

Univerza
v Ljubljani

Fakulteta za
*gradbeništvo in
geodezijo*



Jamova 2
1000 Ljubljana, Slovenija
telefon (01) 47 68 500
faks (01) 42 50 681
fgg@fgg.uni-lj.si

UNIVERZITETNI ŠTUDIJ
GRADBENIŠTVA
KONSTRUKCIJSKA SMER

Kandidat:

JAN DOBNIKAR

**UPORABA METOD STROJNEGA UČENJA ZA
RAZVRŠČANJE LESA V TRDNOSTNE RAZREDE**

Diplomska naloga št.: **3216/KS**

**THE USE OF MACHINE LEARNING METHODS FOR
THE GRADING OF TIMBER STRENGTH**

Graduation thesis No.: **3216/KS**

Mentor:
prof. dr. Goran Turk

Predsednik komisije:
izr. prof. dr. Janko Logar

Somentor:
doc. dr. Vlado Stankovski

Ljubljana, 20. 4. 2012

IZJAVE

Podpisani **JAN DOBNIKAR** izjavljam, da sem avtor diplomske naloge z naslovom:
»Uporaba metod strojnega učenja za razvrščanje lesa v trdnostne razrede«.

Izjavljam, da je elektronska različica v vsem enaka tiskani različici.

Izjavljam, da dovoljujem objavo elektronske različice v repozitoriju UL FGG.

Ljubljana, 01. 04. 2012

(podpis)

ZAHVALA

Hvala

vsem prijateljem, ki ste me cele dneve vlekli stran od študijskih obveznosti, ter atiju Alešu in mami Meti, ki sta me istočasno vlekla nazaj. Ko se ozrem nazaj, bom tako lepa leta, kot ste mi naredili ta študentska, težko presegel. Posebna zahvala gre tudi absolutno najboljšima profesorjema na faksu, Goranu Turku in Vladu Stankovskemu.

Krainerju Alešu gre zahvala za vzpodbujanje študentov k intenzivnemu učenju novih področij in širjenju obzorij. Hvala tudi vsem sošolcem za pomoč pri študiju, rad bi se zahvalil še posebej Eriki, Tanji, Mihi in Urošu, brez vaših zapiskov bi cele dneve preživel na predavanjih. Zahvalil bi se tudi Rokotu za rože, ter moji Teji, ki me je uspešno in vztrajno bodrila ob zaključku študijske ere.

BIBLIOGRAFSKO – DOKUMENTACIJSKA STRAN IN IZVLEČEK

UDK:	004.85:624.011.1(043.2)
Avtor:	Jan Dobnikar
Mentor:	prof. dr. Goran Turk, univ.dipl.inž.grad.
Somentor:	doc. dr. Vlado Stankovski, univ.dipl.inž.rač. in inf.
Naslov:	Uporaba metod strojnega učenja za razvrščanje lesa v trdnostne razrede
Obseg in oprema:	63 str., 53 preg., 23 sl., 11 en.
Ključne besede:	trdnostni razred, les, dinamični odziv, elastični modul, strojno učenje, regresijsko drevo, nevronska mreža

Izвлеček:

Diplomska naloga obravnava problem razvrščanja lesa v trdnostne razrede. Vsak material, uporabljen v konstrukciji, mora biti razvrščen v trdnostni razred in ustrezno označen z oznako CE. Za razvrstitev v določen trdnostni razred mora izpolnjevati minimalne zahteve tega razreda glede mehanskih karakteristik. Les je nehomogen anizotropen material. Trdnost lesa znotraj posamezne vrste je med drugim odvisna tudi od dela debla, s katerega odrežemo preizkušaneec, od nahajališča, okolice, prsti, starosti, poškodovanosti, grčavosti itd. Edini način za točno določitev trdnosti lesa je porušitev preizkušanca. Trenutno je v Sloveniji uveljavljeno le vizualno ocenjevanje lesa, ki pa je zelo konzervativno, zato je veliko dobrega lesa prodanega pod zasluženo ceno. V nalogi smo z uporabo metod strojnega učenja razvili sistem, ki je osnova za boljše razvrščanje lesenih elementov v razrede. Analizirano je bilo več kot 5000 lesenih preizkušancev, na katerih so bile opravljene tako neporušne kot porušne preiskave. Rezultati preiskav so bili obdelani s programoma WEKA Toolkit in IBM SPSS 20. Poleg osnovnih statističnih testov so bile za obdelavo uporabljene nevronske mreže in regresijska drevesa. Pokazali smo, da je bil izbrani vzorec dovolj velik, ter da so nevronske mreže primerno orodje za določevanje parametrov za klasificiranje lesa na osnovi rezultatov neporušnih analiz.

BIBLIOGRAPHIC – DOCUMENTALISTIC INFORMATION

UDC:	004.85:624.011.1(043.2)
Author:	Jan Dobnikar
Supervisor:	prof. dr. Goran Turk, univ.dipl.inž.grad.
Co-supervisor:	doc. dr. Vlado Stankovski, univ.dipl.inž.rač. in inf.
Title:	The use of machine learning methods for the grading of timber strength
Notes:	63 p., 53 tab., 23 fig., 11 eq.
Key words:	strength class, timber, dynamic response, module of elasticity, machine learning, regression tree, artificial neural network

Abstract

This work addresses the problem of grading timber into strength classes. All materials, used in the construction, have to be graded and marked with a CE mark. For the classification into a certain strength class, the material has to adhere to minimal mechanical requirements of the class. Timber is an anisotropic, non-homogeneous material. The timber strength within a certain species depends, among other factors, on the part of the log that the sample has been cut off, the environment of its growth, the type of soil, the timber age, the damage to the timber, the presence of knots etc. The only accurate method for determining timber strength is destructive testing. Presently, the only method considered relevant in Slovenia is the visual evaluation of the timber, which is a very conservative method and results in lower prices of good quality timber. The goal of the presented work was to develop a system for timber grading on the basis of results of non-destructive methods, particularly by using the machine learning methods. The analysis sample consists of more than 5000 timber samples, which have been tested with non-destructive, as well as destructive methods. Results of the performed tests were analysed by using the machine learning algorithms contained in the WEKA Toolkit and the statistical methods of the IBM's SPSS 20 program. In order to analyze the collected data, standard statistical tests were performed for the analysis and artificial neural networks, as well as regression trees, were also used. The presented results indicate that the number of the samples was sufficient and that the formalism of the artificial neural networks is the appropriate tool for determining the timber grading parameters on the basis of results of non destructive testing methods.

Kazalo vsebine

1 UVOD	1
2 PREIZKUŠANJE LESENIH VZORCEV	2
2.1 Splošno	2
2.1.1 Namen preiskav	2
2.1.2 Vzorec preizkušancev	3
2.1.3 Vhodni in izhodni podatki analize	3
2.2 Dinamične preiskave	6
2.2.1 Metoda frekvenčnega odziva	6
2.2.2 Vzdolžna impulzna motnja	8
2.3 Vizualni pregled	10
2.4 Globalni modul elastičnosti	13
2.5 Porušne preiskave	15
2.6 Statistična analiza vzorcev	16
3 STROJNO UČENJE	19
3.1 Teoretične osnove	19
3.1.1 Ideja in zasnova	19
3.1.2 Kratek pregled algoritmov in metod strojnega učenja	21
3.2 Izbira in opis algoritmov	22
3.2.1 Regresijska drevesa	22
3.2.1.1 Splošno	22
3.2.1.2 Parametri za izračun	23
3.2.2 Umetne nevronske mreže	25
3.2.2.1 Splošno	25
3.2.2.2 Parametri za izračun	30
3.3 Izbira orodja - WEKA Toolkit	32
3.3.1 Splošno	32
3.3.2 Regresijska drevesa in nevronske mreže	33
3.3.2.1 Regresijska drevesa	33
3.3.2.2 Nevronske mreže	35
4 REZULTATI	36
4.1 Porazdelitve vrednosti	36
4.2 Linearna regresija	38
4.3 Rezultati nevronskih mrež	42
4.4 Rezultati regresijskih dreves	49

4.5 Determinacijski koeficienti R^2	56
4.6 Preglednica vpliva parametrov	58
4.7 Razvrščanje vzorcev lesa v trdnostne razrede	60
5 ZAKLJUČEK	61
VIRI	62

Kazalo slik

Slika 1: Načini prostoležečega podpiranja preizkušanca.....	6
Slika 2: Primer spektra odziva	6
Slika 3: Nihalo z eno prostorsko stopnjo	7
Slika 4: Časovni odziv sistema na pulz.....	7
Slika 5: Shema merjenja longitudinalnega širjenja valov	8
Slika 6: Shema delovanja enosmernega ultrazvočnega sistema	8
Slika 7: Merjenje širine branike.....	10
Slika 8: Vizualno merjenje zavitosti, loka in sablje	11
Slika 9: Razpokan les	11
Slika 10: Vizualno merjenje grč	12
Slika 11: Shema preizkušanja za določitev globalnega modula elastičnosti	13
Slika 12: Merjenje modula elastičnosti v tekočem traku	13
Slika 13: Primer generaliziranega in prenaučnega sistema.....	20
Slika 14: Regresijsko drevo za trdnost bora pri nateznem preizkusu	23
Slika 15: Biološki nevron.....	25
Slika 16: Umetni nevron	26
Slika 17: Pragovna funkcija.....	27
Slika 18: Linearna funkcija	27
Slika 19: Eksponentna sigmoidna funkcija	28
Slika 20: Tangentna sigmoidna funkcija	28
Slika 21: Shema nevronske mreže.....	29
Slika 22: Program weka toolkit.....	32
Slika 23: Razvrščanje v razrede	60

Kazalo preglednic

Preglednica 1: Poreklo preizkušancev	3
Preglednica 2: Vrste preizkušancev	3
Preglednica 3: Vhodni in izhodni podatki	4
Preglednica 4: Razvrščanje lesa po standardu Evrokod 5, po [12]	5
Preglednica 5: Celoten vzorec	16
Preglednica 6: Bor, upogibni preizkus	17
Preglednica 7: Bor, natezni preizkus	17
Preglednica 8: Smreka, upogibni preizkus	18
Preglednica 9: Smreka, natezni preizkus	18
Preglednica 10: Bor – upogib, porazdelitev trdnosti in MOE	36
Preglednica 11: Bor – nateg, porazdelitev trdnosti in MOE	36
Preglednica 12: Smreka – upogib, porazdelitev trdnosti in MOE	37
Preglednica 13: Smreka – nateg, porazdelitev trdnosti in MOE	37
Preglednica 14: Linearna regresija, bor, upogibni preizkus, MOE	38
Preglednica 15: Linearna regresija, bor, upogibni preizkus, trdnost	38
Preglednica 16: Linearna regresija, bor, natezni preizkus, MOE	39
Preglednica 17: Linearna regresija, bor, natezni preizkus, trdnost	39
Preglednica 18: Linearna regresija, smreka, upogibni preizkus, MOE	40
Preglednica 19: Linearna regresija, smreka, upogibni preizkus, trdnost	40
Preglednica 20: Linearna regresija, smreka, natezni preizkus, MOE	41
Preglednica 21: Linearna regresija, smreka, natezni preizkus, trdnost	41
Preglednica 22: Parametri nevronske mreže, bor, upogibni preizkus, MOE	42
Preglednica 23: Parametri nevronske mreže, bor, upogibni preizkus, MOE 12 %	43
Preglednica 24: Parametri nevronske mreže, bor, upogibni preizkus, trdnost	43
Preglednica 25: Parametri nevronske mreže, bor, upogibni preizkus, trdnost 12 %	44
Preglednica 26: Parametri nevronske mreže, bor, natezni preizkus, MOE 12 %	44
Preglednica 27: Parametri nevronske mreže, bor, natezni preizkus, trdnost	44
Preglednica 28: Parametri nevronske mreže, bor, natezni preizkus, trdnost 12 %	45
Preglednica 29: Parametri nevronske mreže, smreka, upogibni preizkus, MOE	45
Preglednica 30: Parametri nevronske mreže, smreka, upogibni preizkus, MOE 12 %	46
Preglednica 31: Parametri nevronske mreže, smreka, upogibni preizkus, trdnost	46
Preglednica 32: Parametri nevronske mreže, smreka, upogibni preizkus, trdnost 12 %	46
Preglednica 33: Parametri nevronske mreže, smreka, natezni preizkus, MOE	47
Preglednica 34: Parametri nevronske mreže, smreka, natezni preizkus, MOE 12 %	47
Preglednica 35: Parametri nevronske mreže, smreka, natezni preizkus, trdnost	48
Preglednica 36: Parametri nevronske mreže, smreka, natezni preizkus, trdnost 12 %	48
Preglednica 37: Parametri regresijskih dreves, bor, upogibni preizkus, MOE	49
Preglednica 38: Parametri regresijskih dreves, bor, upogibni preizkus, MOE 12 %	50
Preglednica 39: Parametri regresijskih dreves, bor, upogibni preizkus, trdnost	50
Preglednica 40: Parametri regresijskih dreves, bor, upogibni preizkus, trdnost 12 %	51
Preglednica 41: Parametri regresijskih dreves, bor, natezni preizkus, MOE 12 %	51
Preglednica 42: Parametri regresijskih dreves, bor, natezni preizkus, trdnost	52
Preglednica 43: Parametri regresijskih dreves, bor, natezni preizkus, trdnost 12 %	52
Preglednica 44: Parametri regresijskih dreves, smreka, upogibni preizkus, MOE	52
Preglednica 45: Parametri regresijskih dreves, smreka, upogibni preizkus, MOE 12 %	53
Preglednica 46: Parametri regresijskih dreves, smreka, upogibni preizkus, trdnost	53
Preglednica 47: Parametri regresijskih dreves, smreka, upogibni preizkus, trdnost 12 %	53
Preglednica 48: Parametri regresijskih dreves, smreka, natezni preizkus, MOE	54
Preglednica 49: Parametri regresijskih dreves, smreka, natezni preizkus, MOE 12 %	54

Preglednica 50: Parametri regresijskih dreves, smreka, natezni preizkus, trdnost	54
Preglednica 51: Parametri regresijskih dreves, smreka, natezni preizkus, trdnost 12 %	55
Preglednica 52: Determinacijski koeficienti R^2	56
Preglednica 53: Vpliv parametrov	59

1 UVOD

V diplomski nalogi sem obravnaval problem razvrščanja lesa v trdnostne razrede.

Vsak material, uporabljen v konstrukciji, mora biti razvrščen in označen z oznako CE. Za razvrstitev v določen trdnostni razred mora izpolnjevati minimalne zahteve tega razreda glede mehanskih karakteristik. Te vrednosti projektant upošteva pri statičnem računu. Dobra razvrstitev tako pomeni trdno in ekonomično gradnjo.

Les pa je nehomogen anizotropen material. Za razliko od betona in jekla, kjer lahko trdnost materiala dobro predvidimo izkustveno z znanimi recepturami priprave, drevesa le redko zrastejo v kontroliranih okoliščinah. Trdnost lesa znotraj posamezne vrste je tako odvisna že od dela debla, s katerega odrežemo preizkušane, kaj šele od nahajališča, okolice, prsti, starosti, poškodovanosti, grčavosti itd. Preprosta metoda določevanja trdnosti na način, da iz enega debla izločimo in porušimo dve deski, dobljeno trdnost pa apliciramo na celotno deblo, ne pride v poštev.

Še vedno je edini način za točno določitev trdnosti lesa porušitev preizkušanca. Uveljavile pa so se nekatere neporušne metode, n.pr. ugotavljanje hitrosti širjenja zvoka, iskanje lastne frekvence, vizualno ocenjevanje ipd., s katerimi lahko trdnost lesa ocenimo. Opisane so v poglavju 2. Več kot opravimo neporušnih preiskav, bolj zanesljivo informacijo o posameznem preizkušancu dobimo. Trenutno je v Sloveniji uveljavljeno le vizualno ocenjevanje lesa, ki pa je zelo konzervativno, zato je veliko dobrega lesa prodanega pod zasluženo vrednostjo. Cilj naloge je ugotoviti, kako se na osnovi ostalih neporušnih preiskav lesene preizkušance čim natančneje razvršča v razrede.

V diplomski nalogi bom analiziral več kot 5000 lesenih preizkušancev, na katerih so bile opravljene tako neporušne kot porušne preiskave. Iskal bom čim boljše korelacijo med porušnimi in neporušnimi preiskavami. Vsi podatki tvorijo veliko bazo podatkov, ki jo bom statistično analiziral. Produkt analize bo sistem, po katerem bomo lahko preizkušancu, na katerem so bile opravljene le neporušne preiskave, določili trdnost in modul elastičnosti, ter ocenili napako. Z neporušnimi preiskavami se namreč lahko oceni vsak konstrukcijski kos lesa, ki zapusti lesno predelovalni obrat.

Možnih statističnih pristopov k analizi je veliko. Izbral sem nekaj najbolj uveljavljenih. Izračunal bom medsebojno korelacijo vseh parametrov ter določil približno rešitev s pomočjo linearne regresije. Za pridobitev bolj natančnih rezultatov pa bom uporabil algoritme strojnega učenja. Izbral sem metodi umetnih nevronske mreže ter regresijskih dreves. Metodi sta opisani v poglavju 3.

Dober sistem se odraža z visokim determinacijskim koeficientom. Določi se ga s testiranjem sistema na preizkušancih, ki v vzpostavitev sistema niso bili vključeni. Determinacijski koeficienti so predstavljeni v poglavju 4.

2 PREIZKUŠANJE LESENIH VZORCEV

2.1 Splošno

2.1.1 Namen preiskav

Cilj preiskav lesenih vzorcev je določiti mehanske lastnosti preizkušancev ter jih razvrstiti v trdnostne razrede. Bolj kot so metode določevanja natančne, manjši raztros imajo rezultati, posledično pa je več lesa pravilno razvrščenega v višje razrede. Tako je les izkoriščen v konstruktivnem in ekonomskem smislu.

Les je anizotropen naraven material, čigar mehanske lastnosti so odvisne ne le od smeri obremenitve, temveč tudi od trenutne vlažnosti, trajanja obremenitve, lokalnih nepravilnosti (grč), razmaka med letnicami (gostote) itd. Proizvodnja lesne mase se, za razliko od jeklenih ali betonskih elementov, ne vrši v kontroliranem okolju. Zato se mehanske lastnosti preizkušancev med seboj razlikujejo znotraj iste vrste z istega območja pogozditve, pa tudi znotraj posameznega debla. Zaradi tako velikega raztrosa mehanskih lastnosti je potrebno vsak kos lesa pred vgradnjo pregledati. Optimalen nadzor kakovosti bi bil natančno določiti trdnost vsakega elementa, vendar se trdnost lahko ugotovi le z destruktivnimi preiskavami, torej s poružitvijo elementa.

V pomoč so nam nedestruktivne preiskave, s katerimi lahko določimo nekatere karakteristike posameznega preizkušanca. Na podlagi teh sklepamo o trdnosti elementa, kar je eden od osnovnih kriterijev za razvrstitev v trdnostne razrede. Z nedestruktivnimi preiskavami, opisanimi v poglavju 2.2, lahko v lesno predelovalnem obratu zelo hitro pregledamo celotno zalogo lesa. Problem pa je, da korelacija med nedestruktivnimi ter porušnimi preiskavami ni popolna. Za čim boljšo povezavo med podatki smo v diplomski nalogi uporabili nevronske mreže in regresijska drevesa, ki so že uveljavljene metode statističnih analiz.

2.1.2 Vzorec preizkušancev

Za analizo smo imeli na voljo vzorec s 5231 preizkušanci. Na vseh so bili izvedeni najprej nedestruktivni testi, nato pa še destruktivni. Poreklo preizkušancev je 22 rastišč iz desetih držav, ki so prikazane v preglednici 1:

Preglednica 1: Poreklo preizkušancev

Država	Število vzorcev
Švica	441
Finska	252
Francija	238
Poljska	1087
Romunija	401
Rusija	107
Švedska	836
Slovenija	1266
Slovaška	198
Ukrajina	405

Preizkušanci so bili vrste bor in smreka. Porušne preiskave so bile na nekaterih vzorcih narejene upogibno, na drugih pa natezno. Celoten vzorec smo zato razdelili v štiri skupine, število vzorcev v posamezni skupini je prikazano v preglednici 2:

Preglednica 2: Vrste preizkušancev

vrsta lesa	način preizkušanja	
	natezno	upogibno
bor	1085	428
smreka	1478	2311

Za nastavitev stroja za avtomatično razvrščanje lesa potrebujemo po standardu vsaj 900 preizkušancev. Iz preglednice 2 je razvidno, da je v našem primeru preizkušancev dovolj v vseh skupinah, razen pri upogibnem obremenjevanju preizkušancev bora.

Veliko število rastišč je vzrok za raznolikost vzorca, torej povečuje raztros rezultatov. Korelacijski koeficienti so zato nekoliko manjši, kot če bi za vzorec vzeli drevesa z le enega rastišča. Zato pa kot produkt obširnejše analize dobimo splošna pravila za razvrščanje preizkušancev določene vrste drevesa v trdnostne razrede. Če želimo v sistem razvrščanja vključiti tudi značilnosti, specifične za posamezna rastišča, moramo bodisi lokacijo preizkušanca vnesti kot vhodni podatek, bodisi bazo podatkov fizično ločiti ter analizirati vsak del posebej.

2.1.3 Vhodni in izhodni podatki analize

V izbranih statističnih analizah se za vsak preizkušanec vnese nabor vhodnih ter izhodnih podatkov, rezultat analize pa so računske povezave med njimi. Vhodni podatki so vsi podatki, ki jih lahko hitro pridobimo v lesno predelovalnem obratu, torej osnovne

geometrijske karakteristike, teža, vsebnost vlage ter rezultati nedestruktivnih preiskav. Izhodna podatka sta globalni modul elastičnosti, ki ga je v primerjavi z ostalimi nedestruktivnimi preiskavami zamudno izmeriti, ter trdnost, ki je dobljena s porušno preiskavo. Podatki so predstavljeni v preglednici 3:

Preglednica 3: Vhodni in izhodni podatki

	Ime	Opis
Vhodni podatki	Width	Širina preizkušanca v [mm]
	Thickness	Debelina preizkušanca v [mm]
	Length	Dolžina preizkušanca v [mm]
	Weight	Teža preizkušanca v [g]
	Moisture	Vsebnost vlage v preizkušancu [%]
	Frequency	Lastna frekvenca preizkušanca v [Hz]
	Running_time	Čas potovanja zvoka po preizkušancu v [μ s]
	tKAR	Delež grč v prerezu [-]
Izhodni podatki	gMOE	Globalni modul elastičnosti preizkušanca v [N/mm^2]
	gMOE_12	Globalni modul elastičnosti preizkušanca, preračunan na 12% vlažnosti, v [N/mm^2]
	Strength	Porušna trdnost preizkušanca v [N/mm^2]
	Strength_12	Porušna trdnost preizkušanca, preračunana na 12% vlažnosti, v [N/mm^2]

Namen naloge je pripraviti matematični model, ki ustrezno opisuje odvisnost vhodnih in izhodnih podatkov. Ker smo k problemu pristopili tudi z uporabo regresijskih dreves (poglavje 3.2.1), lahko za vhodne podatke uporabimo vseh osem spremenljivk, za izhodne pa le po eno naenkrat. Tako je potrebno za eno vrsto drevesa in eno vrsto porušne obremenitve izdelati dve drevesni strukturi odločanja (ter še dve za podatke, ki so preračunani na vsebnost vlage 12%). Nezmožnost upoštevanja vseh izhodnih podatkov naenkrat ne ovira funkcionalnosti sistema, saj so za potrebe razvrščanja lesa v trdnostne razrede trdnost, modul elastičnosti ter gostota preizkušanca trije neodvisni parametri. Pogoji za razvrščanje so prikazani v preglednici 4.

Preglednica 4: Razvrščanje lesa po standardu SIST EN 338:2010 [1]

	Karakteristične vrednosti po razredih									
	C14	C16	C18	C20	C22	C24	C27	C30	C35	C40
Upogib [N/mm ²]	14	16	18	20	22	24	27	30	35	40
Nateg vzp. z vlakni [N/mm ²]	8	10	11	12	13	14	16	18	21	24
Modul elastičnosti [kN/mm ²]	7	8	9	9.5	10	11	11.5	12	13	14

V splošnem nevronske mreže (poglavje 3.2.2) sicer omogočajo vnos več izhodnih podatkov naenkrat, vendar se za njihovo uporabo nisem odločil. Programsko orodje, v katerem je delo izvedeno, WEKA Toolkit (poglavje 3.3), v času izdelave diplomske naloge arhitekture nevronskih mrež z več izhodnimi podatki ni omogočalo. Drugi razlog pa je primerljivost koeficientov korelacije med nevronskimi mrežami in regresijskimi drevesi. Trdnost ima precej večji raztros kot modul elastičnosti, zato so korelacijski koeficienti nekoliko nižji. Rezultatov analize, kjer nastopata oba izhodna podatka, se zato ne da neposredno primerjati z rezultati analize, kjer nastopa le eden.

Najprej smo vključili podatke iz vseh držav (preglednico 1) v isto analizo. Naknadno pa smo v naboru podatkov pustili le preizkušance slovenskega porekla. Namen tega je bil izdelati statistični odločitveni model z nekoliko višjo natančnostjo, uporaben za preizkušance s področja Slovenije. Ta statistični vzorec vsebuje 1060 smrekovih preizkušancev s treh območij: z Notranjske, Osrednjeslovenske ter Koroške.

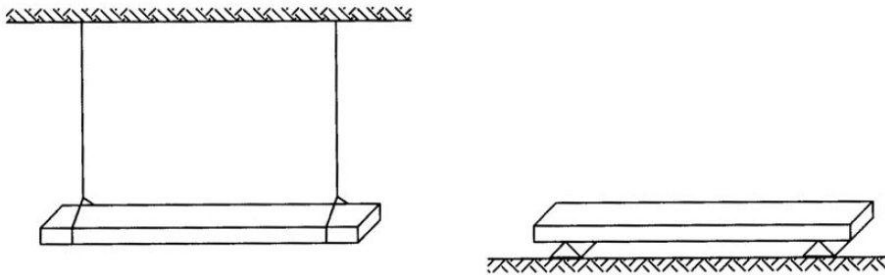
Ker so bili testi izvedeni v različnih državah, je bil predhodno opravljen pregled vseh sodelujočih laboratorijev. Ugotovljeno je bilo, da bistvenih razlik v funkcionalnosti opreme ni, zato je možno preizkušance primerjati med sabo in jih statistično analizirati skupaj. 78 preizkušancev smo morali iz analize izločiti zaradi pomanjkanja podatkov, zaradi neprimernosti preizkušanca za analizo, ali pa zaradi tehničnih težav med testiranjem.

2.2 Dinamične preiskave

2.2.1 Metoda frekvenčnega odziva

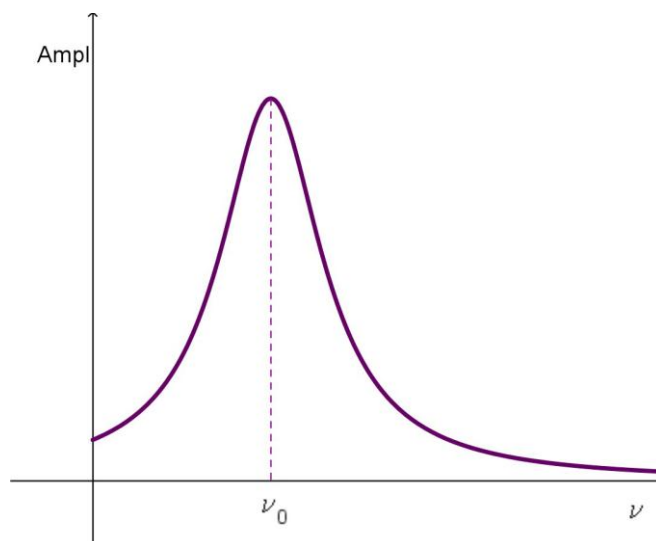
Ena od metod za oceno elastičnega modula preizkušanca je metoda frekvenčnega odziva. Cilj metode je ugotoviti lastno (oziroma resonančno) frekvenco preizkušanca. Teoretična povezava med lastno frekvenco in elastičnim modulom je odvisna še od ploščine in vztrajnostnega momenta prereza, gostote preizkušanca, razdalje med podporami ter načina podpiranja.

Ločimo dva načina ugotavljanja lastne frekvence preizkušanca. Pri obeh je pomembno, da je preizkušanec vpet čim bolj idealno, kar pomeni, da morajo podpore čim bolj simulirati idealizirano računsko podporo. Najlažje simuliramo prosto podprte sisteme. V laboratoriju se preizkušanec velikokrat na koncih obesi z laksom, lahko pa ga tudi položimo na podpore iz mehke gume ali pene, kot je prikazano na sliki 1. V splošnem bi lahko testirali preizkušanec, vpet na kakršenkoli način, vendar je zagotavljanje idealnih robnih pogojev v vseh ostalih primerih težje.

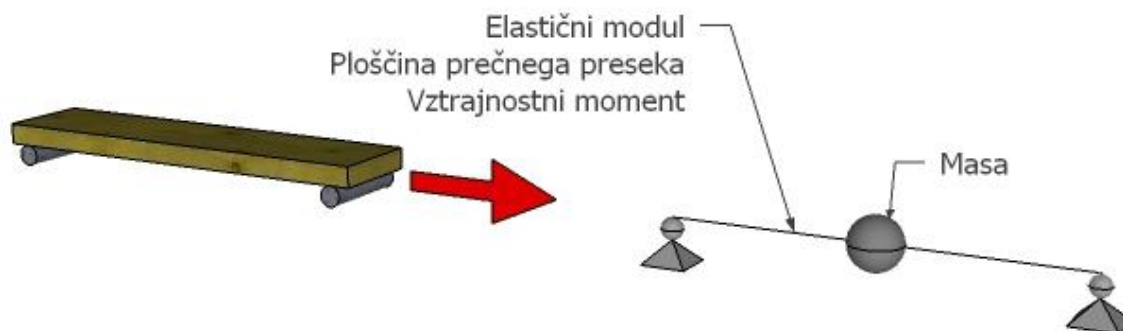


Slika 1: Načini prostoležečega podpiranja preizkušanca [2]

Pri prvem načinu ugotavljanja lastne frekvence preizkušanca le-tega vzbujamo s sinusnim signalom in istočasno merimo odziv preizkušanca. Ko zvezno spreminjamo frekvenco sinusnega signala, dobimo spekter odziva, iz katerega lahko razberemo lastno frekvenco preizkušanca. Primer idealnega spektra je prikazan na sliki 2. Nihanje sistema prostoležečega ali prostovpetega nosilca z enakomerno porazdeljeno maso lahko idealiziramo v nihanje z eno prostorsko stopnjo, kot je prikazano na sliki 4.

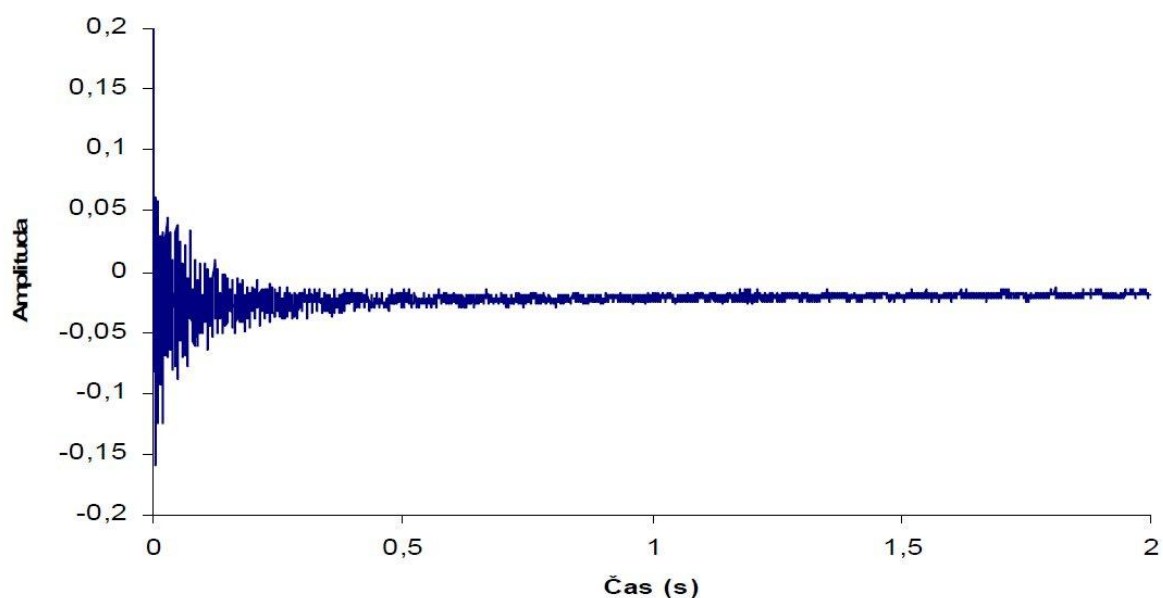


Slika 2: Primer spektra odziva



Slika 3: Nihalo z eno prostorsko stopnjo

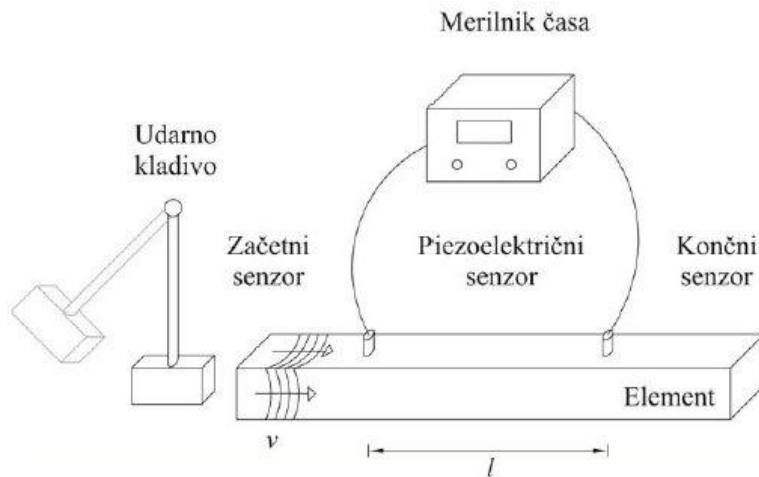
Drugi način ugotavljanja lastne frekvence preizkušanca je, da v preizkušane v silimo kratkotrajno motnjo – impulz, ter merimo časovni odziv nihanja. Tak odziv v splošnem merimo z osciloskopom. Iz oscilograma lahko razberemo precej podatkov o sistemu, dušenje, lastno frekvenco itd. Primer oscilograma je prikazan na sliki 4. Pri ugotavljanju lastne frekvence preizkušancev, ki so analizirani v diplomski nalogi, je bila uporabljena ta metoda.



Slika 4: Časovni odziv sistema na pulz [3]

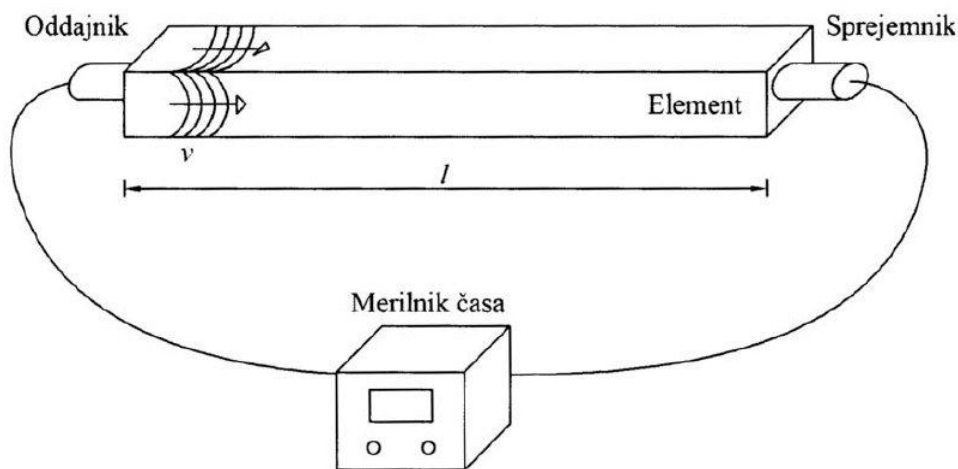
2.2.2 Vz dolžna impulzna motnja

Dinamični modul elastičnosti (in posledično tudi trdnost) preizkušanca lahko ocenimo tudi na osnovi longitudinalne impulzne motnje. Na enem koncu preizkušanca vzbudimo longitudinalno tlačno motnjo, na drugem pa pritrdimo merilnik, ki zazna časovno razliko med vzbuditvijo ter prihodom tlačnih valov do merilnika. Shema delovanja je prikazana na sliki 5. Poznamo tudi metodo t.i. 'dvojnega preleta', kjer sta merilnik in vzbujevalnik na isti strani; ko zgostitveni valovi pridejo do konca preizkušanca, se odbijejo in potujejo nazaj. Taka metoda ima zaradi večjega časovnega intervala višjo natančnost.



Slika 5: Shema merjenja longitudinalnega širjenja valov [2]

Zgostitve lahko vzbudimo bodisi z udarnim kladivom, bodisi z ultrazvokom. Princip delovanja je isti, le da ultrazvok deluje pri višjih frekvencah. Pri obeh načinih se lahko uporablja metoda dvojnega preleta. Prednost ultrazvoka je, da deluje na manjši površini, zato se lahko uporablja pri merjenju vezanih plošč in lepljenega lesa. Shematično je prikazan na sliki 6. Ultrazvok odkriva tudi anomalije v lesu, saj se signal od nepravilnosti odbija nazaj k oddajniku, do sprejemnika pa pride šibkejši signal. Ultrazvok, ki deluje dvosmerno, ima le eno sondo, ki istočasno deluje kot oddajnik in sprejemnik signala.



Slika 6: Shema delovanja enosmernega ultrazvočnega sistema [2]

Ne glede to, katero metodo smo izbrali, smo izmerili čas širjenja longitudinalnih zgostitvenih valov. Hitrost širjenja se izračuna po enačbi (1):

$$v = \frac{l}{\Delta t} \quad (1)$$

v hitrost širjenja valov

l dolžina preizkušanca

Δt čas širjenja valov

V idealnih pogojih dinamični elastični modul izračunamo po enačbi (2):

$$MOE = v^2 \times \rho \quad (2)$$

MOE dinamični elastični modul

ρ gostota preizkušanca

Pri preizkušancih, analiziranih v tej diplomski nalogi, so se zgostitve vzbujele z udarnim kladivom.

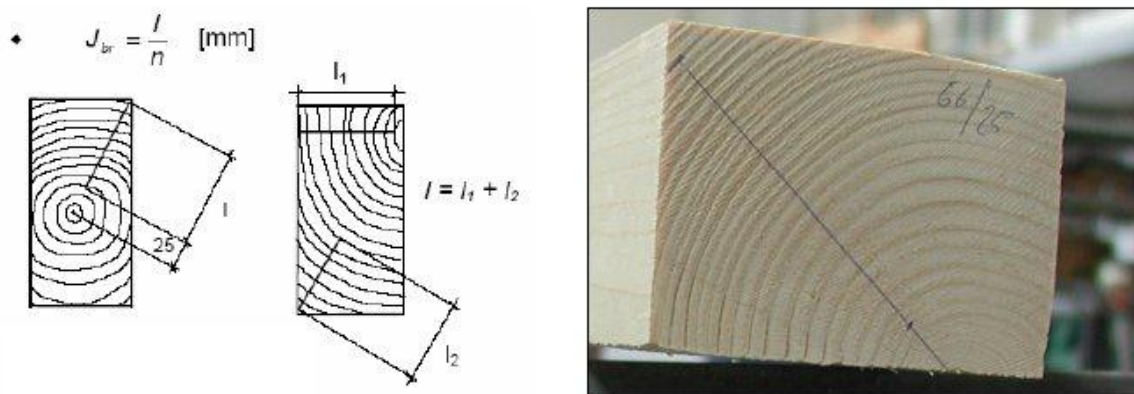
2.3 Vizualni pregled

Vizualno ocenjevanje lesa je najstarejša metoda razvrščanja, ki je v uporabi še danes. Največja prednost metode je, da je enostavna, zato se uporablja predvsem na manjših lesno predelovalnih obratih, kot dobra podpora ostalim preiskavam pa tudi na večjih. Slabost je, da je zamudna, lahko pa ocenimo le karakteristike zunanje površine, o notranjosti preizkušanca ne dobimo nikakršne informacije. Prav tako vizualni pregled ni najboljša preiskava s kontroliranim raztrosom, zato je razvrščanje preizkušancev v trdnostne razrede izključno po tej metodi zelo konzervativno in neekonomično. Trenutno v Sloveniji z vizualno preiskavo lesa ne moremo razvrstiti v razred, višji od C24.

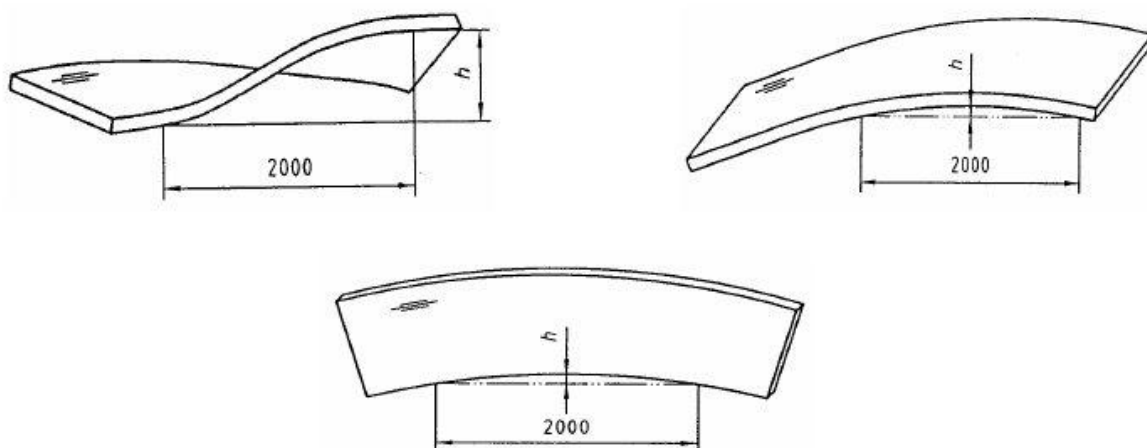
Zaradi množice vrst dreves, raznolikosti rastišč ter različnih lokalnih dimenzijskih in namembnostnih zahtev v Evropi poznamo več pravilnikov za razvrščanje po vizualnem pregledu. Vizualno lahko ocenimo:

- količino ter mesto grč,
- poškodbe preizkušanca,
- zunanji biološki razkroj,
- poškodbe vlaken,
- razpoke,
- zavitost vlaken,
- zavitost preizkušanca ter
- hitrost rasti.

Poleg vseh naštetih lastnosti lahko ocenimo tudi nekaj nestandardiziranih karakteristik, ki jih moramo smiselno ovrednotiti, saj zmanjšujejo trdnost in elastični modul lesa. To so n.pr. mehanske poškodbe, razni vključki, poškodbe zaradi parazitov, nenormalen potek stržena ipd. Nekaj primerov je prikazano na slikah od 7 do 9.



Slika 7: Merjenje širine branike [4]

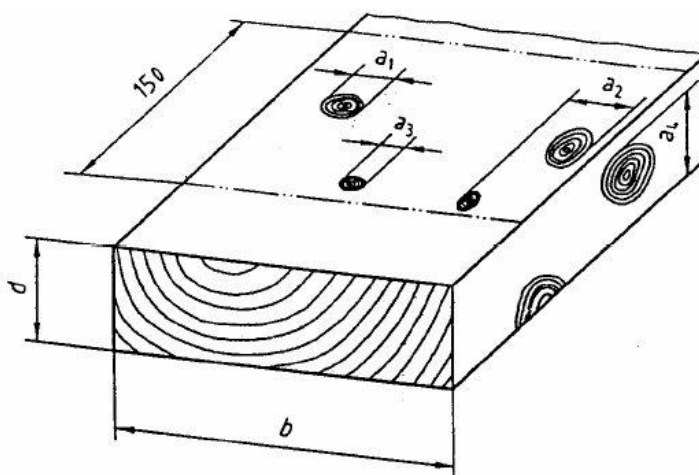


Slika 8: Vizualno merjenje zavitosti, loka in sablje [4]



Slika 9: Razpokan les

Preizkušanci, vključeni v analizo diplomske naloge, bistvenih mehanskih, bioloških in geometrijskih pomanjkljivosti niso imeli. Zato je bila edina karakteristika, ki jo je bilo smiselno kvantitativno opredeliti, razmerje grč v prerezu. Kratica tKAR (poglavje 2.1.3, preglednica 3) pomeni 'total Knot Area Ratio', torej razmerje med vsoto površin vseh grč v prerezu in površino prereza. Pri merjenju razmerja grč v prerezu se zabeleži tudi pozicija največje grče vzdolž preizkušanca, vendar ta podatek z mehanskimi karakteristikami preizkušanca nima nobene povezave, zato ga v analizi nismo upoštevali. Shematično je merjenje prikazano na sliki 10:



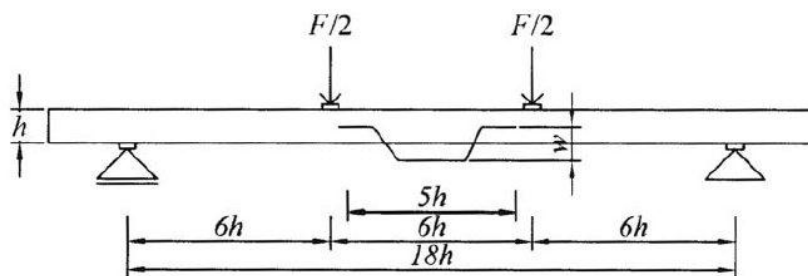
Slika 10: Vizualno merjenje grč [4]

Koeficient grč tKAR se izračuna po enačbi 3:

$$tKAR = \frac{\sum_i (a_i)^2}{4 \cdot b \cdot d} \quad (3)$$

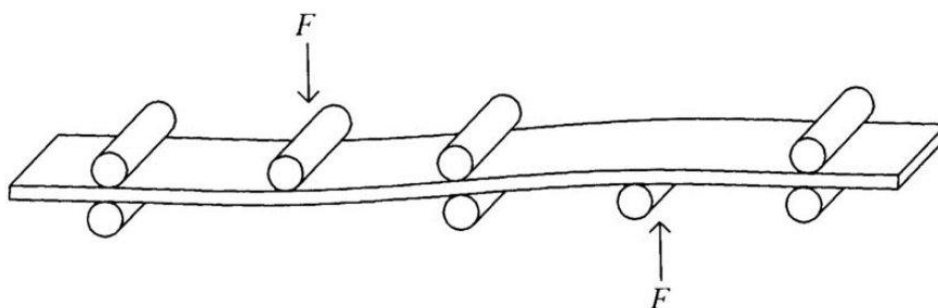
2.4 Globalni modul elastičnosti

Ker je ugotavljanje modula elastičnosti neporušna preiskava, lahko tudi iz tega podatka sklepamo o trdnosti vsakega preizkušanca. Ugotavljamo ga lahko s preiskavo, ki je shematično prikazana na sliki 11. Preizkušanec višine 'h' ter dolžine '18h' prostoležeče podpremo, ter obremenimo z enakima silama na tretjinah razpona z vsake strani. Merimo pomik na sredini razpona.



Slika 11: Shema preizkušanja za določitev globalnega modula elastičnosti [5]

Opisan način preizkušanja je v uporabi v laboratorijih. V lesno predelovalnih obratih se velikokrat uporablja sistem valjev, ki delujejo po principu tekočega traku. Skoznje spustimo lesene preizkušance, naprava pa zazna debelino ter pomik ob določeni sili, iz česar avtomatično izračuna globalni modul elastičnosti za vsak preizkušanec. Po takem postopku določimo elastični modul položenih elementov, medtem ko se laboratorijsko elastični modul določa na pokončnih elementih. Delovanje je prikazano na sliki 12:



Slika 12: Merjenje modula elastičnosti v tekočem traku [2]

Izhodni podatek preiskave je graf, ki prikazuje pomik sredine nosilca v odvisnosti od sile, s katero na tretjinah razpona pritiskamo na preizkušane. Največja dovoljena obremenitev ne sme povzročiti napetosti, večjih od 40% porušne trdnosti. Le tako namreč preizkušane ostane v linearnem elastičnem območju. Ko poznamo zvezo med silo in pomiki, lahko globalni modul elastičnosti izračunamo po formuli:

$$E_g = \frac{l^3 (F_2 - F_1)}{b h^3 (w_2 - w_1)} \left[\left(\frac{3a}{4l} \right) - \left(\frac{a}{l} \right)^3 \right], \quad (4)$$

kjer je:

- F_1, F_2 sila na začetku in koncu obremenjevanja,
- w_1, w_2 pomik na začetku in na koncu obremenjevanja,
- b širina preizkušance,
- h višina preizkušance ter
- l razdalja med podporama.

Če privzamemo, da je $a = 6h$ in $l = 18h$, enačbo (4) poenostavimo in dobimo enačbo (5):

$$E_g = \frac{810 (F_2 - F_1)}{b (w_2 - w_1)}. \quad (5)$$

2.5 Porušne preiskave

Vse nedestruktivne metode določanja trdnosti lesenih preizkušancev so približek dejanski vrednosti porušne trdnosti. Edini točni način za določitev trdnosti lesenih preizkušancev so porušne preiskave. Ker po porušni preiskavi preizkušanec ni več uporaben, trdnost preizkušancev pa varira tudi znotraj istega debela, jih v obsežni meri izvajamo le za potrebe določitve korelacije med porušnimi in neporušnimi preiskavami ter za določitev pomena neporušnih preiskav.

Ločimo več načinov porušnih preiskav lesenih preizkušancev. V naši raziskavi so bili opravljeni natezni ter upogibni preizkusi. Pri obeh variantah je potrebno v napravo za preizkušanje vnesti zelene pomike, ter meriti silo, ki je za ta pomik potrebna. Če vnesemo silo, ki jo želimo doseči, ter merimo pomike, se lahko zgodi, da ob bližanju porušni sili element že popusti, naprava pa še vedno želi doseči vnaprej podano maksimalno silo, zato se pomik zelo hitro, nekontrolirano poveča.

Geometrija upogibnega preizkusa je identična geometriji, s katero se določi globalni statični modul elastičnosti. Preizkušanec obremenjujemo na tretjini razpona z vsake strani zato, da je območje maksimalnega momenta konstantno po celi srednji tretjini. Tako do odpovedi prvega vlakna pride zaradi najšibkejšega vlakna, in ne zaradi lokalnega največjega momenta. Če pride do odpovedi prvega vlakna v območju izven srednje tretjine, se preizkušanec zavrže. Ko imamo maksimalno silo, ki jo preizkušanec prenese, trdnost izračunamo po enačbi 5:

$$\sigma_{max} = \frac{18 \cdot F_{max}}{b \cdot h}, \quad (6)$$

pri čemer smo upoštevali, da je razdalja med podporama enaka $l = 18 \cdot h$.

Pri nateznem preizkusu se preizkušanec vpne v dvoje jeklenih čeljusti. Prav tako kot pri upogibnem preizkusu, mora tudi pri nateznem do odpovedi priti v srednji tretjini preizkušanca. S tem se izključi možnost odpovedi vlaken zaradi vpliva nazobčanih jeklenih čeljusti. Ko imamo maksimalno porušno silo, se trdnost preizkušanca izračuna po formuli:

$$\sigma_{max} = \frac{F_{max}}{b \cdot h} \quad (7)$$

Zaradi nelinearne anizotropne mehanike lesa tudi pri istem preizkušancu pri upogibnem preizkusu ne dobimo enakih vrednosti trdnosti lesa kot pri nateznem. Zato smo preizkušance razdelili tudi po vrsti destruktivne preiskave.

3 STROJNO UČENJE

3.1 Teoretične osnove

3.1.1 Ideja in zasnova

Že leta 1956 je John McCarthy z ameriškega MIT uvedel izraz 'umetna inteligenca' [6]. Definiral ga je kot 'znanost in inženiring izgradnje inteligentnih strojev'.

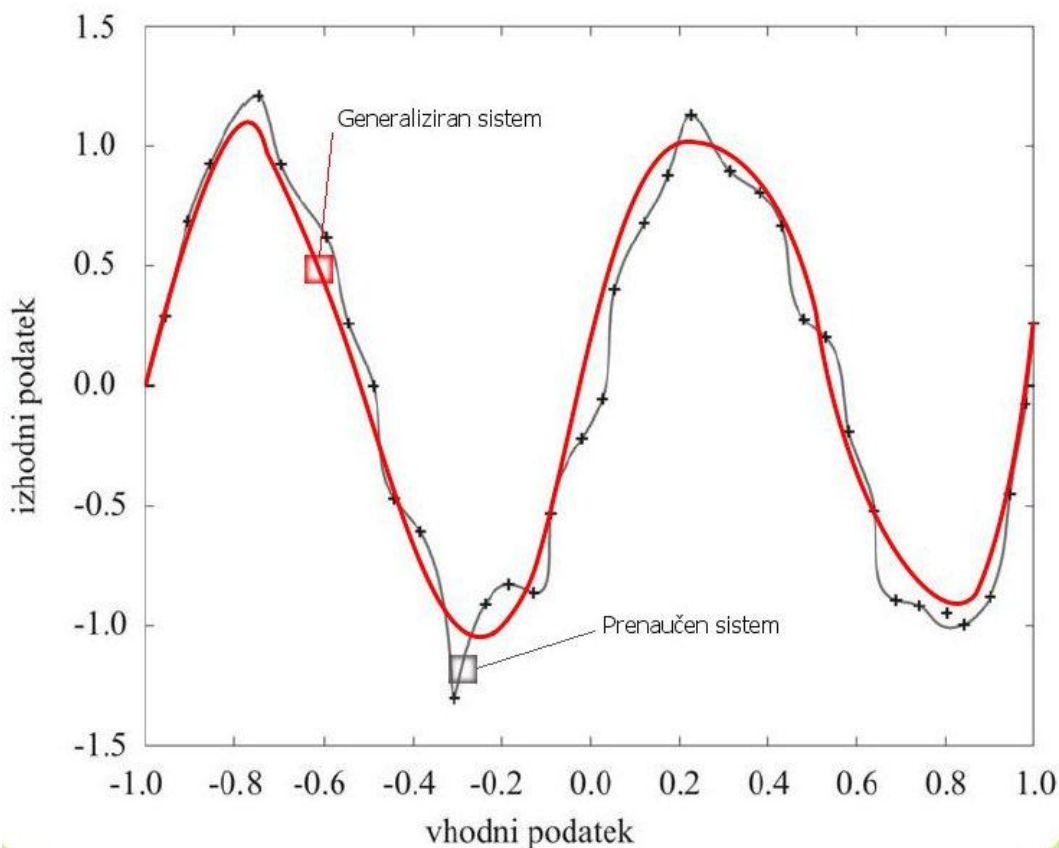
Ideja umetne inteligence je posnemati delovanje človeških možganov. Ta človeški organ ima značilnost, da tudi v vnaprej nepredvideni situaciji zna odreagirati smiselno in logično, opirajoč se na pretekle izkušnje in znanja. Tak način delovanja je za konvencionalne računalniške programe neizvedljiv. Ko se vnaprej napisana skripta, torej program, znajde v nepredvideni situaciji, preneha z delovanjem.

Programi umetne inteligence so zmožni podati določen odziv na vhodne podatke, ki jih predhodno pri pisanju programa nismo predvideli; to pomeni, da, za razliko od običajnih programov, rešijo tudi nepredvidene probleme. Pri oblikovanju odziva se opirajo na obstoječe podatkovne baze, torej na 'preteklo znanje in izkušnje'. Proces, v katerem program prepozna vzorce ter korelacije med spremenljivkami podatkovne baze, imenujemo strojno učenje [6].

Cilj razvijalcev umetne inteligence je narediti sisteme, ki bodo sposobni prepoznati zakonitosti posameznih problemov in ponuditi rešitve zanj. Znanost, ki se ukvarja z razvojem programov umetne inteligence, se najbolj osredotoča na dve področji, opisani v spodnjih dveh odstavkih. Prednost uporabe teh metod je, da vsaka odločitev ali napoved sistema temelji na vnaprej določeni metodi, kar prepreči subjektivnost odločitve.

Prvo je avtomatično prepoznavanje kompleksnih vzorcev znotraj obstoječih podatkovnih baz. To pomeni, da ob dovolj dobri podatkovni bazi program prepozna povezave, ki jih je sicer težko prepoznati in modelirati. Angleško temu procesu rečemo 'Knowledge discovery in databases' ali 'KDD'.

Drugo področje je napovedovanje odziva. Ko program umetne inteligence pozna zakonitosti povezav med spremenljivkami podatkovne baze, lahko empirično napove odziv na nove vhodne podatke. Značilnost dobro naučenega sistema je tudi generaliziranje ustvarjenih pravil. To pomeni, da je program uspel pravilno prepoznati zakonitosti problema, in jih uspešno ločil od posameznih odstopanj. Če je sistem dobro naučen, in novi vhodni podatki dobljeni na enak način kot tisti, ki smo jih uporabili za učenje, lahko sklepamo, da bo tudi napoved izhodnih podatkov zelo podobna dejanskemu stanju. Primera generaliziranega ter prenačenega sistema sta prikazana na sliki 13.



Slika 13: Primer generaliziranega in prenaučenege sistema [2]

Vsako učenje je iterativni proces, zato pri strojnem učenju želimo, da ta proces konvergira. Postavlja pa se osnovno vprašanje kriterija, kdaj lahko za program rečemo, da se uspešno uči. Prof. Thomas M. Mitchell z univerze Carnegie Mellon je podal znano definicijo [7], ki se glasi: Računalnik se uspešno uči iz izkušenj, če se njegova uspešnost pri opravljanju nalog zaradi teh izkušenj izboljšuje.

3.1.2 Kratek pregled algoritmov in metod strojnega učenja

Obstoječi algoritmi se delijo v več skupin, katerih lastnosti so opisane v spodnjih odstavkih.

Nadzorovano učenje je proces, pri katerem se funkcije povezav inducirajo iz znanega in nadzorovanega nabora podatkov. Nabor podatkov je sestavljen iz množice učnih primerov, vsak primer pa je 'par', saj ga sestavljajo vhodni in izhodni podatki. Ideja je, da se na osnovi učnih primerov inducira taka funkcija, ki za vsak nov primer, za katerega izhodnih podatkov nimamo, čim natančneje napove vrednosti izhodnih podatkov. Najbolj znani algoritmi so umetne nevronske mreže, Bayesova statistika, regresijska drevesa, Gaussova regresija ipd.

Nenadzorovano učenje se osredotoča na iskanje smiselnih povezav znotraj podatkovne baze. To pomeni, da nabor podatkov ni sestavljen iz parov vektorjev, torej vhodnih ter izhodnih podatkov, temveč so vektorji le vhodni. Znotraj nabora teh vektorjev algoritmi nenadzorovanega učenja iščejo povezave med spremenljivkami. Kot je omenjeno v poglavju 3.1.1, so taki procesi znani pod imenom 'Knowledge discovery in databases'. Poleg umetnih nevronskih mrež sta najbolj znana algoritma nenadzorovanega učenja tudi gručenje podatkov, angl. 'Data clustering', ter algoritem maksimizacije pričakovanega, angl. 'Expectation-maximization algorithm'.

Delno nadzorovano učenje je, kot že ime pove, kombinacija nadzorovanega in nenadzorovanega učenja. Praktični primeri so pokazali, da je pridobivanje podatkov za nadzorovano učenje velikokrat precej dražje od pridobivanja podatkov za nenadzorovano. Če pa večjemu številu učnih primerov za nenadzorovano učenje dodamo manjšo količino učnih primerov za nadzorovano, se natančnost in učinkovitost procesa učenja bistveno izboljša. Logičen pristop je izbira algoritmov, ki so zmožni inducirati funkcije povezav kljub manjkajočim vrednostim, kot n.pr. Gibbsovo vzorčenje ter algoritem maksimizacije pričakovanega.

Vzpodbujevalno učenje je vrsta strojnega učenja, ki je največkrat zasnovana kot Markovski odločitveni proces (po ruskem matematiku Andreyu Markovu). Učenje znotraj tega procesa poteka tako, da je v vsakem koraku učenja sistem v nekem stanju 's', v katerem je na voljo več odločitev 'o'. Ko sistem izbere eno od odločitev 'o', se sistem preoblikuje v novo stanje 's_i'. Tranzicijo stanja se lahko opiše s funkcijo $F(s, s_i)$. Neodvisni vzporedni proces vsako stanje kvantitativno oceni, zato se lahko kvantitativno oceni tudi uspešnost tranzicije. Na podlagi teh ocen se sistem nauči, katere odločitve so najboljše v posameznih korakih. Tak način sprejemanja odločitev je v uporabi na mnogo področjih: v informatiki, robotiki, pri snovanju strategij in taktik, pri snovanju umetne inteligence računalniških iger, optimizaciji ipd.

3.2 Izbira in opis algoritmov

Ker poznamo mnogo različnih algoritmov strojnega učenja, se postavlja vprašanje, katerega izbrati za uporabo pri analizi v diplomski nalogi. Za analizo imamo na voljo pare vhodnih (podatki o geometriji, teži in vlažnosti preizkušancev ter rezultati nedestruktivnih testov) in izhodnih (destruktivne preiskave) vektorjev, zato je logično izbirati procese nadzorovanega učenja. Empirično povezavo med vhodnim in izhodnim nizom podatkov lahko simuliramo na več različnih analitičnih načinov. V praksi so se kot dovolj dobre dokazale umetne nevronske mreže, ki so predstavljene v poglavju 3.2.2. Umetne nevronske mreže so bile v preteklosti že uporabljene pri simuliranju povezav med nedestruktivnimi in destruktivnimi metodami preizkušanja lesenih vzorcev, vendar le redko za tako velik vzorec preizkušancev ter nikoli za tako velik vzorec slovenskih preizkušancev.

Niz podatkov, ki sestavljajo izhodne vektorje, smo razdelili na več primerov tako, da smo v vsakem paru ohranili vektor vhodnih podatkov cel, od vektorja izhodnih podatkov pa smo enkrat ohranili le skalar 'gMOE', drugič pa 'Strength' (glej pogl. 2.1.3). To smo storili pri vsaki od štirih možnosti (poglavje 2.1.2, preglednica 2). Razdelitev podatkov na le eno neznanko, torej na en izhodni podatek, omogoča tudi iskanje odvisnosti z regresijsko analizo, produkt katere je odločitveno drevo.

3.2.1 Regresijska drevesa

3.2.1.1 Splošno

Algoritmi, ki generirajo odločitvena drevesa (tudi 'TDIDT' – Top Down Induction of Decision Trees) sodijo po klasifikaciji sistemov strojnega učenja v skupino propozicijskega učenja [8]. Algoritmi so zasnovani po logiki 0. reda, kar pomeni, da pri generiranju odločitvenih pravil gledajo en korak naprej.

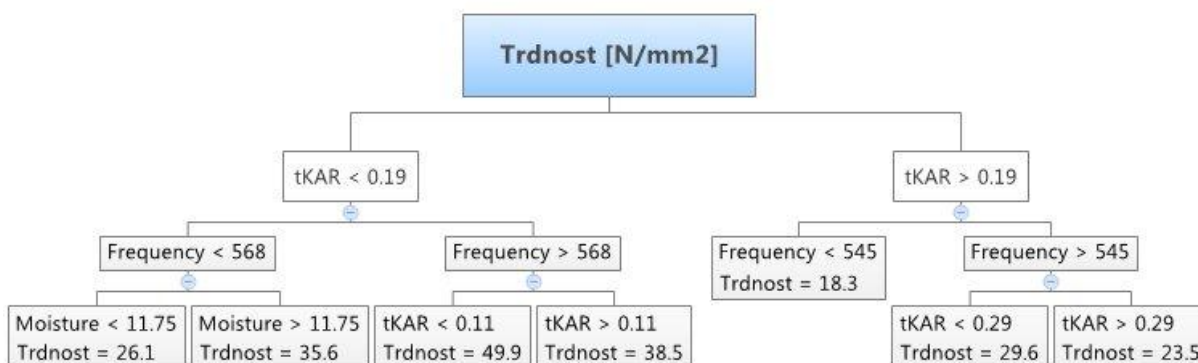
Vhodno-izhodni par podatkov sestavlja vhodni vektor (več vhodnih podatkov) ter izhodni skalar (podatek je en sam). Odločitveno drevo se sestavi tako, da algoritem v vsakem koraku izbere eno spremenljivko iz nabora vhodnih podatkov, ter zanjo postavi kriterij – določeno vrednost te spremenljivke [9]. Na ta način se celoten nabor parov razdeli na dva dela, na manjše in večje od tega kriterija. Iz enega 'debla' smo tako dobili dve 'veji'.

Kriterij je s poizkušanjem določen tako, da v vsaki od podvej obstaja še ena spremenljivka (lahko tudi ista kot prej), ki ima po delitvi iz debla v dve veji čim večjo normalizirano standardno deviacijo. Ko naprej delimo po tej spremenljivki, spremljamo obnašanje izhodnega skalarja. Manj kot se prepletajo meje nastalih listov, boljše delitev smo dobili. To pomeni, da algoritem gleda le en korak naprej.

V vsakem koraku se najprej izračuna prva delitev po vseh možnih spremenljivkah, nato pa še druga delitev po vseh možnih spremenljivkah. Po drugi delitvi se gleda prepletanje mej listov drevesa. Šele ko smo preizkusili vse možne variante, se izbere spremenljivko, znotraj katere je določen kriterij dejanske delitve. Ko je delitev opravljena, se vsak del obravnava neodvisno od drugega. To pomeni, da so na določeni globini drevesne strukture delitve v splošnem opravljene po različnih spremenljivkah.

Kadarkoli delitev zaključimo, dobimo 'liste', ki določajo meje izhodnega skalarja, katerega vhodni par je sledil vejam od debla do lista. Zanima nas, kako dobro te meje napovejo vrednost izhodnega podatka poljubnega preizkušanca. Na mestu je vprašanje, kdaj delitev ustaviti. Če pustimo delitev brez omejitev, se drevo razvije v t.i. 'popolno drevo', kjer je na vsakem listu po en vhodno-izhodni par. Taka delitev že ob majhnem raztrosu rezultatov ni dobra, saj je drevo preveč prilagojeno vzorcu, na osnovi katerega se je induciralo.

Drevesa lahko omejimo na dva načina. Prvi je s postavljanjem omejitvenih pogojev, drugi pa z naknadnim rezanjem. Omejitveni pogoji so dveh vrst. Omejujemo lahko bodisi minimalno število preizkušancev na enem listu, bodisi maksimalni razpon izhodnih vrednosti. Naknadno rezanje pomeni, da izgradimo popolno drevo, nato pa se odločimo za vsak primer posebej, katere veje so nezanesljive, in jih izločimo. V splošnem velja, da se naknadno rezanje bolje obnese pri generalizaciji manjših vzorcev preizkušancev. Vzorec, obravnavan v tej diplomski nalogi, je dovolj velik, da naknadno rezanje ni smiselno. Možna je tudi kombinacija obeh načinov. Primer regresijskega drevesa je prikazan na sliki 14. Napovedani parameter je trdnost natezno obremenjenih borovih preizkušancev, na vsakem vozlišču je postavljen kriterij delitve, na zadnjih listih pa tudi vrednost napovedane trdnosti.



Slika 14: Regresijsko drevo za trdnost bora pri nateznem preizkusu

Korelacija se izračuna s postopkom 'cross-validation'. Pri tem postopku program naključno razdeli primerke v toliko enako velikih skupin, kot smo si izbrali. Eno od skupin začasno odstrani iz analize, pravila za delitev drevesa pa se generirajo na preostalih primerkih. Ko so pravila za delitve drevesa ustvarjena, delitev ustavljena, regresijsko drevo pa ustrezno porezano, algoritem preveri delovanje na primerkih, ki so bili predhodno izločeni. Nato se postopek ponovi tako, da se izločeni primerki vrnejo celotnemu vzorcu, izloči pa se nova skupina. Postopek tako ponovimo tolikokrat, na kolikor delov smo razdelili celoten vzorec. Končna pravila za razdelitev in rezanje drevesa ter končna korelacija so aritmetična sredina vrednosti, ki smo jih dobili v posameznih ponovitvah. Pogosta je delitev vzorca na deset delov.

Korelacijo določimo na primerkih, ki v analizo niso bili zajeti. Zato lahko predpostavljamo, da koeficient korelacije med vrednostmi, ki so predpisane na posameznih listih drevesa, ter dejanskimi izhodnimi vrednostmi, ni previsok, saj drevo ni prilagojeno posameznim primerkom validacijske skupine.

3.2.1.2 Parametri za izračun

V tem poglavju so predstavljeni parametri algoritma, ki inducira odločitveno drevo. S spreminjanjem teh parametrov poskušamo izdelati čim boljše drevo; kriterij za kakovost je koeficient korelacije med drevesom, ki se je induciralo iz učnega nabora vhodno-izhodnih parov, ter naborom testnih parov, ki v učenje niso bili vključeni. Vrednosti najboljših parametrov za izračun so odvisni od števila vhodno-izhodnih parov, oblike in velikosti raztrosa, korelacije med spremenljivkami in drugega. Zato je optimizacija parametrov iterativen proces poizkušanja. Splošne metode za konvergiranje k optimalni vrednosti ni. Predstavljeni so parametri, ki jih omogoča program WEKA Toolkit (pogl. 3.3). Izbira parametrov za problem diplomske naloge je skupaj s pripadajočimi rezultati opisana v poglavju 4.4.

Globina drevesa (maxDepth) – S tem parametrom lahko omejimo število delitev. Z nastavitvijo '-1' je globina neomejena. S tem kriterijem največkrat ustavimo algoritem takrat, ko želimo, da je za določen primer možen grafični prikaz celotnega odločitvenega drevesa.

Število primerkov na listu (minNum) – Deljenje v podveje lahko ustavimo tudi z omejitvijo števila preizkušancev na posameznem listu. Če je po določeni delitvi na enem od nastalih listov preizkušancev manj od izbranega števila, to delitev algoritem pozabi in se v tej veji vrne korak nazaj.

Interval izhodnega skalarja (minVarianceProp) – Algoritem se lahko ustavi tudi v primeru, da je v nastalem listu primerkov veliko, vendar so njihove vrednosti izhodnega skalarja (v našem primeru trdnost ali modul elastičnosti) tako blizu skupaj, da je nadaljna delitev nesmiselna. Na tem mestu določimo interval izhodnega podatka na enem listu; če so vrednosti izhodnih podatkov vseh primerkov na tem listu znotraj intervala, se nadaljnje deljenje ne izvede. Taka zaustavitev deljenja je v našem primeru najbolj smiselna omejitev, saj razvrščamo lesene preizkušance v trdnostne razrede, ki imajo prav tako omejene intervale za trdnost in elastični modul.

Rezanje drevesa (noPruning) – Kadar ustvarjeno drevo kljub vsem prejšnjim omejitvam ustvari zelo natančne liste, so ponavadi pravila, ki veljajo za elemente tega lista, zelo specifična, prilagojena učnemu naboru podatkov, in zato neuporabna za splošne primere. Z omogočeno funkcijo avtomatičnega rezanja drevesa program iz nabora podatkov sam izloči določeno število primerkov, ki jih uporabi za preverjanje pravil posameznega lista. Kadar so pravila preveč specifična, veje poreže.

Število delitev (numFolds) – Tu lahko nastavljam, na koliko enakih delov bomo razdelili celoten nabor primerkov. Če n.pr. izberemo število 5, se bo naših 5000 preizkušancev naključno razdelilo v 5 skupin po 1000. Štiri od teh skupin bodo generirale pravila za delitve; končna pravila bodo povprečje vseh štirih. Peta skupina bo preizkuševalni nabor, ki bo pokazal, ali so pravila preveč specifična ali ne.

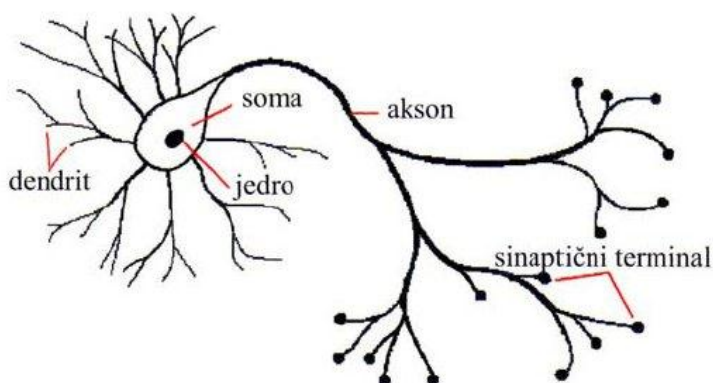
3.2.2 Umetne nevronske mreže

3.2.2.1 Splošno

Umetne nevronske mreže pri zastavljenih problemih delujejo podobno kot človeški možgani. Želja po posnemanju delovanja ni nova, znanstvenike in razvijalce namreč že dolgo navdihuje zmožnost človeških možganov pri reševanju novih problemov. Kreativnega in inovativnega razmišljanja zaenkrat še ne znamo programirati, lahko pa z algoritmi strojnega učenja oponašamo učenje iz izkušenj, ter znotraj danega vzorca ugotavljamo povezave med podatki, ki so algoritmu na voljo. Preprost primer učenja iz izkušenj je količina in vrsta oblek, ki si jih moramo obleči ob določenem vremenu. Četudi še nikoli nismo videli točno določene kombinacije vhodnih podatkov, torej temperature, vlage, dežja ali sonca, ter časa, ki ga načrtujemo preživeti na prostem, izbira garderobe zaradi izkušenj, ki jih imamo s podobnimi pogoji, in tvorijo naučen sistem, ni zahteven problem. Primer učenja sistema znotraj danega vzorca je razumevanje deklet. Naloga je poiskati smiselne povezave med vhodnimi signali, ki jih lahko posameznik inducira. Praktični problem je, da so določeni signali smatrani za dobre/slabe le v določenih okoliščinah, v določeni povezavi z drugimi signali pa so razumljeni povsem nasprotno. Praksa kaže, da zaradi velikega raztrosa in množice naključnih dejavnikov tak sistem ni nikoli dobro naučen.

Biološki možgani so kompleksen sistem več milijard med seboj povezanih nevronov. Vsak nevron je majhna procesna enota, zgrajena iz treh delov: dendritov (receptorji nevronskega signala), celičnega telesa (procesna enota) ter daljšega aksona (prevodnik signala), ki se konča s sinapsami (oddajniki nevronskega signala, ki se navezujejo na dendrite drugih nevronov). Naloga nevrona je, da zbere signale prek vseh dendritov, in se na podlagi več faktorjev odloči o svojem odzivu, torej ali bo poslal svoj signal naprej ostalim nevronom, ter kako močan bo ta signal.

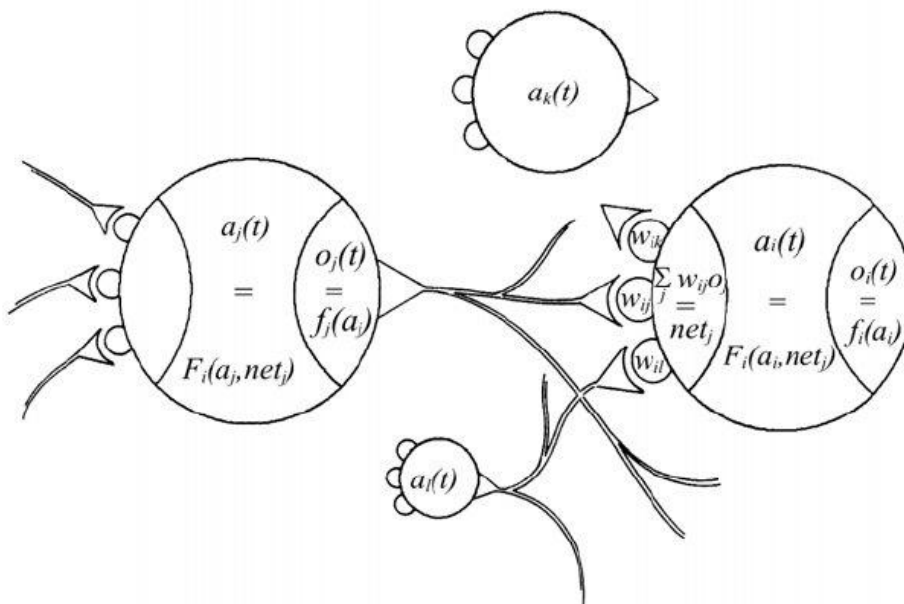
Jakost signala se meri v količini kemičnih spojin, ki jih prenaša akson. Osnovno učenje možganov pomeni, da se jakost signala spreminja in prilagaja. Tehnično gledano je največja razlika med umetno in biološko nevronske mrežo ta, da je umetna precej manjša, povezave pa so vnaprej predpostavljene. Biološka nevronska mreža je veliko večja in bolj kompleksna, struktura povezav pa se spreminja. Povezave, po katerih signali ne tečejo pogosto oziroma so šibki (periferne stvari v življenju pozabljamo), hitreje razpadejo od povezav, ki jih uporabljamo redno, oziroma je nekoč čez njih teklen močan signal (n.pr. spomin za celo življenje). Nove povezave se naključno nenehno vzpostavljajo. Biološki nevron je shematično predstavljen na sliki 15.



Slika 15: Biološki nevron [10]

V primeru, obravnavanem v diplomski nalogi, smo imeli za vsak preizkušanelec na voljo vhodno-izhodne pare podatkov. Podatki o geometriji, teži, grčah ter rezultati nedestruktivnih preiskav so vhodni podatki. Izhodni del podatkov so podatki o trdnosti in globalnemu elastičnemu modulu. Tip biološkega učenja, ki smo ga posnemali s strojnim učenjem, je torej učenje iz izkušenj.

Umetni nevron deluje podobno kot naravni. V vsako vozlišče iz vnaprej določenih drugih nevronov pride signal določene jakosti. Za vsako navezavo predhodnega nevrona (dendrit) ima obravnavani nevron (celično telo) zabeleženo utež, s katero pomnoži jakost signala s posameznega predhodnega nevrona. Vsota uteženih signalov določi, ali je vhodni signal dovolj močan, da doseže prag jakosti, in bo obravnavani nevron sploh oddal signal naprej (prek aksona v nadaljne nevrone), ter če ga bo, kako močan bo. Proces učenja nevronske mreže poteka tako, da algoritem na posameznih dendritih spreminja in optimizira uteži ter pragove jakosti. Vsak nevron se lahko navezuje tudi na njegove predhodnike. Nevronska mreža je shematično prikazana na sliki 16:



Slika 16: Umetni nevron [10]

kjer je:

$a_j(t)$ stanje aktivnosti j -tega nevrona v določenem času, to je vrednost v j -tem nevronu

w_{ij} utež povezave med i -tim in j -tim nevronom

net_j vrednost vhodnega signala v j -tem nevronu

a_i prag i -tega nevrona

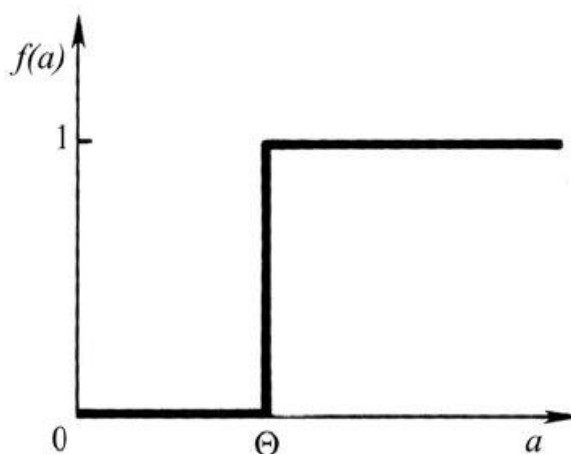
$f_i(a_i)$... izhodna funkcija i -tega nevrona

o_i izhodni signal i -tega nevrona

Ko je prag jakosti dosežen, se seštevek uteženih vhodnih signalov preračuna v izhodni signal. Jakost izhodnega signala lahko določamo z različnimi funkcijami [7]. Najbolj uporabljane so:

- pragovna funkcija (enačba (8)), prikazana na sliki 17:

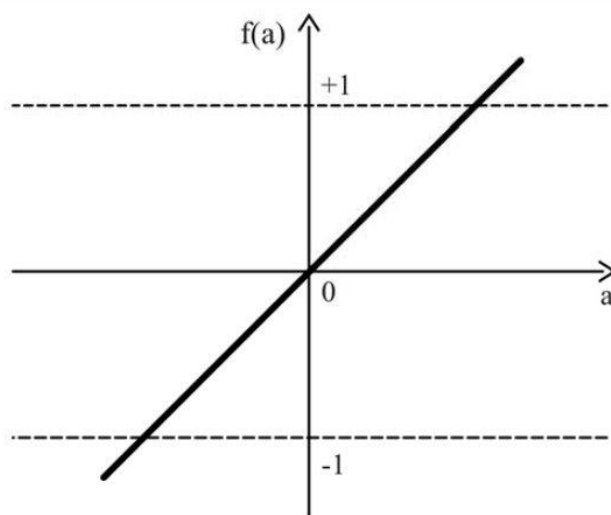
$$f(a) = \begin{cases} 1, & a \geq \Theta \\ 0, & a < \Theta \end{cases} \quad (8)$$



Slika 17: Pragovna funkcija

- linearna funkcija (enačba (9)), prikazana na sliki 18:

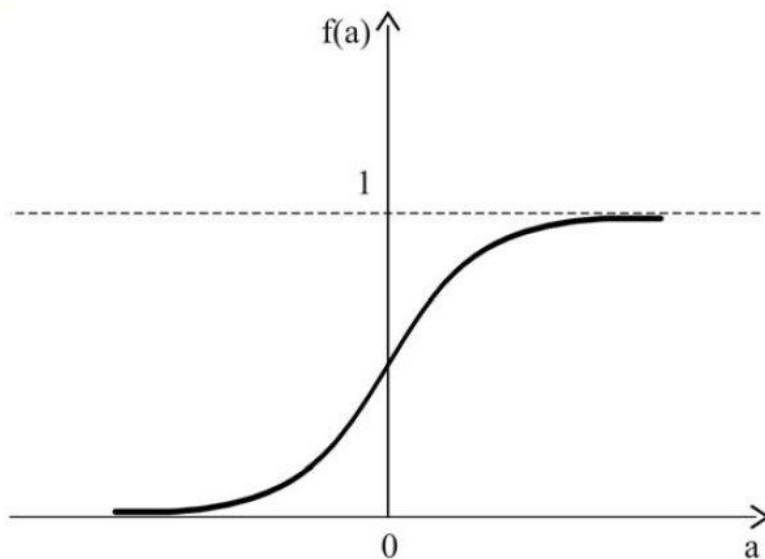
$$f(a) = k \cdot a \quad (9)$$



Slika 18: Linearna funkcija

- eksponentna sigmoidna funkcija (enačba (10)), prikazana na sliki 19:

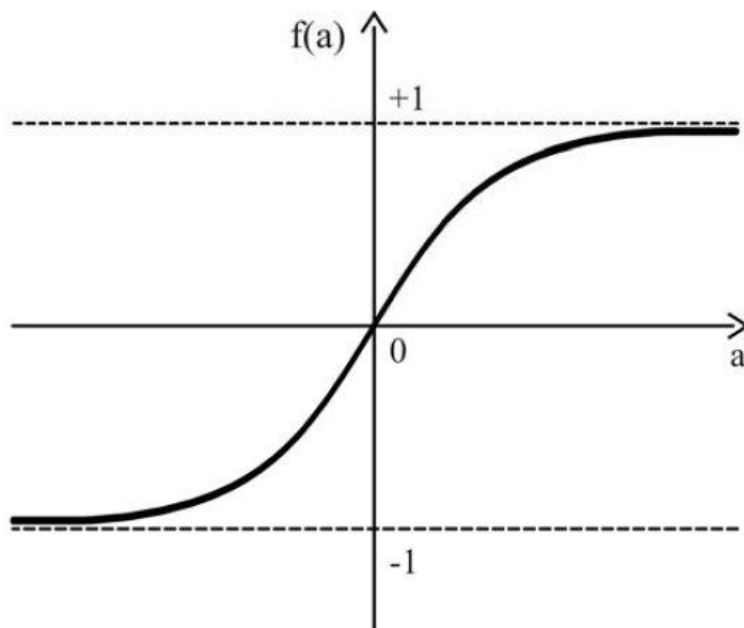
$$f(a) = \frac{1}{1 + e^{-a}} \quad (10)$$



Slika 19: Eksponentna sigmoidna funkcija

- tangentna sigmoidna funkcija, prikazana na sliki 20:

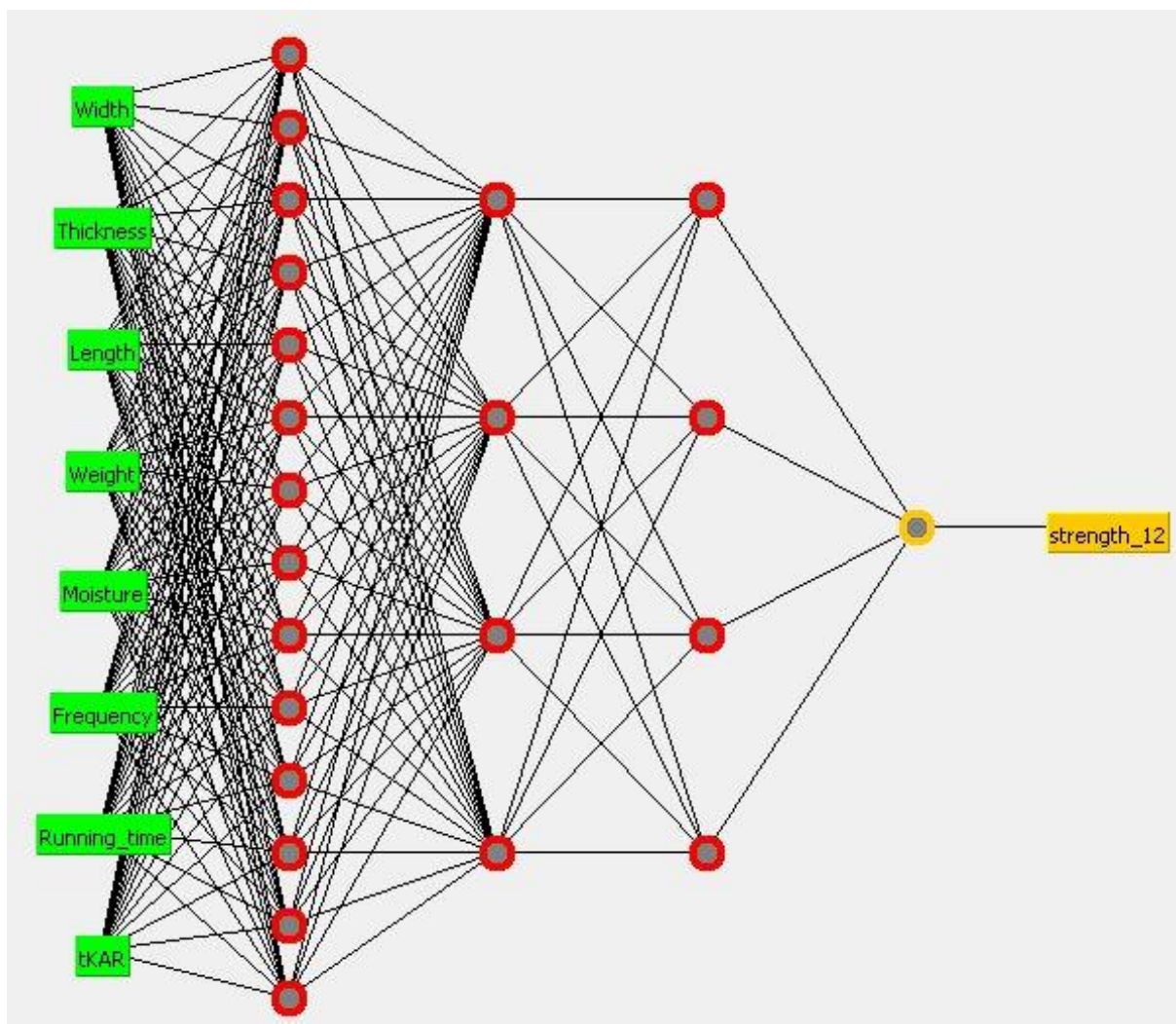
$$f(a) = \frac{2}{1 + e^{-2a}} - 1 \quad (11)$$



Slika 20: Tangentna sigmoidna funkcija

V diplomski nalogi smo za določanje jakosti signala med vsemi povezavami med nevroni uporabili tangentno sigmoidno funkcijo. Izkušnje ([2], [10], [11], [12]) kažejo, da so za primere ocenjevanja trdnostnega razreda lesa najbolj natančne, za razliko od eksponentnih sigmoidnih pa tudi dopuščajo negativne vrednosti.

Primer strukture nevronske mreže je prikazan na sliki 21. Prikazana je nevronska mreža z osmimi vhodnimi podatki (zeleni parametri), enim izhodnim ('strength_12') ter tremi skritimi plastmi. Na prvi je 14 vozlišč, na drugi ter tretji pa po 4.



Slika 21: Shema nevronske mreže

Plod naučene nevronske mreže lahko v n -dimenzionalnem prostoru predstavimo kot ploskev nad večdimenzionalno ravnino. Število dimenzij prostora je določeno s številom vhodnih parametrov, število hiper ploskev pa je enako številu izhodnih parametrov (v primeru, da izhodne parametre obravnavamo ločeno). Podobno rešitev bi dobili z aproksimacijo ploskev, kjer z metodo najmanjših kvadratov iščemo ploskev, ki čim bolje posploši in aproksimira oblak točk v prostoru.

Uspešna iteracija je odvisna od optimizacije uteži na dendritih nevrona. Optimizacija poteka iterativno tako, da ob danih utežeh algoritem določi vrednost izhodne spremenljivke, potem pa primerja rezultate z dejanskimi podatki na izhodnem nivoju. V naslednjem koraku utež spremenimo tako, da ji prištejemo popravek, odvisen od gradienta napake ter velikosti učnega koraka.

Pri mrežah s skritimi sloji (glej sliko 21) je sprememba uteži nekoliko bolj zahtevna, saj za posamezno utež na nevronu skritega sloja nimamo nobene dejanske vrednosti, na katero bi se lahko oprli za izračun napake. Zato algoritem pošlje signal v vzvratni smeri (od izhodnih proti vhodnim podatkom). V zadnjem sloju (izhodni nevroni) se napaka in gradient izračunata analogno kot pri mrežah brez skritih slojev. V vsakem koraku razrešujemo en skriti sloj. Posamezen sloj razrešimo tako, da določimo gradiente napake s pomočjo verižnega pravila za odvajanje, korekcijo uteži pa s pomočjo gradientov in uteži predhodno razrešenega sloja.

Velikost koraka in število iteracij igra pomembno vlogo pri določevanju dobre aproksimacijske ravnine. Če je velikost koraka prevelika ali število iteracij premajhno, globalni ekstremi niso dobro določeni, rešitev ne konvergira, lokalni ekstremi pa so izpuščeni. Če je velikost koraka premajhna ali pa število iteracij preveliko, postane mreža prenaučena, to pomeni, da se predobro prilagaja izbranemu vzorcu preizkušancev, in za splošno uporabo ni dobra.

Koeficient korelacije se izračuna po metodi 'cross validation', podobno kot pri regresijskih drevesih, kar je opisano v poglavju 3.2.1.1. Naknadno lahko koeficient korelacije določimo tudi s pomočjo neodvisnega vzorca primerkov, ki pri generiranju uteži mreže niso bili uporabljeni (glej pogl. 3.2.2.2, Velikost validacijskega nabora).

3.2.2.2 Parametri za izračun

V tem poglavju so predstavljeni parametri algoritma, ki inducira uteži na vozliščih umetne nevronske mreže. Z variranjem teh parametrov poskušamo izdelati čim boljše mrežo; kriterij za kvaliteto je koeficient korelacije med mrežo, ki se je inducirala iz učnega nabora vhodno-izhodnih parov, ter naborom testnih parov, ki v učenje niso bili vključeni. Vrednosti najboljših parametrov za izračun so odvisne od števila vhodno-izhodnih parov, oblike in velikosti raztrosa, korelacije med spremenljivkami in drugega. Zato je optimizacija parametrov iterativen proces poizkušanja. Splošna metoda za konvergiranje parametra k optimalni vrednosti ne obstaja. Predstavljeni so parametri, ki jih omogoča program WEKA Toolkit (pogl. 3.3). Izbira parametrov za problem diplomske naloge je skupaj s pripadajočimi rezultati opisana v poglavju 4.3.

Grafični vmesnik (GUI – Graphical User Interface) – Kadar je izbrana ta opcija, se pred vsakim izračunom pokaže grafična shema arhitekture mreže. Urejanje skritih plasti, vozlišč in povezav med vozlišči je v tej shemi pregledno in enostavno. Če opcije ne izberemo, program avtomatično ustvari predpisano število skritih plasti, predpisano število vozlišč v posamezni plasti, ter generira vse možne povezave med vhodnimi podatki, vozlišči ter izhodnim podatkom.

Avtomatična izgradnja mreže (autoBuild) – Pri vklopljeni opciji program avtomatično ustvari predpisano število skritih plasti, predpisano število vozlišč v posamezni plasti, ter generira vse možne povezave med vhodnimi podatki, vozlišči ter izhodnim podatkom. Če opcija ni izbrana, moramo mrežo ročno zasnovati z grafičnim vmesnikom. Če je grafični vmesnik izključen, je ta opcija vključena avtomatično.

Drobljenje (decay) – Če je ta opcija omogočena, se učni korak zmanjša ob vsaki ponovitvi iteracije. Tako se lahko prepreči divergiranje rešitev. Opcija zelo dobro deluje ob majhnem raztrosu in močni korelaciji osnovnih spremenljivk. Pri večjem raztrosu obstaja nevarnost, da se mreža prenauča zelo hitro.

Skrite plasti (hiddenLayers) – Na to mesto se vnaša število skritih plasti, ki jih želimo imeti v umetni nevronske mreži, ter število vozlišč na posamezni plasti. Program dovoljuje tudi vnos številke 0. Izkušnje kažejo, da za problem lesenih preizkušancev, katerih trdnost želimo

oceniti, arhitektura mreže ne igra bistvene vloge. Dobre rezultate so dale na primer nevronske mreže s tremi skritimi plastmi ter razporeditvijo vozlišč 14 – 4 – 4 [11].

Učni korak (learningRate) – Velikost koraka pri iterativnem iskanju najboljše rešitve. Premajhen korak za smiselno rešitev zahteva veliko število ponovitev, lahko pa tudi pripelje do prenaučeni mreže. Pri prevelikem koraku rešitev ne konvergira. Velikostni razred koraka se od primera do primera povsem razlikuje.

Vztrajnost (momentum) – Vpliv popravkov na uteži. Preveliki popravki povzročijo divergiranje rešitev, premajhni pa mrežo lokalno prenaučijo, še preden so odkriti globalni maksimumi. Velikostni razred vztrajnosti se od primera do primera povsem razlikuje.

Nominalni filter (nominalToBinaryFilter) – Če so v naboru podatkov nekateri parametri podani nazivno, jih ta opcija spremeni v števila, kar pohitri proces izračuna. V našem primeru nazivnih parametrov ni bilo.

Normalizacija nazivnih parametrov (normalizeAttributes) – Program lahko, preden požene algoritem, normalizira vse nazivne parametre tako, da zavzamejo vrednosti med –1 in 1. Zaradi tega predhodnega koraka se izračun mreže konča nekoliko hitreje.

Normalizacija numeričnih parametrov (normalizeNumericClass) – Program lahko, preden požene algoritem, normalizira vse numerične parametre tako, da zavzamejo vrednosti med –1 in 1. Zaradi tega predhodnega koraka se izračun mreže konča nekoliko hitreje. To je le vmesni korak, končni rezultat je ponovno preračunan na originalne vrednosti.

Ponovni začetek izračuna (reset) – Če je ta opcija izbrana, in program ugotovi, da algoritem iteracije divergira, avtomatično zmanjša učni korak in ponovno zažene algoritem.

Generator naključnih števil (seed) – Naključna števila so osnova za določanje začetnih uteži, ter za izbiro preizkušancev, ki jih program iz učnega nabora podatkov prestavi v testni nabor. Na njihove vrednosti lahko vplivamo s številko, ki jo vnesemo v to polje.

Število ponovitev iteracije (trainingTime) – Ob vsaki ponovitvi algoritma izračuna popravke in na novo določi uteži na vozliščih. Preveliko število ponovitev lahko privede do prenaučeni mreže, premajhno pa do slabih približkov. Velikostni razred števila ponovitev se od primera do primera povsem razlikuje.

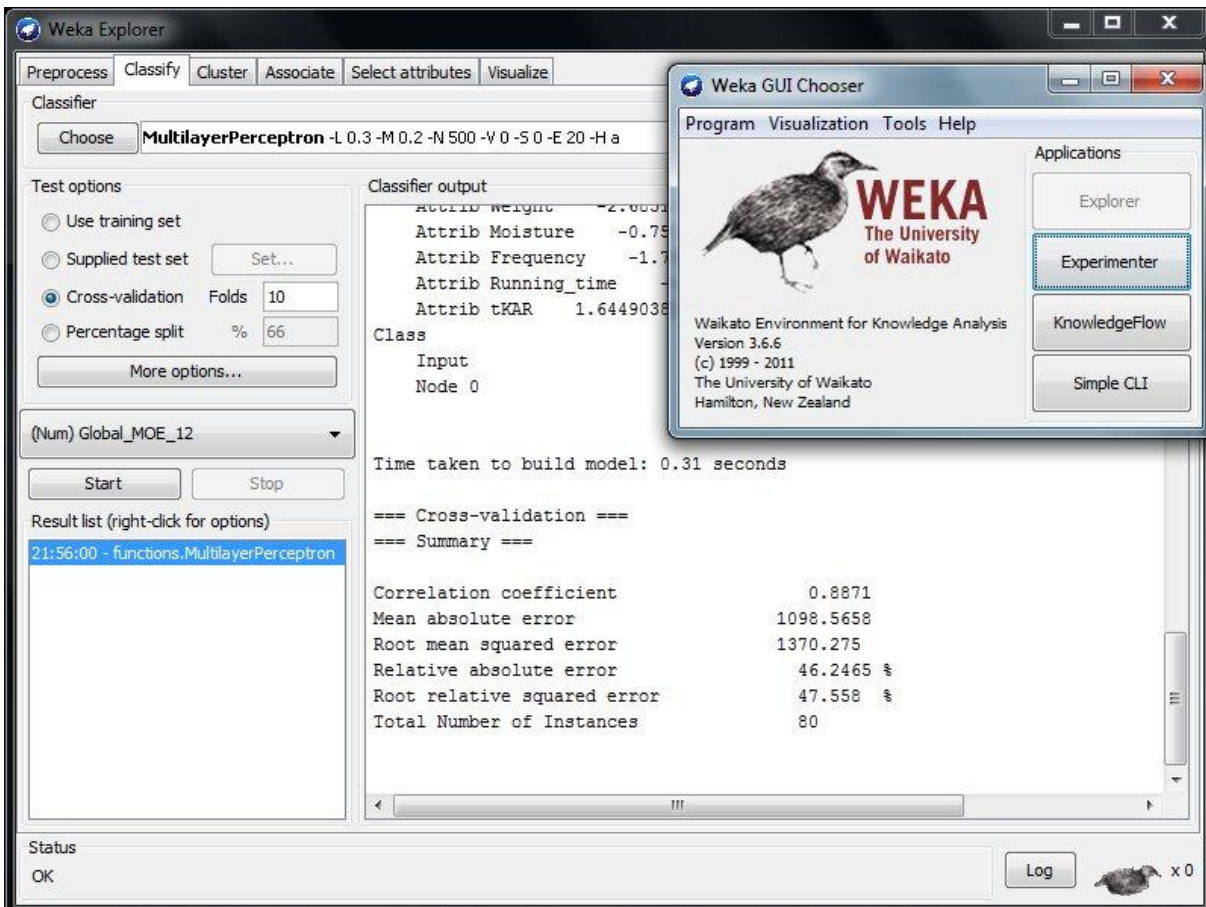
Velikost validacijskega nabora (validationSetSize) – Program lahko določen del nabora podatkov povsem izloči iz analize, ter na koncu, ko je mreža že dokončno izdelana, na njih preveri, ali je dobra tudi za splošne, še ne uporabljene primere. Postopek ni isti kot 'X-fold cross validation' (glej pogl. 3.2.1.1). Premajhen validacijski nabor podatkov lahko pomeni, da smo pri izbiri primerkov imeli smolo, in izbrali tako specifične, da ne kažejo prave vrednosti kakovosti ustvarjene mreže. Prevelik validacijski nabor pomeni, da je za učenje mreže ostalo premalo primerkov. V diplomski nalogi smo izbrali velikost validacijskega vzorca 20 odstotkov, vendar smo to storili ročno, z ločitvijo podatkov na dve datoteki. Tako smo dosegli kontrolo nad validacijskim naborom, ki je v tem primeru vedno isti (program izbira naključno, ob vsakem izračunu druge primerke), to pa omogoča boljšo primerljivost med mrežami, ki so se generirale ob različnih parametrih (število ponovitev, geometrija, učni korak, vztrajnost itd.)

Ustavitev algoritma (validationThreshold) – Algoritem se lahko ustavi, če je doseženo določeno število zaporednih iteracij, ob katerih je rezultat, glede na validacijski nabor, vedno slabši.

3.3 Izbira orodja – WEKA Toolkit

3.3.1 Splošno

Programski paket WEKA [14] je zbirka algoritmov strojnega učenja. Vsebuje orodja za predprocesiranje podatkov, klasifikacijo, grozdenje, regresijo, ugotavljanje povezav znotraj nabora podatkov, ter vizualizacijo podatkovne baze. Prikazan je na sliki 22.



Slika 22: Program WEKA Toolkit

Razdeljen je na štiri sklope:

‘Explorer’ – Osnovni modul programa. Prvi zavihek je namenjen predprocesiranju podatkov, torej osnovnim operacijam s podatkovno bazo ter osnovnim izračunom porazdelitev, korelacij, deviacij ipd. Na zadnjem zavihku se nahajajo orodja za vizualizacijo podatkovne baze. Vsi ostali zavihki so namenjeni algoritmom strojnega učenja. Zavihek ‘Classify’ je namenjen metodam nadzorovanega učenja, tu tudi izberemo način testiranja naučenega sistema (‘X-fold cross validation’, zunanji nabor podatkov ipd.). V zavihku ‘Cluster’ lahko izbiramo med metodami nenadzorovanega učenja. Preostali zavihki so namenjeni splošnim statističnim analizam.

‘Experimenter’ – Ta del programa omogoča, da na več bazah podatkov poženemo več različnih analiz, in z enostavno primerjavo rezultatov hitro ugotovimo, katera metoda je za analizo določene baze najprimernejša. Pri izdelavi diplomske naloge tega dela programa nismo uporabljali, saj je nestabilno deloval že pri posameznih analizah.

‘Knowledge flow’ – V tem delu programa lahko ustvarjamo kompleksne sisteme vzporednih in zaporednih analiz več baz podatkov, avtomatiziramo vizualizacijo ipd. Je alternativa delu programa ‘Experimenter’, ena od razlik pa je, da ima ‘Knowledge flow’ grafični vmesnik.

‘Simple CLI’ – Preprost in zanesljiv konzolni vmesnik. Je zelo priročen ob uporabi navodil iz literature, saj posameznih gumbov za ukaze ni treba iskati po celem programu, ampak se jih enostavno vpiše v ukazno vrstico.

Podatkovno bazo moramo vnesti v datoteko tipa *.arff, ki je tekstovna. Pred vsako vrstico, za katero ne želimo, da je del analize, vnesemo znak ‘%’. Naslov baze vpišemo z ‘@relation naslov_baze’. Spremenljivke definiramo z ‘@attribute ime_spremenljivke tip_spremenljivke’. Za tip spremenljivke imamo na voljo opcije ‘numeric’, ‘real’, ‘string’, ‘class’ and ‘date’. Ker so vse spremenljivke, s katerimi smo delali, numerične in zvezno porazdeljene, smo pri vseh izbrali tip spremenljivke ‘real’.

Po začetnih definicijah vnesemo bazo. Začnemo z vrstico ‘@data’. Vsaka nova vrstica pomeni nov primerek. Spremenljivke ločujemo z vejico, decimalno ločilo je pika. Na mesto manjkajočih vrednosti vpisujemo ‘?’. Konca podatkovne baze ni potrebno označevati.

3.3.2 Regresijska drevesa in nevronske mreže

V tem poglavju je predstavljen natančen postopek, s katerim generiramo regresijsko drevo ali nevronske mreže. Priprava baze podatkov je predstavljena v poglavju 3.3.1. V obeh primerih smo za validacijo že naučenega sistema ročno ločili 20 odstotkov primerkov od ostalega vzorca.

V programu WEKA Explorer v zavihku ‘Preprocess’ izberemo ‘Open file...’ ter poiščemo datoteko z bazo podatkov s končnico ‘*.arff’. Če predhodno tega nismo storili, lahko v tem oknu izbrišemo parametre, ki v analizo ne sodijo. Izbira ‘Edit’ pa nudi vpogled v bazo podatkov, kjer lahko urejamo podatke o posameznih preizkušancih, ter definiramo, kateri parameter je izhodni podatek analize.

3.3.2.1 Regresijska drevesa

V zavihek ‘Classify’ lahko s pritiskom na gumb ‘Choose’ izbiramo med algoritmi strojnega učenja, ki so za izbrano podatkovno bazo na voljo. Ker so v obravnavani bazi podatkov vsi parametri realna števila, lahko izberemo drevo tipa ‘REPTree’. V primerjavi z drugim drevesom, ki je na voljo, torej tip drevesa ‘M5P’, nudi več kriterijev zaustavitve delitve lista (glej pogl. 3.2.1.2). Parametre nastavljam s klikom na okence desno od gumba ‘Choose’, kjer so vrednosti izbranih parametrov tudi zapisane. V tem zavihku tudi izbiramo način izračuna koeficienta korelacije (z navzkrižnim preverjanjem, z validacijskim vzorcem itd.).

Ko poženemo algoritem, se v istem zavihku v polju 'Classifier output' pojavi izdelano regresijsko drevo. Pravila za deljenje so oblike:

tKAR < 0.19

| Frequency < 567.5 : 30.93 (56/105.02) [23/156.29]

| Frequency >= 567.5

| | tKAR < 0.11 : 49.6 (37/149.37) [26/183.83]

| | tKAR >= 0.11 : 38.76 (78/124.14) [37/103.14]

tKAR >= 0.19

| Frequency < 544.5 : 18.63 (281/39.69) [124/51.47]

| Frequency >= 544.5

| | tKAR < 0.28 : 30.47 (127/75.69) [67/57.82]

| | tKAR >= 0.28

| | | Length < 4464 : 23.62 (115/50.98) [69/61.37]

| | | Length >= 4464 : 28.71 (30/40.99) [16/47.32]

Število navpičnih črt pred parametrom pove, na katerem nivoju drevesa se vrši delitev. Ko se list ne deli več, se poleg zadnjega parametra izpiše pričakovana vrednost primerka, ki je razvrščen na ta list. Cele številke v oklepajih pomenijo število primerkov iz nabora za generiranje pravil, ki so se na list uvrstili. Decimalne številke v oklepajih bi morale predstavljati vsoto kvadratov razlik med vrednostjo ocenjevanega parametra primerkov in na listu predpisano vrednostjo, vendar v času izdelave diplomske naloge ta del še ni deloval zanesljivo.

Na dnu polja 'Classifier output' se izpiše tudi koeficient korelacije, povprečna absolutna napaka, povprečje korenov kvadratov razlik ter relativna napaka.

3.3.2.2 Nevronske mreže

V zavihek 'Classify' lahko s klikom na gumb 'Choose' izbiramo med algoritmi strojnega učenja, ki so za izbrano podatkovno bazo na voljo. V mapi 'Functions' izberemo 'MultilayerPerceptron'. Parametre nastavljamo s klikom na okence desno od gumba 'Choose', kjer so vrednosti izbranih parametrov tudi zapisane. V tem zavihku tudi izbiramo način izračuna koeficienta korelacije (z navzkrižnim preverjanjem, z validacijskim vzorcem itd.).

Ko poženemo algoritem, se v istem zavihku v polju 'Classifier output' pojavi izdelana nevronska mreža. Mreža je podana v obliki:

Sigmoid Node 10

Inputs	Weights
Threshold	-2.215270465017345
Attrib Width	0.2948800612723481
Attrib Thickness	0.9616124776549114
Attrib Length	-0.5295465201127751
Attrib Weight	-0.10047073542593579
Attrib Moisture	-0.1018862335359702
Attrib Frequency	-2.9904737125428276
Attrib Running_time	-0.4317660192533671
Attrib tKAR	1.624760753150424

Sigmoid Node 11

Inputs	Weights
Threshold	-2.31337998451882
Node 1	2.771080837408178
Node 2	1.1502037757223456
Node 3	2.577888859571096
Node 4	2.0286129421706196
Node 5	0.9539624637935127
Node 6	-1.0708915279630833
Node 7	1.8421952690170305
Node 8	2.4059091546967055
Node 9	0.6554275608847763
Node 10	1.279786998806342

Za vsako vozlišče (n.pr. node 10) je napisan tip funkcije izhodnega signala (sigmoid), jakost, pri dosegu katere se vozlišče aktivira in pošlje signal (threshold), ter uteži vpliva signalov predhodnih vozlišč (vsi 'Attrib' ali Node 1–10).

Na dnu polja 'Classifier output' se izpiše tudi koeficient korelacije, povprečna absolutna napaka, povprečje korenov kvadratov razlik ter relativna napaka.

4 REZULTATI

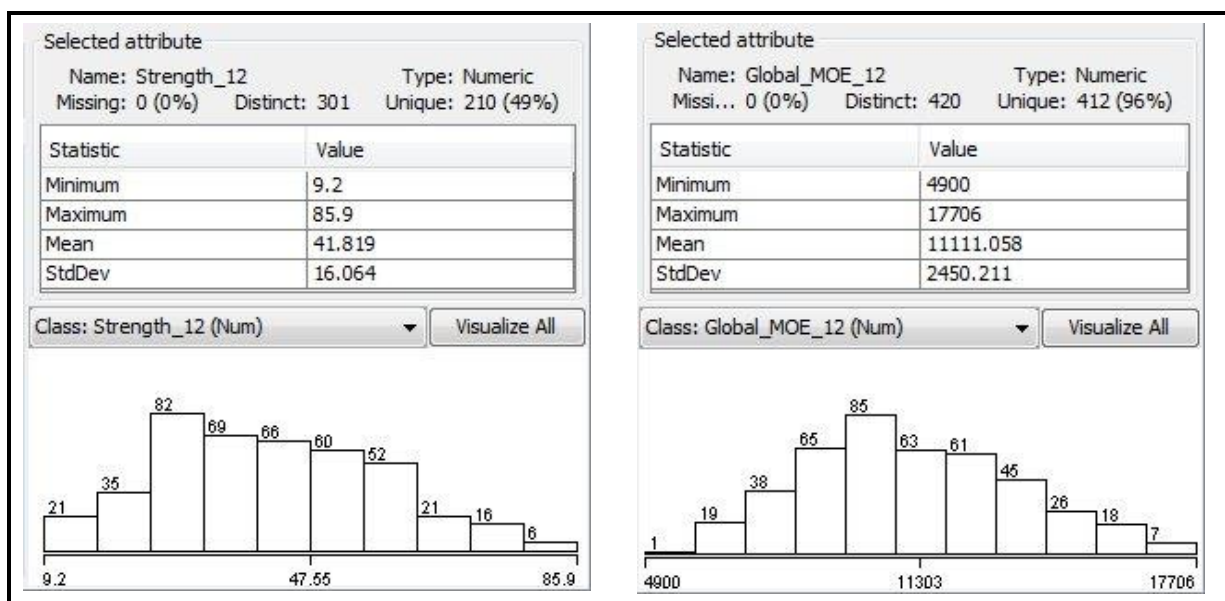
4.1 Porazdelitve vrednosti

V preglednicah od 10 do 13 so prikazane vzorčne porazdelitve izhodnih podatkov, torej globalnega modula elastičnosti ter trdnosti, preračunanih na 12% vlažnost.

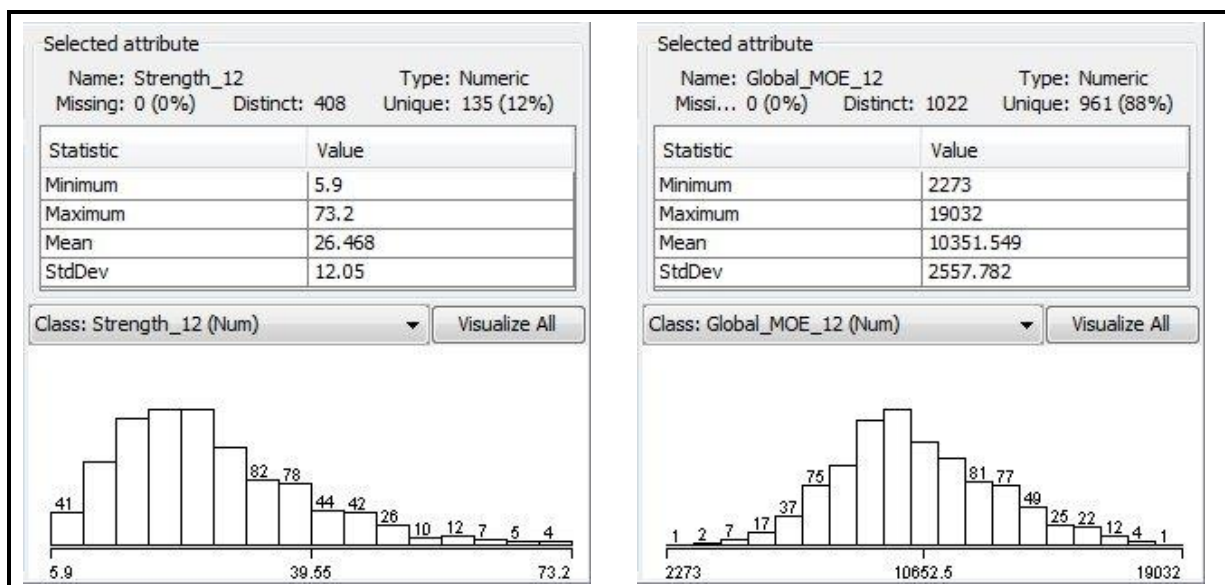
Iz vsakega nabora preizkušancev smo ročno naključno ločili 20% preizkušancev, ki v proces strojnega učenja niso bili vključeni. Služili so kot validacijski vzorec.

Trdnost in modul elastičnosti sta izražena v $[N/mm^2]$.

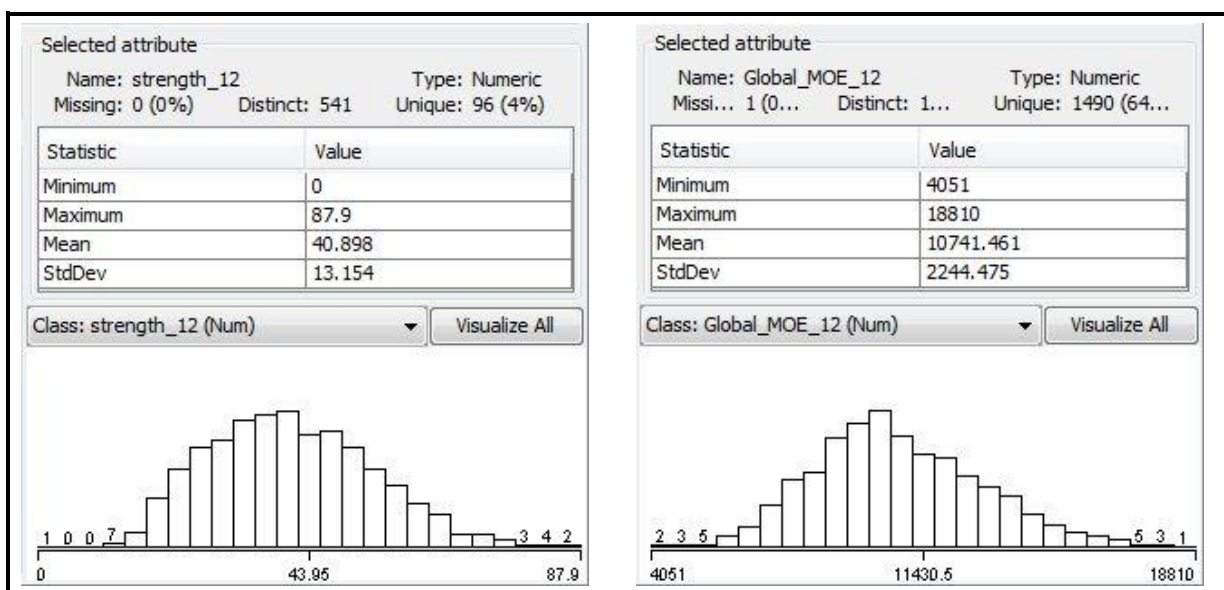
Preglednica 10: Bor – upogib, porazdelitev trdnosti in MOE



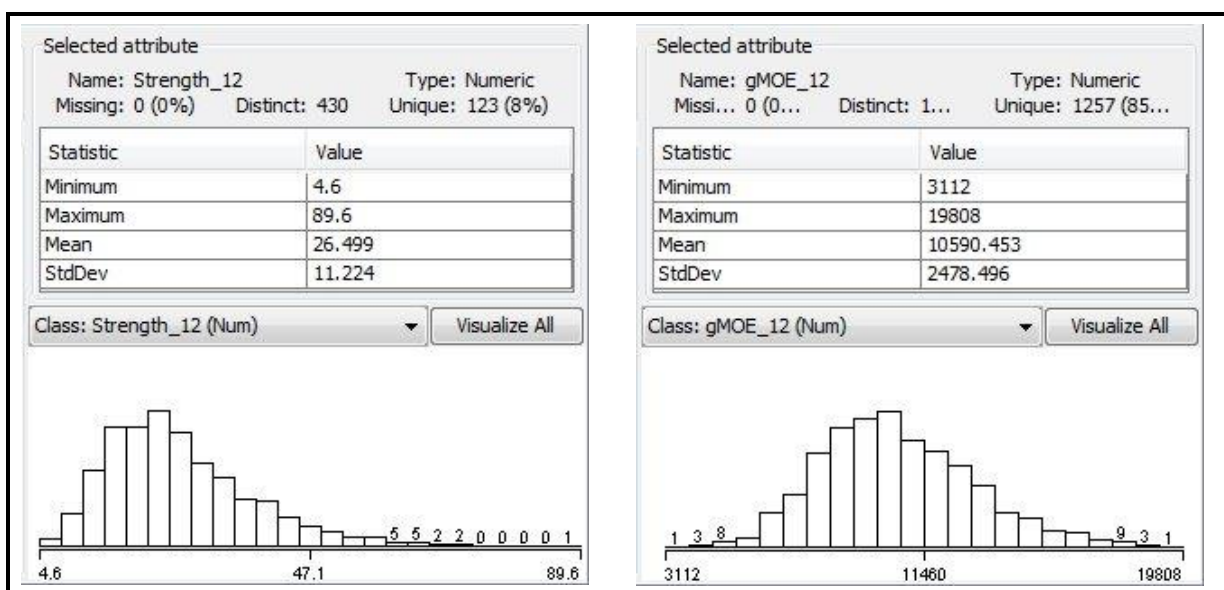
Preglednica 11: Bor – nateg, porazdelitev trdnosti in MOE



Preglednica 12: Smreka – upogib, porazdelitev trdnosti in MOE



Preglednica 13: Smreka – nateg, porazdelitev trdnosti in MOE



4.2 Linearna regresija

Linearna regresija je metoda, s katero iščemo tisto hiper ravnino, ki se čim boljše prilega množici točk v večdimenzionalnem prostoru. Število dimenzij prostora je enako vsoti vhodnih in izhodnih parametrov. Glavna razlika med nevronske mreže in linearno regresijo je ta, da nevronska mreža inducira poljubno ukrivljeno ploskev, linearna regresija pa ravnino. Izračun je zato precej lažji, uveljavljena metoda je metoda najmanjših kvadratov. V preglednicah od 14 do 21 so predstavljeni rezultati linearne regresije.

Preglednica 14: Linearna regresija, bor, upogibni preizkus, MOE

Global_MOE = -79.9162 * Width + -72.1862 * Thickness + 1.9272 * Length + 0.7705 * Weight + -115.5025 * Moisture + 22.9044 * Frequency + -12.5977 * Running_time + 5787.7792	Global_MOE_12 = -11.7416 * Width + 118.4555 * Thickness + 5.3299 * Length + 27.0209 * Frequency + -5.7682 * Running_time + -2626.9744 * tKAR + -24463.212
Determinacijski koeficient: 0.753	Determinacijski koeficient: 0.748
Povprečna absolutna napaka: 1059	Povprečna absolutna napaka: 1225
Povprečna relativna napaka: 9.5 %	Povprečna relativna napaka: 11.0 %

Preglednica 15: Linearna regresija, bor, upogibni preizkus, trdnost

Strength = -0.3183 * Width + -0.3503 * Thickness + 0.0154 * Length + 0.0022 * Weight + 0.1328 * Frequency + -44.2075 * tKAR + -56.2995	Strength_12 = -0.2902 * Width + -0.455 * Thickness + 0.0148 * Length + 0.0025 * Weight + 0.1451 * Frequency + -40.9677 * tKAR + -67.1269
Determinacijski koeficient: 0.707	Determinacijski koeficient: 0.706
Povprečna absolutna napaka: 8.12	Povprečna absolutna napaka: 9.31
Povprečna relativna napaka: 19.4 %	Povprečna relativna napaka: 22.3 %

Preglednica 16: Linearna regresija, bor, natezni preizkus, MOE

Izveden je bil le preizkus Local_MOE, iz katerega se je Global_MOE_12 izračunal naknadno. Rezultati Local_MOE niso posebej obdelani.	$\begin{aligned} \text{Global_MOE_12} = & \\ & -55.3544 * \text{Width} + \\ & -132.3125 * \text{Thickness} + \\ & \quad 3.484 * \text{Length} + \\ & \quad 0.6068 * \text{Weight} + \\ & -114.1362 * \text{Moisture} + \\ & \quad 28.2419 * \text{Frequency} + \\ & \quad -3.9898 * \text{Running_time} + \\ & -4166.3233 * \text{tKAR} + \\ & -9135.3082 \end{aligned}$
	Determinacijski koeficient : 0.783
	Povprečna absolutna napaka: 925
	Povprečna relativna napaka: 8.94 %

Preglednica 17: Linearna regresija, bor, natezni preizkus, trdnost

$\begin{aligned} \text{Strength} = & \\ & -0.238 * \text{Width} + \\ & -0.629 * \text{Thickness} + \\ & \quad 0.0123 * \text{Length} + \\ & \quad 0.0021 * \text{Weight} + \\ & \quad 0.3412 * \text{Moisture} + \\ & \quad 0.1038 * \text{Frequency} + \\ & \quad -0.0088 * \text{Running_time} + \\ & -38.7679 * \text{tKAR} + \\ & -36.872 \end{aligned}$	$\begin{aligned} \text{Strength_12} = & \\ & -0.1821 * \text{Width} + \\ & -0.5562 * \text{Thickness} + \\ & \quad 0.0127 * \text{Length} + \\ & \quad 0.0019 * \text{Weight} + \\ & \quad 0.1164 * \text{Frequency} + \\ & -39.9869 * \text{tKAR} + \\ & -58.0516 \end{aligned}$
Determinacijski koeficient: 0.687	Determinacijski koeficient: 0.723
Povprečna absolutna napaka: 5.06	Povprečna absolutna napaka: 4.66
Povprečna relativna napaka: 19.1 %	Povprečna relativna napaka: 17.6 %

Preglednica 18: Linearna regresija, smreka, upogibni preizkus, MOE

Global_MOE = -38.3547 * Width + -144.2967 * Thickness + 3.2657 * Length + 0.5527 * Weight + 225.2692 * Moisture + 30.9628 * Frequency + -1.8759 * Running_time + -1140.6142 * tKAR + -17757.4771	Global_MOE_12 = -39.4746 * Width + -146.5503 * Thickness + 3.0409 * Length + 0.5615 * Weight + 347.0377 * Moisture + 30.3267 * Frequency + -2.2799 * Running_time + -1502.5721 * tKAR + -17471.3427
Determinacijski koeficient: 0.766	Determinacijski koeficient: 0.826
Povprečna absolutna napaka: 765	Povprečna absolutna napaka: 683
Povprečna relativna napaka: 7.1 %	Povprečna relativna napaka: 6.4 %

Preglednica 19: Linearna regresija, smreka, upogibni preizkus, trdnost

Strength = -0.1815 * Width + -0.4413 * Thickness + 0.0174 * Length + 0.0018 * Weight + 1.1928 * Moisture + 0.1458 * Frequency + -33.8058 * tKAR + -102.6336	strength_12 = -0.1323 * Width + -0.4744 * Thickness + 0.0165 * Length + 0.0019 * Weight + 1.1184 * Moisture + 0.1517 * Frequency + 0.0219 * Running_time + -33.8367 * tKAR + -125.4747
Determinacijski koeficient: 0.579	Determinacijski koeficient: 0.570
Povprečna absolutna napaka: 6.69	Povprečna absolutna napaka: 6.88
Povprečna relativna napaka: 16.4 %	Povprečna relativna napaka: 16.8 %

Preglednica 20: Linearna regresija, smreka, natezni preizkus, MOE

<p>Global_MOE =</p> <p>-46.8774 * Width + 17.1379 * Length + 0.5547 * Weight + -301.8005 * Moisture + 40.8197 * Frequency + 17.7661 * Running_time + 761.4279 * tKAR + -93733.625</p>	<p>gMOE_12 =</p> <p>-50.1002 * Width + -107.9943 * Thickness + 3.6231 * Length + 0.6839 * Weight + 120.8522 * Moisture + 31.8968 * Frequency + -4.8802 * Running_time + -1388.3757 * tKAR + -17488.0642</p>
Determinacijski koeficient: 0.785	Determinacijski koeficient: 0.799
Povprečna absolutna napaka: 985	Povprečna absolutna napaka: 790
Povprečna relativna napaka: 9.3 %	Povprečna relativna napaka: 7.5 %

Preglednica 21: Linearna regresija, smreka, natezni preizkus, trdnost

<p>Strength =</p> <p>-0.1493 * Width + -0.1806 * Thickness + 0.0138 * Length + 0.0017 * Weight + 0.1508 * Frequency + 0.0162 * Running_time + -22.2071 * tKAR + -119.2055</p>	<p>Strength_12 =</p> <p>-0.1302 * Width + -0.1968 * Thickness + 0.0115 * Length + 0.0019 * Weight + 0.1471 * Frequency + 0.0198 * Running_time + -22.9876 * tKAR + -115.0483</p>
Determinacijski koeficient: 0.663	Determinacijski koeficient: 0.717
Povprečna absolutna napaka: 4.78	Povprečna absolutna napaka: 4.23
Povprečna relativna napaka: 18.0 %	Povprečna relativna napaka: 16.0 %

4.3 Rezultati nevronske mreže

Uporabili smo umetno nevronske mrežo s tremi skritimi plastmi. Na prvi je bilo 14 vozlišč, na drugi in tretji pa po 4. Drobljenja in nominalnega filtra nismo uporabili. Pri vseh variantah smo izbrali normalizacijo nazivnih in numeričnih parametrov. Ponovni začetek izračuna (opcija 'reset') ni bil izbran, spreminjanja koeficienta generatorja naključnih števil pa rezultata bistveno ne spremeni. Ker smo 20 % primerkov že ročno ločili od celotnega vzorca, smo velikost validacijskega vzorca izbral 0.

Ostali parametri, pripadajoče absolutne in relativne napake, ter koeficienti korelacije, so prikazani v preglednicah od 22 do 36.

Preglednica 22: Parametri nevronske mreže, bor, upogibni preizkus, MOE

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.3	0.2	500	816	7,3%	0.861
2	0.3	0.2	50	2262	20,4%	0.476
3	0.3	0.2	5000	917	8,3%	0.781
4	0.3	0.2	2000	852	7,7%	0.830
5	0.3	0.2	1000	809	7,3%	0.857
6	0.2	0.2	1000	801	7,2%	0.869
7	0.1	0.2	1000	792	7,1%	0.861
8	0.2	0.1	1000	770	6,9%	0.874
9	0.02	0.01	10000	764	6,9%	0.839

Preglednica 23: Parametri nevronske mreže, bor, upogibni preizkus, MOE 12 %

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.2	0.1	1000	752	6,8%	0.859
2	0.22	0.1	1000	765	6,9%	0.852
3	0.18	0.1	1000	748	6,7%	0.863
4	0.18	0.12	1000	748	6,7%	0.861
5	0.18	0.1	1500	790	7,1%	0.850

Preglednica 24: Parametri nevronske mreže, bor, upogibni preizkus, trdnost

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.18	0.12	1000	8.19	19,6%	0.699
2	0.20	0.12	1000	8.30	19,9%	0.692
3	0.16	0.12	1000	8.18	19,6%	0.704
4	0.16	0.12	1500	8.41	20,1%	0.667
5	0.16	0.10	1500	8.37	20,0%	0.671
6	0.16	0.10	5000	10.91	26,1%	0.458
7	0.16	0.10	500	7.91	18,9%	0.721
8	0.16	0.10	400	7.83	18,7%	0.724
9	0.18	0.06	300	7.80	18,7%	0.726

Preglednica 25: Parametri nevronske mreže, bor, upogibni preizkus, trdnost 12 %

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.18	0.06	300	8.61	20,6%	0.500
2	0.18	0.06	1000	8.74	20,9%	0.504
3	0.18	0.10	500	8.66	20,7%	0.510
4	0.30	0.15	500	9.00	21,5%	0.510
5	0.3	0.15	1500	11.1	26,6%	0.379

Preglednica 26: Parametri nevronske mreže, bor, natezni preizkus, MOE 12%

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.18	0.1	1000	737	7,1%	0.852
2	0.18	0.1	1500	749	7,2%	0.848
3	0.18	0.1	300	753	7,3%	0.857
4	0.2	0.3	300	748	7,2%	0.857

Preglednica 27: Parametri nevronske mreže, bor, natezni preizkus, trdnost

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.18	0.06	300	4.57	17,2%	0.746
2	0.18	0.06	500	4.64	17,5%	0.733
3	0.18	0.06	200	4.61	17,4%	0.746
4	0.18	0.16	300	4.56	17,2%	0.745

Preglednica 28: Parametri nevronske mreže, bor, natezni preizkus, trdnost 12 %

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.18	0.06	300	4.15	15,7%	0.771
2	0.16	0.06	500	4.18	15,8%	0.776
3	0.16	0.05	700	4.25	16,0%	0.774
4	0.10	0.10	700	4.17	15,7%	0.774
5	0.18	0.1	200	4.28	16,2%	0.755

Preglednica 29: Parametri nevronske mreže, smreka, upogibni preizkus, MOE

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.2	0.3	300	663	6,2%	0.810
2	0.2	0.3	1000	636	5,9%	0.821
3	0.15	0.3	1000	644	6,0%	0.821
4	0.15	0.6	1000	638	5,9%	0.821
5	0.15	0.6	5000	675	6,3%	0.824
6	0.15	0.6	15000	714	6,6%	0.810
7	0.7	0.7	500	824	7,7%	0.752

Preglednica 30: Parametri nevronske mreže, smreka, upogibni preizkus, MOE 12 %

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.2	0.3	1000	612	5,7%	0.876
2	0.3	0.4	1000	617	5,7%	0.863
3	0.15	0.2	1500	585	5,4%	0.878
4	0.15	0.2	300	634	5,9%	0.865

Preglednica 31: Parametri nevronske mreže, smreka, upogibni preizkus, trdnost

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.2	0.3	1000	6.63	16,2%	0.598
2	0.2	0.3	300	6.71	16,4%	0.585
3	0.3	0.2	300	6.83	16,7%	0.584
4	0.1	0.06	1500	6.53	16,0%	0.604
5	0.02	0.03	1500	6.66	16,3%	0.591
6	0.02	0.03	200	10.8	26,4%	0.194

Preglednica 32: Parametri nevronske mreže, smreka, upogibni preizkus, trdnost 12 %

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.2	0.3	1000	6.66	16,3%	0.587
2	0.2	0.3	300	6.66	16,3%	0.590
3	0.4	0.3	300	7.53	18,4%	0.587
4	0.2	0.45	300	6.65	16,3%	0.587

Preglednica 33: Parametri nevronske mreže, smreka, natezni preizkus, MOE

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R ²
1	0.2	0.3	300	613	5,8%	0.884
2	0.3	0.3	300	712	6,7%	0.876
3	0.15	0.2	300	607	5,7%	0.889
4	0.15	0.2	1000	564	5,3%	0.901
5	0.15	0.15	1000	566	5,3%	0.899
6	0.1	0.2	1000	554	5,2%	0.904

Preglednica 34: Parametri nevronske mreže, smreka, natezni preizkus, MOE 12 %

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R ²
1	0.1	0.2	1000	690	6,5%	0.826
2	0.1	0.2	500	731	6,9%	0.823
3	0.15	0.2	1000	719	6,8%	0.814
4	0.05	0.2	1000	686	6,5%	0.832
5	0.05	0.15	1000	689	6,5%	0.832

Preglednica 35: Parametri nevronske mreže, smreka, natezni preizkus, trdnost

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.2	0.3	300	4.47	16,9%	0.697
2	0.2	0.3	1000	4.37	16,5%	0.706
3	0.15	0.3	1000	4.32	16,3%	0.712
4	0.15	0.25	1000	4.3	16,2%	0.712
5	0.13	0.18	2000	4.44	16,8%	0.701
6	0.13	0.18	400	4.44	16,8%	0.697

Preglednica 36: Parametri nevronske mreže, smreka, natezni preizkus, trdnost 12 %

Primer	Učni korak	Vztrajnost	Število iteracij	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	0.15	0.25	1000	5.26	19,8%	0.759
2	0.2	0.25	1000	5.88	22,2%	0.752
3	0.12	0.2	1000	4.87	18,4%	0.762
4	0.12	0.2	2000	5.10	19,2%	0.719

4.4 Rezultati regresijskih dreves

Pri vseh regresijskih drevesih smo izbrali možnost naknadnega rezanja drevesa, saj ustrezno posplošena drevesa niso prenaučena, zato nove preizkušance razvrščajo bolje. Variranje parametra generatorja naključnih števil rezultata ne spremeni bistveno.

Vsi ostali parametri so predstavljeni v preglednicah od 37 do 51.

Parameter največje globine, nastavljene na -1 pomeni, da globina ni omejena, ter da se delitev ustavi zaradi drugih kriterijev.

Preglednica 37: Parametri regresijskih dreves, bor, upogibni preizkus, MOE

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	-1	1	0.001	147	1432	12,9%	0.549
2	-1	3	0.001	47	1463	13,2%	0.539
3	-1	5	0.001	29	1575	14,2%	0.472
4	-1	3	0.01	47	1463	13,2%	0.539
5	-1	3	0.1	45	1469	13,2%	0.534
6	5	3	0.1	23	1750	15,8%	0.407
7	4	3	0.1	17	1895	17,1%	0.264

Preglednica 38: Parametri regresijskih dreves, bor, upogibni preizkus, MOE 12 %

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel.. napaka	Determinacijski koeficient R^2
1	-1	1	0.001	131	1413	12,7%	0.591
2	-1	3	0.01	29	1509	13,6%	0.543
3	-1	5	0.05	17	1548	13,9%	0.527
4	5	5	0.05	15	1536	13,8%	0.526
5	4	5	0.05	11	1667	15,0%	0.494
6	3	5	0.05	7	1819	16,4%	0.452

Preglednica 39: Parametri regresijskih dreves, bor, upogibni preizkus, trdnost

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel.. napaka	Determinacijski koeficient R^2
1	-1	1	0.001	97	9.74	23,3%	0.508
2	-1	3	0.01	27	9.23	22,1%	0.530
3	-1	5	0.05	23	9.48	22,7%	0.504
4	4	5	0.05	11	10.90	26,1%	0.277

Preglednica 40: Parametri regresijskih dreves, bor, upogibni preizkus, trdnost 12 %

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R ²
1	-1	1	0.001	115	9.62	23,0%	0.362
2	-1	3	0.01	35	9.45	22,6%	0.372
3	-1	5	0.05	11	10.29	24,6%	0.264
4	-1	1	0.2	59	9.40	22,5%	0.388
5	-1	1	0.3	45	8.74	20,9%	0.444
6	-1	3	0.3	29	9.12	21,8%	0.397
7	5	3	0.3	9	10.21	24,4%	0.281

Preglednica 41: Parametri regresijskih dreves, bor, natezni preizkus, MOE 12 %

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R ²
1	-1	1	0.001	249	1210	11,7%	0.635
2	-1	3	0.001	87	1236	11,9%	0.627
3	-1	1	0.2	81	1256	12,1%	0.602
4	-1	3	0.3	33	1286	12,4%	0.598
5	4	3	0.3	13	1384	13,4%	0.527

Preglednica 42: Parametri regresijskih dreves, bor, natezni preizkus, trdnost

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	-1	1	0.001	173	6.59	24,9%	0,511
2	-1	1	0.3	65	6.29	23,7%	0,521
3	-1	5	0.3	31	6.13	23,1%	0,552
4	-1	10	0.3	19	6.44	24,3%	0,516
5	5	10	0.3	13	6.36	24,0%	0,527

Preglednica 43: Parametri regresijskih dreves, bor, natezni preizkus, trdnost 12 %

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	-1	1	0.001	245	5.84	22,0%	0.588
2	-1	1	0.3	43	6.45	24,3%	0.504
3	-1	5	0.3	31	6.56	24,8%	0.479
4	-1	10	0.4	13	6.61	24,9%	0.473

Preglednica 44: Parametri regresijskih dreves, smreka, upogibni preizkus, MOE

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	-1	1	0.001	443	1006	9,4%	0.654
2	-1	1	0.25	59	1047	9,7%	0.635
3	-1	20	0.25	31	1041	9,7%	0.627
4	4	20	0.2	25	1040	9,7%	0.626
5	3	20	0.2	15	1141	10,6%	0.560

Preglednica 45: Parametri regresijskih dreves, smreka, upogibni preizkus, MOE 12 %

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R ²
1	-1	1	0.001	501	944	8,8%	0,672
2	-1	1	0.3	41	969	9,0%	0,674
3	-1	20	0.3	29	989	9,2%	0,658
4	3	20	0.3	13	1119	10,4%	0,569

Preglednica 46: Parametri regresijskih dreves, smreka, upogibni preizkus, trdnost

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R ²
1	-1	1	0.001	429	7.93	19,4%	0,426
2	-1	1	0.25	131	7.43	18,2%	0,482
3	-1	15	0.25	35	7.45	18,2%	0,472
4	4	15	0.25	21	7.58	18,5%	0,468
5	3	15	0.25	15	7.85	19,2%	0,433

Preglednica 47: Parametri regresijskih dreves, smreka, upogibni preizkus, trdnost 12 %

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R ²
1	-1	1	0.001	329	7.87	19,2%	0.412
2	-1	1	0.3	95	7.52	18,4%	0.453
3	-1	30	0.3	21	7.72	18,9%	0.448
4	-1	30	0.38	17	7.76	19,0%	0.442
5	3	30	0.38	11	7.93	19,4%	0.423

Preglednica 48: Parametri regresijskih dreves, smreka, natezni preizkus, MOE

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel.. napaka	Determinacijski koeficient R^2
1	-1	1	0.001	249	1012	9,6%	0.687
2	-1	1	0.3	27	1104	10,4%	0.686
3	-1	15	0.33	17	1213	11,5%	0.616
4	3	15	0.3	13	1185	11,2%	0.634

Preglednica 49: Parametri regresijskih dreves, smreka, natezni preizkus, MOE 12 %

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel.. napaka	Determinacijski koeficient R^2
1	-1	1	0.001	393	1229	11,6%	0.582
2	-1	1	0.15	119	1192	11,3%	0.610
3	-1	1	0.35	29	1298	12,3%	0.555
4	-1	20	0.35	11	1347	12,7%	0.514

Preglednica 50: Parametri regresijskih dreves, smreka, natezni preizkus, trdnost

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel.. napaka	Determinacijski koeficient R^2
1	-1	1	0.001	273	5.86	22,1%	0.500
2	-1	1	0.3	23	5.66	21,4%	0.546
3	-1	20	0.3	11	5.90	22,3%	0.523

Preglednica 51: Parametri regresijskih dreves, smreka, natezni preizkus, trdnost 12 %

Primer	Največja globina	Min. št. preizk. na listu	Min. interval izh. skal.	Velikost drevesa	Pov. abs. napaka	Pov. rel. napaka	Determinacijski koeficient R^2
1	-1	1	0.001	283	5.27	19,9%	0.598
2	-1	1	0.25	67	5.12	19,3%	0.612
3	-1	20	0.3	21	5.18	19,5%	0.594
4	3	20	0.3	13	5.17	19,5%	0.594
5	3	50	0.5	7	5.90	22,3%	0.490

4.5 Determinacijski koeficienti R^2

Preglednica 52: Determinacijski koeficienti R^2

	Linearna regresija	Nevronska mreža	Regresijsko drevo
BOR – upogib			
Trdnost	0.707	0.726	0.530
Trdnost – 12%	0.706	0.510	0.444
MOE	0.753	0.869	0.549
MOE – 12%	0.748	0.863	0.591
BOR – nateg			
Trdnost	0.687	0.746	0.552
Trdnost – 12%	0.723	0.776	0.588
MOE – 12%	0.783	0.857	0.635
SMREKA – upogib			
Trdnost	0.579	0.604	0.482
Trdnost – 12%	0.570	0.590	0.453
MOE	0.766	0.824	0.654
MOE – 12%	0.826	0.878	0.674
SMREKA – nateg			
Trdnost	0.663	0.712	0.546
Trdnost – 12%	0.717	0.762	0.612
MOE	0.785	0.904	0.687
MOE – 12%	0.799	0.832	0.610
SMREKA – upogib, trdnost, SLO			
vsi SLO skupaj	0.587	0.600	
samo Notranjska	0.585	0.585	
samo Koroška	0.526	0.546	
samo Osrednjeslov.	0.608	0.609	

Komentar

Pokazalo se je, da so spremenljivke med seboj precej linearno odvisne, saj smo že s preprosto linearno regresijo dosegli zelo dobre rezultate.

Nevronske mreže so se izkazale kot zelo dobro orodje za določevanje trdnosti in modula elastičnosti iz podatkov, ki so pri pregledu preizkušancev v splošnem na voljo. Zaradi velikega vzorca je rešitev konvergirala znotraj širokega spektra vhodnih parametrov. Mreža ne daje dobrih rezultatov le v primeru vnosa ekstremnih parametrov. Zelo veliko nastavitvev parametrov je rezultiralo v korelaciji, manj kot 3 odstotke manjši od največje dosežene.

Ustvarjena regresijska drevesa niso dosegla tako visokih korelacij. Razlog gre iskati v dejstvu, da se odločitvena drevesa dobro obnesejo v primerih, ko vrednost enega parametra v enih primerih bistveno vpliva na rezultat, v drugih pa zelo malo, ali pa celo nasprotno (n.pr. če je vhodni parameter diskretna spremenljivka). Zveznih, skoraj linearnih funkcij, regresijska drevesa ne rešujejo dobro.

Pri analizi vzorca z območja Slovenije smo ugotovili, da imajo pri teh vhodnih podatkih preizkušanci z določenega območja lahko manjši raztros in zato nekoliko višji koeficient korelacije. Je pa za učenje dobre nevronske mreže po naši oceni potrebnih vsaj 1000 preizkušancev s posameznega področja v primeru smreke in vsaj 500 v primeru bora. Bor namreč raste počasneje in bolj enakomerno, zato je tudi raztros rezultatov manjši.

4.6 Preglednica vpliva parametrov

Za štiri primere smo izračunali vpliv vhodnih parametrov. Pri linearni regresiji to storimo tako, da povprečje posameznega parametra pomnožimo s faktorjem, ki smo ga dobili kot rezultat linearne regresije. Za vsak parameter je ustrezen faktor razviden v preglednicah v poglavju 4.2. Absolutne vrednosti dobljenih produktov seštejemo, prištejemo še konstantni člen, posamezen produkt faktorja s parametrom množimo s 100 in delimo s to vsoto, rezultat pa pomeni vpliv posameznega parametra v procentih. Vplivi so predstavljeni v preglednici 53.

V isti preglednici so predstavljeni tudi kriteriji, ki med prvimi delijo regresijsko drevo. Kriteriji delitve so v splošnem na istem nivoju različni. Prej kot je parameter uporabljen kot kriterij delitve, bolj pomemben je za končni rezultat.

Preglednica 53: Vpliv parametrov

	Linearna regresija		Regresijska drevesa	
	Parameter	Odstotek vpliva	Parameter	Nivo delitve
BOR – upogib, trdnost, 12% vlažnost	Lastna frekvenca	28.6%	tKAR	1
	Konstantni člen	20.9%	Lastna frekvenca	2
	Dolžina	17.6%	tKAR	2
	Širina	13.5%	Širina	3
	Teža	10.3%	tKAR	3
	Debelina	6.0%		
	tKAR	3.0%		
BOR – nateg, MOE, 12% vlažnost	Lastna frekvenca	23,9%	Lastna frekvenca	1
	Dolžina	21.2%	Lastna frekvenca	2
	Konstantni člen	13.5%	tKAR	2
	Teža	12.4%	Teža	3
	Širina	11.8%	Dolžina	3
	Debelina	8.6%		
	Čas širjenja zvoka	4.8%		
	Vsebnost vlage	2.0%		
tKAR	1.7%			
SMREKA – upogib, MOE, 12% vlažnost	Lastna frekvenca	25.7%	Lastna frekvenca	1
	Konstantni člen	22.9%	Lastna frekvenca	2
	Dolžina	16.1%	Teža	3
	Teža	9.9%	Dolžina	3
	Debelina	9.3%		
	Širina	8.0%		
	Vsebnost vlage	5.4%		
	Čas širjenja zvoka	2.2%		
	tKAR	0.5%	Lastna frekvenca	1
SMREKA – nateg, trdnost, 12% vlažnost	Konstantni člen	35.0%	Lastna frekvenca	2
	Lastna frekvenca	28.4%	tKAR	2
	Dolžina	14.7%	Teža	3
	Teža	6.9%	Dolžina	3
	Širina	5.9%		
	Čas širjenja zvoka	4.5%		
	Debelina	2.7%		
	tKAR	2.0%		

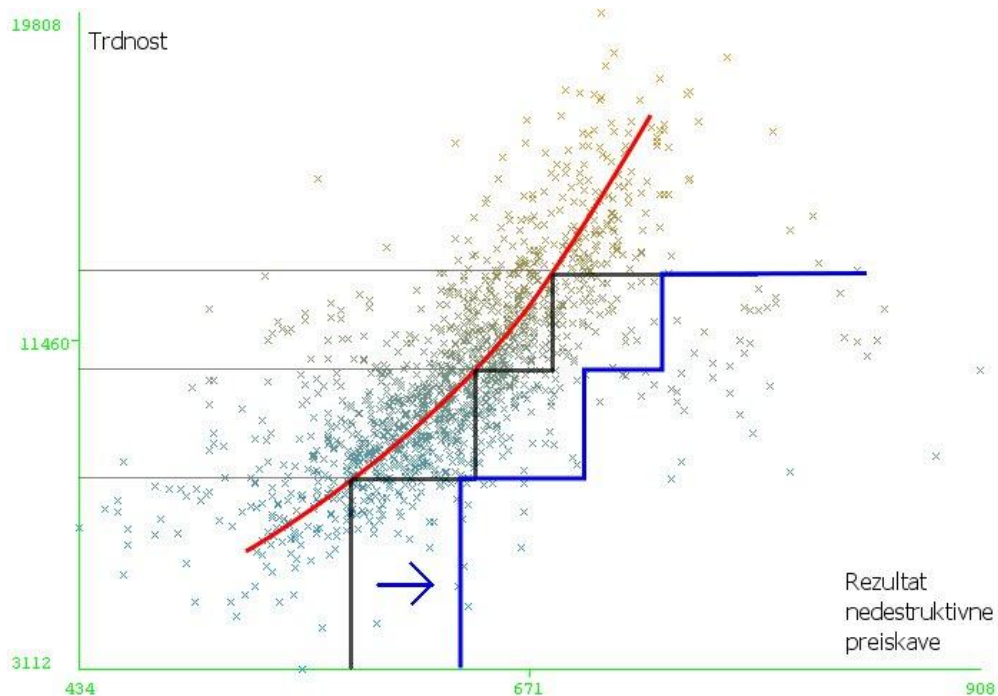
Opazimo, da v obeh primerih kot eden najpomembnejših parametrov nastopa lastna frekvenca preizkušanca. Prav tako visoko kotirajo tudi parametri, ki neposredno ne vplivajo na trdnost in elastični modul lesa, vendar jih nujno potrebujemo za normiranje ostalih parametrov, to so geometrijske karakteristike preizkušancev. Presenečenje pa je koeficient grč tKAR, ki po metodi regresijskih dreves kotira zelo visoko, linearna regresija pa ga praktično zanemari. Tako velika razlika lahko pomeni, da bi morali grče numerično opisovati na popolnoma drugačen način.

4.7 Razvrščanje vzorcev lesa v trdnostne razrede

Ko smo po eni od izbranih metod dobili statistično povezavo med vhodnimi in izhodnimi podatki, lesenih preizkušancev zaradi raztrosa še ne moremo razvrščati v razrede. Če bi uspeli posamezen preizkušanec z nedestruktivnimi metodami analizirati tako dobro, da bi ustvarjen statistični sistem nudil 100% korelacijo, bi bilo za razvrščanje novih primerkov to dovolj. Ker pa je nemogoče, da bi bili vhodni podatki tako natančni, okrog rešitve obstaja določen raztros.

Kot smo omenili v poglavju 3.2.2.1, z nevronske mreže poiščemo večdimenzionalno ploskev nad večdimenzionalno ravnino. Ta ravnina predstavlja nekakšno sredino oblaka točk (vsaka točka v tem večdimenzionalnem prostoru predstavlja podatke enega primerka). Zato približno polovica točk leži pod, polovica pa nad ploskvijo.

Ko razvrščamo lesene preizkušance v razrede, ploskev spremenimo v stopničasto linearno funkcijo tako, kot je prikazano na sliki 23. Sive vodoravne črte predstavljajo razrede, v katere razvrščamo les. Ker standard zahteva, da 95% preizkušancev ustreza minimalnim kriterijem razreda, moramo nastalo stopničasto funkcijo premakniti v smeri naraščanja razredov na oseh vhodnih podatkov tako, da bo pod funkcijo ostalo le še 5% primerkov. Linearna translacija je za eno dimenzijo prikazana na sliki 23.



Slika 23: Razvrščanje v razrede

5 ZAKLJUČEK

V diplomski nalogi sem ugotovil, da je bil vzorec preizkušancev dovolj velik za praktično uporabo. Opravljene statistične metode so dovolj dobre za določitev trdnosti in elastičnega modula preizkušancev ter za zanesljivo določitev raztrosa.

Potrdili smo tezo, da se umetne nevronske mreže lahko uporabljajo pri avtomatičnem razvrščanju lesenih preizkušancev v trdnostne razrede. Ko so poleg arhitekture in uteži umetne nevronske mreže izdelana še pravila za razvrščanje, se lahko sistem vgradi v stroj za avtomatično razvrščanje lesa, ki na podlagi rezultatov nedestruktivnih preiskav vsak preizkušanec hitro in zanesljivo uvrsti v enega od izbranih trdnostnih razredov.

Regresijska drevesa so se zaradi linearnosti in zveznosti problema pri razvrščanju odrezala nekoliko slabše kot nevronske mreže. Opaziti pa je bilo, da korelacijski koeficient tudi pri zelo grobem rezanju drevesa ostaja približno enak. Regresijska drevesa so bila koristna tudi zato, ker smo pri delu z njimi opazili, da velika odstopanja v izračunu nastanejo zaradi edine količine, ki v svoji informaciji linearizira nelinearen problem – tKAR. Sklepamo, da bi z izboljšanjem postopka pridobivanja informacije o grčah preizkušanca izboljšali rezultate.

Potrdilo se je, da imajo borovi preizkušanci zaradi počasnejše in enakomernejše rasti manjši raztros rezultatov kot preizkušanci iz smreke. Višji koeficient korelacije smo, kljub manjšemu številu primerkov, dosegli v vseh primerih.

V naslednjem koraku želimo z istimi metodami razred preizkušancev določiti direktno, brez vmesnega računanja trdnosti in modula elastičnosti. V račun želimo vključiti tudi poreklo preizkušancev. Sklepamo, da bodo koeficienti korelacije pri uporabi nevronske mreže primerljivi s tistimi, ki jih bomo dobili pri uporabi regresijskih dreves.

VIRI

- [1] SIST EN 338:2003. Konstrukcijski les – Trdnostni razredi.
- [2] Šinkovec, N. 2008. Uporaba umetnih nevronske mreže za oceno trdnosti lesenih elementov. Diplomsko naloga. Ljubljana, Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo: 91 str.
- [3] Čepon, M. 2006. Nedestruktivne metode določanja relevantnih gradacijskih lastnosti žaganega lesa. Diplomsko naloga. Ljubljana, Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za lesarstvo: 83 str.
- [4] DIN 4074-1:2009. Vizualno razvrščanje konstrukcijskega žaganega lesa.
- [5] SIST EN 408:2010. Lesene konstrukcije – Konstrukcijski les in lepljeni lamelirani les – Ugotavljanje nekaterih fizikalnih in mehanskih lastnosti.
- [6] Caudill, M., Butler, C. 1991. Naturally intelligent systems. Cambridge, Massachusetts, Massachusetts Institute of Technology: 2–36.
- [7] Topping, B.H.V., Bahreininejad, A. 1997. Neural Computing for Structural Mechanics. United Kingdom, Edinburgh, Saxe–Coburg Publications: 178 str.
- [8] Stankovski, V. 1995. Modeliranje populacijske dinamike jelenjadi glede na razvoj gozda. Diplomsko delo. Ljubljana, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko: 72 str.
- [9] Smrke, D., Demšar, J., Bratko, I., Kersnič, B., Stankovski, V. 1998. Analiza implantacij kolčnih endoprotez z regresijskimi drevesi = Regression tree analysis of indications concerning hip arthroplasty. Informatica medica slovenica 5, 1–2: 111–116.
- [10] Reščič, D. 2004. Strojno razvrščanje lesa z uporabo umetnih nevronske mreže. Diplomsko naloga. Ljubljana, Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo: 77 str.
- [11] Esteban, L.G., Fernández, F. G., de Palacios, P. 2009. MOE Prediction in *Abies pinsapo* Boiss. timber: Application of an artificial neural network using non-destructive testing. Computers and structures 87: 1360–1365.

- [12] Jordan, R., Feeney, F., Nesbitt N., Evertsen, J.A. 1998. Classification of wood species by neural network analysis of ultrasonic signals. *Ultrasonics* 36: 219–222.

- [13] IBM. 2012. SPSS software for predictive analytics: aplikacija IBM SPSS 20. <http://www-01.ibm.com/software/analytics/spss/> (Pridobljeno 4. 4. 2012.)

- [14] Frank, E., Holmes, G., Mayo, M., Pfahringer, B., Smith, T., Witten, I. 2012. Data Mining with Open Source Machine Learning Software: aplikacija WEKA 3. <http://www.cs.waikato.ac.nz/ml/weka/> (Pridobljeno 4. 4. 2012.)