

Dwie metody sztucznej inteligencji do klasyfikacji rodzajów uszkodzeń przewodów wodociągowych

Two AI methods for classification of water pipes damage

MAŁGORZATA KUTYŁOWSKA, WOJCIECH CIEŻAK

DOI 10.36119/15.2024.2.5

Stan techniczny infrastruktury komunalnej jest nieustannie w centrum zainteresowania z uwagi na konieczność utrzymania ustabilizowanej gospodarki wodnej, która może być zaburzona w sytuacji występowania wielu awarii i uszkodzeń sieci wodociągowej. Nie tylko liczba, ale również rodzaj uszkodzeń jest zmienną, która powinna być przedmiotem analiz technicznych, ekonomicznych i niezawodnościowych. W pracy przedstawiono porównanie wyników klasyfikacji uszkodzeń przewodów wodociągowych z wykorzystaniem dwóch metod opartych o założenia sztucznej inteligencji, czyli metody drzew klasyfikacyjnych i algorytmu sieci neuronowych. Celem pracy było sprawdzenie, czy trafność klasyfikacji w oparciu o perceptron wielowarstwowy będzie większa niż uzyskano to kilka lat temu w autor-skich badaniach podczas analizy metody drzew klasyfikacyjnych. Do modelowania zastosowano dane eksploatacyjne z istniejącego jednego z większych systemów wodociągowych w Polsce. Przeanalizowano kilka wariantów różniących się od siebie rozmiarem wektora wejściowego. Uzyskane wyniki nie są zadowalające, gdyż tylko najbardziej licznie występujące w zbiorze danych uszkodzenia, zostały zaklasyfikowane poprawnie w ok. 80%, a w przypadku innych rodzajów uszkodzeń dokładność klasyfikacji była minimalna. W związku z tym konieczna wydaje się zmiana podejścia w kolejnych etapach pracy nad zagadnieniem przewidywania zmiennych jakościowych.

Słowa kluczowe: uszkodzenia, przewody wodociągowe, sieci neuronowe, drzewa klasyfikacyjne

The technical condition of municipal infrastructure is continually in the centre of attention. It is necessary to maintain water management in stable conditions. Many failures of water pipes could lead to destabilize the whole water management in the cities. Not only the number of damage, but also their kinds should be analysed using technical, economical and reliability analysis. The comparison of classification results of kinds of water pipes damage using two artificial intelligence methods (classification trees and neural networks) was presented in the paper. The aim of the work was to check if classification accuracy using multilayer perceptron is higher than in other original investigations carried out several years ago when classification trees were analysed. Exploitation data from water supply system were used for modelling purposes. Several configurations with different size of input vector were investigated. Obtained results are not satisfactory. Only the most numerous classified in approx. 80%. For other kinds of failures the classification accuracy was minimal. It is necessary to change, in the next work stages, the investigation approach of quality variables prediction.

Keywords: damage, water pipes, neural networks, classification trees

Wprowadzenie

Historia rozwoju wodociągów i kanalizacji pokazuje, że od tysiącleci ludzkość starała się podnieść jakość życia poprzez stosowanie różnych i coraz nowocześniejszych rozwiązań technicznych mających na celu dostawę czystej wody i bezpieczne odprowadzenie ścieków [1]. Współczesne zakłady produkcji wody i oczyszczalnie ścieków są niezbędnymi elementami zapewniającymi prawidłowe funkcjonowanie nie tylko wielkich aglomeracji miejskich, ale także mniejszych jednostek osadniczych, gdzie coraz częściej stosuje się nowatorskie technologie jako narzędzie wspomagające eksploatację [2]. Jednak nie wszystkie zdarzenia eksploatacyjne można przewidzieć, zaplanować

i zamodelować z wieloletnim wyprzedzeniem, np. nagle sytuacje awaryjne, mające charakter zmiennej losowej, są jednym z czynników, które oddziałują na konieczność zmiany podejścia w zarządzaniu systemami wodno-kanalizacyjnymi. Większość eksploataatorów boryka się z problemem nieprzewidywalności, gdyż przystosowywanie infrastruktury całego systemu zaopatrzenia w wodę nie zawsze nadąża za relatywnie szybko zmieniającymi się warunkami zewnętrznymi (np. zmiany klimatu, zmiany w strukturze rozborów wody, zmiany charakteru odbiorców wody i ich liczba [3]) i występującymi awariami. Stąd wynika konieczność zastosowania monitoringu, z wykorzystaniem metod informatycznych, np. wycieków z sieci, a tym samym strat wody [4,

5], które mają negatywny wpływ nie tylko na odbiorców, ale również na środowisko [6], nawet jeśli mamy do czynienia jedynie z awarią na przyłączy domowym [7]. Oczywiście należy pamiętać, że prawidłowa eksploatacja powinna również obejmować obiekty sieciowe, np. hydranty [8], których regularna inspekcja jest konieczna z uwagi na rolę, jaką pełnią w systemie, zwłaszcza obecnie, kiedy mamy do czynienia z często występującymi suszami [9]. Wszystko to dowodzi konieczności zarządzania infrastrukturą podziemną w sposób nowoczesny i inteligentny biorąc pod uwagę specyfikę regionu [10], tworząc mapy ryzyka i modele matematyczne [11] oraz dokonując oceny niezawodności z uwzględnieniem czynnika ludzkiego [12], co jest niezbędne zwłaszcza

dr hab. inż. Małgorzata Kutylowska, prof. uczelni <https://orcid.org/0000-0001-8425-9041>,

dr inż. Wojciech Cieżak e-mail: wojciech.ciezak@pwr.edu.pl, <https://orcid.org/0000-0001-6210-8728> – Wydział Inżynierii Środowiska, Politechnika Wrocławska, Wrocław

Adres do korespondencji/ Corresponding author: malgorzata.kutylowska@pwr.edu.pl

w obecnych czasach, gdy woda staje się dobrem, o które należy dbać z uwagi na szeroko pojęte zmiany klimatyczne i niewielkie zasoby wodne [13, 14]. A jak zaznaczono w pracy [15] większość strat wody jest wynikiem awarii, więc zasadne wydaje się wdrożenie do procesu eksploatacji tak obecnie modnej sztucznej inteligencji [16, 17] w celu wykrywania nieszczelności i uszkodzeń. Również metody matematycznego modelowania umożliwiają klasyfikację rodzajów uszkodzeń przewodów wodociągowych, co zostało przedstawione w monografii [18], w której wspomniano również o konieczności wykonania dalszych badań, a mianowicie porównania metod drzew klasyfikacyjnych z algorytmem sztucznych sieci neuronowych, które jako uniwersalne narzędzie prognostyczne, również z powodzeniem sprawdzają się w zadaniach regresyjnych [19]. Celem niniejszego opracowania jest porównanie wyników jakości i trafności klasyfikacji rodzajów uszkodzeń przewodów wodociągowych z wykorzystaniem wspomnianych wcześniej sieci neuronowych i drzew klasyfikacyjnych. Obie te metody należą do szeroko pojętej sztucznej inteligencji, która obecnie jest coraz częściej stosowana w wielu dziedzinach nauki, co jednak nie powinno w całości zastępować racjonalnych działań człowieka, który z racji doświadczenia i świadomości możliwości wystąpienia nagłych zagrożeń jest niezbędny w sytuacjach konieczności podejmowania kluczowych decyzji w procesie eksploatacji. Sztuczna inteligencja może być narzędziem pomocnym, ale nie jedynym zwłaszcza gdy chodzi o strategiczne obiekty komunalne infrastruktury krytycznej [19].

Metodyka badań

Historyczne dane eksploatacyjne (z ośmiu lat przełomu dwóch pierwszych dekad XXI w.) wybranego systemu wodociągowego posłużyły jako poligon badawczy. Analizowano przewody magistralne, rozdzielcze i przyłącza, które wykonane były z żeliwa szarego, żeliwa sferoidalnego, stali, stali ocynkowanej, azbestocementu, PVC i PE. Na koniec analizowanego okresu długość przewodów magistralnych, rozdzielczych i przyłączy wynosiła odpowiednio 80 km, 560 km i 292 km. Ponad 58% całkowitej długości sieci wodociągowej zostało wykonane w latach 60–90. XX wieku. Jednak z uwagi na rozbudowę miasta, jak i na konieczność wymiany najstarszych odcinków co roku odnotowywano przyrost długości sieci oraz zmianę struktury wiekowej i materiałowej. Przewody o średnicy 300 mm i większej były traktowane jako magistrale. W analizowanym czasie zakres

średnic przewodów rozdzielczych wahał się w granicach 50–250 mm, a przyłącza 20–110 mm. W niektórych rejonach miasta sieć wodociągowa budowana była niedługo po II Wojnie Światowej i przewody o relatywnie małych średnicach, poniżej 100 mm, pełniły rolę rurociągów rozdzielczych.

Klasyfikację rodzajów uszkodzeń przeprowadzono w programie Statistica z wykorzystaniem metody drzew klasyfikacyjnych (CT) i sztucznych sieci neuronowych (SSN). Przed przystąpieniem do modelowania ujednolicono zapis dotyczący nazw rodzaju uszkodzenia, tworząc autorskie skrócone nazwy na podstawie informacji dostępnych w dzienniku awarii, które w wielu miejscach miały charakter opisowy i w takiej formie nie nadawały się do zastosowania w procesie modelowania. Należy jednak wyjaśnić, co rozumiane jest pod pojęciem niektórych rodzajów uszkodzeń. Rozszczelnienie złącza dotyczy uszkodzenia złącza kielichowego. Uszkodzenie uszczelnienia, dalej nazywane po prostu „uszczelnienie”, dotyczyło awarii opisanej w protokołach jako „wypchnięte uszczelnienie”. Uszkodzenie „przeciek” w dziennikach awarii posiadało dokładniejszy opis, np. przeciek na połączeniu, przeciek zasuwy, przeciek na kompensatorze. W związku z koniecznością wprowadzenia jednolitych nazw do programu Statistica tego typu awarie zaklasyfikowano po prostu jako „przeciek”. Pozostałe przypisane nazwy rodzajów uszkodzeń nie wymagają dalszych wyjaśnień. Zakresy zmienności predyktorów i jakościowej zmiennej zależnej (rodzaj uszkodzenia) przedstawiono w tabeli 1.

Tabela 1. Predyktory i zmienna zależna
Table 1. Predictors and dependent variable

Materiał	Przykrycie, m	Ciśnienie, MPa	Rok budowy	Średnica, mm	Typ przewodu	Pora roku	Rodzaj uszkodzenia
stal, stal ocynkowana, azbestocement, żeliwo, PE, PVC	1–5,5	0,18–0,79	1926–2011	20–800	przewód magistralny, przewód rozdzielczy, przyłącze	zima, lato	korozja, pęknięcie, pęknięcie podłużne, pęknięcie poprzeczne, przeciek, rozszczelnienie złącza, uszczelnienie

W metodzie sztucznych sieci neuronowych stworzono cztery konfiguracje macierzy zmiennych niezależnych (A1, A1g, A3, A9). Nie budowano osobnych modeli np. dla przyłączy i przewodów rozdzielczych (innymi słowy typ przewodu był zmienną niezależną i zaliczał się do wektora wejściowego). W przypadku drzew klasyfikacyjnych, omówionych w [18], odpowiedniki dla podanych wyżej konfiguracji są nazywane 1, 1g, 3 oraz 9 i mają takie wektory zmiennych niezależnych, jak w metodzie sieci neuronowych. W związku z tym porównanie wyników klasyfikacji dwoma metodami (CT i SSN) jest jak najbardziej zasadne. Tworzenie różnych konfiguracji miało na celu spraw-

dzenia, czy usunięcie lub dodanie do wektora predyktorów pewnych informacji o sieci wodociągowej ma wpływ na wyniki, jakość i dokładność modelowania. W tabeli 2 przedstawiono, które ze zmiennych niezależnych tworzyły wektor predyktorów w zależności od konfiguracji podczas analizy jakościowej (klasyfikacja rodzajów uszkodzeń).

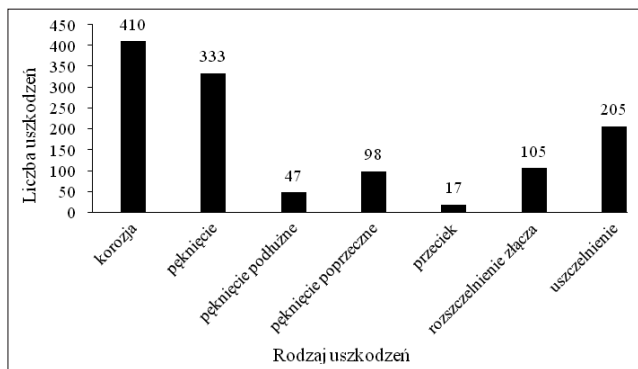
Tabela 2. Zmienne niezależne
Table 2. Independent variables

Zmienna	Konfiguracje			
	A1	A1g	A3	A9
Materiał	+	+	+	+
Średnica	+	+	+	+
Typ przewodu	+	+	+	+
Rok budowy	+	+	+	-
Ciśnienie	+	+	+	+
Przykrycie	+	-	+	+
Pora roku	-	-	+	-

Sposób modelowania za pomocą drzew klasyfikacyjnych został dokładnie opisany w monografii [18], więc w tym opracowaniu przedstawiona zostanie tylko metodyka modelowania z wykorzystaniem wielowarstwowego perceptronu (MLP), gdyż znacząco się ona różni od podejścia CT. Dla każdej konfiguracji i dla danych z ośmiu lat eksploatacji utworzono zbiór uczący i testowy. Dane podzielono losowo, a zatem zarówno w zbiorze uczącym, jak i testowym znalazły się informacje z początku i końca analizowanego okresu, co zwiększa reprezentatywność obu zbiorów, a to powinno mieć przełożenie na zdolności generalizacyjne sieci neuronowych. Do budowy modeli SSN zastosowano

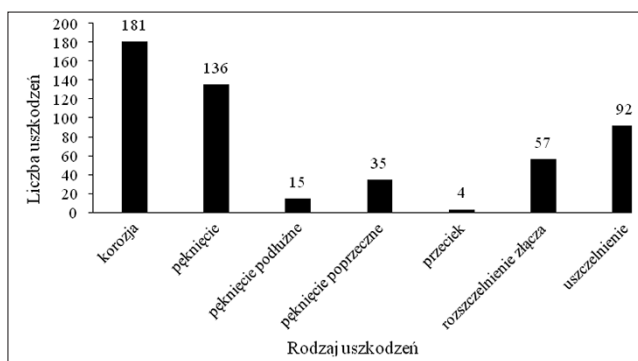
1215 przypadków, które tworzyły zbiór uczący (liczbę poszczególnych rodzajów uszkodzeń przedstawiono na rysunku 1).

Walidacji łącznie 80 modeli (po 20 dla każdej konfiguracji zmiennych wejściowych) dokonano na zbiorze testowym (520 przypadków, z uwzględnieniem podziału na rodzaje uszkodzeń i ich liczbę – rysunek 2). W tym opracowaniu ukazano jedynie wyniki testowania dla wybranej sieci charakteryzującej się najwyższą jakością klasyfikacji w każdej analizowanej konfiguracji. Liczba neuronów ukrytych wahała się w zakresie 1–30. Zastosowano następujące funkcje aktywacji neuronów w warstwie ukrytej i wyjściowej: liniowa, logistyczna tangens hiperboliczny,



Rys. 1.
Liczba rodzajów uszkodzeń w zbiorze uczącym

Fig. 1. Number of kind of damages in learning set



Rys. 2.
Liczba rodzajów uszkodzeń w zbiorze testowym

Fig. 2. Number of kind of damages in testing set

wykładnicza. Zastosowano redukcję wartości wag w granicach 0,0001–0,001. Zmiany w wartościach wag, które określają połączenia pomiędzy poszczególnymi neuronami, wpływają na jakość modelu i dopasowanie danych. Liczba neuronów wejściowych była zmienna w zależności od konfiguracji, gdyż zmieniał się rozmiar wektora zmiennych niezależnych. Należy pamiętać, że zmienne jakościowe, np. pora roku są definiowane przez dwa neurony; typ przewodu przez trzy, itd., co zwiększa rozmiar wektora wejściowego. W przypadku zmiennych ilościowych takie zjawisko nie występuje i np. ciśnienie (jako zmienna ilościowa) jest definiowane przez jeden neuron. Liczba neuronów wyjściowych jest zależna od rodzaju klasyfikowanej zmiennej jakościowej. W przedstawianej analizie klasyfikacji zostało poddane siedem rodzajów uszkodzeń, stąd liczba neuronów w warstwie wyjściowej wynosiła właśnie 7.

Wyniki i dyskusja

Modelowanie za pomocą metody CT i SSN odbywa się kilkietapowo. Najpierw budowanych jest kilka modeli, na podstawie oceny jakości klasyfikacji dla próby uczącej wybierany jest jeden, służący do walidacji na zbiorze testowym. Oba analizowane w tej pracy algorytmy klasyfikacji różnią się metodyką modelowania oraz sposobem przedstawienia wyników. Dla przykładu jakość modelu SSN opisywana jest jako liczba przypadków poprawnie zaklasyfikowanych odniesiona do całkowitej liczby przypadków (tabela 3). Im większa wartość jakości modelu, tym lepsze wyniki klasyfikacji

Tabela 3. Struktury sieci neuronowych i jakość klasyfikacji – zbiór uczący
Table 3. Neural networks' structures and quality of classification – learning set

Nr sieci	A1	Jakość	A1g	Jakość	A3	Jakość	A9	Jakość
1	MLP 13-11-7	46,2	MLP 12-11-7	45,9	MLP 15-11-7	46,5	MLP 12-11-7	45,8
2	MLP 13-26-7	56,0	MLP 12-18-7	55,6	MLP 15-25-7	57,0	MLP 12-18-7	55,8
3	MLP 13-9-7	56,3	MLP 12-3-7	56,0	MLP 15-27-7	56,0	MLP 12-3-7	55,6
4	MLP 13-21-7	56,1	MLP 12-13-7	54,9	MLP 15-2-7	55,2	MLP 12-13-7	54,8
5	MLP 13-3-7	56,0	MLP 12-25-7	55,9	MLP 15-15-7	55,8	MLP 12-25-7	55,7
6	MLP 13-5-7	55,6	MLP 12-12-7	56,6	MLP 15-5-7	56,0	MLP 12-12-7	56,7
7	MLP 13-9-7	55,3	MLP 12-15-7	55,6	MLP 15-11-7	56,4	MLP 12-15-7	55,5
8	MLP 13-16-7	56,0	MLP 12-5-7	55,8	MLP 15-20-7	61,2	MLP 12-5-7	55,6
9	MLP 13-13-7	56,2	MLP 12-24-7	57,9	MLP 15-25-7	56,0	MLP 12-24-7	58,0
10	MLP 13-4-7	56,1	MLP 12-24-7	56,0	MLP 15-30-7	56,1	MLP 12-24-7	55,6
11	MLP 13-21-7	56,0	MLP 12-11-7	54,8	MLP 15-20-7	56,1	MLP 12-11-7	54,8
12	MLP 13-3-7	55,1	MLP 12-19-7	48,1	MLP 15-3-7	55,1	MLP 12-19-7	48,3
13	MLP 13-29-7	56,0	MLP 12-25-7	57,8	MLP 15-21-7	55,9	MLP 12-25-7	58,1
14	MLP 13-29-7	55,6	MLP 12-3-7	55,2	MLP 15-6-7	56,5	MLP 12-3-7	56,4
15	MLP 13-15-7	56,1	MLP 12-10-7	55,8	MLP 15-8-7	55,0	MLP 12-10-7	55,5
16	MLP 13-11-7	56,0	MLP 12-12-7	56,0	MLP 15-5-7	56,8	MLP 12-12-7	56,0
17	MLP 13-13-7	56,7	MLP 12-14-7	56,5	MLP 15-18-7	56,3	MLP 12-14-7	56,4
18	MLP 13-9-7	57,4	MLP 12-13-7	55,6	MLP 15-22-7	56,8	MLP 12-13-7	55,5
19	MLP 13-5-7	54,8	MLP 12-5-7	55,8	MLP 15-27-7	46,5	MLP 12-5-7	55,6
20	MLP 13-21-7	56,0	MLP 12-22-7	56,8	MLP 15-20-7	56,0	MLP 12-22-7	56,6

zmiennej zależnej. W tabeli 3 wybrano dla każdej konfiguracji model SSN charakteryzujący się najwyższą wartością jakości, choć i tak wyniki te kształtują się na średnim poziomie, o czym będzie jeszcze mowa w dalszej części opracowania. Dla konfiguracji A1 sieć z dziewięcioma neuronami ukrytymi w jednej warstwie aktywowanymi funkcją wykładniczą charakteryzowała się jakością na poziomie 57,4. I wcale nie musi być tak, że większa złożoność modelu jest skorelowana z lepszą jakością, co widać na przykładzie konfiguracji A3 i A9. W konfiguracji A3 na wejściu jest więcej zmiennych,

ale mniej neuronów ukrytych niż w przypadku konfiguracji A9, a jakość na nieco wyższym poziomie. Zresztą i tak, z dokładnością inżynierską, są to wyniki porównywalne, oscylujące wokół wartości 60.

W przypadku drzew klasyfikacyjnych do oceny jakości modelu służy koszt resubstytucji [20], czyli stosunek błędnie zaklasyfikowanych przypadków do wszystkich przypadków. Zatem im większa wartość kosztu resubstytucji, tym gorsze wyniki klasyfikacji zmiennej zależnej. W tabeli 4 przedstawiono wyniki dla czterech konfiguracji tak, jak zostało to zaproponowane w [18], wybrano model charakteryzujący się najniższym kosztem. W każdej konfiguracji najniższy koszt obserwowany był w przypadku drzew o najbardziej skomplikowanej strukturze, czyli z największą liczbą węzłów końcowych. Analizując wyniki jakości modelowania na etapie uczenia sieci neuronowych (tabela 3) wnioski nie są już tak oczywiste, gdyż w zależności od konfiguracji optymalny model charakteryzował się inną strukturą.

W ocenie przydatności analizowanych w tym opracowaniu metod należy brać pod uwagę wyniki pochodzące z etapu walidacji modeli, czyli wdrożenia ich na zbiorze testowym. W tabeli 5 ukazano rezultaty przewidywania metodą SSN, całkowitą liczbę poprawnie zaklasyfikowanych rodzajów uszkodzeń w zależności od czterech konfiguracji (A1, A1g, A3 i A9) oraz procent poprawnych wyników łącznie dla wszystkich rodzajów uszkodzeń. Zaś w tabeli 6 zestawiono skróconą formę wyników klasyfikacji metodą CT [18] dla czterech konfiguracji (1, 1g, 3 i 9).

Tabela 4. Koszt resubstytucji w metodzie CT – zbiór uczący
 Table 4. Resubstitution cost in CT method – learning set

Nr drzewa	Liczba węzłów końcowych	Koszt resubstytucji	Nr drzewa	Liczba węzłów końcowych	Koszt resubstytucji
1			1g		
1	14	0,415	1	13	0,416
2	12	0,416	2	11	0,417
3	10	0,421	3	5	0,437
4	8	0,427	4	2	0,452
5	5	0,437	5	1	0,663
6	2	0,452			
7	1	0,663			
3			9		
1	15	0,416	1	22	0,416
2	11	0,420	2	14	0,417
3	9	0,423	3	13	0,418
4	5	0,437	4	11	0,421
5	2	0,452	5	5	0,436
6	1	0,663	6	2	0,452
			7	1	0,663

Tabela 5. Wyniki klasyfikacji metodą SSN – próba testowa
 Table 5. Classification results using SSN method – testing set

A1; MLP 13-9-7, funkcje aktywacji: wykładnicza – warstwa ukryta i wyjściowa								
	korozja	pęknięcie	pęknięcie podłużne	pęknięcie poprzeczne	przeciek	rozszczelnienie złącza	uszczelnienie	Razem
Razem	181	136	15	35	4	57	92	520
Poprawne	149	117	0	0	0	3	23	292
Niepoprawne	32	19	15	35	4	54	69	228
Poprawne (%)	82,32	86,03	0,00	0,00	0,00	5,26	25,00	56,15
Niepoprawne (%)	17,68	13,97	100,00	100,00	100,00	94,74	75,00	43,85
A1g; MLP 12-24-7, funkcje aktywacji: wykładnicza – warstwa ukryta i logistyczna – warstwa wyjściowa								
	korozja	pęknięcie	pęknięcie podłużne	pęknięcie poprzeczne	przeciek	rozszczelnienie złącza	uszczelnienie	Razem
Razem	181	136	15	35	4	57	92	520
Poprawne	147	109	2	0	0	3	24	285
Niepoprawne	34	27	13	35	4	54	68	235
Poprawne (%)	81,22	80,15	13,33	0,00	0,00	5,26	26,09	54,81
Niepoprawne (%)	18,78	19,85	86,67	100,00	100,00	94,74	73,91	45,19
A3; MLP 15-20-7, funkcje aktywacji: tangens hiperboliczny – warstwa ukryta i wyjściowa								
	korozja	pęknięcie	pęknięcie podłużne	pęknięcie poprzeczne	przeciek	rozszczelnienie złącza	uszczelnienie	Razem
Razem	181	136	15	35	4	57	92	520
Poprawne	154	94	3	0	0	4	28	283
Niepoprawne	27	42	12	35	4	53	64	237
Poprawne (%)	85,08	69,12	20,00	0,00	0,00	7,02	30,43	54,42
Niepoprawne (%)	14,92	30,88	80,00	100,00	100,00	92,98	69,57	45,58
A9; MLP 12-25-7, funkcje aktywacji: wykładnicza – warstwa ukryta i logistyczna – warstwa wyjściowa								
	korozja	pęknięcie	pęknięcie podłużne	pęknięcie poprzeczne	przeciek	rozszczelnienie złącza	uszczelnienie	Razem
Razem	181	136	15	35	4	57	92	520
Poprawne	146	113	0	0	0	3	24	286
Niepoprawne	35	23	15	35	4	54	68	234
Poprawne (%)	80,66	83,09	0,00	0,00	0,00	5,26	26,09	55,00
Niepoprawne (%)	19,34	16,91	100,00	100,00	100,00	94,74	73,91	45,00

Wyniki porównania dwóch metod klasyfikacyjnych nie są zbyt obiecujące. W monografii [18] wskazano na konieczność zastosowania w przyszłości m. in. właśnie algorytmu sztucznych sieci neuronowych jako narzędzia być może bardziej uniwersalnego.

Jednakże, jak widać po rezultatach zawartych w tabelach 5 i 6, poprawna klasyfikacja rodzajów uszkodzeń ogółem kształtuje się na podobnym poziomie w przypadku obu metodologii modelowania, czyli wynosi ok. 55%, co jest wynikiem niezadowalającym z uwagi na brak ewentualnych możliwości aplikacyjnych. Z punktu widzenia eksploatatora sieci prawidłowa klasyfikacja rodzaju uszkodzenia powinna prowadzić do wyboru odpowiedniej metody naprawy bądź rekonstrukcji przewodu wodociągowego. To zaś jest związane z ponoszonymi kosztami, których poprawne oszacowanie jest zagadnieniem nietrywialnym i bardzo istotnym, zwłaszcza gdy awariom i wymianie podlegają przewody magistralne. Analiza tabel 5 i 6 pokazuje, iż jedynie takie uszkodzenia, jak korozja i pęknięcie zostały zaklasyfikowane poprawnie w ok. 80%, co jednak nie uprawnia do postawienia stwierdzenia, że metody CT i SSN są w pełni poprawne i całkowicie za-

lecane do klasyfikowania rodzajów uszkodzeń przewodów wodociągowych. Znowu pojawia się zagadnienie, poruszane już wielokrotnie, a mianowicie reprezentatywność, zmienność i liczność zbioru danych. Uszkodzenia „korozja” i „pęknięcie” stanowiły od-

powiednio niecałe 35% i 26% całkowitej liczby uszkodzeń w zbiorze testowym, czyli te dwa rodzaje uszkodzeń były najczęściej rejestrowane w dziennikach awarii na przestrzeni analizowanych ośmiu lat eksploatacji. Podobnie zresztą było w zbiorze uczącym, w którym korozja stanowiła niecałe 34%, a przeciek – 27% wszystkich przypadków. Natomiast błędnie klasyfikowane uszkodzenie „pęknięcie podłużne” stanowiło jedynie niecałe 4% całkowitej liczby przypadków w zbiorze uczącym i niecałe 3% w zbiorze testowym. Jeśli modele zostały stworzone z wykorzystaniem danych, w których przeważały dwa rodzaje uszkodzenia, to nie powinno dziwić, że wyniki dla zbioru testowego są miarodajne jedynie tylko dla tych dwóch najbardziej licznie występujących rodzajów uszkodzeń. Z jednej strony takie spojrzenie na zagadnienie jest uzasadnione, z drugiej jednak dąży się do tego, aby modele oparte o algorytmy sztucznej inteligencji posiadały zdolności generalizacyjne przynajmniej w obrębie stosowności w tym samym systemie wodociągowym. Próba wyodrębnienia każdego rodzaju uszkodzenia i tworzenia oddzielnych modeli dla każdego z nich wydaje się pomysłem nie do końca stosownym i akceptowalnym przy założeniu konieczności uzyskania uniwersalnych wyników i początkowo postawionej tezy o możliwościach aplikacyjnych metod czarnej skrzynki podczas eksploatacji wybranego systemu zaopatrzenia w wodę. Na trafność i dokładność klasyfikacji nie ma również wpływu rozmiar wektora zmiennych niezależnych, gdyż w przypadku każdej analizowanej konfiguracji inna liczba predyktorów uczestniczyła w tworzeniu modelu SSN. Natomiast w metodzie CT do testowania wybrano modele najbardziej złożone, z największą liczbą węzłów końcowych i o najmniejszym koszcie resubstytucji. Wcześniej przeprowadzone autorskie analizy na tym samym zbiorze danych eksploatacyjnych wykazały, że w przypadku klasyfikacji rodzajów uszkodzeń mniej złożone drzewa generują wyniki o jeszcze mniejszej trafności niż ok. 55%. Zatem zmniejszanie złożoności modelu klasyfikującego nie prowadzi do zwiększenia dokładności modelowania. Nieco inne obserwacje dotyczące architektury modeli predykcyjnych były poczynione dla zadania regresyjnego [18], które w swej istocie jest łatwiejsze w opisie zarówno matematycznym, jak i inżynierskim. Z przeprowadzonych wcześniej analiz [18] wynika, że klasyfikacja rodzajów uszkodzeń metodą CT w podziale na poszczególne typy przewodów wodociągowych (oddzielne modele dla magistral, przyłączy i przewodów rozdzielczych) również nie była zbyt dokładna, gdyż podobnie, jak w opisanym tutaj przypadku trafność

Tabela 6. Wyniki klasyfikacji metodą CT – próba testowa
Table 6. Classification results using CT method – testing set

1								
	korozja	pęknięcie	pęknięcie podłużne	pęknięcie poprzeczne	przeciek	rozszczelnienie złącza	uszczelnienie	Razem
Poprawne (%)	81,22	77,94	0,00	5,71	0,00	0,00	39,13	55,96
Niepoprawne (%)	18,78	22,06	100,00	94,29	100,00	100,00	60,87	44,04
1g								
	korozja	pęknięcie	pęknięcie podłużne	pęknięcie poprzeczne	przeciek	rozszczelnienie złącza	uszczelnienie	Razem
Poprawne (%)	81,22	80,88	0,00	5,71	0,00	0,00	29,35	55,00
Niepoprawne (%)	18,78	19,12	100,00	94,29	100,00	100,00	70,65	45,00
3								
	korozja	pęknięcie	pęknięcie podłużne	pęknięcie poprzeczne	przeciek	rozszczelnienie złącza	uszczelnienie	Razem
Poprawne (%)	81,22	82,35	0,00	0,00	0,00	0,00	31,52	55,38
Niepoprawne (%)	18,78	17,65	100,00	100,00	100,00	100,00	68,48	44,62
9								
	korozja	pęknięcie	pęknięcie podłużne	pęknięcie poprzeczne	przeciek	rozszczelnienie złącza	uszczelnienie	Razem
Poprawne (%)	81,77	76,47	0,00	17,14	0,00	0,00	25,00	54,04
Niepoprawne (%)	18,23	23,53	100,00	82,86	100,00	100,00	75,00	45,96

ogólna modelu nie przekraczała 73% dla przyłączy, a dla przewodów magistralnych i rozdzielczych wynosiła ok. 40%.

Podsumowanie

Porównanie dwóch metod modelowania, drzew regresyjnych oraz sztucznych sieci neuronowych, nie jest zagadnieniem nowym [21]. Jednak, jak wynika z literatury przedmiotu, większość rozwiązywanych tymi dwoma metodami problemów miała charakter zadania regresyjnego, czyli predykcji zmiennej ilościowej. Stąd również w kontekście klasyfikacji zmiennej jakościowej (rodzaj uszkodzeń przewodów wodociągowych) postanowiono dokonać porównania tych dwóch algorytmów sztucznej inteligencji. Otrzymane wyniki kształtują się na średnim poziomie. Nie jest to zbyt optymistyczny wniosek płynący z zastosowania modeli CT do oceny i klasyfikacji rodzajów uszkodzeń rurociągów. Świadczy to o małej elastyczności tego algorytmu klasyfikującego. Trudno na obecnym etapie badań wyrokować, czy zastosowanie np. lasu losowego poprawiłoby trafność wyników predykcyjnych. Zatem problem niedokładnej klasyfikacji rodzajów uszkodzeń wydaje się na razie na tym etapie nadal nierozwiązany. Celem jest dalsze doskonalenie warsztatu, aby znaleźć takie narzędzie, które umożliwiłoby rozwiązanie zadania klasyfikacyjnego, które z uwagi na konieczność usprawniania procesu zarządzania i eksploatacji systemów zaopatrzenia w wodę, jest zadaniem istotnym. Niezawodność działania systemów wodociągowych pod kątem redukcji strat wody i liczby oraz rodzaju awarii są elementami kluczowymi w nowoczesnie zarządzanym przedsiębiorstwach wodociągowych na terenach miejskich i wiejskich [22], które powinny dbać nie tylko o stan technicz-

ny infrastruktury, ale również o ochronę zasobów dyspozycyjnych poprzez działanie naprawcze na sieci i redukujące niekontrolowane straty wody. W związku z tym zastosowanie sztucznej inteligencji do klasyfikacji rodzajów uszkodzeń wydaje się być zadaniem jak najbardziej zasadnym. Prawidłowe oszacowanie rodzaju uszkodzenia może mieć wpływ na wybór odpowiedniej metodologii naprawy, odnowy lub wręcz rekonstrukcji przewodu, co w sposób oczywisty przekłada się na możliwość bardziej uzasadnionego wydatkowania kosztów związanych z modernizacją systemu dystrybucji wody oraz na usprawnienie działań naprawczych prowadzonych przez przedsiębiorstwa wodociągowe.

LITERATURA

- [1] Kwietniewski M., Kowalski D. 2022.: Wodociągi i kanalizacje – od starożytności do współczesności. Gaz, Woda i Technika Sanitarna, (9): 30-37.
- [2] Kasino W., Suchorab P., Kowalski D. 2021.: Inteligentny system zarządzania siecią wodociągową w Żywcu. Gaz, Woda i Technika Sanitarna, (10): 16-19.
- [3] Rak J., Ustrobiński M., Potyczny R., Stręk M. 2023.: Analiza zużycia wody w Rzeszowie w czasie kryzysu uchodźczego z Ukrainy. Instal, (1): 44-46, DOI: 10.36119/15.2023.1.6
- [4] Kowalski D., Kowalska B., Suchorab P. 2022.: Smart water supply system: a quasi intelligent diagnostic method for a distribution network. Applied Water Science, (12): 135, <https://doi.org/10.1007/s13201-022-01656-w>.
- [5] Kowalski D., Suchorab P. 2023.: The impact assessment of water supply DMA formation on the monitoring system sensitivity. Applied Sciences, (13): 1554, <https://doi.org/10.3390/app13031554>.
- [6] Iwanek M. 2022.: Parameters characterizing leakages from damaged water pipes in the aspect of environmental security. Applied Water Science, (12): 126, <https://doi.org/10.1007/s13201-022-01641-3>.
- [7] Iwanek M., Suchorab P., Suchorab Z. 2019.: FEM simulation of water lost through damaged

household water connection. MATEC Web of Conferences, (252): 05008, <https://doi.org/10.1051/mateconf/201925205008>.

- [8] Piórek D., Stańczyk J., Licznar P. 2022.: Ocena parametrów wydajności i ciśnienia sieci wodociągowej na cele przeciwpożarowe w świetle przeprowadzonych testów hydrantowych. Instal, (12): 71-76, DOI: 10.36119/15.2022.12.12.
- [9] Miszta-Kruk K., Kaznowska E. 2019.: Problematyka zjawiska suszy a systemy zaopatrzenia w wodę. Instal, (12): 48-50, DOI: 10.36119/15.2019.12.7.
- [10] Leite M., Felipe E., Fogaça M., Maher R. 2022.: A comparison of Brazilian, North American, and English water supply regulation models. Groundwater for Sustainable Development, (18): 100748, <https://doi.org/10.1016/j.gsd.2022.100748>.
- [11] Boryczko K., Piegdoń I., Szpak D., Żywiec J. 2021.: Risk assessment of lack of water supply using the hydraulic model of the water supply. Resources, (10): 43, <https://doi.org/10.3390/resources10050043>.
- [12] Żywiec J., Tchórzewska-Cieślak B. 2023.: Porównanie wybranych metod oceny niezawodności operatora systemów technicznych na przykładzie systemu zbiorowego zaopatrzenia w wodę. Instal, (5): 42-48, DOI: 10.36119/15.2023.5.7.
- [13] Żywiec J., Szpak D., Piegdoń I., Boryczko K., Pietrucha-Urbaniak K., Tchórzewska-Cieślak B., Rak J. 2023.: An approach to assess the water resources reliability and its management. Resources, (12): 4, <https://doi.org/10.3390/resources12010004>.
- [14] Boryczko K., Rak J. 2020.: Method for assessment of water supply diversification. Resources, (9): 87, DOI: 10.3390/resources9070087.
- [15] Kwietniewski M., Świercz P., Chudziński J. 2022.: Modern methods for monitoring water leakages in water networks. Studia Geotechnica et Mechanica, (44:1): 53-65, <https://doi.org/10.2478/sgem-2022-0001>.
- [16] Vanijirattikharn R., Khomsay S., Kitbutrattaw N., Khomsay K., Supachukul U., Udomsuk S., Suwattikul J., Oumtrakul N., Anusart K. 2022.: AI-based acoustic leak detection in water distribution systems. Results in Engineering, (15): 100557, <https://doi.org/10.1016/j.rineng.2022.100557>.
- [17] Robles-Velasco A., Muñuzuri J., Onieva L., Cortés P. 2021.: An evolutionary fuzzy system to support the replacement policy in water supply networks: The ranking of pipes according to their failure risk. Applied Soft Computing, (111): 107731, <https://doi.org/10.1016/j.asoc.2021.107731>.
- [18] Kutylowska M. 2019.: Metody regresyjne i klasyfikacyjne w analizie i ocenie poziomu awaryjności przewodów wodociągowych. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
- [19] Kutylowska M., Cieżak W. 2023.: Sztuczna sieć neuronowa jako narzędzie wspomagające w analizie awaryjności przewodów wodociągowych. Instal, (3): 41-45, DOI: 10.36119/15.2023.3.7.
- [20] Breiman L., Friedman J. H., Olshen R. A., Stone C. J. 1984.: Classification and regression trees. Chapman & Hall/CRC, Boca Raton, USA.
- [21] Kim J.-W., Pachepsky Y. 2010.: Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. Journal of Hydrology, (394): 305-314, DOI: 10.1016/j.jhydrol.2010.09.005.
- [22] Flis A., Gęca W., Iwanek M., Kowalski D. 2023.: Analiza strat wody w dwóch wodociągach wiejskich. Gaz, Woda i Technika Sanitarna, (1): 8-14, DOI: 10.15199/17.2023.1.2.