



Zastosowanie sztucznych sieci neuronowych do prognozy ilości odpadów bytowo-gospodarczych

Anna Chmieleńska-Bernacka, Robert Sidelko
Politechnika Koszalińska

1. Wstęp

W styczniu 2012 r. weszły w życie zmiany do ustawy z dnia 13 września 1996 r. o utrzymaniu czystości i porządku w gminach, znowelizowane ponownie w styczniu 2013 r. (tekst jednolity: Dz. U. 2012 r. poz. 391). Powyższe zmiany rewolucjonizują system gospodarki odpadami. Zgodnie z wprowadzanymi zmianami gminy staną się właścicielami odpadów, a w związku z tym przejmą pełną kontrolę w zakresie gospodarki odpadami na swoim terenie. Wdrożenie nowego systemu wraz z ustaleniem opłaty pobieranej od mieszkańców za gospodarowanie odpadami komunalnymi, musi nastąpić do dnia 1 lipca 2013 r. Opłata za gospodarowanie odpadami, tzw. podatek śmieciowy, stanowić będzie wartość, jaką za korzystanie z systemu zapłaci każde gospodarstwo domowe, a suma opłat powinna zrekompensować wydatki poniesione na prowadzenie całego systemu przez gminę. Sposób naliczenia opłaty za gospodarowanie odpadami został ściśle określony przez ustawę o utrzymaniu czystości i porządku w gminie w art. 6j ust. 1 i ust. 2. W każdym z ustawowych wariantów wyliczenie opłaty stanowi iloczyn stawki i czynnika różnicującego. Czynnikiem różnicującym może być: powierzchnia lokalu mieszkalnego, ilość zużytej wody z danej nieruchomości, liczba mieszkańców zamieszkujących daną nieruchomość lub jednostkowo gospodarstwo domowe (dopuszcza się stosowanie więcej niż jednej metody ustalenia opłat na obszarze gminy – 6k ust. 1. pkt. 1). Żadna z zaproponowanych przez ustawodawcę metod nie jest w odczuciu społecznym uważana za sprawiedliwą, z racji tego, iż aktualna konstruk-

cja prawna powoduje obciążenie finansowe za fakt wytwarzania odpadów, a nie za ilość wytworzonych odpadów – stanowiącą rzeczywiste korzystanie ze środowiska. Powyższe rodzi wiele dyskusji na temat wad i zalet każdej z metod, sprowadzając ten problem jedynie do fiskalnego znaczenia egzekwowalności opłaty.

Gospodarka odpadami wymaga znacznych nakładów finansowych, które szacuje się w Polsce na 650–890 mln zł/rok, co stanowi 8–10% wszystkich wydatków na ochronę środowiska [2]. Tworzenie systemu gospodarki odpadami, obok kryteriów ekonomicznych, musi uwzględniać także kryteria akceptowalności społecznej i efektywności ekologicznej. Podstawą racjonalnego planowania gospodarki odpadami jest tzw. jednostkowy wskaźnik nagromadzenia odpadów, którego prawidłowy dobór jest najważniejszym zadaniem etapu planistycznego [1, 4]. Do grup uwarunkowań mających wpływ na ilość wytwarzanych odpadów należą czynniki: ekonomiczne, społeczne i infrastrukturalne. Samo wskazanie grup elementów mających wpływ na zmianę ilości wytwarzanych odpadów jest niewystarczające, bowiem nieznana jest siła ich wzajemnego oddziaływania [5].

Wybór metody pozwalającej opracować model prognozujący ilość wytwarzanych odpadów w gospodarstwie domowym, stanowiący podstawę planowania gospodarki na danym obszarze np. gminy, powinien uwzględniać szereg cech, co do których przewidujemy ich istotny wpływ na wynik końcowy. Ze względu na zdefiniowanie szeregu cech o charakterze zmiennych ilościowych i jakościowych, wzajemnie skorelowanych, interesującą alternatywą w tym przypadku jest zastosowanie sztucznych sieci neuronowych (SSN).

2. Badania pilotażowe

Badanie pilotażowe wykonano w okresie wrzesień–listopad 2012 roku, w ramach projektu pt. „Określenie strumienia odpadów komunalnych zmieszanych, jako podstawa tworzenia systemów gospodarki odpadami i edukacji ekologicznej w Gminie Sianów”, finansowanego przez Wojewódzki Fundusz Ochrony Środowiska i Gospodarki Wodnej w Szczecinie oraz Gminę Sianów. Celem badań było opracowanie modelu prognozującego masę odpadów bytowo-gospodarczych gromadzonych w gospodarstwie domowym, w określonym czasie, z uwzględnieniem

czynników determinujących ilość wytwarzanych odpadów, przy wykorzystaniu sztucznej inteligencji. Zakres badań obejmował:

- ustalenie kryterium wyboru poszczególnych gospodarstw domowych uczestniczących w programie badań,
- określenie liczby gospodarstw domowych objętych monitoringiem ilości wytwarzanych odpadów,
- opracowanie ankiet do wypełnienia przez wytypowane gospodarstwa,
- badania terenowe polegające na inwentaryzacji zbieranych odpadów,
- analizę statystyczną uzyskanych danych wraz z opracowaniem elektronicznej bazy danych,
- analizę numeryczną przy wykorzystaniu sztucznej inteligencji,
- interpretację uzyskanego modelu neuronowego.

Programem badań objęto 101 gospodarstw domowych w siedmiu jednostkach osiedleńczych gminy Sianów, dobierając próbę tak, by uwzględniała ona różny charakter zabudowy (zabudowa jednorodzinna/wielorodzinna) oraz różne jednostki administracyjne (miast/wieś). Inwentaryzację wytworzonych odpadów przez poszczególne gospodarstwa domowe prowadzono w okresie 4 tygodni. Każdy przypadek opisano za pomocą piętnastu cech stanowiących zmienne, podzielone w trakcie analizy numerycznej na tzw. zmienne zależne (objaśniane) i zmienne niezależne (objaśniające). Zmiennymi uwzględnianymi w analizie były:

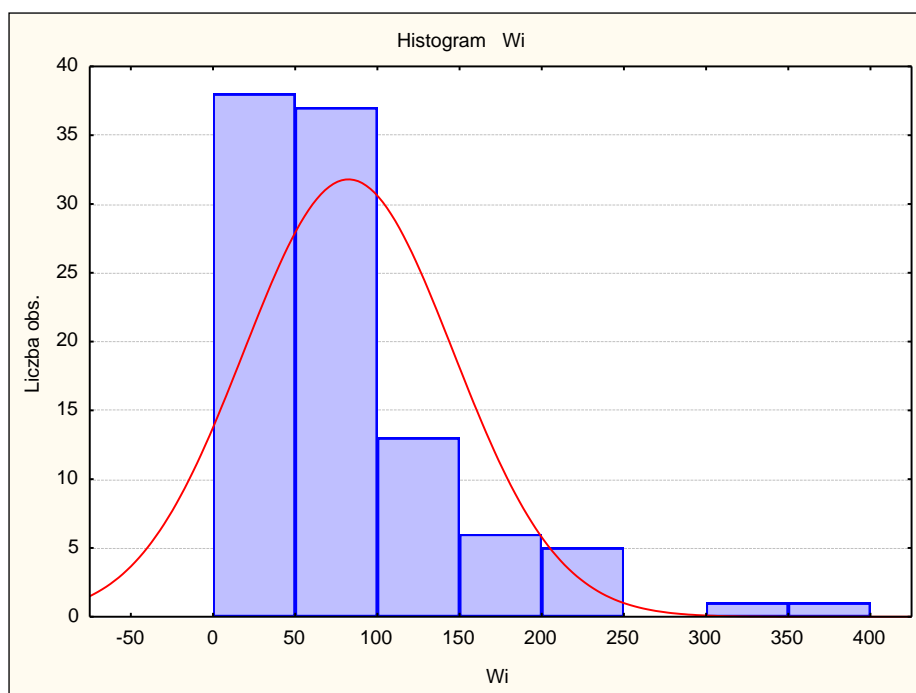
- liczba osób w gospodarstwie domowym [osób/ gosp.] – zmienna 1,
- ilość gromadzonych odpadów w gospo. [kg/4 tyg.] – zmienna 2,
- liczba dzieci do 3 lat [dzieci/ gosp.] – zmienna 3,
- liczba dzieci powyżej 3 lat [dzieci/ gosp.] – zmienna 4,
- dochód netto [zł/ gosp.] – zmienna 5,
- liczba osób pracujących [osób/ gosp.] – zmienna 6,
- liczba osób uczących się [osób/ gosp.] – zmienna 7,
- samoocena świadomości ekologicznej w skali 1-10 [-] – zmienna 8,
- jednostka administracyjna miasto/wieś [m lub w] – zmienna 9,
- rodzaj zabudowy wielo/jednorodzinna [w lub j] – zmienna 10,
- media 1 tj. wod.-kan. [tak lub nie] – zmienna 11,
- media 2 tj. gaz sieciowy [tak lub nie] – zmienna 12,
- powierzchnia mieszkalna gospodarstwa [m^2] – zmienna 13,
- zużycie wody w gospodarstwie [m^3 / 4 tyg.] – zmienna 14.

3. Analiza wyników badań

Wskaźnikiem, którego wyznaczenie stanowiło główny cel badań była jednostkowa ilość wytwarzanych odpadów. Ze względu na zastosowanie sztucznej sieci neuronowej do modelowania związku pomiędzy określonymi w ankietach parametrami, konieczne było dostosowanie dostępnej bazy do wymogów warunkujących zastosowanie SSN. Jednym z nich jest proporcja pomiędzy liczbą przypadków (1 przypadek – jedno gospodarstwo), a liczbą zmiennych objaśniających. Generalnie należy przyjąć, że im więcej danych zostanie wykorzystanych w procesie uczenia, tym lepsza zdolność generalizacji opracowanej sieci, mówiąc inaczej uzyskane prognozy będą dokładniejsze. Trudno przyjąć jednoznaczne kryterium pozwalające ustalić minimalną liczbę przypadków w zależności od liczby zmiennych na wejściu (*zmienne niezależne*) i wyjściu sieci (*zmienne zależne*). W tym względzie można posłużyć się regułami heurystycznymi wskazującymi, że na każdą zmienną wejściową powinna przypadać seria, co najmniej 10 przypadków uczących, gdzie jeden przypadek zawiera po jednej wartości dla każdego wejścia sieci. Przykładowo, przyjmując 10 wejść sieci, należy dysponować 100 przypadkami uczącymi [6].

Kryterium dostępnej ilości przypadków i liczby zmiennych wymusiło *preprocessing*, w efekcie którego zrezygnowano z pięciu zmiennych (numery: 4, 7, 8, 9 i 12) oraz dokonano kompilacji dwóch zmiennych (numery: 1 i 2), uzyskując zmienną wyrażającą ilość odpadów wytwarzanych przez jednego członka rodziny w ciągu roku, a więc jednostkowy wskaźnik nagromadzenia odpadów, W_i . Na rysunku nr 1, przedstawiono nomogram z naniesioną krzywą rozkładu normalnego, obrazującą liczebność przypadków w poszczególnych przedziałach wartości wskaźnika W_i . Przypadki nr 34 i 100, ze względu na dużą wartość tego wskaźnika, przekraczającą 300 [Mg/miesz. · rok.] potraktowano jako odstające i nie uwzględniono ich w dalszej analizie. Nie uwzględniono również, jako niewiarygodne przypadki dla których obliczona wartość wskaźnika W_i wyniosła poniżej 30 [kg/miesz. · rok]. Ostatecznie liczba przypadków zweryfikowanych pozytywnie wyniosła 88. Do kompilacji sztucznej sieci neuronowej wykorzystano opcję automatycznego poszukiwania sieci, dostępną w zakładce *Data Mining/Automatyczne sieci neuronowe* pakietu *Statistica*.

Zmiennymi stanowiącymi dane wejściowe sieci były wartości zmiennych o charakterze ilościowym (zmiennie o numerach: 1, 3, 5, 6, 13 i 14) oraz zmiennie o charakterze jakościowym (zmiennie o numerach: 10 i 12). Zmienną na wyjściu sieci był jednostkowy wskaźnik nagromadzenia odpadów.



Rys. 1. Ilości przypadków w kolejnych przedziałach wartości wskaźnika
Fig. 1. Number of cases for next ranges of W_i coefficient

Wyboru sieci dokonano w oparciu o interpretację współczynników określających jakość procesu uczenia i testowania, które w przypadku sieci RBF 10-17-1 były największe i wyniosły odpowiednio: 0,671 i 0,503. Symbol sieci RBF, oznacza sieć o architekturze radialnej zawierającą odpowiednio: 10 i 17 neuronów w warstwie wejściowej i ukrytej oraz 1 neuron w warstwie wyjściowej. Zestawienie wyników działania sieci w formie arkuszy predykcji dla wszystkich przypadków ze zbioru testującego, przedstawiono w tabeli 1.

Tabela 1. Arkusz predykcji SSN dla zbioru testowego
Table 1. A sheet of NN prediction for testing set

Przypadek nazwa	Arkusz predykcji dla W_i (sianow)					
	W_i Zm.zal	W_i - Wyjście 1. RBF 10-17-1	W_i - Wyjście 2. RBF 10-17-1	W_i - Wyjście 3. RBF 10-17-1	W_i - Wyjście 4. RBF 10-12-1	W_i - Wyjście 5. RBF 10-17-1
6	111,9000	83,06009	39,8348	77,1416	92,7943	78,38142
8	66,7680	83,06773	39,8198	77,1788	88,4985	77,36205
13	50,3400	83,06690	39,8233	76,3127	85,4188	77,29399
25	111,2000	87,17113	58,8034	96,8475	100,8421	90,43367
26	64,7250	83,08083	40,9979	83,2242	75,9130	86,08078
39	70,8000	83,09886	37,7199	104,7886	74,1579	82,18413
48	18,4800	83,54632	66,8594	75,8124	75,8636	89,00770
53	22,0800	68,36345	39,8704	77,3393	77,8378	82,13079
66	52,9700	84,23907	43,9187	86,9171	78,7053	87,13773
71	118,8000	85,88396	39,8826	78,3222	74,4898	82,14947
88	176,7000	88,55379	114,2031	103,6862	93,2661	93,85262
89	75,1800	86,15344	78,9071	54,0719	75,3411	92,67472

4. Dyskusja wyników badań

Wykorzystując sztuczne sieci neuronowe do modelowania procesu, w którym szereg zmiennych wpływa na wartość zmiennej przewidywanej, oczekuje się uzyskanie optymalnej prognozy nie tylko w kontekście testów prowadzonych w trakcie procesu uczenia, a więc ograniczonych do określonej próby statystycznej (baza danych), ale przede wszystkim wiarygodnych prognoz dla każdego przypadku w całej populacji. Zdolność sieci neuronowej do generalizacji w omawianych badaniach, oceniono poprzez wyłączenie z procesu uczenia dwóch przypadków o numerach 5 i 46. Oba przypadki różniły się zasadniczo większością zmiennych, ostatecznie zakwalifikowanych do procesu uczenia SSN, w sposób następujący:

	Liczba osób	Dzieci >3lat	Dochód [zł]	Osób Pracuj.	Rodz. Zabud.	Media 2	Pow. Miesz.	Zuż. wody	W_i
Nr 5	5	0	3000	2	W	tak	47,77	8	97,2
Nr 46	3	1	2000	2	J	nie	250	5	52,0

W tabeli 2 przedstawiono wynik prognozy wybranej SSN (RBF 10-17-1) dla dwóch powyższych przypadków, uzyskując wartości wskaźnika W_i odpowiednio dla przypadku numer 5 – 83,04 i przypadku numer 46 – 84,34. W obu przypadkach prognoza przewiduje inne, niż

deklarowane wartości wskaźnika W_i , zaniżając je nieznacznie w pierwszym (numer 5) i podwyższając w drugim przypadku (numer 46).

Tabela 2. Wynik prognozy wskaźnika W_i dla dwóch przypadków: 5(1) i 46(2)
Table 2. Result of W_i coefficient prognosis for two cases: 5 (1) and 46 (2)

#	1.Wi	osob	dzieci <3	dochod	prac/uc...	pow_r
1	83,0411...	5,000000	0,000000	3000,00...	3,000000	47,770
2	84,3371...	3,000000	1,000000	2000,00...	2,000000	250,00

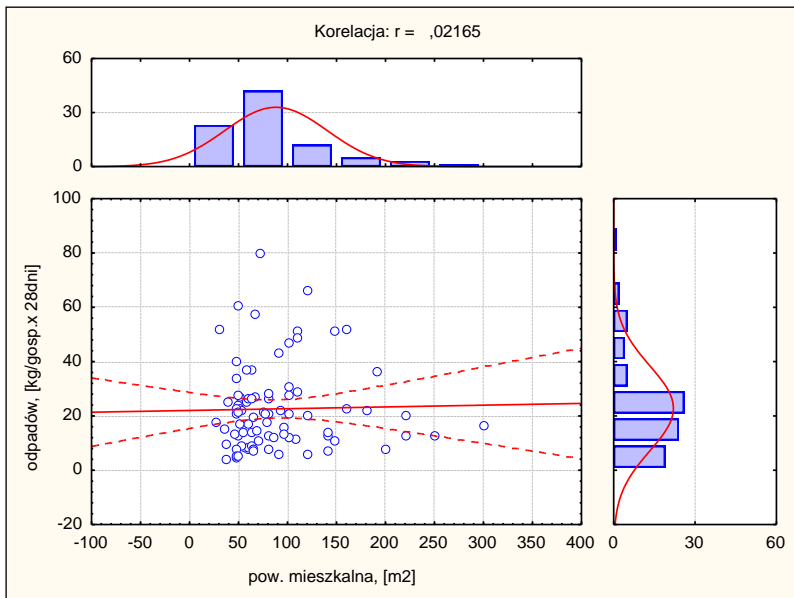
Dokonując oceny uzyskanego rezultatu działania sieci należy wziąć pod uwagę dwa czynniki. Po pierwsze, wiarygodności informacji na przykład, co do podawanej w ankietach ilości gromadzonych w gospodarstwie odpadów. Deklarowane w tym zakresie dane nie muszą być danymi rzeczywistymi z różnych względów, chociażby nierzetelnego gromadzenia odpadów w okresie prowadzenia badań poprzez wydzielenie ze strumienia odpadów składników zagospodarowywanych we własnym zakresie. Po drugie, fakt zastosowania algorytmu matematycznego wykorzystywanego do tworzenia modelu nieliniowego, opisującego zmienną objaśnianą w funkcji n zmiennych objaśniających.

Model regresyjny SSN podaje prognozę, jako najbardziej prawdopodobną wynikającą z posiadanego „doświadczenia”. Owe doświadczenie jest utrwalone w strukturze sieci poprzez określenie wag sygnałów synaptycznych w wyniku uczenia SSN, w trakcie wielokrotnej prezentacji dostępnych przypadków na etapie trenowania sieci. W konsekwencji SSN podaje prognozę obarczoną pewnym błędem statystycznym, a nie błędem metodycznym. Zmniejszenie błędu prognozy jest możliwe wy-

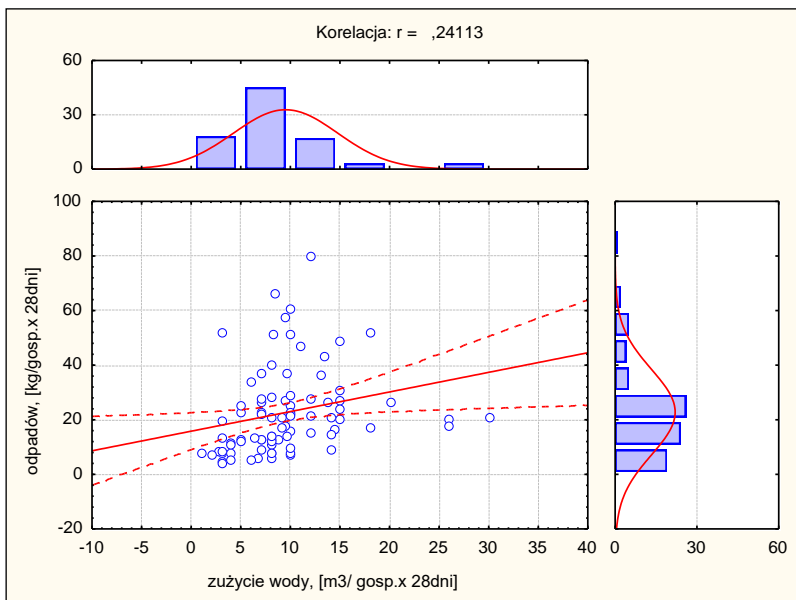
łącznie poprzez udostępnienie w procesie uczenia SSN większej liczby przypadków. Na końcowy wynik mają wpływ wszystkie zmienne niezależne, aczkolwiek część z nich może mieć charakter zmiennych dominujących, a część zmiennych dyskretnych niejednokrotnie istotnie wpływających na wartość prognozy. Taką zmienną może być np. zmienna jakościowa *media-2*, która określa dostęp danego gospodarstwa do sieci gazowniczej. Skutkiem braku możliwości korzystania z gazu sieciowego jest wykorzystywanie tradycyjnego źródła energii w postaci paliwa stałego, pozostawiającego odpad po jego wykorzystaniu. W konsekwencji, sieć neuronowa prognozuje znacznie większy wskaźnik W_i dla przypadku numer 46, gdyż na podstawie „własnego doświadczenia” gospodarstwa wykorzystujące inne źródło energii, niż gaz wytwarzają więcej odpadów.

Równoległe do procesu modelowania prognozy wytwarzanych odpadów bytowo-gospodarczych w gospodarstwach domowych, dokonano oceny możliwości szacowania ilości wytwarzanych odpadów dla 101 przypadków, objętych badaniem pilotażowym, w odniesieniu do metod naliczania opłaty za wytwarzane odpady, zaproponowanych w znowelizowanej ustawie z dnia 13 września 1996 r. o utrzymaniu czystości i porządku w gminach (tekst jednolity: Dz. U. 2012 r. poz. 391). Zgromadzone w tym zakresie wyniki badań wskazują, że szacowanie ilości wytwarzanych odpadów w zależności od powierzchni mieszkalnej, czy ilości zużytej wody jest niewiarygodne. Świadczy o tym niska korelacja pomiędzy poszczególnymi zmiennymi, wynosząca odpowiednio: $k = 0,02$ i $k = 0,24$, co według przyjętych zasad należy uznać za słabą korelację [3, 7]. Wykres krzywej regresji dla obu współzależności przedstawiono na rysunkach nr 2 i 3. Najwyższą korelację ($k = 0,49$) stwierdzono pomiędzy ilością wytwarzanych odpadów, a ilością osób w gospodarstwie domowym.

Z przedstawionych danych wynika, że uzależnienie wysokości opłaty od ilości wytwarzanych w danym gospodarstwie domowym odpadów bytowo-gospodarczych, szacowanych przy założeniu proporcjonalnego wzrostu niezależnie od rodzaju przyjętego czynnika różnicującego, jest nieuzasadnione. Błędnie ustalona ilość wytwarzanych odpadów, nie daje w konsekwencji możliwości opodatkowania opłatą adekwatną do rzeczywistego korzystania ze środowiska.



Rys. 2. Regresja opisująca zależność ilości odpadów od pow. mieszkalnej
Fig. 2. Regression between amount of waste and flat area



Rys. 3. Regresja opisująca zależność ilości odpadów od zużycia wody
Fig. 3. Regression between amount of waste and water consumption

Literatura

1. **Kempa E.S.** *Gospodarka odpadami miejskimi*. Arkady. Warszawa 1983.
2. **Konieczna R., Kulczycka J.** *Analiza ekonomiczna systemów gospodarki odpadami*. cz. II. IGSMiE PAN, Karków 2011.
3. **Piekarski J.**: *Zastosowanie metod numerycznych do modelowania procesu filtracji grawitacyjnej*. Rocznik Ochrona Środowiska (Annual Set of Environment Protection), 12 (2011).
4. **Sidelko R.** *Jednostkowy wskaźnik nagromadzenia odpadów komunalnych na przykładzie Koszalina*. VII Konf. naukowo-techniczna. Koszalin 2011.
5. **Talalaj I.A.** *Wpływ wybranych czynników społeczno-ekonomicznych na zmiany ilości odpadów komunalnych w województwie podlaskim*. Inżynieria Ekologiczna. Nr 25 (2011).
6. **StatSoft.** *Sieci neuronowe*. Electronic textbook. <http://www.statsoft.pl>
7. **Stanisz A.** *Przystępny kurs statystyki*. StatSoft Polska. Kraków 1998.

Neural Network Application for Prediction of Amount of Municipal Waste

Abstract

In article the results of neural network (NN) application connected with prediction of amount of municipal waste was presented. To neural network training about one hundred cases was used. One case mean household where a individual questionnaire enclosed fourteen questions was filed up. Finally only nine features treated as independent variables was qualified. Neural network verification depended on comparing real and predicted value for two cases removed during NN teaching process.