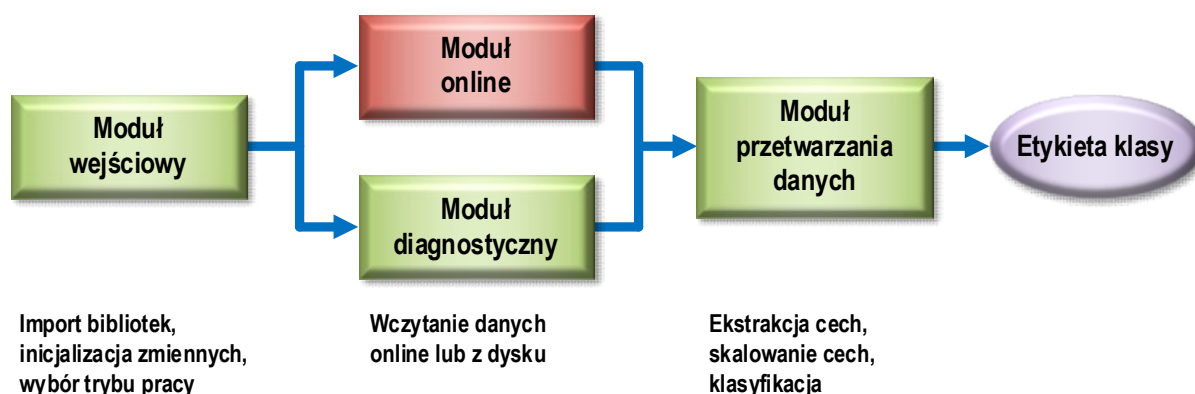


## PREDYKCJA

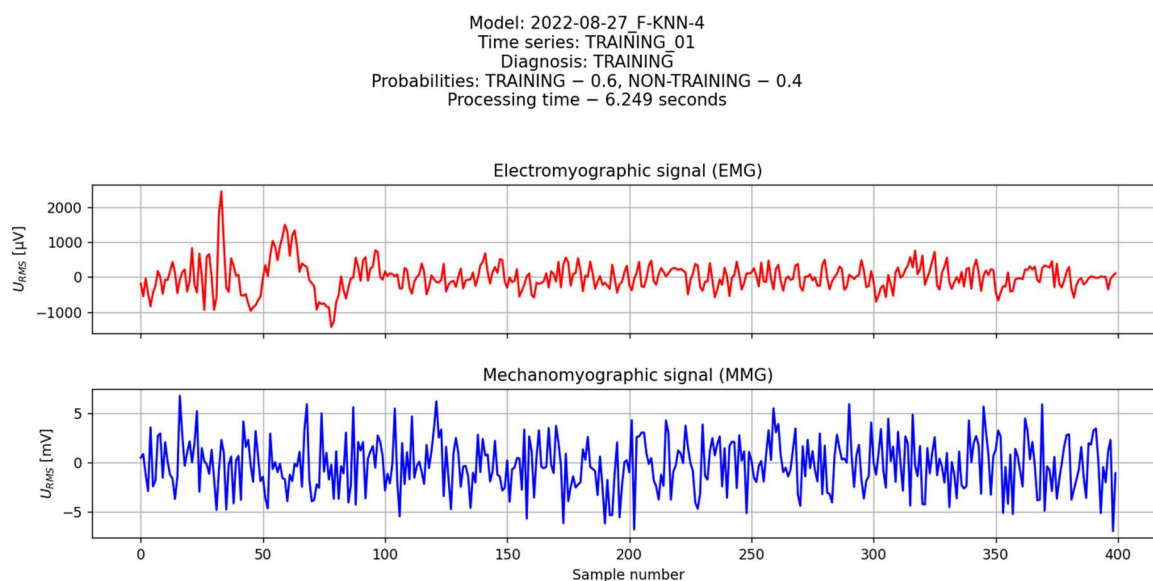
Na rys. 1 przedstawiono ogólną koncepcję systemu do klasyfikacji nowych szeregów czasowych. Na początku systemu znajduje się moduł wejściowy, który odpowiada za import niezbędnych pakietów, zainicjowanie zmiennych globalnych oraz wybór trybu pracy. Zaproponowano 2 tryby pracy systemu: tryb online oraz tryb diagnostyczny (opcjonalny).



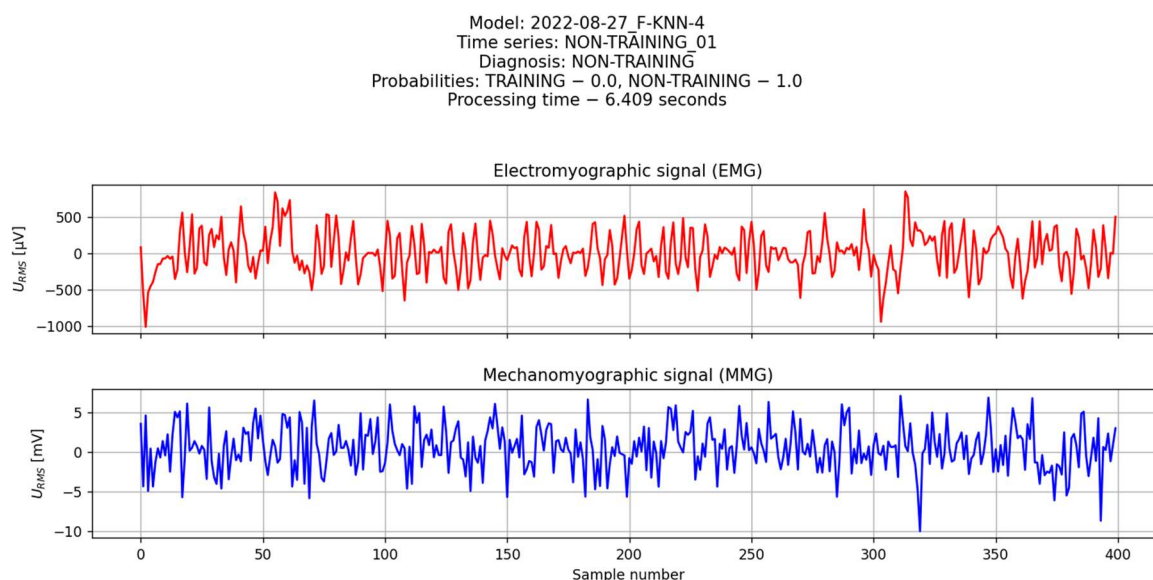
Rys. 1. Schemat blokowy systemu predykcji

Po wybraniu trybu diagnostycznego, z dysku zostają załadowane szeregi czasowe pozwalające sprawdzić poprawność działania modułu przetwarzania danych. Moduł ten jest newralgiczną częścią całego systemu, dlatego jego prawidłowe funkcjonowanie jest kwestią priorytetową. Moduł przetwarzania danych dokonuje ekstrakcji cech na podstawie wejściowych szeregów czasowych, przeskalowuje te cechy w sposób identyczny, jak w przypadku zbioru treningowego oraz dokonuje klasyfikacji sygnału wejściowego do jednej z dwóch klas – TRENUJĄCY lub NIETRENUJĄCY. Przykładowe wyniki uzyskane w trybie diagnostycznym przedstawiono na rys. 2 i 3.

Po wybraniu głównego trybu pracy, czyli trybu online, sygnał z interfejsu radiowego zostaje przekazany na wejście modułu online. Moduł ten odpowiada za wypełnienie danymi okna pomiarowego o szerokości  $W = 400$  próbek i przekazanie tych danych na wejście modułu przetwarzania danych.



Rys. 2. Przykładowy wynik uzyskany w trybie diagnostycznym dla szeregu czasowego, który w rzeczywistości odpowiadał osobie TRENUJĄCEJ



Rys. 3. Przykładowy wynik uzyskany w trybie diagnostycznym dla szeregu czasowego, który w rzeczywistości odpowiadał osobie NIETRENUJĄCEJ

Algorytm pracy systemu predykcji przedstawiono na rys. 6. Po uruchomieniu aplikacji i zakończeniu procesów inicjalizacji, element decyzyjny modułu wejściowego sprawdza warunek uruchomienia trybu online. Jeżeli warunek ten jest spełniony, sterowanie zostaje przekazane do wejściowego elementu decyzyjnego modułu online. Jeżeli nie jest spełniony warunek zakończenia pracy, uruchomiona zostaje pętla akwizycji danych. Po zapisaniu nowej próbki w buforze, następuje sprawdzenie, czy zostało wypełnione okno pomiarowe o szerokości  $W = 400$  próbek. Jeżeli nie, w pętli są rejestrowane kolejne próbki danych

z interfejsu radiowego. Po wypełnieniu okna pomiarowego, następuje sprawdzenie, czy moduł przetwarzania danych zakończył swoją pracę i może przyjąć kolejną porcję danych. Jeżeli moduł ten jest jeszcze zajęty, moduł online oczekuje 1 sekundę, po czym ponownie sprawdza dostępność modułu przetwarzania danych. Po pozytywnej weryfikacji tego warunku, następuje skopiowanie zawartości bufora pomiarowego do ramki danych pandas znajdującej się na wejściu modułu przetwarzania danych. Następnie, bufor zostaje wyczyszczony i wprowadzone zostaje opóźnienie  $T$  w pracy modułu online. Po jego zakończeniu, następuje wznowienie działania pętli akwizycji danych. Wprowadzenie opóźnienia  $T$  daje możliwość dopasowania częstości aktualizacji informacji o etykiecie klasy (diagnozy) do realnych potrzeb i oczekiwań użytkownika systemu. W module online urządzenia docelowego może pojawić się potrzeba wprowadzenia dodatkowego procesu, odpowiedzialnego za realizację operacji związanych z akwizycją danych. Będzie to konieczne, jeżeli 2 sygnały rejestrowane naprzemiennie będą przesyłane w kanale komunikacyjnym z wykorzystaniem jednego strumienia danych. Wówczas, po odbiorze strumienia będzie potrzebna jego rozdzielania na 2 oddzielne sygnały. Ta operacja może wnieść dodatkowe opóźnienie w pracy modułu online.

Dane przekazane na wejście modułu przetwarzania danych mają postać ramki danych pandas o kształcie (400, 2) (rys. 4).

	EMG	MMG
0	-185.798699	0.459992
1	-548.817509	0.839228
2	-38.232591	-1.218936
3	-467.113595	-2.926587
4	-832.508168	3.534091
...	...	...
395	3.143432	-2.037885
396	-352.509815	1.096999
397	-46.910026	2.278163
398	41.675549	-7.007799
399	108.241371	-1.104590

[400 rows x 2 columns]

Rys. 4. Budowa i przykładowa zawartość wejściowej ramki danych

Znaczenie kolumn jest następujące:

- **EMG** – kolejne wartości zmierzonego sygnału EMG.
- **MMG** – kolejne wartości zmierzonego sygnału MMG.

Następnie, dane zostają sformatowane zgodnie z wymaganiami pakietu *tsfresh*, odpowiedzialnego za ekstrakcję cech. Podczas formatowania dodane zostają 2 kolumny, w efekcie czego ramka uzyskuje kształt (400, 4), co przedstawiono na rys. 5.

Znaczenie nowych kolumn jest następujące:

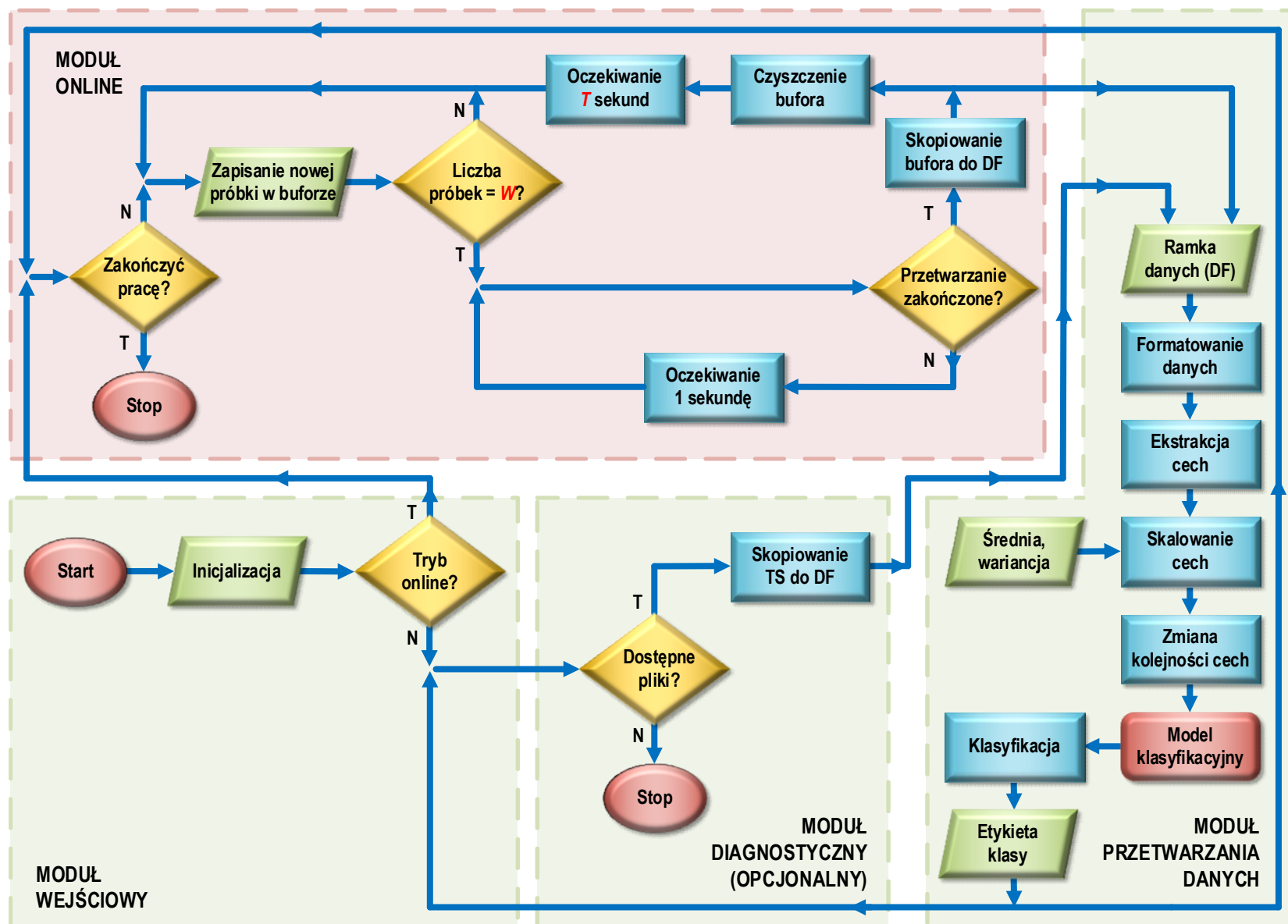
- **id** – identyfikator obserwacji (w przypadku danych online może to być kolejny numer podszeregu wyciętego z ciągłego strumienia danych).
- **time** – kolejny numer próbki w zakresie  $0 \div 399$ .

	id	time	EMG	MMG
0	1	0	-185.798699	0.459992
1	1	1	-548.817509	0.839228
2	1	2	-38.232591	-1.218936
3	1	3	-467.113595	-2.926587
4	1	4	-832.508168	3.534091
...	...	...	...	...
395	1	395	3.143432	-2.037885
396	1	396	-352.509815	1.096999
397	1	397	-46.910026	2.278163
398	1	398	41.675549	-7.007799
399	1	399	108.241371	-1.104590

[400 rows x 4 columns]

Rys. 5. Budowa i przykładowa zawartość ramki danych po formatowaniu

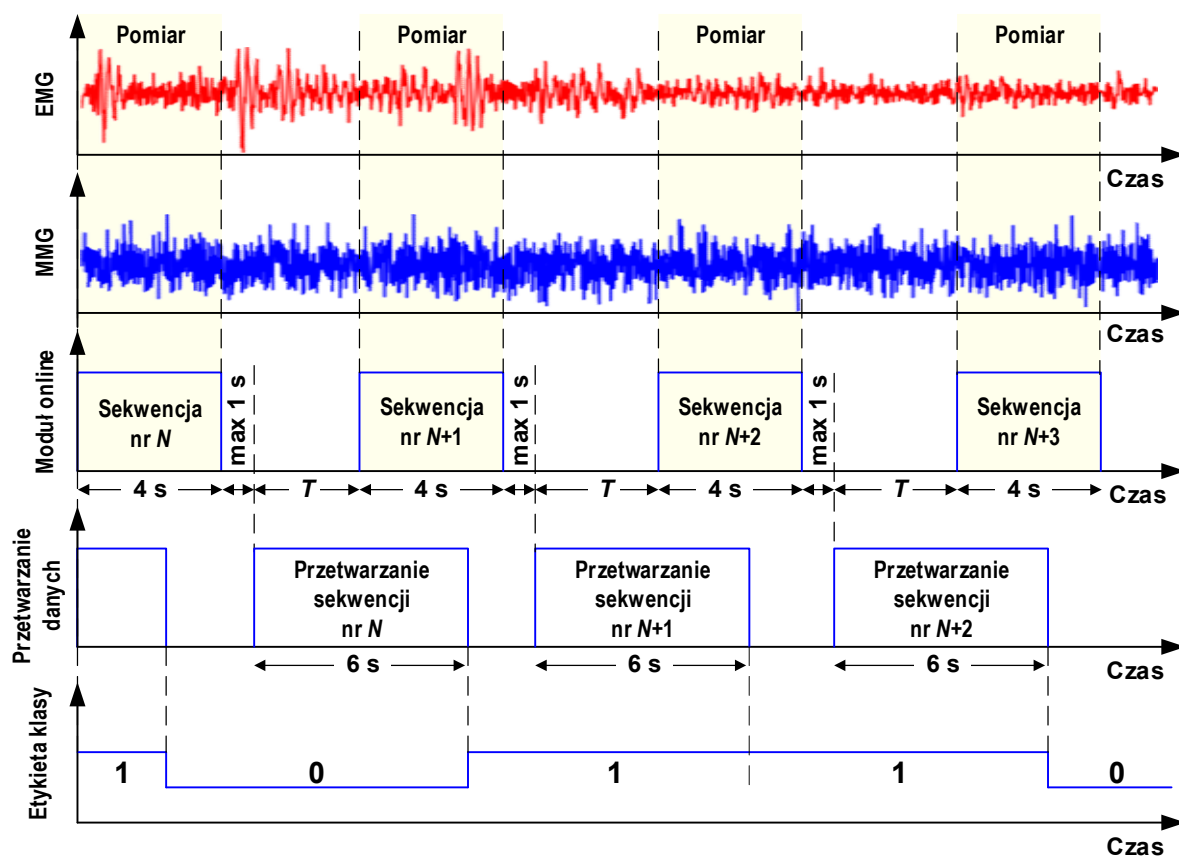
Po ekstrakcji istotnych cech, ich wartości liczbowe zostają przeskalowane w identyczny sposób, jak zostało to uczynione z danymi treningowymi. Przeskalowanie polega na wykonaniu standaryzacji klasycznej z wykorzystaniem plików wartości średnich i wariancji utworzonych na etapie wstępnego przetwarzania danych. Kolejną operacją jest ustawienie cech według rankingu określonego przez metodę selekcji zastosowaną do budowy modelu. Tak przygotowane dane podawane są na wejście modelu klasyfikacyjnego, który generuje etykietę klasy oraz prawdopodobieństwo przynależności analizowanej obserwacji do tej klasy.



**Oznaczenia:** TS – szereg czasowy w postaci pliku tekstowego, DF – ramka danych pandas,  $W$  – szerokość okna pomiarowego (liczba próbek),  $T$  – opóźnienie akwizycji nowego fragmentu strumienia danych.

Rys. 6. Algorytm pracy systemu predykcji

Testowanie pracy algorytmu przeprowadzono na komputerze z procesorem Intel Core i5-3470 3.20 GHz. Podczas tej operacji zmierzono czasy wykonania najważniejszych procesów. Wspomniane czasy mogą się różnić, w zależności od szybkości procesora i wydajności całego systemu komputerowego. Podczas testów, krytyczna okazała się ekstrakcja cech realizowana przez pakiet *tsfresh*. W tym przypadku, czas ekstrakcji cech z jednego szeregu czasowego wyniósł ok. 3 sekundy, co dla dwóch sygnałów (EMG, MMG) dało łączne opóźnienie równe ok. 6 sekund. Opóźnienia wnoszone przez inne operacje przetwarzania danych były pomijalnie małe (pojedyncze milisekundy). Na rys. 7 przedstawiono sekwencję procesów realizowanych przez system predykcji, uwzględniającą zależności czasowe.



Rys. 7. Zależności czasowe w pracy systemu predykcji (procesor Intel Core i5-3470 3.20 GHz, 32 GB RAM.)