyskanego z walidacji krzyżowej dla każdego z zestawów danych. Wyniki te, wraz z nazwami modeli, przedstawiono w Tabeli 2. Wizualizację uzyskanych metryk dla każdego zestawu danych zaprezentowano na Rycinie 5, natomiast p-wartości uzyskane z testu Wilcoxona dla porównań metryk otrzymanych z różnych zestawów danych przedstawiono na Rycinie 6.

Tabela 2. Wartości metryk uzyskane dla poszczególnych zbiorów danych wraz z modelem, dla którego zostały otrzymane i który otrzymał najmniejszą wartość MAPE, przedstawione jako średnia  $\pm$  odchylenie standardowe z 10-krotnej walidacji krzyżowej. Najkorzystniejsze wartości metryk są wyróżnione pogrubieniem.

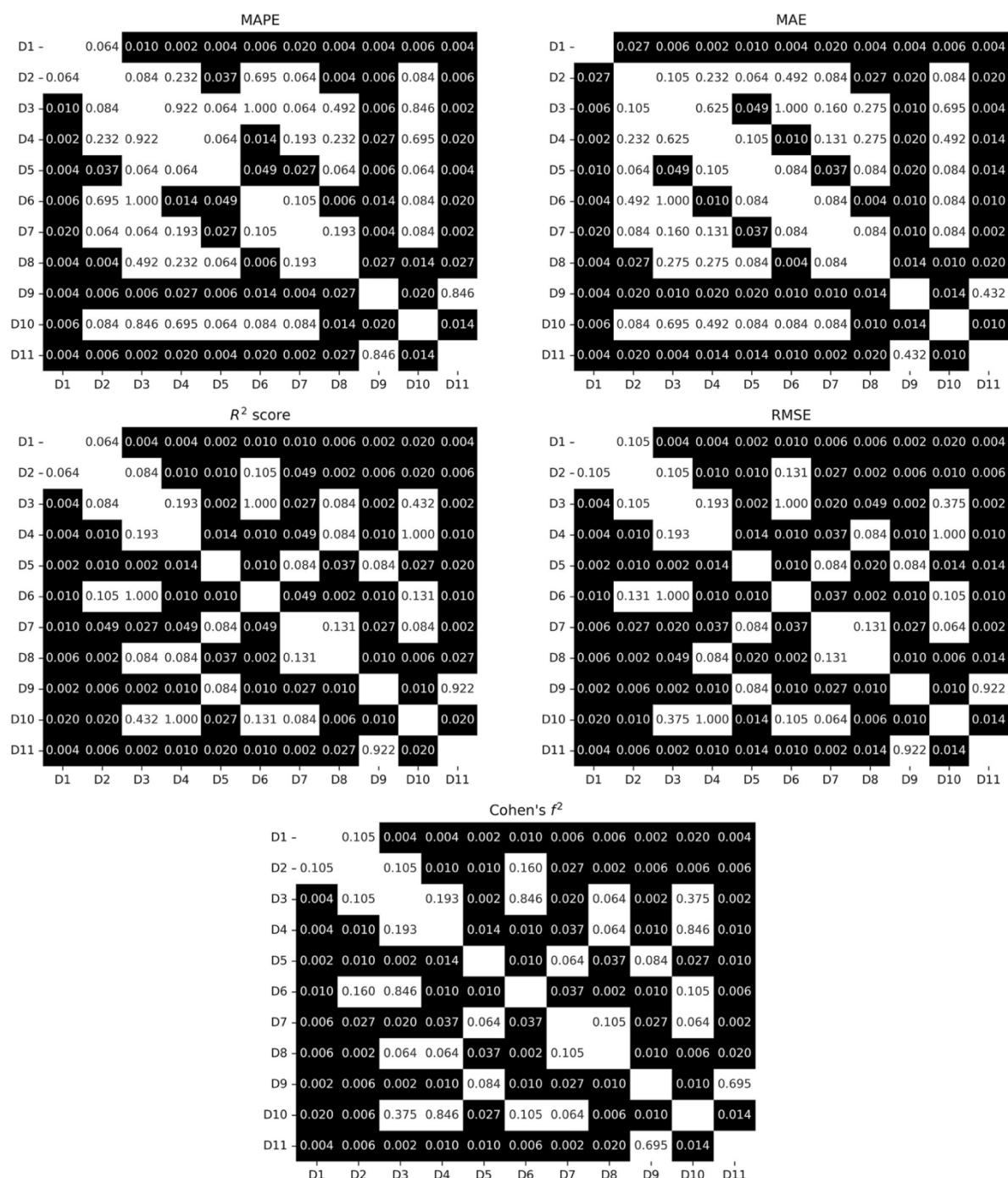
| Zestaw danych | MAPE [%]         | $R^2$           | MAE [ml/min/kg] | RMSE [ml/min/kg] | $f^2$ Cohena    | Rodzaj modelu    |
|---------------|------------------|-----------------|-----------------|------------------|-----------------|------------------|
| D1            | 12,52 $\pm$ 2,11 | 0,26 $\pm$ 0,09 | 5,50 $\pm$ 0,80 | 6,84 $\pm$ 0,81  | 0,37 $\pm$ 0,17 | Ridge regression |
| D2            | 11,95 $\pm$ 1,84 | 0,31 $\pm$ 0,07 | 5,24 $\pm$ 0,84 | 6,61 $\pm$ 0,71  | 0,47 $\pm$ 0,16 | Huber regression |



|     |                         |                        |                    |                    |                        |                               |
|-----|-------------------------|------------------------|--------------------|--------------------|------------------------|-------------------------------|
| D3  | 11,63 ±<br>1,84         | 0,34 ±<br>0,05         | 5,13 ± 0,75        | 6,45 ± 0,65        | 0,53 ±<br>0,12         | Bayesian<br>ARD<br>regression |
| D4  | 11,51 ±<br>1,72         | 0,36 ±<br>0,06         | 5,07 ± 0,68        | 6,36 ± 0,64        | 0,58 ±<br>0,15         | Bayesian<br>ARD<br>regression |
| D5  | 10,86 ±<br>1,23         | 0,44 ±<br>0,06         | 4,78 ± 0,52        | 5,95 ± 0,51        | 0,80 ±<br>0,19         | Bayesian<br>ARD<br>regression |
| D6  | 11,67 ±<br>1,72         | 0,34 ±<br>0,07         | 5,15 ± 0,69        | 6,46 ± 0,64        | 0,53 ±<br>0,15         | Lasso<br>regression           |
| D7  | 11,10 ±<br>1,26         | 0,42 ±<br>0,08         | 4,90 ± 0,53        | 6,07 ± 0,50        | 0,74 ±<br>0,23         | Bayesian<br>ARD<br>regression |
| D8  | 11,36 ±<br>1,49         | 0,38 ±<br>0,06         | 4,99 ± 0,61        | 6,29 ± 0,59        | 0,61 ±<br>0,14         | Bayesian<br>ARD<br>regression |
| D9  | 10,54 ±<br>1,20         | <b>0,47 ±<br/>0,06</b> | 4,64 ± 0,49        | <b>5,78 ± 0,50</b> | 0,91 ±<br>0,19         | Bayesian<br>ARD<br>regression |
| D10 | 11,50 ±<br>1,49         | 0,36 ±<br>0,07         | 5,06 ± 0,62        | 6,37 ± 0,59        | 0,57 ±<br>0,16         | Bayesian<br>ARD<br>regression |
| D11 | <b>10,51 ±<br/>1,24</b> | 0,47 ±<br>0,07         | <b>4,63 ± 0,52</b> | 5,78 ± 0,52        | <b>0,91 ±<br/>0,23</b> | Bayesian<br>ARD<br>regression |



Rycina 5. Wykresy skrzypcowe przedstawiające obliczone metryki dla każdego zestawu danych z wizualizacją metryk uzyskanych w każdej iteracji 10-krotnej walidacji krzyżowej. Czarne kropki reprezentują metryki uzyskane z zestawów danych niezawierających cech oddechowych, natomiast czerwone kropki reprezentują te, które owe cechy zawierają.

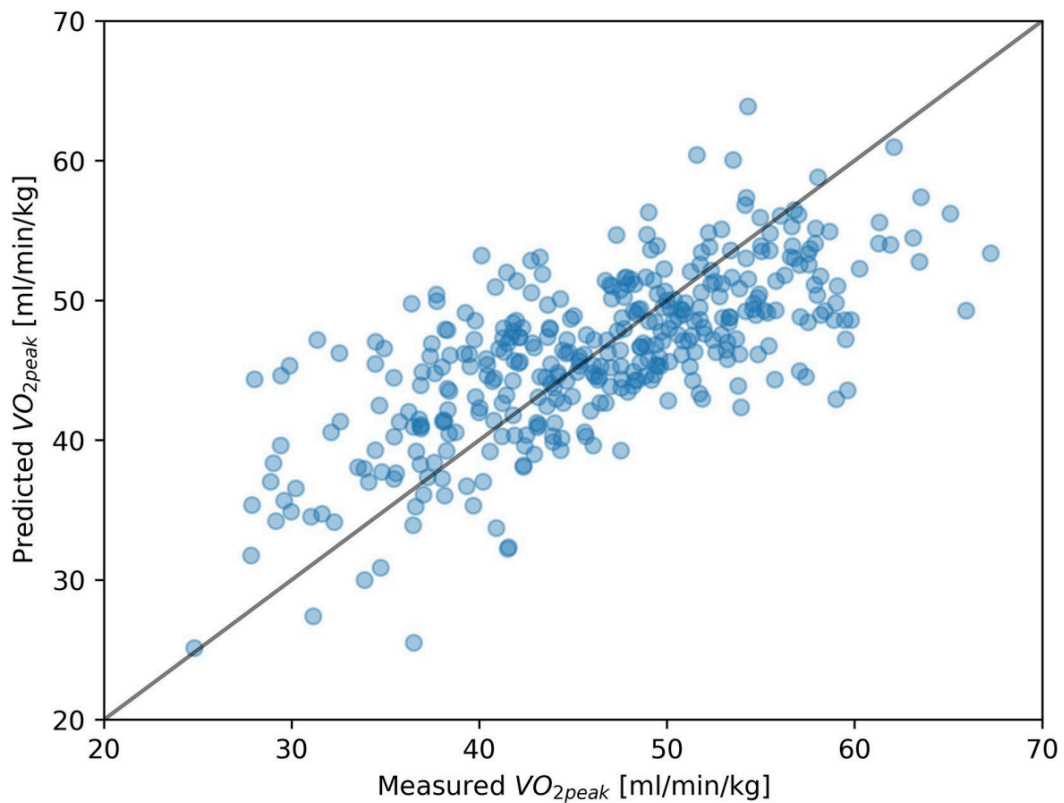


Rycina 6. P-wartości uzyskane z testu Wilcoxon dla porównań metryk otrzymanych z różnych zestawów danych. P-wartości mniejsze niż 0,05 oznaczono czarnym tłem.

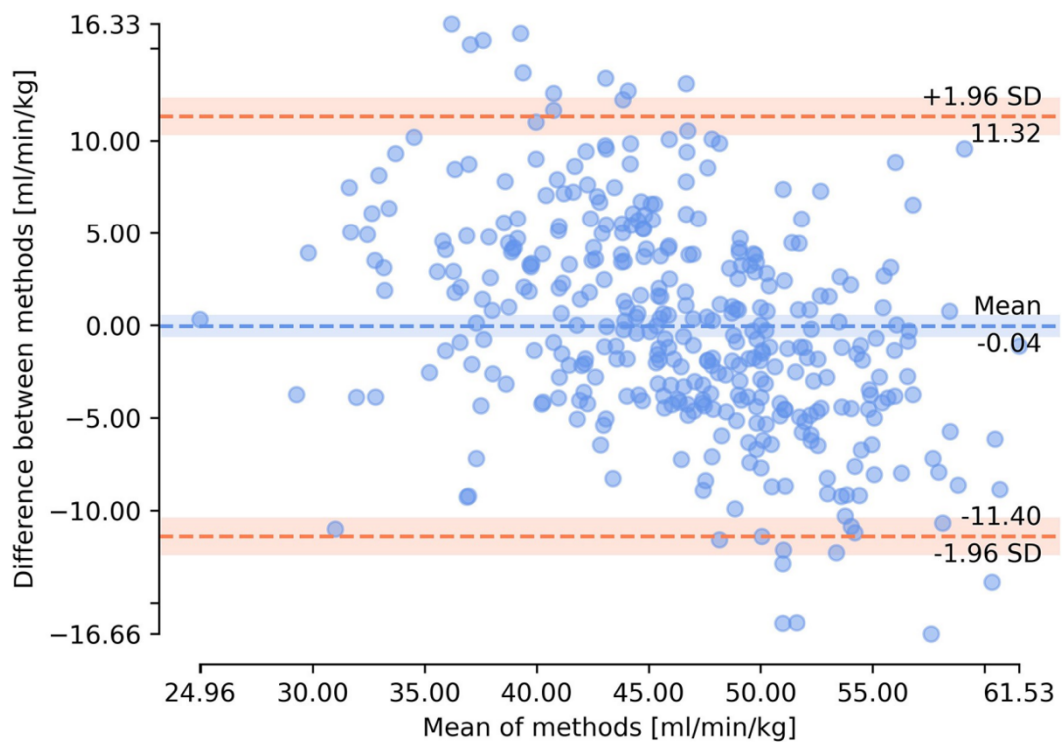
Najniższe wartości MAPE i MAE (odpowiednio 10,51% i 4,63) osiągnięto dla zestawu danych D11 (zestaw ten zawierał dane demograficzne wraz z cechami kardiologicznymi i oddechowymi z ostatnich 30 sekund rozgrzewki i testu CPET do 85% wartości  $HR_{max}$  wyliczonej na podstawie wieku), podczas gdy najniższy RMSE i najwyższą wartość współczynnika determinacji  $R^2$  (odpowiednio 5,78 i 0,47) uzyskano dla zestawu D9 (zestaw danych składał się z danych demograficznych wraz z cechami kardiologicznymi

i oddechowymi z testu CPET do 85% wartości  $HR_{max}$  wyliczonej na podstawie wieku). Najgorszą predykcję  $VO_{2peak}$  pod względem wszystkich metryk uzyskano, korzystając z zestawu danych D1 (wyłącznie dane demograficzne).

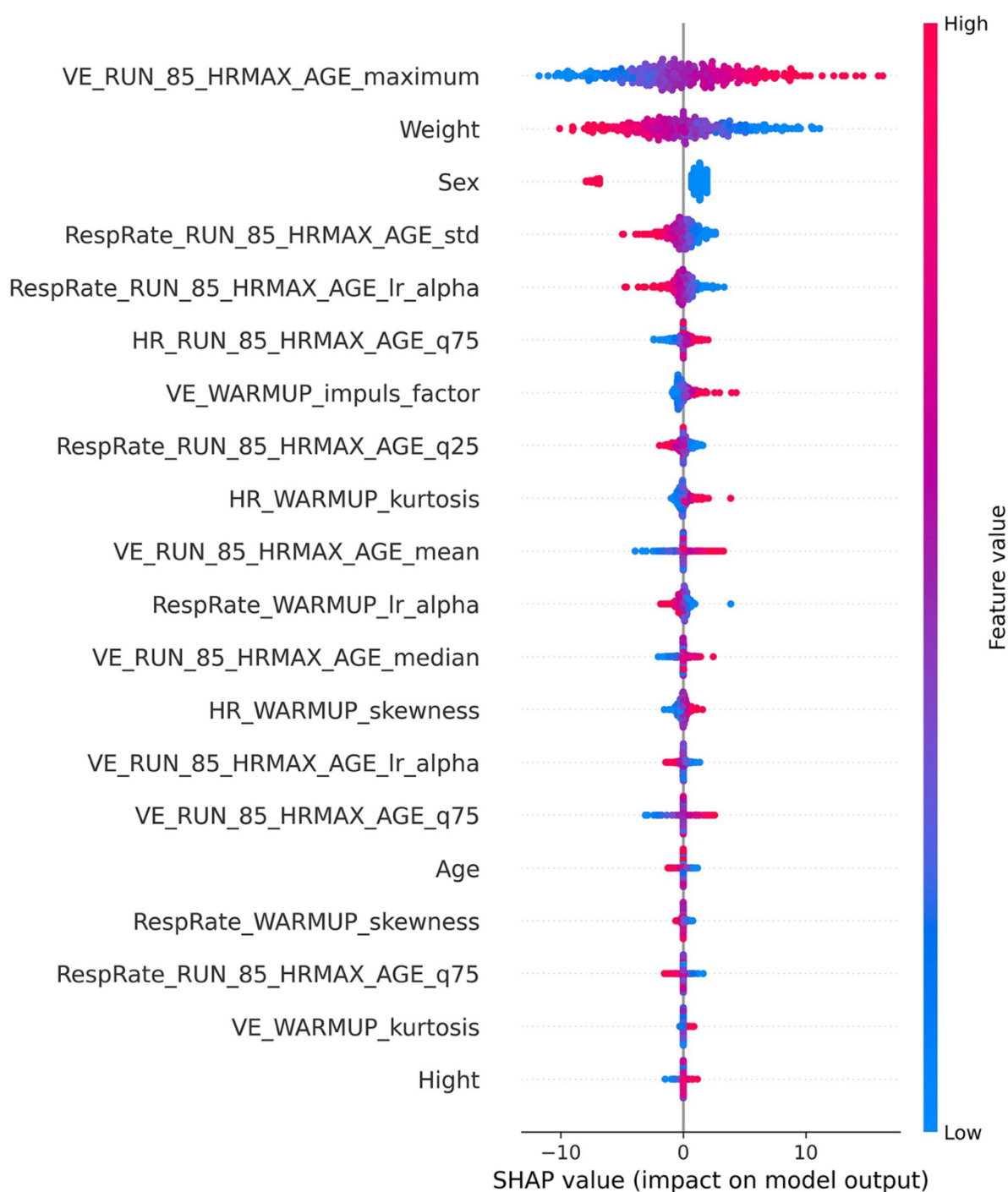
Wyniki uzyskane dla D11 były statystycznie znacząco lepsze pod względem wszystkich metryk w porównaniu z resztą zestawów danych, z wyjątkiem zestawu D9. Metryki  $R^2$  i RMSE dla zestawów danych zawierających cechy oparte o parametry oddechowe z submaksymalnej części CPET (niezależnie od sposobu wyznaczania  $HR_{max}$ ) były statystycznie istotnie lepsze od metryk uzyskanych dla zestawów danych, gdzie parametry były wyliczane z analogicznych okresów próby, ale nie zawierających cech oddechowych. Dla MAPE i MAE, zestawy danych zawierające cechy oddechowe obliczone do 85% maksymalnej wartości tętna wyliczonej na podstawie wieku wykazały znacząco lepsze metryki niż zestawy zawierające wyłącznie cechy kardiologiczne z tego okresu. Zmierzone wartości  $VO_{2peak}$  oraz wartości przewidziane dla zestawu danych, który uzyskał najniższy wynik MAPE (D11), zostały zwizualizowane na Rycinie 7. Współczynnik korelacji zgodności Lina między przewidywanymi a zmierzonymi wartościami  $VO_{2peak}$  wyniósł 0,66. Wykres Blanda-Altmana dla tego zestawu danych przedstawiono na Rycinie 8, natomiast wartości Shapleya na Rycinie 9. Nie stwierdzono statystycznie znaczących różnic w przypadku metryk uzyskanych dla uczestników płci męskiej i żeńskiej.



Rycina 7. Wykres przedstawiający rzeczywiste i przewidziane przez model wartości  $VO_{2peak}$  dla zestawu danych D11. Czarna ciągła linia reprezentuje funkcję  $y=x$ .



Rycina 8. Wykres Blanda-Altmana na podstawie wyników uzyskanych dla zestawu danych D11.



Rycina 9. Wartości Shapleya uzyskane dla zestawu danych D11. Opis poszczególnych cech znajduje się w Załączniku 1 do pracy [71].

### 3.3. Dyskusja

W badaniu predykcji  $VO_{2peak}$  najlepsze rezultaty pod względem MAPE uzyskano wykorzystując zbiór danych zawierający dane demograficzne, parametry kardiologiczne uzyskane z ostatnich 30 sekund rozgrzewki oraz z części testu wysiłkowego do momentu

uzyskania 85%  $HR_{max}$  wyliczonego na podstawie wieku oraz parametrów oddechowych z tych samych okresów. Dzięki włączeniu cech oddechowych do modelowania uzyskano w tym wypadku statystycznie istotną poprawę wyników. Zastosowanie progu 85%  $HR_{max}$  wyliczanego na podstawie wieku, potencjalnie umożliwia zastosowanie zaproponowanej metody w praktyce klinicznej, w związku z brakiem konieczności uprzedniego wykonania testu maksymalnego do precyzyjnego określenia  $HR_{max}$ . Wykonywanie testów submaksymalnych zamiast maksymalnych może być wskazane ze względu na występujące choroby układu krążenia, układu oddechowego lub układu mięśniowo-szkieletowego, a także ze względu na rygorystyczny program treningowy. Co więcej, wykorzystanie parametrów opartych na HR, RespRate i VE, pozwala na przeprowadzenie testu z wykorzystaniem aparatury prostszej względem tej standardowo stosowanej podczas testów wysiłkowych, chociażby ze względu na brak konieczności analizy gazów oddechowych. Uzyskane wyniki są porównywalne lub lepsze w stosunku do niektórych innych metod predykcji  $VO_{2peak}$  [79], [80], [81], jednakże w dostępnej literaturze znajdują się także techniki, które osiągnęły lepsze rezultaty [82], [83]. Warto zauważyć, że różnice w skuteczności poszczególnych metod mogą wynikać z wielu czynników, w tym ze specyfiki grupy badanej, dokładności użytych urządzeń czy też zastosowanego protokołu testowego. Zaproponowano również możliwość dalszej poprawy jakości modelowania poprzez uwzględnienie informacji o współzależnościach krążeniowo-oddechowych. Ponieważ niedostępne były surowe sygnały EKG/RRi oraz surowe krzywe oddechowe, niemożliwe było obliczenie bardziej zaawansowanych parametrów, szczególnie z domeny przyczynowej i informacyjnej, które pozwoliłyby na zweryfikowanie tej hipotezy. Innym ograniczeniem tego badania była wielkość próby, gdyż użyto tylko 369 rekordów z początkowej bazy danych zawierającej 992 nagrania, po zastosowaniu opisanych kryteriów wykluczenia i wizualnej inspekcji sygnałów. Co więcej, zestaw danych był niezrównoważony pod względem płci sportowców, gdyż zawierał rejestracje pochodzące od 275 mężczyzn i 52 kobiet. Większa i bardziej zrównoważona pod względem płci baza danych mogłaby pozwolić na uzyskanie predykcji obciążonych mniejszym błędem. Brakowało również informacji o ilości aktywności sportowej podejmowanej przez uczestników, co mogło wprowadzać niespójności w badanej populacji. Dodatkowo, równanie użyte do określenia  $HR_{max}$  na podstawie wieku (220-wiek) może być również traktowane jako ograniczenie, ponieważ istnieją równania z mniejszym błędem predykcji tej wartości. Jednak użyte w badaniu równanie jest najpopularniejsze i charakteryzuje się prostotą stosowania. Sprawdzone również wyłącznie jeden próg dla testów submaksymalnych, wynoszący 85%  $HR_{max}$ . Jest to zmienna, która również mogłaby zostać zoptymalizowana. Zaprezentowana metodologia mogłaby być

z łatwością zastosowana przy użyciu urządzenia Pneumonitor, ze względu na funkcjonalności jednoczesnej rejestracji sygnałów EKG i IP, co umożliwia wyliczenie cech wykorzystanych do trenowania modeli ML (po uprzedniej kalibracji na potrzeby wyznaczenia VE) oraz nie wymaga użycia zaawansowanej aparatury do analizy gazów i ciasno przylegającej maski, która zaburza naturalny przepływ gazów oddechowych oraz może być problematyczna do zastosowania np. u dzieci. Dalsze badania w tym zakresie powinny skupić się na sprawdzeniu wpływu parametrów z domeny przyczynowej i informacyjnej oraz określeniu optymalnego progu procentowego  $HR_{max}$  pod względem minimalizacji błędu predykcji oraz wymaganego wysiłku. Wartościowe byłoby także zbadanie wpływu poziomu aktywności fizycznej badanych na dokładność predykcji oraz przeprowadzenie modelownia dla innych form testów wysiłkowych (np. test na cykloergometrze), aby określić optymalne warunki do przewidywania  $VO_{2peak}$ . Przydatność opisanej metody może również zależeć od powtarzalności wyników, które również wymagają dalszego sprawdzenia. Przy wysokiej powtarzalności metoda ta mogłaby być użyteczna w praktyce klinicznej, np. do śledzenia zmian w CRF podczas rehabilitacji kardiologicznej lub obozów treningowych sportowców, bez konieczności przeprowadzania pełnego testu wysiłkowego.



## 4. Granger causality test with nonlinear neural-network-based methods: Python package and simulation study

### 4.1. Materiały i metody

Celem artykułu było przedstawienie zaproponowanej metody analizy przyczynowości opartej na nieliniowym modelowaniu z zastosowaniem modeli uczenia maszynowego i opracowanego w języku Python pakietu *nonlincausality*, wraz z prezentacją zastosowania na zasymulowanych szeregach czasowych.

Stworzony pakiet składał się z 8 funkcji, które można podzielić na funkcje testujące związek przyczynowy między dwoma seriami czasowymi oraz funkcje mierzące zmianę przyczynowości w czasie. Pakiet opiera się na podejściu zaproponowanym przez Grangera, gdzie o występowaniu zależności przyczynowej  $Y \rightarrow X$  świadczy sytuacja, gdy przeszłe wartości  $Y$  pozwalają na dokładniejszą predykcję obecnej wartości  $X$  względem predykcji opartej wyłącznie na przeszłości  $X$ . Jednakże w celu przezwyciężenia ograniczeń związanych z zastosowaniem modeli autoregresji liniowej zastosowano modele oparte o sieci neuronowe. Pakiet pozwala na zastosowanie modeli takich jak perceptron wielowarstwowy (MLP, z ang. multilayer perceptron), sieć rekurencyjna z komórkami długiej pamięci krótkotrwałej (LSTM, z ang. long-short term memory) oraz z komórkami bramkowanych jednostek rekurencyjnych (GRU, z ang. gated recurrent unit). MLP jest prostą siecią neuronową zbudowaną z warstw pojedynczych perceptronów, która umożliwia modelowanie zależności nieliniowych. Komórki LSTM zawierają trzy bramki, które kontrolują przepływ informacji, umożliwiając "zapamiętywanie" lub "zapominanie" informacji z poprzednich momentów czasowych, co pozwala na zniwelowanie problemu zanikającego gradientu, występującego w podstawowych sieciach rekurencyjnych. Komórka GRU, zaproponowana przez Cho et al. w 2014 [84], jest podobna do komórki LSTM, jednakże ma względem niej uproszczoną strukturę i charakteryzuje się dwiema bramkami, co redukuje złożoność obliczeniową. Ponadto pakiet umożliwia analizę przyczynowości z wykorzystaniem modelu autoregresyjnej zintegrowanej średniej kroczącej (ARIMA, z ang. autoregressive integrated moving average). Do oceny istotności statystycznej poprawy predykcji przy uwzględnieniu przeszłych wartości  $Y$  użyto testu Wilcoxona, przy użyciu którego porównywane są wartości absolutne błędu modelu wykorzystującego oba szeregi czasowe  $X$  i  $Y$  z wartościami otrzymanymi dla modelu używającego wyłącznie szeregu  $X$ . Hipoteza zerowa zakłada, że mediana błędu absolutnego dla modelu opartego na przeszłych wartościach  $X$  jest równa lub mniejsza niż dla modelu

opartego na przeszłych wartościach  $X$  i  $Y$ . Cały pakiet został napisany w języku Python z wykorzystaniem biblioteki Keras do modelowania z użyciem sieci neuronowych. Od czasu opublikowania pracy pakiet został zaktualizowany do wersji 2.0.2, gdzie zredukowano liczbę funkcji z zachowaniem wszystkich opisanych funkcjonalności, zwiększono możliwości kontroli procesu treningu sieci neuronowej, dodano możliwość modyfikacji parametrów regularyzacji oraz umożliwiono modelowanie z wykorzystaniem dowolnych modeli uczenia maszynowego dostępnych w bibliotece Scikit-learn.

Na potrzeby przedstawienia wykorzystania opracowanego pakietu do wykrywania przyczynowości za pomocą różnych modeli, wygenerowano dwa sygnały  $X$  i  $Y$  zgodnie ze wzorami:

$$Y(t) = \cos(t) + \sin(0.15 * t) + E_1(t) \quad (3)$$

$$X(t) = 2 * Y(t - 100)^3 - 5 * Y(t - 100)^2 + 0.3 * Y(t - 100) + 2 + E_2(t) \quad (4)$$

Analizę przeprowadzono dla dwóch przypadków: obecności zależności przyczynowej z wykorzystaniem wygenerowanych sygnałów i braku związku między szeregami czasowymi (w tym wypadku zamieniono sygnał  $Y$  na losowy szum). Zastosowano również dwie wartości analizowanych opóźnień równych 50 i 150. Pierwsze 70% sygnałów zostało użyte jako dane treningowe, a pozostałe 30% jako zbiór testowy. Do analizy przyczynowości wykorzystano 3 sieci neuronowe z utworzonego pakietu, o następujących architekturach:

- dwie warstwy LSTM z 10 komórkami każda (LSTM), po których następuje jeden neuron wyjściowy z liniową funkcją aktywacji;
- dwie warstwy GRU z 10 komórkami każda (GRU), po których następuje jeden neuron wyjściowy z liniową funkcją aktywacji;
- dwie warstwy w pełni połączone ze 100 neuronami każda (MLP) i funkcją aktywacji ReLU, po których następuje jeden neuron wyjściowy z liniową funkcją aktywacji.

Dla każdej sieci zastosowano do regularyzacji technikę porzucania (ang. dropout) z wartością współczynnika porzucania równą 0,01. Sieci neuronowe były trenowane przez 150 epok, z współczynnikiem uczenia równym 0,001 i 0,0001 odpowiednio dla pierwszych 50 epok oraz ostatnich 100. Aby uzyskać najdokładniejsze modele, sieci LSTM i GRU były tworzone i trenowane 2 razy, a model MLP 5 razy (przy wykorzystaniu parametru *run*). Do ostatecznych testów przyczynowości wybrano modele (jeden oparty na  $X$  i jeden oparty na  $X$  oraz  $Y$ ) z najmniejszą wartością sumy kwadratów błędów (RSS, z ang. residua sum of squares). Różnica

w wartości parametru *run* była spowodowana długim czasem potrzebnym na trening LSTM i GRU.

Do analizy przyczynowości zastosowano również model autoregresji liniowej (AR), stosowany w tradycyjnym podejściu Grangera oraz metodę wielkoskalowej nieliniowej przyczynowości Grangera (lsNGC, z ang. large-scale nonlinear Granger causality), która wykorzystuje sieć z uogólnionymi radialnymi funkcjami bazowymi (GRBF, z ang. generalized radial basis functions) [85]. Jakość predykcji wszystkich metod została skwantyfikowana za pomocą średniego błędu kwadratowego (MSE, z ang. mean square error), średniego błędu bezwzględnego (MAE, z ang. mean absolute error) oraz mediany błędu bezwzględnego (MedAE, z ang. median absolute error). Ponadto oceniono, czy użycie proponowanych metod skutkuje poprawą dokładności modeli predykcyjnych w porównaniu z AR lub lsNGC. W tym celu użyto testu rang Wilcoxona do porównania wartości bezwzględnych błędów uzyskanych z odpowiadających sobie modeli (np. model MLP bazujący na przeszłości  $X$  i model AR bazujący również na przeszłości  $X$ ). Do oceny wielkość efektu włączenia przeszłości szeregu czasowego  $Y$  do predykcji dla każdej metody użyto  $d$  Cohena. W celu lepszego zobrazowania wyników utworzono wykresy wartości przewidywanych w stosunku do wartości prawdziwych oraz wykresy błędu predykcji względem wartości przewidywanych dla każdego modelu. Przyjęty poziom istotności wynosi 0,05. Wszystkie analizy zostały przeprowadzone przy użyciu języka Python w wersji 3.7.10.

## 4.2. Wyniki

Wszystkie metryki błędów obliczone na zbiorze testowym dla każdej metody analizy przyczynowości, dla opóźnienia równego 50 i 150, przedstawiono w Tabeli 3. P-wartości uzyskane z testu Wilcoxona, użytego do oceny obecności przyczynowości  $Y \rightarrow X$  dla każdej metody i każdego opóźnienia, przedstawiono w Tabeli 4. Przykładowe wykresy wartości zestawu testowego i wartości przewidzianych z wykorzystaniem sieci neuronowych dla opóźnienia równego 50 i 150, przedstawiono na Rycinie 10.

Tabela 3. Metryki błędu uzyskane na zbiorze testowym dla każdego modelu utworzonego na podstawie 50 i 150 wartości przeszłych w przypadku, gdy  $Y \rightarrow X$ . Najmniejsza metryka błędu dla danego przypadku jest wyróżniona pogrubieniem.

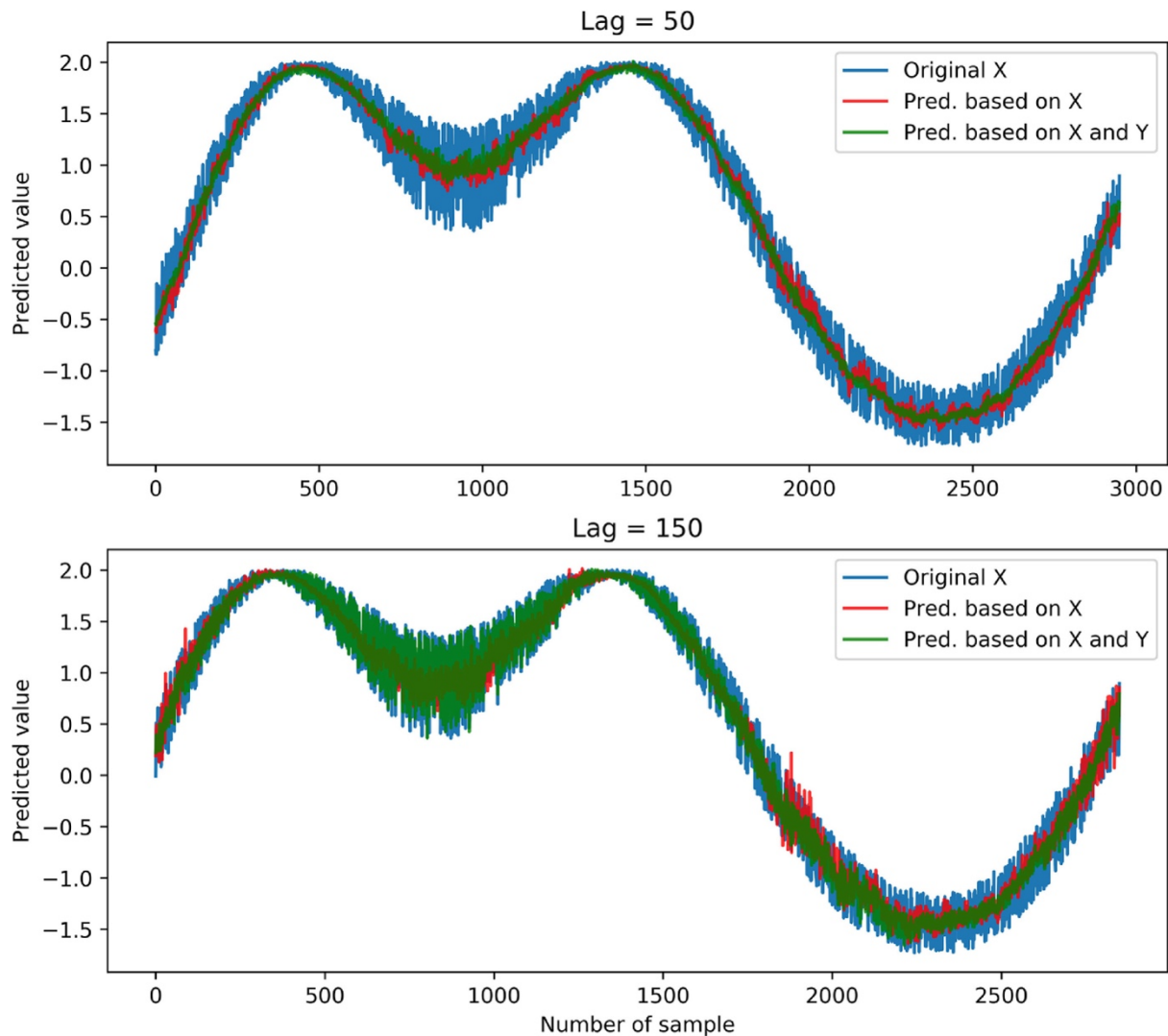
| Opóźnienie | Metryka | Sygnały | LSTM         | GRU          | MLP          | AR           | GRBF         |
|------------|---------|---------|--------------|--------------|--------------|--------------|--------------|
| 50         | MSE     | X       | 0,042        | <b>0,041</b> | 0,042        | <b>0,041</b> | 0,057        |
|            |         | X and Y | 0,039        | 0,039        | 0,039        | 0,041        | <b>0,038</b> |
|            | MAE     | X       | 0,162        | <b>0,161</b> | 0,162        | <b>0,161</b> | 0,188        |
|            |         | X and Y | 0,158        | 0,158        | 0,161        | 0,161        | <b>0,155</b> |
|            | MedAE   | X       | 0,139        | 0,138        | <b>0,136</b> | <b>0,136</b> | 0,153        |
|            |         | X and Y | 0,137        | <b>0,135</b> | 0,137        | 0,140        | <b>0,135</b> |
| 150        | MSE     | X       | <b>0,041</b> | <b>0,041</b> | 0,045        | <b>0,041</b> | 0,254        |
|            |         | X and Y | 0,039        | 0,038        | <b>0,011</b> | 0,041        | 0,036        |
|            | MAE     | X       | <b>0,159</b> | 0,161        | 0,166        | 0,160        | 0,413        |
|            |         | X and Y | 0,156        | 0,156        | <b>0,078</b> | 0,160        | 0,151        |
|            | MedAE   | X       | 0,135        | 0,136        | <b>0,132</b> | 0,137        | 0,379        |
|            |         | X and Y | 0,134        | 0,134        | <b>0,056</b> | 0,136        | 0,130        |

Tabela 4. P-wartości dla każdego modelu i każdego testowanego opóźnienia uzyskane z testu Wilcoxona dla  $Y \rightarrow X$ . Przypadki, w których wykryto związek przyczynowy, są wyróżnione pogrubieniem.

| Opóźnienie | LSTM              | GRU               | MLP               | AR    | GRBF              |
|------------|-------------------|-------------------|-------------------|-------|-------------------|
| 50         | <b>&lt; 0,001</b> | <b>&lt; 0,001</b> | <b>0,025</b>      | 0,703 | <b>&lt; 0,001</b> |
| 150        | <b>&lt; 0,001</b> | <b>&lt; 0,001</b> | <b>&lt; 0,001</b> | 0,261 | <b>&lt; 0,001</b> |

W przypadku opóźnienia równego 50, wszystkie modele z utworzonego pakietu uzyskały podobne wyniki pod względem metryk błędów. Modele AR miały podobną dokładność przewidywania, podczas gdy modele GRBF miały tendencję do uzyskiwania największych metryk błędów dla modelu opartego tylko na poprzednich wartościach X i najmniejszych w przypadku modelu opartego na obu sygnałach. Dla opóźnienia 150, NN i AR uzyskały przeważnie podobne wyniki, z wyjątkiem modelu MLP opartego na X i Y, który przewyższył wszystkie inne modele i uzyskał najmniejszy błąd we wszystkich trzech metrykach. W przypadku większego opóźnienia, model GRBF odnotował spadek jakości predykcji dla modelu opartego tylko na sygnale X. W przypadku testu przyczynowości, wszystkie nieliniowe metody uzyskały p-wartość mniejszą niż przyjęty poziom istotności, co oznacza,

że w przypadku użycia modeli z opracowanego pakietu lub modelu GRBF istnieją podstawy do odrzucenia hipotezy o braku zależności przyczynowej  $Y \rightarrow X$ , nawet dla opóźnienia mniejszego niż rzeczywiste opóźnienie między szeregami czasowymi. Model autoregresyjny używany w pierwotnej metodzie Grangera nie wykazał występowania zależności przyczynowej dla żadnego z podanych opóźnień.



Rycina 10. Wykres prezentujący dane testowe oraz wartości przewidziane przez model oparty wyłącznie na przeszłych wartościach szeregu czasowego X oraz model bazujący na przeszłości szeregów X i Y.

### 4.3. Dyskusja

Praca opisująca opracowany pakiet *nonlincausality* prezentuje nową metodologię korzystającą z sieci neuronowych do badania zależności przyczynowych, która radzi sobie z ograniczeniami klasycznych modeli autoregresyjnych stosowanych w metodzie Grangera. Modele AR zakładają liniową zależność między analizowanymi szeregami czasowymi, w związku z czym

mogą nie wykrywać bardziej złożonych relacji, które są obserwowane w wielu dziedzinach, takich jak medycyna czy ekonomia. Sieci neuronowe w zaprezentowanym pakiecie nie tylko pozwalają na wykrywanie bardziej skomplikowanych współzależności, ale również pozwalają na obejście ograniczenia wymogu stacjonarności analizowanych szeregów czasowych. W tej publikacji modele oparte na sieciach neuronowych były w stanie wskazać występowanie związków przyczynowych dla zależności nieliniowych, co nie było możliwe z wykorzystaniem tradycyjnego podejścia Grangera oraz przeważnie osiągały najmniejsze wartości metryk błędów w porównaniu z innymi analizowanymi algorytmami. Porównując uzyskane wyniki z metodą lsNGC, wytrenowane modele MLP charakteryzowały się znacznie mniejszym błędem predykcji dla większej z analizowanych wartości opóźnienia. Co więcej, modele utworzone z wykorzystaniem opracowanego pakietu nie wykazały związku przyczynowego w sytuacji, gdy zastąpiono wygenerowany szereg czasowy  $Y$  losowymi wartościami, podczas gdy metoda lsNGC wskazała taką zależność dla opóźnienia równego 150. Ograniczeniem zaproponowanego podejścia jest jednak złożoność obliczeniowa i czasochłonność treningu sieci neuronowych, w związku z częstą koniecznością dostrojenia hiperparametrów modeli. Podjęto również wysiłki w celu zaadresowania tego problemu poprzez umożliwienie wykorzystania prostszych obliczeniowo modeli uczenia maszynowego dostępnych w bibliotece Scikit-learn, które również bazują na podejściach nieliniowych, jednakże ta aktualizacja pakietu została dodana po opublikowaniu omawianej pracy i wykorzystanie tej możliwości nie zostało w niej zaprezentowane. Dodatkowo ograniczeniem tej pracy jest zastosowanie sygnałów zasymulowanych a nie rzeczywistych oraz w związku z wykorzystaniem bardziej zaawansowanych metod modelowania, konieczność pogłębionej analizy uzyskiwanych wyników w celu kontroli zjawiska niedotrenowania (ang. underfitting) lub przetrenowania (ang. overfitting) modeli ML. Niewątpliwą zaletą opracowanego pakietu jest jego uniwersalność pod względem zastosowania, publiczna dostępność i możliwość wykorzystania w dowolnych dziedzinach nauki, takich jak ekonomia, neurobiologia czy fizjologia. Zgodnie z planami przedstawionymi w tej publikacji, w późniejszym czasie wykorzystano opracowany pakiet do badania zależności przyczynowych między sygnałami krążeniowo-oddechowymi.

## 5. Causality in cardiorespiratory signals in pediatric cardiac patients

### 5.1. Materiały i metody

Kolejna praca miała na celu zastosowanie opracowanej metody analizy przyczynowości opartej o sieci neuronowe (NNGC) i porównanie jej z trzema innymi metodami - jednej liniowej i dwoma nieliniowymi: przyczynowości Grangera (GC), jądrowej przyczynowości Grangera (KGC, z ang. kernel Granger causality) oraz lsNGC - do oceny i ilościowego określenia niemiarowości oddechowej w grupie pediatrycznych pacjentów kardiologicznych. W tym celu w grupie 20 pacjentów pediatrycznych Kliniki Kardiologii Wieku Dziecięcego i Pediatrii Ogólnej Warszawskiego Uniwersytetu Medycznego wykonano pomiary z wykorzystaniem urządzenia Pneumonitor, które pozwala na jednoczesną akwizycję sygnału elektrokardiograficznego oraz pneumografii impedancyjnej [34], [35], [36]. Charakterystykę grupy przedstawiono w Tabeli 5.

Tabela 5. Charakterystyka grupy badanej z podziałem na płeć przedstawiona jako wartość średnia  $\pm$  odchylenie standardowe.

|                          | Wszyscy          | Chłopcy (12)     | Dziewczynki (8)  |
|--------------------------|------------------|------------------|------------------|
| Wiek [lata]              | 12,9 $\pm$ 3,5   | 14,3 $\pm$ 3,3   | 10,8 $\pm$ 2,6   |
| Masa [kg]                | 58,8 $\pm$ 23,2  | 65,2 $\pm$ 22,9  | 46,0 $\pm$ 18,1  |
| Wzrost [cm]              | 160,1 $\pm$ 17,2 | 166,1 $\pm$ 16,3 | 151,0 $\pm$ 16,3 |
| BMI [kg/m <sup>2</sup> ] | 21,5 $\pm$ 5,0   | 22,7 $\pm$ 5,1   | 19,3 $\pm$ 4,1   |

Z zarejestrowanego sygnału EKG, poprzez interpolację sześcienną wyznaczonych interwałów RR (RRi), uzyskano tachogram (RR). W związku z zastosowaną konfiguracją elektrod, sygnał IP był traktowany jako ekwiwalent sygnału objętości oddechowej (TV, z ang. tidal volume). Częstotliwość próbkowania obu szeregów czasowych została zredukowana do 25 Hz, aby osiągnąć kompromis między złożonością obliczeniową a zachowaniem rozdzielczości czasowej. Tak uzyskane szeregi czasowe zostały poddane analizie przyczynowości z wykorzystaniem wyżej wymienionych metod. Analiza przyczynowości była przeprowadzana dla opóźnienia równego 1 sekundzie dla każdej metody. Architektura sieci neuronowych użytych w NNGC składała się z 2 w pełni połączonych warstw ukrytych, zawierających po 20 neuronów, z funkcją aktywacji ReLU w warstwach ukrytych i liniową funkcją aktywacji w warstwie wyjściowej. Sieci były trenowane przez 100 epok, ze współczynnikiem uczenia równym odpowiednio  $10^{-4}$  i  $10^{-5}$  dla pierwszej i drugiej połowy procesu uczenia. Zbadano

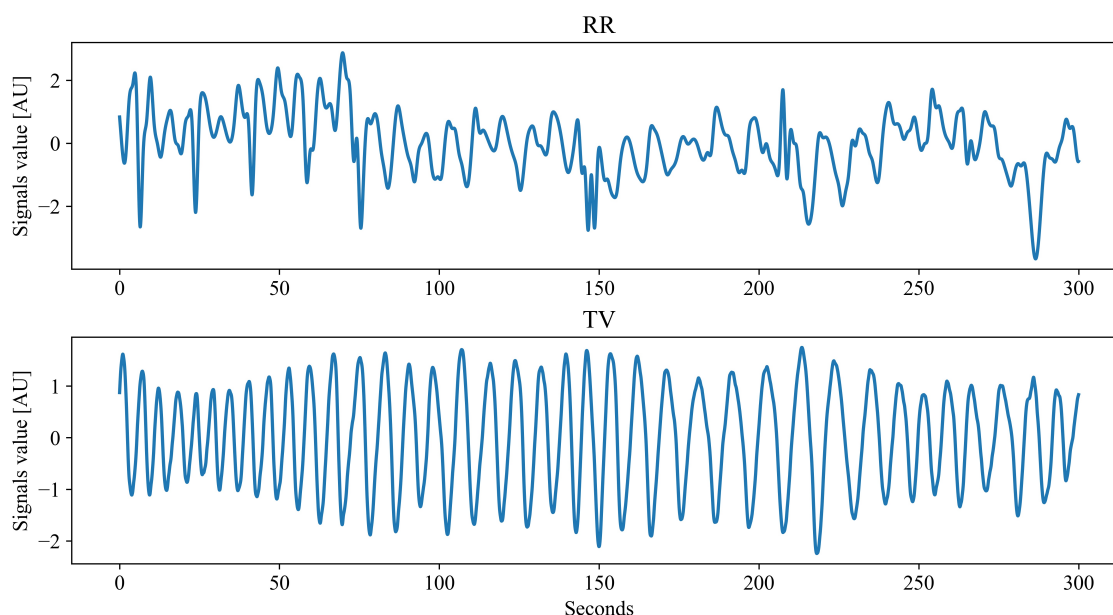
również zależności między otrzymanymi wartościami przyczynowości w obu kierunkach a danymi demograficznymi. Korzystając ze współczynnika korelacji Spearmana, obliczono korelację między wartościami przyczynowości a wiekiem pacjentów oraz ich wskaźnikiem masy ciała (BMI, z ang. body mass index), jak również między wartościami przyczynowości w różnych kierunkach. Sprawdzono również występowanie różnicy w wartościach przyczynowości w zależności od płci pacjentów przy użyciu testu Manna-Whitneya.

Zależności czasowe między sygnałami były także zbadane za pomocą pakietu *tempord* [86], z zastosowaniem metody modelowania liniowego z progiem 0,9 i przesunięciem czasowym od -2 do 2 sekund. Wyniki uzyskane z analizy porządków czasowych zostały również porównane z wynikami z grupy 10 elitarnych sportowców i 10 zdrowych studentów przedstawionymi w [86].

## 5.2. Wyniki

Przykład uzyskanego tachogramu i sygnału oddechowego, przedstawiono na Rycinie 11. Dla tych sygnałów dokonano analizy przyczynowości pomiędzy nimi dla każdego z pacjentów z osobna, wykorzystując różne metody oraz analizując zależność zarówno od TV do RR jak i od RR do TV. Uzyskane wartości liczbowe określające siłę zależności przyczynowych (stopień, w jakim przeszłość szeregu czasowego Y poprawia predykcję wartości szeregu X) zostały zilustrowane na Rycinie 12. Dla dwunastu pacjentów przyczynowość TV→RR była wyższa niż w kierunku odwrotnym, a dla ośmiu pacjentów większe wartości przyczynowości RR→TV odnotowano przy użyciu metod GC i NNGC. W przypadku metody lsNGC proporcje były odwrotne: dla ośmiorga dzieci większe wartości przyczynowości zaobserwowano dla kierunku od TV do RR, a dla dwanaściorga od RR do TV. Przyczynowość między sygnałami była statystycznie istotna w obu kierunkach niezależnie od użytej metody, z wyjątkiem przyczynowości TV→RR według lsNGC u pacjenta nr 4.



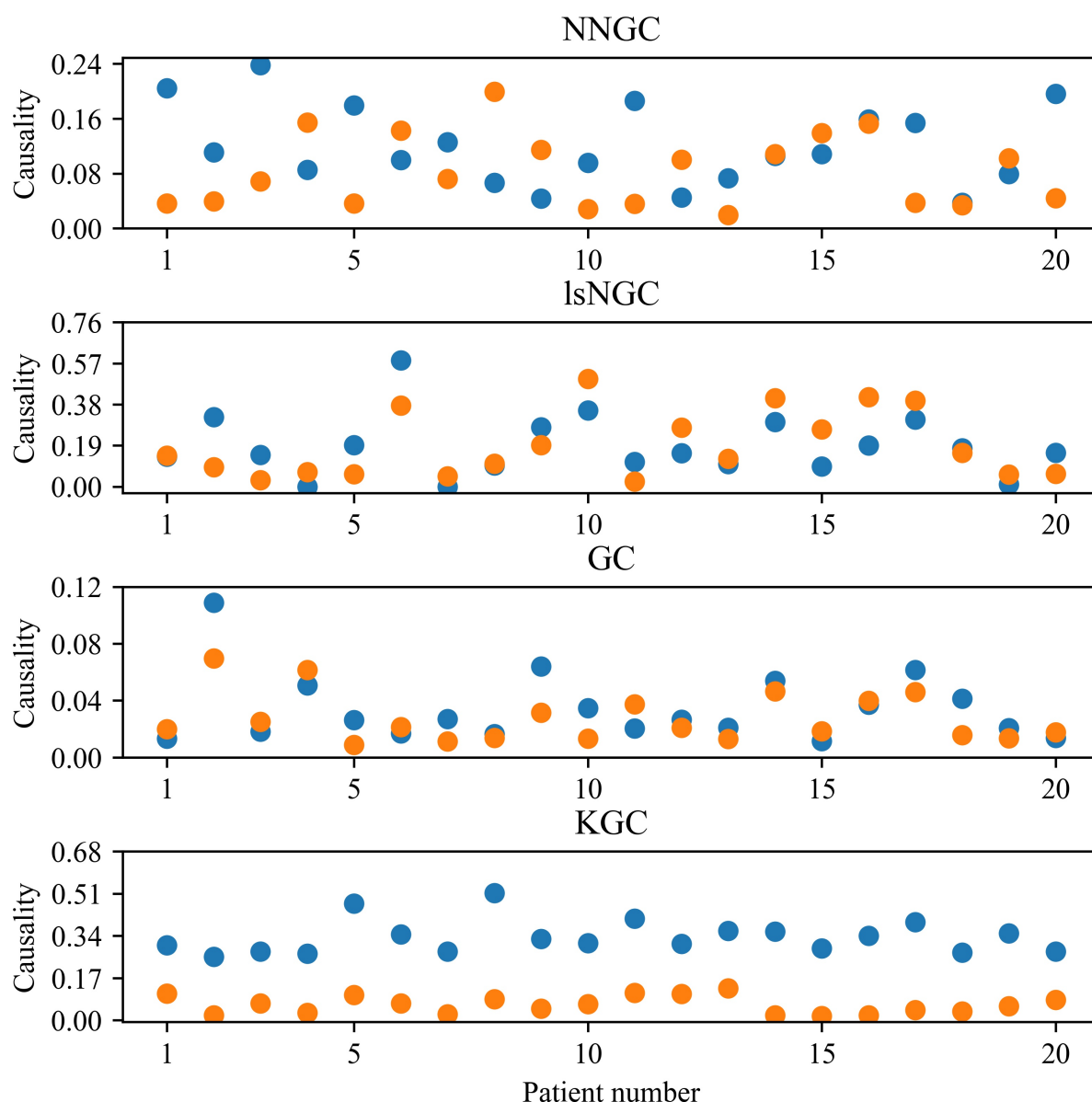


Rycina 11. Przykładowe przebiegi tachogramu i sygnału oddechowego od pacjenta #8.

Wykonano analizę korelacji pomiędzy wartościami przyczynowości a wiekiem i BMI pacjentów oraz między wartościami przyczynowości w obu kierunkach. Wyniki tej analizy przedstawiono w Tabeli 6. Niezależnie od metody, nie stwierdzono statystycznie istotnej korelacji między wiekiem a kierunkiem przyczynowości. Zaobserwowano natomiast istotną statystycznie ujemną korelację między BMI a wartością przyczynowości TV→RR wyliczoną metodą NNGC oraz dodatnią korelację między BMI a RR→TV wyliczoną metodą lsNGC. Dla metod lnNGC i GC również stwierdzono istotne korelacje między TV→RR a RR→TV. Zaobserwowano również statystycznie istotne różnice w przyczynowości między chłopcami a dziewczętami dla metod lsNGC i GC (p-wartości < 0,05).

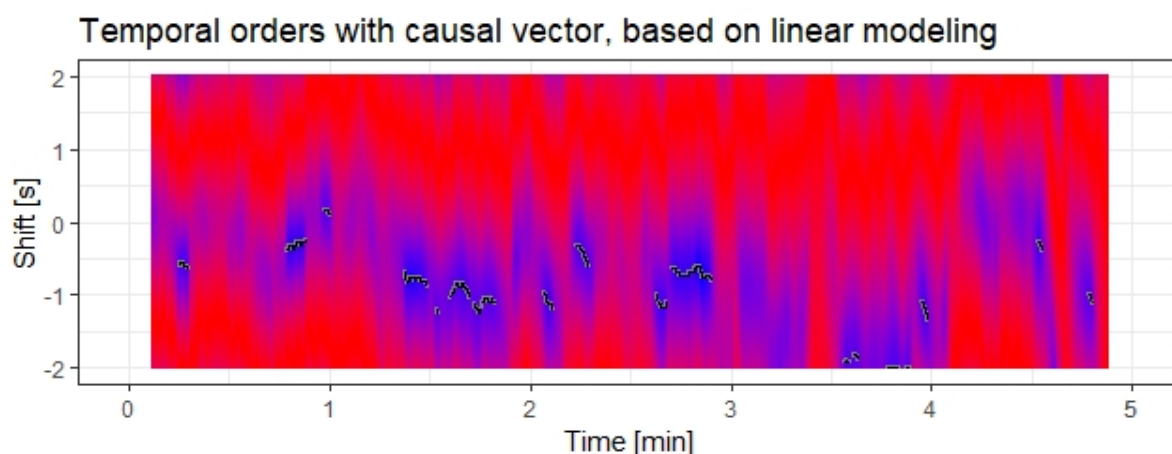
Tabela 6. Wartości współczynników korelacji rangowej Spearmana między wartościami przyczynowości a wiekiem i BMI oraz między wartościami przyczynowości w przeciwnych kierunkach. W nawiasach podano uzyskane p-wartości.

|       | Wiek         |              | BMI           |              | Przyczynowości |
|-------|--------------|--------------|---------------|--------------|----------------|
|       | TV→RR        | RR→TV        | TV→RR         | RR→TV        |                |
| GC    | 0,30 (0,20)  | 0,16 (0,49)  | 0,25 (0,30)   | 0,12 (0,63)  | 0,45 (<0,05)   |
| KGC   | -0,38 (0,10) | -0,30 (0,21) | -0,31 (0,20)  | -0,33 (0,17) | 0,40 (0,08)    |
| lsNGC | 0,17 (0,47)  | 0,13 (0,60)  | 0,15 (0,53)   | 0,52 (<0,05) | 0,59 (<0,01)   |
| NNGC  | -0,35 (0,14) | 0,20 (0,39)  | -0,46 (<0,05) | 0,04 (0,88)  | -0,19 (0,43)   |



Rycina 12. Wartości przyczynowości uzyskane dla różnych metod dla każdego pacjenta. Niebieskie kropki reprezentują wartości przyczynowości TV→RR, natomiast pomarańczowe odpowiadają wartościom przyczynowości RR→TV.

Przykładowy wykres uzyskany z analizy porządków czasowych z wykorzystaniem pakietu *tempord* został przedstawiony na Rycinie 13. Zgodnie z oczekiwaniami, wektory przyczynowości uzyskane wśród pacjentów kardiologicznych były znacznie krótsze i charakteryzowały się mniejszą ciągłością niż te uzyskane w populacji profesjonalnych sportowców oraz zdrowych studentów [86]. Średni wektor przyczynowości wynosił  $-72 \pm 668$  milisekund (średnia  $\pm$  odchylenie standardowe). Dla sześciu z dwudziestu pacjentów nie zaobserwowano w ogóle wektora przyczynowości.



Rycina 13. Wykres porządków czasowych z zaznaczonym wektorem przyczynowości uzyskany przy użyciu pakietu *tempord* dla pacjenta #8.

### 5.3. Dyskusja

Statystycznie istotna zależność przyczynowa między sygnałami RR a TV została wykryta, a co za tym idzie skwantyfikowana, przez każdą z zastosowanych metod przynajmniej w jednym kierunku. Metoda KGC jako jedyna w każdym przypadku wskazywała na silniejszy związek przyczynowy od TV do RR. Metody GC i NNGC częściej identyfikowały większą przyczynowość od TV do RR, podczas gdy lsNGC częściej wskazywała na większą przyczynowość od RR do TV. Co więcej, stwierdzono statystycznie istotną negatywną korelację między BMI a wartościami przyczynowości od TV do RR dla NNGC, co może świadczyć o silnym związku tej miary z aktywnością autonomicznego układu nerwowego, ponieważ podobną zależność obserwuje się między wartościami BMI a parametrami HRV, które zależą właśnie od pracy tego układu [87], [88].

Analiza porządków czasowych również zdaje się mieć potencjał jako narzędzie diagnostyczne, w związku z widoczną różnicą otrzymanych wykresów w porównaniu do przebiegów wektorów przyczynowych u sportowców i zdrowych studentów zaprezentowanych w [86]. Ze względu na jednolitość grupy badanej i brak uprzednich badań w tym zakresie, nie było możliwości analogicznego sprawdzenia dla badania zmiany przyczynowości w czasie z wykorzystaniem pakietu *nonlincausality*, jednakże pokazano podobną możliwość tego narzędzia przy wizualizacji współzależności.

Ponieważ każda z badanych metod jest zdolna do wykrywania zależności przyczynowych między analizowanymi sygnałami krążeniowo-oddechowymi, wszystkie mogą być potencjalnie użyteczne, dostarczając różnych informacji diagnostycznych. Stwierdzona

przyczynowość od TV do RR jest zgodna z oczekiwaniami w związku z występowaniem zjawiska niemiarowości oddechowej. Wykryta przyczynowość od RR do TV, która może wydawać się sprzeczna ze zjawiskiem RSA, poza fizjologiczną dwukierunkową zależnością między pracą serca a układu oddechowego [89], może być związana z charakterystyką zastosowanych metod, które badają zależności przyczynowe bazując na analizie sygnałów. W wypadku sygnału RR i TV może dochodzić do występowania lokalnych maksimów sygnału RR przed wystąpieniem lokalnych maksimów sygnału TV [90], [91], co może być interpretowane przez użyte metody jako wystąpienie przyczynowości. Dlatego też interpretacja zależności przyczynowych powinna być wykonywana w połączeniu z aktualną wiedzą z danej domeny. Pomimo nieoczywistego charakteru wykrytych zależności RR→TV, mogą one stanowić istotną informację diagnostyczną, ponieważ opisują wcześniej nieparametryzowane współzależności krążeniowo-oddechowe. Głównym ograniczeniem tego badania jest stosunkowo mała i jednorodna grupa badawcza. Brak grupy kontrolnej zdrowych osób uniemożliwia określenie przydatności obliczonych wartości przyczynowości jako biomarkera dla problemów kardiologicznych. Ograniczeniem omawianych metod analizy przyczynowości jest ich zależność od wyboru parametrów, co może mieć wpływ na końcowy wynik, a także niedeterministyczny charakter trenowania sieci neuronowych. W celu dalszej walidacji zaprezentowanych metod, a także innych sposobów parametryzacji zależności krążeniowo-oddechowych, wykonano dalsze badania opisane w kolejnej pracy.

## 6. Prediction of the duration of maximal exercise test in professional adolescent football players based on the cardiorespiratory signals – a pilot study

### 6.1. Materiały i metody

Celem tego artykułu było zbadanie jakości przewidywania wydolności krążeniowo-oddechowej pod względem czasu trwania próby wysiłkowej, przy wykorzystaniu modeli uczenia maszynowego trenowanych na podstawie cech krążeniowo-oddechowych obliczonych z sygnałów pozyskanych podczas spoczynku w pozycji leżącej wśród grupy wytrenowanych młodych sportowców oraz ocena wpływu cech z domeny przyczynowej i informacyjnej na generowane predykcje. W badaniu wzięło udział 36 osób aktywnie trenujących piłkę nożną w wieku od 10 do 15 lat, których dane demograficzne zostały przedstawione w Tabeli 7. W trakcie badania najpierw, przez co najmniej 5 minut w spoczynku w pozycji leżącej rejestrowano sygnały krążeniowo-oddechowe, po czym przeprowadzono test wysiłkowy na cykloergometrze, zaczynając od obciążenia równego 40 W i zwiększając obciążenie o 40 W co 3 minuty. Zawodnicy mieli wykonywać test do wyczerpania. Mierzono czas trwania próby wysiłkowej od momentu rozpoczęcia do momentu odmowy dalszego wysiłku przez osobę wykonującą test. Badanie zostało zatwierdzone przez Komisję Bioetyczną Warszawskiego Uniwersytetu Medycznego (KB/70/2021) oraz uzyskano pisemną zgodę na udział w badaniu od opiekuna prawnego każdego z uczestników. Do rejestracji sygnałów krążeniowo-oddechowych użyto urządzenia Pneumonitor, tak jak w poprzedniej pracy. Z zarejestrowanych sygnałów uzyskano szereg czasowy RR oraz TV w sposób analogiczny jak w [73], z tą różnicą, że sygnał TV został dodatkowo wyfiltrowany za pomocą filtru pasmowo-przepustowego o częstotliwościach odcięcia 0,05 i 0,67 Hz, co odpowiadało 3 i 40 oddechom na minutę. Modelowanie z wykorzystaniem metod uczenia maszynowego przeprowadzono przy użyciu parametrów demograficznych, HRV, oddechowych oraz w domenie przyczynowej i informacyjnej jako wektorów cech wejściowych. W celu ograniczenia zbioru cech uwzględniono tylko te cechy, które miały współczynnik korelacji Pearsona w stosunku do czasu trwania CPET większy niż 0,2. Ocena dokładności predykcji była przeprowadzana za pomocą MAPE, MAE, RMSE, współczynnika determinacji  $R^2$  oraz współczynnika korelacji Pearsona, z zastosowaniem techniki walidacji krzyżowej z zestawem testowym składającym się z jednej próbki (leave-one-out). Wszystkie metryki były obliczane po walidacji, na podstawie porównania wyników predykcji uzyskanych na próbkach testowych z rzeczywistymi

wartościami czasu trwania testu. Aby zbadać wpływ poszczególnych cech na wyniki predykcji, zastosowano techniki XAI w postaci wartości Shapley'a obliczonych dla danych testowych. Wszystkie analizy wykonano w języku Python 3.10.8.

Tabela 7. Charakterystyka grupy badanej zaprezentowana jako średnia  $\pm$  odchylenie standardowe. Wartości  $VO_{2max}$  zostały otrzymane na podstawie testu Yoyo i nie zostały uwzględnione w przeprowadzonej analizie.

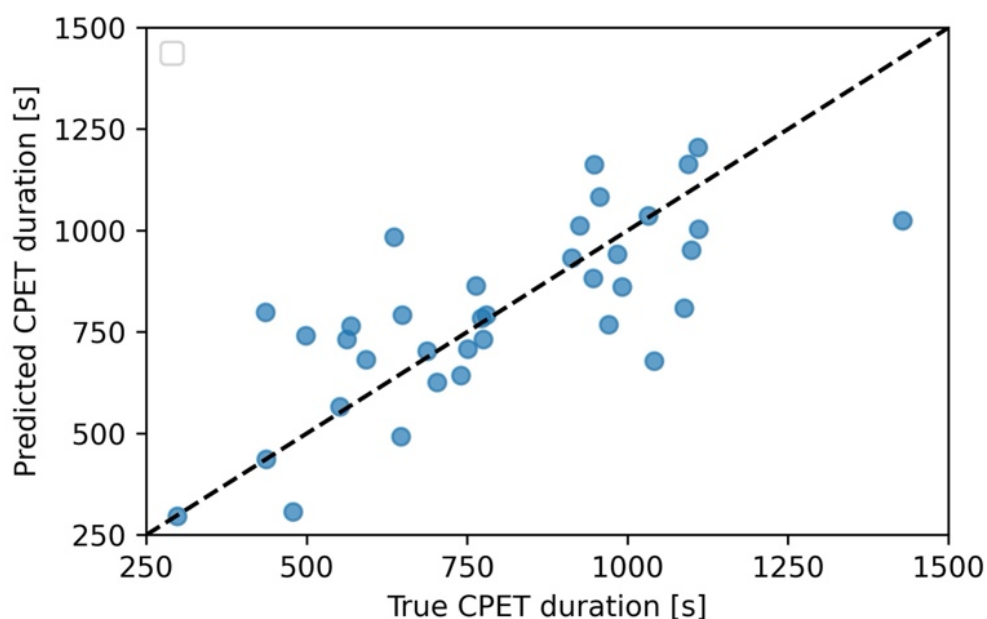
| Wiek [lata]    | Masa [kg]       | Wzrost [cm]      | BMI [kg/m <sup>2</sup> ] | $VO_{2max}$<br>[ml/min/kg] |
|----------------|-----------------|------------------|--------------------------|----------------------------|
| 13,3 $\pm$ 1,5 | 56,7 $\pm$ 14,0 | 169,1 $\pm$ 13,0 | 19,4 $\pm$ 2,4           | 48,7 $\pm$ 5,5             |

## 6.2. Wyniki

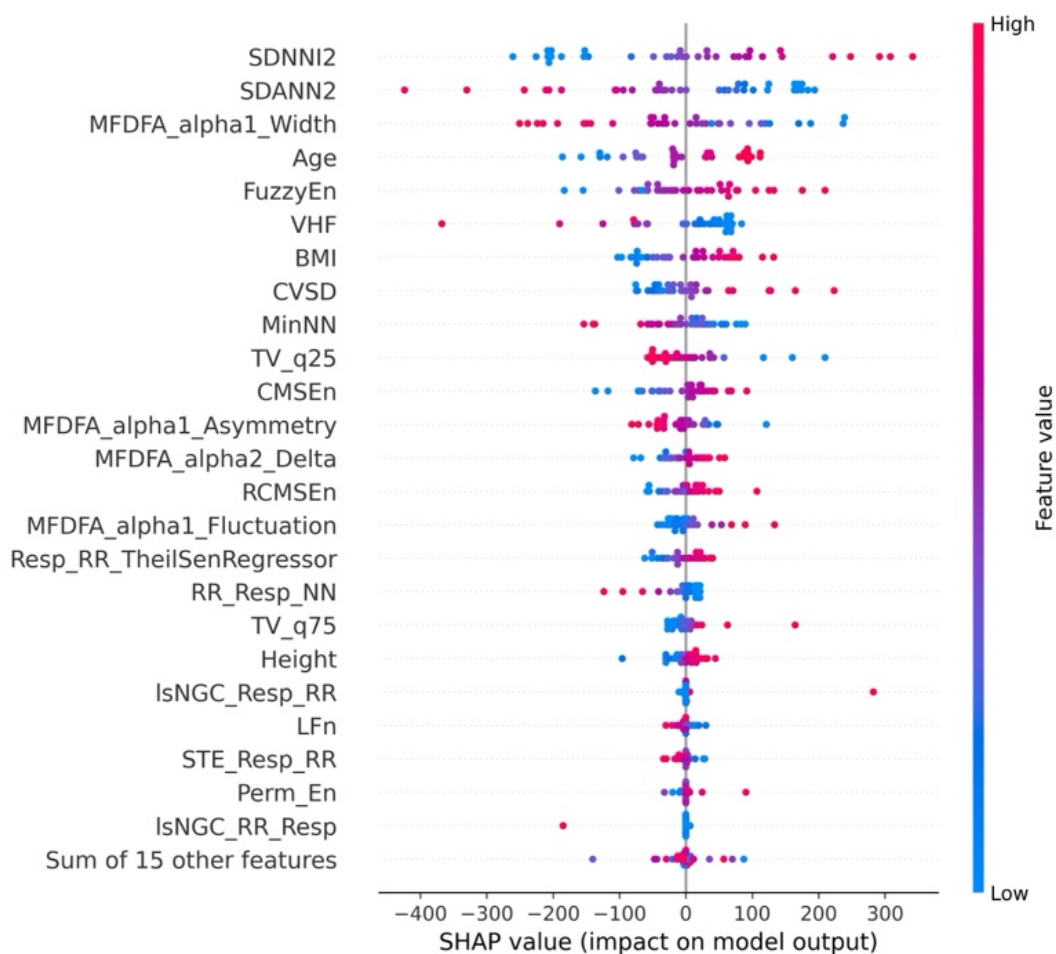
Początkowy zestaw danych składający się z 153 cech po procesie selekcji ograniczono do 39. Najlepsze wyniki pod względem MAPE uzyskano dla modelu regresji liniowej z zastosowaną regularyzacją Lasso, z parametrem alfa równym 1,65. Uzyskane metryki przedstawiono w Tabeli 8. Zależność między przewidywanymi wartościami a rzeczywistymi czasami trwania CPET została przedstawiona na Rycinie 14. Istotność cech oceniona za pomocą wartości Shapley'a obliczonych na danych testowych została zobrazowana na Rycinie 15. Najbardziej wpływową cechą była średnia odchylen standardowych RRI wyodrębnionych z 2-minutowych segmentów (SDNNI2), jednak również dane demograficzne (wiek i BMI) oraz nieliniowe cechy HRV miały znaczący wpływ na wyniki predykcji. W przypadku cech przyczynowych największy wpływ na predykcję miała wartość przyczynowości od sygnału oddechowego do tachogramu, oceniona za pomocą regresora Theil-Sen ( $Resp \rightarrow RR_{TSR}$ ). Dla cech z domeny informacyjnej była to wartość symbolicznej entropii transferowej (ang. symbolic transfer entropy) w tym samym kierunku ( $Resp \rightarrow RR_{STE}$ ).

Tabela 8. Wartości metryk uzyskane z walidacji leave-one-out.

| MAPE [%] | MAE [s] | RMSE [s] | $R^2$ | R Pearsona |
|----------|---------|----------|-------|------------|
| 17,4     | 129,1   | 170,3    | 0,52  | 0,74       |



Rycina 14. Wykres rzeczywistego i przewidywanego czasu trwania CPET. Przerywana czarna linia reprezentuje funkcję  $y=x$ .



Rycina 15. Wartości Shapley'a uzyskane na danych testowych. Nazwy cech są wyjaśnione w [92].

### 6.3. Dyskusja

W wykonanym badaniu przedstawiono predykcję czasu trwania CPET na podstawie parametrów zebranych podczas spoczynku w pozycji leżącej z umiarkowaną dokładnością, przy MAPE wynoszącym 17,4%. Dla zastosowanej metody wyższe przewidywane wartości uzyskiwano dla uczestników starszych, z większym BMI, wyższymi wartościami SDNNI2 lub Fuzzy Entropy oraz niższymi wartościami SDANN2 lub VHF. W przeciwieństwie do rezultatów zaprezentowanych w [71], gdzie wartości  $VO_{2peak}$  były przewidywane na podstawie submaksymalnego CPET, cechy związane z oddechem nie miały kluczowego znaczenia dla wyników modelu. Najbardziej wpływowym parametrem oddechowym był 25-centyl względnej wartości TV, którego wysokie wartości wskazywały na krótszy przewidywany czas trwania próby. Cechy z domeny przyczynowej i informacyjnej również okazały się cennymi predyktorami wyników: wzrost wartości  $Resp \rightarrow RR_{TSR}$  był związany ze wzrostem przewidywanego czasu testu, podczas gdy dla najbardziej wpływowej cechy informacyjnej  $Resp \rightarrow RR_{STE}$ , zależność była odwrotna. Cechy te, parametryzujące zależności krążeniowo-oddechowe, mogą być interpretowane jako liczbowe określenie zjawiska niemiaryowości oddechowej [73], a ich istotność sugeruje związek między sprawnością fizyczną a interakcją między układami sercowym i oddechowym.

Przewidywanie wydolności fizycznej jest istotne nie tylko dla sportowców, ale również dla pacjentów z chorobami układu krążenia [93] oraz dla oceny wykonywania zadań fizycznych [94]. W zaprezentowanej pracy ważną zaletą zastosowanego podejścia do predykcji wydolności jest użycie pneumografii impedancyjnej. Jest to prostsza i tańsza alternatywa dla zaawansowanych systemów analizy gazów, które wymagają wykorzystania specjalnej maski zakłócającej naturalny przepływ wdychanego i wydychanego powietrza.

Wyniki przedstawione w tym badaniu są porównywalne pod względem uzyskanego wskaźnika  $R^2$  z innymi badaniami, które miały na celu przewidywanie rezultatów w wyścigu kolarskim [95] lub wyniku sześciominutowego testu marszu [93]. Uzyskane rezultaty są suboptymalne, a ich praktyczne zastosowanie wymagałoby dalszego modelowania, w szczególności bazującego na większej liczbie danych. Głównym ograniczeniem badania jest stosunkowo mała liczba uczestników z perspektywy treningu modeli uczenia maszynowego. Ponadto, CPET przeprowadzono na cykloergometrze, co nie odzwierciedla głównej aktywności podejmowanej przez badanych. Badanie przeprowadzono wyłącznie na mężczyznach, co z kolei ogranicza możliwość generalizacji wyników na całą populację. Warto również zauważyć, że różne



metody pozyskiwania sygnałów oddechowych, takie jak pomiar przepływu powietrza czy ekstrakcja sygnału oddechowego z sygnału EKG, mogą wpłynąć na kwantyfikację CRC [95], a co za tym idzie na wynik predykcji, co również mogłoby podlegać dalszym badaniom.

Zaprezentowane badanie pokazuje możliwość przewidywania czasu trwania maksymalnych testów wysiłkowych u młodych piłkarzy poprzez wykorzystanie kombinacji cech krążeniowo-oddechowych związanych z HRV, cechami oddechowymi oraz parametrami z domeny przyczynowej i informacyjnej w połączeniu z technikami uczenia maszynowego. Jednakże, aby zwiększyć dokładność i zastosowanie modelu, konieczne są dalsze badania z większą liczbą próbek i zróżnicowanymi aktywnościami fizycznymi.

## 7. Machine Learning Classification of Pediatric Health Status Based on Cardiorespiratory Signals with Causal and Information Domain Features Applied—An Exploratory Study

### 7.1. Materiały i metody

Ostatnia praca wchodząca w skład prezentowanego cyklu miała na celu ewaluację jakości klasyfikacji pacjentów pediatrycznych w zależności od stanu zdrowia na podstawie parametrów krążeniowo-oddechowych przy użyciu uczenia maszynowego, zbadanie wpływu parametrów z domeny przyczynowej i informacyjnej na jakość predykcji oraz ustalenie preferowanego zestawu cech, które mogłyby zostać wykorzystane do dalszego rozwoju w bardziej wyspecjalizowanych zadaniach klasyfikacyjnych lub regresyjnych, związanych z oceną postępów jednostki w trakcie treningu lub rehabilitacji, bądź diagnozowaniem określonych stanów zdrowia. Wszystkie osoby uczestniczące w badaniu poddano rejestracji sygnałów EKG oraz pneumografii impedancyjnej przez co najmniej 5 minut w spoczynku, w pozycji leżącej na plecach, przy użyciu urządzenia Pneumonitor. Uczestnicy badania zostali przydzieleni do odpowiedniej grupy na podstawie następujących kryteriów:

- Grupa *Cardiac* – osoby z problemem kardiologicznym w trakcie leczenia.
- Grupa *Healthy* – osoby bez problemów kardiologicznych, prowadzące siedzący tryb życia lub podejmujące rekreacyjną aktywność fizyczną, zgodnie z klasyfikacją przedstawioną przez McKay’a [96].
- Grupa *Sport* – wytrenowani młodzi sportowcy [96], [97], trenujący piłkę nożną, zrzeszeni w klubie sportowym, z co najmniej 3-letnim doświadczeniem treningowym i regularnie trenujący ~3 razy w tygodniu w celu rywalizacji.

Grupa sportowców charakteryzowała się średnim doświadczeniem treningowym wynoszącym  $5,82 \pm 1,19$  lat i maksymalnym pobraniem tlenu ( $VO_{2max}$ ) na poziomie  $46,55 \pm 4,42$  mL/kg/min. Wiek, waga, wzrost i BMI uczestników z poszczególnych grup porównano przy użyciu testu Kruskala-Wallisa ze względu na rozkład danych inny niż normalny. Pełne informacje na temat demografii badanych grup zostały przedstawione w Tabeli 9. Badanie otrzymało aprobatę dwóch komisji bioetycznych. Zebrano pisemną zgodę na udział w badaniu od opiekunów prawnych osób niepełnoletnich oraz bezpośrednio od uczestników, jeśli byli oni starsi niż 16 lat.

Tabela 9. Charakterystyka grupy badanej z podziałem uwzględniającym stan zdrowia, przedstawiona jako wartość średnia  $\pm$  odchylenie standardowe.

|                          | <i>Cardiac</i>   | <i>Healthy</i>   | <i>Sport</i>     | Wszyscy          |
|--------------------------|------------------|------------------|------------------|------------------|
| N                        | 29               | 62               | 44               | 135              |
| Chłopcy/dziewczynki      | 20/9             | 33/29            | 44/0             | 97/38            |
| Wiek [lata]              | 13,1 $\pm$ 3,5   | 11,0 $\pm$ 2,2   | 13,3 $\pm$ 1,4   | 12,2 $\pm$ 2,6   |
| Masa [kg]                | 57,1 $\pm$ 21,0  | 43,5 $\pm$ 12,1  | 57,2 $\pm$ 13,6  | 50,9 $\pm$ 16,4  |
| Wzrost [cm]              | 160,4 $\pm$ 17,2 | 151,2 $\pm$ 13,1 | 169,4 $\pm$ 12,7 | 159,1 $\pm$ 16,0 |
| BMI [kg/m <sup>2</sup> ] | 21,3 $\pm$ 4,4   | 18,7 $\pm$ 3,2   | 19,6 $\pm$ 2,5   | 19,5 $\pm$ 3,4   |

Sygnały EKG i IP zarejestrowano z częstotliwością próbkowania 250 Hz. Sygnał IP poddano filtracji pasmowo-przepustowej z częstotliwościami odcięcia 0,05 i 0,67 Hz, co odpowiada zakresowi od 3 do 40 oddechów na minutę. Z sygnału EKG, na podstawie automatycznej detekcji i ręcznej korekty przeprowadzonej przez doświadczonego lekarza, wyodrębniono interwały RR, a następnie dokonano interpolacji sześciennych tych interwałów w celu uzyskania tachogramu z taką samą częstotliwością próbkowania jak sygnał oddechowy. Częstotliwość próbkowania obu sygnałów zredukowano następnie do 25 Hz, aby zmniejszyć złożoność obliczeniową. Stacjonarność RRi potwierdzono przy użyciu testu Phillipsa-Perrona dla każdej osoby badanej. Z tak przygotowanych sygnałów oraz szeregów RRi obliczono trzy typy parametrów krążeniowo-oddechowych: HRV (w dziedzinie czasu, częstotliwości oraz parametry nieliniowe), parametry oddechowe oraz parametry z dziedziny przyczynowej i informacyjnej. Parametry HRV obliczono za pomocą pakietu *Neurokit2* [9], z wyjątkiem parametrów symboliki dynamicznej, które zostały wyliczone za pomocą własnej implementacji. Z sygnału oddechowego wyznaczono charakterystyki statystyczne, takie jak częstość oddechów (RespRate), względna objętość oddechowa (TV) oraz stosunek czasu wdechu do wydechu. W kontekście zależności przyczynowych między sygnałem opisującym czynność układu krążenia i sygnałem oddechowym, obliczono parametry bazując na GC, opisanego wcześniej pakietu *nonlincausality* z zastosowaniem różnych modeli ML, KGC oraz lsNGC. Parametry z dziedziny informacyjnej oparto głównie na analizie entropii oraz na prostych statystykach, takich jak najwyższy współczynnik korelacji Pearsona między sygnałami, dla opóźnienia czasowego między -1 a 1 sekundą. W analizie wykorzystano tylko te cechy, dla których można było obliczyć wartości liczbowe dla wszystkich uczestników. W rezultacie dla każdej osoby obliczono łącznie 157 cech, w tym 5 demograficznych (wiek,

waga, wzrost, płeć, BMI), 102 kardiologicznych, 18 oddechowych i 32 z dziedziny przyczynowej i informacyjnej.

Na podstawie wymienionych parametrów utworzono cztery zbiory danych (oznaczone jako D1-D4) według różnych typów cech, które posłużyły jako dane wejściowe do modeli uczenia maszynowego zgodnie z Tabelą 10. Jako etykiety do trenowania modeli ML zastosowano informacje o przypisanej grupie.

Tabela 10. Informacja o cechach zawartych w zestawach danych D1-D4, gdzie „+” oznacza obecność danego rodzaju parametrów.

| Zestaw danych | Dane demograficzne | Parametry kardiologiczne | Parametry oddechowe | Parametry z domeny przyczynowej i informacyjnej |
|---------------|--------------------|--------------------------|---------------------|-------------------------------------------------|
| D1            | +                  | +                        |                     |                                                 |
| D2            | +                  | +                        | +                   |                                                 |
| D3            | +                  | +                        | +                   | +                                               |
| D4            |                    | +                        | +                   | +                                               |

Po wykonanej analizie dla zestawów cech D1-D4 utworzono dwa nowe zestawy cech D5 i D6, składające się z 35 najistotniejszych cech określonych na podstawie wartości Shapely’a z zestawów odpowiednio D3 i D4. W analizie wykorzystano wiele popularnych algorytmów uczenia maszynowego, w tym regresję logistyczną (z regularyzacją Ridge i Lasso), drzewa decyzyjne, maszynę wektorów nośnych, lasy losowe, wzmocnienia gradientowe (ang. gradient boosting), naiwny klasyfikator Bayesa, algorytmu K-najbliższych sąsiadów, algorytm AdaBoost, XGBoost oraz perceptron wielowarstwowy. Dla każdego algorytmu zastosowano optymalizację hiperparametrów. Do sprawdzenia jakości klasyfikacji zastosowano 10-krotną walidację krzyżową. W każdej iteracji algorytmu walidacji obliczano następujące metryki dla zbioru testowego: dokładność, precyzję, czułość, współczynnik F1, współczynnik korelacji Matthews (MCC, z ang. Matthews correlation coefficient) oraz pole pod krzywą charakterystyki operacyjnej odbiornika (ROC, z ang. receiver operating characteristic) (AUC, z ang. area under curve). Średnie wartości metryk z walidacji krzyżowej traktowano jako ostateczną miarę dla algorytmu i zestawu cech. W celu zwiększenia zbioru treningowego zastosowano technikę nadpróbkiowania mniejszości syntetycznych (SMOTE, z ang. synthetic minority oversampling technique) na zbiorze treningowym podczas każdej iteracji walidacji

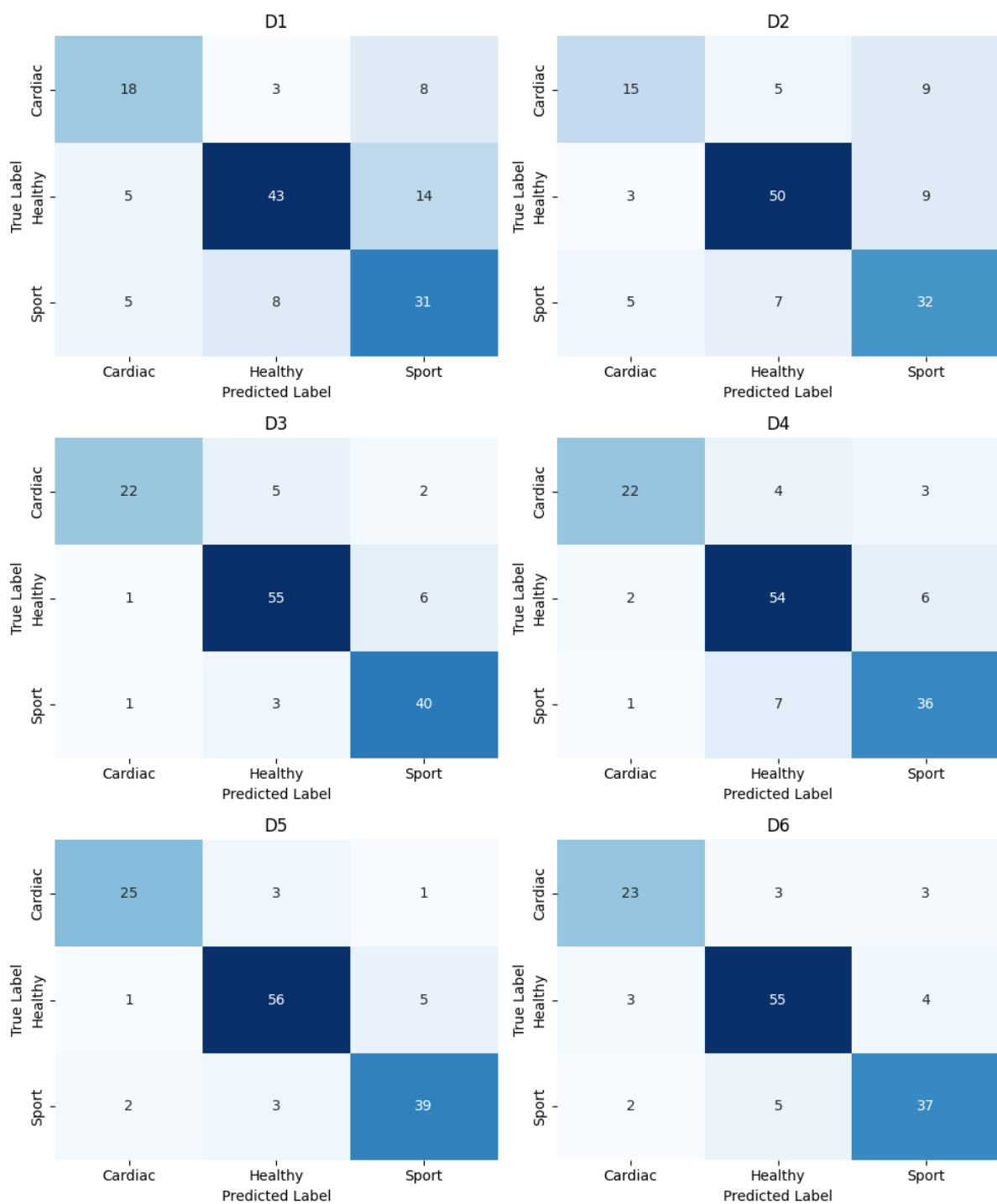
krzyżowej. Dla każdego zestawu danych wybrano najlepszy algorytm na podstawie najwyższej wartości dokładności, a jego wyniki poddano dalszej analizie. Metryki z poszczególnych iteracji walidacji krzyżowej porównano między zestawami danych przy użyciu testu Wilcozona dla par obserwacji, aby określić, czy włączenie określonych typów cech poprawiło wydajność klasyfikacji. W celu zbadania istotności poszczególnych cech i ich wpływu na wyniki otrzymywane z modelu, wykorzystano narzędzia XAI dla zbioru danych, który osiągnął najlepsze wyniki pod względem dokładności. Zbadano istotność zmiennych na poziomie modelu w oparciu o permutacje oraz wartości Shapleya do oceny wpływu zmiennych na wynik dla poszczególnych osób badanych. Założono poziom istotności na poziomie 0,05. Analizę przeprowadzono przy użyciu języka Python w wersji 3.10.8.

## 7.2. Wyniki

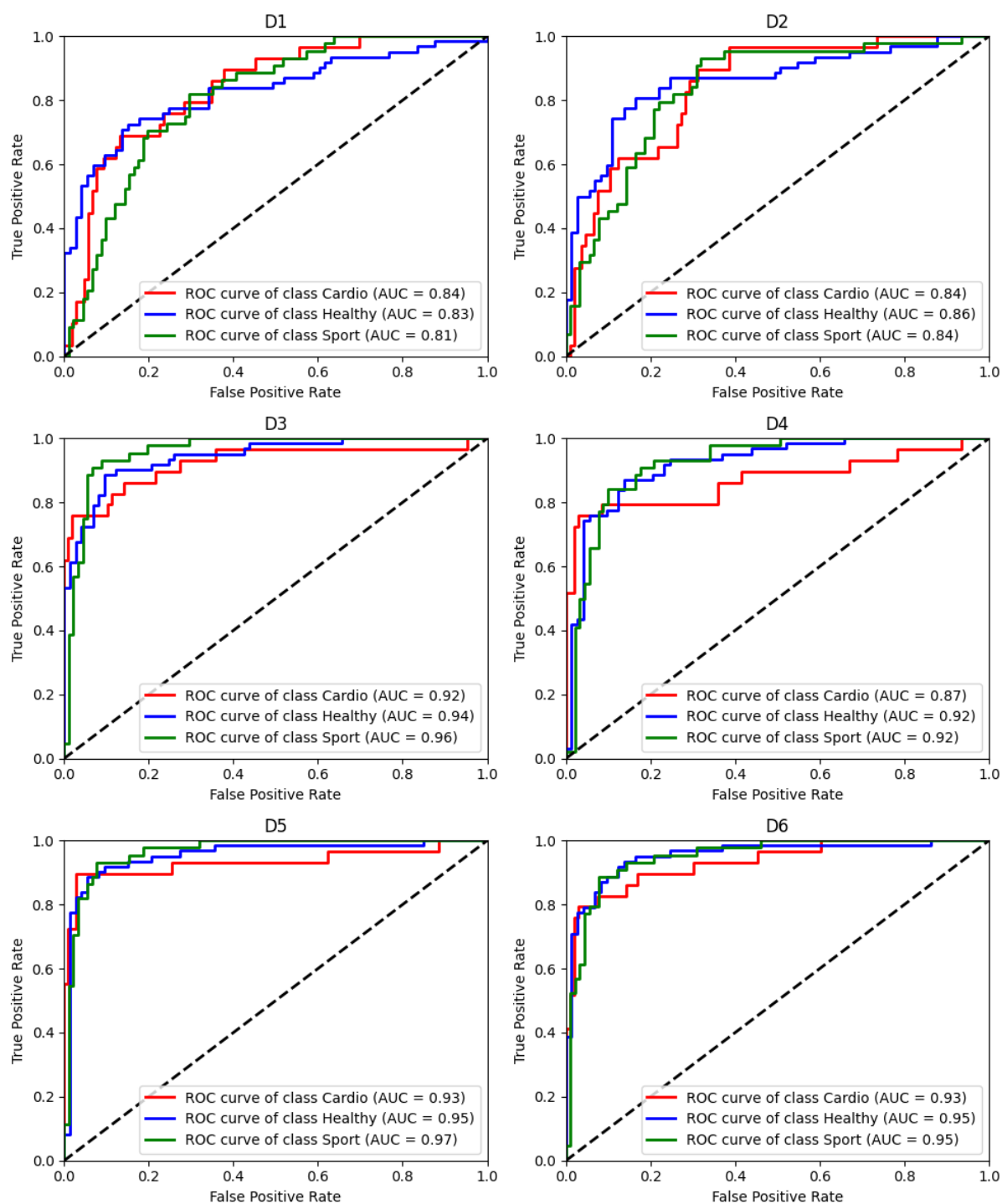
Metryki dla każdego zestawu danych uzyskane z klasyfikacji przez modele, które uzyskały najwyższą dokładność zostały zaprezentowane w Tabeli 11. Najbardziej korzystne wartości wszystkich metryk otrzymano dla zestawu danych D5, który składał się z 35 najistotniejszych cech demograficznych, krążeniowych, oddechowych oraz z domeny przyczynowej i informacyjnej. Wykres prezentujący zbiorcze macierze pomyłek otrzymane w wyniku 10-krotnej walidacji krzyżowej oraz krzywe ROC zaprezentowano odpowiednio na Rycinie 16 i 17.

Tabela 11. Średnia  $\pm$  odchylenie standardowe metryk uzyskanych z 10-krotnej walidacji krzyżowej dla danego algorytmu uczenia maszynowego z zastosowaniem techniki SMOTE, ze strategią nadpróbkowania zaprezentowaną jako liczba próbek treningowych odpowiednio dla grup *Cardiac* / *Healthy* / *Sport*.

|                | D1                | D2                  | D3                | D4                | D5                | D6                |
|----------------|-------------------|---------------------|-------------------|-------------------|-------------------|-------------------|
| Dokładność [%] | 68,3 $\pm$ 8,1    | 72,0 $\pm$ 8,7      | 86,7 $\pm$ 8,4    | 83,1 $\pm$ 11,5   | 89,1 $\pm$ 9,6    | 85,3 $\pm$ 10,0   |
| AUC            | 83,2 $\pm$ 6,7    | 85,2 $\pm$ 6,5      | 94,2 $\pm$ 5,2    | 90,1 $\pm$ 8,3    | 95,8 $\pm$ 5,7    | 94,1 $\pm$ 5,7    |
| Czułość [%]    | 67,6 $\pm$ 9,6    | 68,1 $\pm$ 10,9     | 85,1 $\pm$ 9,6    | 81,6 $\pm$ 11,2   | 88,9 $\pm$ 10,2   | 84,0 $\pm$ 9,9    |
| Precyzja [%]   | 66,9 $\pm$ 12,7   | 70,8 $\pm$ 13,0     | 89,5 $\pm$ 8,6    | 85,6 $\pm$ 11,3   | 89,6 $\pm$ 11,1   | 86,9 $\pm$ 10,6   |
| MCC            | 0,516 $\pm$ 0,132 | 0,566 $\pm$ 0,140   | 0,801 $\pm$ 0,133 | 0,742 $\pm$ 0,180 | 0,835 $\pm$ 0,151 | 0,778 $\pm$ 0,152 |
| Miara F1       | 0,659 $\pm$ 0,109 | 0,676 $\pm$ 0,114   | 0,856 $\pm$ 0,095 | 0,823 $\pm$ 0,111 | 0,885 $\pm$ 0,109 | 0,843 $\pm$ 0,102 |
| Algorytm       | XGBoost           | Logistic Regression | Gradient Boosting | Gradient Boosting | Gradient Boosting | Gradient Boosting |
| SMOTE          | 200 / 200 / 200   | 200 / 200 / 150     | 200 / 200 / 200   | 200 / 200 / 200   | 200 / 200 / 200   | 200 / 200 / 200   |



Rycina 16. Skumulowane macierze pomyłek otrzymane z 10-krotnej walidacji krzyżowej na zbiorach testowych dla każdego zbioru danych.

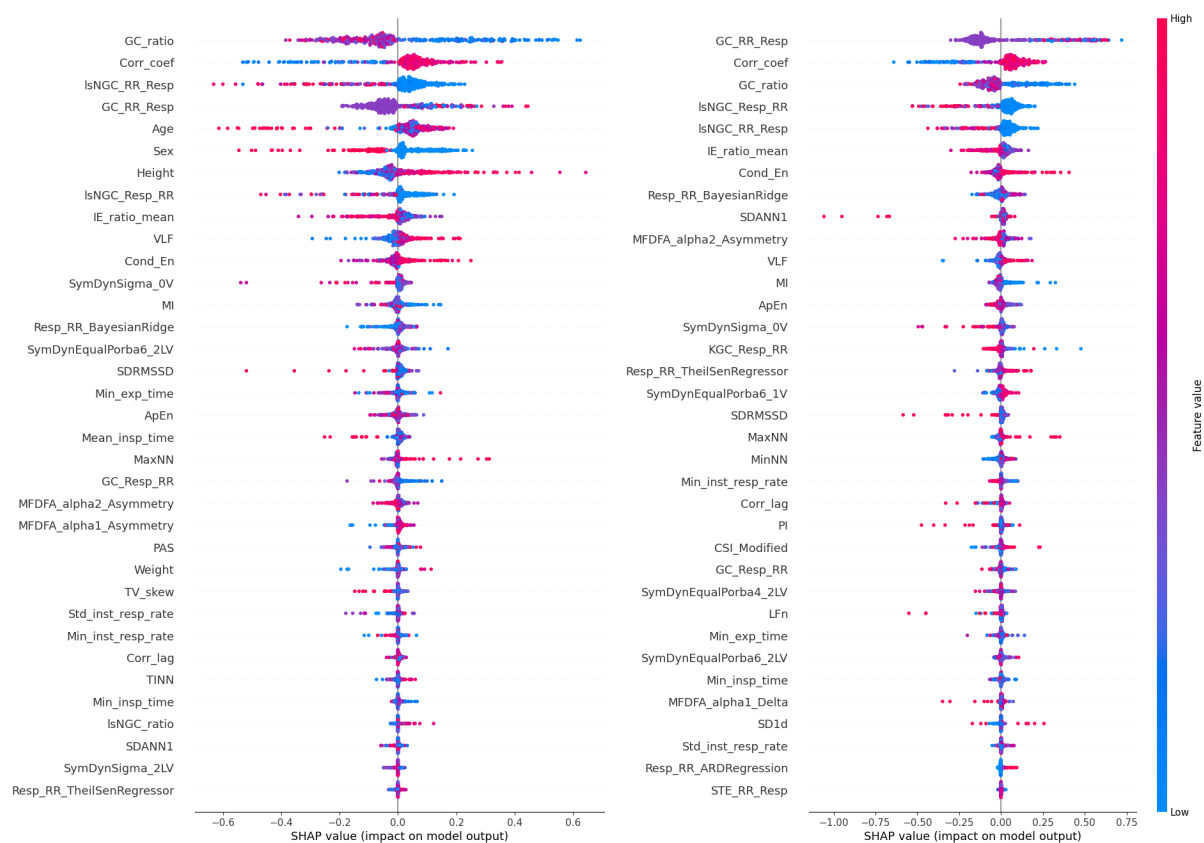


Rycina 17. Krzywe ROC wraz z wartościami AUC dla każdej klasy na podstawie podejścia jeden przeciw wszystkim (ang. one vs all) dla każdego zbioru danych.

Wartości Shapley'a uzyskane z analizy XAI dla D5 i D6 zostały zaprezentowane na Rycinie 18. Cechy mające największe znaczenie, określone na podstawie wartości Shapleya, to odpowiednio: stosunek GC z sygnału oddechowego do tachogramu ( $\text{Resp} \rightarrow \text{RR}$ ) oraz z tachogramu do sygnału oddechowego ( $\text{RR} \rightarrow \text{Resp}$ ), najwyższe wartości współczynnika korelacji Pearsona między sygnałem oddechowym a sygnałem opisującym czynność układu



krążenia dla opóźnień od -1 do 1 sekundy (CorrCoef), wartość przyczynowości RR→Resp dla metody lsNGC oraz RR→Resp dla metody GC. Analiza istotności zmiennych oparta na permutacjach wykazała, że te cechy odgrywają kluczową rolę w odróżnianiu osób z grupy *Healthy* od grupy *Sport* (z wyjątkiem cechy CorrCoef w przypadku zestawu danych D5). Natomiast w przypadku rozróżniania grupy *Cardiac* od pozostałych grup, największe znaczenie miały cechy CorrCoef oraz lsNGC RR→Resp. Test Kruskala-Wallisa wykazał statystycznie istotne różnice między grupami pod względem wieku, wagi, wzrostu i wskaźnika BMI.



Rycina 18. Wartości Shapleya uzyskane dla danych testowych w 10-krotnej walidacji krzyżowej dla zestawów D5 (po lewej) i D6 (po prawej).

### 7.3. Dyskusja

W tym badaniu sprawdzono jakość klasyfikacji dzieci i młodzieży do jednej z trzech grup w zależności od stanu zdrowia na podstawie parametrów krążeniowo-oddechowych. Uwzględnienie 35 najbardziej istotnych cech spośród wszystkich parametrów (D5), w tym parametrów z domeny przyczynowej i informacyjnej, pozwoliło na uzyskanie najlepszych wyników pod względem każdej zastosowanej metryki oraz pod względem kształtu krzywych ROC przy średniej dokładności równej 89,1%. Niewiele gorsze wyniki zostały otrzymane dla zestawu D6, który zawierał wyłącznie najistotniejsze parametry krążeniowo-oddechowe bez

informacji demograficznych, co pozwala potencjalnie na zastosowanie prezentowanej metody bez potrzeby dodatkowych pomiarów tych cech. Otrzymanie statystycznie istotnie lepszych wyników dla zbiorów danych zawierających cechy z domeny przyczynowej i informacyjnej względem zestawów danych niezawierających tych cech świadczą o ich przydatności i powiązaniu z szeroko pojętym w tym badaniu stanem zdrowia. Co więcej, w przypadku błędnej klasyfikacji, osoby z grupy *Sport* były częściej klasyfikowane jako *Healthy* niż jako *Cardiac*. Analogicznie pacjenci z grupy *Cardiac* byli częściej przypisywani do grupy *Healthy* niż do grupy *Sport*. Świadczy to prawdopodobnie o większym oddaleniu w wielowymiarowej przestrzeni parametrów między osobami z grupy *Cardiac* i *Sport* oraz mniejszym dystansie między grupą *Healthy* i *Sport* niż *Healthy* i *Cardiac*, jako że przy nieprawidłowej klasyfikacji osoby zdrowe były częściej oznaczane jako należące do grupy *Sport*.

Uwzględnienie parametrów oddechowych w modelowaniu pozwoliło na uzyskanie wyższej dokładności, jednakże w odróżnieniu do zaprezentowanej wcześniejszej pracy [71] różnica między wynikami uzyskanymi dla zestawu danych zawierającego parametry kardiologiczne i zestawu rozszerzonego o parametry oddechowe (D1 i D2) nie była istotna statystycznie. Istotna poprawa predykcji przynależności do odpowiedniej grupy została uzyskana dopiero dla zestawów danych zawierających parametry z domeny przyczynowej i informacyjnej (D3-D6). Przyczyna tej poprawy może być związana z dodatkową informacją diagnostyczną o stanie zdrowia zawartą w parametrach związanych z CRC. Na podstawie parametrów HRV można uzyskać informacje o wpływie w sensie przyczynowym oddechu na pracę serca (głównie za pomocą parametrów z dziedziny częstotliwości) [13], jednakże pełniejsza informacja o CRC może być uzyskana wyłącznie przy uwzględnieniu zarówno sygnału opisującego czynność układu krążenia jak i sygnału oddechowego.

Pomimo prężnego rozwoju zastosowania metod sztucznej inteligencji w wielu gałęziach medycyny, wciąż wykorzystanie parametrów związanych z CRC w rozwiązaniach opartych na uczeniu maszynowym pozostaje tematem wymagającym dalszych badań. Obecnie dostępne badania wykazały, że CRC odgrywa ważną rolę w medycynie sportowej [37], [98], np. umożliwiając różnicowanie między sportowcami a osobami nieuprawiającymi sportu [99], jako wczesny marker dysfunkcji autonomicznej układu sercowego u pacjentów z cukrzycą typu 2 [100] oraz w badaniach nad obturacyjnym bezdechem sennym [101], [102]. Zjawisko sprzężenia krążeniowo-oddechowego znalazło swoje zastosowanie również w personalizacji treningu oddechowego [103]. Te właściwości CRC wraz z uzyskanymi wynikami podkreślają potencjał tych parametrów do wykorzystania w szczególności w połączeniu z metodami

uczenia maszynowego. Dzięki technikom ML istnieje możliwość uproszczenia wielu parametrów kardiologicznych, oddechowych i przyczynowych oraz informacyjnych, które mogą być dla lekarza trudne do interpretacji ze względu na ich mnogość, do pojedynczego wyniku modelu uczenia maszynowego jako nowego parametru, który byłby łatwiejszy do interpretacji.

Zastosowanie narzędzi XAI w omawianym badaniu pozwoliło na głębsze zrozumienie, które cechy mają największy wpływ na wyniki modeli oraz potwierdziło istotność parametrów przyczynowych i z domeny informacyjnej w kontekście badanego zadania klasyfikacji, zarówno pod względem wartości Shapley'a jak i istotności zmiennych opartych na permutacjach. Wyniki wykazały większą istotność cech związanych z zależnością  $RR \rightarrow Resp$  niż cech parametryzujących zależność  $Resp \rightarrow RR$ . Jak opisano w sekcji 5.3., pomimo nieintuicyjnego charakteru tej współzależności w kontekście zjawiska niemiarowości oddechowej, przyczyną takiego stanu może być przesunięcie lokalnych ekstremów sygnału RR względem TV oraz fizjologiczna dwukierunkowa współzależność między pracą układu oddechowego i krążeniowo-oddechowego.

Brak kobiet w grupie sportowców, różne licznosci poszczególnych grup, różnice w parametrach demograficznych oraz zróżnicowanie schorzeń wśród grupy *Cardiac* stanowią ograniczenia przeprowadzonego badania. Większa liczba osób z bardziej zrównoważonym rozkładem parametrów demograficznych i większą jednorodnością problemów zdrowotnych mogłaby korzystnie wpłynąć na proces trenowania modeli uczenia maszynowego i na uzyskane metryki.

Niniejsze badanie nie tylko prezentuje klasyfikacje młodych osób ze względu na różny stan zdrowia, co może być przydatne przy wstępnej ocenie pacjentów, ale również zwraca uwagę na istotność parametrów związanych ze zjawiskiem sprzężenia krążeniowo-oddechowego. Szczególnie istotna jest tutaj wykazana znacząca poprawa w jakości klasyfikacji, uzyskana dla zestawów obejmujących cechy związane z CRC względem zarówno zestawu D1, który wykorzystuje popularnie stosowane parametry HRV, jak i zestawu D2, który dodatkowo wykorzystuje parametry oddechowe. Dodatkowo dzięki zastosowaniu narzędzi XAI wyłonione zostały najistotniejsze parametry, przy użyciu których udało się uzyskać jeszcze lepsze rezultaty. Wyniki te mogą zostać wykorzystane w dalszych badaniach i tworzeniu modeli predykcyjnych do monitorowania zmian parametrów krążeniowo-oddechowych u osób podczas treningu sportowego lub rehabilitacji kardiologicznej, a także w kontekście CRF i konkretnych schorzeń kardiologicznych. Dalsze badania związane z cechami przyczynowymi

i z domeny informacyjnej parametryzującymi CRC, mogą również skupiać się na utworzeniu norm oraz siatek centylowych dla owych parametrów, co pozwoliłoby na lepszą ich interpretację w praktyce klinicznej.

## 8. Podsumowanie

Przeprowadzone badania, opisane w niniejszej rozprawie, pozwoliły na opracowanie nowych metod analizy zależności przyczynowych oraz zbadanie istotności aktywności oddechowej i parametrów z domeny przyczynowej oraz informacyjnej dla sygnałów krążeniowo-oddechowych w kontekście stanu zdrowia osoby badanej.

Wyniki uzyskane w pracach [71], [75] w grupie odpowiednio 327 i 135 osób ze średnim bezwzględnym błędem procentowym równym 10,51% oraz dokładnością wynoszącą 89,1% pozwalają stwierdzić, iż udało się potwierdzić postawioną we wstępie hipotezę badawczą.

W ramach niniejszej rozprawy doktorskiej, dokonano znaczących postępów w zakresie analizy współzależności krążeniowo-oddechowych, wprowadzając nowe metody i narzędzia badawcze oraz walidując ich użyteczność w kontekście oceny stanu zdrowia. Opisane wyżej publikacje rozszerzają obecny stan wiedzy w dziedzinie inżynierii biomedycznej o nowatorskie podejścia do analizy sygnałów krążeniowo-oddechowych, ze szczególnym uwzględnieniem metod uczenia maszynowego, analizy przyczynowości oraz wyjaśnialnej sztucznej inteligencji. W niniejszym cyklu prac przedstawiono praktyczne zastosowanie algorytmów uczenia maszynowego w predykcji parametrów wydolnościowych, jak również wprowadzono nowe metody oparte na nieliniowej analizie przyczynowości, umożliwiające identyfikację złożonych relacji między układem oddechowym a krążeniowym, niedostępnych dla tradycyjnych modeli liniowych. Ponadto wykazano diagnostyczną wartość zjawiska sprzężenia krążeniowo-oddechowego, które zostało opisane liczbowo z wykorzystaniem parametrów z domeny przyczynowej i informacyjnej. Wyniki prezentowanych badań otwierają nowe możliwości dla personalizacji opieki zdrowotnej i zastosowań w rehabilitacji oraz ocenie zdolności wysiłkowej. Zaprezentowane wyniki wzbogaciły literaturę naukową o dowody na istotność uwzględnienia parametrów związanych z układem oddechowym oraz przede wszystkim, tych określających współzależności między układem krążeniowym i oddechowym w analizie sygnałów fizjologicznych.

W ocenie Autora główne osiągnięcia pracy obejmują:

- Przeprowadzenie analizy predykcji  $VO_{2peak}$  na podstawie różnych zestawów danych z submaksymalnego testu wysiłkowego wraz z badaniem istotności cech oddechowych na jakość predykcji.

- Podkreślenie potrzeby poszukiwania nowych parametrów krążeniowo-oddechowych, które umożliwiłyby uzyskanie dokładniejszej informacji diagnostycznej, istotnej z klinicznego punktu widzenia.
- Opracowanie pakietu *nonlincausality* do badania nieliniowych zależności przyczynowych z wykorzystaniem modeli uczenia maszynowego, który może zostać zastosowany w analizie sygnałów krążeniowo-oddechowych.
- Wykonanie analizy współzależności w sygnałach krążeniowo-oddechowych u pediatrycznych pacjentów kardiologicznych, z wykorzystaniem liniowych i nieliniowych metod analizy przyczynowości.
- Porównanie różnych możliwości parametryzacji zależności przyczynowych w sygnałach krążeniowo-oddechowych.
- Przeprowadzenie predykcji czasu trwania maksymalnego testu wysiłkowego u młodych piłkarzy, z wykorzystaniem szerokiego zestawu cech krążeniowo-oddechowych, obejmującego cechy przyczynowe i informacyjne, jako prezentacja zastosowania zróżnicowanego zestawu cech krążeniowo-oddechowych do oceny wydolności.
- Klasyfikację stanu zdrowia na podstawie parametrów krążeniowo-oddechowych, z uwzględnieniem cech z domeny przyczynowej i informacyjnej wraz z badaniem istotności cech oddechowych oraz z domeny przyczynowej i informacyjnej na jakość predykcji.
- Wykazanie istotnego wpływu cech związanych z CRC na jakość predykcji, a także wyodrębnienie zestawu najistotniejszych cech o potencjalnie największej wartości diagnostycznej.

Zaprezentowane prace wskazują również na przydatność urządzeń integrujących w sobie jednoczesny pomiar kilku sygnałów takich jak np. Pneumonitor. Dzięki symultanicznej akwizycji zarówno sygnału EKG, jak i pneumografii impedancyjnej umożliwia on nie tylko analizę parametrów kardiologicznych i oddechowych, co zazwyczaj wymaga zastosowania osobnych aparatur, ale również współzależności między rejestrowanymi sygnałami, co pozwala na uzyskanie dodatkowej informacji diagnostycznej. Ponadto, integracja pomiarów w jednym urządzeniu może zwiększyć komfort pacjenta i uprościć procedury diagnostyczne, zmniejszając potrzebę korzystania z wielu różnych urządzeń jednocześnie, a także minimalizuje ryzyko błędów wynikających z niespójności pomiarów.

Ograniczeniem w pracy [71] był brak dostępu do surowych sygnałów EKG/interwałów RR i krzywych oddechowych, co uniemożliwiło obliczenie bardziej zaawansowanych parametrów z domeny informacyjnej i przyczynowej, a także konieczność dużego ograniczenia liczby próbek oraz nierównomierny rozkład płci uczestników. W publikacji prezentującej pakiet *nonlincausality* [72] pewnym ograniczeniem było zastosowanie wyłącznie sygnałów symulowanych, natomiast w [73] stosunkowo mała i jednorodna grupa badawcza oraz brak grupy kontrolnej zdrowych osób, który uniemożliwił ocenę przydatności obliczonych wartości przyczynowości jako biomarkera schorzeń kardiologicznych. Głównym ograniczeniem badania zaprezentowanego w [74] była niewielka z punktu widzenia uczenia maszynowego liczba osób badanych, a także fakt, że testy wysiłkowe przeprowadzono na cykloergometrze, co nie odzwierciedlało głównej aktywności badanych, dodatkowo badanie obejmowało wyłącznie mężczyzn. W pracy [75] ograniczeniami są: brak kobiet w grupie *Sport*, zróżnicowanie liczebności grup oraz parametrów demograficznych, a także heterogeniczność problemów zdrowotnych w grupie kardiologicznej, co mogło negatywnie wpłynąć na efektywność modeli ML.

Uwzględniając powyższe, dalsze badania powinny skupić się na analizie istotności współzależności krążeniowo-oddechowych w kontekście konkretnych schorzeń lub dyscyplin sportowych, co pozwoliłoby na precyzyjne dopasowanie modeli uczenia maszynowego do specyficznych potrzeb diagnostycznych lub treningowych. Ważnym krokiem byłoby również opracowanie norm i siatek centylowych dla parametrów z domeny przyczynowej i informacyjnej, co umożliwiłoby bardziej dokładne porównania międzysobnicze oraz monitorowanie zmian tych parametrów w czasie np. podczas rehabilitacji kardiologicznej lub w trakcie przygotowań do zawodów sportowych. Istotną trudnością w analizie parametrów krążeniowo-oddechowych jest ich mnogość, co utrudnia interpretację przez człowieka. Co za tym idzie kolejnym kierunkiem badań mogłoby być uproszczenie zestawu parametrów krążeniowo-oddechowych poprzez redukcję ich liczby do jednego wskaźnika z wykorzystaniem technik uczenia maszynowego, co znacząco ułatwiłoby interpretację uzyskiwanych wyników. Integracja tych elementów w spójny system mogłaby prowadzić do stworzenia zaawansowanego narzędzia diagnostyczno-monitorującego, które precyzyjnie dokonywałoby pomiaru, parametryzacji, selekcji najistotniejszych w danym zastosowaniu cech oraz prezentowałoby wyniki w przejrzysty sposób, aby efektywnie monitorować stan zdrowia i wydolność danej osoby w czasie.

Podsumowując, niniejsza rozprawa doktorska dostarcza istotnych narzędzi i wiedzy, które mogą znaleźć zastosowanie zarówno w badaniach naukowych, jak i praktyce klinicznej oraz sportowej. Zastosowanie nowatorskich metod analizy sygnałów krążeniowo-oddechowych z uwzględnieniem domeny przyczynowej i informacyjnej oraz wykorzystanie zaawansowanych technik uczenia maszynowego pozwoliło na uzyskanie nowych możliwości diagnostycznych oraz zwiększenie precyzji oceny stanu zdrowia i wydolności. Przeprowadzone badania wskazują na potencjał takich podejść w personalizacji opieki zdrowotnej, monitorowaniu postępów treningowych czy ocenie skuteczności rehabilitacji. Opracowane metody stanowią solidną podstawę do dalszych badań w zakresie analizy sygnałów fizjologicznych, rozwijania bardziej zaawansowanych modeli predykcyjnych oraz tworzenia praktycznych rozwiązań wspierających opiekę nad pacjentami i sportowcami.



## 9. Bibliografia

- [1] H. De Cannière *et al.*, “Wearable monitoring and interpretable machine learning can objectively track progression in patients during cardiac rehabilitation,” *Sensors (Switzerland)*, vol. 20, no. 12, 2020, doi: 10.3390/s20123601.
- [2] A. Nazaret, S. Tonekaboni, G. Darnell, S. Y. Ren, G. Sapiro, and A. C. Miller, “Modeling personalized heart rate response to exercise and environmental factors with wearables data,” *NPJ Digit Med*, vol. 6, no. 1, 2023, doi: 10.1038/s41746-023-00926-4.
- [3] C. Serantoni, G. Zimatore, G. Bianchetti, A. Abeltino, M. De Spirito, and G. Maulucci, “Unsupervised Clustering of Heartbeat Dynamics Allows for Real Time and Personalized Improvement in Cardiovascular Fitness,” *Sensors*, vol. 22, no. 11, 2022, doi: 10.3390/s22113974.
- [4] P. Melillo *et al.*, “Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis,” *PLoS One*, vol. 10, no. 3, 2015, doi: 10.1371/journal.pone.0118504.
- [5] D. Z. H. Levett *et al.*, “Perioperative cardiopulmonary exercise testing (CPET): consensus clinical guidelines on indications, organization, conduct, and physiological interpretation,” *Br J Anaesth*, vol. 120, no. 3, 2018, doi: 10.1016/j.bja.2017.10.020.
- [6] M. D. Fridman, P. Tsoukas, A. Jeewa, R. S. M. Yeung, B. D. Gamulka, and B. W. McCrindle, “Differentiation of COVID-19–Associated Multisystem Inflammatory Syndrome From Kawasaki Disease With the Use of Cardiac Biomarkers,” *Canadian Journal of Cardiology*, vol. 39, no. 6, 2023, doi: 10.1016/j.cjca.2022.11.012.
- [7] M. D. Patel, K. Mariano, T. Dunbar, T. T. Cornell, R. Pun, and B. Haileselassie, “Cardiac Dysfunction Identified by Strain Echocardiography Is Associated With Illness Severity in Pediatric Sepsis,” *Pediatric Critical Care Medicine*, vol. 21, no. 4, 2020, doi: 10.1097/PCC.0000000000002247.
- [8] M. P. Tarvainen, J. P. Niskanen, J. A. Lipponen, P. O. Ranta-aho, and P. A. Karjalainen, “Kubios HRV - Heart rate variability analysis software,” *Comput Methods Programs Biomed*, vol. 113, no. 1, 2014, doi: 10.1016/j.cmpb.2013.07.024.

- [9] D. Makowski *et al.*, “NeuroKit2: A Python toolbox for neurophysiological signal processing,” *Behav Res Methods*, vol. 53, no. 4, 2021, doi: 10.3758/s13428-020-01516-y.
- [10] M. Brennan, M. Palaniswami, and P. Kamen, “Do existing measures of Poincaré plot geometry reflect nonlinear features of heart rate variability?,” *IEEE Trans Biomed Eng*, vol. 48, no. 11, 2001, doi: 10.1109/10.959330.
- [11] A. Delgado-Bonal and A. Marshak, “Approximate entropy and sample entropy: A comprehensive tutorial,” 2019. doi: 10.3390/e21060541.
- [12] T. Penzel, J. W. Kantelhardt, L. Grote, J. H. Peter, and A. Bunde, “Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea,” *IEEE Trans Biomed Eng*, vol. 50, no. 10, 2003, doi: 10.1109/TBME.2003.817636.
- [13] F. Shaffer and J. P. Ginsberg, “An Overview of Heart Rate Variability Metrics and Norms,” 2017. doi: 10.3389/fpubh.2017.00258.
- [14] S. Zeid *et al.*, “Heart rate variability: reference values and role for clinical profile and mortality in individuals with heart failure,” *Clinical Research in Cardiology*, 2023, doi: 10.1007/s00392-023-02248-7.
- [15] T. Pham, Z. J. Lau, S. H. A. Chen, and D. Makowski, “Heart rate variability in psychology: A review of hrv indices and an analysis tutorial,” 2021. doi: 10.3390/s21123998.
- [16] M. B. A. Mol *et al.*, “Heart-rate-variability (HRV), predicts outcomes in COVID-19,” *PLoS One*, vol. 16, no. 10 October, 2021, doi: 10.1371/journal.pone.0258841.
- [17] J. Aftyka, J. Staszewski, A. Dębiec, A. Pogoda-Wesołowska, and J. Żebrowski, “Heart rate variability as a predictor of stroke course, functional outcome, and medical complications: A systematic review,” 2023. doi: 10.3389/fphys.2023.1115164.
- [18] A. Sethi, C. W. Callaway, E. Sejdić, L. Terhorst, and E. R. Skidmore, “Heart Rate Variability Is Associated with Motor Outcome 3-Months after Stroke,” *Journal of Stroke and Cerebrovascular Diseases*, vol. 25, no. 1, 2016, doi: 10.1016/j.jstrokecerebrovasdis.2015.09.005.

- [19] L. Stepanyan and G. Lalayan, "Heart rate variability features and their impact on athletes' sports performance," *Journal of Physical Education and Sport*, vol. 23, no. 8, 2023, doi: 10.7752/jpes.2023.08247.
- [20] E. Mosley and S. Laborde, "A scoping review of heart rate variability in sport and exercise psychology," *Int Rev Sport Exerc Psychol*, 2022, doi: 10.1080/1750984X.2022.2092884.
- [21] S. A. Shah, C. Velardo, A. Farmer, and L. Tarassenko, "Exacerbations in chronic obstructive pulmonary disease: Identification and prediction using a digital health system," *J Med Internet Res*, vol. 19, no. 3, 2017, doi: 10.2196/jmir.7207.
- [22] D. E. O'Donnell, "Ventilatory limitations in chronic obstructive pulmonary disease," 2001. doi: 10.1097/00005768-200107001-00002.
- [23] A. S. Ginsburg, J. L. Lenahan, R. Izadnegahdar, and J. M. Ansermino, "A systematic review of tools to measure respiratory rate in order to identify childhood pneumonia," 2018. doi: 10.1164/rccm.201711-2233CI.
- [24] C. Varon *et al.*, "A Comparative Study of ECG-derived Respiration in Ambulatory Monitoring using the Single-lead ECG," *Sci Rep*, vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-62624-5.
- [25] B. Cairo *et al.*, "Comparison Between ECG-Derived Respiration and Respiratory Flow for the Assessment of Cardiorespiratory Coupling Before and After Cardiopulmonary Exercise Test Protocol," in *Computing in Cardiology*, 2022. doi: 10.22489/CinC.2022.103.
- [26] J. M. Ernst, D. A. Litvack, D. L. Lozano, J. T. Cacioppo, and G. G. Berntson, "Impedance pneumography: Noise as signal in impedance cardiography," *Psychophysiology*, vol. 36, no. 3, 1999, doi: 10.1017/S0048577299981003.
- [27] V. P. Seppä, M. Uitto, and J. Viik, "Tidal breathing flow-volume curves with impedance pneumography during expiratory loading," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2013. doi: 10.1109/EMBC.2013.6610032.
- [28] V. P. Seppä, J. Viik, A. Naveed, J. Väisänen, and J. Hyttinen, "Signal waveform agreement between spirometer and impedance pneumography of six chest band

- electrode configurations,” in *IFMBE Proceedings*, 2009. doi: 10.1007/978-3-642-03885-3\_191.
- [29] J. H. Houtveen, P. F. C. Groot, and E. J. C. De Geus, “Validation of the thoracic impedance derived respiratory signal using multilevel analysis,” *International Journal of Psychophysiology*, vol. 59, no. 2, 2006, doi: 10.1016/j.ijpsycho.2005.02.003.
  - [30] V. P. Seppä, J. Hyttinen, M. Uitto, W. Chrapek, and J. Viik, “Novel electrode configuration for highly linear impedance pneumography,” *Biomedizinische Technik*, 2013, doi: 10.1515/bmt-2012-0068.
  - [31] M. Młyńczak, W. Niewiadomski, M. Zylński, and G. Cybulski, “Assessment of calibration methods on impedance pneumography accuracy,” *Biomedizinische Technik*, 2015, doi: 10.1515/bmt-2015-0125.
  - [32] M. Młyńczak and G. Cybulski, “Flow parameters derived from impedance pneumography after nonlinear calibration based on neural networks,” in *BIOSIGNALS 2017 - 10th International Conference on Bio-Inspired Systems and Signal Processing, Proceedings; Part of 10th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2017*, 2017. doi: 10.5220/0006146800700077.
  - [33] V. P. Seppä, J. Viik, and J. Hyttinen, “Assessment of pulmonary flow using impedance pneumography,” *IEEE Trans Biomed Eng*, vol. 57, no. 9, 2010, doi: 10.1109/TBME.2010.2051668.
  - [34] M. Młyńczak, M. Zylinski, W. Niewiadomski, and G. Cybulski, “Ambulatory Devices Measuring Cardiorespiratory Activity with Motion,” in *BIODEVICES 2017 - 10th International Conference on Biomedical Electronics and Devices, Proceedings; Part of 10th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2017*, 2017. doi: 10.5220/0006111700910097.
  - [35] M. C. Młyńczak, W. Niewiadomski, M. Zylński, and G. P. Cybulski, “Ambulatory impedance pneumography device for quantitative monitoring of volumetric parameters in respiratory and cardiac applications,” in *Computing in Cardiology*, 2014.
  - [36] J. S. Gąsior *et al.*, “Validity of the Pneumonitor for RR intervals acquisition for short-term heart rate variability analysis extended with respiratory data in pediatric cardiac patients,” *Kardiol Pol*, vol. 81, no. 5, 2023, doi: 10.33963/KP.a2023.0070.

- [37] R. M. de Abreu, B. Cairo, and A. Porta, “On the significance of estimating cardiorespiratory coupling strength in sports medicine,” *Frontiers in Network Physiology*, vol. 2, 2023, doi: 10.3389/fnetp.2022.1114733.
- [38] P. G. Katona and F. Jih, “Respiratory sinus arrhythmia: noninvasive measure of parasympathetic cardiac control,” *J Appl Physiol*, 1975, doi: 10.1152/jappl.1975.39.5.801.
- [39] S. Schulz *et al.*, “Cardiovascular and cardiorespiratory coupling analyses: A review,” 2013. doi: 10.1098/rsta.2012.0191.
- [40] C. W. J. Granger, “Investigating Causal Relations by Econometric Models and Cross-spectral Methods,” *Econometrica*, vol. 37, no. 3, p. 424, 1969, doi: 10.2307/1912791.
- [41] C. A. Sims, “Money, Income, and Causality,” *American Economic Review*, vol. 62, no. 4, pp. 540–552, 1972, doi: 10.1126/science.151.3712.867-a.
- [42] A. von Eye, W. Wiedermann, and E. Y. Mun, “Granger Causality-Statistical Analysis Under a Configural Perspective,” *Integr Psychol Behav Sci*, vol. 48, no. 1, 2014, doi: 10.1007/s12124-013-9243-1.
- [43] L. Faes, G. Nollo, and A. Porta, “Information domain approach to the investigation of cardio-vascular, cardio-pulmonary, and vasculo-pulmonary causal couplings,” *Front Physiol*, vol. 2 NOV, 2011, doi: 10.3389/fphys.2011.00080.
- [44] A. Porta, G. Baselli, F. Lombardi, N. Montano, A. Malliani, and S. Cerutti, “Conditional entropy approach for the evaluation of the coupling strength,” *Biol Cybern*, vol. 81, no. 2, 1999, doi: 10.1007/s004220050549.
- [45] M. M. Platiša, N. N. Radovanović, R. Pernice, C. Barà, S. U. Pavlović, and L. Faes, “Information-Theoretic Analysis of Cardio-Respiratory Interactions in Heart Failure Patients: Effects of Arrhythmias and Cardiac Resynchronization Therapy,” *Entropy*, vol. 25, no. 7, 2023, doi: 10.3390/e25071072.
- [46] A. Schumann, B. Fleckenstein, and K. J. Bär, “Nonlinear causal influences assessed by mutual compression entropy,” *Current Directions in Biomedical Engineering*, vol. 2, no. 1, 2016, doi: 10.1515/cdbme-2016-0049.
- [47] B. Cairo, V. Bari, F. Gelpi, B. De Maria, and A. Porta, “Assessing cardiorespiratory interactions via lagged joint symbolic dynamics during spontaneous and controlled

- breathing,” *Frontiers in Network Physiology*, vol. 3, 2023, doi: 10.3389/fnetp.2023.1211848.
- [48] M. G. Rosenblum, L. Cimponeriu, A. Bezerianos, A. Patzak, and R. Mrowka, “Identification of coupling direction: Application to cardiorespiratory interaction,” *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, vol. 65, no. 4, 2002, doi: 10.1103/PhysRevE.65.041909.
- [49] Y. Wang, W. Shi, and C. H. Yeh, “A Novel Measure of Cardiopulmonary Coupling During Sleep Based on the Synchrosqueezing Transform Algorithm,” *IEEE J Biomed Health Inform*, vol. 27, no. 4, 2023, doi: 10.1109/JBHI.2023.3237690.
- [50] J. Li, X. Zhang, W. Shi, and C. H. Yeh, “A novel dynamic cardiorespiratory coupling quantification method reveals the effect of aging on the autonomic nervous system,” *Chaos*, vol. 33, no. 12, 2023, doi: 10.1063/5.0156340.
- [51] M. Shehab *et al.*, “Machine learning in medical applications: A review of state-of-the-art methods,” 2022. doi: 10.1016/j.compbimed.2022.105458.
- [52] N. Hong *et al.*, “State of the Art of Machine Learning-Enabled Clinical Decision Support in Intensive Care Units: Literature Review,” 2022. doi: 10.2196/28781.
- [53] A. P. Susanto, D. Lyell, B. Widyantoro, S. Berkovsky, and F. Magrabi, “Effects of machine learning-based clinical decision support systems on decision-making, care delivery, and patient outcomes: a scoping review,” 2023. doi: 10.1093/jamia/ocad180.
- [54] M. Rana and M. Bhushan, “Machine learning and deep learning approach for medical image analysis: diagnosis to detection,” *Multimed Tools Appl*, vol. 82, no. 17, 2023, doi: 10.1007/s11042-022-14305-w.
- [55] S. Suganyadevi, V. Seethalakshmi, and K. Balasamy, “A review on deep learning in medical image analysis,” *Int J Multimed Inf Retr*, vol. 11, no. 1, 2022, doi: 10.1007/s13735-021-00218-1.
- [56] E. Sajno, S. Bartolotta, C. Tuena, P. Cipresso, E. Pedroli, and G. Riva, “Machine learning in biosignals processing for mental health: A narrative review,” 2023. doi: 10.3389/fpsyg.2022.1066317.
- [57] Q. Xiao *et al.*, “Deep Learning-Based ECG Arrhythmia Classification: A Systematic Review,” 2023. doi: 10.3390/app13084964.

- [58] D. Fraile Navarro *et al.*, “Clinical named entity recognition and relation extraction using natural language processing of medical free text: A systematic review,” 2023. doi: 10.1016/j.ijmedinf.2023.105122.
- [59] E. Hossain *et al.*, “Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review,” 2023. doi: 10.1016/j.compbiomed.2023.106649.
- [60] M. Rosoł, J. S. Gąsior, J. Łaba, K. Korzeniewski, and M. Młyńczak, “Evaluation of the performance of GPT-3.5 and GPT-4 on the Polish Medical Final Examination,” *Sci Rep*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-46995-z.
- [61] J. Kufel *et al.*, “What Is Machine Learning, Artificial Neural Networks and Deep Learning?—Examples of Practical Applications in Medicine,” 2023. doi: 10.3390/diagnostics13152582.
- [62] S. J. Maceachern and N. D. Forkert, “Machine learning for precision medicine,” 2021. doi: 10.1139/gen-2020-0131.
- [63] D. Mendhe, A. Dogra, P. S. Nair, S. Punitha, K. S. Preetha, and S. B. G. T. Babu, “AI-Enabled Data-Driven Approaches for Personalized Medicine and Healthcare Analytics,” in *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, 2024, pp. 1–5. doi: 10.1109/ICONSTEM60960.2024.10568722.
- [64] H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and U. R. Acharya, “Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022),” 2022. doi: 10.1016/j.cmpb.2022.107161.
- [65] A. S. Albahri *et al.*, “A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion,” *Information Fusion*, vol. 96, 2023, doi: 10.1016/j.inffus.2023.03.008.
- [66] J. K. Kim and S. Kang, “Neural Network-Based Coronary Heart Disease Risk Prediction Using Feature Correlation Analysis,” *J Healthc Eng*, vol. 2017, 2017, doi: 10.1155/2017/2780501.

- [67] I. Or, I. Omri, R. R., S. M., G. H., and S. M., “A Machine Learning Approach to Classify Exercise Limitation Severity Using Cardiopulmonary Exercise Testing - Development and Validation,” *Med Res Arch*, vol. 11, no. 11, 2023.
- [68] N. Cauwenberghs *et al.*, “Integrative Interpretation of Cardiopulmonary Exercise Tests for Cardiovascular Outcome Prediction: A Machine Learning Approach,” *Diagnostics*, vol. 13, no. 12, 2023, doi: 10.3390/diagnostics13122051.
- [69] G. Varisco *et al.*, “Central apnea detection in premature infants using machine learning,” *Comput Methods Programs Biomed*, vol. 226, 2022, doi: 10.1016/j.cmpb.2022.107155.
- [70] M. A. Raja, R. Loughran, and F. M. Caffery, “A review of applications of artificial intelligence in cardiorespiratory rehabilitation,” 2023. doi: 10.1016/j.imu.2023.101327.
- [71] M. Rosoł, M. Petelczyc, J. S. Gąsior, and M. Młyńczak, “Prediction of peak oxygen consumption using cardiorespiratory parameters from warmup and submaximal stage of treadmill cardiopulmonary exercise test,” *PLoS One*, vol. 19, no. 1, p. e0291706, Jan. 2024, doi: 10.1371/journal.pone.0291706.
- [72] M. Rosoł, M. Młyńczak, and G. Cybulski, “Granger causality test with nonlinear neural-network-based methods: Python package and simulation study.,” *Comput Methods Programs Biomed*, vol. 216, 2022, doi: 10.1016/j.cmpb.2022.106669.
- [73] M. Rosoł, J. S. Gąsior, I. Walecka, B. Werner, G. Cybulski, and M. Młyńczak, “Causality in cardiorespiratory signals in pediatric cardiac patients,” in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022, pp. 355–358. doi: 10.1109/EMBC48229.2022.9871750.
- [74] M. Rosoł, J. S. Gąsior, K. Korzeniewski, J. Łaba, R. Makuch, and M. Młyńczak, “Prediction of the duration of maximal exercise test in professional adolescent football players based on the cardiorespiratory signals – a pilot study,” in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024, pp. 1–4. doi: 10.1109/EMBC53108.2024.10782255.
- [75] M. Rosoł *et al.*, “Machine Learning Classification of Pediatric Health Status Based on Cardiorespiratory Signals with Causal and Information Domain Features Applied—An Exploratory Study,” *J Clin Med*, vol. 13, no. 23, 2024, doi: 10.3390/jcm13237353.



- [76] D. Mongin, J. García-Romero, and J. R. Alvero-Cruz, "Treadmill Maximal Exercise Tests from the Exercise Physiology and Human Performance Lab of the University of Malaga (version 1.0.1) PhysioNet," <https://physionet.org/content/treadmill-exercise-cardioresp/1.0.1/>.
- [77] D. Mongin, C. Chabert, D. S. Courvoisier, J. García-Romero, and J. R. Alvero-Cruz, "Heart rate recovery to assess fitness: comparison of different calculation methods in a large cross-sectional study," *Research in Sports Medicine*, vol. 31, no. 2, 2023, doi: 10.1080/15438627.2021.1954513.
- [78] R. A. Robergs, D. Dwyer, and T. Astorino, "Recommendations for improved data processing from expired gas analysis indirect calorimetry," *Sports Medicine*, vol. 40, no. 2, 2010, doi: 10.2165/11319670-000000000-00000.
- [79] C. A. Selland, J. Kelly, K. Gums, J. R. Meendering, and M. Vukovich, "A Generalized Equation for Prediction of VO<sub>2</sub>peak from a Step Test," *Int J Sports Med*, vol. 42, no. 9, 2021, doi: 10.1055/a-1310-3900.
- [80] A. Szijarto *et al.*, "Deep-learning based prediction of peak oxygen uptake in athletes using 2D echocardiographic videos," *Eur Heart J Cardiovasc Imaging*, vol. 24, no. Supplement\_1, p. jead119.244, Jun. 2023, doi: 10.1093/ehjci/jead119.244.
- [81] R. E. Klaren, G. P. Horn, B. Fernhall, and R. W. Motl, "Accuracy of the VO<sub>2</sub>peak prediction equation in firefighters," *Journal of Occupational Medicine and Toxicology*, vol. 9, no. 1, 2014, doi: 10.1186/1745-6673-9-17.
- [82] S. A. BILLINGER, E. VAN SWEARINGEN, M. MCCLAIN, A. A. LENTZ, and M. B. GOOD, "Recumbent Stepper Submaximal Exercise Test to Predict Peak Oxygen Uptake," *Med Sci Sports Exerc*, vol. 44, no. 8, 2012, [Online]. Available: [https://journals.lww.com/acsm-msse/Fulltext/2012/08000/Recumbent\\_Stepper\\_Submaximal\\_Exercise\\_Test\\_to.17.aspx](https://journals.lww.com/acsm-msse/Fulltext/2012/08000/Recumbent_Stepper_Submaximal_Exercise_Test_to.17.aspx)
- [83] A. A. Herda, A. A. Lentz, A. E. Mattlage, J. F. Sisante, and S. A. Billinger, "Cross-validation of the recumbent stepper submaximal exercise test to predict peak oxygen uptake in older adults," *Phys Ther*, vol. 94, no. 5, 2014, doi: 10.2522/ptj.20130307.
- [84] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *EMNLP 2014 - 2014 Conference on Empirical*

- Methods in Natural Language Processing, Proceedings of the Conference*, 2014. doi: 10.3115/v1/d14-1179.
- [85] A. Wismüller, A. M. Dsouza, M. A. Vosoughi, and A. Abidin, “Large-scale nonlinear Granger causality for inferring directed dependence from short multivariate time-series data,” *Sci Rep*, 2021, doi: 10.1038/s41598-021-87316-6.
  - [86] M. Mlynczak, “Temporal orders and causal vector for physiological data analysis,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2020. doi: 10.1109/EMBC44109.2020.9176842.
  - [87] A. Plaza-Florido *et al.*, “The role of heart rate on the associations between body composition and heart rate variability in children with overweight/Obesity: The active brains project,” *Front Physiol*, vol. 10, no. JUL, 2019, doi: 10.3389/fphys.2019.00895.
  - [88] A. Molfino, A. Fiorentini, L. Tubani, M. Martuscelli, F. R. Fanelli, and A. Laviano, “Body mass index is related to autonomic nervous system activity as measured by heart rate variability,” *Eur J Clin Nutr*, vol. 63, no. 10, 2009, doi: 10.1038/ejcn.2009.35.
  - [89] A. Porta *et al.*, “Cardiovascular control and time domain granger causality: Insights from selective autonomic blockade,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1997, 2013, doi: 10.1098/rsta.2012.0161.
  - [90] M. Młyńczak and H. Krysztofiak, “Cardiorespiratory temporal causal links and the differences by sport or lack thereof,” *Front Physiol*, 2019, doi: 10.3389/fphys.2019.00045.
  - [91] U. Freyschuss and A. Melcher, “Sinus arrhythmia in man: Influence of tidal volume and oesophageal pressure,” *Scand J Clin Lab Invest*, vol. 35, no. 6, 1975, doi: 10.1080/00365517509095772.
  - [92] Maciej Rosoł, “Repozytorium utworzone na potrzeby EMBC2024.”
  - [93] X. Chen *et al.*, “Prediction of 6-minute walk performance in patients with peripheral artery disease,” *J Vasc Surg*, vol. 66, no. 4, 2017, doi: 10.1016/j.jvs.2017.03.438.
  - [94] S. L. Sutterfield *et al.*, “Prediction of Planetary Mission Task Performance for Long-Duration Spaceflight,” *Med Sci Sports Exerc*, vol. 51, no. 8, 2019, doi: 10.1249/MSS.0000000000001980.

- [95] P. Leo, J. Spragg, J. Wakefield, and J. Swart, "Predictors of cycling performance success: Traditional approaches and a novel method to assess performance capacity in U23 road cyclists," *J Sci Med Sport*, vol. 26, no. 1, 2023, doi: 10.1016/j.jsams.2022.11.005.
- [96] A. K. A. McKay *et al.*, "Defining Training and Performance Caliber: A Participant Classification Framework," *Int J Sports Physiol Perform*, vol. 17, no. 2, 2022, doi: 10.1123/ijsp.2021-0451.
- [97] C. G. S. Araújo and J. Scharhag, "Athlete: A working definition for medical and health sciences research," 2016. doi: 10.1111/sms.12632.
- [98] R. M. de Abreu *et al.*, "Cardiorespiratory coupling is associated with exercise capacity in athletes: A cross-sectional study," *Respir Physiol Neurobiol*, vol. 320, p. 104198, Feb. 2024, doi: 10.1016/J.RESP.2023.104198.
- [99] R. M. de Abreu *et al.*, "Cardiorespiratory coupling strength in athletes and non-athletes," *Respir Physiol Neurobiol*, vol. 305, 2022, doi: 10.1016/j.resp.2022.103943.
- [100] C. D. Da Silva *et al.*, "Cardiorespiratory coupling as an early marker of cardiac autonomic dysfunction in type 2 diabetes mellitus patients," *Respir Physiol Neurobiol*, vol. 311, 2023, doi: 10.1016/j.resp.2023.104042.
- [101] S. Hietakoste *et al.*, "Acute cardiorespiratory coupling impairment in worsening sleep apnea-related intermittent hypoxemia," *IEEE Trans Biomed Eng*, 2023, doi: 10.1109/TBME.2023.3300079.
- [102] H. Yoon *et al.*, "Sleep-Dependent Directional Coupling of Cardiorespiratory System in Patients With Obstructive Sleep Apnea," *IEEE Trans Biomed Eng*, vol. 65, no. 12, 2018, doi: 10.1109/TBME.2018.2819719.
- [103] J. Cui *et al.*, "A wearable system for cardiopulmonary assessment and personalized respiratory training," *Future Generation Computer Systems*, vol. 112, 2020, doi: 10.1016/j.future.2020.06.039.

## 10. Opinia Komisji Bioetycznej



### Komisja Bioetyczna przy Warszawskim Uniwersytecie Medycznym

Tel.: 022/ 57 - 20 -303  
Fax: 022/ 57 - 20 -165

ul. Żwirki i Wigury nr 61  
02-091 Warszawa

e-mail: komisja.bioetyczna@wum.edu.pl  
www.komisja-bioetyczna.wum.edu.pl

KB/ 40 /2021

Komisja Bioetyczna przy Warszawskim Uniwersytecie Medycznym  
w dniu 14 czerwca 2021 r. po zapoznaniu się z wnioskiem:

Prof. dr hab.n.med. Bożena Werner,  
Klinika Kardiologii Wieku Dziecięcego  
i Pediatrii Ogólnej  
ul. Żwirki i Wigury 63A, 02-091 Warszawa

dotyczącym: wyrażenia opinii w sprawie badania pt.: „Ocena zgodności i przydatności wyników analizy zmienności zatokowego rytmu serca oraz częstotliwości i objętości oddechowej uzyskanych za pomocą urządzenia Pneumonitor w grupie pacjentów pediatrycznych kardiologicznych”

- Badanie może być prowadzone wyłącznie w okresie obowiązywania polisy ubezpieczeniowej.

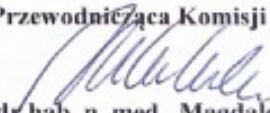
#### wyraża następującą opinię

- stwierdza, że jest ono dopuszczalne i zgodne z zasadami naukowo-etycznymi\*.
- ~~- stwierdza, że jest ono niedopuszczalne i niezgodne z zasadami naukowo-etycznymi.\*~~

#### Uwagi Komisji – *verte*

Komisja działa na podstawie art.29 ustawy z dnia 5.12.1996r. o zawodzie lekarza /Dz.U.nr 28/97 poz.152 wraz z późn.zm./, zarządzenia MZiOS z dn.11.05.1999r. w sprawie szczegółowych zasad powoływania i finansowania oraz trybu działania komisji bioetycznych /Dz.U.nr 47 poz.480/, Ustawy prawo farmaceutyczne z dnia 6 września 2001r. (Dz.U.Nr 126, poz. 1381 z późn. zm.) oraz Zarządzenie nr 56/2007 z dnia 15 października 2007r. w sprawie działania Komisji Bioetycznej przy Warszawskim Uniwersytecie Medycznym /Regulamin Komisji Bioetycznej przy Warszawskim Uniwersytecie Medycznym/.  
Komisja działa zgodnie z zasadami GCP .

Przewodnicząca Komisji Bioetycznej

  
Prof. dr hab. n. med. Magdalena Kuźma-Kozakiewicz

\*niepotrzebne skreślić

## 11. Oświadczenia współautorów publikacji

### Oświadczenie współautorów publikacji

Niniejszym oświadczam, że w publikacji:

M. Rosoł, M. Petelczyc, J. S. Gąsior, and M. Młyńczak, "Prediction of peak oxygen consumption using cardiorespiratory parameters from warmup and submaximal stage of treadmill cardiopulmonary exercise test," PLoS One, vol. 19, no. 1, p. e0291706, Jan. 2024, doi: 10.1371/journal.pone.0291706.

wkład merytoryczny autorów wyglądał następująco:

Konceptualizacja: M.R. i M.P.

Metodologia: M.R., M.P., J.S.G. i M.M.

Gromadzenie danych: M.R.

Obróbka danych: M.R. i M.P.

Oprogramowanie: M.R.

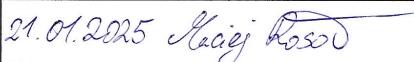
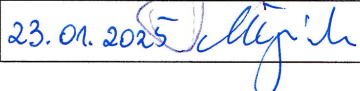
Formalna analiza: M.R.

Manuskrypt — przygotowanie pierwszego draftu: M.R.

Manuskrypt — recenzja i edycja: M.R., M.P., J.S.G. i M.M.

Wizualizacje: M.R.

Procentowy wkład autorski Macieja Rosoła wynosi 75%.

| Imię i nazwisko współautora | Data i podpis współautora                                                                       |
|-----------------------------|-------------------------------------------------------------------------------------------------|
| Maciej Rosoł                | 21.01.2025  |
| Monika Petelczyc            | 23.01.2025  |
| Jakub S. Gąsior             | 22.01.2025  |
| Marcel Młyńczak             | 23.01.2025  |

## Oświadczenie współautorów publikacji

Niniejszym oświadczam, że w publikacji:

M. Rosoł, M. Młyńczak, and G. Cybulski, "Granger causality test with nonlinear neural-network-based methods: Python package and simulation study.," Comput Methods Programs Biomed, vol. 216, 2022, doi: 10.1016/j.cmpb.2022.106669.

wkład merytoryczny autorów wyglądał następująco:

Konceptualizacja: M.R. i M.M.

Metodologia: M.R.

Gromadzenie danych: M.R.

Obróbka danych: M.R.

Oprogramowanie: M.R.

Formalna analiza: M.R.

Manuskrypt — przygotowanie pierwszego draftu: M.R.

Manuskrypt — recenzja i edycja: M.R., M.M. i G.C.

Wizualizacje: M.R.

Procentowy wkład autorski Macieja Rosoła wynosi 80%.

| Imię i nazwisko współautora | Data i podpis współautora |
|-----------------------------|---------------------------|
| Maciej Rosoł                | 21.01.2025 Maciej Rosoł   |
| Marcel Młyńczak             | 23.01.2025 mlynczak       |
| Gerard Cybulski             | 22.01.2025 G. Cybulski    |

## Oświadczenie współautorów publikacji

Niniejszym oświadczam, że w publikacji:

M. Rosoł, J. S. Gąsior, I. Walecka, B. Werner, G. Cybulski, and M. Młyńczak, "Causality in cardiorespiratory signals in pediatric cardiac patients," Annu Int Conf IEEE Eng Med Biol Soc, vol. 2022, 2022, doi: 10.1109/EMBC48229.2022.9871750

wkład merytoryczny autorów wyglądał następująco:

Konceptualizacja: M.R. i M.M.

Metodologia: M.R.

Gromadzenie danych: M.R., J.S.G., I.W. i B.W.

Obróbka danych: M.R.

Oprogramowanie: M.R.

Formalna analiza: M.R.

Manuskrypt — przygotowanie pierwszego draftu: M.R.

Manuskrypt — recenzja i edycja: M.R., J.S.G., I.W., B.W., G.C. i M.M.

Wizualizacje: M.R.

Procentowy wkład autorski Macieja Rosoła wynosi 75%.

| Imię i nazwisko współautora | Data i podpis współautora  |
|-----------------------------|----------------------------|
| Maciej Rosoł                | 21.01.2025 Maciej Rosoł    |
| Jakub S. Gąsior             | 22.01.2025 J. Gąsior       |
| Iwona Walecka               | 22.01.2025 Iwona Walecka   |
| Bożena Werner               | 22.01.2025 B. Werner       |
| Gerard Cybulski             | 22.01.2025 Gerard Cybulski |
| Marcel Młyńczak             | 23.01.2025 Marcel Młyńczak |

## Oświadczenie współautorów publikacji

Niniejszym oświadczam, że w publikacji:

M. Rosoł, J. S. Gąsior, K. Korzeniewski, J. Łaba, R. Makuch, and M. Młyńczak, "Prediction of the duration of maximal exercise test in professional adolescent football players based on the cardiorespiratory signals – a pilot study," in 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2024, pp. 1–4. doi: 10.1109/EMBC53108.2024.10782255.

wkład merytoryczny autorów wyglądał następująco:

Konceptualizacja: M.R. i M.M.

Metodologia: M.R.

Gromadzenie danych: M.R., J.S.G., K.K., J.L. i R.M.

Obróbka danych: M.R. i J.S.G.

Oprogramowanie: M.R.

Formalna analiza: M.R.

Manuskrypt — przygotowanie pierwszego draftu: M.R.

Manuskrypt — recenzja i edycja: M.R., J.S.G., K.K., J.L., R.M. i M.M.

Wizualizacje: M.R.

Procentowy wkład autorski Macieja Rosoła wynosi 75%.

| Imię i nazwisko współautora | Data i podpis współautora             |
|-----------------------------|---------------------------------------|
| Maciej Rosoł                | 21.01.2025 <i>Maciej Rosoł</i>        |
| Jakub S. Gąsior             | 22.01.2025 <i>J. Gąsior</i>           |
| Kacper Korzeniewski         | 23.01.2025 <i>Kacper Korzeniewski</i> |
| Jonasz Łaba                 | 22.01.2025 <i>Jonasz Łaba</i>         |
| Robert Makuch               | 24.01.2025 <i>R. Makuch</i>           |
| Marcel Młyńczak             | 23.01.2025 <i>M. Młyńczak</i>         |



## Oświadczenie współautorów publikacji

Niniejszym oświadczam, że w publikacji:

M. Rosoł, J. S. Gąsior, K. Korzeniewski, J. Łaba, R. Makuch, B. Werner and M. Młyńczak, "Machine Learning Classification of Pediatric Health Status Based on Cardiorespiratory Signals with Causal and Information Domain Features Applied—An Exploratory Study," J Clin Med, vol. 13, no. 23, 2024, doi: 10.3390/jcm13237353.

wkład merytoryczny autorów wyglądał następująco:

Konceptualizacja: M.R. i M.M.

Metodologia: M.R.

Gromadzenie danych: M.R., J.S.G., K.K., J.L., R.M. i B.W.

Obróbka danych: M.R.

Oprogramowanie: M.R.

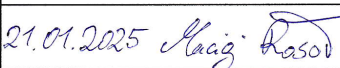
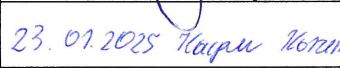
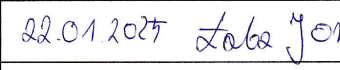
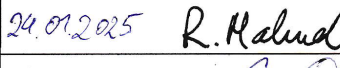
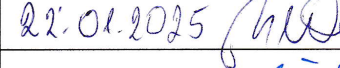
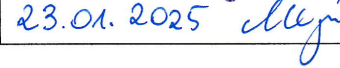
Formalna analiza: M.R.

Manuskrypt — przygotowanie pierwszego draftu: M.R.

Manuskrypt — recenzja i edycja: M.R., J.S.G., K.K., J.L., R.M., B.W. i M.M.

Wizualizacje: M.R.

Procentowy wkład autorski Macieja Rosoła wynosi 75%.

| Imię i nazwisko współautora | Data i podpis współautora                                                                       |
|-----------------------------|-------------------------------------------------------------------------------------------------|
| Maciej Rosoł                | 21.01.2025  |
| Jakub S. Gąsior             | 22.01.2025  |
| Kacper Korzeniewski         | 23.01.2025  |
| Jonasz Łaba                 | 22.01.2025  |
| Robert Makuch               | 24.01.2025  |
| Bożena Werner               | 22.01.2025  |
| Marcel Młyńczak             | 23.01.2025  |

## 12. Pozostałe publikacje

1. **M. Rosol**, J. S. Gąsior, J. Łaba, K. Korzeniewski, and M. Młyńczak, "Evaluation of the performance of GPT-3.5 and GPT-4 on the Polish Medical Final Examination," *Sci Rep*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-46995-z. (140 MNiSW, 3,8 IF)
2. M. Młyńczak, **M. Rosol**, A. Spinelli, et al., "Obstetric Anal Sphincter Injury Detection Using Impedance Spectroscopy with the ONIRY Probe," *Applied Sciences*, vol. 11, no. 2, 2021, doi: 10.3390/app11020637. (100 MNiSW, 2,5 IF)
3. J. S. Gąsior, **M. Rosol**, M. Młyńczak, et al., "Reliability of Symbolic Analysis of Heart Rate Variability and Its Changes During Sympathetic Stimulation in Elite Modern Pentathlon Athletes: A Pilot Study," *Front Physiol*, vol. 13, 2022, doi: 10.3389/fphys.2022.829887. (100 MNiSW, 3,2 IF)
4. M. Młyńczak, **M. Rosol**, K. Korzeniewski, et al., "Determination of anal sphincter injury location using impedance spectroscopy in obstetric patients," in 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2024, pp. 1–4. doi: 10.1109/EMBC53108.2024.10782955. (20 MNiSW)
5. J. S. Gąsior, M. Młyńczak, **M. Rosol**, et al., "Validity of the Pneumonitor for RR intervals acquisition for short-term heart rate variability analysis extended with respiratory data in pediatric cardiac patients," *Kardiol Pol*, vol. 81, no. 5, 2023, doi: 10.33963/KP.a2023.0070. (100 MNiSW, 3,7 IF)
6. J. S. Gąsior, M. Młyńczak, **M. Rosol**, P. Wieniawski, R. Pietrzak, and B. Werner, "Validity of the Pneumonitor for Analysis of Short-Term Heart Rate Asymmetry Extended with Respiratory Data in Pediatric Cardiac Patients," *J Clin Med*, vol. 13, no. 16, 2024, doi: 10.3390/jcm13164654. (140 MNiSW, 3,0 IF)
7. J. S. Gąsior, M. Gąsienica-Józkowy, M. Młyńczak, **M. Rosol**, et al., "Heart rate dynamics and asymmetry during sympathetic activity stimulation and post-stimulation recovery in ski mountaineers—a pilot exploratory study," *Front Sports Act Living*, vol. 6, 2024, doi: 10.3389/fspor.2024.1336034. (20 MNiSW, 2,3 IF)
8. B. J. Bednarski, Ł. E. Lepak, J. J. Łyskawa, P. Pieńczuk, **M. Rosol**, and R. S. Romaniuk, "Influence of IQT on research in ICT," *International Journal of Electronics and Telecommunications*, vol. 68, no. 2, 2022, doi: 10.24425-ijet.2022.139876/963. (70 MNiSW, 0,5 IF)
9. K. Korzeniewski, J. S. Gąsior, M. Mikielwicz, **M. Rosol**, R. Makuch, M. Młyńczak, „Exploration of Heart Rate Variability for the Prediction of Performance in Youth Footballers,” *Computing in Cardiology*, 2024, vol. 51, doi: 10.22489/cinc.2024.136.

## RESEARCH ARTICLE

# Prediction of peak oxygen consumption using cardiorespiratory parameters from warmup and submaximal stage of treadmill cardiopulmonary exercise test

Maciej Rosol<sup>1\*</sup>, Monika Petelczyc<sup>2</sup>, Jakub S. Gąsior<sup>3</sup>, Marcel Młyńczak<sup>1</sup>

**1** Faculty of Mechatronics, Institute of Metrology and Biomedical Engineering, Warsaw University of Technology, Warsaw, Poland, **2** Faculty of Physics, Warsaw University of Technology, Warsaw, Poland, **3** Department of Pediatric Cardiology and General Pediatrics, Medical University of Warsaw, Warsaw, Poland

\* [maciej.rosol.dokt@pw.edu.pl](mailto:maciej.rosol.dokt@pw.edu.pl)



## OPEN ACCESS

**Citation:** Rosol M, Petelczyc M, Gąsior JS, Młyńczak M (2024) Prediction of peak oxygen consumption using cardiorespiratory parameters from warmup and submaximal stage of treadmill cardiopulmonary exercise test. PLoS ONE 19(1): e0291706. <https://doi.org/10.1371/journal.pone.0291706>

**Editor:** Mario André da Cunha Espada, Instituto Politécnico de Setúbal, PORTUGAL

**Received:** September 4, 2023

**Accepted:** December 19, 2023

**Published:** January 10, 2024

**Copyright:** © 2024 Rosol et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The Python code used for the analysis is available at <https://github.com/mrosol/VO2peak-prediction>. The data used in the study is available at <https://physionet.org/content/treadmill-exercise-cardioresp/1.0.1/> (DOI: [10.13026/7ezk-j442](https://doi.org/10.13026/7ezk-j442)).

**Funding:** Research was funded by POB Biotechnology and biomedical engineering of Warsaw University of Technology within the Excellence Initiative: Research University (IDUB)

## Abstract

This study investigates the quality of peak oxygen consumption ( $VO_{2peak}$ ) prediction based on cardiac and respiratory parameters calculated from warmup and submaximal stages of treadmill cardiopulmonary exercise test (CPET) using machine learning (ML) techniques and assesses the importance of respiratory parameters for the prediction outcome. The database consists of the following parameters: heart rate (HR), respiratory rate (RespRate), pulmonary ventilation (VE), oxygen consumption ( $VO_2$ ) and carbon dioxide production ( $VCO_2$ ) obtained from 369 treadmill CPETs. Combinations of features calculated based on the HR, VE and RespRate time-series from different stages of CPET were used to create 11 datasets for  $VO_{2peak}$  prediction. Thirteen ML algorithms were employed, and model performances were evaluated using cross-validation with mean absolute percentage error (MAPE),  $R^2$  score, mean absolute error (MAE), and root mean squared error (RMSE) calculated after each iteration of the validation. The results demonstrated that incorporating respiratory-based features improves the prediction of  $VO_{2peak}$ . The best results in terms of  $R^2$  score (0.47) and RMSE (5.78) were obtained for the dataset which included both cardiac- and respiratory-based features from CPET up to 85% of age-predicted  $HR_{max}$ , while the best results in terms of MAPE (10.5%) and MAE (4.63) were obtained for the dataset containing cardiorespiratory features from the last 30 seconds of warmup. The study showed the potential of using ML models based on cardiorespiratory features from submaximal tests for prediction of  $VO_{2peak}$  and highlights the importance of the monitoring of respiratory signals, enabling to include respiratory parameters into the analysis. Presented approach offers a feasible alternative to direct  $VO_{2peak}$  measurement, especially when specialized equipment is limited or unavailable.

programme. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## 1. Introduction

Peak oxygen consumption ( $VO_{2peak}$ ) obtained through cardiopulmonary exercise test (CPET) is the popular measure of cardiorespiratory fitness [1]. It is a reliable predictor of cardiac events [2, 3], as well as lung cancer [4] and liver transplantation survival [5] and risk of postoperative complications [6]. Moreover,  $VO_{2peak}$  is a predictor of sport performance [7–9] and physical task performance during spaceflight [10]. Although CPET is the most reliable form of test, it is costly, requires specialized personnel and advanced equipment [11].

While conducting CPET, heart rate (HR) data are usually obtained through electrocardiography (ECG), while respiratory rate (RespRate) and pulmonary ventilation (VE) are gathered using tight-fitting masks. Nevertheless, this data can be acquired with relative ease, using heart rate monitors or smartwatches in case of HR, and impedance pneumography (IP) in case of RespRate and VE [12, 13]. Moreover, CPET is physically demanding as assumes the participants' exhaustion and thus it is contraindicated for patients with acute myocardial infarction, unstable angina, uncontrolled arrhythmia causing symptoms or hemodynamic compromise, uncontrolled asthma, and other pathological conditions [11]. Maximal cardiopulmonary exercise test might also interfere with an athletes training program [14].

Actually, due to: a) the growing development of aforementioned measurement devices, b) availability of simply field-based tests such as incremental shuttle walk test [15, 16] and c) new statistical prediction models and equations, clinicians and/or researchers are able to estimate  $VO_{2peak}$  and/or  $VO_{2max}$  (it is however not the subject of this study), based on selected parameters without performing maximal CPET [17–22]. Unfortunately, estimated  $VO_{2peak}$  using, e.g., only 6-min Walk Test distance demonstrated poor agreement with measured  $VO_{2peak}$  from a CPET [23]. Addition of other data such as demographic, anthropometric, and functional characteristics improved the accuracy of  $VO_{2peak}$  estimate based on walking tests at least in elderly patients with stable coronary artery disease (model with all variables explained 73% of  $VO_{2peak}$  variance) [24]. Thus, estimation of peak oxygen consumption based on combination of demographic factors and cardiac parameters obtained during submaximal (or even not) physical effort is possible, however, may be biased.

Reliable and accurate estimation of  $VO_{2peak}$  without performing maximal CPET may require more input physiological data to perform more sophisticated analyses. Thus, the development of new prediction models or equations, which will be able to accurately estimate  $VO_{2peak}$  and/or  $VO_{2max}$ , and will not relies on performing maximal CPET, is still ongoing [18, 25]. In recent years with the growth of the popularity of ML tools incorporated during the data analysis phase, those techniques were also utilized for the prediction of  $VO_2$  kinetics and  $VO_{2max}$  [26, 27]. ML models were also used by Szijarto *et al.* for prediction of  $VO_{2peak}$  based on the anthropometric data and 2D echocardiography (2DE) [28]. This approach was more accurate than a model based on anthropometric factors, however, it required performing a 2DE examination with sophisticated equipment and a trained physician. Importantly, not only the model or prediction algorithm might be important in terms of the prediction accuracy, but also the features used for the training. There are existing studies utilizing respiratory rate and ML for prediction of oxygen uptake dynamics during CPET [29–31]. However, to the best of our knowledge, there have been no previous studies utilizing features from cardiorespiratory time-series obtained from submaximal CPET, for the prediction of  $VO_{2peak}$  using ML models.

The aim of this paper was hence to investigate the quality of  $VO_{2peak}$  prediction by models based on cardiac and respiratory features obtained from different stages of CPET. Additionally, we assessed the importance of respiratory-based features included in the models for  $VO_{2peak}$  prediction.

## 2. Materials and methods

### 2.1. Data and study population

The study was performed on the publicly available database of cardiorespiratory time-series acquired during treadmill maximal cardiopulmonary exercise tests presented by *Mongin et al.* [32, 33]. The database comprises 992 recordings from experiments undertaken among amateur and professional athletes in the Exercise Physiology and Human Performance Lab of the University of Málaga between 2008 and 2018 with two types of protocols: a continuous increase of treadmill speed and a graded approach. In the database, one may find two forms of protocols on the treadmill continuous (ramping) and step-by-step incremental effort [32, 33]. Discussing the consistency of the analysis, we decided to limit our study only to experiments with continuous increase of the speed. In general, the protocol has: warmup lasting 8–10 minutes at 5 km/h (recording covers only about two last minutes), incremental effort with a 1 km/h/min speed increase and recovery. In the latter phase, the treadmill speed was set back to the initial 5 km/h speed [33]. The length of recordings from warmup differs between cases but was satisfactory for our modeling purposes. During recovery, subjects were asked to walk. Participants were instructed to go beyond exhaustion and the test was considered maximal if the increase of  $\text{VO}_2$  was less than 2.1 mL/kg/min between successive stages. Then the effort test was stopped “to avoid the vasovagal syncope” [32, 33] and the recovery started. The study was conducted according to the principles of the Declaration of Helsinki, the study protocol was approved by the Research Ethics Committee of the University of Málaga, written informed consent was obtained from the participants and all the data were analyzed anonymously.

During each test, the following cardiorespiratory time-series were acquired: heart rate (HR), respiratory rate (RespRate), pulmonary ventilation (VE), oxygen uptake ( $\text{VO}_2$ ) and carbon dioxide production ( $\text{VCO}_2$ ). Data were acquired on a breath-to-breath basis. HR was monitored via a 12-lead ECG (Mortara Instrument, Inc., USA), while respiratory signals were obtained using the CPX MedGraphics gas analyzer system (Medical Graphics Corporation, USA) [32].

Participants between 18 and 40 years old were chosen for the analysis reducing the sample size to 692. Tests only with continuous increasing speed were selected in order to obtain more consistent conditions along the study population. In result, 485 recordings have left. Next, subjects who were determined as outliers based on the 1.5 interquartile range method in terms of weight, height, and  $\text{VO}_{2\text{peak}}$ , with respect to the given sex, were excluded from the study, limiting to 462 recordings. Furthermore, the obtained data was visually evaluated in order to discard measurements during which there were visible artefacts in HR acquisition (e.g., sudden drop of over 30 bpm or lack of continuity of HR time-series during CPET probably due to electrode detachment); ultimately 369 recordings became background for the analysis. The final recordings belong to 327 unique subjects (42 subjects had more than one test) including 275 men and 52 women. The demographic summary of the final group is presented in Table 1.

Table 1. Descriptive statistics of the study population.

|       | Age [years]            | Height [cm]               | Weight [kg]            | BMI                    | $\text{VO}_{2\text{peak}}$ [ml/min/kg] |
|-------|------------------------|---------------------------|------------------------|------------------------|----------------------------------------|
| Men   | 27.3 ± 5.8 (18.0–39.8) | 177.4 ± 6.3 (160.5–193.0) | 76.6 ± 8.3 (55.3–97.0) | 24.3 ± 2.2 (17.9–31.4) | 47.7 ± 7.5 (28.9–67.3)                 |
| Woman | 26.9 ± 6.3 (18.0–40.0) | 165.2 ± 6.1 (154.0–178.0) | 62.2 ± 8.2 (46.0–83.0) | 22.8 ± 2.3 (18.0–29.6) | 38.1 ± 6.3 (24.8–53.8)                 |
| All   | 27.3 ± 5.9 (18–40)     | 175.5 ± 7.6 (154.0–193.0) | 74.5 ± 9.7 (46.0–97.0) | 24.1 ± 2.3 (17.9–31.3) | 46.3 ± 8.1 (24.8–67.3)                 |

<https://doi.org/10.1371/journal.pone.0291706.t001>

## 2.2. Modeling

Based on the aforementioned dataset, we decided to investigate the quality of  $VO_{2peak}$  prediction from different stages of CPET based on cardiac and respiratory parameters, and to assess the importance of respiratory-based features included in the modeling of  $VO_{2peak}$ . For this purpose, we utilized recorded time-series of HR, RespRate, and VE.  $VO_{2peak}$  was determined as the maximal value of the signal obtained after a 15-breath  $VO_2$  moving average window according to the recommendation presented by *Roberts et al.* [34].

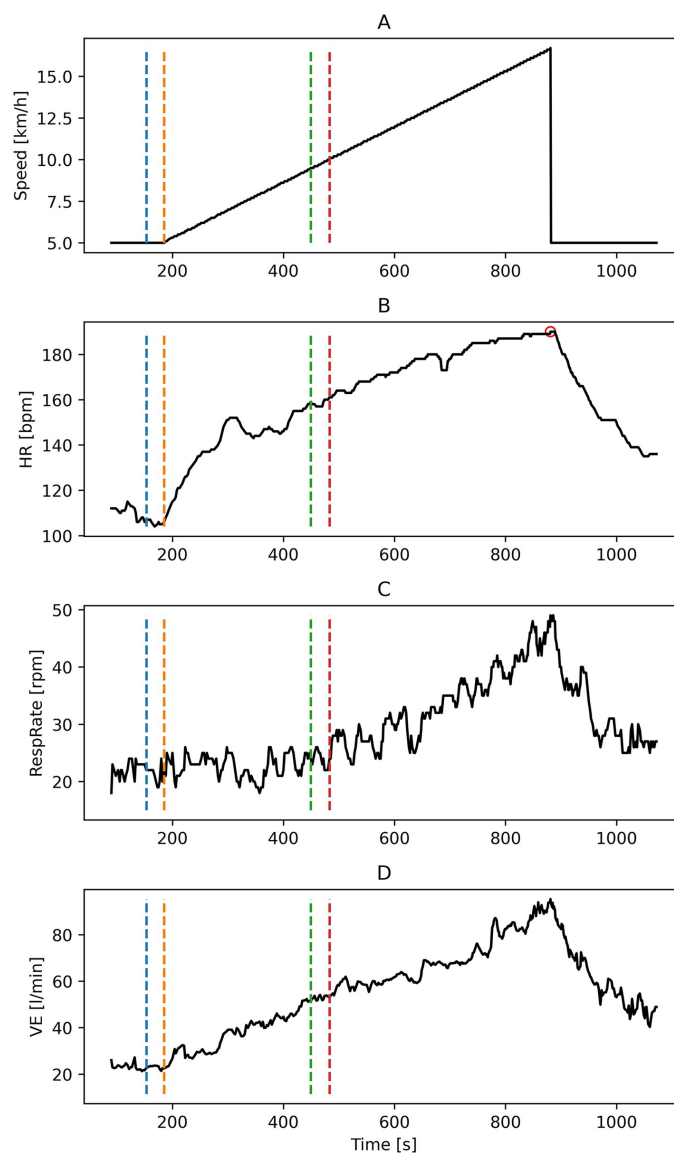
As features for ML models, basic statistics such as mean, standard deviation, maximal and minimal value, median, 25<sup>th</sup> and 75<sup>th</sup> quantile, skewness, kurtosis, coefficient from linear regression, impulse and shape factors were calculated for HR, RespRate, and VE, for a given stage of the maximal CPET. On this basis, 11 datasets were created based on different combinations of parameters and CPET stages, as presented in Table 2. Our research is focused on the submaximal stage from the cardiopulmonary exercise test, which equals 85% of the maximal measured and age-predicted  $HR_{max}$  as a threshold. Studied value of HR termination is commonly used in submaximal testing [35–37]. We also used both actual  $HR_{max}$  obtained during the treadmill cardiopulmonary exercise test, and age-predicted  $HR_{max}$  (220-age) in order to provide insights about the utility of the prediction of  $VO_{2peak}$  in submaximal tests without prior knowledge about the value of  $HR_{max}$  for a given subject. The example plot of the signals, alongside the threshold for all the stages of the CPET for which the features were calculated, is presented in Fig 1.

The 10-fold cross-validation (CV) was used to assess the accuracy of the prediction. In each iteration, standardization of the non-categorical features based on the mean and standard deviation from the training dataset was performed. The only feature that was not standardized was participants' sex: -1 was assigned to male, and 1 to female subjects. Different ML algorithms, commonly used for regression problems, were utilized: Linear, Lasso and Ridge Regression, Random Forest, XGBoost, Multilayer perceptron, Epsilon-Support Vector Regression, Bayesian Ridge Regression, Bayesian Automatic Relevance Determination (ARD) Regression, Gaussian Process Regression, Gradient Boosting for Regression, Huber Regression and Theil-Sen Estimator [38–40]. The hyperparameter tuning was performed for each algorithm using the grid-search technique. In each iteration of the validation, metrics like mean absolute percentage error (MAPE),  $R^2$  score, mean absolute error (MAE), root mean squared error (RMSE) and Cohen's  $f^2$  for effect size were calculated. The best model for each dataset was determined based on the lowest MAPE

Table 2. Characteristics of all datasets with an indication of features belonging to individual datasets.

| Dataset | Subjects' demography (age, weight, height, sex) | HR features from the last 30 s of warmup | RespRate and VE features from the last 30 s of warmup | HR features from CPET up to 85% of $HR_{max}$ | RespRate and VE features from CPET up to 85% of $HR_{max}$ | HR features from CPET up to 85% of age-predicted $HR_{max}$ | RespRate and VE features from CPET up to 85% of age-predicted $HR_{max}$ |
|---------|-------------------------------------------------|------------------------------------------|-------------------------------------------------------|-----------------------------------------------|------------------------------------------------------------|-------------------------------------------------------------|--------------------------------------------------------------------------|
| D1      | +                                               |                                          |                                                       |                                               |                                                            |                                                             |                                                                          |
| D2      | +                                               | +                                        |                                                       |                                               |                                                            |                                                             |                                                                          |
| D3      | +                                               | +                                        | +                                                     |                                               |                                                            |                                                             |                                                                          |
| D4      | +                                               |                                          |                                                       | +                                             |                                                            |                                                             |                                                                          |
| D5      | +                                               |                                          |                                                       | +                                             | +                                                          |                                                             |                                                                          |
| D6      | +                                               | +                                        |                                                       | +                                             |                                                            |                                                             |                                                                          |
| D7      | +                                               | +                                        | +                                                     | +                                             | +                                                          |                                                             |                                                                          |
| D8      | +                                               |                                          |                                                       |                                               |                                                            | +                                                           |                                                                          |
| D9      | +                                               |                                          |                                                       |                                               |                                                            | +                                                           | +                                                                        |
| D10     | +                                               | +                                        |                                                       |                                               |                                                            | +                                                           |                                                                          |
| D11     | +                                               | +                                        | +                                                     |                                               |                                                            | +                                                           | +                                                                        |

<https://doi.org/10.1371/journal.pone.0291706.t002>



**Fig 1. Typical representation of the time-series for participants with selected fragments used in the analysis.** Part A presents the linearly increasing treadmill speed, part B heart rate fluctuations, part C respiratory rate and part D pulmonary ventilation kinetics. The segment between the blue and orange dashed lines is the last 30 seconds of warmup. The segment between the orange and green lines corresponds to the section of CPET up to 85% of the age-predicted  $\text{HR}_{\text{max}}$ . Finally, the segment between the orange and red lines corresponds to the increasing workload in CPET up to 85% of the measured  $\text{HR}_{\text{max}}$ , which is marked with a red circle on the HR plot.

<https://doi.org/10.1371/journal.pone.0291706.g001>

score (which was chosen arbitrarily) obtained from the cross-validation. For the best model, Lin concordance correlation coefficient was calculated, and results were visualized as the dependency between predicted and actual values of  $\text{VO}_{2\text{peak}}$  and as Bland-Altman plot. Moreover, the difference in the values of metrics for male and female was tested.

Metrics obtained from all datasets were pairwise compared using the Wilcoxon signed-rank test. The significance level was set to 0.05. For the calculations, Python 3.9.13 was used. The whole modeling pipeline is presented in Fig 2.

### 2.3. Explainable AI

In order to investigate the importance of the individual features used for ML modeling, explainable artificial intelligence (XAI) tools were applied. For this purpose, the Dalex Python package was used [41]. During each iteration of the cross-validation, Shapley values and model-level variable importance based on drop-out loss values were calculated on the test set. After the whole cross-validation, all Shapley values for each sample and feature, as well as mean variable importance values were visualized. For the variable importance, *model\_parts* function of *dalex.Explainer* class was used. 30 permutation rounds were performed on each variable with MAE as a loss function and no data sampling (argument *N* was equal to *None*) due to the small number of samples.

## 3. Results

The metrics obtained for the best algorithm in terms of the lowest MAPE from the cross-validation for each dataset are presented in Table 3 alongside the model names. The violin-plots of the obtained metrics for each dataset were visualized in Fig 3. The p-values from the Wilcoxon signed-rank test from a pairwise comparison of the metrics are presented in Fig 4.

The lowest MAPE and MAE (10.51% and 4.63, respectively) were obtained for dataset D11 (demographic data along with cardiac and respiratory features from the last 30 seconds of warmup and CPET up to 85% of age-predicted  $\text{HR}_{\text{max}}$ ), while the lowest RMSE and highest  $R^2$  score (5.78 and 0.47, respectively) were obtained for D9 (demographic data along with cardiac and respiratory features from CPET up to 85% of age-predicted  $\text{HR}_{\text{max}}$ ). The worst prediction of  $\text{VO}_{2\text{peak}}$  in terms of all metrics was achieved by using the D1 (demographic data) dataset. Results obtained for D11 were statistically significantly better in terms of all metrics than results for all the rest of the datasets excluding D9 as presented in Fig 4. Regarding  $R^2$  score and RMSE metrics, datasets that included respiratory-based features from the part of CPET (irrespective of  $\text{HR}_{\text{max}}$  determination, whether measured or estimated) showed statistically significant superiority over datasets lacking features based on VE and respiratory rate during the corresponding period. Similarly, for MAPE and MAE, datasets containing respiratory-based features calculated up to 85% of age-predicted  $\text{HR}_{\text{max}}$  demonstrated significantly better metrics than datasets without such features. The effect size was large (Cohen's  $f^2 > 0.35$ ) [42] for all iterations of cross-validations in the case D4, D5, D7, D8, D9 and D11.

The measured values of  $\text{VO}_{2\text{peak}}$  and values predicted for the dataset that obtained the lowest MAPE score (D11) were visualized in Fig 5. The Lin concordance correlation coefficient between predicted and measured  $\text{VO}_{2\text{peak}}$  values was 0.66. The Bland-Altman plot for this dataset is presented in Fig 6. There was no statistically significant difference in case of metrics obtained for male and female subjects. The results of the comparison of metrics obtained for male and female are presented in Table 4.

As the smallest mean MAPE was obtained for D11, Shapley values and feature importance were visualized for this dataset in Figs 7 and 8, respectively. The discussion of the XAI results can be found in the next section.



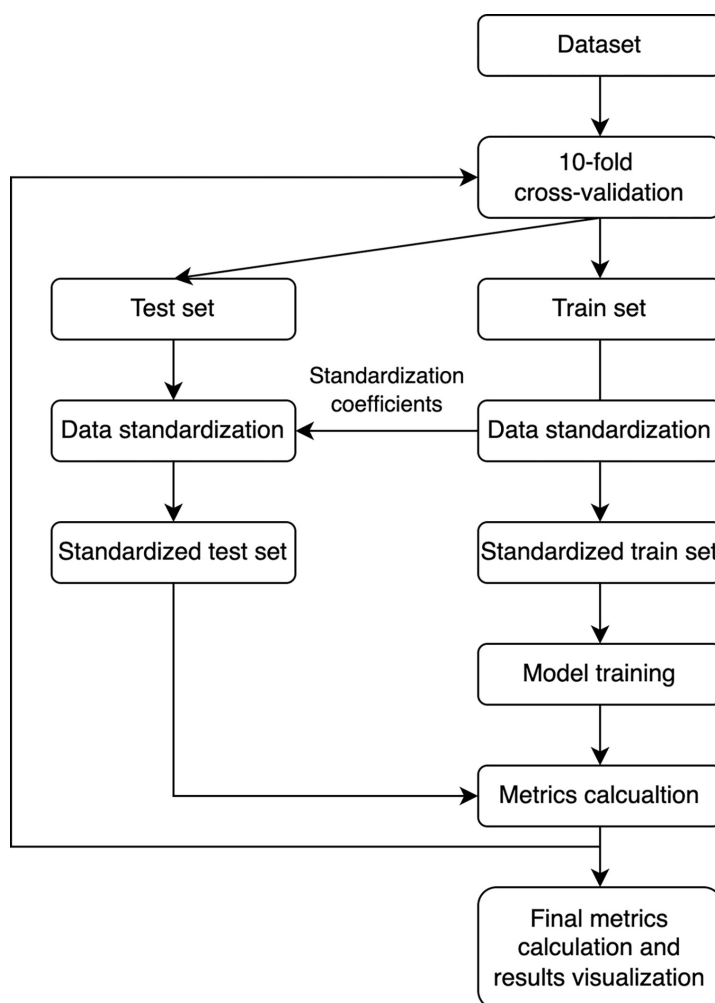


Fig 2. Modeling pipeline applied for each dataset and algorithm.

<https://doi.org/10.1371/journal.pone.0291706.g002>

#### 4. Discussion

Considering the features calculated from HR, VE, and RespRate time-series (attainable without the specialized equipment used in CPET), it is possible to predict  $VO_{2peak}$  from a submaximal test relying on age-predicted  $HR_{max}$ , achieving a mean absolute percentage error of 10.51% (for D11), using Bayesian ARD regression method. The addition of respiratory-based parameters resulted in an improvement of prediction compared to datasets based solely on the

**Table 3. Mean and standard deviation of metrics from cross-validation for each dataset for the model which resulted in the lowest MAPE for the given dataset. The highest metric values were highlighted.**

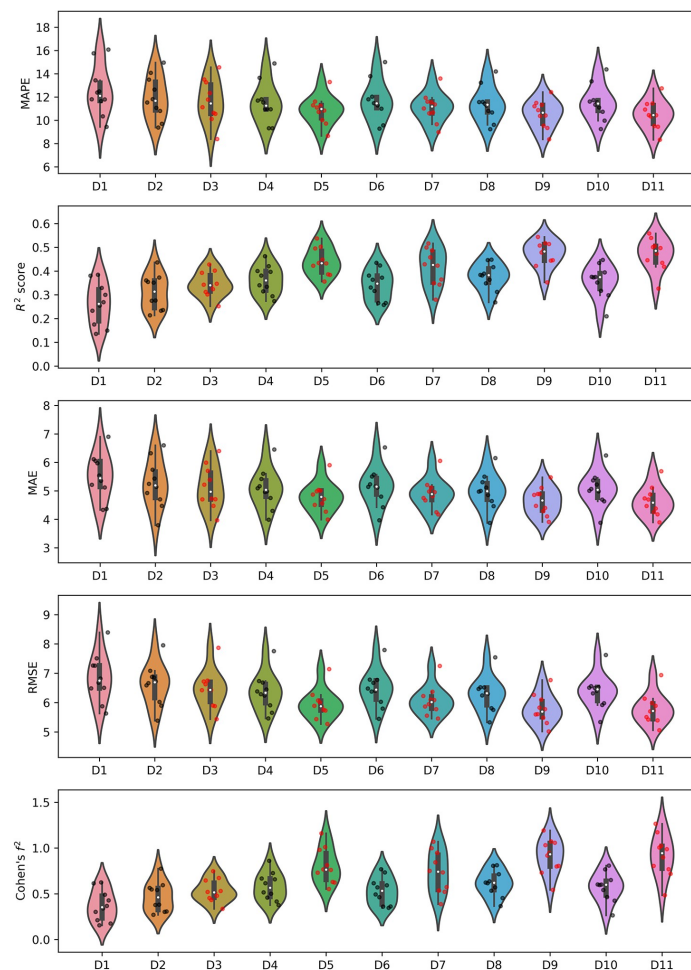
|     | MAPE [%]            | R <sup>2</sup>     | MAE [ml/min/kg]    | RMSE [ml/min/kg]   | Cohen's f <sup>2</sup> | Model                   |
|-----|---------------------|--------------------|--------------------|--------------------|------------------------|-------------------------|
| D1  | 12.52 ± 2.11        | 0.26 ± 0.09        | 5.50 ± 0.80        | 6.84 ± 0.81        | 0.37 ± 0.17            | Ridge regression        |
| D2  | 11.95 ± 1.84        | 0.31 ± 0.07        | 5.24 ± 0.84        | 6.61 ± 0.71        | 0.47 ± 0.16            | Huber regression        |
| D3  | 11.63 ± 1.84        | 0.34 ± 0.05        | 5.13 ± 0.75        | 6.45 ± 0.65        | 0.53 ± 0.12            | Bayesian ARD regression |
| D4  | 11.51 ± 1.72        | 0.36 ± 0.06        | 5.07 ± 0.68        | 6.36 ± 0.64        | 0.58 ± 0.15            | Bayesian ARD regression |
| D5  | 10.86 ± 1.23        | 0.44 ± 0.06        | 4.78 ± 0.52        | 5.95 ± 0.51        | 0.80 ± 0.19            | Bayesian ARD regression |
| D6  | 11.67 ± 1.72        | 0.34 ± 0.07        | 5.15 ± 0.69        | 6.46 ± 0.64        | 0.53 ± 0.15            | Lasso regression        |
| D7  | 11.10 ± 1.26        | 0.42 ± 0.08        | 4.90 ± 0.53        | 6.07 ± 0.50        | 0.74 ± 0.23            | Bayesian ARD regression |
| D8  | 11.36 ± 1.49        | 0.38 ± 0.06        | 4.99 ± 0.61        | 6.29 ± 0.59        | 0.61 ± 0.14            | Bayesian ARD regression |
| D9  | 10.54 ± 1.20        | <b>0.47 ± 0.06</b> | 4.64 ± 0.49        | <b>5.78 ± 0.50</b> | 0.91 ± 0.19            | Bayesian ARD regression |
| D10 | 11.50 ± 1.49        | 0.36 ± 0.07        | 5.06 ± 0.62        | 6.37 ± 0.59        | 0.57 ± 0.16            | Bayesian ARD regression |
| D11 | <b>10.51 ± 1.24</b> | 0.47 ± 0.07        | <b>4.63 ± 0.52</b> | 5.78 ± 0.52        | <b>0.91 ± 0.23</b>     | Bayesian ARD regression |

<https://doi.org/10.1371/journal.pone.0291706.t003>

corresponding stage of the treadmill cardiopulmonary exercise test in 4 out of 5 cases in terms of R<sup>2</sup> score and RMSE, and 2 out of 5 cases in terms of MAPE and MAE. When limiting treadmill cardiopulmonary exercise test to 85% of age-based HR<sub>max</sub>, the inclusion of features based on VE and RespRate improved the prediction in terms of all the specified metrics. The fact that the best results were achieved for the dataset considering 85% aged-based HR<sub>max</sub> and parameters obtained from easily accessible time-series indicates the possibility of using the presented method in clinical practice to determine  $VO_{2peak}$  without the prior knowledge of the actual HR<sub>max</sub> value and the necessity to perform a maximal treadmill cardiopulmonary exercise test.

Obtaining  $VO_{2peak}$  from maximal CPET might be costly, time-consuming and in some cases impossible or contraindicated to carried out due to observed cardiac or pulmonary dysfunction, musculoskeletal diseases, or strict training programs. Therefore, there is a growing interest in the prediction of  $VO_{2peak}$  and/or  $VO_{2max}$  from submaximal tests [14, 43–49]. Our study focused on investigating ML algorithms to predict  $VO_{2peak}$  with the set of features, which could be obtained using simpler techniques than commonly used spirometry, and the significance of incorporating respiration into the prediction process. The presented results are similar or superior compared to some other presented  $VO_{2peak}$  prediction methods like WFI  $VO_{2peak}$  prediction equation, deep-learning model based on 2DE, or regression models from PACER 20-m shuttle run [19, 28, 50–53]. However, in the existing literature, there are also techniques, which managed to obtain better performance like regression models based on submaximal exercise test protocol using a total body recumbent stepper [54–56]. Nonetheless, in those studies more heterogeneous groups of patients were present in terms of age or health status (patients after heart failure or individuals with low to moderate risk of cardiovascular diseases). Further improvement of the prediction of  $VO_{2peak}$  might be achieved by increasing sample size, and inclusion of other parameters based on the raw signals (especially ECG) like HRV and parameters from information and causal domain [57–60].

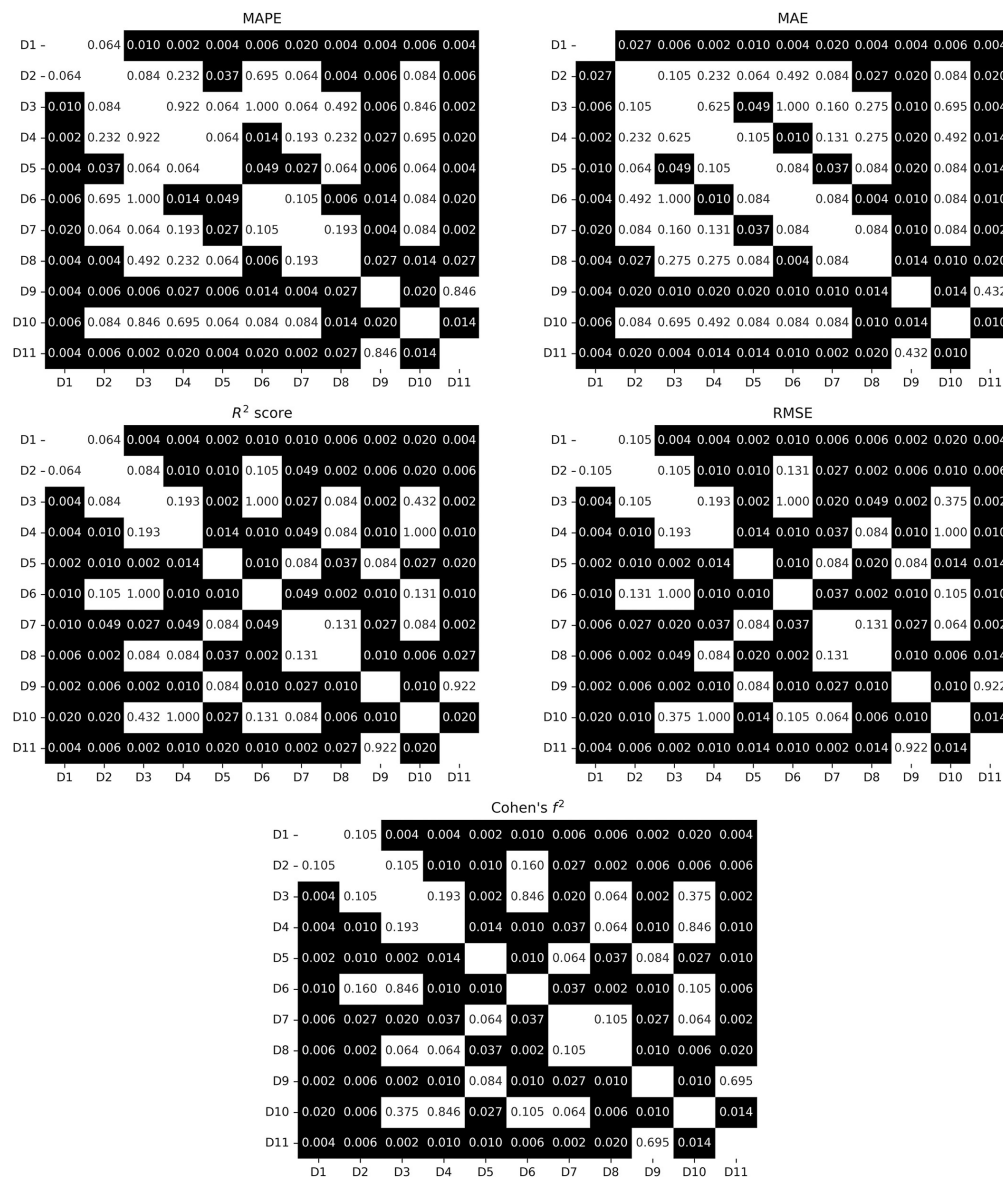
Another notable aspect of the study was the utilization of XAI tools, specifically Shapley values and model-level variable importance, to obtain insights into the feature importance for prediction. For most datasets (including D9 and D11, which produced the best results) Bayesian ARD Regression model was used, which has an ability to automatically determine the relevance of each feature, effectively pruning irrelevant or redundant information, while accentuating the impactful variables [61]. In our analysis, we found that the top five most influential features were consistent between Shapley values and variable importance. The most



**Fig 3.** Violin-plots of the calculated metrics for each dataset with the visualization of the metrics obtained in each iteration of the 10-fold cross-validation. Black dots represent metrics obtained from datasets without respiratory-based features, while red dots represent these that include such features.

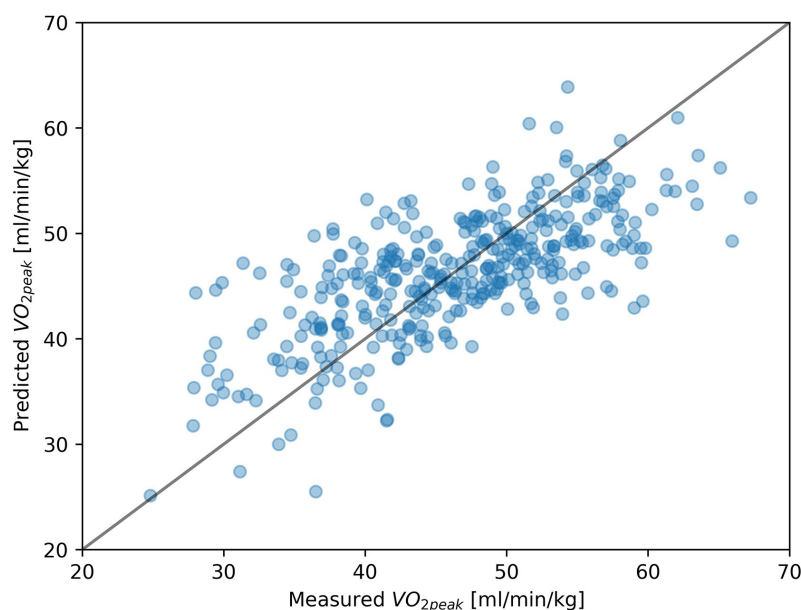
<https://doi.org/10.1371/journal.pone.0291706.g003>

impactful feature of the prediction was the maximal value of VE during the test, up to 85% of age-predicted  $HR_{max}$ . Additionally, subjects' weight and sex influenced the prediction results, with higher  $VO_{2peak}$  observed in lighter individuals and males compared to females. The importance of weight as a predictor for  $VO_{2peak}$  in the presented study was probably due to the utilization of peak oxygen consumption in relation to mass and expressed in ml/min/kg, which is in line with the results presented by Loftin et al. [62]. There are also multiple studies



**Fig 4.** The p-values from Wilcoxon signed-rank test from pairwise comparison of the metrics obtained from different datasets. P-values smaller than 0.05 are marked with a black background.

<https://doi.org/10.1371/journal.pone.0291706.g004>



**Fig 5. The plot of measured and predicted  $VO_{2peak}$  values for dataset D11.** The solid black line represents the function where predicted value is equal to the measured one.

<https://doi.org/10.1371/journal.pone.0291706.g005>

presenting the difference in  $VO_{2peak}$  between male and female [63, 64], thus the influence of the patients sex on the prediction seem to be natural. Notably, there was no significant difference between the metrics obtained for both sexes, indicating the robustness of the models in this regard. Patients age was not among the most influential features, however a higher patients age tend to result in the lower value of the predicted  $VO_{2peak}$ , which seem to be in line with the results of other studies [64, 65]. ML algorithms offer the advantage of processing and analyzing vast amounts of data at incredible speed, enabling them to identify complex patterns and relationships that may not be captured by humans. Thus, ML allows for the extraction of information on the influence of the almost unlimited number of features from big datasets on the  $VO_{2peak}$  values, which would be impossible just by human-based analysis. Notably, 13 out of the 20 features with the highest Shapley values and 10 out of the 15 features with the highest variable importance score were related to respiratory signals. Those findings seem to be in line with results presented in other studies, where the importance of respiratory signals in the context of oxygen consumption was presented [31, 66, 67]. The presented configuration offers the benefit of avoiding monitoring  $O_2$  consumption and  $CO_2$  production through laboratory device, instead allowing for the application of less sophisticated respiratory monitoring techniques, such as IP. Simultaneous acquisition of both ECG and IP can be performed using e.g., Pneumonitor device, which was recently developed and designed for research in the fields of physiology and sports medicine [12, 13, 68]. Thus, all the cardiorespiratory features under current study could be obtained using Pneumonitor without any additional equipment. However, it is important that the presented results are based on the CPET performed on a treadmill and

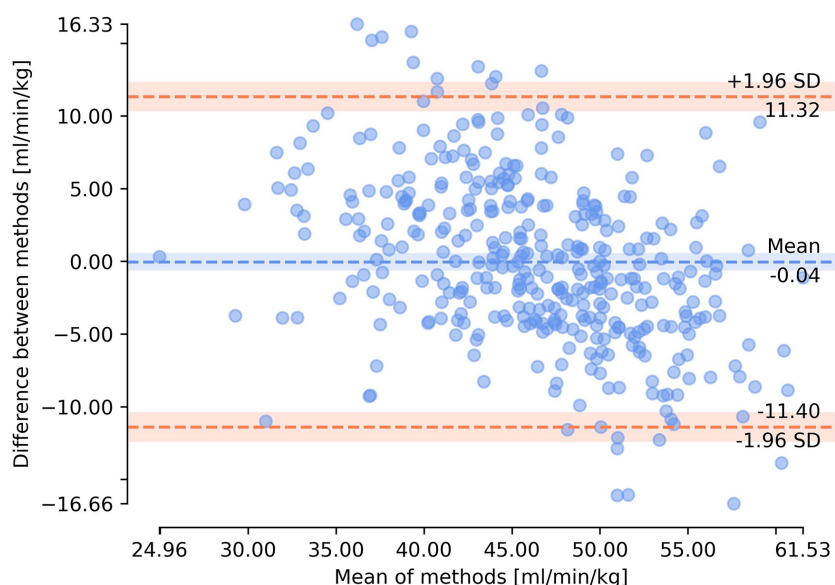


Fig 6. Bland-Altman plot of determined in CPET directly and predicted  $VO_{2peak}$  values based on the results for dataset D11.

<https://doi.org/10.1371/journal.pone.0291706.g006>

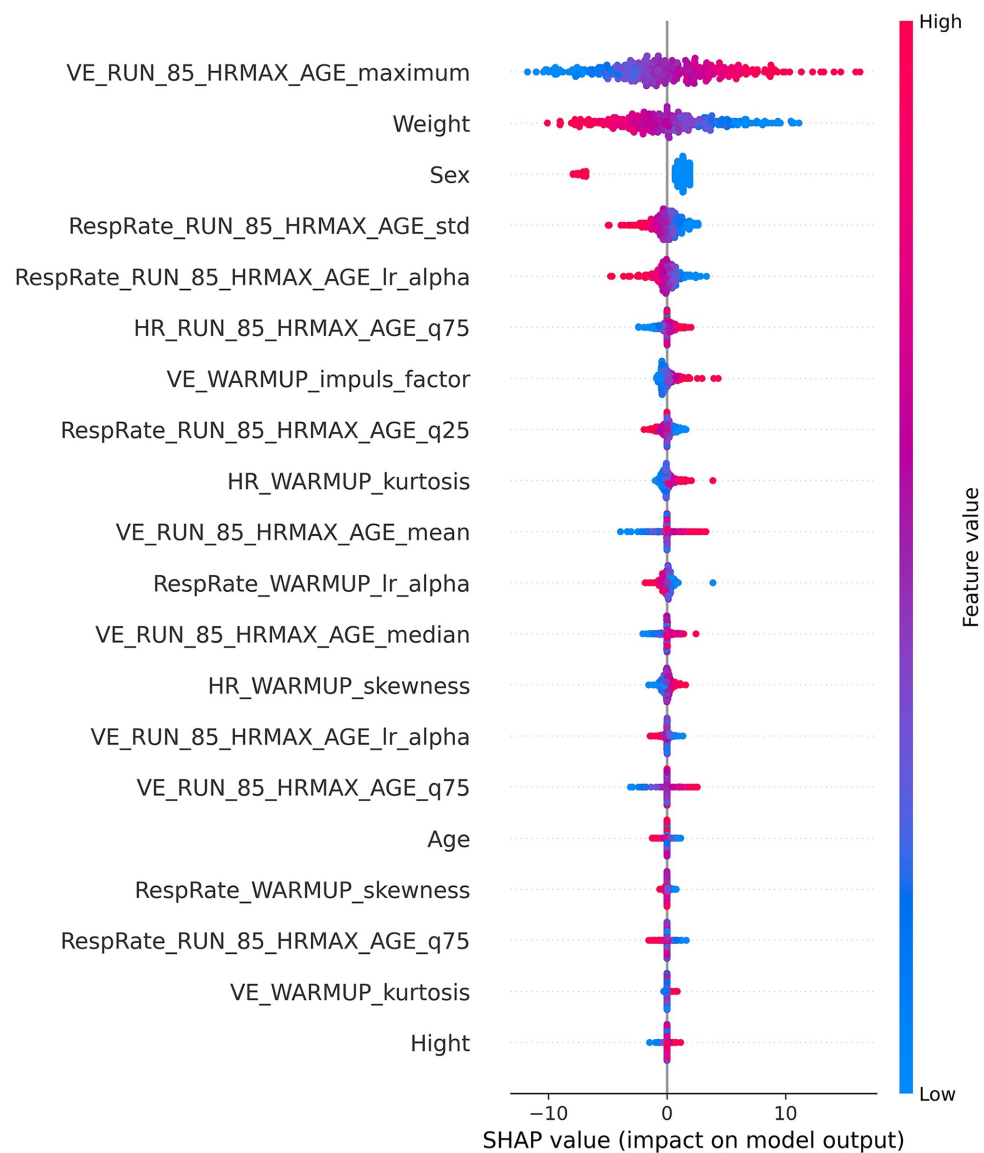
machine learning models training in this study should not be used on a data from tests performed using other exercise modalities, as it influences the cardiorespiratory parameters [69]. Moreover, the profile of the study population in terms of age and fitness level of subjects should be considered if applying the obtained models, as it has an influence on the  $VO_{2peak}$  values [70].

There are several limitations of the study. First of all, the raw ECG/RR-intervals signals and raw respiratory curves were unavailable, and thus more sophisticated parameters especially from information and causal domains, which could provide additional insights into the predictive models could have not been calculated. Moreover, the sample size in this study was limited, as only 369 recordings from the initial database of 992 CPET recordings were used for analysis after applying exclusion criteria based on outlier detection methods and visual inspection of the signals. Furthermore, the dataset was imbalanced in terms of patients' sex as there were 275 men and 52 women. A larger and more balanced dataset could prove beneficial for ML model training. There was also lack of information about the amount of sport activity undertaken by the participants, which might introduce inconsistency in the study population. Moreover, the equation used for determination of age-predicted  $HR_{max}$  might be also treated

Table 4. Mean and standard deviation of metrics from cross-validation for male and female for dataset D11 with p-value from Wilcoxon signed-rank test.

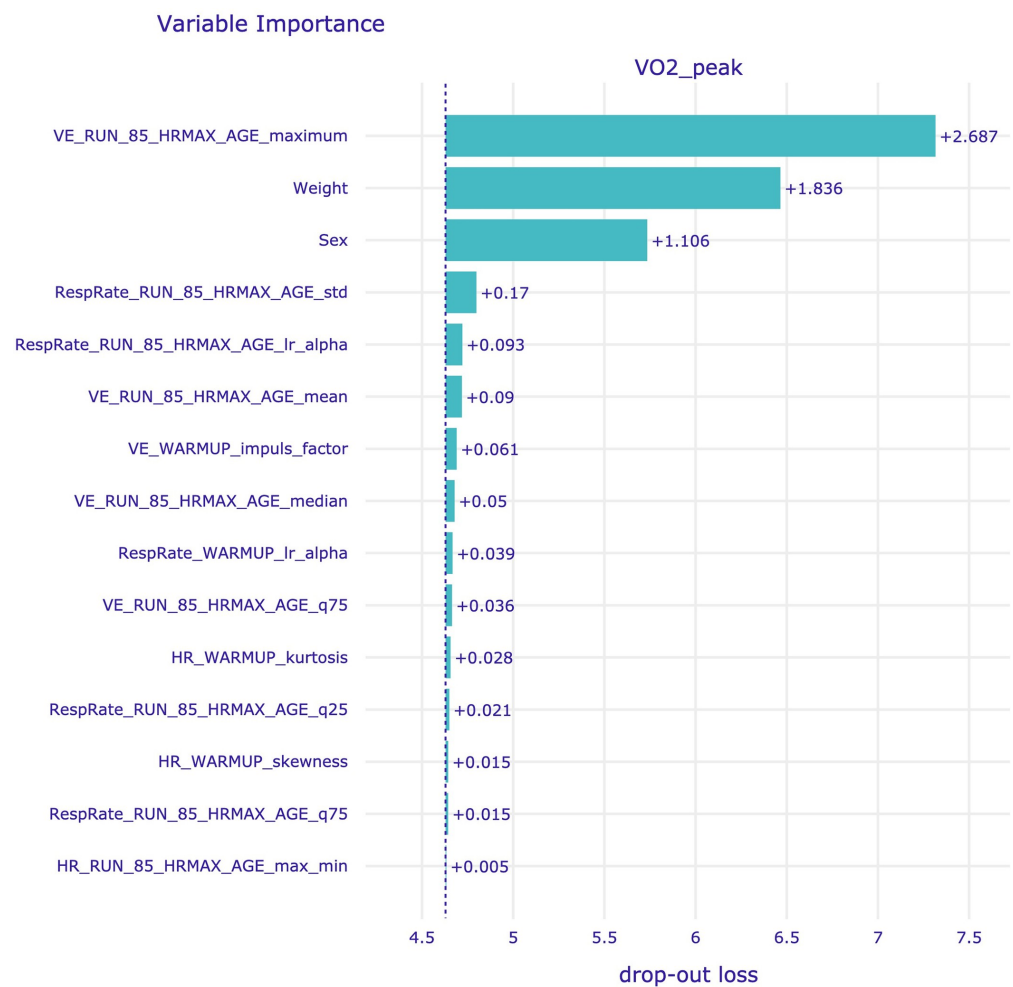
|         | MAPE         | R2           | MAE         | RMSE        | Cohens f2   |
|---------|--------------|--------------|-------------|-------------|-------------|
| Male    | 10.29 ± 1.30 | 0.35 ± 0.12  | 4.69 ± 0.58 | 5.86 ± 0.51 | 0.58 ± 0.29 |
| Female  | 11.77 ± 2.03 | -0.16 ± 1.24 | 4.25 ± 0.70 | 5.16 ± 1.13 | 0.43 ± 0.76 |
| P-value | 0.105        | 0.232        | 0.160       | 0.105       | 0.557       |

<https://doi.org/10.1371/journal.pone.0291706.t004>



**Fig 7. Shapley values obtained for dataset D11.** Feature names are explained in the [S1 Appendix](#).

<https://doi.org/10.1371/journal.pone.0291706.g007>



**Fig 8. Variable importance for dataset D11.** Feature names are explained in the [S1 Appendix](#).  
<https://doi.org/10.1371/journal.pone.0291706.g008>

as a limitation as there exist equations with smaller errors [71]. However, the equation used in this study is the most popularity and characterized by simplicity of application. Additionally, one approach of age-predicted  $HR_{max}$  calculation and one threshold of  $HR_{max}$  were introduced. Some of these limitations could be overcome by the usage of the Pneumonitor device, which allows for the simultaneous acquisition of raw ECG and IP signals [68]. Thus, the pulmonary activity (including RespRate and VE) can be monitored without the usage of sophisticated apparatus for gas analysis and tight-fitting masks may stress some groups of patients



(e.g., children). Future studies may explore the optimal percentage of  $HR_{max}$  and different approaches of estimation of age-predicted  $HR_{max}$ . It would be also valuable to study the influence of the subjects' physical activity level on the prediction accuracy as well as other than treadmill forms of cardiopulmonary exercise tests in order to determine the optimal settings for the prediction of  $VO_{2peak}$  for clinical practice. The utility of the described method may also depend on the reproducibility of the results, which need further testing. With high reproducibility, the method could be useful in clinical practice for e.g., tracking the changes of the CRF during training camps of athletes without performing full CPET.

This study expands the discussion on predicting cardiorespiratory fitness by highlighting the important role of submaximal testing and incorporating respiratory signals in the prediction process. The presented analysis indicates that the inclusion of respiratory parameters might improve the quality of the  $VO_{2peak}$  prediction in a group of athletic subjects aged between 18 and 40 years old when performing a submaximal test on a treadmill. The use of a submaximal test based on age-predicted  $HR_{max}$  and the utilization of cardiological and respiratory parameters that can be obtained without specialized CPET equipment is an advantage of the presented approach and facilitates its potential application in clinical practice.

## Supporting information

**S1 Appendix. List of feature names.**  
(DOCX)

## Author Contributions

**Conceptualization:** Maciej Rosoł, Monika Petelczyc.

**Data curation:** Maciej Rosoł, Monika Petelczyc.

**Formal analysis:** Maciej Rosoł.

**Methodology:** Maciej Rosoł, Monika Petelczyc, Jakub S. Gąsior, Marcel Młyńczak.

**Software:** Maciej Rosoł.

**Validation:** Maciej Rosoł.

**Visualization:** Maciej Rosoł.

**Writing – original draft:** Maciej Rosoł.

**Writing – review & editing:** Maciej Rosoł, Monika Petelczyc, Jakub S. Gąsior, Marcel Młyńczak.

## References

1. Lee J, Zhang XL. Physiological determinants of  $VO_{2max}$  and the methods to evaluate it: A critical review. Vol. 36, Science and Sports. 2021.
2. Laukkanen JA, Rauramaa R, Salonen JT, Kurl S. The predictive value of cardiorespiratory fitness combined with coronary risk evaluation and the risk of cardiovascular and all-cause death. J Intern Med. 2007; 262(2). <https://doi.org/10.1111/j.1365-2796.2007.01807.x> PMID: 17645594
3. Laukkanen JA, Kurl S, Salonen R, Rauramaa R, Salonen JT. The predictive value of cardiorespiratory fitness for cardiovascular events in men with various risk profiles: A prospective population-based cohort study. Eur Heart J. 2004; 25(16). <https://doi.org/10.1016/j.ehj.2004.06.013> PMID: 15321701
4. Jones LW, Watson D, Herndon JE, Eves ND, Hailthcock BE, Loewen G, et al. Peak oxygen consumption and long-term all-cause mortality in nonsmall cell lung cancer. Cancer. 2010; 116(20). <https://doi.org/10.1002/cncr.25396> PMID: 20597134

5. Bernal W, Martin-Mateos R, Lipcsey M, Tallis C, Woodsford K, McPhail MJ, et al. Aerobic capacity during cardiopulmonary exercise testing and survival with and without liver transplantation for patients with chronic liver disease. *Liver Transplantation*. 2014; 20(1). <https://doi.org/10.1002/lt.23766> PMID: 24136710
6. Snowden CP, Prentis JM, Anderson HL, Roberts DR, Randles D, Renton M, et al. Submaximal cardiopulmonary exercise testing predicts complications and hospital length of stay in patients undergoing major elective surgery. *Ann Surg*. 2010; 251(3). <https://doi.org/10.1097/SLA.0b013e3181cf811d> PMID: 20134313
7. Schabort EJ, Killian SC, St Clair Gibson A, Hawley JA, Noakes TD. Prediction of triathlon race time from laboratory testing in national triathletes. *Med Sci Sports Exerc*. 2000; 32(4). <https://doi.org/10.1097/00005768-200004000-00018> PMID: 10776905
8. Billat VL, Demarle A, Slawinski J, Paiva M, Koralsztajn JP. Physical and training characteristics of top-class marathon runners. *Med Sci Sports Exerc*. 2001; 33(12). <https://doi.org/10.1097/00005768-200112000-00018> PMID: 11740304
9. Staib JL, Im J, Caldwell Z, Rundell KW. Cross-Country Ski Racing Performance Predicted by Aerobic and Anaerobic Double Poling Power. *J Strength Cond Res*. 2000; 14(3).
10. Sutterfield SL, Alexander AM, Hammer SM, Dildier KD, Caldwell JT, Barstow TJ, et al. Prediction of Planetary Mission Task Performance for Long-Duration Spaceflight. *Med Sci Sports Exerc*. 2019; 51(8). <https://doi.org/10.1249/MSS.0000000000001980> PMID: 30882564
11. Levett DZH, Jack S, Swart M, Carlisle J, Wilson J, Snowden C, et al. Perioperative cardiopulmonary exercise testing (CPET): consensus clinical guidelines on indications, organization, conduct, and physiological interpretation. *Br J Anaesth*. 2018; 120(3). <https://doi.org/10.1016/j.bja.2017.10.020> PMID: 29452805
12. Mlyńczak M, Zyliński M, Niewiadomski W, Cybulski G. Ambulatory Devices Measuring Cardiorespiratory Activity with Motion. In: *BIODEVICES 2017 - 10th International Conference on Biomedical Electronics and Devices, Proceedings; Part of 10th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2017*. 2017.
13. Młyńczak MC, Niewiadomski W, Zyliński M, Cybulski GP. Ambulatory impedance pneumography device for quantitative monitoring of volumetric parameters in respiratory and cardiac applications. In: *Computing in Cardiology*. 2014.
14. Petelczyc M, Kotlewski M, Bruhn S, Weippert M. Maximal oxygen uptake prediction from submaximal bicycle ergometry using a differential model. *Sci Rep [Internet]*. 2023; 13(1):11289. Available from: <https://doi.org/10.1038/s41598-023-38089-7>
15. Marsico A, Corso SD, de Carvalho EF, Arakelian V, Phillips S, Stibulov R, et al. A more effective alternative to the 6-minute walk test for the assessment of functional capacity in patients with pulmonary hypertension. *Eur J Phys Rehabil Med*. 2021; 57(4).
16. Alves R, Lima MM, Fonseca C, Dos Reis R, Figueiredo PH, Costa H, et al. Peak oxygen uptake during the incremental shuttle walk test in a predominantly female population with Chagas heart disease. *Eur J Phys Rehabil Med*. 2016;52(1).
17. Buttar KK, Saboo N, Kacker S. A Review: Maximal Oxygen Uptake ( $VO_{2Max}$ ) and Its Estimation Methods. *International Journal of Physical Education, Sports and Health*. 2019; 6(6).
18. Lima LP, Leite HR, de Matos MA, Neves CDC, da Silva Lage VK, da Silva GP, et al. Cardiorespiratory fitness assessment and prediction of peak oxygen consumption by Incremental Shuttle Walking Test in healthy women. *PLoS One*. 2019; 14(2). <https://doi.org/10.1371/journal.pone.0211327> PMID: 30730949
19. Selland CA, Kelly J, Gums K, Meendering JR, Vukovich M. A Generalized Equation for Prediction of  $VO_{2peak}$  from a Step Test. *Int J Sports Med*. 2021; 42(9).
20. Cureton KJ, Sloniger MA, O'Bannon JP, Black DM, McCormack WP. A generalized equation for prediction of  $VO_{2peak}$  from 1-mile run/walk performance. *Med Sci Sports Exerc*. 1995; 27(3). PMID: 7752874
21. Cooper KD, Shafer AB. Validity and Reliability of the Polar A300's Fitness Test Feature to Predict  $VO_{2max}$ . *Int J Exerc Sci*. 2019; 12(4). PMID: 30899351
22. HERNANDEZ B, Roberts B, Kodidhi A, Roelle L, Miller N, Littell LM, et al. EVALUATING ACCURACY OF ESTIMATED  $VO_{2MAX}$  WITH WRIST WORN POLAR IGNITE COMPARED TO PEAK  $VO_{2}$  ON FORMAL CARDIOPULMONARY EXERCISE TESTING IN HEALTHY AND FONTAN PEDIATRIC PATIENTS. *J Am Coll Cardiol*. 2023; 81(8).
23. Chirico D, Davidson TW, Terada T, Scott K, Keast ML, Reid RD, et al. Using the 6-min walk test to monitor peak oxygen uptake response to cardiac rehabilitation in patients with heart failure. *J Cardiopulm Rehabil Prev*. 2020; 40(6). <https://doi.org/10.1097/HCR.0000000000000517> PMID: 33031135

24. Mandic S, Walker R, Stevens E, Nye ER, Body D, Barclay L, et al. Estimating exercise capacity from walking tests in elderly individuals with stable coronary artery disease. *Disabil Rehabil.* 2013; 35(22). <https://doi.org/10.3109/09638288.2012.759629> PMID: 23600709
25. Wiecha S, Kasiak PS, Szwed P, Kowalski T, Cieřliński I, Postula M, et al.  $VO_{2max}$  prediction based on submaximal cardiorespiratory relationships and body composition in male runners and cyclists: a population study. Löllgen H, Barton M, Löllgen H, Knechtel B, editors. *Elife* [Internet]. 2023; 12:e86291. Available from: <https://doi.org/10.7554/eLife.86291> PMID: 37162318
26. Ashfaq A, Cronin N, Müller P. Recent advances in machine learning for maximal oxygen uptake ( $VO_{2max}$ ) prediction: A review. Vol. 28, *Informatics in Medicine Unlocked*. 2022.
27. Hedge ET, Amelard R, Hughson RL. Prediction of oxygen uptake kinetics during heavy-intensity cycling exercise by machine learning analysis. *J Appl Physiol.* 2023; 134(6). <https://doi.org/10.1152/japplphysiol.00148.2023> PMID: 37199779
28. Szijarto A, Tokodi M, Fabian A, Lakatos BK, Shiida K, Tolvaj M, et al. Deep-learning based prediction of peak oxygen uptake in athletes using 2D echocardiographic videos. *Eur Heart J Cardiovasc Imaging* [Internet]. 2023 Jun 1; 24(Supplement\_1):jead119.244. Available from: <https://doi.org/10.1093/ehjci/jead119.244>
29. Zignoli A, Fornasiero A, Ragni M, Pellegrini B, Schena F, Biral F, et al. Estimating an individual's oxygen uptake during cycling exercise with a recurrent neural network trained from easy-to-obtain inputs: A pilot study. *PLoS One.* 2020; 15(3). <https://doi.org/10.1371/journal.pone.0229466> PMID: 32163443
30. Amelard R, Hedge ET, Hughson RL. Temporal convolutional networks predict dynamic oxygen uptake response from wearable sensors across exercise intensities. *NPJ Digit Med.* 2021; 4(1). <https://doi.org/10.1038/s41746-021-00531-3> PMID: 34764446
31. Wang Z, Zhang Q, Lan K, Yang Z, Gao X, Wu A, et al. Enhancing instantaneous oxygen uptake estimation by non-linear model using cardio-pulmonary physiological and motion signals. *Front Physiol.* 2022;13. <https://doi.org/10.3389/fphys.2022.897412> PMID: 36105296
32. Mongin D, Chabert C, Courvoisier DS, Garcia-Romero J, Alvero-Cruz JR. Heart rate recovery to assess fitness: comparison of different calculation methods in a large cross-sectional study. *Research in Sports Medicine.* 2023; 31(2).
33. Mongin D, Garcia-Romero J, Alvero-Cruz JR. <https://physionet.org/content/treadmill-exercise-cardioresp/1.0.1/>. 2021. Treadmill Maximal Exercise Tests from the Exercise Physiology and Human Performance Lab of the University of Malaga (version 1.0.1) PhysioNet.
34. Robergs RA, Dwyer D, Astorino T. Recommendations for improved data processing from expired gas analysis indirect calorimetry. *Sports Medicine.* 2010; 40(2). <https://doi.org/10.2165/11319670-000000000-00000> PMID: 20092364
35. Noonan V, Dean E. Submaximal exercise testing: Clinical application and interpretation. Vol. 80, *Physical Therapy.* 2000. PMID: 10911416
36. Shushan T, Lovell R, Buchheit M, Scott TJ, Barrett S, Norris D, et al. Submaximal Fitness Test in Team Sports: A Systematic Review and Meta-Analysis of Exercise Heart Rate Measurement Properties. Vol. 9, *Sports Medicine—Open.* 2023. <https://doi.org/10.1186/s40798-023-00564-w> PMID: 36964427
37. Leopold E, Tuller T, Scheinowitz M. A computational predictor of the anaerobic mechanical power outputs from a clinical exercise stress test. *PLoS One.* 2023; 18(5 MAY). <https://doi.org/10.1371/journal.pone.0283630> PMID: 37146031
38. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2016.
39. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research.* 2011;12.
40. Chollet F. Keras (2015). URL <http://keras.io>. 2017;
41. Baniecki H, Kretowicz W, Piatyszek P, Wisniewski J, Biecek P. dalex: Responsible machine learning with interactive explainability and fairness in python. *Journal of Machine Learning Research.* 2021; 22.
42. Lorah J. Effect size measures for multilevel models: definition, interpretation, and TIMSS example. *Large Scale Assess Educ.* 2018; 6(1).
43. Mart MF, Ely EW, Tolle JJ, Patel MB, Brummel NE. Physiologic responses to exercise in survivors of critical illness: an exploratory pilot study. *Intensive Care Medicine Experimental.* 2022; 10(1). <https://doi.org/10.1186/s40635-022-00461-8> PMID: 36008625
44. Baldasseroni S, Silverii MV, Pratesi A, Burgisser C, Orso F, Lucarelli G, et al. Cardiac Rehabilitation in Advanced aGE after PCI for acute coronary syndromes: predictors of exercise capacity improvement in the CR-AGE ACS study. *Aging Clin Exp Res.* 2022; <https://doi.org/10.1007/s40520-022-02130-y> PMID: 35451734

45. Izquierdo MC, Lopes S, Teixeira M, Polónia J, Alves AJ, Mesquita-Bastos J, et al. The Chester step test is a valid tool to assess cardiorespiratory fitness in adults with hypertension: reducing the gap between clinical practice and fitness assessments. Vol. 42, Hypertension Research. 2019. <https://doi.org/10.1038/s41440-019-0316-5> PMID: 31451720
46. Garcia-Tabar I, Iturricastillo A, Castellano J, Cadore EL, Izquierdo M, Setuain I. Predicting Cardiorespiratory Fitness in Female Soccer Players: The Basque Female Football Cohort Study. *Int J Sports Physiol Perform*. 2022;17(1).
47. Altmann S, Neumann R, Härtel S, Woll A, Buchheit M. Using submaximal exercise heart rate for monitoring cardiorespiratory fitness changes in professional soccer players: A replication study. *Int J Sports Physiol Perform*. 2021;16(8).
48. Sartor F, Vernillo G, De Morree HM, Bonomi AG, La Torre A, Kubis HP, et al. Estimation of maximal oxygen uptake via submaximal exercise testing in sports, clinical, and home settings. Vol. 43, Sports Medicine. 2013. <https://doi.org/10.1007/s40279-013-0068-3> PMID: 23821468
49. Albouaini K, Egred M, Alahmar A, Wright DJ. Cardiopulmonary exercise testing and its application. Vol. 83, Postgraduate Medical Journal. 2007.
50. Klaren RE, Horn GP, Fernhall B, Mott RW. Accuracy of the  $\text{VO}_{2\text{peak}}$  prediction equation in firefighters. *Journal of Occupational Medicine and Toxicology*. 2014; 9(1). <https://doi.org/10.1186/1745-6673-9-17> PMID: 24860611
51. Mahar MT, Welk GJ, Rowe DA, Crotts DJ, McIver KL. Development and Validation of a Regression Model to Estimate  $\text{VO}_{2\text{peak}}$  from PACER 20-m Shuttle Run Performance. *J Phys Act Health*. 2016; 3(s2).
52. Loe H, Nes BM, Wisløff U. Predicting  $\text{VO}_{2\text{peak}}$  from submaximal-and peak exercise models: The HUNT 3 fitness study, Norway. *PLoS One*. 2016; 11(1).
53. Harmsen WJ, Ribbers GM, Slaman J, Heijnenbroek-Kal MH, Khajeh L, van Kooten F, et al. The six-minute walk test predicts cardiorespiratory fitness in individuals with aneurysmal subarachnoid hemorrhage. *Top Stroke Rehabil*. 2017; 24(4). <https://doi.org/10.1080/10749357.2016.1260263> PMID: 27915583
54. BILLINGER SA VAN SWEARINGEN E, MCCLAIN M, LENTZ AA, GOOD MB. Recumbent Stepper Submaximal Exercise Test to Predict Peak Oxygen Uptake. *Med Sci Sports Exerc* [Internet]. 2012; 44(8). Available from: [https://journals.lww.com/acsm-msse/Fulltext/2012/08000/Recumbent\\_Stepper\\_Submaximal\\_Exercise\\_Test\\_to\\_17.aspx](https://journals.lww.com/acsm-msse/Fulltext/2012/08000/Recumbent_Stepper_Submaximal_Exercise_Test_to_17.aspx)
55. Herda AA, Lentz AA, Matlage AE, Sisante JF, Billinger SA. Cross-validation of the recumbent stepper submaximal exercise test to predict peak oxygen uptake in older adults. *Phys Ther*. 2014; 94(5). <https://doi.org/10.2522/ptj.20130307> PMID: 24435104
56. Deka P, Pozehl BJ, Pathak D, Williams M, Norman JF, Alonso WW, et al. Predicting maximal oxygen uptake from the 6 min walk test in patients with heart failure. *ESC Heart Fail*. 2021; 8(1).
57. Mlynczak M. Temporal orders and causal vector for physiological data analysis. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. 2020.
58. Rosol M, Mlynczak M, Cybulski G. Granger causality test with nonlinear neural-network-based methods: Python package and simulation study. *Comput Methods Programs Biomed*. 2022;216. <https://doi.org/10.1016/j.cmpb.2022.106669> PMID: 35151111
59. Schulz S, Adochiei FC, Edu IR, Schroeder R, Costin H, Bär KJ, et al. Cardiovascular and cardiorespiratory coupling analyses: A review. Vol. 371, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2013. <https://doi.org/10.1098/rsta.2012.0191> PMID: 23858490
60. Rosol M, Gasior JS, Walecka I, Werner B, Cybulski G, Mlynczak M. Causality in cardiorespiratory signals in pediatric cardiac patients. *Annu Int Conf IEEE Eng Med Biol Soc*. 2022;2022. <https://doi.org/10.1109/EMBC48229.2022.9871750> PMID: 36085711
61. Wipf D, Nagarajan S. A new view of automatic relevance determination. In: Advances in Neural Information Processing Systems 20—Proceedings of the 2007 Conference. 2008.
62. Loftin M, Sothorn M, Trosclair L, O'Hanlon A, Miller J, Udall J. Scaling  $\text{Vo}_2$  peak in obese and non-obese girls. *Obes Res*. 2001; 9(5). <https://doi.org/10.1038/oby.2001.36> PMID: 11346670
63. Welde B, Morseth B, Handegård BH, Lagestad P. Effect of Sex, Body Mass Index and Physical Activity Level on Peak Oxygen Uptake Among 14–19 Years Old Adolescents. *Front Sports Act Living*. 2020;2.
64. Byrne EA, Fleg JL, Vaitkevicius P V., Wright J, Porges SW. Role of aerobic capacity and body mass index in the age-associated decline in heart rate variability. *J Appl Physiol*. 1996;81(2).
65. Jackson AS, Beard EF, Wier LT, Ross RM, Stuteville JE, Blair SN. Changes in aerobic power of men, ages 25–70 yr. *Med Sci Sports Exerc*. 1995;27(1). PMID: 7898326

66. Gastinger S, Sorel A, Nicolas G, Gratas-Delamarche A, Prioux J. A comparison between ventilation and heart rate as indicator of oxygen uptake during different intensities of exercise. *J Sports Sci Med*. 2010;9(1). PMID: [24149394](#)
67. Beltrame T, Amelard R, Wong A, Hughson RL. Prediction of oxygen uptake dynamics by machine learning analysis of wearable sensors during activities of daily living. *Sci Rep*. 2017; 7.
68. Gašior JS, Młyńczak M, Rosol M, Wieniawski P, Walecka I, Cybulski G, et al. Validity of the Pneumonitor for RR intervals acquisition for short-term heart rate variability analysis extended with respiratory data in pediatric cardiac patients. *Kardiol Pol*. 2023; 81(5). <https://doi.org/10.33963/KP.a2023.0070> PMID: [36929303](#)
69. Cornelis N, Buys R. The effects of exercise modality on maximal and submaximal exercise parameters obtained by graded maximal exercise testing. *Int J Cardiol*. 2016;222. <https://doi.org/10.1016/j.ijcard.2016.07.190> PMID: [27513650](#)
70. Resaland GK, Andersen LB, Mamen A, Anderssen SA. Effects of a 2-year school-based daily physical activity intervention on cardiorespiratory fitness: The Sogndal school-intervention study. *Scand J Med Sci Sports*. 2011; 21(2). <https://doi.org/10.1111/j.1600-0838.2009.01028.x> PMID: [19895384](#)
71. Lach J, Wiecha S, Śliż D, Price S, Zaborski M, Cieśliński I, et al. HR Max Prediction Based on Age, Body Composition, Fitness Level, Testing Modality and Sex in Physically Active Population. *Front Physiol*. 2021; 12. <https://doi.org/10.3389/fphys.2021.695950> PMID: [34393819](#)



# Granger causality test with nonlinear neural-network-based methods: Python package and simulation study



Maciej Rosol<sup>\*</sup>, Marcel Młynczak, Gerard Cybulski

Faculty of Mechatronics, Institute of Metrology and Biomedical Engineering, Warsaw University of Technology, Warsaw, Poland

## ARTICLE INFO

**Article history:**  
Received 7 September 2021  
Revised 25 January 2022  
Accepted 26 January 2022

**Keywords:**  
Granger causality  
Time series  
Neural networks  
Python  
Prediction models

## ABSTRACT

**Background and objective:** Causality defined by Granger in 1969 is a widely used concept, particularly in neuroscience and economics. As there is an increasing interest in nonlinear causality research, a Python package with a neural-network-based causality analysis approach was created. It allows performing causality tests using neural networks based on Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), or Multilayer Perceptron (MLP). The aim of this paper is to present the nonlinear method for causality analysis and the created Python package.

**Methods:** The created functions with the autoregressive (AR) and Generalized Radial Basis Functions (GRBF) neural network models were tested on simulated signals in two cases: with nonlinear dependency and with absence of causality from Y to X signal. The train-test split (70/30) was used. Errors obtained on the test set were compared using the Wilcoxon signed-rank test to determine the presence of the causality. For the chosen model, the proposed method of study the change of causality over time was presented.

**Results:** In the case when X was a polynomial of Y, nonlinear methods were able to detect the causality, while the AR model did not manage to indicate it. The best results (in terms of the prediction accuracy) were obtained for the MLP for the lag of 150 (MSE equal to 0.011, compared to 0.041 and 0.036 for AR and GRBF, respectively). When there was no causality between the signals, none of the proposed and AR models did indicate false causality, while it was detected by GRBF models in one case. Only the proposed models gave the expected results in each of the tested scenarios.

**Conclusions:** The proposed method appeared to be superior to the compared methods. They were able to detect non-linear causality, make accurate forecasting and not indicate false causality. The created package enables easy usage of neural networks to study the causal relationship between signals. The neural-networks-based approach is a suitable method that allows the detection of a nonlinear causal relationship, which cannot be detected by the classical Granger method. Unlike other similar tools, the package allows for the study of changes in causality over time.

© 2022 The Authors. Published by Elsevier B.V.  
This is an open access article under the CC BY-NC-ND license  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## 1. Introduction

### 1.1. Granger causality

The causality concept, which is discussed in this paper was presented by Sir Clive Granger in 1969 [1]. Nowadays it is widely used in economics [2–5] and neuroscience [6–14]. Recently, there has also been a growing interest in Granger causality in the field of physiology, where it can be used for searching the cause of phenomena and even as a physiological marker [15–20]. In this ap-

proach, a causal relationship (where the time series Y is the cause of the time series X, denoted by  $Y \rightarrow X$ ) occurs if the variance of the prediction error of the Y based on the past of all available information (U) is statistically significantly smaller than the prediction error variance of the X based on the past of all available information except for the time series Y (Eq. (1)). In practice, some defined by researcher number of lag values of X and Y time series are treated as all available information, while past lags of X are treated as all available information except for Y.

$$\sigma^2(X|U) < \sigma^2(X|U - Y) \quad (1)$$

To test hypotheses about causality 2 linear autoregressive models (AR) and covariance stationary time series X and Y are assumed. The first model predicts the current value of time series

<sup>\*</sup> Corresponding author.  
E-mail address: [maciej.rosol.dokt@pw.edu.pl](mailto:maciej.rosol.dokt@pw.edu.pl) (M. Rosol).

X based only on  $p$  lagged values of X and Y (Eq. (2)), the second one predicts the current value of Y, based on the same values but with different coefficients (Eq. (3)). In the equations shown below A are the regression coefficients and E are the regression errors.

$$X(t) = \sum_{j=1}^p A_{11,j}X(t-j) + \sum_{j=1}^p A_{12,j}Y(t-j) + E_1(t) \quad (2)$$

$$Y(t) = \sum_{j=1}^p A_{21,j}X(t-j) + \sum_{j=1}^p A_{22,j}Y(t-j) + E_2(t) \quad (3)$$

If in the first equation the variance of error  $E_1$  is statistically significantly smaller for the model taking into account the variable Y (coefficients  $A_{12}$  different from zero) than the variance of the same model without taking into account the variable Y (coefficients  $A_{12}$  equal to zero), it means that the variable Y is the G-cause of the variable X [21]. Similarly, if the variance of the error  $E_2$  in the second equation is smaller to a statistically significant degree for the model including the variable X (coefficients  $A_{21}$  other than zero) from the variance of the prediction error for a model in which the variable X was not included (coefficients  $A_{21}$  equal to zero), it means that the variable X is the G-cause of the variable Y.

To test if variable Y is G-causing variable X, the F-test can be applied [22]. First, the residual sums of squares (RSS) are calculated for the model making prediction only on past values of X (Eq. (4)) and for the model that uses for prediction past values of X and Y time series (Eq. (5)) [23].

$$RSS_1 = \sum_{t=1}^T E_X(t)^2 \quad (4)$$

$$RSS_2 = \sum_{t=1}^T E_{X,Y}(t)^2 \quad (5)$$

Based on this, the test statistic for the F-test ( $S$ ) can be computed according to Eq. (6), where  $l$  is equal to the number of considered lags, and  $T$  is equal to the number of predicted values [24].

$$S_1 = \frac{(RSS_1 - RSS_2)/l}{RSS_2/(T - 2l - 1)} \sim F_{l,T-2l-1} \quad (6)$$

The value of the  $S_1$  test statistic is consistent with the Fisher distribution with

$l$  and  $(T-2l-1)$  degrees of freedom. To test whether Y causes X ( $Y \rightarrow X$ ), the F-test is performed under the null hypothesis that Y does not cause X and the alternative hypothesis that Y is causing X.

Also, the Chi-squared test may be applied with test statistics computed according to Eq. (7) [23].

$$S_1 = \frac{T(RSS_1 - RSS_2)}{RSS_2} \sim \chi^2(p) \quad (7)$$

Causality can not only be tested for occurrence but also quantified according to Eq. (8), where  $\sigma^2 E_X$  means variance of the error obtained from the model based only on the past values of X and  $\sigma^2 E_{X,Y}$  is the variance of the error obtained from the model based on the past of both signals [6,7,25].

$$F_{Y \rightarrow X} = \ln \frac{\sigma^2 E_X}{\sigma^2 E_{X,Y}} \quad (8)$$

If three time series (X-Z) are available, the mutual Granger causality analysis for two variables for each pair is not able to show some relationships between the data. For example, if Y is the signal causing Z, and Z is the signal causing X, but Y is not the signal causing X, then in the case of a two-variable analysis, this will be indistinguishable from the situation where Y is the signal causing both Z and X. This situation is presented in Fig. 1. With a

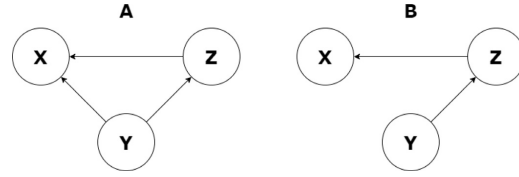


Fig. 1. Two cases of dependencies are indistinguishable for the two-way Granger causality analysis and in both cases, the result of the analysis would suggest the presence of causality presented in Diagram A.

causality analysis for two time series for both presented cases, the obtained result will suggest the presence of the relationship presented in part A of Fig. 1.

In order to distinguish such situations, a conditional Granger causality test can be used. It is able to indicate whether the past of the Y signal helps to reduce the variance of the prediction error of X predicted from the past of X and Z time series. In the case of conditional causality analysis, the linear autoregressive equations presented in Eq. (2) are extended by the sum of the products of the respective coefficients ( $A_{12}$ ) and the third variable Z, so that this variable is used for forecasting the present value of X, as presented in Eq. (9).

$$X(t) = \sum_{j=1}^p A_{11,j}X(t-j) + \sum_{j=1}^p A_{12,j}Y(t-j) + \sum_{j=1}^p A_{13,j}Z(t-j) + E_1(t) \quad (9)$$

Similar to the causality analysis for two time series, two models are created in the conditional causality analysis. The first one does not take into account the past of variable Y in the model (coefficients  $A_{12}$  equal to 0), the second one takes into account the past of all variables (coefficients  $A_{12}$  different from 0). If error  $E_1$  is statistically significantly smaller for the second model it means that Y is causing X conditioned on Z, which is written as  $Y \rightarrow X|Z$ .

## 1.2. Nonlinear methods for Granger causality analysis

The main limitation of the approach presented by Granger is the usage of the linear model. Due to the fact that many real data turn out to be non-stationary processes, this introduces a large limitation for the above-mentioned methods, which assume the stationarity of the tested time series. It is possible to overcome the issue of non-stationarity of the time series by the usage of vector error correction models however, the modeling is still using linear autoregression [26,27]. The limitation coming from the usage of the AR model is that more complex causality dependencies may not be captured by this method [21,28]. Thus, many researchers use other models for prediction instead of linear autoregression. One of the approaches of nonlinear causality testing is Kernel Granger Causality (KGC) [29,30]. It is based on the transformation of the data using a specified kernel function. The inner product of the data and kernel function is used to perform the linear regression. The nonlinearity of this method is controlled by the choice of the kernel function. KGC can be applicable for multivariate problems, it also allows for quantification of the causality and does not suffer from overfitting problem which is a common issue for many methods [31]. Another approach that is growing in popularity is to use neural networks as a forecasting method [14,32–37]. Proposed in 2017, Causal Relationship Estimation by Artificial Neural Network (CREANN) method uses weights of Multilayer Perceptron to assess the causality of the individual lags [36]. This



method is robust with respect to signal-to-noise ratio and model order however, it does not make statistical inference about causality. Another recently proposed method for causality assessment is large-scale nonlinear Granger causality (lsNGC) [14]. In this approach, the Generalized Radial Basis Functions neural network is used as a prediction model. In the experiments conducted by Wis-müller et al. [14], this method outperformed other methods compared, including KGC. The lsNGC was implemented in Python and made available to researchers to use [38]. Other existing Python solutions using neural networks to study causality use lasso regularization [39,40]. By using this technique, some input weights are zeroed and those  $Y$  lag values whose weights are not equal to zero are considered to be the values that cause the time series  $X$ . The results obtained using this approach do not allow for statistical inference using statistical tests and depend on the value of the lambda parameter corresponding to the lasso penalty.

The created Python package presented in this paper was designed to test for causality with precise forecasting thanks to the usage of neural networks. By the usage of dropout and out-of-sample testing, the overfitting problem, which might lead to false causality detection, can be omitted. Moreover, the package allows the quantification of the change in causality over time (which is not enabled by any other available package). Created functions allow for the forecasting using Multilayer Perceptron (MLP) and recurrent neural networks, in particular Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). It also supports the usage of ARIMA model, which is not in the scope of this paper. The approach proposed by the authors does not make the obtained results dependent on an additional parameter or chosen kernel and allows for statistical inference. The package is designed to help scientists use more complex models in terms of Granger causality in an easy user-friendly way without very specific programming knowledge, as well as study causality changes over time, which is not provided by any other framework. It was designed to study the relationship between biological signals, however, it can be widely used in any field of science. The paper aims at presenting the method and corresponding Python package, along with a simulation study showing its usage and relevance.

## 2. Methods

### 2.1. Used models

#### 2.1.1. Multilayer perceptron

MLP is a type of artificial neural network, which is built from many single perceptrons organized in layers. MLP are the most popular artificial neural networks used for forecasting, but also in the field of physiology [41,42]. Each perceptron in the first layer takes as an input the features vector and calculates its output as presented in Eq. (23), where  $w$  are weights and  $f$  is the activation function.

$$y = f(w_0 + w_1 * x_{t-1} + w_2 * x_{t-2} + \dots + w_p * x_{t-p}) \quad (23)$$

Neurons in successive layers take the output values of neurons from previous layers  $y$  as the input vector. In the created package, the activation function in hidden layers is the Rectified Linear Unit (ReLU) function, while in the output layer it is a linear function. The architecture of a simple MLP model based only on past values of  $X$  and predicting the current value of  $X$  for the Granger causality test is presented in Fig. 2.

#### 2.1.2. Long short-term memory

The LSTM is a type of gated recurrent neural network. It is often used for sequence data analysis like text and speech recognition, time series forecasting, or physiological data analysis [43–45]. The LSTM network includes a system of gate units, thanks to which

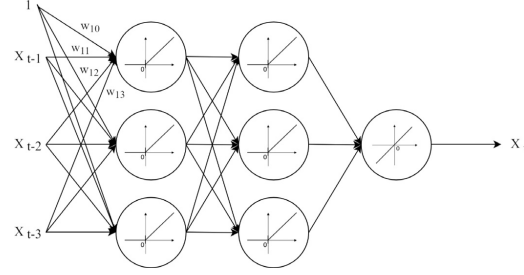


Fig. 2. MLP model which is forecasting the current value of  $X$  based on 3 past values of  $X$  with activation functions and example weights shown.

it is able to control the flow of information, and thus "remember" or "forget" information from previous moments in time. What is more, this type of network is not affected by the gradient vanishing or explosion problem, which is common for recurrent neural networks [46]. The big advantage of using LSTM in causality testing is that, unlike linear regression models, those types of networks do not assume the stationarity of the predicted time series. The LSTM cell takes as input as the input vector  $x(t)$ , the value of the long-term state from the previous time point  $s(t-1)$  and the value of the short-term state from the previous time point  $h(t-1)$ . Instead, it returns the value of the short-term  $h(t)$  and long-term  $s(t)$  state at the current moment in time. The first gate is the forget gate, which controls what part of the information from the past is "remembered" and which one is "forgotten". For this purpose, a sigmoid function is used, and the value passed further from the forget gate is calculated based on Eq. (24).

$$f^{(t)} = \sigma(U^f x^{(t)} + W^f h^{(t-1)} + b^f) \quad (24)$$

Another important gate is the input gate. It determines the degree of status update at a given moment in time. For this, it uses the sigmoid function to select the values that should be used to update the state at a given moment from the input vector and the output vector for the previous moment of time. The formula describing the operation of the input gate is presented in Eq. (25), where  $U^i$  is the gate input weights,  $W^i$  is the recursive weights, and  $b^i$  is the bias of the input gate. For the state to be updated, the candidate values are also calculated using the hyperbolic tangent for the input vector and the output vector of the previous time moment, taking into account the appropriate weights  $U^s$  and  $W^s$  and the bias  $b^s$  as shown in Eq. (26).

$$i^{(t)} = \sigma(U^i x^{(t)} + W^i h^{(t-1)} + b^i) \quad (25)$$

$$\tilde{s}^{(t)} = \tanh(U^s x^{(t)} + W^s h^{(t-1)} + b^s) \quad (26)$$

Updating the LSTM cell state takes place by summing the product of the result obtained on the forget gate  $f^{(t)}$  and the state value at the previous time point  $s^{(t-1)}$  with the product of the result obtained on the input gate with the candidate values. The formula for updating the state is shown in Eq. (27).

$$s^{(t)} = f^{(t)} s^{(t-1)} + i^{(t)} \tilde{s}^{(t)} \quad (27)$$

The last gate in the LSTM cell is the output gate, the operation of which is described in Eq. (28), where  $U^o$  is the gate input weights,  $W^o$  is the recursive weights, and  $b^o$  is the bias of the output gate. Like the other gates, it uses a sigmoidal function to control which state values at a given point in time will be included in the calculation of the final output of the network.

$$o^{(t)} = \sigma(U^o x^{(t)} + W^o h^{(t-1)} + b^o) \quad (28)$$



The computation of the output of the LSTM cell is done by multiplying the hyperbolic tangent of the cell's state and the result obtained from the output gate as shown in Eq. (29).

$$h^{(t)} = o^{(t)} \tanh(s^{(t)}) \quad (29)$$

Thanks to the use of the three above-described gates, the LSTM neural network is able to control the flow of information between consecutive time moments by "remembering" or "forgetting" them. What is more, LSTM networks have greater ease in recognizing long-term dependencies than simple recursive network architectures [47].

### 2.1.3. Gated recurrent unit

GRU neural network was presented by Cho et al. in 2014 [48]. GRU, like LSTM, is a kind of gated recursive neural network. The GRU network is characterized by the use of two gates - a reset gate and an update one. Such a recursive neural network architecture also prevents gradient vanishing and exploding effects [47]. In the GRU cell, the update gate is a kind of counterpart of the forget gate and input gate in the LSTM cell. The update gate controls which information is "forgotten" and which new information from a given point in time will be taken into account in further calculations. The operation of the update gate is based on calculating the value of the sigmoid function for the input data and the cell result from the previous time moment, taking into account the appropriate weights  $U^u$  and  $W^u$  and the bias  $b^u$ , as shown in Eq. (30) [49].

$$u^{(t)} = \sigma(U^u x^{(t)} + W^u h^{(t-1)} + b^u) \quad (30)$$

The reset gate, on the other hand, only serves to control the amount of past information that will be used to compute the candidate state [50]. Like the other gates, the reset gate uses a sigmoidal function with appropriate weights  $U^r$  and  $W^r$  and bias  $b^r$  as presented in Eq. (31).

$$r^{(t)} = \sigma(U^r x^{(t)} + W^r h^{(t-1)} + b^r) \quad (31)$$

After calculating the value of the reset gate, the candidate state value is calculated, which will be used in the next step to calculate the value of the GRU cell state at a given time moment. The candidate state value is calculated using the hyperbolic tangent function, according to Eq. (32), where  $U^h$  and  $W^h$  are the weights, while  $b^h$  corresponds to the bias.

$$\tilde{h}^{(t)} = \tanh(U^h x^{(t)} + W^h (r^{(t)} h^{(t-1)} + b^h)) \quad (32)$$

The value of the state  $h^t$  at a given point in time is computed using the value of the update gate, the value of the candidate state and the value of the state at the previous point in time, according to Eq. (33).

$$h^{(t)} = (1 - u^{(t)}) h^{(t-1)} + u^{(t)} \tilde{h}^{(t)} \quad (33)$$

Thanks to the use of a reset gate and an update gate, the GRU network is able to "forget" and "remember" information at successive time points. GRU networks, as well as LSTM networks, are not affected by the gradient vanishing or explosion problem and do not assume the stationarity of the predicted time series. The recursive neural network of the GRU type has similar advantages to the LSTM type network, however, thanks to the use of two gates instead of three, the GRU network has fewer parameters, which simplifies the learning process and reduces computational complexity.

### 2.2. Statistical significance assessment

The F-test is based on residual sums of squares, which are minimized by linear autoregression models using the least-squares method. This test cannot be performed in the case of using neural networks for prediction, because those models are fitted using

other methods as the "Adam" algorithm and thus value obtained from Eq. (6) for neural networks may not follow the assumed F-distribution [51].

Therefore, we decided to test if the error obtained by the model, which uses both time series for prediction is significantly smaller than the error obtained by the model using only one time series as an input by using Wilcoxon signed-rank test [52].

The null hypothesis is that the median absolute error for a model based on past values of X is equal or smaller than for a model based on past values of X and Y, while the alternative hypothesis is that the model based on past values of both X and Y time series has a smaller median absolute error. In order to examine this hypothesis differences between absolute errors of both models are calculated. Then ranks are assigned to obtained numbers based on their absolute values. The sign of the obtained difference is allocated to the rank and the sum of positive and negative ranks are calculated. In Python implementation, the sum of positive ranks is taken as a statistic value ( $T$ ) based on which p-value is calculated [53]. If the number of differences ( $n$ ) is less than or equal to 25 then the p-value is derived from the tables. If  $n$  is greater than 25, then a normal approximation is applied. In this case, the test statistic ( $z$ ) is calculated according to Eq. (34) and follows the normal distribution [54].

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \quad (34)$$

### 2.3. Description of the created package

Taking all considerations together, we created the Python package, which aims to help researchers to study causality using non-linear prediction methods [55,56]. The use of neural networks for prediction has been implemented using the Keras library [57]. The package consists of 8 functions, which can be split into 2 different types. The first type of functions are simply functions for testing the causality relationship between two time series (with possible additional time series in terms of conditional causality) using the above-mentioned methods and the ARIMA model, which is not in the scope of this paper. The first input of those functions is `NumPy ndarray` with 2 columns, where each column represents one time series, and the number of rows depends on the number of times steps. The second input may be an `int`, `list`, `tuple`, or `NumPy ndarray`. If it is an `int`, then the causality test is made for lags in the range from 1 to the given number, in other cases, the test is made only for numbers given in this variable. In functions using neural networks for prediction additional required inputs are describing the architecture of the network. As an output functions of this type return a `dictionary`, where the keys are the number of lags for which the test was performed. Each key stores a `list`, which contains test results, the model for prediction of X fitted only on X time series and the model for prediction of X fitted on X and Y time series, individual errors and RSS of both models, in case of functions using neural networks additional history of fitting the first model and history of fitting the second model are also stored under this key.

The second type of function is focused on measuring the change of causality over time. The first input for those functions is the same as in the first type, but the number of columns may be greater than 2. If so, then causality between each pair of time series (columns) is measured. As those functions measure the change of causality over time two windows were applied. The first `w1` is responsible for the number of time steps from which the causality is calculated, and the second window `w2` relates to the number of time steps by which the window `w1` is moved after the causality is calculated. So the first value of causality is calculated for time moments from 1st to `w1`th, the second value is calculated for

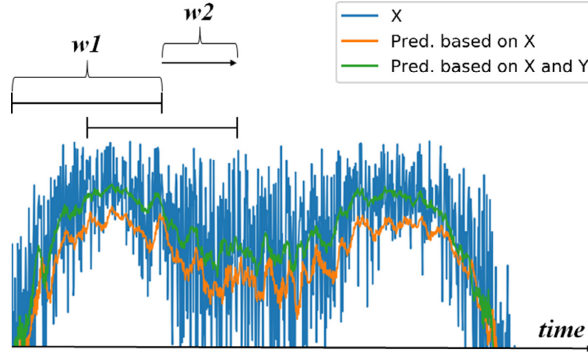


Fig. 3. Visualization of windows used for calculation of change of causality over time.

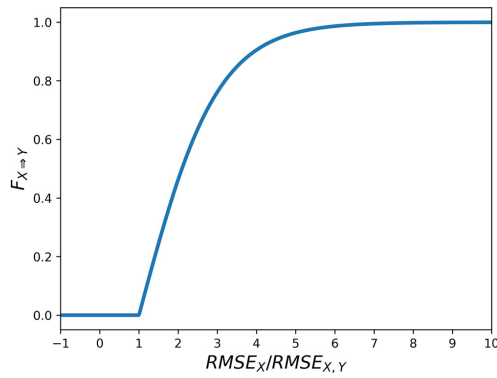


Fig. 4. Graph of proposed causality measure from the quotient of  $RMSE_X$  and  $RMSE_{X,Y}$ .

time moments from  $w2$ th to  $(w1+w2)$ th and so on as presented in Fig. 3.

As the measure presented in Eq. (8) is not applicable for the analysis of the change of causality over time, the new measure of this phenomenon was proposed. This measure of causality is based on the root mean square errors (RMSE) obtained from both models from the currently analyzed window  $w1$ . RMSE used for calculation of causality in the first window is presented in Eqs. (35) and (36).

$$RMSE_X = \sqrt{\sum_{t=1}^{w1} E_X(t)^2 / w1} \quad (35)$$

$$RMSE_{X,Y} = \sqrt{\sum_{t=1}^{w1} E_{X,Y}(t)^2 / w1} \quad (36)$$

The measure proposed by authors is shown in Eq. (37). If the obtained value is smaller than 0 then the result is changed to 0, because a negative number has no sense in terms of causality. The plot of values of the proposed measure in relation to the quotient of  $RMSE_X$  and  $RMSE_{X,Y}$  is presented in Fig. 4.

$$F_{Y \rightarrow X} = \frac{2}{1 + e^{\left(-\frac{RMSE_X}{RMSE_{X,Y}} + 1\right)}} - 1 \quad (37)$$

### 3. Results

#### 3.1. Simulated signals

In this paper, we would like to present the usage and possibilities of the created package in terms of detecting the causality by different models in two cases - with the presence of the causality between the signals and with no dependencies between the time series. The results of the causality detection and forecasting accuracy were compared with linear and nonlinear methods. What is more, we would like to present the unique feature of illustrating the change of causality over time using the chosen model. To present the performance of the package two signals with 10,000 samples each were generated, where  $Y$  is a periodic function and  $X$  is a polynomial function of  $Y$  delayed by 100-time steps as presented in Eqs. (38) and (39) and in Listing 1. Some random noise ( $E_1$  and  $E_2$ ) also was added to both signals in order to make them more real-like. Both  $X$  and  $Y$  time series were visualized in Fig. 5.

$$Y(t) = \cos(t) + \sin(0.15 * t) + E_1(t) \quad (38)$$

$$X(t) = 2 * Y(t - 100)^3 - 5 * Y(t - 100)^2 + 0.3 * Y(t - 100) + 2 + E_2(t) \quad (39)$$

The first 70% of the signals were used as training data and the remaining 30% was used as a test dataset. The signals prepared in this way can be used for the causality analysis using the first kind of functions included in the developed module. It was decided to perform a causality test for two lags - smaller and greater than the actual delay. The lags were equal to 50 and 150.

Using the created Python module (version 1.0.3), three types of neural networks with the following architectures were applied for causality analysis:

- Two LSTM layers with 10 cells each (LSTM) followed by one output neuron with linear activation function;
- Two GRU layers with 10 cells each (GRU) followed by one output neuron with linear activation function;
- Two fully connected layers with 100 neurons each (MLP) and ReLU activation function followed by one output neuron with linear activation function.

For each network dropout regularization technique was used with a dropout rate equal to 0.01 (arguments *NN\_config* and *NN\_neurons* for MLP and *Dropout\_rate* in case of GRU and LSTM). Neural networks were trained for 150 epochs, with a learning rate

```
y = np.cos(np.linspace(0,20,10100)) + np.sin(np.linspace(0,3,10100)) - 0.2*np.random.random(10100)
x = 2*y**3 - 5*y**2 + 0.3*y + 2 - 0.05*np.random.random(10100)
```

Listing 1. Generation of two signals.

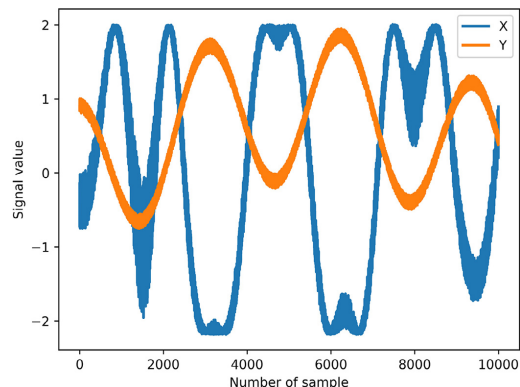


Fig. 5. Visualization of signals X and Y.

equal to 0.001 and 0.0001 for the first 50 epochs and the last 100 epochs, respectively. To obtain the most accurate models LSTM and GRU neural network models were created and trained 2 times (parameter *run*) and the MLP model 5 times. For final causality testing models (one based on X and one based on both X and Y) with the smallest RSS were chosen. The difference in the value of the *run* parameter was due to the long time needed for LSTM and GRU training. The usage of the created functions is presented in Listing 2.

Regarding the detection of causality and the prediction performance, the presented methods were compared with models used in two other methods for causality testing, both of which allow for out-of-sample forecasting and are as well implemented in Python. The first of these methods was a nonmodified Granger test, which uses linear autoregression (AR) for prediction. The second one was large-scale nonlinear Granger causality (lsNGC) [14,38], which is based on a Generalized Radial Basis Functions (GRBF) neural network. Its parameters  $c_f$  and  $c_g$ , which are the number of hidden layer neurons in the GRBF networks were set to 25. All models were fitted on the training set and used on the test set. The absolute values of the prediction errors obtained on a test set from a model based on the past of X and a model based on the past of both signals were compared using Wilcoxon signed-rank test, to determine the presence of causality. Performance of all methods was quantified using mean squared error (MSE), mean absolute error (MAE) and median absolute error (MedAE). Moreover, it was assessed if the usage of the proposed results in improvement of the prediction accuracy models in comparison to AR or GRBF network. For this purpose, Wilcoxon signed-rank test was used to compare the absolute values of errors obtained from corresponding models (e.g. MLP model based on past of X and AR model based also on a past of X). The effect size of including the past of signal Y in prediction for each method was assessed using Cohen's d. In order to better visualize the results plots of predicted values versus the true values of X and plots of prediction error versus the predicted X values were prepared for each model. The steps of the entire process of causality analysis were presented in Fig. 6.

In the next step of the analysis, the Y time series was replaced with the random noise to compare the results from the tests in the

case where the causality relation should be detected with the case where there is no relation between the signals. This analysis was performed as well using the first 70% of the signal as the training set and the last 30% as the test set. The same metrics and tests were performed as described in the previous paragraph.

The second type of function in the package can be used to examine the change of causality over time. In order to simulate such a change of causality relation the first 50% of the test Y signal was changed to random noise (Fig. 7), so there would be no causal relation between X and Y at the first half of the test time series. To present this feature of the package the MLP architecture which obtained the highest Cohen's d was chosen. Mentioned in Section 2.2. Window values were set to 30 and 1 for  $w_1$  and  $w_2$ , respectively. The example usage of the function for assessment of the change of causality over time was presented in Listing 3.

The assumed significance level is equal to 0.05. All analyses were performed using Python version 3.7.10.

### 3.2. Presence of causality from Y to X

Each of the designed functions generates a plot of original testing signal X, values predicted by the model based only on past values of X and values predicted by the model based on past values of X and Y. For each function based on neural networks, it is possible to obtain a history of the fitting from the output of the function and create the learning curve. Sample plots of the original and predicted data obtained from function *nonlincausalityNN* for lag equal to 50 and 150 are presented in Fig. 8 while learning curves of the models returned by this function are presented in Fig. 9.

All error metrics calculated on a testing set for each model for lag equal to 50 and 150 are presented in Table 1. P-values obtained from the Wilcoxon signed-rank test used to assess the presence of causality from Y to X for each method and each lag are presented in Table 2.

In case of lag equal to 50, all models from a created package (NN) obtained similar results in terms of error metrics. The AR models had a similar accuracy of prediction, while GRBF models tend to obtain the biggest error metrics for model based only on the past values of X and the smallest one in the case of model based on both signals, while model based on both time series obtained slightly smaller error metrics than LSTM, GRU and AR models. For the lag of 150, NN and AR mostly obtained similar results, except for MLP model based on X and Y which outperformed all other models and got the smallest error in the case of all three metrics. In the case of the bigger lag GRBF model had a drop in performance for the model based only on X signal. In the case of the causality test, all nonlinear methods obtained a p-value smaller than the assumed significance level, thus in the case of using any of the presented models or GRBF the causality relationship between signals was detected, even for lag smaller, than the actual delay between the time series. The autoregressive model used in the state-of-the-art Granger method did not capture the causality for any given lag.

P-values obtained from the Wilcoxon signed-rank test used to assess the improvement in prediction due to usage of neural networks over autoregressive and GRBF models are presented in Table 3 and Table 4, respectively. Cohen's d used to assess the effect size of incorporating the past of Y signal into the models are presented in Table 5. Plots of predicted values from actual values are presented in Figs. 10, and 11. for lag 50 and 150, respectively and

```
lags = [50,150]

results_LSTM = nlc.nonlincausalityLSTM(data_train,maxlag=lags,xtest=data_test,LSTM_layers=2,
                                       LSTM_neurons=[10,10],run=2,epochs_num=[50,100],
                                       learning_rate=[0.001,0.0001],batch_size_num=128,
                                       Dropout_rate=0.01, plot=True,verbose=False)

results_GRU = nlc.nonlincausalityGRU(data_train,maxlag=lags,xtest=data_test,GRU_layers=2,
                                     GRU_neurons=[10,10],run=2,epochs_num=[50,100],
                                     learning_rate=[0.001,0.0001],batch_size_num=128,plot=True,
                                     Dropout_rate=0.01, verbose=False)

results_NN =
nlc.nonlincausalityNN(data_train,maxlag=lags,xtest=data_test,NN_config=['d','dr','d','dr'],
NN_neurons=[100,0.01,100,0.01],run=5,epochs_num=[50,100],
learning_rate=[0.001,0.0001], batch_size_num=128, plot=True,
verbose=False)
```

Listing 2. Using the functions for causality testing.

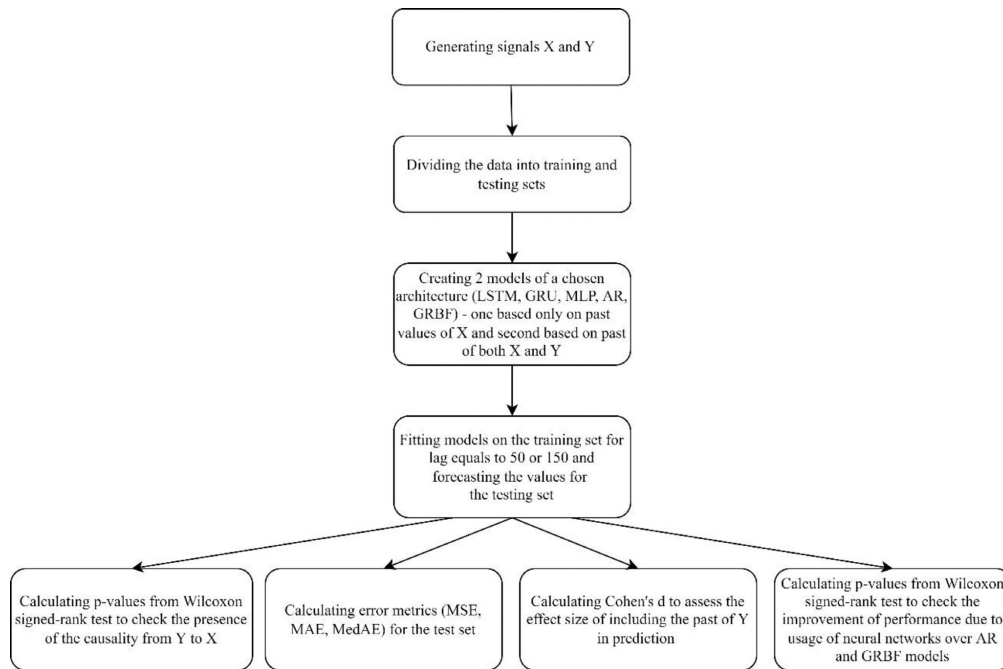


Fig. 6. The process scheme of the data analysis from generating the data up to the causality and prediction assessment. In the case of real-world application, the process would be the same, but X and Y would be the data obtained from the study.

plots of the prediction error from predicted values are presented in Figs. 12 and 13.

The usage of neural networks over AR did not statistically significantly improve the prediction of the X signal if taking only past values of X as an input. On the other hand, the usage of NN over GRBF with one signal as an input always results in significantly

better prediction accuracy. If the model was based on both X and Y signals the usage of neural networks instead of the autoregressive model improved the prediction in 5 out of 6 cases (only the MLP model for 50 lags did not obtain significantly better results). In the case of GRBF, only MLP model for lag equal to 150 outperformed it. For both lags, the effect size of incorporating past of Y

```
res_meas_NN = nlc.nonlincausalitymeasureNN(data_train, lags, w1=30,w2=1, xtest=data_test_measure,
NN_Config=['d','dr','d','dr'], NN_neurons=[100,0.01,100,0.01], run=5, epochs_num=[50,100],
learning_rate=[0.001,0.0001], batch_size_num=128, verbose=False, plot=True,
plot_with_xtest=True)
```

Listing 3. Usage of the function for measuring the causality change over time with MLP models.

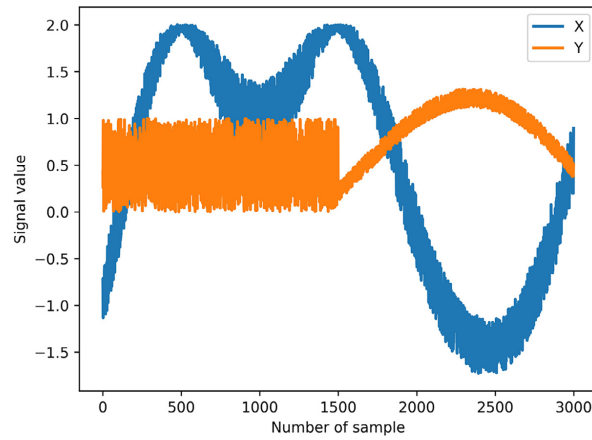


Fig. 7. Visualization of test signals X and Y with the first 50% of the test Y signal changed to random noise.

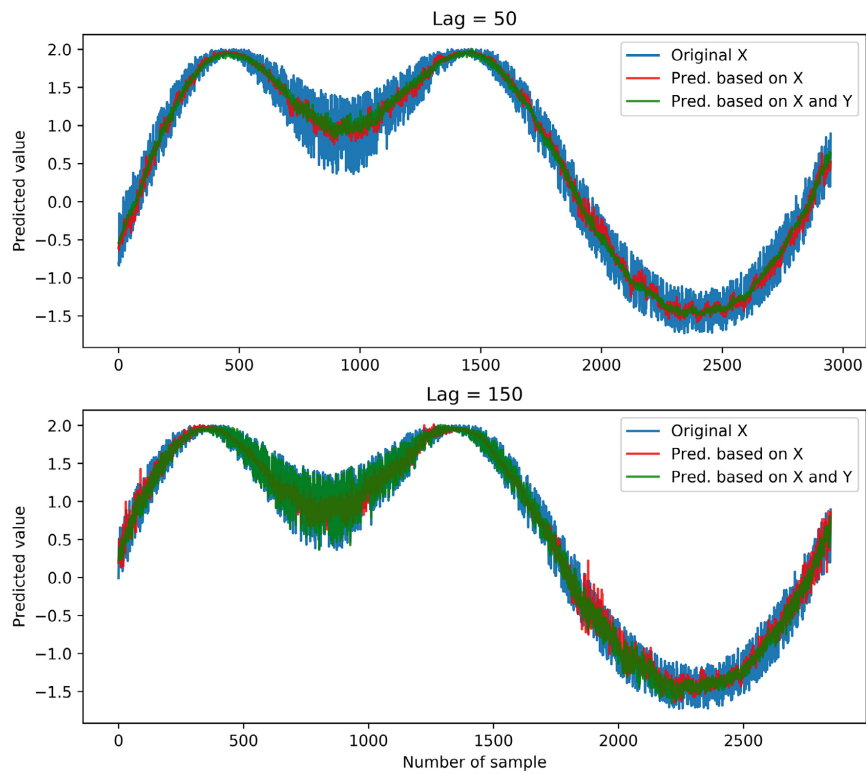


Fig. 8. The plot of the original value of X, predicted values from the model based on the past values of X and predicted values from the model based on past values of X and Y. If there is causality from Y to X, then the values predicted based on both signals should be closer to the actual values than the values predicted only with the past X values (prediction error of model based on both signals should be lower, then the error of model based only on X).

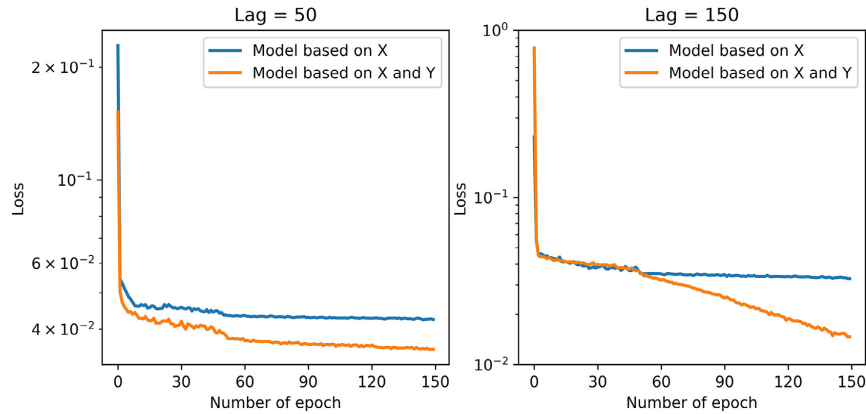


Fig. 9. Plot presenting the dependencies between loss and number of epochs for lag equal to 50 and 150.

Table 1

Error metrics (MSE, MAE and MedAE) obtained on a test set for each model created based on 50 and 150 past values in the case where  $Y \rightarrow X$ . The smallest error metric for a given case is shown in bold.

| Lag value | Metrics | Model trained on | LSTM         | GRU          | MLP          | AR           | GRBF         |
|-----------|---------|------------------|--------------|--------------|--------------|--------------|--------------|
| 50        | MSE     | X                | 0.042        | <b>0.041</b> | 0.042        | <b>0.041</b> | 0.057        |
|           |         | X and Y          | 0.039        | 0.039        | 0.039        | 0.041        | <b>0.038</b> |
|           | MAE     | X                | 0.162        | <b>0.161</b> | 0.162        | <b>0.161</b> | 0.188        |
|           |         | X and Y          | 0.158        | 0.158        | 0.161        | 0.161        | <b>0.155</b> |
|           | MedAE   | X                | 0.139        | 0.138        | <b>0.136</b> | <b>0.136</b> | 0.153        |
|           |         | X and Y          | 0.137        | <b>0.135</b> | 0.137        | 0.140        | <b>0.135</b> |
| 150       | MSE     | X                | <b>0.041</b> | <b>0.041</b> | 0.045        | <b>0.041</b> | 0.254        |
|           |         | X and Y          | 0.039        | 0.038        | <b>0.011</b> | 0.041        | 0.036        |
|           | MAE     | X                | <b>0.159</b> | 0.161        | 0.166        | 0.160        | 0.413        |
|           |         | X and Y          | 0.156        | 0.156        | <b>0.078</b> | 0.160        | 0.151        |
|           | MedAE   | X                | 0.135        | 0.136        | <b>0.132</b> | 0.137        | 0.379        |
|           |         | X and Y          | 0.134        | 0.134        | <b>0.056</b> | 0.136        | 0.130        |

Table 2

P-values for each model and each tested lag obtained from the Wilcoxon signed-rank test in case where  $Y \rightarrow X$ . Cases, where a causal relationship has been detected, are shown in bold.

| Lag value | LSTM           | GRU            | MLP            | AR    | GRBF           |
|-----------|----------------|----------------|----------------|-------|----------------|
| 50        | < <b>0.001</b> | < <b>0.001</b> | <b>0.025</b>   | 0.703 | < <b>0.001</b> |
| 150       | < <b>0.001</b> | < <b>0.001</b> | < <b>0.001</b> | 0.261 | < <b>0.001</b> |

Table 3

P-values form a comparison of the models performances, with the alternative hypothesis that error obtained from neural networks models is smaller than for autoregressive ones in a case where  $Y \rightarrow X$ . Cases, where usage of the neural network resulted in significantly better performance, are shown in bold.

| Lag value | Model trained on | LSTM           | GRU            | MLP            |
|-----------|------------------|----------------|----------------|----------------|
| 50        | X                | 0.969          | 0.899          | 0.944          |
|           | X and Y          | < <b>0.001</b> | < <b>0.001</b> | 0.068          |
| 150       | X                | 0.227          | 0.814          | 0.991          |
|           | X and Y          | <b>0.010</b>   | <b>0.008</b>   | < <b>0.001</b> |

signal into prediction was the highest for MLP and the smallest for AR. The biggest Cohen's d equal to 0.659 was obtained for MLP for lag equal to 150. Prediction error was equally distributed around 0 for the whole range of predicted values.

Table 4

P-values form a comparison of the model performances, with the alternative hypothesis that absolute error obtained from neural networks models is smaller than for GRBF ones in a case where  $Y \rightarrow X$ . Cases, where usage of the neural network resulted in significantly better performance, are shown in bold.

| Lag value | Model trained on | LSTM           | GRU            | MLP            |
|-----------|------------------|----------------|----------------|----------------|
| 50        | X                | < <b>0.001</b> | < <b>0.001</b> | < <b>0.001</b> |
|           | X and Y          | 0.991          | 0.998          | 1.000          |
| 150       | X                | < <b>0.001</b> | < <b>0.001</b> | < <b>0.001</b> |
|           | X and Y          | 1.000          | 1.000          | < <b>0.001</b> |

Table 5

Cohen's d for the case where  $Y \rightarrow X$ . The highest Cohen's d for each lag are shown in bold.

| Lag value | LSTM  | GRU   | MLP          | AR    | GRBF  |
|-----------|-------|-------|--------------|-------|-------|
| 50        | 0.047 | 0.046 | <b>0.049</b> | 0.005 | 0.013 |
| 150       | 0.037 | 0.044 | <b>0.659</b> | 0.001 | 0.054 |

### 3.3. Absence of causality from Y to X

The metrics calculated for the analysis, where there was no relationship between the time series (with the Y changed to random noise) are presented in Table 6. The results from the Wilcoxon signed-rank test used to test for the presence of causality are presented in Table 7.

For the lag of 50 neural networks obtained the same or higher error measure values than autoregressive models. For lag equal to

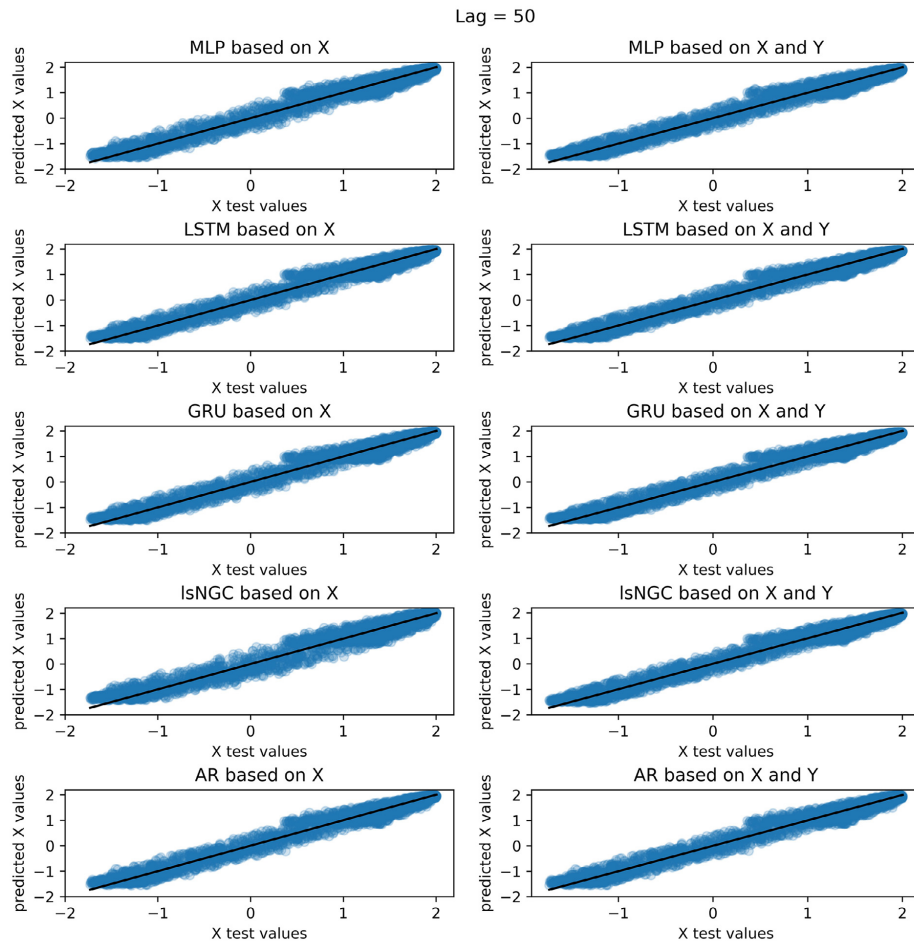


Fig. 10. The plot of predicted values from true X values for lag equals 50. The solid black line represents  $y$  equals to  $x$ .

Table 6

Error metrics (MSE, MAE and MedAE) obtained on a test set for each model created based on 50 and 150 past values in the case where there was no causality relation between signals. The smallest error metric for a given case is shown in bold.

| Lag value | Measure | Model trained on | LSTM         | GRU          | MLP          | AR           | GRBF  |
|-----------|---------|------------------|--------------|--------------|--------------|--------------|-------|
| 50        | MSE     | X                | 0.042        | <b>0.041</b> | 0.042        | <b>0.041</b> | 0.055 |
|           |         | X and Y          | 0.042        | <b>0.041</b> | 0.042        | <b>0.041</b> | 0.055 |
|           | MAE     | X                | 0.162        | <b>0.161</b> | 0.163        | <b>0.161</b> | 0.184 |
|           |         | X and Y          | 0.162        | <b>0.161</b> | 0.164        | <b>0.161</b> | 0.184 |
|           | MedAE   | X                | 0.138        | 0.137        | <b>0.136</b> | <b>0.136</b> | 0.152 |
|           |         | X and Y          | 0.138        | 0.137        | 0.138        | <b>0.135</b> | 0.150 |
| 150       | MSE     | X                | <b>0.041</b> | <b>0.041</b> | 0.045        | <b>0.041</b> | 0.278 |
|           |         | X and Y          | <b>0.041</b> | <b>0.041</b> | 0.049        | <b>0.041</b> | 0.276 |
|           | MAE     | X                | 0.161        | <b>0.159</b> | 0.168        | 0.160        | 0.443 |
|           |         | X and Y          | <b>0.160</b> | <b>0.160</b> | 0.173        | 0.162        | 0.442 |
|           | MedAE   | X                | 0.137        | <b>0.134</b> | 0.135        | 0.137        | 0.417 |
|           |         | X and Y          | 0.137        | <b>0.136</b> | 0.140        | 0.137        | 0.411 |



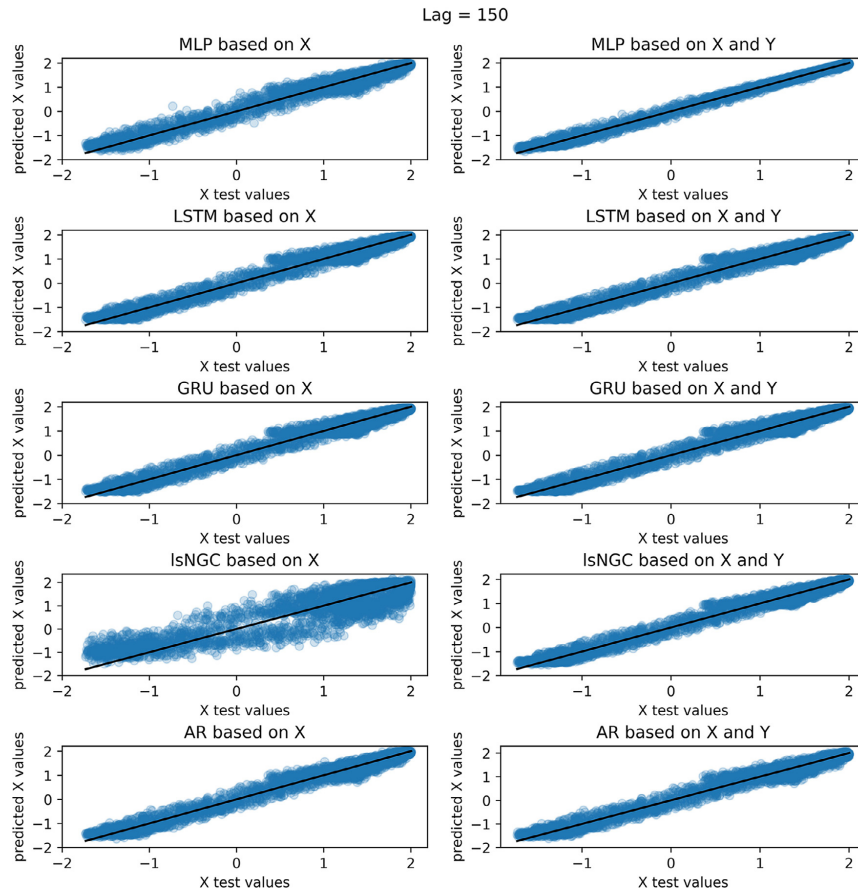


Fig. 11. The plot of predicted values from true X values for lag equals 150. The solid black line represents  $y$  equals to  $x$ .

Table 7

P-values for each model and each tested lag obtained from the Wilcoxon signed-rank test in the case where there was no causality relation between signals. Cases, where a causal relationship has been detected, are shown in bold.

| Lag value | LSTM  | GRU   | MLP   | AR    | GRBF         |
|-----------|-------|-------|-------|-------|--------------|
| 50        | 0.558 | 0.267 | 0.930 | 1.000 | 0.127        |
| 150       | 0.073 | 0.592 | 0.999 | 1.000 | <b>0.024</b> |

150 the measures were mostly higher for MLP, mostly lower for GRU and mostly equal to those obtained by AR for LSTM. The GRBF compared to other models obtained slightly higher error metrics for lag equal to 50 and several times larger for lag equal to 150. Causality was not detected using any of the models except the GRBF model for a delay of 150, where false causality was indicated.

P-values obtained from the Wilcoxon signed-rank test used to assess the improvement in prediction due to usage of neural networks over autoregressive and IsNGC models are presented in Table 8 and Table 9, respectively. Cohens'd used to assess the ef-

Table 8

P-values from a comparison of the model performances, with the alternative hypothesis that error obtained from neural networks models is smaller than for autoregressive ones in the case where there was no causality relation between signals. Cases, where usage of the neural network resulted in significantly better performance, are shown in bold.

| Lag value | Model trained on | LSTM         | GRU          | MLP   |
|-----------|------------------|--------------|--------------|-------|
| 50        | X                | 0.982        | 0.941        | 0.999 |
|           | X and Y          | 0.857        | 0.486        | 1.000 |
| 150       | X                | 0.607        | 0.219        | 1.000 |
|           | X and Y          | <b>0.041</b> | <b>0.038</b> | 1.000 |

fect size of incorporating the past of Y signal (random noise in this case) into the model are presented in Table 10.

The only models which performed statistically significantly better than AR were GRU and LSTM models based on both X and Y for lag equal to 150. All NN models performed better than GRBF models in all cases. The highest Cohens'd was obtained again for MLP models, but all values were much smaller compared to those in the case where  $Y \rightarrow X$ .



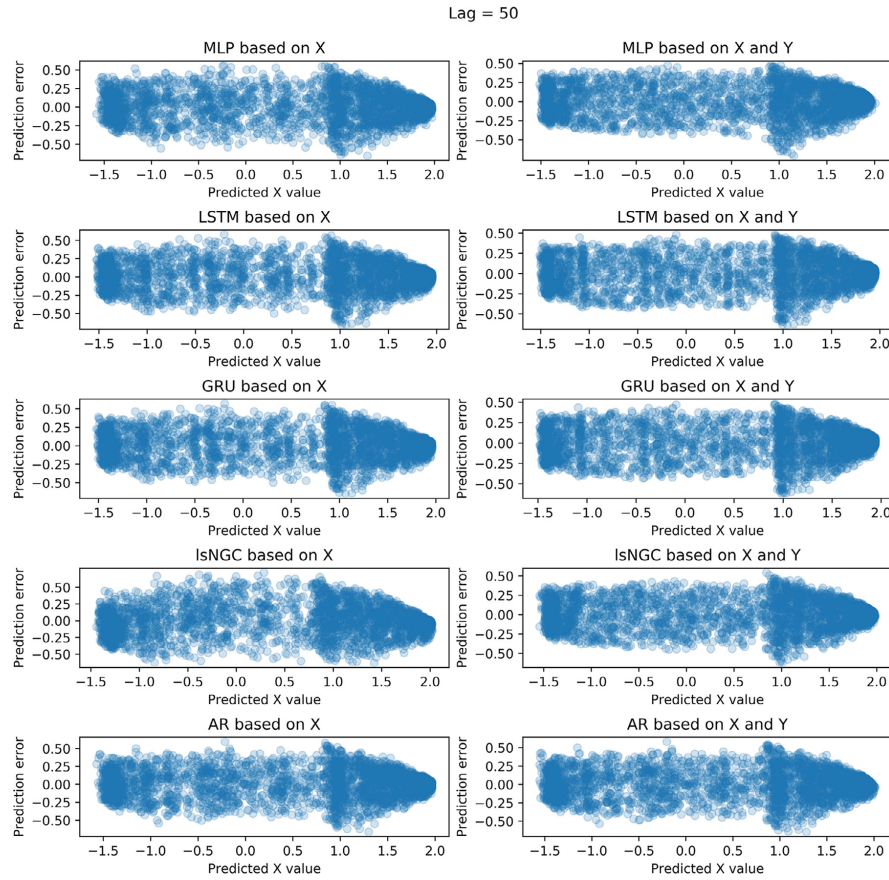


Fig. 12. The plot of prediction errors from predicted X values for lag equals 50.

Table 9

*P*-values form a comparison of the model performances, with the alternative hypothesis that error obtained from neural networks models is smaller than for GRBF ones in the case where there was no causality relation between signals. Cases, where usage of the neural network resulted in significantly better performance, are shown in bold.

| Lag value | Model trained on | LSTM           | GRU            | MLP            |
|-----------|------------------|----------------|----------------|----------------|
| 50        | X                | < <b>0.001</b> | < <b>0.001</b> | < <b>0.001</b> |
|           | X and Y          | < <b>0.001</b> | < <b>0.001</b> | < <b>0.001</b> |
| 150       | X                | < <b>0.001</b> | < <b>0.001</b> | < <b>0.001</b> |
|           | X and Y          | < <b>0.001</b> | < <b>0.001</b> | < <b>0.001</b> |

Table 10

Cohens'd for the case where there was no causality relation between signals. The highest Cohens'd for each lag are shown in bold.

| Lag value | LSTM    | GRU     | MLP          | AR      | GRBF    |
|-----------|---------|---------|--------------|---------|---------|
| 50        | < 0.001 | < 0.001 | <b>0.004</b> | < 0.001 | < 0.001 |
| 100       | < 0.001 | < 0.001 | <b>0.005</b> | < 0.001 | < 0.001 |

### 3.4. Change of the causality over time

The analysis of the change of causality over time was performed using the same MLP architecture as for causality testing which managed to obtain the biggest Cohen's d in the analysis. The measure of causality according to Eq. (37) and its change over time along with the original time series for better visualization of the dependency for 50 lags was presented in Fig. 14 and for 150 lags in Fig. 15.

For lag equal to 150 when the causality relation from signal Y to signal X appears (after the random noise) there is a visible increase in the measure of causality for  $Y \rightarrow X$ . After the random noise the measure for  $Y \rightarrow X$  tends to be much higher, than for  $X \rightarrow Y$  (as there is no such causality). In case of lag equal to 50 in part of the plot where there is the random noise, the measures are similar for both  $X \rightarrow Y$  and  $Y \rightarrow X$ , while after this part the measure for  $Y \rightarrow X$  seems to be slightly bigger than for  $X \rightarrow Y$  (but smaller than for the lag of 150). In the middle of the signals for both lags, there is an increase in causality from X to Y, which means that in that part the model was able to benefit from using the X signal in forecasting the Y in terms of decreasing the prediction error. The maximum of the causality measure for the lag of 50 is equal to 0.324 and 0.139

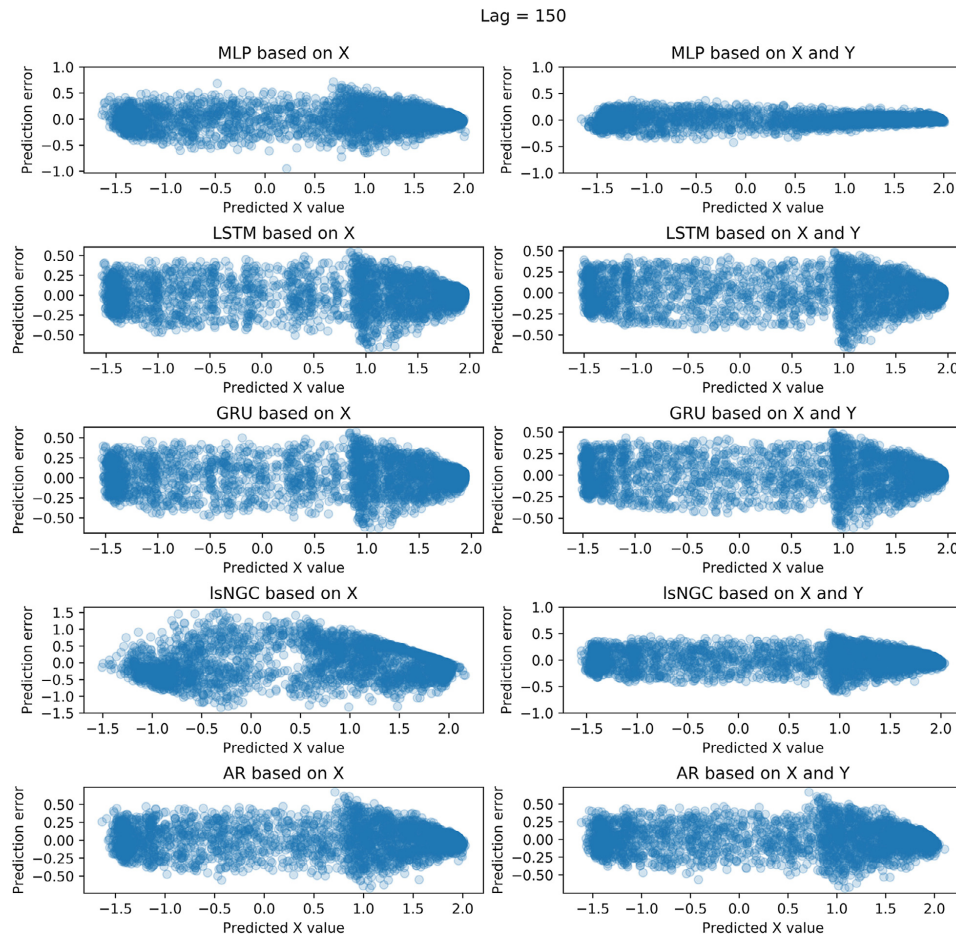


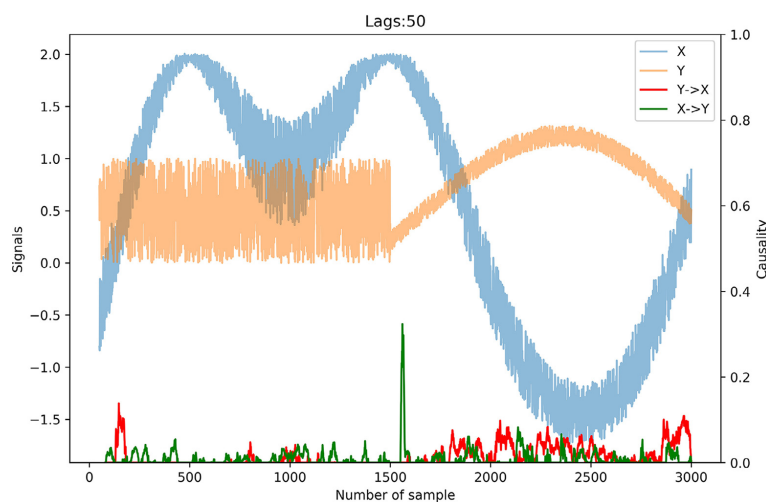
Fig. 13. The plot of prediction errors from predicted X values for lag equals 150.

for  $X \rightarrow Y$  and  $Y \rightarrow X$ , respectively and for lag equal to 150 those values were equal to 0.318 and 0.620. The mean value of the causality for lag equal to 50 was equal to 0.009 and 0.014 for  $X \rightarrow Y$  and  $Y \rightarrow X$ , respectively, while for lag equal to 150 those values were equal to 0.007 for  $X \rightarrow Y$  and 0.128 for  $Y \rightarrow X$ . However, it should be emphasized that the mean and maximum values of the causality measure are showing the gain from incorporating the other variable into prediction and those values cannot be interpreted as a presence of causality between signals on their own. As showed in the previous sections the assessment of the presence of causality is performed using Wilcoxon signed-rank test.

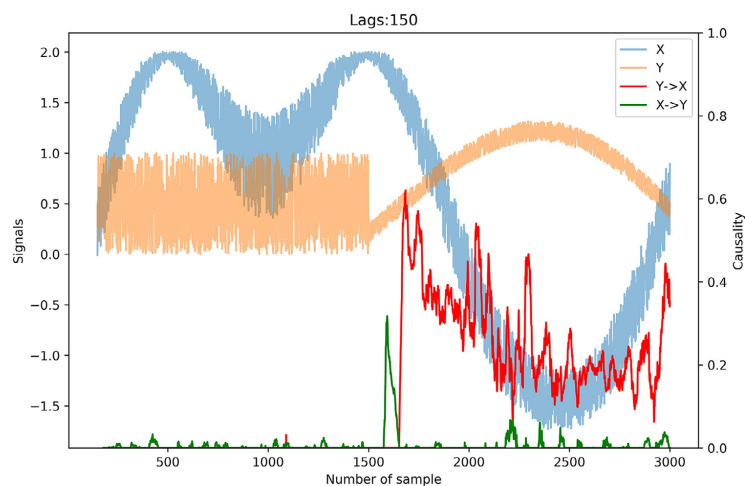
#### 4. Discussion

The methods presented in this paper are focused on overcoming the main disadvantages of the non-modified Granger method based on linear autoregressive models and proposing an alternative for existing nonlinear methods. The first weakness of the linear Granger causality test is that of using AR models the state-of-the-

art Granger approach may not capture the more complex, nonlinear causality dependencies like polynomial, exponential, logarithmic, or others that are visible in many biomedical, physiological, economic and social measurements/signals. What is more, the AR models are assuming the stationarity of the signals, which is problematic as most real-life signals are not stationary (however this problem might be also addressed by usage of vector error correction model [26,27], but this approach is still assuming the linearity of the dependency and is not implemented in the existing Python packages). Thanks to the usage of neural networks models it is possible not only to detect the causal relations, which are nonlinear but also to test the dependencies between nonstationary time series and obtain very accurate forecasting results. In order to test the proposed approach and the created Python package, two time series were simulated, where the signal X was a polynomial of the Y signal and was delayed in relation to it by 100-time steps. The data analysis was performed in 2 different ways. In the first approach, both models (based on the past of X and the past of X and Y) were trained on the first 70% of the data and testing for causal-



**Fig. 14.** The measure of causality (Eq. (37)) from signal Y to signal X ( $Y \rightarrow X$ ) and from signal X to Y ( $X \rightarrow Y$ ), along with signal X and Y for analysis for 50 lags. The axis on the left represents the signal values, and the axis on the right represents the causality value.



**Fig. 15.** The measure of causality (Eq. (37)) from signal Y to signal X ( $Y \rightarrow X$ ) and from signal X to Y ( $X \rightarrow Y$ ), along with signal X and Y for analysis for 150 lags. The axis on the left represents the signal values, and the axis on the right represents the causality value.

ity on the remaining 30% to check the robustness of the methods. The second analyses were similar to the first one, but the Y signal was changed to a random noise so there was no causality relation between X and Y. This change was made to test if the proposed methods do not indicate the false causality relations. The proposed models were compared with the autoregressive model used in the traditional Granger method and with the Generalized Radial Basis Functions neural network used in IsNGC approach [14].

The created functions based on neural networks were able to detect the causality relation between time series for lag both smaller and greater, than the actual delay between the two signals. The GRBF models also indicated the presence of true causality, unlike the AR ones, which were not able to detect this depen-

dency. The biggest difference between the results for lag bigger and smaller than the actual delay between X and Y signals were the p-value, Cohens'd and error metrics for MLP models. The p-value was many orders of magnitude smaller for the lag equal to 150, while Cohen's d was much higher for this lag value. The MLP model based on both X and Y obtained much smaller error metrics in case of the lag equal to 150 compared to the results for lag equal to 50. In the case of lag equal to 150, there was a large drop in the precision of GRBF model based on the past of X compared to the same model but for a smaller lag value. In the case of other models, the performance was rather similar irrespective of the value of the lag. The plot of predicted value from the actual value was much more centered around the line  $y$  equals to  $x$  for

MLP model using both signals for lag equal to 150, which also indicates that the most accurate prediction was for this model. In the case of GRBF model using only X signal for lag equal to 150, the points are the least concentrated around the mentioned line, which also confirms its lowest accuracy. When it comes to the plot of the prediction error from the predicted values, the MLP model based on X and Y for lag equals 150 is distinguished, where the error at higher predicted values is much smaller than in the case of other models, while for GRBF based on X there is visible more dispersion of the points, which indicates the lowest prediction performance. The best results in terms of Cohen's d and error metrics were obtained for MLP for lag equal to 150 for models based on both X and Y signals. The usage of NN instead of AR mostly resulted in more accurate results in the case of models based on both signals, while compared to GRBF usage of NN was always more efficient for models based only on the past of X.

When Y time series was changed to a random noise to test if proposed methods do not indicate the false dependency when there is no causal relation between examined time series, all neural networks obtained from the created package and AR models did not detect any causality between the signals for any lag value. In the case of GRBF the false causality was detected for lag equal to 150. Error metrics and Cohen's d obtained by NN and AR models were similar to each other, while for GRBF the prediction accuracy was the lowest especially in case of lag equal to 150.

Compared to the autoregressive models used in traditional Granger causality analysis the proposed approach was characterized by better or similar prediction performance and discovery of nonlinear causality relations undetected by the AR models. In the case of comparison with the nonlinear approach using GRBF the proposed solution was always superior in prediction based on the past of only one variable and did not indicate any false causality, which appeared to be the issue in one case for GRBF models. The advantage of the proposed method over other approaches using non-linear data transforms such as KGC is the lack of assumptions about the data transformation (kernel used). The prepared Python package is superior to other existing solutions that use neural networks with lasso regularization to study causality [39,40] as it allows for statistical inference and the results are independent of the lambda parameter used in the lasso regularization. As the neural networks applied in this research are already used in the analysis of physiological data [41,42,45], the created package can also find wide application in the study of causal relationships in physiology, but also other scientific fields.

To our knowledge, a novelty of the created package is the proposed method of studying the change in causality over time, which may allow for a better understanding of the causal relationships between the signals. For both lags, the causality values seem to be bigger for the part of the signal without random noise. As expected, the plots of the change of causality over time show much more causal dependency for the lag equal to 150, than for lag equal to 50. For the higher lag, there is very clearly visible the moment when the causality relationship appears. The value of the causality measure varies over time, probably due to the added noise to the signal and to the out-of-sample testing. This feature of the package can be very useful especially in the case of signals which dependence varies over time.

The limitations of the prepared package are the computational complexity and time-consuming involved in training neural networks. However, the benefits of using neural networks seem to outweigh these issues. The limitation of the study is the use of simulated data, so further studies on real-world data are planned. In the near future, we plan to focus on using the created package to investigate the causal relationships between biomedical signals and their dependence on various vital parameters as a continuation of research conducted by Młyńczak and Krysztofak [16,17].

We plan, among others, to use the package in the analysis of the causality between the respiratory signal from impedance pneumography (tidal volume equivalent) and the cardiological signal from the ECG (mainly RR-intervals and tachogram as their interpolation) and to investigate the causality phenomenon in various groups of patients (network physiology paradigm [58]).

## 5. Conclusion

The usage of neural networks in causality testing allows capturing the nonlinear causal dependencies, which are not detected by the AR model used in the state-of-the-art Granger method. In the case when there is no causal relation neural-network-based methods do not indicate false causality, which might be an issue while using GRBF model for prediction. Usage of neural networks allowed to provide better prediction results especially in the case of multilayer perceptron taking as an input past values of both time series for a lag value greater than the actual delay. This model obtained the highest effect size of incorporating the past of Y signal into the prediction model. The measure of the change of causality over time seems to be a valid feature that allows to better understand the detected dependencies between the signals. The created package can be widely used (is available in PyPI [56]) in the analysis of signals in different scientific fields like neuroscience, physiology, or economy. Thanks to the proposed method, it is possible to study nonlinear dependencies, study causality changes over time, and unlike similar nonlinear approaches, it is easily usable thanks to the package created in Python.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This research was not financially supported by any institution or organization. No ethical approval was required. The authors declare that they have no conflict of interests.

## References

- [1] C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* 37 (1969) 424, doi:10.2307/1912791.
- [2] R.P. Maradana, R.P. Pradhan, S. Dash, D.B. Zaki, K. Gaurav, M. Jayakumar, A.K. Sarangi, Innovation and economic growth in European economic area countries: the Granger causality approach, *IIMB Manag. Rev.* (2019), doi:10.1016/j.iimb.2019.03.002.
- [3] M.O. Appiah, Investigating the multivariate Granger causality between energy consumption, economic growth and CO<sub>2</sub> emissions in Ghana, *Energy Policy* (2018), doi:10.1016/j.enpol.2017.10.017.
- [4] V. Troster, M. Shahbaz, G.S. Uddin, Renewable energy, oil prices, and economic activity: a Granger-causality in quantiles analysis, *Energy Econ.* (2018), doi:10.1016/j.eneco.2018.01.029.
- [5] R.P. Pradhan, M.B. Arvin, S. Bahmani, Are innovation and financial development causative factors in economic growth? Evidence from a panel Granger causality test, *Technol. Forecast. Soc. Change* (2018), doi:10.1016/j.techfore.2018.01.024.
- [6] S. Hu, Y. Cao, J. Zhang, W. Kong, K. Yang, Y. Zhang, X. Li, More discussions for granger causality and new causality measures, *Cogn. Neurodyn.* 6 (2012) 33–42, doi:10.1007/s11571-011-9175-8.
- [7] M. Ding, Y. Chen, S.L. Bressler, Granger causality: basic theory and application to neuroscience, in: *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, Wiley-VCH Verlag GmbH & Co. KGaA, 2006, pp. 437–460.
- [8] S. Cekić, D. Grandjean, O. Renaud, Time, frequency, and time-varying Granger-causality measures in neuroscience, *Stat. Med.* (2018), doi:10.1002/sim.7621.
- [9] F. Azarmi, S.N. Miri Ashtiani, A. Shalbaf, H. Behnam, M.R. Daliri, Granger causality analysis in combination with directed network measures for classification of MS patients and healthy controls using task-related fMRI, *Comput. Biol. Med.* (2019), doi:10.1016/j.combiomed.2019.103495.
- [10] Z. Abbasvandi, A.M. Nasrabadi, A self-organized recurrent neural network for estimating the effective connectivity and its application to EEG data, *Comput. Biol. Med.* (2019), doi:10.1016/j.combiomed.2019.05.012.

- [11] M.G. Tana, R. Sclocco, A.M. Bianchi, GMAC: a Matlab toolbox for spectral Granger causality analysis of fMRI data, *Comput. Biol. Med.* (2012), doi:10.1016/j.combiomed.2012.07.003.
- [12] R.M. Demir, M.S. Özderem, C. Bayrak, E. Mendi, Determination of ECoG information flow activity based on Granger causality and Hilbert transformation, *Comput. Methods Programs Biomed.* (2013), doi:10.1016/j.cmpb.2013.08.011.
- [13] Y. Gao, X. Wang, T. Potter, J. Zhang, Y. Zhang, Single-trial EEG emotion recognition using Granger causality/transfer entropy analysis, *J. Neurosci. Methods* (2020), doi:10.1016/j.jneumeth.2020.108904.
- [14] A. Wismüller, A.M. Dsouza, M.A. Vosoughi, A. Abidin, Large-scale nonlinear Granger causality for inferring directed dependence from short multivariate time-series data, *Sci. Rep.* (2021), doi:10.1038/s41598-021-87316-6.
- [15] L. Faes, G. Nollo, A. Porta, Non-uniform multivariate embedding to assess the information transfer in cardiovascular and cardiorespiratory variability series, *Comput. Biol. Med.* (2012), doi:10.1016/j.combiomed.2011.02.007.
- [16] M. Młyńczak, H. Krystofiak, Discovery of causal paths in cardiorespiratory parameters: a time-independent approach in elite athletes, *Front. Physiol.* (2018), doi:10.3389/fphys.2018.01455.
- [17] M. Młyńczak, H. Krystofiak, Cardiorespiratory temporal causal links and the differences by sport or lack thereof, *Front. Physiol.* (2019), doi:10.3389/fphys.2019.00045.
- [18] A.D. Orjuela-Cañón, A. Cerquera, J.A. Freund, G. Juliá-Seirdá, A.G. Ravelo-García, Sleep apnea: tracking effects of a first session of CPAP therapy by means of Granger causality, *Comput. Methods Programs Biomed.* (2020), doi:10.1016/j.cmpb.2019.105235.
- [19] A.D. Jaimes-Albarracín, A.D. Orjuela-Canon, A.L. Jutínico, M.A. Bazurto, E. Dueñas, Brain and heart physiological networks analysis employing neural networks granger causality, in: *Proceedings of the 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2021, doi:10.1109/NER49283.2021.9441375.
- [20] C. Corbier, F. Chouchou, F. Roche, J.C. Barthélémy, V. Pichot, Causal analyses to study autonomic regulation during acute head-out water immersion, head-down tilt and supine position, *Exp. Physiol.* (2020), doi:10.1113/EP088640.
- [21] A. Seth, Granger causality, *Scholarpedia* 2 (2007) 1667, doi:10.4249/scholarpedia.1667.
- [22] C.A. Sims, Money, income, and causality, *Am. Econ. Rev.* 62 (1972) 540–552, doi:10.1126/science.151.3712.867-a.
- [23] SAS, Granger causality test SAS, (2021), <https://support.sas.com/rnd/app/ets/examples/granger/index.htm> (accessed December 11, 2021).
- [24] S.L. Bressler, A.K. Seth, Wiener-Granger causality: a well established methodology, *Neuroimage* 58 (2011) 323–329, doi:10.1016/j.neuroimage.2010.02.059.
- [25] J. Geweke, Measurement of linear dependence and feedback between multiple time series, *J. Am. Stat. Assoc.* 77 (1982) 304–313, doi:10.1080/01621459.1982.10477803.
- [26] A.E. Obayelu, A.S. Salau, Agricultural response to prices and exchange rate in Nigeria: application of co-integration and vector error correction model (VECM), *J. Agric. Sci.* (2010), doi:10.1080/09766898.2010.11884656.
- [27] F.F.A.H. Asani, N.S. Baharuddin, N. Jusoh, Z. Mohamad, N. Shamsudin, K. Jusoff, A vector error correction model (VECM) approach in explaining the relationship between interest rate and inflation towards exchange rate volatility in Malaysia, *World Appl. Sci. J.* 12 (2011) 49–56.
- [28] N. Ancona, D. Marinazzo, S. Stramaglia, Radial basis function approach to nonlinear Granger causality of time series, *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* (2004) 70, doi:10.1103/PhysRevE.70.056221.
- [29] D. Marinazzo, M. Pellicoro, S. Stramaglia, Kernel method for nonlinear Granger causality, *Phys. Rev. Lett.* (2008), doi:10.1103/PhysRevLett.100.144103.
- [30] D. Marinazzo, M. Pellicoro, S. Stramaglia, Kernel-Granger causality and the analysis of dynamical networks, *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* (2008), doi:10.1103/PhysRevE.77.056215.
- [31] N. Nicolaou, T.G. Constantinou, A nonlinear causality estimator based on non-parametric multiplicative regression, *Front. Neuroinform.* (2016), doi:10.3389/fninf.2016.00019.
- [32] A. Montalto, S. Stramaglia, L. Faes, G. Tessitore, R. Prevete, D. Marinazzo, Neural networks with non-uniform embedding and explicit validation phase to assess Granger causality, *Neural Netw.* 71 (2015) 159–171, doi:10.1016/j.neunet.2015.08.003.
- [33] A. Attanasio, U. Triacca, Detecting human influence on climate using neural networks based Granger causality, *Theor. Appl. Climatol.* 103 (2011) 103–107, doi:10.1007/s00704-010-0285-8.
- [34] T. Li, G. Li, T. Xue, J. Zhang, Analyzing brain connectivity in the mutual regulation of emotion-movement using bidirectional Granger causality, *Front. Neurosci.* 14 (2020) 369, doi:10.3389/fnins.2020.00369.
- [35] Y. Huang, Z. Fu, C.L.E. Franzke, Detecting causality from time series in a machine learning framework, *Chaos* 30 (2020) 063116, doi:10.1063/5.0007670.
- [36] N. Talebi, A.M. Nasrabadi, I. Mohammad-Rezazadeh, Estimation of effective connectivity using multi-layer perceptron artificial neural network, *Cogn. Neurodyn.* (2018), doi:10.1007/s11571-017-9453-1.
- [37] N. Talebi, A.M. Nasrabadi, I. Mohammad-Rezazadeh, R. Cohen, NCREANN: non-linear causal relationship estimation by artificial neural network; applied for autism connectivity study, *IEEE Trans. Med. Imaging* (2019), doi:10.1109/TMI.2019.2916233.
- [38] Large-Scale Nonlinear Granger Causality, 2021, <https://github.com/Large-scale-causality-inference/Large-scale-nonlinear-causality> (accessed December 11, 2021).
- [39] R. Marcinkevičs, D. Miladinović, Granger-causal inference in time series for identifying molecular fingerprints during sleep, <https://github.com/i6092467/NNCC-SLIMMBA> (accessed December 11, 2021).
- [40] A. Tank, I. Covert, N. Foti, A. Shojai, E. Fox, Neural Granger Causality, *IEEE Trans Pattern Anal Mach Intell* (2021), doi:10.1109/TPAMI.2021.3065601.
- [41] S. Kaushik, A. Choudhury, P.K. Sheron, N. Dasgupta, S. Natarajan, L.A. Pickett, V. Dutt, AI in healthcare: time-series forecasting using statistical, neural, and ensemble architectures, *Front. Big Data* (2020), doi:10.3389/fdata.2020.00004.
- [42] K.M. Dalmeida, G.L. Masala, HRV features as viable physiological markers for stress detection using wearable devices, *Sensors* (2021), doi:10.3390/s21082873.
- [43] Y. Yu, X. Si, C. Hu, J. Zhang, A review of recurrent neural networks: LSTM cells and network architectures, *Neural Comput.* (2019), doi:10.1162/neco.2019.01199.
- [44] J. Cao, Z. Li, J. Li, Financial time series forecasting model based on CEEMDAN and LSTM, *Phys. A Stat. Mech. Appl.* (2019), doi:10.1016/j.physa.2018.11.061.
- [45] T. Umetsu, A. Sano, R.W. Picard, Daytime data and LSTM can forecast tomorrow's stress, health, and happiness, in: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, doi:10.1109/EMBC.2019.8856862.
- [46] S. Xiang, Y. Qin, C. Zhu, Y. Wang, H. Chen, Long short-term memory neural network with weight amplification and its application into gear remaining useful life prediction, *Eng. Appl. Artif. Intell.* 91 (2020) 103587, doi:10.1016/j.engappai.2020.103587.
- [47] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.
- [48] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, in: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, doi:10.3115/v1/d14-1179.
- [49] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Gated feedback recurrent neural networks, in: *Proceedings of the 32nd International Conference on ICML*, 2015.
- [50] J.C. Heck, F.M. Salem, Simplified minimal gated unit variations for recurrent neural networks, in: *Proceedings of the IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2017, doi:10.1109/MWSCAS.2017.8053242.
- [51] R. Lund, Time series analysis and its applications: with R examples, 2007, 10.1198/jasa.2007.s209.
- [52] F. Wilcoxon, Individual comparisons by ranking methods, *Biom. Bull.* 1 (1945) 80–83, doi:10.2307/3001968.
- [53] The Wilcoxon Signed-Rank Test - Python implementation, (2021), <https://github.com/scipy/scipy/blob/v1.6.0/scipy/stats/morestats.py#L2809-L3059> (accessed December 11, 2021).
- [54] R.H. Riffenburgh, *Statistics in Medicine*, Elsevier Inc., 2006, doi:10.1016/B978-0-12-088770-5.X5036-9.
- [55] M. Rosol, Nonlincausality, Python package - github, (2021), <https://github.com/mrosol/Nonlincausality> (accessed December 11, 2021).
- [56] M. Rosol, Nonlincausality, Python package - PyPI, (2021), <https://pypi.org/project/nonlincausality/> (accessed December 11, 2021).
- [57] F. Chollet & others, Keras library, (2015), <https://keras.io/> (accessed December 11, 2021).
- [58] P.C. Ivanov, R.P. Bartsch, Network physiology: mapping interactions between networks of physiological networks, *Underst. Complex Syst.* (2014), doi:10.1007/978-3-319-03518-5\_10.



# Causality in cardiorespiratory signals in pediatric cardiac patients

Maciej Rosol<sup>1</sup>, Jakub S. Gąsior<sup>2</sup>, Iwona Walecka<sup>2</sup>, Bożena Werner<sup>2</sup>, Gerard Cybulski<sup>1</sup>, Marcel Młynczak<sup>1</sup>

**Abstract**— Four different Granger causality-based methods - one linear and three nonlinear (Granger Causality, Kernel Granger Causality, large-scale Nonlinear Granger Causality, and Neural Network Granger Causality) were used for assessment and causal-based quantification of the respiratory sinus arrhythmia (RSA) in the group of pediatric cardiac patients, based on the single-lead ECG and impedance pneumography signals (the latter as the tidal volume curve equivalent). Each method was able to detect the dependency (in terms of causal inference) between respiratory and cardiac signals. The correlations between quantified RSA and the demographic parameters were also studied, but the results differ for each method.

**Clinical relevance**— The presented methods (among which NNGC seems to be the most valid) allow for quantification of RSA and study of dependency between tidal volume and RR intervals, which may help to better understand association between respiratory and cardiovascular systems in different populations.

## I. INTRODUCTION

Although the circulatory and respiratory systems serve two distinct functions, their activity is deeply interconnected and interrelated. Respiratory sinus arrhythmia (RSA), a phenomenon whereby RR intervals shorten during inhalations and lengthen during exhalations, is one example of such interactions [1]. This physiological phenomenon is rooted in many mechanisms, among which is the activity of the sympathetic and parasympathetic nervous systems [2,3]. Heart rate variability (HRV) parameters reflect the fluctuation of RR intervals and have wide applicability in medicine and sports as a means of assessing autonomous nervous system activity [4,5]. HRV parameters can be calculated from ECG or from photoplethysmography signal [6,7].

RSA can be used as a biomarker reflecting conditions like diabetes, sleep apnea, cardiomyopathy, anxiety disorders or heart failure [8-10]. Approaches based on the Granger causality concept are among the viable methods of RSA assessment and quantification [11]. In a particular Granger causality analysis, two models are used. The first one predicts the values of X signal (i.e., the tachogram/RR intervals signal) based on the past values of this signal, while the second model tries to fit the same value but based on the past values of X and Y signals (i.e., both tachogram and respiratory signal). Y is said to cause X (denoted as  $Y \rightarrow X$ ) if the prediction error of the second model is statistically significantly smaller than the prediction error of the first model. The measure of RSA would

depend on the improvement in the prediction accuracy after incorporating the Y signal (e.g., respiratory signal) for the prediction of X signal (in this case the tachogram). Causality can be assessed and quantified from a respiratory signal to a tachogram, but also in the opposite direction.

The aim of this paper is to present and compare the results from different causality-based methods (one linear and three nonlinear) used in analysis of short-term RSA phenomenon among a group of pediatric cardiac patients.

## II. METHODS

### A. Conducted study

This study analyzed data recorded among 20 patients (a demographic breakdown is presented in Table 1). In order to be included in the study, patients had to be between 7 and 15 years old and had to have a current diagnosis of cardiac disease (as well as a signed consent form). Patients were excluded if they:

- presented symptoms of infection,
- had been diagnosed with disorders potentially affecting the autonomous nervous system, or
- were unable to perform physical activity.

TABLE I. THE DEMOGRAPHIC DETAILS OF THE STUDY GROUP (AS MEAN  $\pm$  STANDARD DEVIATION). THE NUMBER OF PARTICIPANTS PER SEX IS GIVEN IN THE BRACKETS.

| Demographic data         |                  |                  |                  |
|--------------------------|------------------|------------------|------------------|
|                          | All              | Boys (12)        | Girls (8)        |
| Age [years]              | 12.9 $\pm$ 3.5   | 14.3 $\pm$ 3.3   | 10.8 $\pm$ 2.6   |
| Weight [kg]              | 58.8 $\pm$ 23.2  | 65.2 $\pm$ 22.9  | 46.0 $\pm$ 18.1  |
| Hight [cm]               | 160.1 $\pm$ 17.2 | 166.1 $\pm$ 16.3 | 151.0 $\pm$ 16.3 |
| BMI [kg/m <sup>2</sup> ] | 21.5 $\pm$ 5.0   | 22.7 $\pm$ 5.1   | 19.3 $\pm$ 4.1   |

The study was approved by the Ethics Committee of the Medical University of Warsaw (permission: KB/70/2021).

In this study the cardiac activity was registered through single-lead ECG, while the respiratory signal (tidal volume equivalent) was obtained from impedance pneumography (IP). The latter measurement was obtained using a Pneumonitor 2 [12]. Tetrapolar configuration was used for IP recordings with sinusoidal application current, which amplitude was up to 1 mA, and frequency was equal to 100 kHz. The electrodes were placed according to [13], with the

<sup>1</sup>M. Rosol, M. Młynczak and G. Cybulski are with the Warsaw University of Technology, Faculty of Mechatronics, Institute of Metrology and Biomedical Engineering, 8 Boboli Street, 02-525 Warsaw, Poland. maciej.rosol.dokt@pw.edu.pl

<sup>2</sup>J.S. Gąsior, I. Walecka and B. Werner are with the Medical University of Warsaw, Department of Pediatric Cardiology and General Pediatrics, Żwirki i Wigury 63a, 02-091 Warsaw, Poland

receiving electrodes placed on a midaxillary line around 5<sup>th</sup> and 6<sup>th</sup> rib and the application electrodes attached to the proximal side of the arm at the level of receiving electrodes. Both signals had sampling frequency of 250Hz. Standard Holter-type Ag/AgCl electrodes were used.

All measurements were conducted in a diagnostic room in the Department of Pediatric Cardiology and General Pediatrics of the Medical University of Warsaw. The study protocol consisted of:

- obtaining a signed examination consent form from the patient's parent or legal guardian,
- presenting the parent/legal guardian with the conditions for the patient's participation in the study,
- presenting the testing apparatus and carrying out the process of getting the patient acclimated to the test conditions,
- commencing the recording at rest until rest values were reached,
- acquiring signals for analysis over the span of 5 minutes.

#### B. Methods used

The RR signal was obtained by detecting the R peaks in ECG using Stationary Wavelet Transform [14,15] and then performing cubic interpolation. The IP signal was filtrated using high-pass and low-pass Butterworth filters with cutoff frequencies equal to 0.1 Hz and 30 Hz, respectively. Based on [16], it was assumed that by using the aforementioned electrode configuration, the linear fitting provides the best compatibility between IP and pneumotachometry so that the IP signal was treated as relative tidal volume equivalent (TV). Both RR and TV signals were down-sampled to 25 Hz to obtain the compromise between computational complexity reduction and time resolution preservation (for analytical methods utilized).

For the assessment and quantification of the RSA the following methods were used:

- Traditional Granger causality analysis (GC) [17],
- Kernel Granger Causality (KGC) [18],
- Large scale Nonlinear Granger Causality (lsNGC) [19], and
- *nonlincausality* Python package using neural networks for prediction in term of Granger causality (NNGC) [20].

Moreover, the change of the causality relationship between the signals over time was studied using *nonlincausality* approach with windows  $w_1$  and  $w_2$ , equal to 25 (1 second) and 1 (40 milliseconds) samples, respectively. The temporal orders between signals were also studied with the *tempord* package [21] using linear-based LM method with 0.9 threshold and time shift between -2 and 2 seconds.

All of the above-mentioned methods were used in quantifying RSA for each patient; subsequently, because all the methods used allow for statistical inference, the statistical significance of these findings was assessed. Analysis of causality was performed for lag equal to 1 second for each method. In the case of the KGC method, the chosen kernel was gaussian and the order parameter was 25. For lsNGC, the  $c_f$  and  $c_g$  parameters, which corresponds to the number of

hidden neurons in the Generalized Radial Basis Functions neural network were set to 25. The architecture of neural networks used in NNGC consisted of 2 hidden layers with 20 neurons with the ReLU activation function in hidden layers and linear activation function in the output layer. Networks were trained for 100 epochs with learning rate equal to 1e-4 and 1e-5 for the first and second half of the learning process, respectively.

Dependencies of causality values in both directions were also studied in light of demographic data. Correlation between causality metrics and the patient's age and body mass index (BMI), as well as between measures of causality in different directions was calculated using Spearman's correlation coefficient. Mann-Whitney U rank test was used to check whether the patients' sex accounted for any differences in causality measures.

The results obtained from temporal orders analysis were also compared to the results from the group of 10 elite athletes and 10 healthy students presented in [21].

The significance level was assumed to be 0.05. All calculations were run using Python 3.7 (calculation of GC, lsNGC and NNGC), R version 3.6.1 (temporal orders) and MATLAB 2019b (KGC).

### III. RESULTS

#### A. Causality analysis

An example of the obtained tachogram and respiratory signal is presented in Fig. 1.

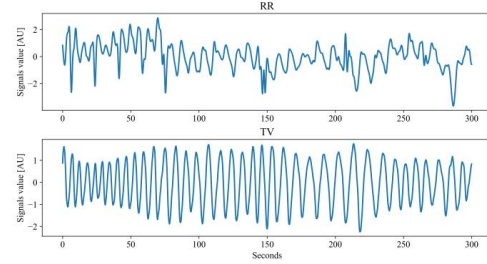


Figure 1. Sample plots of tachogram and respiratory signal from the patient #8.

For each patient the causality between both signals was quantified using each method and in both directions, and those values were plotted in Fig. 2. For twelve patients the causality from TV to RR was higher than from RR to TV, and for eight RR→TV causality was higher in case of GC and NNGC. For lsNGC method, the proportions were the opposite as eight children obtained higher causality values for the direction from TV to RR and twelve of them - from RR to TV.

Causality between signals was statistically significant in both directions and regardless of the method used, an exception being the TV→RR causality according to lsNGC in patient #4.

The results of the correlation analysis are presented in Table 2. Irrespective of the method used, there was no statistically significant correlation between age and either direction of causality. There was a significant negative correlation

( $p$ -value  $< 0.05$ ) between BMI and the TV $\rightarrow$ RR causality value calculated using NNGC (Spearman's rho equal to -0.46), and a positive one between BMI and RR $\rightarrow$ TV calculated using lsNGC (Spearman's rho equal to 0.52). For lnNGC and GC approaches, there were also significant correlations ( $p$ -values  $< 0.01$  and  $< 0.05$ , respectively) between TV $\rightarrow$ RR and RR $\rightarrow$ TV with Spearman's rho equal to 0.59 and 0.45, respectively. There was a statistically significant difference in causality for boys and girls for lsNGC and GC ( $p$ -values  $< 0.05$ ).

#### B. Temporal causality

Changes of causality over time calculated with *nonlincausality* package (NNGC approach) were plotted for each patient in both directions (TV $\rightarrow$ RR, and RR $\rightarrow$ TV). For each patient, the signal of change in causality over time was highly variable for both directions of causality, regardless of the patient's age, sex, or BMI. A sample plot of this change can be seen in Fig. 3. The temporal orders were also visualized for each patient and the example plot is presented in Fig. 4.

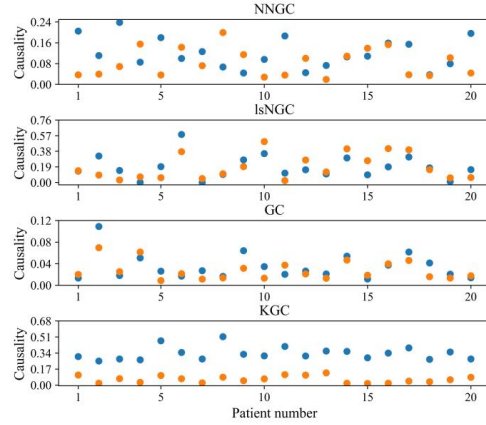


Figure 2. Causality values assessed with different methods for each individual patient. The blue dots are representing the quantified causality values from TV to RR, while orange ones stand for causality from RR to TV.

TABLE II. SPEARMAN'S CORRELATION COEFFICIENTS WITH  $p$ -VALUES IN BRACKETS BETWEEN CAUSALITY AND AGE OR BMI AND BETWEEN TWO DIRECTIONS OF CAUSALITY FOR DIFFERENT METHODS.

|       | Age                 |                     | BMI                  |                     | Causalities         |
|-------|---------------------|---------------------|----------------------|---------------------|---------------------|
|       | TV $\rightarrow$ RR | RR $\rightarrow$ TV | TV $\rightarrow$ RR  | RR $\rightarrow$ TV |                     |
| GC    | 0.30<br>(0.20)      | 0.16<br>(0.49)      | 0.25<br>(0.30)       | 0.12<br>(0.63)      | 0.45<br>( $<0.05$ ) |
| KGC   | -0.38<br>(0.10)     | -0.30<br>(0.21)     | -0.31<br>(0.20)      | -0.33<br>(0.17)     | 0.40<br>(0.08)      |
| lsNGC | 0.17<br>(0.47)      | 0.13<br>(0.60)      | 0.15<br>(0.53)       | 0.52<br>( $<0.05$ ) | 0.59<br>( $<0.01$ ) |
| NNGC  | -0.35<br>(0.14)     | 0.20<br>(0.39)      | -0.46<br>( $<0.05$ ) | 0.04<br>(0.88)      | -0.19 (0.43)        |

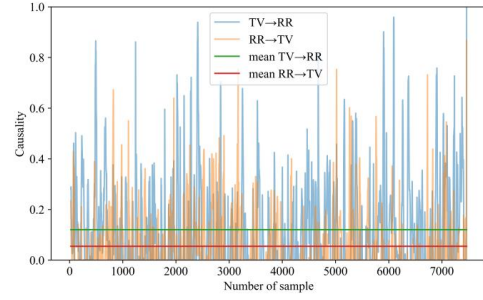


Figure 3. Causality changes over time for RR $\rightarrow$ TV and TV $\rightarrow$ RR for patient #10.

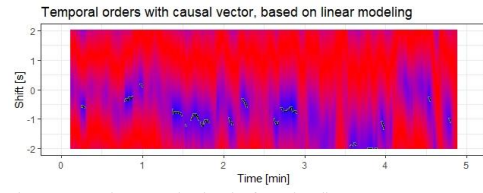


Figure 4. Sample temporal order plot for patient #8.

In line with expectations, the causal vectors obtained among cardiological patients were much smaller and less continuous than those collected in studies with subject populations of elite athletes and healthy students (presented in [21]). The average causal vector was  $-72 \pm 668$  milliseconds (mean  $\pm$  standard deviation). For six of the twenty patients there was no causal vector at all.

#### IV. DISCUSSION

All of the used methods were able to detect a dependency between RR and TV in at least one direction. The KGC method stands out as the only one which in every case indicated a higher causal relationship from TV to RR. Meanwhile, the GC and NNGC methods detected TV $\rightarrow$ RR as higher more frequently, while lsNGC found more cases with higher causality from RR to TV. The negative correlation between BMI and causality derived from NNGC data seems to be in line with the findings of Mazurak *et al.* [22], which demonstrated that decreasing a patient's body mass would result in increased HRV measures. Given that the directions of the causalities for GC and lsNGC are intercorrelated it is questionable whether using both measures would be useful in case of those methods. When it comes to NNGC and KGC, the usage of both directions of causality measures may lead to new and additional information about patients.

Based on a plot of causality figures for each patient and method, it is apparent that each method interprets the two directions of causality differently. Because each one is capable of detecting causality between the analyzed signals, they could all potentially be useful, as each might provide different diagnostic information. That said, the NNGC



approach appears to be the most valid as it indicates a correlation of RSA with BMI while the two directions of causality measurements are not intercorrelated.

It is also to be emphasized that the detected causality from RR to TV does not indicate that the work of the cardiovascular system is driving the work of the respiratory system, but it provides additional information regarding the interdependencies between signals from those systems. Moreover, due to the complexity of the relationships between those systems, the usage of nonlinear modeling methods seems to be a more appropriate choice.

Furthermore, it seems that the analysis of temporal orders may be a valuable tool capable of distinguishing between different group of patients. Although further studies on the matter are undoubtedly needed, analysis of causality's change over time might prove to be a source of additional information, as it can help visualize relationships between various signals and causality measures.

The foremost limitation of this study lies in its relatively small study group as well as in the homogeneity of this group. Because the study sample lacked a control group of healthy subjects, the utility of the calculated causality values as a biomarker for cardiac autonomic functions abnormalities could not be determined. Therefore, further studies (ones involving larger and more varied sample populations) will have to be conducted in order to ascertain the possible utility of each method and each direction of causality in terms of quantification of RSA. The main limitation of the mentioned techniques is their dependance of the choice of input, which might have an impact on the final result.

## V.CONCLUSION

RSA was detected and quantified for each patient using four different causality-based methods. There appears to be no correlation between the results and the patient's age, but some correlation with BMI (in the case of NNGC and IsNGC) and with sex (in the case of GC and IsNGC) may be demonstrable. Further studies on a larger group including healthy subjects are needed to assess the utility of each method in terms of RSA.

## REFERENCES

- [1] G. G. BERTSON, J. T. CACIOPPO, and K. S. QUIGLEY, "Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications," *Psychophysiology*, 1993, doi: 10.1111/j.1469-8986.1993.tb01731.x.
- [2] P. G. Katona and F. Jih, "Respiratory sinus arrhythmia: noninvasive measure of parasympathetic cardiac control," *J. Appl. Physiol.*, 1975, doi: 10.1152/jappl.1975.39.5.801.
- [3] F. M. Fouad, R. C. Tarazi, and C. M. Ferrario, "Assessment of parasympathetic control of heart rate by a noninvasive method," *Am. J. Physiol. - Heart, Circ. Physiol.*, 1984, doi: 10.1152/ajpheart.1984.246.6.h838.
- [4] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: A review," *Medical and Biological Engineering and Computing*, 2006, doi: 10.1007/s11517-006-0119-0.
- [5] V. P. Da Silva, N. A. De Oliveira, H. Silveira, R. G. T. Mello, and A. C. Deslandes, "Heart rate variability indexes as a marker of chronic adaptation in athletes: A systematic review," *Annals of Noninvasive Electrocardiology*, 2015, doi: 10.1111/anec.12237.
- [6] M. Bolanos, H. Nazeran, and E. Haltiwanger, "Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals," in *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 2006, doi: 10.1109/IEMBS.2006.260607.
- [7] B. Vescio, M. Salsone, A. Gambardella, and A. Quattrone, "Comparison between electrocardiographic and earlobe pulse photoplethysmographic detection for evaluating heart rate variability in healthy subjects in short- and long-term recordings," *Sensors (Switzerland)*, 2018, doi: 10.3390/s18030844.
- [8] B. F. Giraldo, M. F. Pericas, R. Schroeder, and A. Voss, "Respiratory Sinus Arrhythmia Quantified with Linear and Non-Linear Techniques to Classify Dilated and Ischemic Cardiomyopathy," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018, doi: 10.1109/EMBC.2018.8513199.
- [9] J. Morales, P. Borzée, D. Testelmans, B. Buyse, S. Van Huffel, and C. Varon, "Linear and Non-linear Quantification of the Respiratory Sinus Arrhythmia Using Support Vector Machines," *Front. Physiol.*, 2021, doi: 10.3389/fphys.2021.623781.
- [10] M. Peltola *et al.*, "Respiratory sinus arrhythmia as a predictor of sudden cardiac death after myocardial infarction," *Ann. Med.*, 2008, doi: 10.1080/07853890701884659.
- [11] J. Cui, Z. Huang, J. Wu, and H. Jiang, "Cardiopulmonary Resonance Function and Indices—A Quantitative Measurement for Respiratory Sinus Arrhythmia," *Front. Physiol.*, 2020, doi: 10.3389/fphys.2020.00867.
- [12] M. Mlynčzak, M. Zyliński, W. Niewiadomski, and G. Cybulski, "Ambulatory Devices Measuring Cardiorespiratory Activity with Motion," in *BIODEVICES 2017 - 10th International Conference on Biomedical Electronics and Devices, Proceedings, Part of 10th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2017*, 2017, doi: 10.5220/0006111700910097.
- [13] V. P. Seppä, J. Hyttinen, M. Uitto, W. Chrapek, and J. Viik, "Novel electrode configuration for highly linear impedance pneumography," *Biomed. Tech.*, 2013, doi: 10.1515/bmt-2012-0068.
- [14] V. Kalidas and L. Tamil, "Real-time QRS detector using stationary wavelet transform for automated ECG analysis," in *Proceedings - 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering, BIBE 2017*, 2017, doi: 10.1109/BIBE.2017.00-12.
- [15] L. Howell and B. Porr, "py-ecg-detectors package." [Online]. Available: <https://pympi.org/project/py-ecg-detectors/>. [Accessed: 20-Jan-2022].
- [16] M. Mlynčzak, W. Niewiadomski, M. Zyliński, and G. Cybulski, "Assessment of calibration methods on impedance pneumography accuracy," *Biomed. Tech.*, 2015, doi: 10.1515/bmt-2015-0125.
- [17] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, vol. 37, no. 3, p. 424, 1969, doi: 10.2307/1912791.
- [18] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel method for nonlinear Granger causality," *Phys. Rev. Lett.*, 2008, doi: 10.1103/PhysRevLett.100.144103.
- [19] A. Wismüller, A. M. Dsouza, M. A. Vosoughi, and A. Abidin, "Large-scale nonlinear Granger causality for inferring directed dependence from short multivariate time-series data," *Sci. Rep.*, 2021, doi: 10.1038/s41598-021-87316-6.
- [20] M. Rosol, "Nonlinearity - PyPI," 2021. [Online]. Available: <https://pympi.org/project/nonlinearity/>. [Accessed: 20-Jan-2022].
- [21] M. Mlynčzak, "Temporal orders and causal vector for physiological data analysis," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2020, doi: 10.1109/EMBC44109.2020.9176842.
- [22] N. Mazurak *et al.*, "Effect of a weight reduction program on baseline and stress-induced heart rate variability in children with obesity," *Obesity*, 2016, doi: 10.1002/oby.21355.

# Prediction of the duration of maximal exercise test in professional adolescent football players based on the cardiorespiratory signals – a pilot study

Maciej Rosol  
Institute of Metrology and Biomedical  
Engineering  
Warsaw University of Technology  
Warsaw, Poland  
maciej.rosol.dokt@pw.edu.pl

Jakub S. Gąsior  
Department of Pediatric Cardiology  
and General Pediatrics  
Medical University of Warsaw  
Warsaw, Poland  
jakub.gasior@wum.edu.pl

Kacper Korzeniewski  
Institute of Metrology and Biomedical  
Engineering  
Warsaw University of Technology  
Warsaw, Poland  
kacper.korzeniewski.dokt@pw.edu.pl

Jonasz Łaba  
Institute of Metrology and Biomedical  
Engineering  
Warsaw University of Technology  
Warsaw, Poland  
jonasz.laba.dokt@pw.edu.pl

Robert Makuch  
Department of Physical Education  
Pulaski University of Technology and  
Humanities  
Radom, Poland  
r.makuch@uthrad.pl

Marcel Młynczak  
Institute of Metrology and Biomedical  
Engineering  
Warsaw University of Technology  
Warsaw, Poland  
marcel.mlynczak@pw.edu.pl

**Abstract**—This pilot study aimed to evaluate the potential of machine learning models utilizing parameters estimated from cardiorespiratory signals obtained during rest in predicting the duration of maximal cardiopulmonary exercise tests (CPET) in professional adolescent football players. The study involved a group of 36 male athletes, whose cardiac and respiratory signals were recorded in a supine position for at least 5 minutes. Heart rate variability, statistical features from respiratory signal and features from causal and information domains quantifying the interdependency between cardiac and respiratory signals were calculated and later used for the analysis. The most relevant features (Pearson correlation with CPET duration over 0.2) were used for the machine learning modeling with leave-one-out validation applied. Models demonstrated promising results with a mean absolute percentage error of 17%, mean absolute error of 129 seconds, root mean squared error of 170 seconds,  $R^2$  score of 0.52, and a Pearson correlation coefficient of 0.74. Explainable artificial intelligence techniques provided insights into the influence of individual features, showing the primary importance of the cardiac parameters but also highlighting the need for incorporating the information from respiratory signal and cardiorespiratory interdependencies.

**Keywords**—Cardiopulmonary Exercise Testing, Machine Learning, Cardiorespiratory Signals, Causality, Performance Prediction

## I. INTRODUCTION

The maximal cardiopulmonary exercise test (CPET) stands out as a primary tool for assessing the endurance and cardiorespiratory fitness (CRF) of athletes [1]. During such tests, participants undertake physical activity like running on a treadmill or cycling on a cycloergometer with an increasing workload until exhaustion. Throughout the test cardiac and respiratory systems are usually monitored via electrocardiography (ECG) and equipment for gas analysis, respectively. Based on the data gathered during the test, different parameters can be established. The most popular

ones are: maximal or peak oxygen uptake ( $VO_{2max}/VO_{2peak}$ ), maximal heart rate (HR), ventilatory threshold, and duration of the test [2]. However, performing full CPET is physically demanding, which might interfere with the training program and require sophisticated equipment and trained personnel to conduct it. Thus, ongoing research is focused on providing new evaluation approaches based on non-exercise or submaximal tests [3], the utilization of simpler measurement devices [4], the introduction of new cardiorespiratory parameters [5], and the use of machine learning algorithms [6].

In terms of cardiorespiratory parameters, the analysis of Heart Rate Variability (HRV) is utilized as a key set of parameters for understanding and improving athletic performance by monitoring individual adaptation to training [7]. HRV, the physiological phenomenon of variation in the time interval between heartbeats, is recognized as a crucial indicator of autonomic nervous system (ANS) function and overall cardiovascular health. It offers a deeper understanding of an athlete's stress levels, recovery status, and readiness for training or competition [8]. However, while HRV parameters offer significant insights, there are multiple studies suggesting that including respiratory parameters like pulmonary ventilation (VE) in the evaluation process is equally vital as it is also a valid predictor of, e.g., oxygen uptake [9] and  $VO_{2peak}$  [10]. Moreover, recent advances in physiology leverage the interdependence of cardiological and respiratory systems in terms of cardiorespiratory coupling (CRC) by introducing new types of parameters based on both causal and information analysis [5], [11], [12]. Those approaches are based on: 1) improving the prediction of the current value of one signal (e.g., tachogram) by leveraging knowledge about the past values of another signal (e.g., respiration) alongside the values of the first one; 2) analyzing information flow from one signal to another using entropy analysis; 3) employing other techniques for measuring dependence, such as

---

The research was founded by POB Biotechnology and biomedical engineering of Warsaw University of Technology within the Excellence Initiative: Research University (IDUB) programme.

correlation. However, further studies are needed to better understand the utility of parameters from those two domains.

The aim of this paper was to investigate the quality of prediction of athletic performance in terms of the duration of CPET by utilizing machine learning models trained on cardiorespiratory features calculated from signals acquired during rest in a supine position among a group of professional adolescent athletes. Additionally, we assessed the importance of features from causal and information domains included in the models for the duration of CPET prediction.

## II. METHODS

### A. Study design

A total of 36 professional adolescent male football players aged between 10 and 15 years were enrolled in the study. The descriptive statistics of the group are presented in Tab. 1 (presented  $VO_{2max}$  values, determined based on the Yo-yo test, were not used in the analysis). The study consisted of two stages: 1) cardiorespiratory signals were recorded for at least 5 minutes during rest in a supine position; 2) CPET was performed on a cycloergometer starting at 40 Watts and increasing the workload by 40 Watts every 3 minutes. Athletes were instructed to perform the test until exhaustion. The study was approved by the Ethics Committee of the Medical University of Warsaw (permission: KB/70/2021) and a written informed consent form was collected for each participant.

TABLE I. DESCRIPTIVE STATISTICS OF THE STUDY POPULATION PRESENTED AS MEAN  $\pm$  STANDARD DEVIATION.

| Age<br>[years] | Weight<br>[kg]  | Height<br>[cm]   | BMI<br>[kg/m <sup>2</sup> ] | $VO_{2max}$<br>[ml/min/kg] |
|----------------|-----------------|------------------|-----------------------------|----------------------------|
| 13.3 $\pm$ 1.5 | 56.7 $\pm$ 14.0 | 169.1 $\pm$ 13.0 | 19.4 $\pm$ 2.4              | 48.7 $\pm$ 5.5             |

For the cardiorespiratory signals acquisition, the Pneumonitor 3 device was used [13], [14]. This device combines one-lead ECG and impedance pneumography (IP) into a single remote unit designed for heart activity and indirect respiratory function monitoring. Both signals were registered with 250 Hz sampling frequency and saved on the microSD card. For IP, the tetrapolar configuration was used with electrodes placed according to [15], which allowed for the interpretation of the IP signal as an equivalent of relative tidal volume (TV).

### B. Analysis

From the ECG recordings (acquired with 250 Hz sampling frequency), the RR interval time series were obtained using automated detection and manual correction by an experienced clinician. The tachogram time series was obtained by cubic interpolation of the RR interval (RRi) values in milliseconds placed between consecutive R peaks on a time axis. TV signal was filtered using a bandpass filter with cutoff frequencies equal to 0.05 and 0.67 Hz corresponding to 3 and 40 breaths per minute respectively. Thus, two signals with equal length and sampling frequency with a common time scale were obtained. Then, both time series were downsampled to a 25

Hz sampling frequency in order to reduce the computational complexity (only for the calculation of information and causal parameters). The example of the obtained time series is presented in Fig 1. Based on the RRi series time and frequency domain as well as nonlinear HRV parameters were calculated [16]. From TV signal, e.g., statistical features of respiration rate (RespRate), relative TV (indexed by median TV as calibration was not performed and measured impedance signal could not be translated to ml), or inspiration/expiration ratio were calculated. Moreover, the parameters from causal and information domains were calculated based on both tachogram and TV to quantify CRC, which is increasingly recognized as an important, non-invasive indicator of the interplay between cardiac and respiratory functions in sports medicine, reflecting training effects, competition readiness, and adaptations to various physical stimuli [17]. For this purpose methods like Granger causality (GC), nonlinear GC based on machine learning models like support vector machines, random forests, and neural networks [18], kernel GC [19], large-scale nonlinear GC [20], symbolic transfer entropy (STE), conditional entropy, mutual information and others were utilized. Demographic data regarding age, weight, height, and BMI were also utilized. The full list of features alongside their description is presented in the repository, which also contains the code used for the analysis [21]. The duration of the CPET was determined based on the length of the cardiorespiratory recordings conducted during the test, which was started at the very beginning of the test and terminated when the subject was unable to continue.

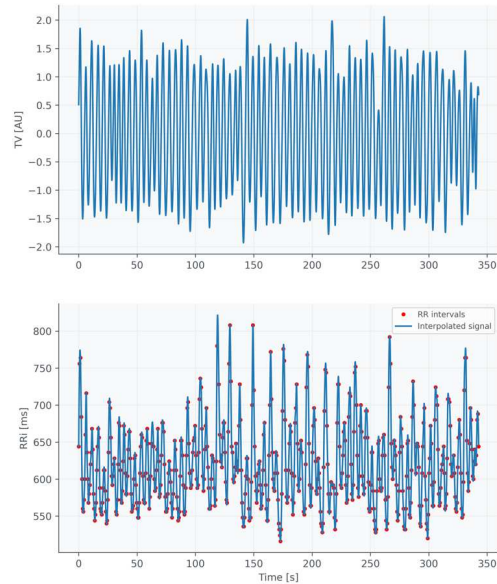


Fig. 1. Example of the obtained RR intervals, tachogram and TV signals based on which all utilized parameters were calculated.

The ML modeling was performed using the aforementioned demographic, HRV, respiratory, causal and information domain-based parameters as input feature vectors. As a first step, only the features with a Pearson correlation coefficient in relation to CPET duration, that was greater than 0.2, were included in the analysis. The following regression models were used for the prediction of the athletic performance in terms of the duration of the CPET: Linear Regression, Random Forest Regressor, Support Vector Regression, Bayesian Ridge Regression, Bayesian Automatic Relevance Determination Regression, Gradient Boosting Regressor, Theil-Sen Regression, Huber Regressor, Multilayer Perceptron, all with hyperparameter tuning applied. The assessment of the prediction accuracy was performed using mean absolute percentage error (MAPE), mean absolute error (MAE), root mean square error (RMSE),  $R^2$  score, and Pearson correlation coefficient with leave-one-out validation technique where each sample of the dataset was used once as the test set while the remaining samples formed the training set, ensuring a thorough and unbiased evaluation of the model's predictive performance on every data point. All metrics were calculated after the validation based on a comparison of the prediction results obtained on the test samples and the actual values of the test duration. In order to investigate the influence of the individual features on the prediction results the explainable artificial intelligence (XAI) was applied in the form of Shapley additive explanations (SHAP) values calculated for the test data. All analyses were performed using Python 3.10.8.

### III. RESULTS

Out of the initial 153 features, 39 were used after the feature selection process. The best results in terms of MAPE were obtained for the linear model with Lasso regularization with an alpha parameter equal to 1.65 applied. The detailed results are presented in Tab. 2. The relationship between the predicted values and the actual CPET durations is presented in Fig. 2.

TABLE II. RESULTS OBTAINED FROM THE LEAVE-ONE-OUT VALIDATION.

| MAPE [%] | MAE [s] | RMSE [s] | $R^2$ score | Pearson R |
|----------|---------|----------|-------------|-----------|
| 17.4     | 129.1   | 170.3    | 0.52        | 0.74      |

The feature importance assessed using Shapley values calculated on the test data is visualized in Fig 3. The most influential feature was the mean of the standard deviations of RRi extracted from 2-minute segments (SDNNI2), however, also demographic data (Age and BMI) and nonlinear HRV features had a notable influence on the prediction results. In terms of the causal features, the greatest impact on the prediction was observed for the value of causality from the respiratory signal to the tachogram, assessed using the Theil-Sen regressor ( $\text{Resp} \rightarrow \text{RR}_{\text{TSR}}$ ). For the information-domain features, it was the STE value in the same direction ( $\text{Resp} \rightarrow \text{RR}_{\text{STE}}$ ).

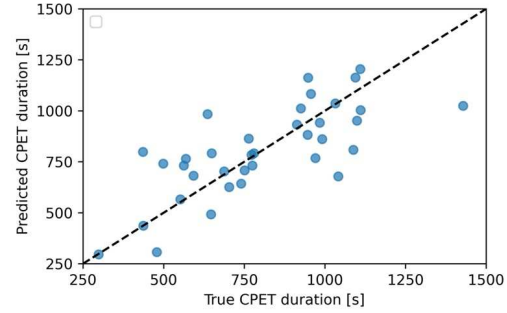


Fig. 2. The plot of actual and predicted CPET duration. The dashed black line represents the line of identity.

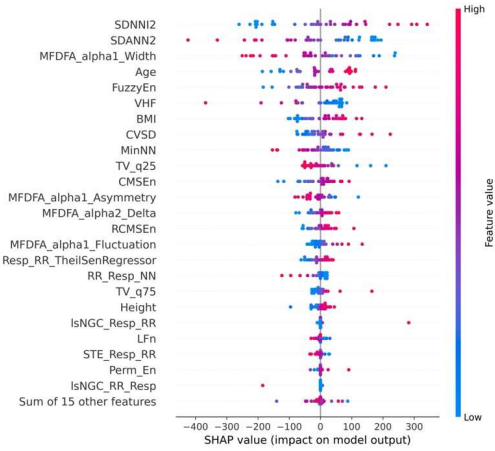


Fig. 3. Shapley values obtained on the test data. Feature names are explained in Appendix 1.

### IV. DISCUSSION

Various cardiorespiratory features used to predict the duration of maximal exercise tests in professional young male football players were analyzed. The main finding is that the duration of the CPET can be predicted based on the parameters collected during rest in a supine position with a moderate accuracy (MAPE of 17.4%).

In the presented setup, the higher predicted values were obtained for older subjects with greater BMI, SDNNI2, or Fuzzy Entropy values, while smaller SDANN2 or VHF HRV parameters. In contrast to the results presented in [10] where  $\text{VO}_{2\text{peak}}$  values were predicted based on submaximal CPET the respiratory-based features were not of primary importance in terms of the impact on the model output. The most influential respiratory parameter was the 25<sup>th</sup> quantile of relative TV, high values of which indicated a smaller predicted value. Causal/information domain-based features also appeared to be

valuable predictors of the performance as the increase of  $\text{Resp} \rightarrow \text{RR}_{\text{TSR}}$  was associated with the increase of predicted test duration, while for the most influential information-domain  $\text{Resp} \rightarrow \text{RR}_{\text{STE}}$ , the dependence was the opposite. As features designated as  $\text{Resp} \rightarrow \text{RR}$  might be interpreted as related to the respiratory sinus arrhythmia phenomenon [12], this suggests a relationship between athleticism and the interplay between the cardiac and respiratory systems. The prediction of physical performance is important for athletes, but also for patients with cardiovascular diseases [22] and physical task performance [23]. The clear advantage of the presented setting is the usage of impedance pneumography, which is a simpler and cheaper alternative for respiratory monitoring to sophisticated gas analyzing systems, which require a tight-fitting face mask that disturbs the natural airflow. The results presented in this study are similar in terms of the obtained  $R^2$  score compared to other studies that aimed at predicting cycling performance success during race [24] or 6-minute walk test [22]. Although the obtained results are suboptimal, the practical use would require further improvements. It is also worth mentioning that approaches for obtaining respiratory signals other than IP, such as airflow measurement or ECG-driven respiration, might affect the CRC quantification [25] and thus also warrant additional studies. The main limitation of the study is a small number of subjects and thus samples for the training of the machine learning models. Moreover, the CPET was performed on a cycleergometer, which does not reflect the primary activity undertaken by subjects. The study was conducted only on male subjects, which limits the generalizability of the results to the entire population. In conclusion, this study demonstrates the feasibility of using a combination of cardiorespiratory features from different modalities and machine learning to predict maximal exercise test duration in professional adolescent male football players. However, further research with larger sample sizes and varied physical activities is needed to enhance the model's accuracy and applicability.

#### REFERENCES

- [1] R. Ross et al., "Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement from the American Heart Association," *Circulation*, vol. 134, no. 24, 2016, doi: 10.1161/CIR.0000000000000461.
- [2] A. Mezzani, "Cardiopulmonary exercise testing: Basics of methodology and measurements," in *Annals of the American Thoracic Society*, 2017, doi: 10.1513/AnnalsATS.201612-997FR.
- [3] K. K. Buttar, N. Saboo, and S. Kacker, "A Review: Maximal Oxygen Uptake (VO<sub>2</sub> Max) and Its Estimation Methods," *International Journal of Physical Education, Sports and Health*, vol. 6, no. 6, 2019.
- [4] K. D. Cooper and A. B. Shafer, "Validity and Reliability of the Polar A300's Fitness Test Feature to Predict VO<sub>2</sub>max," *Int J Exerc Sci*, vol. 12, no. 4, 2019.
- [5] M. Mlynčzak, "Temporal orders and causal vector for physiological data analysis," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2020, doi: 10.1109/EMBC44109.2020.9176842.
- [6] A. Ashfaq, N. Cronin, and P. Müller, "Recent advances in machine learning for maximal oxygen uptake (VO<sub>2</sub> max) prediction: A review," *Informatics in Medicine Unlocked*, vol. 28, 2022, doi: 10.1016/j.imu.2022.100863.
- [7] D. J. Plews, P. B. Laursen, J. Stanley, A. E. Kilding, and M. Buchheit, "Training adaptation and heart rate variability in elite endurance athletes: Opening the door to effective monitoring," *Sports Medicine*, vol. 43, no. 9, 2013, doi: 10.1007/s40279-013-0071-8.
- [8] B. Makivić, M. D. Nikić, and M. S. Willis, "Heart Rate Variability (HRV) as a Tool for Diagnostic and Monitoring Performance in Sport and Physical Activities," *J Exerc Physiol Online*, vol. 16, no. 3, 2013.
- [9] T. Beltrame, R. Amelard, A. Wong, and R. L. Hughson, "Prediction of oxygen uptake dynamics by machine learning analysis of wearable sensors during activities of daily living," *Sci Rep*, vol. 7, 2017, doi: 10.1038/srep45738.
- [10] M. Rosol, M. Petelczyc, J. S. Gąsior, M. Mlynčzak, "Prediction of peak oxygen consumption using cardiorespiratory parameters from warmup and submaximal stage of treadmill cardiopulmonary exercise test" *PLOS ONE* 19(1): e0291706. <https://doi.org/10.1371/journal.pone.0291706>
- [11] M. Baumert, M. Javorka, and M. Kabir, "Joint symbolic dynamics for the assessment of cardiovascular and cardiorespiratory interactions," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 373, no. 2034, 2015, doi: 10.1098/rsta.2014.0097.
- [12] M. Rosol, J. S. Gąsior, I. Walecka, B. Werner, G. Cybulski, and M. Mlynčzak, "Causality in cardiorespiratory signals in pediatric cardiac patients," *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2022, 2022, doi: 10.1109/EMBC48229.2022.9871750.
- [13] M. Mlynčzak, W. Niewiadomski, M. Zylinski, and G. P. Cybulski, "Ambulatory impedance pneumography device for quantitative monitoring of volumetric parameters in respiratory and cardiac applications," in *Computing in Cardiology*, 2014.
- [14] J. S. Gąsior et al., "Validity of the Pneumonitor for RR intervals acquisition for short-term heart rate variability analysis extended with respiratory data in pediatric cardiac patients," *Kardiologia*, vol. 81, no. 5, 2023, doi: 10.33963/KP.a2023.0070.
- [15] V. P. Seppä, J. Hyttinen, M. Uitto, W. Chrapek, and J. Viik, "Novel electrode configuration for highly linear impedance pneumography," *Biomedizinische Technik*, 2013, doi: 10.1515/bmt-2012-0068.
- [16] D. Makowski et al., "NeuroKit2: A Python toolbox for neurophysiological signal processing," *Behav Res* 53, 1689–1696 2021, doi: 10.3758/s13428-020-01516-y
- [17] R. M. de Abreu, B. Cairo, and A. Porta, "On the significance of estimating cardiorespiratory coupling strength in sports medicine," *Frontiers in Network Physiology*, vol. 2, 2023, doi: 10.3389/fnetp.2022.1114733.
- [18] M. Rosol, M. Mlynčzak, and G. Cybulski, "Granger causality test with nonlinear neural-network-based methods: Python package and simulation study," *Comput Methods Programs Biomed*, vol. 216, 2022, doi: 10.1016/j.cmpb.2022.106669.
- [19] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel method for nonlinear Granger causality," *Phys. Rev. Lett.*, 2008, doi: 10.1103/PhysRevLett.100.144103.
- [20] A. Wismüller, A. M. Dsouza, M. A. Vosoughi, and A. Abidin, "Large-scale nonlinear Granger causality for inferring directed dependence from short multivariate time-series data," *Sci Rep*, 2021, doi: 10.1038/s41598-021-87316-6.
- [21] M. Rosol, [https://github.com/mrosol/CPET\\_duration\\_EMBC2024](https://github.com/mrosol/CPET_duration_EMBC2024) (accessed Apr. 27, 2024)
- [22] X. Chen et al., "Prediction of 6-minute walk performance in patients with peripheral artery disease," *J Vasc Surg*, vol. 66, no. 4, 2017, doi: 10.1016/j.jvs.2017.03.438.
- [23] S. L. Sutterfield et al., "Prediction of Planetary Mission Task Performance for Long-Duration Spaceflight," *Med Sci Sports Exerc*, vol. 51, no. 8, 2019, doi: 10.1249/MSS.0000000000001980.
- [24] P. Leo, J. Spragg, J. Wakefield, and J. Swart, "Predictors of cycling performance success: Traditional approaches and a novel method to assess performance capacity in U23 road cyclists," *J Sci Med Sport*, vol. 26, no. 1, 2023, doi: 10.1016/j.jsams.2022.11.005.
- [25] B. Cairo et al., "Comparison Between ECG-Derived Respiration and Respiratory Flow for the Assessment of Cardiorespiratory Coupling Before and After Cardiopulmonary Exercise Test Protocol," in *Computing in Cardiology*, 2022, doi: 10.22489/CinC.2022.103.

Article

# Machine Learning Classification of Pediatric Health Status Based on Cardiorespiratory Signals with Causal and Information Domain Features Applied—An Exploratory Study

Maciej Rosol <sup>1,\*</sup> , Jakub S. Gašior <sup>2</sup> , Kacper Korzeniewski <sup>1</sup>, Jonasz Łaba <sup>1</sup>, Robert Makuch <sup>3</sup> , Bożena Werner <sup>2</sup>  and Marcel Młynczak <sup>1</sup> 

- <sup>1</sup> Institute of Metrology and Biomedical Engineering, Faculty of Mechatronics, Warsaw University of Technology, 02-525 Warsaw, Poland; kacper.korzeniewski.dokt@pw.edu.pl (K.K.); jonasz.laba.dokt@pw.edu.pl (J.L.); marcel.mlynczak@pw.edu.pl (M.M.)
  - <sup>2</sup> Department of Pediatric Cardiology and General Pediatrics, Medical University of Warsaw, 02-091 Warsaw, Poland; jakub.gasior@wum.edu.pl (J.S.G.); bozena.werner@wum.edu.pl (B.W.)
  - <sup>3</sup> Department of Physical Education, Kazimierz Pulaski University of Technology and Humanities in Radom, 26-600 Radom, Poland; r.makuch@uthrad.pl
- \* Correspondence: maciej.rosol.dokt@pw.edu.pl

**Abstract: Background/Objectives:** This study aimed to evaluate the accuracy of machine learning (ML) techniques in classifying pediatric individuals—cardiological patients, healthy participants, and athletes—based on cardiorespiratory features from short-term static measurements. It also examined the impact of cardiorespiratory coupling (CRC)-related features (from causal and information domains) on the modeling accuracy to identify a preferred cardiorespiratory feature set that could be further explored for specialized tasks, such as monitoring training progress or diagnosing health conditions. **Methods:** We utilized six self-prepared datasets that comprised various subsets of cardiorespiratory parameters and applied several ML algorithms to classify subjects into three distinct groups. This research also leveraged explainable artificial intelligence (XAI) techniques to interpret model decisions and investigate feature importance. **Results:** The highest accuracy, over 89%, was obtained using the dataset that included most important demographic, cardiac, respiratory, and interrelated (causal and information) domain features. The dataset that comprised the most influential features but without demographic data yielded the second best accuracy, equal to 85%. Incorporation of the causal and information domain features significantly improved the classification accuracy. The use of XAI tools further highlighted the importance of these features with respect to each individual group. **Conclusions:** The integration of ML algorithms with a broad spectrum of cardiorespiratory features provided satisfactory efficiency in classifying pediatric individuals into groups according to their actual health status. This study underscored the potential of ML and XAI in advancing the analysis of cardiorespiratory signals and emphasized the importance of CRC-related features. The established set of features that appeared optimal for the classification of pediatric patients should be further explored for their potential in assessing individual progress through training or rehabilitation.

**Keywords:** cardiorespiratory parameters; machine learning; causality; XAI; cardiorespiratory coupling; health status



**Citation:** Rosol, M.; Gašior, J.S.; Korzeniewski, K.; Łaba, J.; Makuch, R.; Werner, B.; Młynczak, M. Machine Learning Classification of Pediatric Health Status Based on Cardiorespiratory Signals with Causal and Information Domain Features Applied—An Exploratory Study. *J. Clin. Med.* **2024**, *13*, 7353. <https://doi.org/10.3390/jcm13237353>

Academic Editor: Antonio Jorge Forte

Received: 6 November 2024

Revised: 25 November 2024

Accepted: 29 November 2024

Published: 2 December 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The assessment of cardiovascular function in ambulatory or field conditions (e.g., during physical training, athletic monitoring, or routine primary care visits) has predominantly relied on electrocardiography (ECG), a non-invasive measurement of the electrical activity of the heart. The intervals between the consecutive R peaks from a QRS complex detected from ECG recordings can be used to calculate the heart rate variability (HRV) parameters in the time, frequency, and nonlinear domains, which constitute valuable markers in various

health conditions [1–4]. Importantly, many studies emphasized the value of incorporating respiratory data, such as the respiratory rate (RespRate), tidal volume (TV), and pulmonary ventilation, to enhance the clinical relevance of HRV analysis [2,5–7]. Moreover, respiration acts as a confounder for cardiovascular and cerebrovascular controls [8] and is necessary for the assessment of the baroreflex role [9]. Recently, there has been a growing interest in introducing new cardiorespiratory parameters, which could benefit from the diagnostic information hidden in the interdependence and cooperation of cardiac and respiratory systems [10]. This linkage is known as cardiorespiratory coupling (CRC), which is reflected in phenomena like respiratory sinus arrhythmia (RSA) or baroreceptor coupling [11]. These interdependencies can be quantified based on the HRV associated with breathing [12] or by using parameters from the causal or information domains, which simultaneously utilize both cardiac and respiratory signals for such quantification [13–16]. The causal analysis of cardiorespiratory signals, mostly based on the Granger causality (GC), allows for the identification and quantification of directional influences between the cardiac and respiratory systems. By analyzing the temporal sequence of events, the GC can determine whether changes in one system can improve the prediction of changes in the other, providing insight into the interplay between both heart and lung functions. When testing the causal influence, e.g., from the respiratory signal to the tachogram (denoted as Resp→RR), with this method, two autoregressive models are created. The first model predicts the current value of the tachogram based on the  $p$  defined number of past values of this signal (Equation (1)), while the second model predicts the same current value of the tachogram but based on the past values of the cardiological and respiratory signals (Equation (2)):

$$RR(t) = \sum_{i=1}^p A_i RR(t-i) + \varepsilon_1 \quad (1)$$

$$RR(t) = \sum_{i=1}^p B_i RR(t-i) + \sum_{i=1}^p C_i Resp(t-i) + \varepsilon_2 \quad (2)$$

Then, the measure of the causality Resp→RR can be defined as the logarithm of the ratio of the variances of the models' residuals  $\varepsilon_1$  and  $\varepsilon_2$ , as shown in Equation (3) [17]:

$$GC_{Resp \rightarrow RR} = \ln \frac{\sigma^2 \varepsilon_1}{\sigma^2 \varepsilon_2} \quad (3)$$

While traditional GC relies on linear modeling, more sophisticated nonlinear approaches were also developed to enable the analysis of more complex relationships [18–20]. The information domain quantification of the interdependencies between signals is mostly based on the entropy parameters [21,22]. Both causal- and information-based parameters are commonly applied to detect direct and indirect couplings in time series; thus, they are also useful for CRC quantification [14,23,24]. Notwithstanding, there is a lack of literature on the possible descriptive and diagnostic utility of such parameters.

As computational power increases and more cardiorespiratory parameters become available, the use of machine learning (ML) tools for biomedical data analysis becomes more popular [25]. This trend is advancing the fields of precision and individualized medicine [26,27]. ML algorithms and wearable devices play a crucial role in these contexts, enabling continuous monitoring of physiological signals and advanced analysis of data to support tailored interventions [28,29]. Personalized information about a subject's health status, based on cardiorespiratory data, can be presented either as a continuous parameter (corresponding to a regression problem in ML) or as discrete labels (through ML classification). To achieve precise personalization, it is essential to identify the physiological parameters that most accurately reflect an individual's health condition and enable differentiation between various health statuses. Determining these key parameters enables further tailoring of ML models for personalized insights, preferably based on data gathered from wearable devices. Such insights enable clinicians and coaches to customize inter-



ventions effectively and monitor progress with greater precision. Thus, determining the most relevant features from a broad range of cardiorespiratory data is a critical first step in enhancing diagnostic accuracy and improving individualized care. Despite ML models achieving human-level performance across various tasks, their perception as inscrutable “black boxes” greatly limits the understanding of their decision-making foundations, thus undermining their broader acceptance and application in medicine [30]. To address this issue, the use of explainable artificial intelligence (XAI) techniques has gained popularity. These methodologies play a crucial role in enhancing the interpretability and trustworthiness of ML models, thereby elevating their utility within professional settings. This progress is crucial in bridging the gap between complex ML algorithms and real-world applications, ensuring that their integration into various domains is both effective and ethically responsible [31].

With the growing emphasis on personalized medicine, there is an increasing demand for individualized assessments of health status to optimize treatment, rehabilitation, workouts, and intervention strategies [32–34]. For instance, CRC was recently used to determine the optimal breathing training frequency [35]. Individualized approaches are particularly important in pediatric populations, where only 40% of youth are currently believed to have an optimal cardiorespiratory fitness (CRF) level, a crucial marker of physical and mental health, as well as academic achievement [36]. Furthermore, assessing one’s health status in terms of CRF and muscular fitness is essential for young individuals, as both are positively associated with health-related quality of life, particularly in the physical, psychological, and social domains in this population [37]. Moreover, higher CRF during childhood and adolescence is associated with better cardiometabolic health parameters later in life, emphasizing the long-term benefits of early interventions targeting CRF [38]. These factors highlight the importance of individualized assessments of health status in pediatric populations. In this study, we made an effort to explore the capabilities of ML in classifying the health statuses of pediatric subjects from three distinct groups. This allowed for the identification of an optimal set of cardiorespiratory features and lay the groundwork for further personalized modeling.

This study aimed to evaluate the accuracy of ML techniques in classifying pediatric individuals with respect to their health status—including patients with heart disease, healthy participants, and trained athletes—based on cardiorespiratory features calculated from short-term measurements taken under static conditions. Additionally, this study investigated the importance of CRC-related features by examining their influence on modeling accuracy, hypothesizing that these features capture unique physiological interactions between cardiac and respiratory systems, thereby introducing additional information about the subject’s health status and improving the performance of machine learning models. Moreover, this evaluation was performed to establish a preferred set of features that could be used for further development in more specialized classification or regression tasks related to assessing individual progress through training or rehabilitation or diagnosing specific health conditions.

## 2. Materials and Methods

### 2.1. Study Design

The inclusion criteria for this study were ages between 6 and 18 years old and given written informed consent, while the exclusion criteria were signs of infection and diagnosed additional disorders that may affect the functioning of the autonomic nervous system. Subjects were assigned to three distinct groups (which also served as labels for the ML classification) according to their health status based on the following criteria:

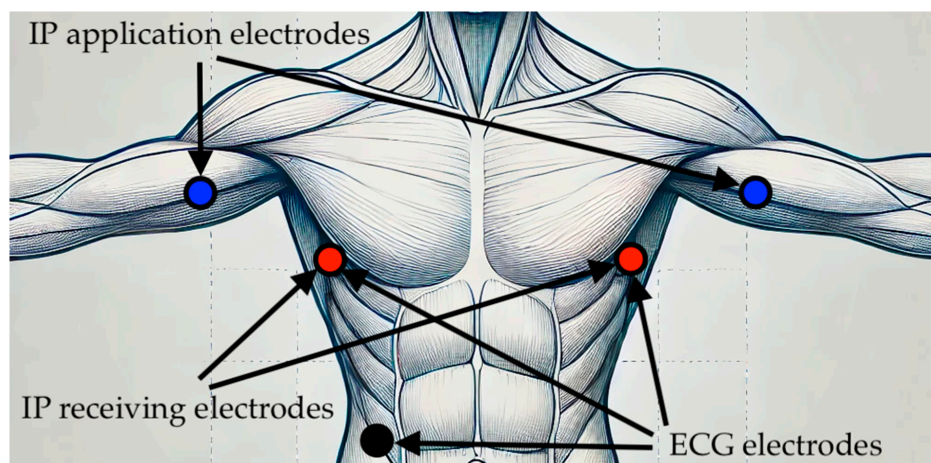
- Cardiac—subjects with an ongoing cardiac disease requiring hospitalization;
- Healthy—subjects without any active heart disease, whether sedentary or recreationally active subjects according to McKay classification [39];



- Sport-trained adolescent athletes [39,40] (soccer players) affiliated with a sports club, with at least 3 years of training experience and regularly training ~3 times per week with a purpose to compete.

For the cardiorespiratory data acquisition, all participants took part in ECG and impedance pneumography (IP) recordings performed for at least 5 min at rest in the supine position using the Pneumonitor device. This apparatus is a recently developed and validated device for cardiorespiratory monitoring that allows for the simultaneous acquisition of these two signals [41–44]. In the IP method, a small electrical current below the tissue excitability threshold is applied through the application electrodes, and the voltage response is measured across the same or an additional pair of electrodes (receiving electrodes). As a person breathes, the air volume in the lungs changes, causing variations in the impedance within the chest, which are measured by the IP technique.

A tetrapolar measurement using a sinusoidal current with an amplitude of up to 1 mA and a frequency of 100 kHz, along with electrode placement configured according to [45], was applied. Based on the findings in [46], it was presumed that such conditions allow for linear fitting to optimally align the IP with direct breathing measurements, e.g., using a facemask or nose cannula. Consequently, this alignment permits the IP signal to be treated as an equivalent to the relative TV. The placement of the electrodes used for the ECG and IP is presented in Figure 1.



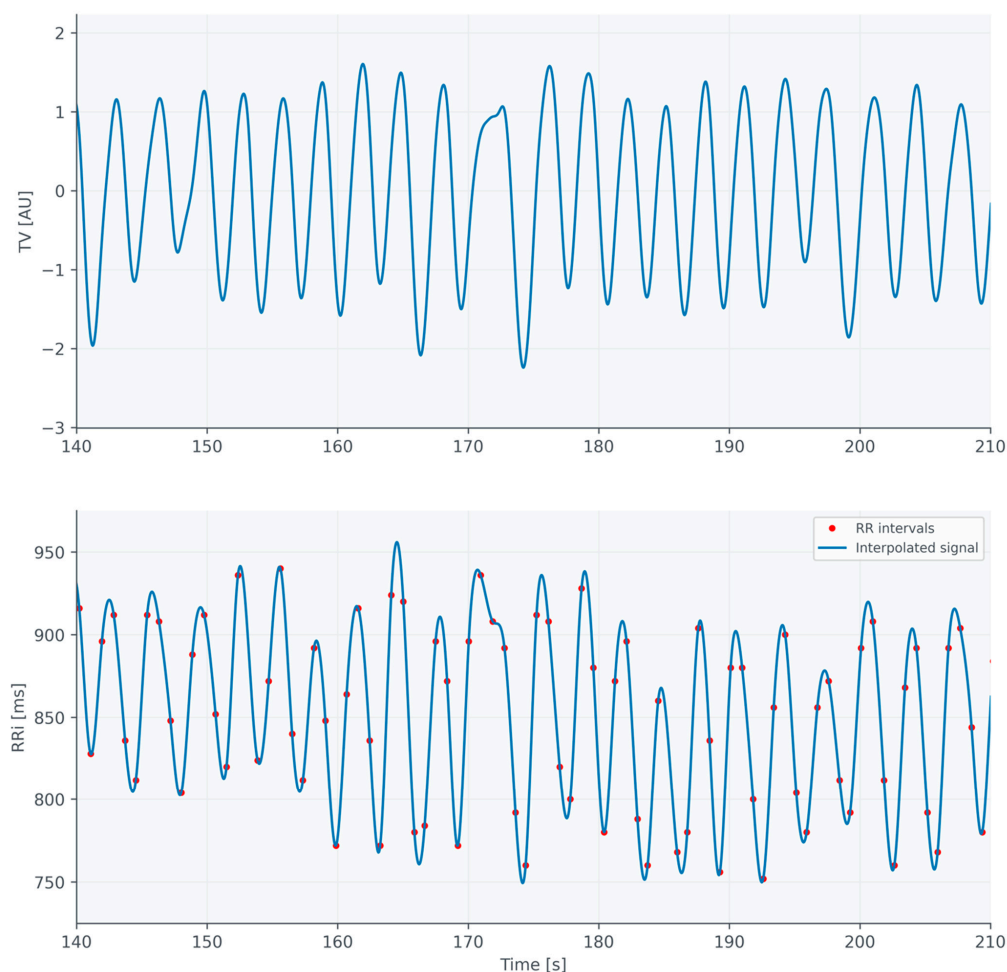
**Figure 1.** Placement of the electrodes used for the ECG and IP measurements.

In terms of ML, the modeling parameters derived from the cardiorespiratory recordings served as the model inputs and information about the group assignment was used as the output. This study was approved by two ethics committees (permissions: KB/55/N02/2019, 5 June 2019 and KB/70/2021, 14 June 2021) and conducted in accordance with the Declaration of Helsinki. Written informed consent forms were obtained from the legal guardians of subjects younger than 16 years old and directly from the subjects themselves if they were 16 years or older.

## 2.2. Signal Processing

Both the ECG and IP were acquired with a 250 Hz sampling frequency. The raw IP signal was filtered with a bandpass filter with cutoff frequencies of 0.05 and 0.67 Hz, corresponding to 3 and 40 breaths per minute, respectively; thus, the respiratory signal (Resp) was obtained. RR intervals (RRi) were extracted from the ECG signal using automatic

detection, followed by manual correction by an experienced physician. The stationarity of the original RRi series was confirmed using the Phillips–Perron test. Such obtained series of RRi were interpolated using cubic interpolation in order to obtain a tachogram time series (RR) with the same sampling as the respiratory signal (which enabled estimating causal and information domain features based on the signals, not only beat-by-beat sequences). Both signals were then down-sampled to 25 Hz to reduce the computational complexity (only for the calculation of a subset of causal and information domain features). Examples of the obtained signals are presented in Figure 2.



**Figure 2.** Examples of TV time series in the top chart and tachogram (blue line) with an individual RRi (red dots) in the bottom chart for Healthy subject #40.

### 2.3. Parameters Calculation

Three types of cardiorespiratory parameters were calculated: HRV (time and frequency domains and nonlinear), respiratory parameters, and parameters from causal and information domains. HRV parameters were calculated using the Neurokit2 package [47], extended

with parameters from symbolic dynamics analysis [48]. From the respiratory signal, statistical characteristics, such as the RespRate, relative TV (indexed by the median TV due to the lack of calibration and the inability to convert the measured impedance signal directly into milliliters), and the inspiration/expiration time ratio, were derived. In terms of the causal relationships between cardiac and respiratory signals, features were calculated using the GC [49], the nonlincausality package with various ML models applied [18,50], the kernel GC [20], and the large-scale nonlinear Granger causality (lsNGC) [19]. Parameters for the information domain were mostly based on entropy analysis, but also simple statistics, like the highest Pearson correlation coefficient between the signals for a time lag between  $-1$  and  $1$  s. The full list of features and their descriptions is presented in Appendix A, while the code used for their calculation is available in the repository [51]. As a result, for each patient, a total of 157 features were calculated, including 5 demographic (age, weight, height, sex, and body mass index), 102 cardiac, 18 respiratory, and 32 causal/information features.

#### 2.4. Modeling

Based on the aforementioned parameters, four datasets (described further using the prefix D) utilized as input for machine learning modeling were created according to different types of features. Dataset D1 included demographic and cardiological features. Dataset D2 contained the same features as D1, with the addition of respiratory features. Dataset D3 expanded further by incorporating causal and information domain features. Finally, dataset D4 consisted of cardiological, respiratory, causal, and information domain features, excluding demographic data. The dataset components are presented in Table 1.

**Table 1.** Description of the content of each dataset based on the type of features, where “+” indicates that the given features are included in the respective dataset.

| Dataset | Demographic Data | Cardiological Features | Respiratory Features | Causal and Information Domain Features |
|---------|------------------|------------------------|----------------------|----------------------------------------|
| D1      | +                | +                      |                      |                                        |
| D2      | +                | +                      | +                    |                                        |
| D3      | +                | +                      | +                    | +                                      |
| D4      |                  | +                      | +                    | +                                      |

Moreover, two more datasets, D5 and D6, were created based on the 35 most influential features determined based on the Shapley values from datasets D3 and D4, respectively, in order to simplify the ML models, potentially further increase their accuracy, and evaluate the approach using features that most accurately reflected an individual’s health condition, making them preferable for future studies. Features for each patient were labeled according to their assigned group (Cardiac/Healthy/Sport). For the classification, various popular machine learning algorithms were utilized, including Logistic Regression (also with Ridge and Lasso regularization), Decision Tree, Support Vector Machine, Random Forest, Gradient Boosting, Naïve Bayes, K-Nearest Neighbors, AdaBoost, XGBoost, and multilayer perceptron. Hyperparameter optimization was applied for each algorithm. To validate the classification, 10-fold cross-validation was performed. In this method the dataset was randomly divided into 10 equal-sized subsets called folds. The ML model was trained on nine of these folds and tested on the remaining fold. This process was repeated 10 times, each time using a different fold as the test set and the remaining folds for training. The final model performance was then calculated as the average of the results from all 10 iterations, providing a more robust estimate of the model’s performance by reducing the variance associated with random sampling of the data into training and test sets.

The following metrics were calculated: accuracy, precision, recall, F1 score, Mathew's correlation coefficient (MCC), and area under the curve (AUC) for each iteration on the test set according to Equations (4)–(8):

$$Accuracy = \frac{1}{n} \sum_{i=1}^n 1(\hat{y}_i = y_i), \quad (4)$$

$$Precision = \frac{TP}{TP + FP}, \quad (5)$$

$$Recall = \frac{TP}{TP + FN}, \quad (6)$$

$$F1score = \frac{2 * TP}{2 * TP + FN + FP}, \quad (7)$$

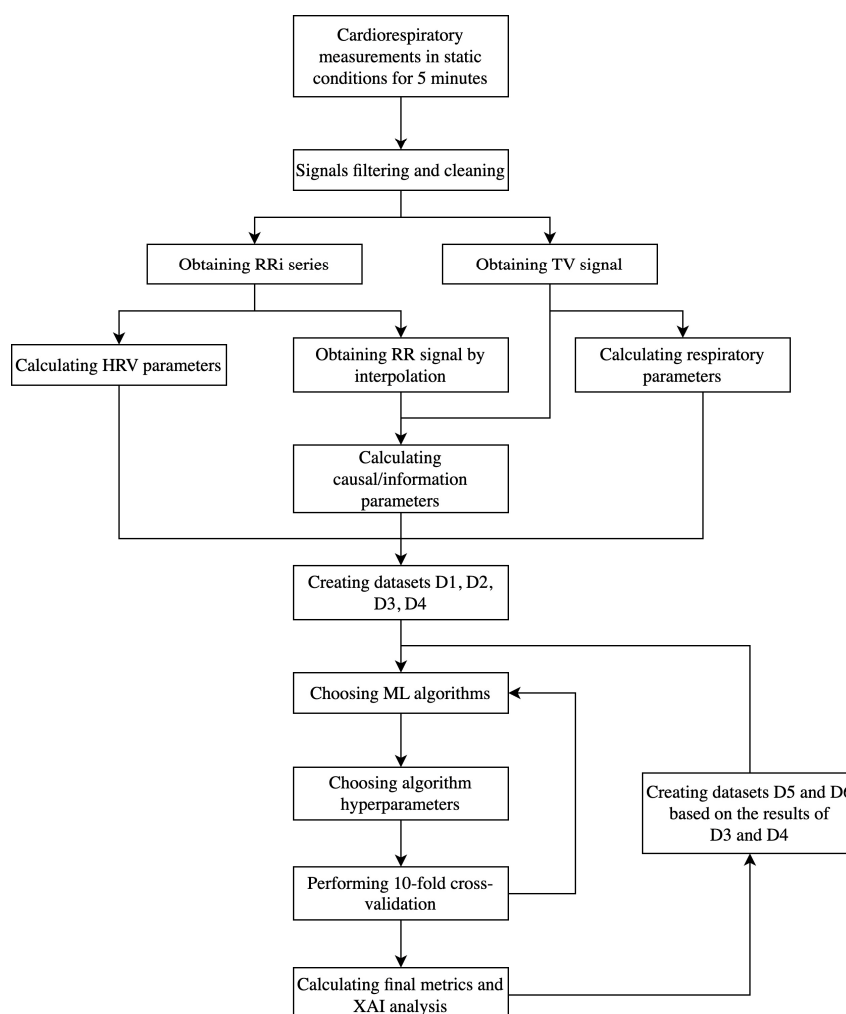
$$MCC = \frac{n * \sum_{i=1}^n 1(\hat{y}_i = y_i) - \sum_k^K p_k * t_k}{\sqrt{(n^2 - \sum_k^K p_k^2) * (n^2 - \sum_k^K t_k^2)}} \quad (8)$$

where  $1(x)$  is the indicator function,  $n$  is the number of samples,  $TP$  is true positive,  $FP$  is false positive,  $FN$  is false negative,  $p_k$  is the number of times class  $k$  was predicted, and  $t_k$  is the number of times class  $k$  truly occurred.

The mean values of the metrics from the cross-validation were treated as a final evaluation of the algorithm. The confusion matrix and receiver operating curve (ROC) were also visualized. In order to increase the training dataset and to handle class imbalance, up-sampling using the synthetic minority oversampling technique (SMOTE) [52] was applied to the training set at each iteration of the validation. The code used for the modeling is presented in [51]. For each dataset, the best algorithm was determined based on the highest accuracy value, whose results were taken for further analysis. The metrics from individual iterations of cross-validation were compared between datasets using the pairwise Wilcoxon signed-rank test to determine whether the inclusion of certain feature types improved the classification performance. The assumed level of significance was 0.05. The analysis was performed using Python 3.10.8. A full diagram of the performed analysis is presented in Figure 3.

## 2.5. Explainable AI

To study the significance of the different features in the machine learning models, tools for XAI were utilized for the four datasets that obtained the best results in terms of accuracy. The Dalex Python package was used to assess which features were the most important for the model's decisions using a permutation-based variable importance analysis [53]. Additionally, Shapley values were applied to understand how each feature influenced the individual predictions, which helped to explain the model's behavior in more detail for individual subjects [54]. During each iteration of the cross-validation, the Shapley values and variable importance were determined based on 30 permutation rounds, using 1-AUC as the loss function for the test set. Following the complete cross-validation process, all the Shapley values for each data point and feature were collated and visualized, along with the average importance values of the variables.



**Figure 3.** Diagram presenting the individual steps of the conducted analysis.

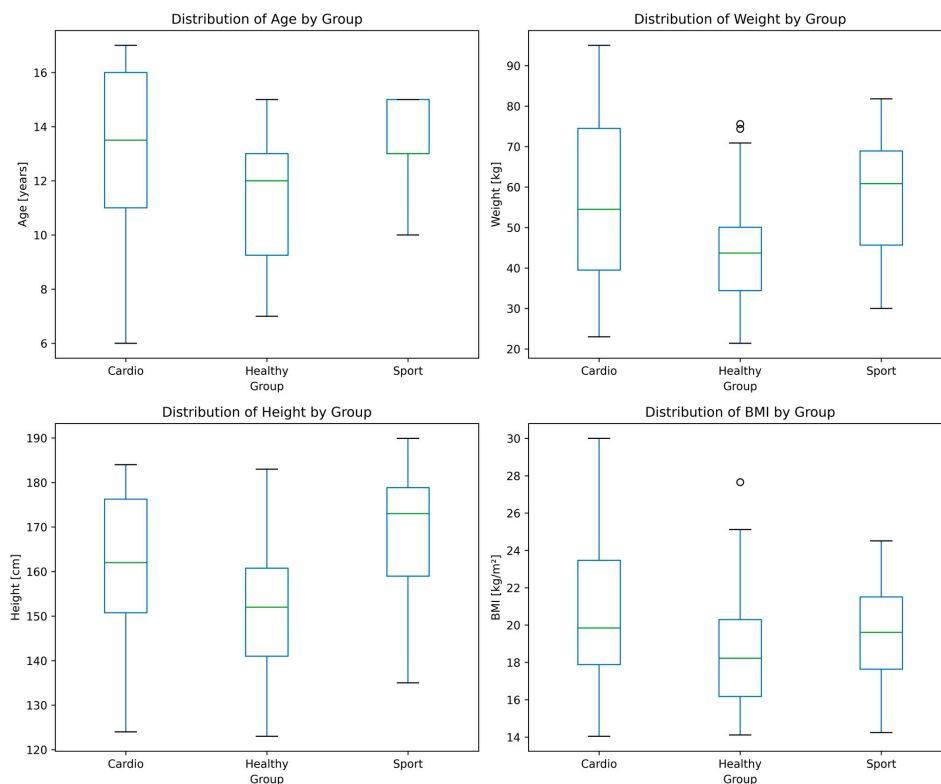
### 3. Results

A total of 135 subjects (97 male and 38 female) were included in this study. The descriptive statistics of all groups are presented in Table 2. The Cardiac group consisted of patients with the following conditions: congenital heart defect (17), cardiomyopathy/myocarditis (8), and arrhythmia (7). The Sport group consisted of individuals with an average training experience of  $5.82 \pm 1.19$  years (range 3–10 years) and a mean maximal oxygen uptake of  $46.55 \pm 4.42$  mL/kg/min (range 39.4–57.9 mL/kg/min). The distributions of age, body mass, height, and body mass index (BMI) are presented in Figure 4. The demographic parameters of the participants were compared using the Kruskal–Wallis test, as the data did not follow a normal distribution. Although this test indicated statistically significant differences between the groups in terms of these parameters, they were widely overlapping.

Assigning each individual subject to a given group based on any individual parameter was not possible; thus, advance machine learning modeling was utilized.

**Table 2.** Descriptive statistics of all three study groups and the overall study population. Values are presented as the mean  $\pm$  standard deviation and the range of the parameter in brackets.

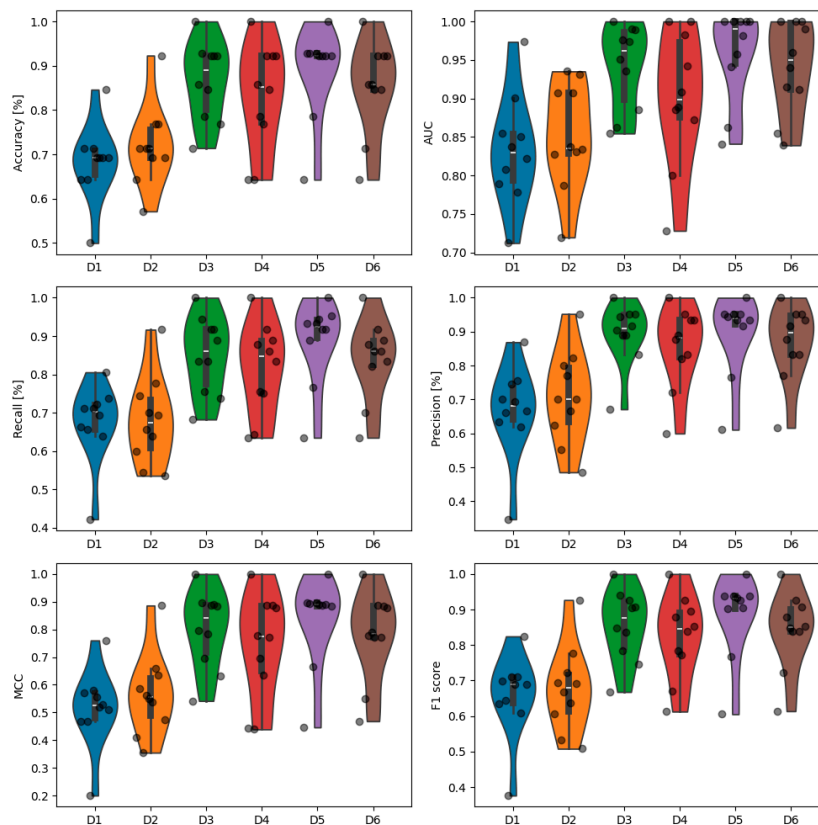
|                        | Cardiac                         | Healthy                         | Sport                           | Overall                         |
|------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| N                      | 29                              | 62                              | 44                              | 135                             |
| Male/female            | 20/9                            | 33/29                           | 44/0                            | 97/38                           |
| Age                    | 13.1 $\pm$ 3.5 (6–17)           | 11.0 $\pm$ 2.2 (7–15)           | 13.3 $\pm$ 1.4 (10–15)          | 12.2 $\pm$ 2.6 (6–17)           |
| Body mass [kg]         | 57.1 $\pm$ 21.0<br>(23.0–95.0)  | 43.5 $\pm$ 12.1<br>(21.4–75.6)  | 57.2 $\pm$ 13.6<br>(30.0–81.8)  | 50.9 $\pm$ 16.4<br>(21.4–95.0)  |
| Height [cm]            | 160.4 $\pm$ 17.2<br>(123–184)   | 151.2 $\pm$ 13.1<br>(123–183)   | 169.4 $\pm$ 12.7<br>(135–190)   | 159.1 $\pm$ 16.0<br>(123–190)   |
| HR [beats/min]         | 72.8 $\pm$ 13.3<br>(56.0–100.5) | 79.4 $\pm$ 10.2<br>(60.7–100.5) | 76.9 $\pm$ 15.0<br>(46.7–121.4) | 77.2 $\pm$ 12.8<br>(46.7–121.4) |
| RMSSD [ms]             | 55.3 $\pm$ 36.8<br>(9.4–140.7)  | 61.8 $\pm$ 34.4<br>(13.0–162.3) | 68.2 $\pm$ 46.7<br>(5.6–178.9)  | 62.5 $\pm$ 39.6<br>(5.6–178.9)  |
| RespRate [breaths/min] | 18.5 $\pm$ 4.6 (7.9–25.4)       | 18.8 $\pm$ 3.5 (10.7–28.5)      | 17.1 $\pm$ 3.5 (10.2–25.8)      | 18.2 $\pm$ 3.8 (7.9–28.5)       |



**Figure 4.** Distributions of the demographical parameters presented as boxplots. The central green line represents the median. Outliers, if present, are shown as individual points.

The metrics obtained for the best algorithm for each dataset alongside the upsampling proportions are presented in Table 3. The best results in terms of all metrics with accu-

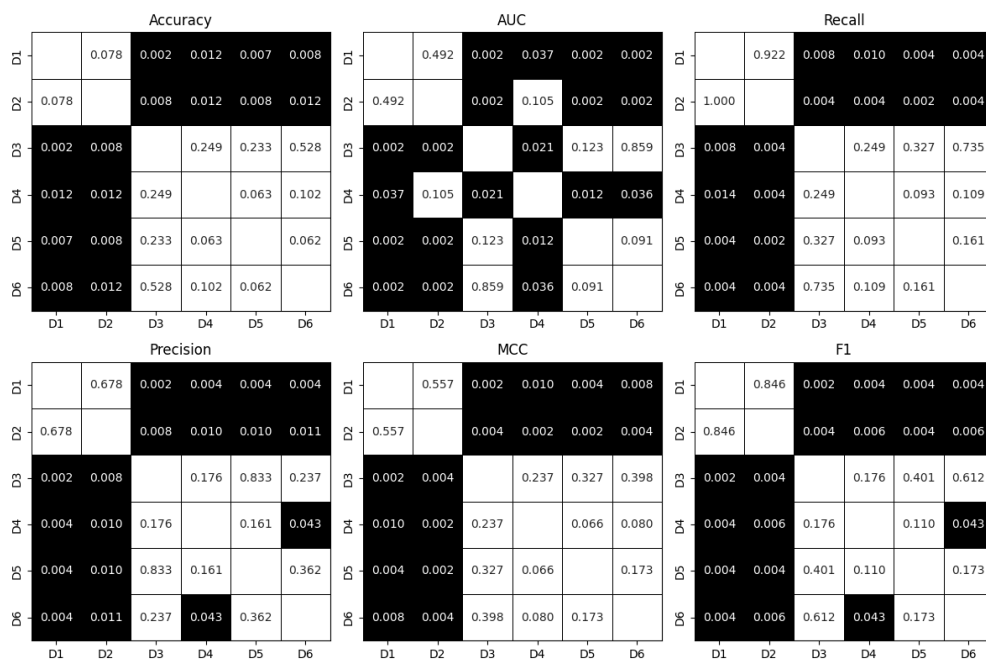
racy equal to 89.1% were obtained for the fifth dataset, which incorporated demographic, cardiac, respiratory, causal, and information domain features while using the Gradient Boosting model. The selection of the most important features resulted in an improvement in the performance, as all the metrics for D5 and D6 were superior compared with the corresponding D3 and D4, respectively. Dataset D6, which did not leverage the demographic data, had an accuracy of 85.3% with the usage of the Gradient Boosting model. The violin plots of the metrics obtained during individual iterations of the 10-fold cross-validation are presented in Figure 5. Datasets D3 to D6 generally showed better performances across most metrics, with D5 typically demonstrating the best overall results. D1 and D2 had lower median values and wider distributions of metrics, indicating poorer and less consistent performance. The pairwise comparison of the obtained metrics between datasets using the Wilcoxon signed-rank test after cross-validation are presented in Figure 6. There was no statistical difference between the metrics for datasets D1 and D2, while all the other datasets had significantly better results than these two (despite the AUC for D4 compared with D2). Moreover, D4 had a significantly smaller AUC compared with D3, D5, and D6. There was also a significant difference in terms of the precision and F1 score between D4 and D6. The use of the limited datasets with the 35 most important features improved the performance, although not statistically significantly.



**Figure 5.** Violin plots of the metric values obtained from the cross-validation for each dataset. The metrics obtained from the individual iterations of 10-fold cross validation are presented as black dots.

**Table 3.** Mean  $\pm$  standard deviation of metrics obtained from the 10-fold cross-validation for the given ML algorithm with the applied SMOTE upsampling technique with the strategy presented as a number of Cardiac/Healthy/Sport training samples.

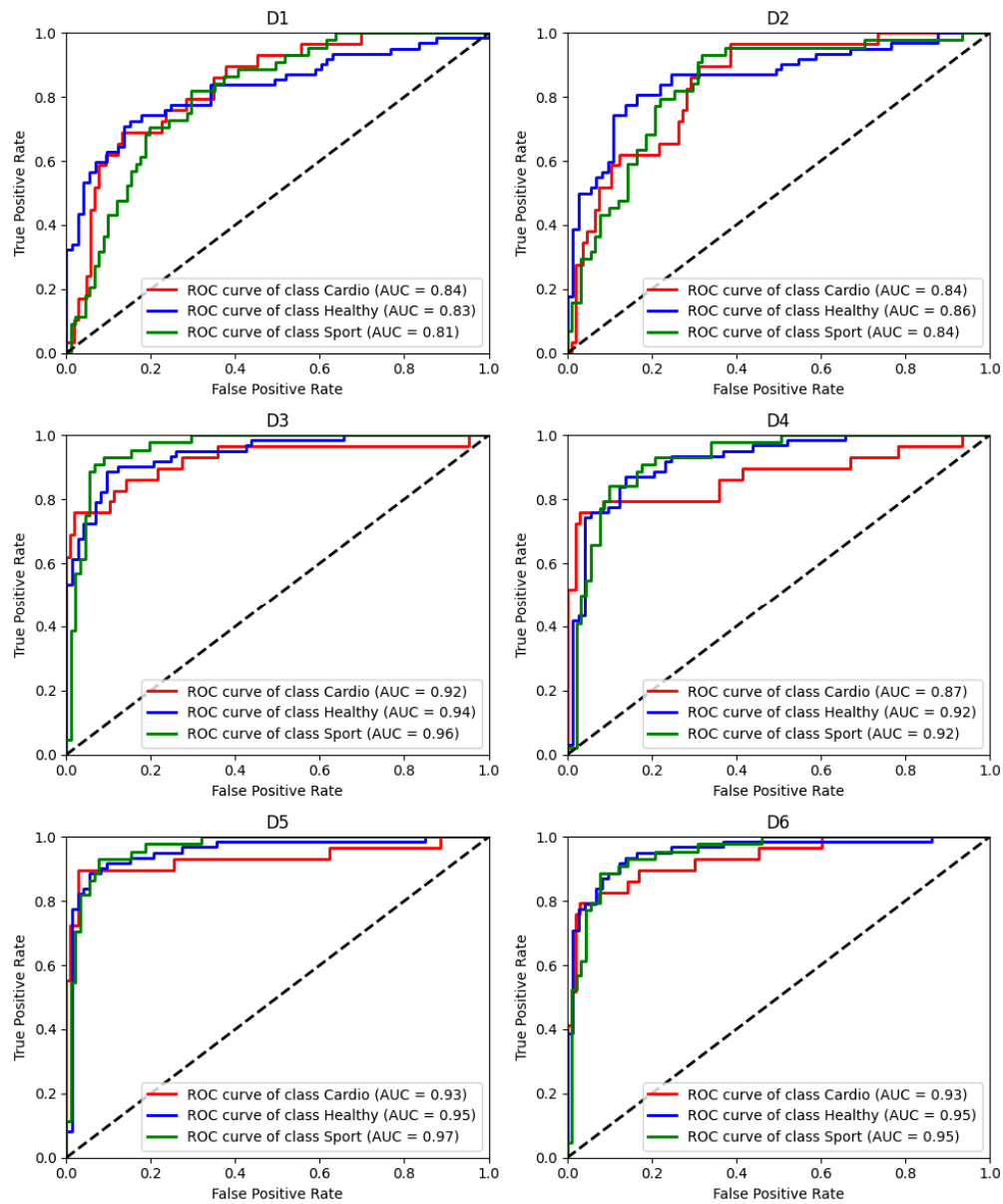
|                     | D1                 | D2                  | D3                | D4                | D5                | D6                |
|---------------------|--------------------|---------------------|-------------------|-------------------|-------------------|-------------------|
| Accuracy [%]        | 68.3 $\pm$ 8.1     | 72.0 $\pm$ 8.7      | 86.7 $\pm$ 8.4    | 83.1 $\pm$ 11.5   | 89.1 $\pm$ 9.6    | 85.3 $\pm$ 10.0   |
| AUC                 | 83.2 $\pm$ 6.7     | 85.2 $\pm$ 6.5      | 94.2 $\pm$ 5.2    | 90.1 $\pm$ 8.3    | 95.8 $\pm$ 5.7    | 94.1 $\pm$ 5.7    |
| Recall [%]          | 67.6 $\pm$ 9.6     | 68.1 $\pm$ 10.9     | 85.1 $\pm$ 9.6    | 81.6 $\pm$ 11.2   | 88.9 $\pm$ 10.2   | 84.0 $\pm$ 9.9    |
| Precision [%]       | 66.9 $\pm$ 12.7    | 70.8 $\pm$ 13.0     | 89.5 $\pm$ 8.6    | 85.6 $\pm$ 11.3   | 89.6 $\pm$ 11.1   | 86.9 $\pm$ 10.6   |
| MCC                 | 0.516 $\pm$ 0.132  | 0.566 $\pm$ 0.140   | 0.801 $\pm$ 0.133 | 0.742 $\pm$ 0.180 | 0.835 $\pm$ 0.151 | 0.778 $\pm$ 0.152 |
| F1 score            | 0.659 $\pm$ 0.109  | 0.676 $\pm$ 0.114   | 0.856 $\pm$ 0.095 | 0.823 $\pm$ 0.111 | 0.885 $\pm$ 0.109 | 0.843 $\pm$ 0.102 |
| ML algorithm        | XGBoost Classifier | Logistic Regression | Gradient Boosting | Gradient Boosting | Gradient Boosting | Gradient Boosting |
| Upsampling strategy | 200/200/200        | 200/200/150         | 200/200/200       | 200/200/200       | 200/200/200       | 200/200/200       |



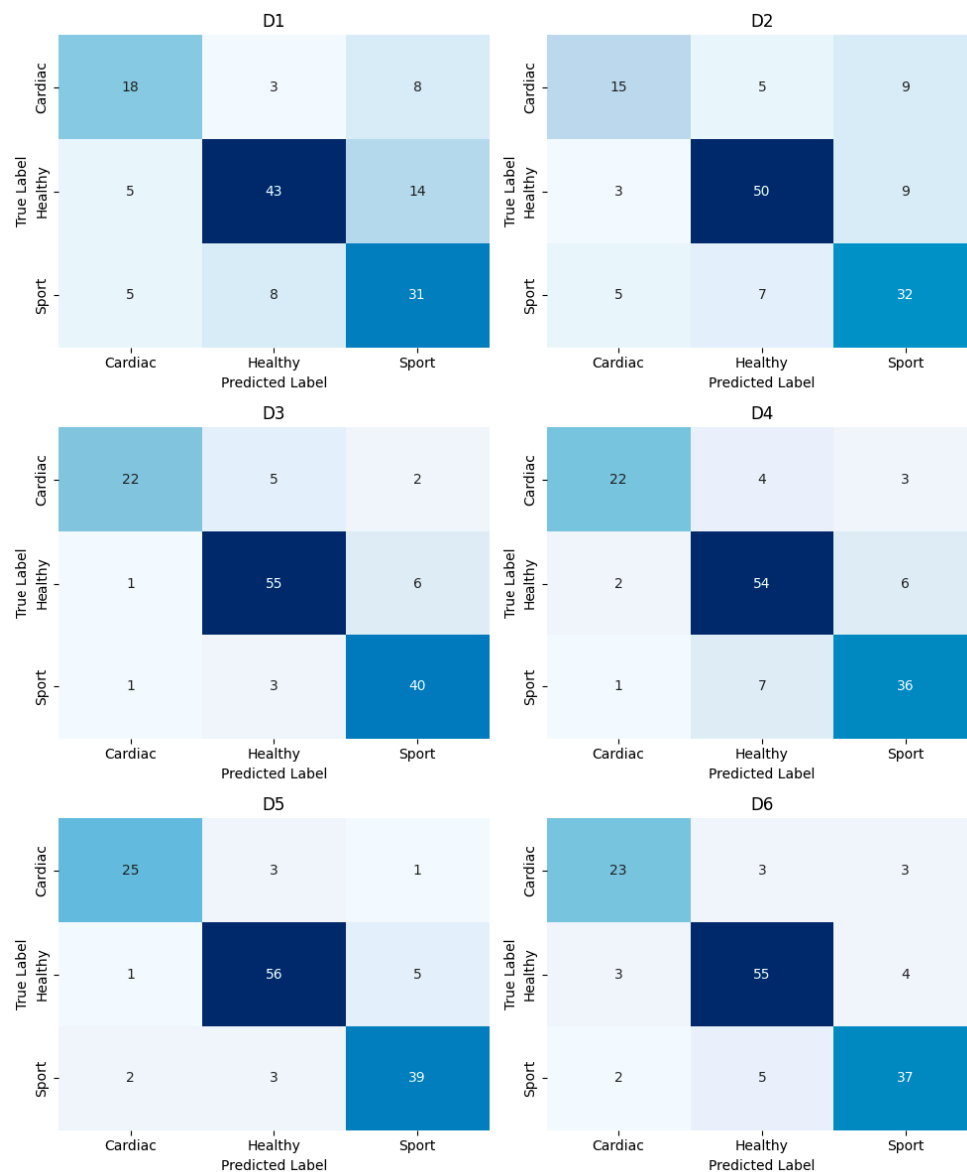
**Figure 6.**  $p$ -values from the Wilcoxon signed-rank test that compared the metrics obtained for individual datasets from individual iterations of 10-fold cross-validation.  $p$ -values smaller than 0.05, indicating statistically significant difference in the metric values, are highlighted with black backgrounds.

The ROC obtained on all predicted values on test sets are presented in Figure 7 for each group based on a one vs. all approach. The cumulative confusion matrices obtained for each dataset after the validation based on the test sets are presented in Figure 8.





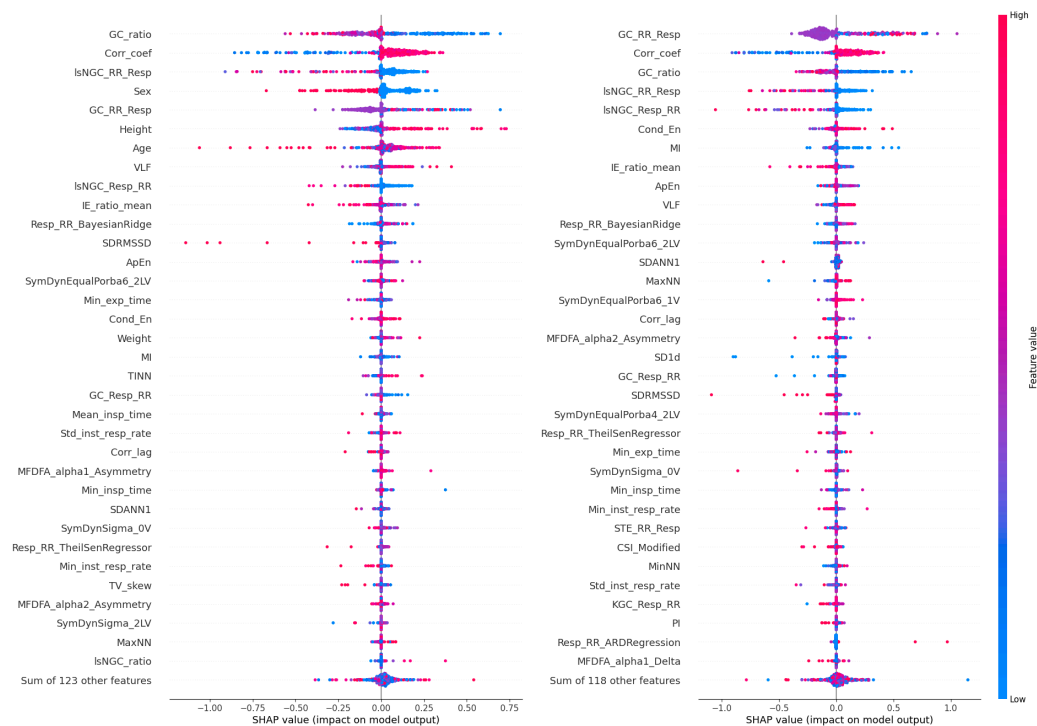
**Figure 7.** ROC and AUC values obtained for each considered dataset. The dashed black line represents the line of identity.



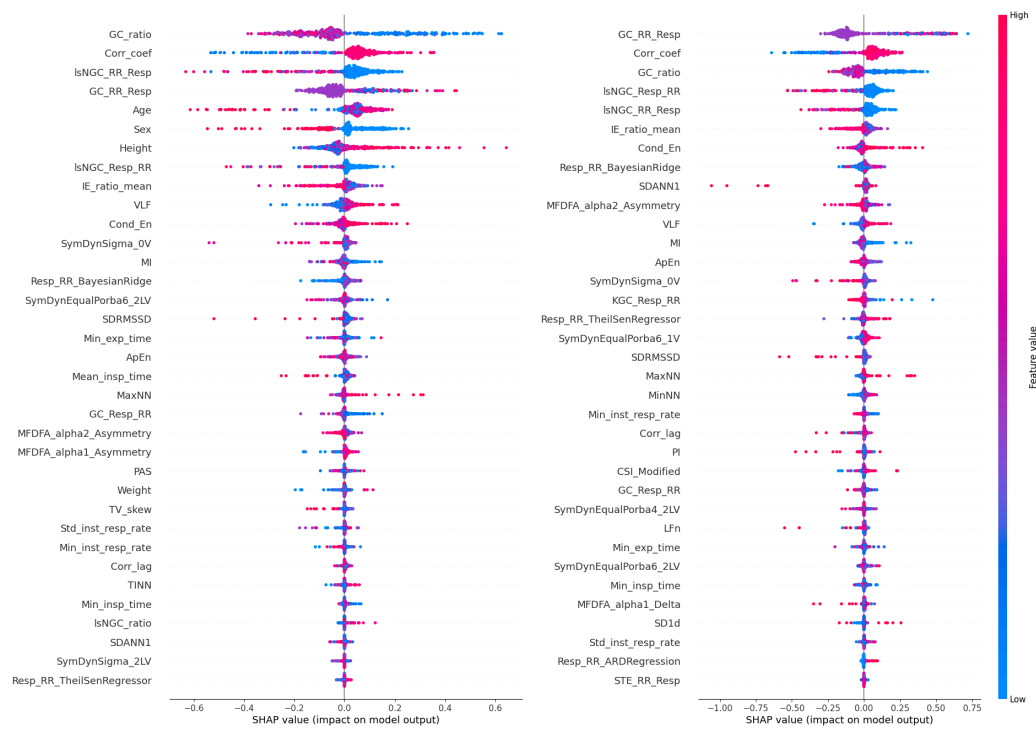
**Figure 8.** Cumulative confusion matrices obtained by summing the confusion matrices from the test set in each iteration of the 10-fold cross-validation for each considered dataset.

The results of the XAI analysis in terms of the Shapley values (presenting the contribution of each feature to the model's predictions for individual samples) for datasets D3 and D4 (which contained all cardiorespiratory features) are presented in Figure 9, while D5 and D6 (which contained the most important features) are presented in Figure 10. Permutation-based variable importance (presenting the overall impact of each feature on

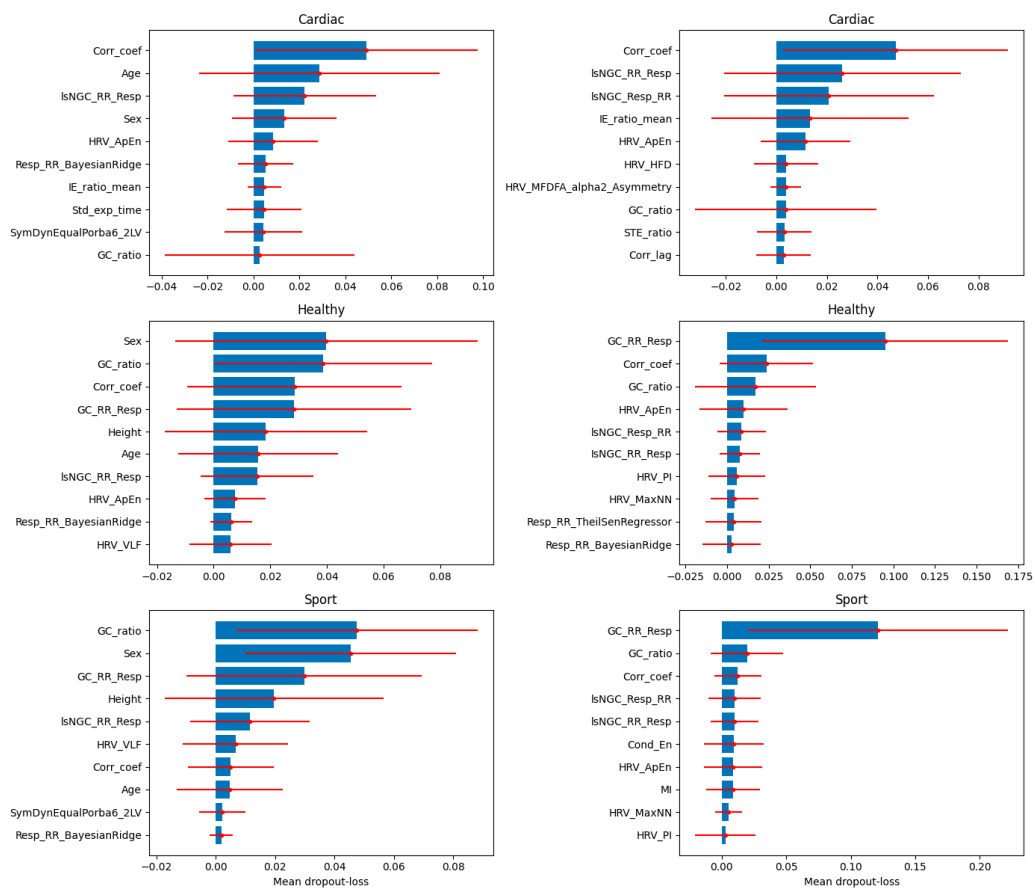
the model's performance) is visualized in Figure 11 for D3 and D4 and in Figure 12 for D5 and D6. For four analyzed datasets, some of the most influential features based on the Shapley values were as follows: the ratio of the GC from the respiratory signal to the tachogram (Resp→RR) by the GC from the tachogram to the respiratory signal (RR→Resp), the highest values of the Pearson correlation coefficient between the respiratory and cardiac signals for a lag between  $-1$  and  $1$  s (CorrCoef), lsNGC RR→Resp, and GC RR→Resp. These features were also indicated as the most influential in the permutation-based variable importance analysis for distinguishing between the individuals from the Healthy and Sport groups (besides CorrCoef for dataset D5). In terms of distinguishing between the Cardiac and other groups, this analysis revealed that the CorrCoef and lsNGC RR→Resp features had the biggest impacts.



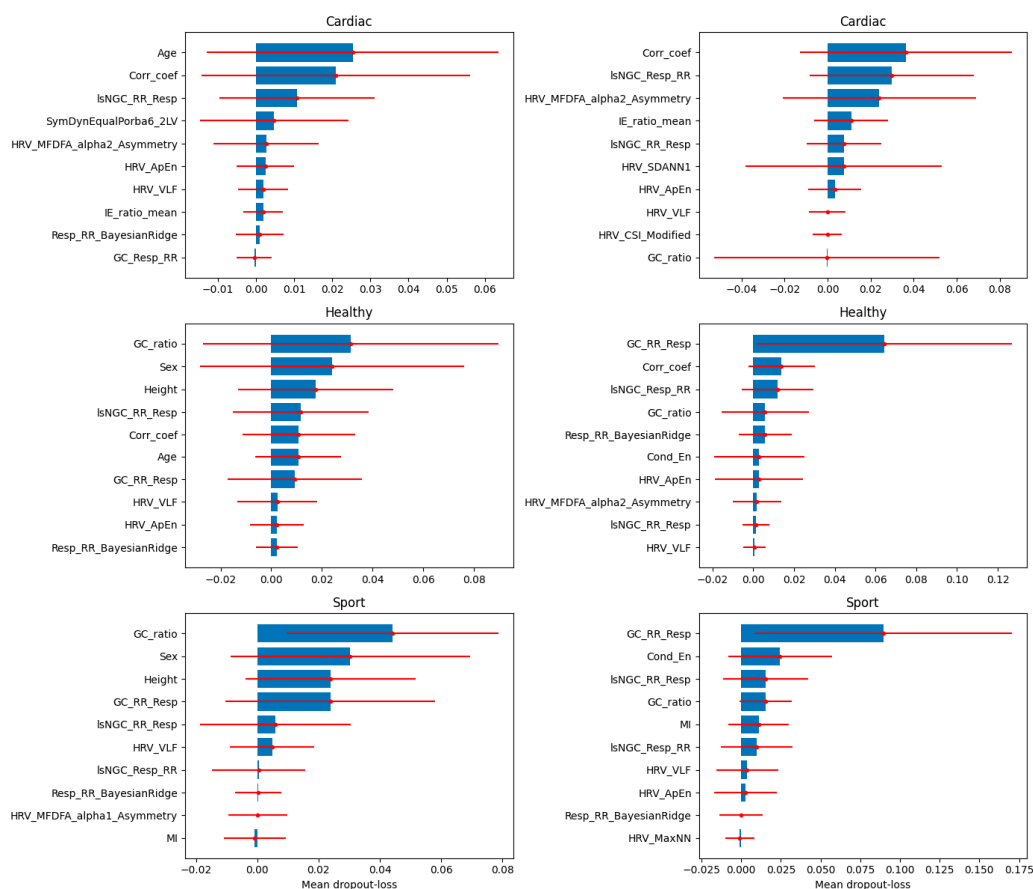
**Figure 9.** Shapley values obtained for the test data from the cross-validation for D3 (on the left) and D4 (on the right). The horizontal axis represents the SHAP value, which reflects the impact of each feature on the model's output. The vertical axis lists the features in order of importance, with the most influential features at the top. The color of each dot represents the feature value for each data point: red dots correspond to high feature values, while blue dots correspond to low feature values.



**Figure 10.** Shapley values obtained for the test data from the cross-validation for D5 (on the left) and D6 (on the right). The horizontal axis represents the SHAP value, which reflects the impact of each feature on the model's output. The vertical axis lists the features in order of importance, with the most influential features at the top. The color of each dot represents the feature value for each data point: red dots correspond to high feature values, while blue dots correspond to low feature values.



**Figure 11.** The mean values of dropout-loss variable importance are presented as bar plots with the standard deviation (red solid lines) for each class separately with a one vs. all approach applied for its calculations. The mean and standard deviation were calculated from the values of variable importance obtained at each iteration of the 10-fold cross-validation. The results for D3 are presented on the left and for D4 on the right.



**Figure 12.** The mean values of the dropout-loss variable importance are presented as bar plots with the standard deviation (red solid lines) for each class separately with a one vs. all approach applied for its calculations. The mean and standard deviation were calculated from the values of variable importance obtained at each iteration of the 10-fold cross-validation. The results for D5 are presented on the left and for D6 on the right.

#### 4. Discussion

We present the classification of young individuals into three distinct groups (Cardiac, Healthy, and Sport) based on cardiorespiratory parameters obtained from 5 min (rest, supine) measurements during spontaneous breathing using ML algorithms. The findings suggest that the integration of diverse cardiorespiratory parameters, including cardiac, respiratory, and causal/information domain features, significantly improved the accuracy and robustness of classification performance. Dataset D5, which incorporated the most influential parameters from all feature types, demonstrated superior performance across various metrics, including accuracy, recall, precision, AUC, MCC, and F1 score, as well as in terms of the shape of the ROC curves. The results obtained for D6 were similar in terms of most metrics, while it did not leverage the demographic information.

The high accuracy and other favorable metrics observed in the D5 dataset highlight the effectiveness of this approach in distinguishing between physiological profiles within

classified groups. Moreover, in the case of misclassification, the Sport subjects were more often labeled as Healthy rather than Cardiac, and the Cardiac patients were more frequently mislabeled as Healthy rather than Sport subjects. This suggests a greater difference between the Cardiac and Sport groups in the feature space, with the Healthy group being somewhere in between, likely closer to the Sport group, as the Healthy subjects were mostly misclassified as Sport individuals. As also suggested in the previous work [55], the inclusion of causal and information domain features significantly improved the predictive models. The imperfect separation of the groups might have been due to changes in the cardiac and respiratory parameters that varied not only with the health status but also with age [56], which made it harder to distinguish the subjects between groups. Additionally, the heterogeneity of health issues in the Cardiac group could also negatively impact the accuracy, as different issues might be characterized by distinct cardiorespiratory profiles.

The observed improvement of classification for datasets containing causal and information features seems to support the initial hypothesis that cardiorespiratory interdependencies provide valuable diagnostic insights. This may be attributed to the additional information about the health status provided by the CRC, particularly the RSA phenomenon in which the change in the heart rate is caused by breathing with shortening of the  $RR_i$  during the inhale and extension during the exhale [57]. Based on the HRV, information about the influence (in the causal sense) of respiration on the cardiac system might be obtained (primarily through frequency domain parameters) [58], although only taking into account the respiratory signal allowed for the full picture of the RSA to be captured. Existing literature seems to support the claim regarding the relevance of information related to CRC, as studies demonstrated that CRC plays an important role in sports medicine [10,59], e.g., allowing for differentiation between athletes and non-athletes [60], as an early marker of cardiac autonomic dysfunction in type 2 diabetes mellitus patients [61] and in research on obstructive sleep apnea [62,63].

The implementation of XAI tools confirmed that the inclusion of causal features was beneficial for the prediction accuracy, as some of them had a meaningful impact on the model output, both in terms of the Shapley values and permutation-based variable importance. Features related to  $RR \rightarrow Resp$  causality had a bigger impact on the model than  $Resp \rightarrow RR$ , which might seem contradictory to the RSA, which may be explained by the fact that the local maxima of the tachogram might occur before the local maxima of the respiratory signal [13,64], as well as physiological bidirectional character of interdependencies between the RR and TV signals [65]. This observation highlights the importance of interpreting causal and information domain features in the context of the underlying data and with respect to the domain knowledge. It is also noteworthy that, although the most influential causal domain features tended to be related to the traditional GC, nonlinear approaches, like lsNGC, were also among the most important parameters, indicating the complexity of the CRC phenomenon. The greater impact of linear features may be attributed to the static measurement conditions without introducing any interventions that could further emphasize the nonlinear relationships. It is also worth mentioning that despite the strong influence of demographic parameters on the model output and their statistical difference between the groups, dataset D6 provided satisfying results that reached over 85% accuracy based solely on features calculated from the cardiorespiratory signals without any information about the subjects' demography. This allowed for the utilization of the method without the need for additional measurements of weight and height or knowledge about the subject's age.

The utilization of ML algorithms with cardiorespiratory data in cardiology, pulmonology, and sports medicine has gained popularity in recent years [52,63–66]. The application of ML algorithms has been found useful in terms of coronary heart disease risk prediction [66], classifying exercise limitation severity [67], identifying integrative cardiopulmonary exercise test (CPET) profiles [68], the prediction of CRF in terms of the peak oxygen consumption [55], and central apnea detection in premature infants [69]. Despite the widespread application of ML in medicine, the integration of CRC-related features

remains underexplored, with only a minority of studies incorporating these features [69]. In this study, we demonstrated that CRC-related features significantly improved the performance of the models, highlighting a gap in the literature and presenting a valuable opportunity for future research to further explore the role of CRC in various clinical and athletic contexts, as well as its impact on predictive modeling performance.

Moreover, the presented results demonstrate the potential of leveraging the ML-assisted evaluation of the health status based on static cardiorespiratory recordings. Such evaluation, which can be widely accessible due to the simplicity of the measurement process; the lack of need for advanced apparatus, like gas analyzers; and the absence of contraindications (as in the case of CPET [70]), is particularly valuable in areas such as pediatric heart transplantation [71], assessment of cardiovascular disease risk in adulthood [72], the monitoring of the cardiac rehabilitation progress [73], the timely identification of pathological conditions prior to sports events [74], and optimizing the training load and avoid overtraining [75]. Health status assessments are especially challenging in the pediatric population due to changes in cardiac and respiratory functions during maturation [76,77]. What is more, the interpretation of multiple cardiac, respiratory, and causal parameters might be challenging for the physician due to their multitude. Therefore, ML tools can simplify the data and provide an output in the form of a new, more interpretable parameter. The improvement in ML performance observed for datasets that contained only the 35 most important features, compared with the corresponding datasets with all cardiorespiratory features, although not statistically significant, highlighted the need for research into identifying the optimal parameter set that would provide the highest diagnostic value.

Models developed in this study, although of the general purpose, could be potentially useful for initial patient screening. Foremost, they could be further personalized and specialized, e.g., based on systematically conducted measurements during training camps or rehabilitation processes with the training/rehabilitation outcome as model targets. After further development for the specific use case, the presented method, integrating various easily accessible cardiorespiratory features and machine learning, would be especially helpful in clinical practice by providing more personalized and precise health assessments. Specifically, it could aid in cardiac rehabilitation by offering a non-invasive, monitoring solution that leverages not only the typically used cardiological parameters (like linear HRV ones), but a broad range of cardiorespiratory features, including nonlinear CRC parameters and machine learning models to track patient progress through the rehabilitation process. The method's ability to classify individuals based on their cardiorespiratory signals could also improve the early detection of potential health issues, enabling timely interventions and more tailored rehabilitation strategies.

Additionally, its application could extend to optimizing training loads in athletes. The ML-assisted parametrization of cardiorespiratory data based on the presented approach would allow coaches and sports physicians to closely monitor athletes' adaptation to training, ensuring they do not exceed their physiological limits and reducing the risk of overtraining or injury. In broader healthcare contexts, this method could be applied to monitor post-operative recovery, where the continuous, non-invasive tracking of cardiorespiratory functions could help detect complications early, such as signs of respiratory distress or cardiovascular instability. However, further studies and model training are needed to optimize the method's predictive power and ensure its accuracy and reliability in those clinical applications.

The limitation of this study was the absence of female subjects in the Sport group, as well as variations in the group sample sizes and demographic parameters, along with the heterogeneity of health issues in the Cardiac group, all of which might have negatively impacted the performance of the ML models. Including patients with arrhythmias could also be seen as a potential limitation. These patients may experience paroxysmal arrhythmias, and the cardiorespiratory parameters measured outside of an arrhythmia episode might not differ significantly from those of healthy subjects. However, the condition itself could indirectly impact the cardiorespiratory profile through lifestyle changes, such as avoiding



physical exercise. A larger sample size with an equal distribution of demographic parameters and increased within-group homogeneity would be beneficial from the perspective of training the machine learning models. Moreover, the fact that subjects in the Sport group only practiced a single sport discipline could also be considered a limitation.

As a result of this study, we not only trained classification models for multiple health conditions that may be useful for initial patient screening but also highlighted the significance of causal and information domain parameters related to CRC and identified a subset of cardiorespiratory features that could be further explored. Our study demonstrated that expanding the most commonly used HRV parameters with respiratory and CRC data could lead to improved subject profiling. These findings have the potential to be leveraged in predictive modeling to monitor parameter trends in individual progress during training or rehabilitation, as well as in the context of CRF and specific cardiac conditions. However, additional research is necessary to further explore these applications.

## 5. Conclusions

This study demonstrated the utilization of ML algorithms with a wide variety of cardiorespiratory features in the classification of pediatric individuals into three groups based on their health statuses while identifying the optimal set of cardiorespiratory features with potential for further use in personalized medical modeling. The results also emphasize the value of including causal and information domain features in the assessment of individuals' health statuses, as these features allowed for significant improvement of the classification accuracy.

**Author Contributions:** Conceptualization, M.R. and M.M.; methodology, M.R.; software, M.R.; validation, M.R., M.M. and J.S.G.; formal analysis, M.R.; investigation, M.R., J.S.G., K.K., J.L., R.M. and B.W.; data curation, M.R. and J.S.G.; writing—original draft preparation, M.R.; writing—review and editing, M.R., J.S.G., M.M., K.K., J.L., B.W. and R.M.; visualization, M.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was founded by the POB Biotechnology and Biomedical Engineering of Warsaw University of Technology within the Excellence Initiative: Research University (IDUB) program.

**Institutional Review Board Statement:** This study was conducted in accordance with the Declaration of Helsinki and approved by two ethics committees (permissions: KB/55/N02/2019, 5 June 2019 and KB/70/2021, 14 June 2021).

**Informed Consent Statement:** Written informed consent forms were obtained from the legal guardians of subjects younger than 16 years old and directly from the subjects themselves if they were 16 years or older.

**Data Availability Statement:** Data and materials used in this study are available upon reasonable request to the corresponding author and under a collaboration agreement.

**Acknowledgments:** Grammarly and GPT-4o were used to check the text's grammatical correctness. GPT-4o was used to prepare Figure 1.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

|      |                               |
|------|-------------------------------|
| AUC  | Area under the curve          |
| BMI  | Body mass index               |
| CRC  | Cardiorespiratory coupling    |
| CRF  | Cardiorespiratory fitness     |
| CPET | Cardiopulmonary exercise test |
| ECG  | Electrocardiography           |
| GC   | Granger causality             |
| HRV  | Heart rate variability        |

|          |                                           |
|----------|-------------------------------------------|
| IP       | Impedance pneumography                    |
| lsNGC    | Large-scale nonlinear Granger causality   |
| MCC      | Mathew's correlation coefficient          |
| ML       | Machine learning                          |
| Resp     | Respiratory signal                        |
| RespRate | Respiratory rate                          |
| ROC      | Receiver operating curve                  |
| RR       | Tachogram time series                     |
| RRi      | RR intervals                              |
| RSA      | Respiratory sinus arrhythmia              |
| SMOTE    | Synthetic minority oversampling technique |
| TV       | Tidal volume                              |
| XAI      | Explainable artificial intelligence       |

## Appendix A

Full list of features used in this study and their descriptions [18–20,47–50,78–81].

### Demography

- Sex
- Age
- Weight
- Height
- BMI: Body mass index

### Cardiac

- MeanNN: The mean of the RR intervals.
- SDNN: The standard deviation of the RR intervals.
- SDANN1: The standard deviation of average RR intervals extracted from 1-min segments of time series data.
- SDNNI1: The mean of the standard deviations of RR intervals extracted from 1-min segments of time series data.
- RMSSD: The square root of the mean of the squared successive differences between adjacent RR intervals.
- SDSD: The standard deviation of the successive differences between RR intervals.
- CVNN: The standard deviation of the RR intervals (SDNN) divided by the mean of the RR intervals (MeanNN).
- CVSD: The root mean square of successive differences (RMSSD) divided by the mean of the RR intervals (MeanNN).
- MedianNN: The median of the RR intervals.
- MadNN: The median absolute deviation of the RR intervals.
- MCVNN: The median absolute deviation of the RR intervals (MadNN) divided by the median of the RR intervals (MedianNN).
- IQRNN: The interquartile range (IQR) of the RR intervals.
- SDRMSSD: SDNN/RMSSD, a time-domain equivalent for the low Frequency-to-High Frequency (LF/HF) Ratio.
- Prc20NN: The 20th percentile of the RR intervals.
- Prc80NN: The 80th percentile of the RR intervals.
- pNN50: The proportion of RR intervals greater than 50 ms, out of the total number of RR intervals.
- pNN20: The proportion of RR intervals greater than 20 ms, out of the total number of RR intervals.
- MinNN: The minimum of the RR intervals.
- MaxNN: The maximum of the RR intervals.
- HTI: The HRV triangular index, measuring the total number of RR intervals divided by the height of the RR intervals histogram.

- TINN: The baseline width of the RR intervals distribution obtained by triangular interpolation.
- VLF: The spectral power of very low frequencies (0.0033 to 0.04 Hz).
- LF: The spectral power of low frequencies (0.04 to 0.15 Hz).
- HF: The spectral power of high frequencies (0.15 to 0.4 Hz).
- VHF: The spectral power of very high frequencies (0.4 to 0.5 Hz).
- TP: The total spectral power.
- LFHF: The ratio obtained by dividing the low frequency power by the high frequency power.
- LFn: The normalized low frequency, obtained by dividing the low frequency power by the total power.
- HFn: The normalized high frequency, obtained by dividing the low frequency power by the total power.
- LnHF: The log transformed HF.
- SD1: Standard deviation perpendicular to the line of identity.
- SD2: Standard deviation along the identity line. Index of long-term HRV changes.
- SD1SD2: ratio of SD1 to SD2.
- S: Area of ellipse described by SD1 and SD2 ( $\pi * SD1 * SD2$ ).
- CSI: The Cardiac Sympathetic Index calculated by dividing the longitudinal variability of the Poincaré plot ( $4*SD2$ ) by its transverse variability ( $4*SD1$ ).
- CVI: The Cardiac Vagal Index equal to the logarithm of the product of longitudinal ( $4*SD2$ ) and transverse variability ( $4*SD1$ ).
- CSI\_Modified: The modified CSI obtained by dividing the square of the longitudinal variability by its transverse variability.
- GI: Guzik's Index.
- SI: Slope Index.
- AI: Area Index.
- PI: Porta's Index.
- SD1d and SD1a: short-term variance of contributions of decelerations (prolongations of RR intervals) and accelerations (shortenings of RR intervals), respectively.
- C1d and C1a: the contributions of heart rate decelerations and accelerations to short-term HRV, respectively.
- SD2d and SD2a: long-term variance of contributions of decelerations (prolongations of RR intervals) and accelerations (shortenings of RR intervals), respectively.
- C2d and C2a: the contributions of heart rate decelerations and accelerations to long-term HRV, respectively.
- SDNNd and SDNNA: total variance of contributions of decelerations (prolongations of RR intervals) and accelerations (shortenings of RR intervals), respectively.
- Cd and Ca: the total contributions of heart rate decelerations and accelerations to HRV.
- PIP: Percentage of inflection points of the RR intervals series.
- IALS: Inverse of the average length of the acceleration/deceleration segments.
- PSS: Percentage of short segments.
- PAS: Percentage of NN intervals in alternation segments.
- DFA\_alpha1: The monofractal detrended fluctuation analysis of the HR signal, corresponding to short-term correlations.
- DFA\_alpha2: The monofractal detrended fluctuation analysis of the HR signal, corresponding to long-term correlations.
- MFDFA\_alpha1\_Width, MFDFA\_alpha1\_Peak, MFDFA\_alpha1\_Mean, MFDFA\_alpha1\_Max, MFDFA\_alpha1\_Delta, MFDFA\_alpha1\_Asymmetry, MFDFA\_alpha1\_Fluctuation, MFDFA\_alpha1\_Increment, MFDFA\_alpha2\_Width, MFDFA\_alpha2\_Peak, MFDFA\_alpha2\_Mean, MFDFA\_alpha2\_Max, MFDFA\_alpha2\_Delta, MFDFA\_alpha2\_Asymmetry, MFDFA\_alpha2\_Fluctuation, MFDFA\_alpha2\_Increment: Indices related to the Multifractal Detrended Fluctuation Analysis.
- ApEn: Approximate entropy.

- SampEn: Sample entropy.
- ShanEn: Shannon entropy.
- FuzzyEn: Fuzzy entropy.
- MSeN: Multiscale entropy.
- CMSeN: Composite Multiscale entropy.
- RCMSeN: Refined Composite Multiscale entropy.
- CD: Correlation Dimension.
- HFD: Higuchi's Fractal Dimension.
- KFD: Katz's Fractal Dimension.
- LZC: Lempel-Ziv Complexity.
- SymDynMaxMin\_0V: Percentage of words in the Max-min method that fall into the 0V family, representing sequences where all three consecutive symbols are equal. This method uses six levels of uniform quantization.
- SymDynMaxMin\_1V: Percentage of words in the Max-min method that fall into the 1V family, which includes sequences with only one variation among three consecutive symbols.
- SymDynMaxMin\_2LV: Percentage of words in the Max-min method that fall into the 2LV family, representing sequences with two variations in the same direction, forming an increasing or decreasing sequence.
- SymDynMaxMin\_2UV: Percentage of words in the Max-min method that fall into the 2UV family, where symbols vary two times in opposite directions, forming a peak or a valley.
- SymDynSigma\_0V: Percentage of words in the  $\sigma$  method that fall into the 0V family. The  $\sigma$  method uses three levels defined by the signal average and its variations shifted up and down by a set factor.
- SymDynSigma\_1V: Percentage of words in the  $\sigma$  method that fall into the 1V family.
- SymDynSigma\_2LV: Percentage of words in the  $\sigma$  method that fall into the 2LV family.
- SymDynSigma\_2UV: Percentage of words in the  $\sigma$  method that fall into the 2UV family.
- SymDynEqualPorba4\_0V: Percentage of words using the Equal-probability method with four quantization levels ( $q = 4$ ) that fall into the 0V family.
- SymDynEqualPorba4\_1V: Percentage of words using the Equal-probability method with four quantization levels that fall into the 1V family.
- SymDynEqualPorba4\_2LV: Percentage of words using the Equal-probability method with four quantization levels that fall into the 2LV family.
- SymDynEqualPorba4\_2UV: Percentage of words using the Equal-probability method with four quantization levels that fall into the 2UV family.
- SymDynEqualPorba6\_0V: Percentage of words using the Equal-probability method with six quantization levels ( $q = 6$ ) that fall into the 0V family.
- SymDynEqualPorba6\_1V: Percentage of words using the Equal-probability method with six quantization levels that fall into the 1V family.
- SymDynEqualPorba6\_2LV: Percentage of words using the Equal-probability method with six quantization levels that fall into the 2LV family.
- SymDynEqualPorba6\_2UV: Percentage of words using the Equal-probability method with six quantization levels that fall into the 2UV family.

#### Respiratory

- RespRate: respiratory rate.
- Std\_inst\_resp\_rate: Standard deviation of instantaneous respiratory rate.
- Min\_inst\_resp\_rate: minimal value of instantaneous respiratory rate.
- Max\_inst\_resp\_rate: maximal value of instantaneous respiratory rate.
- Mean\_insp\_time: mean inspiration time.
- Min\_insp\_time: minimal inspiration time.
- Max\_insp\_time: maximal inspiration time.
- Std\_insp\_time: standard deviation of inspiration time.
- Mean\_exp\_time: mean expiration time.

- Min\_exp\_time: minimal expiration time.
- Max\_exp\_time: maximal expiration time.
- Std\_exp\_time: standard deviation of expiration time.
- TV\_std: standard deviation of tidal volume normalized by median tidal volume.
- TV\_q25: 25th quantile of tidal volume normalized by median tidal volume.
- TV\_q75: 75th quantile of tidal volume normalized by median tidal volume.
- TV\_skew: skewness of tidal volume normalized by median tidal volume.
- TV\_kurtosis: kurtosis of tidal volume normalized by median tidal volume.
- IE\_ratio\_mean: mean inspiration/expiration ratio.

#### Causal/Information

- GC\_RR\_Resp: Granger causality from tachogram to respiratory signal.
- GC\_Resp\_RR: Granger causality from respiratory signal to tachogram.
- STE\_RR\_Resp: Symbolic transfer entropy from tachogram to respiratory signal.
- STE\_Resp\_RR: Symbolic transfer entropy from respiratory signal to tachogram.
- Resp\_RR\_SVR: Granger causality from respiratory signal to tachogram calculated using Support Vector Regression (SVR).
- RR\_Resp\_SVR: Granger causality from tachogram to respiratory signal calculated using Support Vector Regression (SVR).
- Resp\_RR\_BayesianRidge: Granger causality from respiratory signal to tachogram calculated using Bayesian Ridge Regression.
- KGC\_Resp\_RR: Granger causality from respiratory signal to tachogram calculated using Kernel Granger Causality (KGC).
- KGC\_RR\_Resp: Granger causality from Tachogram to respiratory signal calculated using Kernel Granger Causality (KGC).
- RR\_Resp\_GradientBoostingRegressor: Granger causality from tachogram to respiratory signal calculated using Gradient Boosting Regressor.
- Resp\_RR\_GradientBoostingRegressor: Granger causality from respiratory signal to tachogram calculated using Gradient Boosting Regressor.
- RR\_Resp\_TheilSenRegressor: Granger causality from tachogram to respiratory signal calculated using Theil-Sen Regressor.
- Resp\_RR\_TheilSenRegressor: Granger causality from respiratory signal to tachogram calculated using Theil-Sen Regressor.
- RR\_Resp\_ARDRegression: Granger causality from tachogram to respiratory signal calculated using Automatic Relevance Determination (ARD) Regression.
- Resp\_RR\_ARDRegression: Granger causality from respiratory signal to tachogram calculated using Automatic Relevance Determination (ARD) Regression.
- RR\_Resp\_RandomForestRegressor: Granger causality from tachogram to respiratory signal calculated using Random Forest Regression.
- Resp\_RR\_RandomForestRegressor: Granger causality from respiratory signal to tachogram calculated using Random Forest Regression.
- lsNGC\_RR\_Resp: Large scale-nonlinear Granger causality from tachogram to respiratory signal.
- lsNGC\_Resp\_RR: Large scale-nonlinear Granger causality from respiratory signal to tachogram.
- Corr\_coef: Highest values of the Pearson correlation coefficient between respiratory and cardiac signals for lag between  $-1$  and  $1$  s.
- Corr\_lag: Value of the lag for which the highest Pearson correlation coefficient was obtained.
- MI: Mutual information.
- AI: Active information.
- Block\_En: Block entropy.
- Cond\_En: Conditional entropy.
- En\_rate: Entropy rate.
- Trans\_En: Transfer entropy

- Perm\_En: Permutation entropy.
- KGC\_ratio: ratio of KGC\_Resp\_RR and KGC\_RR\_Resp.
- GC\_ratio: ratio of GC\_Resp\_RR and GC\_RR\_Resp.
- STE\_ratios: ratio of STE\_Resp\_RR and STE\_RR\_Resp.
- lsNGC\_ratio: ratio of lsNGC\_Resp\_RR and lsNGC\_RR\_Resp.

## References

1. Zeid, S.; Buch, G.; Velmeden, D.; Söhne, J.; Schulz, A.; Schuch, A.; Tröbs, S.-O.; Heidorn, M.W.; Müller, F.; Strauch, K.; et al. Heart rate variability: Reference values and role for clinical profile and mortality in individuals with heart failure. *Clin. Res. Cardiol.* **2023**, *113*, 1317–1330. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Pham, T.; Lau, Z.J.; Chen, S.H.A.; Makowski, D. Heart rate variability in psychology: A review of hrv indices and an analysis tutorial. *Sensors* **2021**, *21*, 3998. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Mol, M.B.A.; Strous, M.T.A.; van Osch, F.H.M.; Vogelaar, F.J.; Barten, D.G.; Farchi, M.; Foudraïne, N.A.; Gidron, Y. Heart-rate-variability (HRV), predicts outcomes in COVID-19. *PLoS ONE* **2021**, *16*, e0258841. [\[CrossRef\]](#)
4. Stepanyan, L.; Lalayan, G. Heart rate variability features and their impact on athletes' sports performance. *J. Phys. Educ. Sport* **2023**, *23*, 2156–2163. [\[CrossRef\]](#)
5. Shah, S.A.; Velardo, C.; Farmer, A.; Tarassenko, L. Exacerbations in chronic obstructive pulmonary disease: Identification and prediction using a digital health system. *J. Med. Internet Res.* **2017**, *19*, e69. [\[CrossRef\]](#) [\[PubMed\]](#)
6. O'donnell, D. Ventilatory limitations in chronic obstructive pulmonary disease. *Med. Sci. Sports Exerc.* **2001**, *33*, S647–S655. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Ginsburg, A.S.; Lenahan, J.L.; Izadnegahdar, R.; Ansermino, J.M. A systematic review of tools to measure respiratory rate in order to identify childhood pneumonia. *Am. J. Respir. Crit. Care Med.* **2018**, *197*, 1116–1127. [\[CrossRef\]](#)
8. Porta, A.; Gelpi, F.; Bari, V.; Cairo, B.; De Maria, B.; Tonon, D.; Rossato, G.; Ranucci, M.; Faes, L. Categorizing the Role of Respiration in Cardiovascular and Cerebrovascular Variability Interactions. *IEEE Trans. Biomed. Eng.* **2022**, *69*, 2065–2076. [\[CrossRef\]](#)
9. Porta, A.; Bassani, T.; Bari, V.; Pinna, G.D.; Maestri, R.; Guzzetti, S. Accounting for respiration is necessary to reliably infer granger causality from cardiovascular variability series. *IEEE Trans. Biomed. Eng.* **2011**, *59*, 832–841. [\[CrossRef\]](#)
10. de Abreu, R.M.; Cairo, B.; Porta, A. On the significance of estimating cardiorespiratory coupling strength in sports medicine. *Front. Netw. Physiol.* **2023**, *2*, 1114733. [\[CrossRef\]](#)
11. Dick, T.E.; Hsieh, Y.H.; Dhingra, R.R.; Baekey, D.M.; Galán, R.F.; Wehrwein, E.; Morris, K.F. Cardiorespiratory coupling: Common rhythms in cardiac, sympathetic, and respiratory activities. *Prog. Brain Res.* **2014**, *209*, 191–205. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Vinik, A.I. The conductor of the autonomic orchestra. *Front. Endocrinol.* **2012**, *3*, 22505. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Myrńczak, M.; Krysztofiak, H. Cardiorespiratory temporal causal links and the differences by sport or lack thereof. *Front. Physiol.* **2019**, *10*, 45. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Rosol, M.; Gasior, J.S.; Walecka, I.; Werner, B.; Cybulski, G.; Mlynczak, M. Causality in cardiorespiratory signals in pediatric cardiac patients. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, UK, 11–15 July 2022. [\[CrossRef\]](#)
15. Schumann, A.; Fleckenstein, B.; Bär, K.-J. Nonlinear causal influences assessed by mutual compression entropy. *Curr. Dir. Biomed. Eng.* **2016**, *2*, 221–224. [\[CrossRef\]](#)
16. Faes, L.; Porta, A.; Nollo, G. Testing frequency-domain causality in multivariate time series. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1897–1906. [\[CrossRef\]](#)
17. Müller, A.; Kraemer, J.F.; Penzel, T.; Bonnemeier, H.; Kurths, J.; Wessel, N. Causality in physiological signals. *Physiol. Meas.* **2016**, *37*, R46–R72. [\[CrossRef\]](#)
18. Rosol, M.; Mlyńczak, M.; Cybulski, G. Granger causality test with nonlinear neural-network-based methods: Python package and simulation study. *Comput. Methods Programs Biomed.* **2022**, *216*, 106669. [\[CrossRef\]](#)
19. Wismüller, A.; Dsouza, A.M.; Vosoughi, M.A.; Abidin, A. Large-scale nonlinear Granger causality for inferring directed dependence from short multivariate time-series data. *Sci. Rep.* **2021**, *11*, 7817. [\[CrossRef\]](#)
20. Marinazzo, D.; Pellicoro, M.; Stramaglia, S. Kernel method for nonlinear granger causality. *Phys. Rev. Lett.* **2008**, *100*, 144103. [\[CrossRef\]](#)
21. Pompe, B.; Blidh, P.; Hoyer, D.; Eiselt, M. Using mutual information to measure coupling in the cardiorespiratory system. *IEEE Eng. Med. Biol. Mag.* **1998**, *17*, 32–39. [\[CrossRef\]](#)
22. Porta, A.; Guzzetti, S.; Montano, N.; Pagani, M.; Somers, V.; Malliani, A.; Baselli, G.; Cerutti, S. Information domain analysis of cardiovascular variability signals: Evaluation of regularity, synchronisation and co-ordination. *Med. Biol. Eng. Comput.* **2000**, *38*, 180–188. [\[CrossRef\]](#)
23. Schulz, S.; Adochiei, F.-C.; Edu, I.-R.; Schroeder, R.; Costin, H.; Bär, K.-J.; Voss, A. Cardiovascular and cardiorespiratory coupling analyses: A review. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2013**, *371*, 20120191. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Mlynczak, M. Temporal orders and causal vector for physiological data analysis. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Montreal, QC, Canada, 20–24 July 2020. [\[CrossRef\]](#)

25. Krittanawong, C.; Virk, H.U.H.; Bangalore, S.; Wang, Z.; Johnson, K.W.; Pinotti, R.; Zhang, H.; Kaplin, S.; Narasimhan, B.; Kitai, T.; et al. Machine learning prediction in cardiovascular diseases: A meta-analysis. *Sci. Rep.* **2020**, *10*, 16057. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Kufel, J.; Bargieł-Lączek, K.; Kocot, S.; Koźlik, M.; Bartnikowska, W.; Janik, M.; Czogalik, L.; Dudek, P.; Magiera, M.; Lis, A.; et al. What Is Machine Learning, Artificial Neural Networks and Deep Learning?—Examples of Practical Applications in Medicine. *Diagnostics* **2023**, *13*, 2582. [\[CrossRef\]](#) [\[PubMed\]](#)
27. MacEachern, S.J.; Forkert, N.D. Machine learning for precision medicine. *Genome* **2021**, *64*, 416–425. [\[CrossRef\]](#)
28. Chinni, B.K.; Manlhiot, C. Emerging Analytical Approaches for Personalized Medicine Using Machine Learning In Pediatric and Congenital Heart Disease. *Can. J. Cardiol.* **2024**, *40*, 1880–1896. [\[CrossRef\]](#)
29. Hughes, A.; Shandhi, M.H.; Master, H.; Dunn, J.; Brittain, E. Wearable Devices in Cardiovascular Medicine. *Circ. Res.* **2023**, *132*, 652–670. [\[CrossRef\]](#)
30. Loh, H.W.; Ooi, C.P.; Seoni, S.; Barua, P.D.; Molinari, F.; Acharya, U.R. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Comput. Methods Programs Biomed.* **2022**, *226*, 107161. [\[CrossRef\]](#)
31. Albahri, A.; Duham, A.M.; Fadhel, M.A.; Alnoor, A.; Baqer, N.S.; Alzubaidi, L.; Albahri, O.; Alamoodi, A.; Bai, J.; Salhi, A.; et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Inf. Fusion* **2023**, *96*, 156–191. [\[CrossRef\]](#)
32. De Cannière, H.; Corradi, F.; Smeets, C.J.P.; Schoutteten, M.; Varon, C.; Van Hoof, C.; Van Huffel, S.; Groenendaal, W.; Vandervoort, P. Wearable monitoring and interpretable machine learning can objectively track progression in patients during cardiac rehabilitation. *Sensors* **2020**, *20*, 3601. [\[CrossRef\]](#)
33. Nazaret, A.; Tonekaboni, S.; Darnell, G.; Ren, S.Y.; Sapiro, G.; Miller, A.C. Modeling personalized heart rate response to exercise and environmental factors with wearables data. *npj Digit. Med.* **2023**, *6*, 207. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Serantoni, C.; Zimatore, G.; Bianchetti, G.; Abeltino, A.; De Spirito, M.; Maulucci, G. Unsupervised Clustering of Heartbeat Dynamics Allows for Real Time and Personalized Improvement in Cardiovascular Fitness. *Sensors* **2022**, *22*, 3974. [\[CrossRef\]](#)
35. Cui, J.; Huang, Z.; Jiaerken, D.; Fan, Y.; Zhao, S.; Zhang, L.; Wu, J. A wearable system for cardiopulmonary assessment and personalized respiratory training. *Futur. Gener. Comput. Syst.* **2020**, *112*, 1131–1140. [\[CrossRef\]](#)
36. Raghuveer, G.; Hartz, J.; Lubans, D.R.; Takken, T.; Wiltz, J.L.; Mietus-Snyder, M.; Perak, A.M.; Baker-Smith, C.; Pietris, N.; Edwards, N.M. Cardiorespiratory Fitness in Youth: An Important Marker of Health: A Scientific Statement From the American Heart Association. *Circulation* **2020**, *142*, E101–E118. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Bermejo-Cantarero, A.; Álvarez-Bueno, C.; Martínez-Vizcaino, V.; Redondo-Tébar, A.; Pozuelo-Carrascosa, D.P.; Sánchez-López, M. Relationship between both cardiorespiratory and muscular fitness and health-related quality of life in children and adolescents: A systematic review and meta-analysis of observational studies. *Health Qual. Life Outcomes* **2021**, *19*, 127. [\[CrossRef\]](#)
38. García-Hermoso, A.; Ramírez-Vélez, R.; García-Alonso, Y.; Alonso-Martínez, A.M.; Izquierdo, M. Association of Cardiorespiratory Fitness Levels during Youth with Health Risk Later in Life: A Systematic Review and Meta-analysis. *JAMA Pediatr.* **2020**, *174*, 952–960. [\[CrossRef\]](#)
39. McKay, A.K.A.; Stellingwerff, T.; Smith, E.S.; Martin, D.T.; Mujika, I.; Goosey-Tolfrey, V.L.; Sheppard, J.; Burke, L.M. Defining Training and Performance Caliber: A Participant Classification Framework. *Int. J. Sports Physiol. Perform.* **2022**, *17*, 317–331. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Araújo, C.G.S.; Scharhag, J. Athlete: A working definition for medical and health sciences research. *Scand. J. Med. Sci. Sports* **2016**, *26*, 4–7. [\[CrossRef\]](#)
41. Młyńczak, M.; Żyliński, M.; Niewiadomski, W.; Cybulski, G. Ambulatory Devices Measuring Cardiorespiratory Activity with Motion. In *BIODEVICES 2017—10th International Conference on Biomedical Electronics and Devices, Proceedings; Part of 10th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2017*; SciTePress: Setúbal, Portugal, 2017. [\[CrossRef\]](#)
42. Gašior, J.S.; Młyńczak, M.; Rosol, M.; Wieniawski, P.; Walecka, I.; Cybulski, G.; Werner, B. Validity of the Pneumonitor for RR intervals acquisition for short-term heart rate variability analysis extended with respiratory data in pediatric cardiac patients. *Kardiologia Polska* **2023**, *81*, 491–500. [\[CrossRef\]](#)
43. Młyńczak, M.; Cybulski, G. Flow parameters derived from impedance pneumography after nonlinear calibration based on neural networks. In *BIOSIGNALS 2017—10th International Conference on Bio-Inspired Systems and Signal Processing, Proceedings; Part of 10th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2017*; SciTePress: Setúbal, Portugal, 2017; Volume 4. [\[CrossRef\]](#)
44. Młyńczak, M.; Cybulski, G. Decomposition of the cardiac and respiratory components from impedance pneumography signals. In *BIOSIGNALS 2017—10th International Conference on Bio-Inspired Systems and Signal Processing, Proceedings; Part of 10th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2017*; SciTePress: Setúbal, Portugal, 2017; Volume 4. [\[CrossRef\]](#)
45. Seppä, V.-P.; Hyttinen, J.; Uitto, M.; Chrapek, W.; Viik, J. Novel electrode configuration for highly linear impedance pneumography. *Biomed. Eng./Biomed. Tech.* **2013**, *58*, 35–38. [\[CrossRef\]](#)
46. Młyńczak, M.; Niewiadomski, W.; Żyliński, M.; Cybulski, G. Assessment of calibration methods on impedance pneumography accuracy. *Biomed. Eng./Biomed. Tech.* **2016**, *61*, 587–593. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Makowski, D.; Pham, T.; Lau, Z.J.; Brammer, J.C.; Lespinasse, F.; Pham, H.; Schölzel, C.; Chen, S.H.A. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behav. Res. Methods* **2021**, *53*, 1689–1696. [\[CrossRef\]](#) [\[PubMed\]](#)

48. Cysarz, D.; Porta, A.; Montano, N.; Leeuwen, P.; Kurths, J.; Wessel, N. Quantifying heart rate dynamics using different approaches of symbolic dynamics. *Eur. Phys. J. Spec. Top.* **2013**, *222*, 487–500. [\[CrossRef\]](#)
49. Granger, C.W.J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **1969**, *37*, 424–438. [\[CrossRef\]](#)
50. Rosol, M. Nonlincausality—PyPI. 2021. Available online: <https://pypi.org/project/nonlincausality/> (accessed on 28 November 2024).
51. Rosol, M. Classification Code. 2024. Available online: [https://github.com/Mrosol/Cardiac\\_Healthy\\_Sport\\_classification](https://github.com/Mrosol/Cardiac_Healthy_Sport_classification) (accessed on 28 November 2024).
52. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [\[CrossRef\]](#)
53. Baniecki, H.; Kretowicz, W.; Piatyszek, P.; Wisniewski, J.; Biecek, P. Dalex: Responsible machine learning with interactive explainability and fairness in python. *J. Mach. Learn. Res.* **2021**, *22*, 1–7.
54. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *2017*, 4768–4777.
55. Rosol, M.; Petelczyc, M.; Gąsior, J.S.; Młyńczak, M. Prediction of peak oxygen consumption using cardiorespiratory parameters from warmup and submaximal stage of treadmill cardiopulmonary exercise test. *PLoS ONE* **2024**, *19*, e0291706. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Gąsior, J.S.; Sacha, J.; Pawłowski, M.; Zieliński, J.; Jeleń, P.J.; Tomik, A.; Książczyk, T.M.; Werner, B.; Dąbrowski, M.J. Normative values for heart rate variability parameters in school-aged children: Simple approach considering differences in average heart rate. *Front. Physiol.* **2018**, *9*, 1495. [\[CrossRef\]](#)
57. Bernston, G.G.; Cacioppo, J.T.; Quigley, K.S. Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology* **1993**, *30*, 183–196. [\[CrossRef\]](#)
58. Shaffer, F.; Ginsberg, J.P. An Overview of Heart Rate Variability Metrics and Norms. *Front. Public Health* **2017**, *5*, 258. [\[CrossRef\]](#)
59. de Abreu, R.M.; Cairo, B.; Rehder-Santos, P.; da Silva, C.D.; Signini, D.F.; Milan-Mattos, J.C.; Sakaguchi, C.A.; Catai, A.M.; Porta, A. Cardiorespiratory coupling is associated with exercise capacity in athletes: A cross-sectional study. *Respir. Physiol. Neurobiol.* **2024**, *320*, 104198. [\[CrossRef\]](#)
60. de Abreu, R.M.; Porta, A.; Rehder-Santos, P.; Cairo, B.; Sakaguchi, C.A.; da Silva, C.D.; Signini, D.F.; Milan-Mattos, J.C.; Catai, A.M. Cardiorespiratory coupling strength in athletes and non-athletes. *Respir. Physiol. Neurobiol.* **2022**, *305*, 103943. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Da Silva, C.D.; Catai, A.M.; de Abreu, R.M.; Signini, D.F.; Galdino, G.A.M.; Lorevice, L.; Santos, L.M.; Mendes, R.G. Cardiorespiratory coupling as an early marker of cardiac autonomic dysfunction in type 2 diabetes mellitus patients. *Respir. Physiol. Neurobiol.* **2023**, *311*, 104042. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Hietakoste, S.; Armañac-Julián, P.; Karhu, T.; Bailón, R.; Sillanmäki, S.; Töyräs, J.; Leppänen, T.; Myllymaa, S.; Kainulainen, S. Acute cardiorespiratory coupling impairment in worsening sleep apnea-related intermittent hypoxemia. *IEEE Trans. Biomed. Eng.* **2023**, *71*, 326–333. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Yoon, H.; Choi, S.H.; Bin Kwon, H.; Kim, S.K.; Hwang, S.H.; Oh, S.M.; Choi, J.-W.; Lee, Y.J.; Jeong, D.-U.; Park, K.S. Sleep-Dependent Directional Coupling of Cardiorespiratory System in Patients with Obstructive Sleep Apnea. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 2847–2854. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Freyschuss, U.; Melcher, A. Sinus Arrhythmia in Man: Influence of Tidal Volume and Oesophageal Pressure. *Scand. J. Clin. Lab. Invest.* **1975**, *35*, 487–496. [\[CrossRef\]](#)
65. Porta, A.; Castiglioni, P.; Di Rienzo, M.; Bassani, T.; Bari, V.; Faes, L.; Nollo, G.; Cividjan, A.; Quintin, L. Cardiovascular control and time domain Granger causality: Insights from selective autonomic blockade. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2013**, *371*, 20120161. [\[CrossRef\]](#)
66. Kim, J.K.; Kang, S. Neural Network-Based Coronary Heart Disease Risk Prediction Using Feature Correlation Analysis. *J. Health Eng.* **2017**, *2017*, 2780501. [\[CrossRef\]](#)
67. Inbar, O.; Inbar, O.; Reuveny, R.; Segel, M.; Greenspan, H.; Scheinowitz, M. A Machine Learning Approach to Classify Exercise Limitation Severity Using Cardiopulmonary Exercise Testing—Development and Validation. *Med. Res. Arch.* **2023**, *11*. [\[CrossRef\]](#)
68. Cauwenberghs, N.; Sente, J.; Van Criekeing, H.; Sabovčik, F.; Ntalianis, E.; Haddad, F.; Claes, J.; Claessen, G.; Budts, W.; Goetschalckx, K.; et al. Integrative Interpretation of Cardiopulmonary Exercise Tests for Cardiovascular Outcome Prediction: A Machine Learning Approach. *Diagnostics* **2023**, *13*, 2051. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Varisco, G.; Peng, Z.; Kommers, D.; Zhan, Z.; Cottaar, W.; Andriessen, P.; Long, X.; van Pul, C. Central apnea detection in premature infants using machine learning. *Comput. Methods Programs Biomed.* **2022**, *226*, 107155. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Levett, D.; Jack, S.; Swart, M.; Carlisle, J.; Wilson, J.; Snowden, C.; Riley, M.; Danjoux, G.; Ward, S.; Older, P.; et al. Perioperative cardiopulmonary exercise testing (CPET): Consensus clinical guidelines on indications, organization, conduct, and physiological interpretation. *Br. J. Anaesth.* **2018**, *120*, 484–500. [\[CrossRef\]](#)
71. Pastore, E.; Turchetta, A.; Attias, L.; Calzolari, A.; Giordano, U.; Squitieri, C.; Parisi, F. Cardiorespiratory functional assessment after pediatric heart transplantation. *Pediatr. Transplant.* **2001**, *5*, 425–429. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Hauser, C.; Lichtenstein, E.; Nebiker, L.; Streese, L.; Köchli, S.; Infanger, D.; Faude, O.; Hanssen, H. Cardiorespiratory fitness and development of childhood cardiovascular risk: The EXAMIN YOUTH follow-up study. *Front. Physiol.* **2023**, *14*, 1243434. [\[CrossRef\]](#)



73. Akamagwuna, U.; Badaly, D. Pediatric Cardiac Rehabilitation: A Review. *Curr. Phys. Med. Rehabil. Rep.* **2019**, *7*, 67–80. [\[CrossRef\]](#)
74. Adami, P.E.; Squeo, M.R.; Quattrini, F.M.; Di Paolo, F.M.; Pisicchio, C.; Di Giacinto, B.; Lemme, E.; Maestrini, V.; Pelliccia, A. Pre-participation health evaluation in adolescent athletes competing at Youth Olympic Games: Proposal for a tailored protocol. *Br. J. Sports Med.* **2019**, *53*, 1111–1116. [\[CrossRef\]](#)
75. Dürking, P.; Hotho, A.; Holmberg, H.-C.; Fuss, F.K.; Sperlich, B. Comparison of non-invasive individual monitoring of the training and health of athletes with commercially available wearable technologies. *Front. Physiol.* **2016**, *7*, 71. [\[CrossRef\]](#)
76. Gąsior, J.S.; Sacha, J.; Jeleń, P.J.; Pawłowski, M.; Werner, B.; Dąbrowski, M.J. Interaction between heart rate variability and heart rate in pediatric population. *Front. Physiol.* **2015**, *6*, 385. [\[CrossRef\]](#)
77. Fleming, S.; Thompson, M.; Stevens, R.; Heneghan, C.; Plüddemann, A.; Maconochie, I.; Tarassenko, L.; Mant, D. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies. *Lancet* **2011**, *377*, 1011–1018. [\[CrossRef\]](#)
78. Gąsior, J.S.; Rosol, M.; Młyńczak, M.; Flatt, A.A.; Hoffmann, B.; Baranowski, R.; Werner, B. Reliability of Symbolic Analysis of Heart Rate Variability and Its Changes During Sympathetic Stimulation in Elite Modern Pentathlon Athletes: A Pilot Study. *Front. Physiol.* **2022**, *13*, 829887. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Cysarz, D.; Edelhauser, F.; Javorka, M.; Montano, N.; Porta, A. On the Relevance of Symbolizing Heart Rate Variability by Means of a Percentile-Based Coarse Graining Approach. *Physiol. Meas.* **2018**, *39*. [\[CrossRef\]](#) [\[PubMed\]](#)
80. Wismüller, A. Large-Scale Nonlinear Granger Causality Code. 2021. Available online: <https://github.com/Large-scale-causality-inference/Large-scale-nonlinear-causality> (accessed on 28 November 2024).
81. PyInform Package. Available online: <https://elife-asu.github.io/PyInform/> (accessed on 28 November 2024).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.