

ANALIZA OBSZARÓW LEŚNYCH W SYSTEMACH INFORMACJI PRZESTRZENNEJ WSPOMAGANYCH SYSTEMAMI EKSPERTOWYMI I SZTUCZNYMI SIECIAMI NEURONOWYMI

Wiktor Tracz

Katedra Urządzania Lasu, Geomatyki i Ekonomiki Leśnictwa, Szkoła Główna Gospodarstwa Wiejskiego, Warszawa

1 WSTĘP

Las jako składnik żywej przyrody jest układem skomplikowanym i stale zmieniającym się. Współczesna gospodarka leśna pełni zarówno funkcje produkcyjne, jak również: rekreacyjną, ochronną, edukacyjną. Wszystko to komplikuje podejmowanie trafnych i skutecznych decyzji przy zarządzaniu leśnictwem i wymusza stosowanie sprawnych pomocniczych narzędzi i metod.

Różne systemy informatyczne są od dawna wykorzystywane w leśnictwie polskim. Na przykład System Informatyczny Lasów Państwowych (SILP), Systemy Informacji Przestrzennej (SIP) w postaci Leśnej Mapy Numerycznej. Potrzeba wykorzystania takich narzędzi w zarządzaniu gospodarką leśną w Polsce jest oczywista, a korzyści z tego płynące są ogromne. W zagadnieniach bardziej złożonych, nowych i mało zbadanych, które pojawiają się w związku z wielofunkcyjnym podejściem do gospodarki leśnej, zachodzi potrzeba wykorzystania wyrafinowanych technik i metod analitycznych wspomagających SILP i SIP. Metody Sztucznej Inteligencji (SI) są jednymi z takich technik. Stosowanie metod SI razem z SIP w jednym systemie wspomagania decyzji zwiększa możliwości analityczne i wydaje się być bardzo pomocne w rozwiązywaniu wielu słabo ustrukturalizowanych problemów z zakresu leśnictwa.

Praca podejmuje próbę oceny celowości i efektywności wykorzystania Sztucznych Sieci Neuronowych (SN) wspomagających SIP w rozwiązywaniu złożonych problemów przestrzennych leśnictwa [4].

2 MOŻLIWOŚĆ WYKORZYSTANIA SN RAZEM Z SIP W ANALIZACH PRZESTRZENNYCH

Las jest tworem przestrzennym, leży w danym miejscu w przestrzeni geograficznej, zajmuje określoną powierzchnię, jest w relacji przestrzennej z innymi obiektami przestrzennymi. Najefektywniejszym aktualnie narzędziem wspomagającym zarządzanie obiektami i zjawiskami przestrzennymi są Systemy Informacji Przestrzennej. Oparte są na wykorzystaniu technik komputerowych i umożliwiają gromadzenie, udostępnianie, analizowanie i prezentację danych przestrzennych. Dzięki tym bogatym możliwościom SIP jest niezastąpionym narzędziem wspomagającym zarządzanie problemami przestrzennymi w leśnictwie.

W leśnictwie istnieje grupa zagadnień mało ustrukturalizowanych, a zatem trudnych do rozwiązania tylko za pomocą pojedynczych narzędzi informatycznych. Do rozwiązania takich problemów, jak się wydaje, najlepiej nadają się metody SI, które mogą być stosowane wtedy, gdy:

- nie dysponujemy całą wiedzą potrzebną do rozwiązania danego problemu, a czas rozwiązania metodami znanymi (np. algorytmicznymi) jest zbyt długi;
- dane są nie tylko w postaci numerycznej, ale też w postaci symbolicznej;
- informacja jest dostępna raczej w postaci heurystyk niż dobrze zdefiniowanych procedur algorytmicznych;
- wiedza o przedmiocie nie jest pełna, a koszt pozyskania niezbędnych ekspertyz jest duży;
- wiedza jest niepewna.

Podczas rozwiązywania problemów dane przestrzenne poddawane są odpowiedniemu przetwarzaniu i dalej mogą być wykorzystane w SIP. Sieci Neuronowe realizują następujące przetwarzania [5]:

- optymalizację, a w tym rozwiązywanie liniowych i nieliniowych równań;
- klasyfikację, która realizowana jest przez podział zbioru wejściowego na klasy lub kategorie skojarzenia każdego wejścia z kategorią;
- sterowanie, realizowane inteligentnie bez konieczności opracowania modelu, oparte wyłącznie na doświadczeniu;
- przypomnianie polegające na: odzyskiwaniu (albo interpretowaniu) zmagazynowanych w SN informacji, obliczaniu wyjścia dla danego wejścia;
- rozpoznawanie, rozumiane jako klasyfikowanie wejścia, mimo że nie odpowiada ono żadnemu z przechowywanych wzorców;
- skojarzenie, realizowane w następujących wariantach: skojarzenie uszkodzonego (zdeformowanego) wejścia (albo wywołania) z najbliższym przechowywanym wzorcem, skojarzenie między parą wzorców, diagnostyka, analiza;
- estymację, czyli realizacja następujących zadań: aproksymacja, interpolacja, filtrowanie, predykcja, prognozowanie.

Wyżej wymienione przetwarzania realizowane przez SN mogą dostarczać rozwiązań przeznaczonych bezpośrednio do prezentacji w SIP lub wyników pośrednich, które mogą być wykorzystane w dalszych analizach realizowanych za pomocą SIP. Sieci neuronowe, tak jak i SIP, budowane są obecnie jako systemy otwarte. Pozwala to na łatwiejszą i szybszą ich integrację między sobą oraz z innymi technikami w jednym systemie wspomagania.

W pracy podjęto próbę przetestowania możliwości zastosowania sieci neuronowych wspomagających SIP w rozwiązywaniu trudnych zagadnień zarządzania w leśnictwie. Wybrano zagadnienie prognozowania rozwoju populacji szkodliwych owadów drzew leśnych jako reprezentację ważnych zagadnień ochrony lasu.

3 WYKORZYSTANIE SN WSPOMAGAJĄCEJ SIP

Prognozowanie dotyczyło szkodliwego owada drzew iglastych brudnicy mniszki (*Lymantria monacha* L.). Monitoring brudnicy mniszki z zastosowaniem pułapek feromonowych jest stosowany w Polsce w celu stwierdzenia obecności tego szkodnika sosny w terenie, śledzenia przebiegu jego rójki i długoterminowych trendów gęstości populacji oraz wczesnego wykrycia obszarów kompleksów leśnych o podwyższonej liczebności populacji tego owada. Prognoza zagrożenia przez brudnicę mniszkę oparta jest głównie na liczbie odłowionych motyli do pułapek feromonowych, rzadziej na liczbie samic znalezionych na wyznaczonych grupach drzew kontrolnych lub obserwacji samic podczas jednorazowego przejścia przez drzewostan wzdłuż transektu. Prognoza taka jednak nie uwzględnia innych ważnych czynników wpływających na stan populacji owada [1]. Do takich czynników należą:

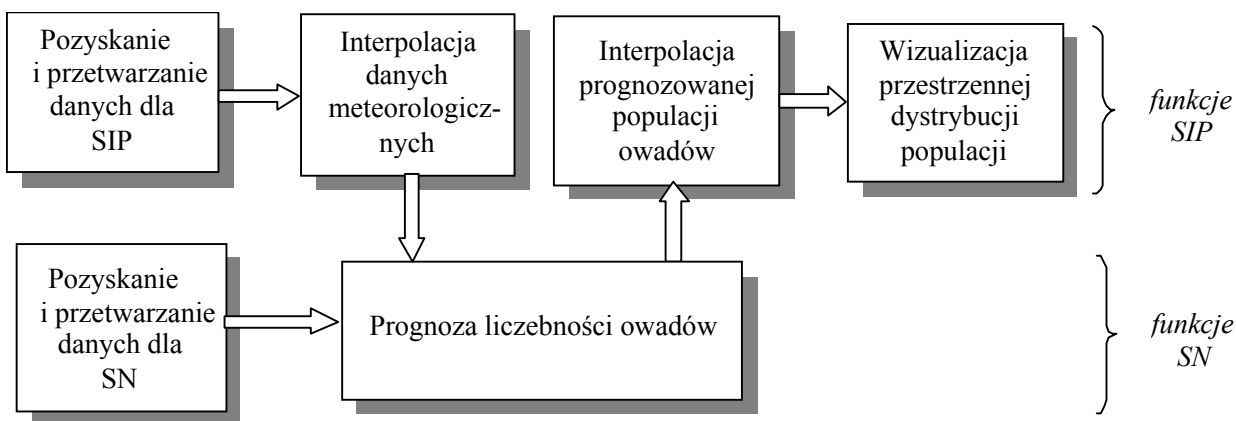
- kondycja drzewostanu (jego stan fizjologiczny), a w związku z tym jego odporność na żery;
- stan populacji owadów pasożytniczych, grzybów oraz gryzoni leśnych i dzików;
- stan fizjologiczny populacji owadów gradacyjnych;
- cykliczność pojawów;
- aktywność słońca;
- fazy gradacji;
- zwalczanie lub jego brak w poprzednich latach.

Celem badań było sprawdzenie efektywności wykorzystania sieci neuronowej wspomagającej SIP w prognozowaniu zagrożenia brudnicą mniszką. Budując model neuronowy zagadnienia prognozowania, starano się z jednej strony w pełni wykorzystać obecny stan wiedzy merytorycznej, z drugiej strony uwzględnić lokalne warunki nadleśnictwa testowego oraz wykorzystać dane już istniejące.

Model prognozowania zagrożenia przez brudnicę mniszkę uwzględniał takie czynniki, jak: odporność drzewostanu na żery owada, stan fizjologiczny populacji brudnicy mniszki, warunki meteorologiczne. Nie mniej istotne czynniki wpływające na stan populacji owada – stan populacji owadów pasożytniczych, aktywność słońca, zwalczanie – nie zostały włączone do

modelu, z powodu braku odpowiednich danych. W badanym okresie czasowym na terenie nadleśnictwa testowego nie wystąpiło zagrożenie, a więc nie były też przeprowadzone zabiegi zwalczające. Wychodząc ze spiralnego modelu lasu [2], odporność drzewostanu została określona przez takie zmienne jak: typ siedliskowy lasu, typ gleby, wystawę i nachylenie terenu. Dane meteorologiczne to: średnia miesięczna temperatura i miesięczna suma opadów w kwietniu i maju. Stan fizjologiczny populacji brudnicy mniszki jest określony przez liczbę odłowionych motyli brudnicy mniszki do pułapek feromonowych. Typ gleby i siedliskowy typ lasu dla testowanych wydzieleń zostały odczytane z bazy danych SILP. Nachylenie i wystawa terenu były uzyskane z warstw informacyjnych utworzonych w SIP z modelu numerycznego terenu dla nadleśnictwa. Dane o liczbie owadów brudnicy mniszki odłowionych w poszczególnych latach do

pułapek feromonowych oraz miejsca ich rozmieszczenia zostały udostępnione przez nadleśnictwo testowe. Dane meteorologiczne pozyskane były z Instytutu Meteorologii i Gospodarki Wodnej (IMiGW). Wartości temperatury zostały zmierzone w dwóch, a wartości opadów w trzech stacjach/posterunkach meteorologicznych rozmieszczonych w okolicach nadleśnictwa testowego. Warunki meteorologiczne są bardzo istotnymi czynnikami warunkującymi rozwój populacji owadów, dlatego należało określić bardziej dokładne wartości tych danych w każdym z wybranych wydzieleń. W tym celu w SIP została utworzona warstwa punktowa pomiarów meteorologicznych i przeprowadzona interpolacja/ekstrapolacja metodą odwrotnych odległości IDW znanych wartości danych meteorologicznych na obszar całego nadleśnictwa. Przepływ danych przedstawia rys. 1.



Rys. 1. Schemat przepływu danych.

Preprocessing danych wykorzystanych do uczenia sieci neuronowych dotyczył: ich kompletności i reprezentatywności pod względem zakresu czasowego (1996-2000), gatunku panującego (So), wieku (40-80 lat), braku zniszczeń pułapek i wartości zerowych. W ostateczności liczba przypadków wynosiła 168. Każdy z nich zawierał takie dane, jak: średnia miesięczna temperatura i suma miesięcznych opadów z kwietnia i maja, typ siedliskowy lasu, typ gleby, wystawa, nachylenie terenu i liczba motyli brudnicy mniszki odłowionych do pułapek feromonowych w roku poprzedzającym rok prognozowany. Zmienna wyjściowa, czyli zmienna prognozowana, to liczba motyli brudnicy mniszki w roku bieżącym.

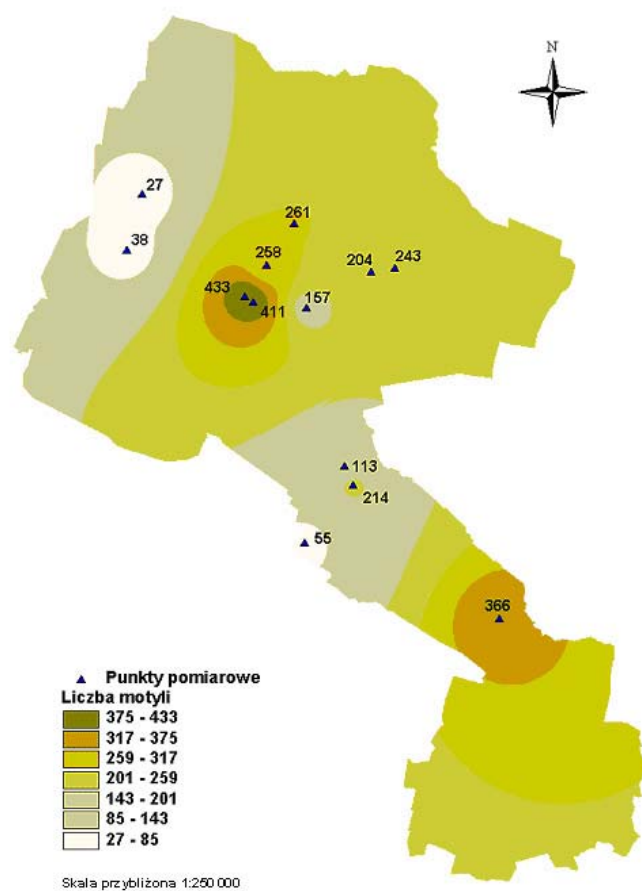
Nauczanie lub trenowanie sieci neuronowej polega na obliczeniu takich wewnętrznych wartości sieci o zadanej architekturze i parametrach uczących, żeby mogła ona jak najlepiej dopasować zaprezentowane jej dane wejściowe (przesłanki) do danych wyjściowych (prognoza). Im dopasowanie jest lepsze, tym mniejszy jest błąd sieci, a większa jej dobroć, i tym dokładniejsze będą oszacowania danych testowych. W obecnych badaniach wykorzystany został program *STATISTICA Sieci Neuronowe* firmy StatSoft. Program udostępnia kilka wskaźników oceny dobroci zbudowanego modelu neuronowego. Dla zagadnień regresyjnych, do których należy badany problem, takimi wskaźnikami są: błąd RMS, iloraz odchylenia, korelacja, rozkład wartości wag, analiza wrażliwości. Do oceny

dobroci modelu i, co za tym idzie, większej trafności prognozy wykorzystano wszystkie ww. wskaźniki. Średniokwadratowy błąd RMS określa przeciętne odchylenie wartości teoretycznych (obliczonych przez sieć) od wartości rzeczywistych. Za najlepsze wskaźniki dobroci nauczonej sieci uważa się iloraz odchylenia standardowego błędów i odchylenia standardowego danych oraz korelację między wartościami rzeczywistymi i obliczonymi przez sieć. Im iloraz odchylenia jest bliższy wartości 0, a korelacja jedności, tym dokładniej prognozuje nauczona sieć. Sieci dobrze nauczonej odpowiadają wykresy słupkowe wskazujące na wyraźną dywersyfikację wartości wag (wykresy wielomodalne). Natomiast sieci źle nauczone (lub nauczone niecałkowicie) charakteryzują się równomiernym rozkładem wartości wag. Analiza wrażliwości z kolei umożliwia ocenę wagi każdej z włączonych do modelu zmiennych, z którą oddziałują one na wynik końcowy.

Zbiór przypadków został podzielony na: podzbiór uczący (130 przypadków), podzbiór walidacyjny (32 przypadki) i podzbiór testowy (6 przypadków). W trakcie uczenia sieci wykorzystywany jest podzbiór uczący. Zbiór testowy służy do oceny dobroci nauczonej sieci. W wyniku przeprowadzonych prób za najlepszą uznano sieć MLP z dwoma warstwami ukrytymi. W uczeniu sieci zastosowano algorytm wstecznej propagacji błędu. Dla podzbioru testowego otrzymano następujące wartości wskaźników dobroci dopasowania: błąd RMS = 17,7; iloraz odchylenia = 0,23; korelacja = 0,97; histogram wartości wag (inaczej rozkład wag) nie jest rozkładem normalnym o wartości średniej 0. Analiza wrażliwości pokazała, że najmniejsze znaczenie (wagę) w prognozie ma zmienna określająca temperaturę w miesiącu maju. Dlatego ostateczny model sieci neuronowej nie uwzględnia tej zmiennej.

Prognozowana przez sieć neuronową liczba owadów dotyczy tylko drzewostanów, w których zostały wyłożone pułapki feromonowe. Aby uzyskać prognozę liczby spodziewanych owadów w pozostałych wydzieleniach, przeprowadzono interpolację wartości obliczonych przez sieć neuronową. W wyniku tej operacji, wykonanej za pomocą SIP, otrzymano warstwę informacyjną rozkładu prawdopodobnego występowania motyli brudnicy mniszki na całym analizowanym obszarze. Dystrybucję przes-

trzną prognozy uzyskanej za pomocą SN po interpolacji w SIP podano na rys. 2.



Rys. 2. Przestrzenna dystrybucja populacji motyli brudnicy mniszki – wynik interpolacji prognozowanych przez SN wartości.

Prognozę sporządzono dla roku, dla którego istniały dane rzeczywiste. Weryfikacji trafności prognozowania przez SN dokonano poprzez porównanie danych rzeczywistych z danymi z prognozy (podzbiór testowy). Różnica między tymi danymi wyniosła średnio 7,27%. Wielkość błędu tego rzędu została uznana przez naukowców z dziedziny ochrony lasu jako mała i tym samym dokładność prognozowania sieci neuronowej jako zadowalająca.

4 PODSUMOWANIE

Analiza wskaźników dobroci nauczonej sieci wskazuje na wiarygodność modelu prognozowania zagrożenia brudnicą mniszką. Z punktu widzenia potencjalnych możliwości sieci neuronowych uzyskany model dla danego problemu nie może być jednak uznany za ostateczny i w peł-

ni kompletny. 130 przypadków, na podstawie których sieć neuronowa zbudowała model zagadnienia, to za mało. Liczba kombinacji możliwych do ułożenia z ośmiu zmiennych jest znacznie większa. Aby uznać wyniki prognozy sieci za w pełni wiarygodne, dane uczące muszą opisywać wszystkie przypadki, które mogą wystąpić w rzeczywistości. Wykorzystanie danych z okresu np. 8-10 lat dla 20-25 wydzieleń mogło by zapewnić wystarczającą ilość informacji, by sieć nauczyła się i rozpoznała większość przypadków. Problem tu występujący jest bardziej problemem kompletności danych i rzetelności ich zbierania, niż problemem ograniczeń samych sieci neuronowych. Dane przeznaczone do uczenia sieci, czyli takie, na podstawie których będzie zbudowany model neuronowy zagadnienia, powinny być jak najpełniejsze i reprezentatywne. Dysponując dobrym modelem, można skutecznie prognozować, korzystając nawet z danych niekompletnych lub niepewnych.

Wykorzystane do uczenia sieci neuronowej dane wyznaczają pole wiarygodnego zastosowania sieci. Wykorzystane w badaniu dane dotyczyły drzewostanów sosnowych w wieku 40-80 lat na terenie przeważnie równinnym i rosnących na wybranych siedliskach. Oznacza to, że prognoza sieci będzie reprezentatywna tylko dla tak opisanych drzewostanów i nie można uznać za poprawną prognozę zagrożenia w drzewostanach w innym wieku niż 40-80 lat, a tym bardziej w terenie górzystym. Dla terenów o innych warunkach meteorologicznych i siedliskowych należało by pozyskać niezbędne dane i zbudować nowy model neuronowy (nauczyć nową sieć), który byłby reprezentatywny dla tego terenu.

Systemy Informacji Przestrzennej mogą być skutecznie wspomagane *Sieciami Neuronowymi* w rozwiązywaniu nieliniowych i słabo sformalizowanych problemów leśnictwa. Zastosowany w badaniu rodzaj współpracy polegający na wymianie plików nie jest jedynym możliwym (a na pewno nie jest najbardziej wygodnym dla użytkownika), lecz wystarczającym do realizacji założeń celu badawczego. Sztuczne sieci neuronowe, tak jak i SIP, są programami komputerowymi, które ciągle są udoskonalane i rozwijane. Podobnie jak oprogramowanie dla SIP, programy – symulatory SN mają interfejs programisty (API). Korzystając z niego, można stworzyć prosty i przyjazny interfejs, nie wprowadzając użytkownika w szczegóły techniczne, jedno-

cznie realizując elastyczną i efektywną współpracę obu narzędzi.

BIBLIOGRAFIA

- 1) Kolk A., Ślusarski S., Sukovata L. 2000. Prognozowanie występowania ważniejszych szkodników sosny – stan obecny i potrzeby badawcze. (w:) Malinowski H. (red.) *Stan i perspektywy badań z zakresu ochrony lasu*. Warszawa. IBL.
- 2) Manion P. 1981. *Tree disease concepts*. Printice Hall Inc., New York.
- 3) *STATISTICA Neural Networks PL*. 2001. *Dokumentacja do programu*. StatSoft Polska Sp. z o.o.
- 4) Tracz W. 2002. *Analiza obszarów leśnych w Systemach Informacji Przestrzennej wspomaganych Systemami Ekspertowymi i Sztucznymi Sieciami Neuronowymi*. Praca doktorska. Warszawa. Wydział Leśny SGGW.
- 5) Zieliński J. 2000. *Inteligentne systemy w zarządzaniu. Teoria i praktyka*. Warszawa. WNT.