

## Wpływ częstotliwości próbkowania na skuteczność systemu NILM z analizą sygnałów w dziedzinie czasu

**Streszczenie.** W artykule przedstawiono metodykę oraz wyniki badań nad wpływem częstotliwości próbkowania na skuteczność systemu NILM. Przeanalizowano zbiór własny zarejestrowany z częstotliwością próbkowania 250 kHz, zawierający zmiany stanu dla 14 urządzeń. Wyznaczono częstotliwość próbkowania powyżej, której nie odnotowuje się poprawy rezultatów identyfikacji dla poszczególnych grup urządzeń.

**Abstract.** This paper presents the methodology and results of a research study on the effect of sampling frequency on NILM system performance. Own dataset recorded at a sampling rate of 250 kHz, containing state changes for 14 appliances, was analyzed. A sampling rate above which there is no improvement in identification accuracy for particular appliance groups was determined. (Impact of sampling rate on the performance of NILM system with signal analysis in time domain).

**Słowa kluczowe:** NILM, rozkład zużycia energii elektrycznej, sygnatury urządzeń, identyfikacja

**Keywords:** NILM, electricity consumption disaggregation, appliance signature, identification

### Wstęp

Zwiększenie świadomości użytkowników o rozkładzie kosztów zużycia energii przez poszczególne Odbiorniki Energii Elektrycznej (OEE) wpływa istotnie na jej oszczędzanie [1]. Rozkład ten może być uzyskany z systemów NILM (ang. Non-Intrusive Load Monitoring), których celem jest określenie stanu pracy OEE w przedziałach czasu oraz wyliczenie kosztów ich pracy [2,3,4,5]. Systemy NILM jako dane wejściowe przyjmują próbki lub wartości parametrów sygnałów napięć i prądów sieci nn (niskiego napięcia). Z nich pozyskiwane są cechy sygnału umożliwiające reprezentację stanu OEE za pomocą wektora wartości (najczęściej liczb rzeczywistych). Ich analiza oraz następujący proces decyzyjny (określenie stanu OEE) realizowane są często przy pomocy metod uczenia maszynowego. Wyższa częstotliwość próbkowania  $f_s$  umożliwia określenie większej liczby cech z sygnału lub uzyskanie bardziej precyzyjnych wartości. Pozwala to metodom MF (ang. Medium Frequency,  $f_s$  powyżej 50 Hz) uzyskiwać lepsze rezultaty od konfiguracji LF (ang. Low Frequency,  $f_s$  poniżej 50 Hz). Jednocześnie konieczność stałej akwizycji próbek z  $f_s$  typową dla metod MF i bieżącego przetwarzania zarejestrowanych próbek sprawia, że system taki jest trudniejszy i droższy w implementacji. Koszt i skuteczność systemów NILM ma kluczowe znaczenie dla upowszechnienia takich systemów. Dostępne zbiory danych dedykowane do badań charakterystycznych dla metod NILM z grupy MF zawierają dane pomiarowe rejestrowane z  $f_s$  od 4 kHz do 250 kHz [6,7,8,9]. Dla niektórych podejść, takich jak analiza zakłóceń EMI [10] czy identyfikacja stanów przejściowych [11] korzystne są dane uzyskane przy jeszcze wyższych  $f_s$ . W literaturze nie weryfikowano wpływu częstotliwości próbkowania sygnału na skuteczność proponowanych metod. Wykorzystując autorski zbiór danych zbadano wpływ  $f_s$  na skuteczność metody NILM wykorzystującej analizę sygnału w dziedzinie czasu [12].

### Dane pomiarowe, metoda identyfikacji oraz metodyka badań

Analizowane dane pomiarowe zostały zebrane z wykorzystaniem układu pomiarowego przedstawionego wcześniej w [2] oraz poddane wstępnej obróbce zgodnie z procedurą zastosowaną w [12]. Częstotliwość próbkowania  $f_s$  wynosiła w tym przypadku 250 kHz. Zgodnie z przyjętymi

załoženiami próbki sygnału prądu  $i(n)$  są przekształcane do macierzy  $I$  o wymiarach  $M \times K$ , gdzie pojedynczy element macierzy oznaczany jest jako  $i_{m,k}$ . Poszczególne wiersze tej macierzy odpowiadają wartościom chwilowym sygnału prądu, które wystąpiły w takim samym odstępie czasu względem momentu  $t_{ok}$ . Czas  $t_{ok}$  określa moment przejścia składowej podstawowej sygnału napięcia  $u(n)$  z wartości ujemnych do dodatnich w okresach  $k = 1, 2, \dots, K-1, K$ . Indeks  $m$  określa numer próbki w pojedynczym okresie  $k$  liczącym  $M$  próbek. Oznacza to, że dla wykorzystywanej  $f_s = 250$  kHz wartość  $M$  wynosi 5000. Zapis macierzowy sygnału umożliwia zastosowanie filtracji medianowej rzędu  $n$ . Polega ona na wyznaczeniu mediany ruchomej dla okna o szerokości  $n$  punktów na wartościach próbek wzdłuż wierszy macierzy  $I$ . Filtracja pozwala na istotne uwypuklenie tych składowych sygnału, które są charakterystyczne dla poszczególnych OEE. Przekłada się to wprost na jakość uzyskanych sygnatur dla poszczególnych OEE.

Opracowany zbiór danych pomiarowych zawiera 2252 zdarzenia, czyli zmiany stanów różnych OEE. Wśród badanych odbiorników znalazły się typowe urządzenia wykorzystywane w gospodarstwie domowym (tab. 1). Badano 14 typowych OEE, które mogą być w danej chwili w stanie włączenia (ON) albo wyłączenia (OFF). W zbiorze są zarejestrowane zdarzenia należące do 28 kategorii (14 odbiorników  $\times$  2 stany). Identyfikator kategorii zdarzenia określany jest dalej jako  $(x,y)$ , gdzie  $x$  jest numerem OEE, zaś  $y \in \{ON, OFF\}$  określa stan urządzenia. Urządzenie o  $x=14$  w tab. 1 podczas części pomiarów pracowało w tle i nie było identyfikowane. Wektorem cech charakteryzującym poszczególne zdarzenia jest wektor zmian  $\Delta I_k^{(x,y)}$

$$(1) \quad \Delta I_k^{(x,y)} = I_{k+P_{x,y}} - I_{k-P_{x,y}}, \text{ dla } k \in N \wedge k \in \langle P_{x,y} + 1, K - P_{x,y} \rangle,$$

gdzie  $P_{x,y}$  jest wartością charakteryzującą czas trwania zmiany stanu dla klasy  $(x,y)$ , zaś wartość  $k$  odpowiada numerowi okresu, w którym rozpoznawane zdarzenie miało miejsce. Identyfikacja zdarzeń realizowana jest z wykorzystaniem algorytmu  $k$ -najbliższych sąsiadów dla  $k=1$ . Wartość parametru  $k$  jest konsekwencją założeń metody opisanej w [11]. Zakłada ona możliwość dodania nowego OEE do systemu na podstawie jednej obserwacji. Jako zbiór treningowy wykorzystano wzorcowe wektory

cech  $\Delta I_{wz\text{or}}^{(x,y)}$ , po jednym wektorze dla każdej z rozpoznawanych klas. Zostały one wyznaczone eksperymentalnie w układzie, w którym był dołączony jedynie OEE o identyfikatorze  $x$ , dla którego wyznaczano wektor cech  $\Delta I_{wz\text{or}}^{(x,y)}$ .

Tabela 1. Zestawienie identyfikowanych urządzeń

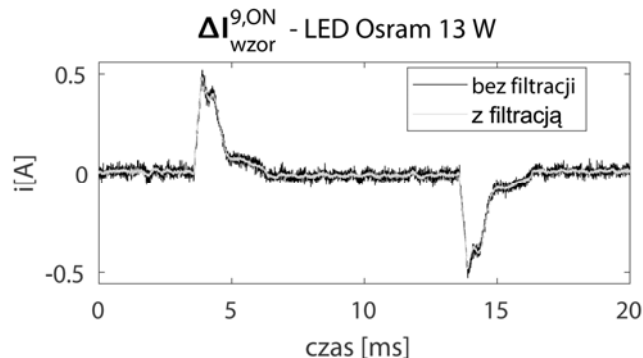
x	nazwa OEE	liczba przykładów	moc [W]
1	żarówka	176	60
2	telewizor LED	220	40
3	4x LED	167	13
4	wyciskarka	185	60
5	LED Philips	146	13
6	wentylator	118	40
7	termowentylator	140	1100
8	odkurzacz	163	1000
9	LED Osram	192	13
10	LED Osram	203	17
11	czajnik	104	1900
12	suszarka	162	120
13	światówka	207	10
14	lodówka	-	-
15	mikrofalówka	153	1400

Zależność skuteczności identyfikacji od częstotliwości próbkowania zbadano poprzez powtarzanie całego procesu identyfikacji na sygnale podpróbkowanym do odpowiedniej  $f_s$ . W identyfikowanych oraz wzorcowych wektorach cech wykorzystano jedynie co  $q$ -tą zamiast wszystkich  $M$  wartości w poszczególnych kolumnach. Przyjęto rząd filtru medianowego  $n=21$ , co odpowiada wyznaczaniu mediany ruchomej za 400 ms. Taka wartość pozwala uwypuklić składowe charakterystyczne dla poszczególnych OEE bez wprowadzania artefaktów spowodowanych zbyt wysokim rzędem filtracji. Wpływ tej filtracji na sygnał pokazano na rys. 1.

Wartości  $q$ , dla których dokonano analizy to 1000, 500, 250, 200, 100, 80, 50, 40, 25, 20, 16, 10, 5, 2, 1. W ten sposób uzyskano zbiory danych z  $f_s$  odpowiednio: 250, 500, 1000, 1250, 2500, 3125, 5000, 6250, 10000, 12500, 15625, 25000, 50000, 125000, 250000 Hz. Taki sposób podpróbkowania wprowadza ryzyko wystąpienia błędu wynikającego z założenia, że przy każdej wartości  $f_s$  momenty  $t_{ok}$  są wyznaczane z taką samą dokładnością jak dla  $f_s = 250$  kHz. W celu uwzględnienia tego efektu wartość indeksu pierwszej próbki  $m_1$  podpróbkowanego wektora zmian wyliczana jest według zależności

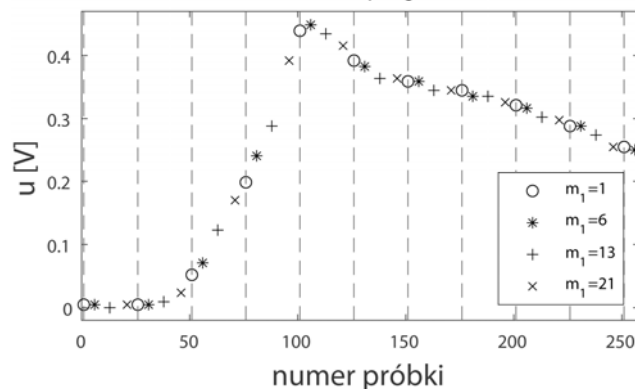
$$(2) \quad m_1 = rand(n), \text{ dla } n \in \langle 1, q \rangle \wedge n \in N$$

gdzie  $rand$  jest generatorem liczb losowych całkowitych o rozkładzie jednostajnym z zadanego zakresu. Wpływ wartości parametru  $m_1$  na przebieg podpróbkowanego sygnału pokazano na rys. 2. Akwizycja sygnału odbywała się dla  $q=25$ , co oznacza że w efekcie uzyskano  $f_s=10$  kHz. W zależności od przyjętej wartości parametru  $m_1$  pozyskiwany sygnał zawiera inne próbki, co jest szczególnie istotne we fragmentach o dużej zmienności wartości sygnału, jak np. obszar od 75 do 125 próbki na rys. 2. Wpływ ten rośnie ze wzrostem wartości parametru  $q$ .

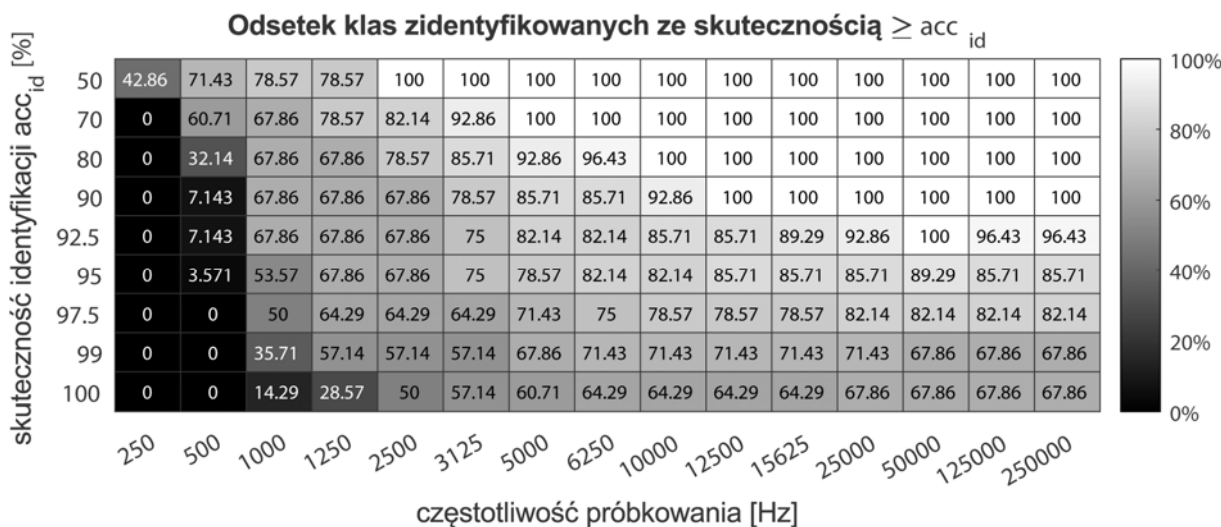


Rys.1. Wpływ filtracji medianowej  $n=21$  na jakość wzorca LED Osram 13 W

Wpływ parametru  $m_1$ ,  $f_s = 250$  kHz,  $q=25$



Rys.2. Wpływ wartości parametru  $m_1$  na przebieg podpróbkowanego sygnału



Rys.3. Wpływ częstotliwości próbkowania  $f_s$  na procent klas zidentyfikowanych ze skutecznością nie mniejszą niż wartość progowa  $acc_{id}$

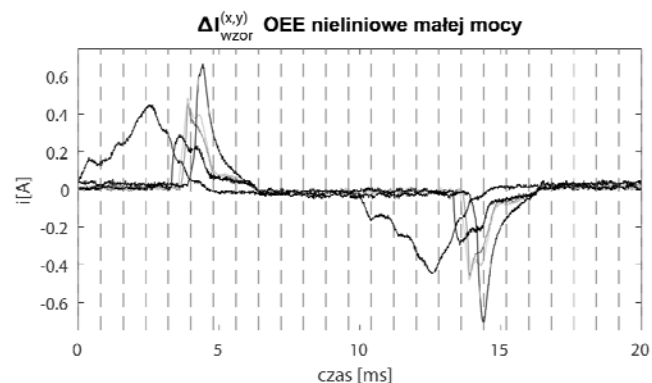
## Wyniki badań

Dla wysokich wartości  $q$  istotny wpływ na uzyskaną skuteczność identyfikacji ma wynik losowań indeksów pierwszych próbek  $m_1$  (zależność (2)), dlatego cały proces został powtórzony 1250 razy dla każdej częstotliwości próbkowania  $f_s$ , a uzyskane wyniki uśredniono. Wpływ  $f_s$  na skuteczność identyfikacji zbadano analizując zmiany odsetka kategorii rozpoznanych z skutecznością nie mniejszą niż skuteczność progowa  $acc_{id}$  w zależności od częstotliwości próbkowania. Jako rozpatrywane wartości progowe  $acc_{id}$  przyjęto skuteczność identyfikacji na poziomie 50, 70, 80, 90, 92.5, 95, 97.5, 99 oraz 100%.

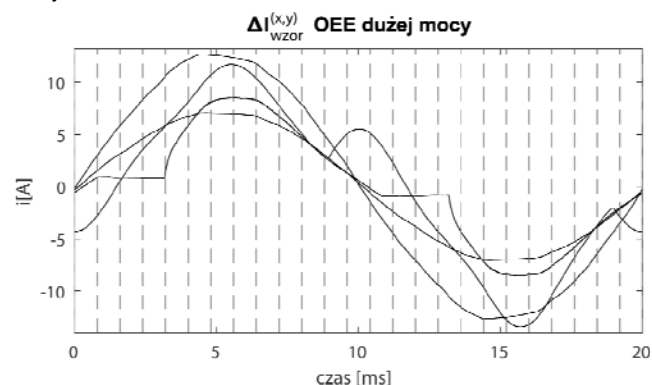
Zgodnie z wynikami przedstawionymi na rys. 3 bardzo dobra ogólna skuteczność identyfikacji jest uzyskiwana przy  $f_s = 12500$  Hz (250 próbek na okres), dla wszystkich kategorii uzyskano skuteczność powyżej 90%. Znaczna część błędów identyfikacji miała miejsce przy rozpoznawaniu urządzeń  $x=5$  oraz  $x=9$ , które mają bardzo podobne sygnatury (rys. 4, odpowiednio linia ciemnoszara dla  $x=5$  oraz jasnoszara dla  $x=9$ ). Poprawę rezultatów różnicznania tych urządzeń można uzyskać podwajając  $f_s$  do 25 kHz, jednak dalsze zwiększanie  $f_s$  nie skutkuje zmianą skuteczności. Bezbłędną identyfikację uzyskano dla 19 z 28 kategorii, zaś skuteczność identyfikacji na poziomie nie mniejszym niż 97.5% uzyskano dla 23 z 28 rozpoznawanych kategorii (82.14%). Warto odnotować, że przy zmianie częstotliwości próbkowania z 1250 Hz (25 próbek na okres) do 3125 Hz (62 próbki na okres) dla klas 5, ON i 5, OFF uzyskano poprawę odpowiednio z 39.07% do 70.57% oraz z 38.89% na 82.12% zaś dla klas 9, ON i 9, OFF poprawę odpowiednio z 38.88% na 76.51% oraz z 39.81 do 79.86%.

Na rys. 4 przedstawiono wzorcowe wektory cech urządzeń nieliniowych małej mocy. Przerywaną pionową linią zaznaczono teoretyczne miejsca próbkowania dla  $f_s = 1250$  Hz. Dla 4 z 5 sygnatur zaledwie 7 z 25 próbek niesie jakąkolwiek informację. Dla porównania, na rys. 5 zestawiono sygnatury urządzeń o bardziej liniowej charakterystyce prądowo-napięciowej. Prawie wszystkie próbki niosą informację umożliwiającą odróżnienie urządzeń. Szczególnie warte uwagi jest zestawienie skuteczności identyfikacji uzyskiwanych dla urządzeń dużej mocy (rys. 6) ze skutecznościami uzyskiwanymi dla nieliniowych urządzeń małej mocy (rys. 7) przy poszczególnych częstotliwościach próbkowania. Dla urządzeń dużej mocy prawie wszystkie próbki w okresie mają wartości istotnie odległe od zera. Skuteczność identyfikacji tych urządzeń uzyskano na poziomie rzędu 99% dla  $f_s = 1250$  Hz (rys 6). Całkowicie bezbłędna ich identyfikacja była możliwa dla  $f_s = 2500$  Hz. Niski pobór prądu badanych urządzeń nieliniowych małej mocy przekłada się na mniejsze odległości między wektorami klasyfikowanymi algorytmem K-najbliższych sąsiadów. Dodatkowo nieliniowa charakterystyka prądowo-napięciowa takich urządzeń sprawia, że istotna część próbek we wzorcowych wektorach cech jak i w rozpoznawanych wektorach cech ma wartości bliskie zera, czyli nie niesie informacji umożliwiającej identyfikację. Podczas gdy dla urządzeń dużej mocy bezbłędną identyfikację uzyskiwano już dla  $f_s = 2500$  Hz, to dla urządzeń nieliniowych małej mocy przy tej częstotliwości próbkowania bezbłędnie rozpoznawano tylko jedną z dziesięciu kategorii (rys. 7). Nawet przy najwyższej częstotliwości próbkowania bezbłędną identyfikację uzyskano jedynie dla 40% kategorii spośród urządzeń nieliniowych małej mocy. Dla  $f_s = 12.5$  kHz oraz  $f_s = 25$  kHz. uzyskano skuteczność identyfikacji na poziomie 90% dla wszystkich urządzeń nieliniowych małej mocy, zaś podniesienie częstotliwości próbkowania do 25 kHz pozwoliło na przekroczenie skuteczności

identyfikacji 92.5% dla dodatkowych dwóch kategorii (włączenia OEE o  $x=5$  oraz wyłączenia OEE o  $x=9$ ). Dla urządzeń nieliniowych małej mocy, czyli najtrudniejszych w identyfikacji, nie uzyskano zwiększenia skuteczności identyfikacji powyżej częstotliwości próbkowania 25 kHz.



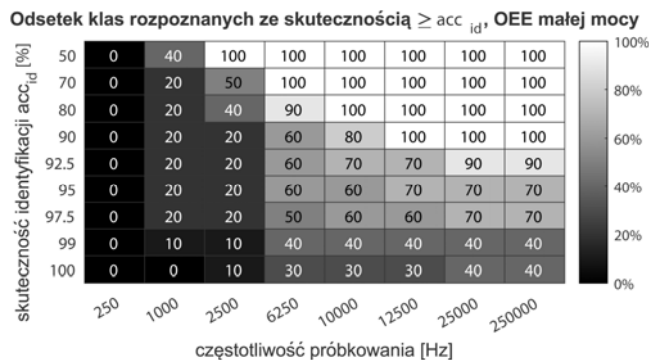
Rys.4. Wzorcowe wektory cech dla nieliniowych urządzeń małych mocy



Rys.5. Wzorcowe wektory cech dla urządzeń dużej mocy



Rys.6. Wpływ częstotliwości próbkowania na skuteczność rozpoznawania OEE dużej mocy



Rys.7. Wpływ częstotliwości próbkowania na skuteczność rozpoznawania nieliniowych OEE małej mocy

## Podsumowanie

Podpróbkując zbiór danych zarejestrowany z częstotliwością próbkowania 250 kHz przeprowadzono analizę porównawczą wpływu  $f_s$  na skuteczność autorskiej metody NILM wykorzystującej analizę sygnałów prądu i napięcia w dziedzinie czasu. Dzięki zastosowaniu losowego wyboru indeksu pierwszej próbki (parametr  $m_1$ ) zapewniono poprawność procesu podpróbkiwania przy typowym nakładzie obliczeniowym (brak konieczności interpolowania wartości próbek). Dla ograniczenia wpływu przypadkowych wartości na wyniki badania proces podpróbkiwania wektorów zmian oraz identyfikacji powtórzono 1250 razy, zatem prezentowane rezultaty są wynikami uśrednionymi.

Stwierdzono istotne różnice we wpływie częstotliwości próbkowania na skuteczność identyfikacji między typami urządzeń. Odbiorniki dużej mocy są rozpoznawane bezbłędnie już przy częstotliwości próbkowania 2500 Hz. Wynika to z faktu, iż prawie wszystkie próbki w okresie 50 Hz (w szczególności dla urządzeń rezystancyjnych) mają wartości istotnie odległe od zera, zaś wysoka wartość elementów klasyfikowanych wektorów jest korzystna w przypadku wykorzystanego klasyfikatora (algorytm K-najbliższych sąsiadów). Należy podkreślić, że metoda utrzymuje wysoką skuteczność nawet jeśli podczas identyfikowanych zmian stanu w tle pracują inne odbiorniki.

Wykazano, że istotnie wyższa częstotliwość próbkowania jest wymagana dla urządzeń małej mocy, w szczególności tych o nieliniowej charakterystyce prądowo-napięciowej. Nawet przy najwyższej częstotliwości próbkowania bezbłędnie rozpoznawano jedynie 4 z 10 kategorii reprezentujących urządzenia nieliniowe małej mocy. To właśnie tego typu są najtrudniejsze do identyfikacji i wyznaczają górne wymagania dla praktycznej implementacji systemu NILM. Dla urządzeń nieliniowych małej mocy najwyższą skuteczność identyfikacji uzyskano dla częstotliwości próbkowania 25 kHz, powyżej tej wartości nie uzyskano poprawy rezultatów.

Konsekwencją nieliniowej charakterystyki prądowo-napięciowej tych urządzeń jest fakt, iż istotna część próbek we wzorcowych wektorach cech jak i w rozpoznawanych wektorach cech ma wartości bliskie zera, czyli nie niesie informacji umożliwiającej rozpoznanie urządzenia. Ponadto mała moc urządzeń przekłada się na mniejsze odległości między wektorami identyfikowanymi algorytmem K-najbliższych sąsiadów. Z tych względów uważamy, że warto w przyszłości powtórzyć opisane eksperymenty z wykorzystaniem bardziej zaawansowanego klasyfikatora, który byłby w stanie nadawać wagę poszczególnych atrybutom i nie byłby tak wrażliwy na zakres wartości klasyfikowanych wektorów.

Rozróżnienie z wysoką skutecznością urządzeń o bardzo zbliżonych sygnaturach jest możliwe już przy częstotliwości próbkowania 12500 Hz, co odpowiada 250 próbkom na okres składowej podstawowej sieci nn. Podwojenie  $f_s$  do 25000 Hz pozwala podnieść skuteczność rozpoznawania tych urządzeń o około 2% i właśnie ta częstotliwość została uznana za optymalną dla zastosowanej metody. Należy zauważyć, że dla 19 z 28 kategorii uzyskiwano 95% skuteczności identyfikacji już przy  $f_s = 1250$  Hz, zaś dla połowy klas bezbłędne rezultaty dla  $f_s = 2500$  Hz, jednak dotyczy to przede wszystkim kategorii urządzeń o liniowej charakterystyce prądowo-napięciowej lub dużej mocy.

Przedstawione rezultaty pozwalają oszacować wymaganą częstotliwość próbkowania dla modułu akwizycji danych w systemie NILM z uwzględnieniem wymagań w zakresie typów urządzeń, które mają być rozpoznawane. Przekłada się ona bezpośrednio na koszt, a więc i powszechność komercyjnego rozwiązania.

Badanie finansowane w ramach projektu POB Energy Politechniki Warszawskiej w ramach programu Inicjatywa Doskonałości: Uczelnia Badawcza (IDUB), umowa 1820/343/Z01/POB7/2021.

**Autorzy:** mgr inż. Krzysztof Dowalla, Politechnika Warszawska, Instytut Radioelektroniki i Technik Multimedialnych, ul. Nowowiejska 15/19 00-665 Warszawa, E-mail: K.Dowalla@ire.pw.edu.pl;  
dr hab. inż. Piotr Biłski, Politechnika Warszawska, Instytut Radioelektroniki i Technik Multimedialnych, ul. Nowowiejska 15/19 00-665 Warszawa, E-mail: P.Bilski@ire.pw.edu.pl;  
dr inż. Robert Łukaszewski, Politechnika Warszawska, Instytut Radioelektroniki i Technik Multimedialnych, ul. Nowowiejska 15/19 00-665 Warszawa, E-mail: R.Lukaszewski@ire.pw.edu.pl;  
dr inż. Augustyn Wójcik, Politechnika Warszawska, Instytut Elektroenergetyki, ul. Koszykowa 75 00-662 Warszawa, E-mail: Augustyn.Wojcik@pw.edu.pl;  
dr hab. inż. Ryszard Kowalik, Politechnika Warszawska, Instytut Elektroenergetyki, ul. Koszykowa 75 00-662 Warszawa, E-mail: Ryszard.Kowalik@pw.edu.pl.

## LITERATURA

- [1] Carrie Armel K., Gupta A., Shrimali G., Albert A., Is disaggregation the holy grail of energy efficiency? The case of electricity, *Energy Policy*, 52 (2013), 213–234, 2013
- [2] Wójcik A., Łukaszewski R., Kowalik R., Winiecki W., Nonintrusive Appliance Load Monitoring: An Overview, Laboratory Test Results and Research Directions, *Sensors*, 19 (2019), no. 16, 3621
- [3] Moreno Jaramillo A. F., Laverty D. M., Morrow D. J., Martinez del Rincon J., Foley A. M., Load modelling and non-intrusive load monitoring to integrate distributed energy resources in low and medium voltage networks, *Renew. Energy*, 179 (2021) 445–466
- [4] Li K., Yin B., Du Z., Sun Y., A Nonintrusive Load Identification Model Based on Time-Frequency Features Fusion, *IEEE Access*, 9 (2021) 1376–1387
- [5] Faustine A., Pereira L., Improved appliance classification in non-intrusive load monitoring using weighted recurrence graph and convolutional neural networks, *Energies*, 13 (2020), No. 13, 1–15
- [6] Iqbal H. K., Malik F. H., Muhammad A., Qureshi M. A., Abbasi M. N., Chishty A. R., A critical review of state-of-the-art non-intrusive load monitoring datasets, *Electric Power Systems Research*, 192 (2020)
- [7] Anderson K., Ocleanu A. F., Benitez D., Carlson D., Rowe A., Bergés M., BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research, *Proc. 2nd KDD Work. Data Min. Appl. Sustain.*, 2012, 1–5
- [8] Medico R. et al., A voltage and current measurement dataset for plug load appliance identification in households, *Sci. Data*, 7 (2020), No. 7, 1–10
- [9] Kahl M., Haq A. U., Kriechbaumer T., Jacobsen H., WHITED - A Worldwide Household and Industry Transient Energy Data Set, *3rd Int. Work. Non-Intrusive Load Monit.*, 2016, no. 4, 1–4
- [10] Dowalla K., Winiecki W., Łukaszewski R., Kowalik R., Identyfikacja odbiorników energii elektrycznej z wykorzystaniem przekształcenia falkowego sygnałów napięcia zasilającego, *Przegląd Elektrotechniczny*, 94 (2018), nr 11, 43–46
- [11] Wójcik A., Winiecki W., Łukaszewski R., Algorytm wyznaczania sygnatury odbiornika energii elektrycznej z wykorzystaniem transformaty falkowej, *Przegląd Elektrotechniczny*, 94 (2018), nr 11, 39–42
- [12] Dowalla K., Biłski P., Łukaszewski R., Wójcik A., Kowalik R., Application of the Time-Domain Signal Analysis for Electrical Appliances Identification in the Non-Intrusive Load Monitoring, *Energies*, 15 (2022), no. 9, 1–20