

**KAUNO TECHNOLOGIJOS UNIVERSITETAS**  
**INFORMATIKOS FAKULTETAS**  
**SISTEMINĖS ANALIZĖS KATEDRA**

**VLADAS CIBULSKIS**

**DIRBTINIŲ NEURONINIŲ TINKLŲ KOLEKTYVŲ FORMAVIMO  
ALGORITMŲ KŪRIMAS**

**Darbo vadovai**  
**prof. A. Verikas**  
**dr. A. Gelžinis**  
**2005 05**

**KAUNAS, 2005**

**KAUNO TECHNOLOGIJOS UNIVERSITETAS  
INFORMATIKOS FAKULTETAS  
SISTEMINĖS ANALIZĖS KATEDRA**

**TVIRTINU  
Katedros vedėjas  
prof. habil. dr. R. Barauskas  
2005 05**

**Dirbtinių neuroninių tinklų kolektyvų formavimo  
algoritmų kūrimas**

**Informatikos mokslo magistro baigiamasis darbas**

**Kalbos konsultante  
Lietuvių k. katedros lekt.  
dr. J. Mikelionienė  
2005 05**

**Darbo vadovai  
prof. A. Verikas  
dr. A. Gelžinis  
2005 05**

**Recenzentas  
dr. A. Lipnickas  
2005 05**

**Atliko  
IFM-9/1 gr. stud.  
V. Cibulskis  
2005 05 23**

**KAUNAS, 2005**

# TURINYS

<b>LENTELIŲ SĄRAŠAS</b> .....	<b>4</b>
<b>1 PRIEDAS</b> .....	<b>4</b>
<b>PAVEIKSLŲ SĄRAŠAS</b> .....	<b>5</b>
<b>PAVEIKSLŲ SĄRAŠAS</b> .....	<b>5</b>
<b>SANTRAUKA</b> .....	<b>6</b>
<b>SUMMARY</b> .....	<b>6</b>
<b>NOMENKLATŪRA</b> .....	<b>7</b>
<b>ĮVADAS</b> .....	<b>8</b>
<b>1. BENDROJI DALIS</b> .....	<b>9</b>
1.1. DIRBTINIAI NEURONINIAI TINKLAI.....	9
1.1.1. <i>Dirbtinių neuroninių tinklų kolektyvai</i> .....	10
1.1.1.1. Vidurkio kolektyvas.....	10
1.2. SKIRTINGŲ TINKLŲ KŪRIMAS.....	12
1.2.1. <i>Požymių atrinkimas</i> .....	12
1.2.1.1. Požymių atmetimo klaidos funkcija.....	13
1.2.1.2. Požymių atmetimas.....	14
1.2.1.3. Sviurių reguliavimo narys.....	15
1.2.1.4. Požymių atmetimo klaidos funkcijos modifikacijos.....	16
1.2.2. <i>Neigiamos koreliacijos mokymas</i> .....	16
1.2.3. <i>Mokymo algoritmai</i> .....	17
1.2.3.1. Vieno tinklo mokymas.....	17
1.2.3.2. Tinklų kolektyvo mokymas.....	18
1.3. KOLEKTYVO NARIŲ SKIRTINGUMO ĮVERTINIMAS.....	19
1.3.1. <i><math>\kappa</math>-statistika</i> .....	20
1.3.2. <i>Q statistika</i> .....	20
1.3.3. <i>Apibendrinimo klaida pagrįsta statistika</i> .....	20
<b>2. EKSPERIMENTINĖ DALIS</b> .....	<b>22</b>
2.1. NAUDOTI DUOMENYS.....	22
2.2. MOKYMO PARAMETRAI.....	22
2.3. EKSPERIMENTO REZULTATAI.....	23
2.3.1. <i>Požymių atrinkimas vienam tinklui</i> .....	23
2.3.1.1. Požymiai neatrenkami.....	23
2.3.1.2. Požymiai atrenkami.....	24
2.3.2. <i>Požymių atrinkimas tinklų kolektyvui</i> .....	28
2.3.3. <i>Tinklų kolektyvo klasifikavimo klaida</i> .....	29
2.3.3.1. Kolektyvas be papildomų narių.....	29
2.3.3.2. Kolektyvas su papildomais nariais.....	29
2.3.4. <i>Skirtingumo palyginimas</i> .....	30
<b>3. IŠVADOS</b> .....	<b>32</b>
<b>4. LITERATŪRA</b> .....	<b>33</b>

## Lentelių sąrašas

LENTELĖ 1. VIENO TINKLO VIDUTINĖ KLASIFIKAVIMO KLAIDA TESTINEI IMČIAI .....	23
LENTELĖ 2. VIENO TINKLO VIDUTINĖ KLASIFIKAVIMO KLAIDA TESTINEI IMČIAI, MOKYMU NAUDOJANT ĮVAIRIUS POŽYMIŲ ATRINKIMO NARIUS.....	24
LENTELĖ 3. POŽYMIŲ ATRINKIMO EIGA <i>DIABETO DIAGNOZAVIMO PROBLEMAI</i> .....	24
LENTELĖ 4. VIDUTINIS ATRINKTŲ POŽYMIŲ SKAIČIUS VIENAM TINKLUI, MOKYMU NAUDOJANT ĮVAIRIUS POŽYMIŲ ATRINKIMO NARIUS.....	25
LENTELĖ 5. REIKŠMINGAIS PRIPAŽINTŲ POŽYMIŲ PASISKIRSTYMAS TARP KOLEKTYVO NARIŲ <i>DIABETO DIAGNOZAVIMO PROBLEMAI</i> SU <i>SR</i> NARIU.....	28
LENTELĖ 6. REIKŠMINGAIS PRIPAŽINTŲ POŽYMIŲ PASISKIRSTYMAS TARP KOLEKTYVO NARIŲ <i>DIABETO DIAGNOZAVIMO PROBLEMAI</i> SU <i>PA+SR</i> NARIU .....	28
LENTELĖ 7. REIKŠMINGAIS PRIPAŽINTŲ POŽYMIŲ PASISKIRSTYMAS TARP KOLEKTYVO NARIŲ <i>DIABETO DIAGNOZAVIMO PROBLEMAI</i> SU <i>PAI+SR</i> NARIU .....	28
LENTELĖ 8. TINKLŲ KOLEKTYVO IR VIENO TINKLO VIDUTINĖ KLASIFIKAVIMO KLAIDA TESTINEI IMČIAI .....	29
LENTELĖ 9. TINKLŲ KOLEKTYVO VIDUTINĖ KLASIFIKAVIMO KLAIDA TESTINEI IMČIAI, MOKYMU NAUDOJANT ĮVAIRIUS PAPILDOMUS NARIUS .....	30

## 1 Priedas

LENTELĖ 1. VIENO TINKLO MOKYMAS, 100 EPOCHŲ .....	34
LENTELĖ 2. VIENO TINKLO MOKYMAS, 50 EPOCHŲ .....	34
LENTELĖ 3. VIENO TINKLO MOKYMAS 100 EPOCHŲ SU <i>SR</i> NARIU.....	34
LENTELĖ 4. VIENO TINKLO MOKYMAS SU <i>PA+SR</i> NARIU ATMETANT POŽYMIUS .....	34
LENTELĖ 5. VIENO TINKLO MOKYMAS SU <i>PAI+SR</i> NARIU ATMETANT POŽYMIUS .....	34
LENTELĖ 6. VIENO TINKLO MOKYMAS SU <i>SR</i> NARIU ATMETANT POŽYMIUS.....	35
LENTELĖ 7. TINKLŲ KOLEKTYVO NKM MOKYMAS .....	35
LENTELĖ 8. TINKLŲ KOLEKTYVO NKM MOKYMAS SU <i>SR</i> PAPILDOMU NARIU NEATMETANT NEREIŠMINGŲ ĮĖJIMŲ .....	35
LENTELĖ 9. TINKLŲ KOLEKTYVO NKM MOKYMAS SU <i>SR</i> PAPILDOMU NARIU ATMETANT NEREIŠMINGUS ĮĖJIMUS .....	35
LENTELĖ 10. TINKLŲ KOLEKTYVO NKM MOKYMAS SU <i>SR</i> PAPILDOMU NARIU ATMETANT NEREIŠMINGUS ĮĖJIMUS, BEI PRADINIŲ APMOKYMU .....	35
LENTELĖ 11. TINKLŲ KOLEKTYVO NKM MOKYMAS SU <i>PA+SR</i> PAPILDOMU NARIU NEATMETANT NEREIŠMINGŲ ĮĖJIMŲ .....	36
LENTELĖ 12. TINKLŲ KOLEKTYVO NKM MOKYMAS SU <i>PA+SR</i> PAPILDOMU NARIU ATMETANT NEREIŠMINGUS ĮĖJIMUS .....	36
LENTELĖ 13. TINKLŲ KOLEKTYVO NKM MOKYMAS SU <i>PA+SR</i> PAPILDOMU NARIU ATMETANT NEREIŠMINGUS ĮĖJIMUS, BEI PRADINIŲ APMOKYMU .....	36
LENTELĖ 14. TINKLŲ KOLEKTYVO NKM MOKYMAS SU <i>PAI+SR</i> PAPILDOMU NARIU NEATMETANT NEREIŠMINGŲ ĮĖJIMŲ .....	36
LENTELĖ 15. TINKLŲ KOLEKTYVO NKM MOKYMAS SU <i>PAI+SR</i> PAPILDOMU NARIU ATMETANT NEREIŠMINGUS ĮĖJIMUS .....	36
LENTELĖ 16. TINKLŲ KOLEKTYVO NKM MOKYMAS SU <i>PAI+SR</i> PAPILDOMU NARIU ATMETANT NEREIŠMINGUS ĮĖJIMUS, BEI PRADINIŲ APMOKYMU .....	37

## Paveikslų sąrašas

1 PAV. BIOLOGINIO NEURONO IR DIRBTINIO NEURONO PALYGINIMAS .....	9
2 PAV. BIOLOGINIŲ NEURONŲ SINAPSĖ .....	9
3 PAV. ADAPTYVIOSIOS SISTEMOS KŪRIMO PROCESO SCHEMA .....	10
4 PAV. STANDARTINIS TRIJŲ SLUOKSNIŲ NEURONINIS TINKLAS .....	13
5 PAV. FUNKCIJOS $f(w)$ GRAFIKAS, KAI $\varepsilon_1 = 10^{-1}$ , $\varepsilon_2 = 10^{-4}$ IR $\beta = 10$ .....	14
6 PAV. POŽYMIŲ NAUDOJIMAS <i>DIABETO DIAGNOZAVIMO PROBLEMAI</i> SPREŠTI SU NARIU: A) <i>PAI + SRI</i> , B) <i>PA + SRI</i> , C) <i>SR</i> .....	25
7 PAV. POŽYMIŲ NAUDOJIMAS <i>VISKONSINO DIAGNOSTINEI KRŪTŲ VĖŽIO PROBLEMAI</i> SPREŠTI SU NARIU A) <i>PAI + SRI</i> , B) <i>PA + SRI</i> , C) <i>SR</i> .....	26
8 PAV. POŽYMIŲ NAUDOJIMAS <i>JAV KONGRESO BALSAVIMO ĮRAŠŲ PROBLEMAI</i> SPREŠTI SU NARIU: A) <i>PAI + SRI</i> , B) <i>PA + SRI</i> , C) <i>SR</i> .....	27
9 PAV. $\rho$ KRITERIAUS PALYGINIMAS SKIRTINGIEMS ALGORITMAMS SPRENDŽIANT <i>DIABETO DIAGNOZAVIMO PROBLEMA</i> .....	31
10 PAV. $\kappa$ STATISTIKOS PALYGINIMAS SKIRTINGIEMS ALGORITMAMS SPRENDŽIANT <i>DIABETO DIAGNOZAVIMO PROBLEMA</i> .....	31
11 PAV. $Q$ STATISTIKOS PALYGINIMAS SKIRTINGIEMS ALGORITMAMS SPRENDŽIANT <i>DIABETO DIAGNOZAVIMO PROBLEMA</i> .....	31

## Santrauka

Ankstesni tyrimai klasifikuojančių neuroninių tinklų kolektyvų srityje parodė, kad efektyvūs kolektyvai turėtų būti sudaryti iš narių, kurie yra ne tik tikslūs, bet ir skirtingi. Šiame darbe analizuojama narių skirtingumo ir tikslumo sąveika sujungiant neuroninių tinklų mokymo, sujungimo į kolektyvą ir nereikšmingų įėjimo kintamųjų eliminavimo žingsnius. Tai pasiekti padeda prie neigiamos koreliacijos mokymo klaidos funkcijos pridedamas papildomas narys, mažinantis neuroninio tinklo įėjimo sluoksnio svorius.

## Summary

### **Algorithms development for creation of artificial neural network committees**

Previous works on classification committees have shown that an efficient committee should consist of networks that are not only very accurate, but also diverse. In this work, aiming to explore trade-off between the diversity and accuracy of committee networks, the steps of neural network training, aggregation of the networks into a committee, and elimination of irrelevant input variables are integrated. To accomplish the elimination, an additional term to the *Negative correlation learning* error function, which forces input weights connected to the irrelevant input variables to decay, is added.

## Nomenklatūra

VKP – vidutinė kvadratinė paklaida.

SKP – suminė kvadratinė paklaida.

NKM – neigiamos koreliacijos mokymas.

PrA – pradinis apmokymas.

PA – požymių atrinkimas.

$I$  – įėjimo neuronų kiekis.

$H$  – paslėptų neuronų kiekis.

$O$  – išėjimo neuronų kiekis.

$B$  – poslinkio (bias) neuronų kiekis.

$N$  – duomenų taškų kiekis.

$K$  – komitetą sudarančių, neuroninių tinklų kiekis.

$C$  – klasių skaičius.

$y(n)$  – laukiamas išėjimas  $n$ -tajam duomenų taškui.

$\hat{y}(n)$  – neuroninio tinklo formuojamas išėjimas  $n$ -tajam duomenų taškui.

$\hat{y}_k(n)$  –  $k$ -tojo kolektyvo nario formuojamas išėjimas  $n$ -tajam duomenų taškui.

$\hat{y}_{kol}(n)$  – kolektyvo formuojamas išėjimas  $n$ -tajam duomenų taškui.

$VKP_k$  –  $k$ -tojo kolektyvo nario vidutinė kvadratinė paklaida.

$\langle VKP \rangle$  – atsitiktinio kolektyvo nario vidutinė kvadratinė paklaida.

$VKP_{kol}$  – kolektyvo vidutinė kvadratinė paklaida.

$w_i^j$  – svoris jungiantis  $i$ -tąjį neuroną su  $j$ -tuoju neuronu.

$e(n)$  – tinklo daroma klaida  $n$ -tajam duomenų taškui.

$E$  – tinklo daroma klaida.

$E_k$  –  $k$ -tojo kolektyvo nario daroma klaida.

$IS(i)$  –  $i$ -tojo įėjimo svarba.

$PA(w)$  – požymių atrinkimo narys.

$PAI(w)$  – modifikuotas požymių atrinkimo narys.

$SR(w)$  – svorių ribojimo narys, taikomas visiems neuroninio tinklo svoriams.

$SRI(w)$  – svorių ribojimo narys, taikomas svoriams tarp paslėpto sluoksnio ir išėjimo sluoksnio, bei poslinkio (bias) svoriams.

## Išvadas

Dirbtiniai neuroniniai tinklai – tai informacijos apdorojimo struktūros, netiksliai imituojančios kai kuriuos gyvųjų organizmų smegenyse vykstančius informacijos apdorojimo procesus. Bet kurios paskirties neuroninis tinklas turi įėjimus, perduodančius kintamųjų vertes iš išorės, ir išėjimus, formuojančius tinklo atsaką. Neuroniniame tinkle neuronus įprasta grupuoti į atskirus sluoksnius. Išėjimo sluoksnio neuronų išėjimo vertės laikomos neuroninio tinklo išėjimo vertėmis. Neuroninis tinklas apmokomas taip, kad minimizuotų skirtumą tarp neuroninio tinklo išėjimo vertės  $\hat{y}(n)$  ir laukiamos vertės  $y(n)$ , kai  $n=1\dots N$ ,  $N$  – duomenų taškų kiekis.

Siekiant pagerinti įvertinimo rezultatus, keleto neuroninių tinklų išėjimai gali būti apjungiami į kolektyvo išėjimą. Daug ankstesnių darbų parodė, kad klasifikavimo užduotį sprendžiantys kolektyvai turi būti ne tik labai tikslūs, bet ir skirtingi ta prasme, kad kolektyvo narių klaidos turi pasireikšti skirtingose įėjimo duomenų erdvės srityse [1].

Manipuliavimas mokymo imtimi [2], skirtingų įėjimo kintamųjų naudojimas ir skirtingos tinklo architektūros parinkimas yra trys populiariausi būdai kurti skirtingus neuroninius tinklus. Liu ir Yao pasiūlė neigiamos koreliacijos mokymo (NKM) algoritmą [3]. Skirtingai nei kiti algoritmai, NKM moko individualius kolektyvo tinklus lygiagrečiai, naudodamas tą pačią mokymo imtį ir tuos pačius įėjimo kintamuosius, visiems tinklams. NKM mokymo metu prie klaidos funkcijos pridedamas papildomas narys, baudžiantis už tinklų koreliaciją. Tuo būdu NKM stengiasi mokyti ir apjungti tinklus į kolektyvą tame pačiame mokymo procese.

Šiame darbe žengiamas dar vienas žingsnis analizuojant kolektyvo narių skirtingumo ir tikslumo sąveiką. Šiam tikslui apjungiami neuroninio tinklo mokymo, tinklų apjungimo į kolektyvą ir nereikšmingų įėjimo kintamųjų atmetimo žingsniai. Norint tai pasiekti, pridedamas dar vienas papildomas narys prie NKM klaidos funkcijos, kuris verčia mažėti įėjimo sluoksnio svorius, besijungiančius su nereikšmingais įėjimo kintamaisiais. Šis spėjimas pagrįstas tuo, kad mažinant svorius, skirtingi įėjimo kintamieji, skirtinguose kolektyvo tinkluose gali pasirodyti naudingi. Jei tai tiesa, tinklų skirtingumas, o tuo pačiu ir kolektyvo klasifikavimo tikslumas, gali būti padidintas.

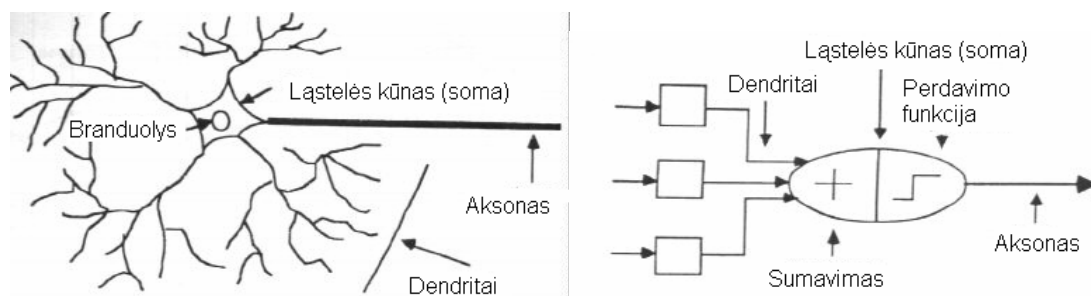
Klasifikatorių skirtingumui patikrinti skirtingi tyrinėtojai naudoja skirtingus metodus. Išsamią skirtingumo matų analizę galima rasti [4]. Dauguma matų yra pagrįsti empirine klasifikavimo klaida.  $\kappa$ -statistika [5] ir Yule's Q [4] statistika, yra du populiariausi matai. Šiame darbe be minėtų matų yra pasiūlytas matas, naudojantis svertine įėjimų įtaka (leverages) pagrįstą apibendrinimo klaidos įvertį.



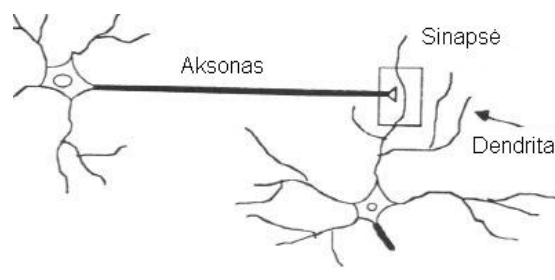
# 1. Bendroji dalis

## 1.1. Dirbtiniai neuroniniai tinklai

Dirbtiniai neuroniniai tinklai – tai informacijos apdorojimo struktūros, netiksliai imituojančios kai kuriuos gyvųjų organizmų smegenyse vykstančius informacijos apdorojimo procesus [6]. Dirbtiniai neuroniniai tinklai sudaromi iš daugelio tarpusavyje sujungtų labai paprastų skaičiavimo elementų. Šie elementai, jungiami vieni su kitais įvairaus stiprumo jungtimis, yra apytikris biologinių neuronų modelis (1 pav.). Skaičiavimo elementus siejančios jungtys yra analogiškos biologinių neuronų sinapsėms (2 pav.). Dirbtiniu neuroniniu tinklu siekiama emuliuoti kai kurias biologinių sistemų savybes. Labiausiai viliojantis atrodo biologinių sistemų gebėjimas mokytis, prisitaikyti ir adaptuotis.

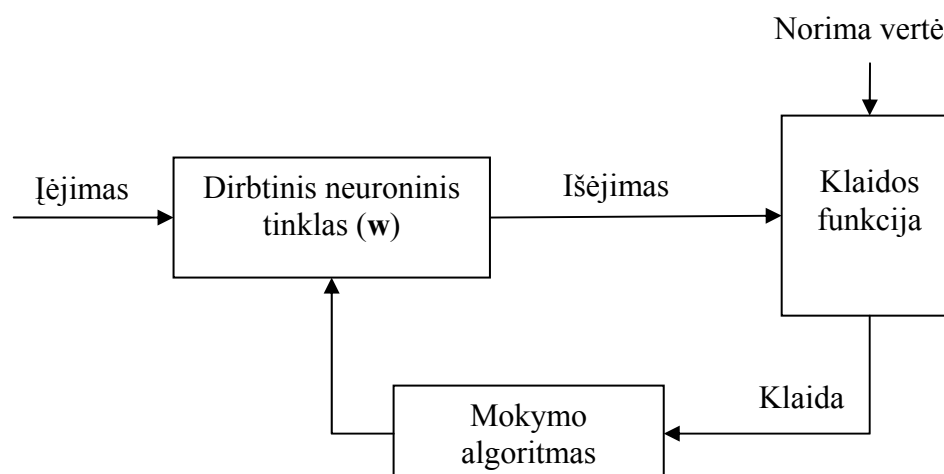


1 pav. Biologinio neuronų ir dirbtinio neuronų palyginimas



2 pav. Biologinių neuronų sinapsė

Mokymo metu smegenyse keičiasi jungčių, siejančių neuronus, stiprumas. Toks jungčių stiprumo kitimas būdingas ir dirbtiniams neuroniniams tinklams. Dirbtiniams neuroniniams tinklams mokytis naudojami duomenų pavyzdžiai. Mokymo metu, duomenims veikiant dirbtinį neuroninį tinklą specialiais algoritmais iteratyvai keičiami jungčių stiprumo koeficientai, vadinami svoriais (3 pav., svoriai pažymėti  $w$ ). Informacija, reikalinga konkrečiam uždaviniui spręsti, sukaupiama svorių vertėse.



3 pav. Adaptyviosios sistemos kūrimo proceso schema

Dar viena labai viliojanti dirbtinių neuroninių tinklų savybė yra ta, kad jie nesudėtingai realizuojami lygiagrečiais skaičiavimais.

### 1.1.1. Dirbtinių neuroninių tinklų kolektyvai

Pagal tuos pačius duomenis apmokant keletą neuroninių tinklų, galima gauti keletą panašios kokybės neuroninių tinklų, darančių beveik to paties dydžio klaidą. Skirtingi neuroniniai tinklai gali būti gaunami, net ir tam pačiam neuroniniam tinklui parinkus skirtingą mokymo algoritmą ar kitas pradines mokymo sąlygas. Be to, atskiri neuroniniai tinklai gali naudoti tų pačių mokymo duomenų kitą įėjimo koordinacių rinkinį. Turint atskirų neuroninių tinklų sprendimus, galima formuluoti bendrą kolektyvinį sprendimą, tikintis, kad kolektyvinis sprendimas bus geresnis nei atskiro kolektyvo nario. Modelis, susidedantis iš keleto atskirų neuroninių tinklų, vadinamas neuroninių tinklų kolektyvu.

#### 1.1.1.1. Vidurkio kolektyvas

Tarkime, kad yra  $K$  modelių, kurie yra vienodi, t. y. tarp jų nėra statistinio skirtumo, kurį būtų galima aptikti. Kiekvienas modelis skaičiuoja išėjimą  $\hat{y}_k$ . Vietoj to, kad pasirinkti viena iš šių modelių atsitiktinai, galima paskaičiuoti jų išėjimų vidurkius:

$$\hat{y}_{kol}(n) = \frac{1}{K} \sum_{k=1}^K \hat{y}_k(n). \quad (1)$$

Šio vidurkio kolektyvo suminė kvadratinė paklaida (SKP) (arba vidutinė kvadratinė paklaida (VKP)) testinei duomenų imčiai visada bus geresnė arba lygi klaidai, kurią darys atsitiktinai pasirinktas modelis [7].

*Įrodymas.* VKP  $k$ -tajam modeliui yra

$$VKP_k = \frac{1}{N} \sum_{n=1}^N (y(n) - \hat{y}_k(n))^2, \quad (2)$$

čia  $N$  – testinės duomenų imties dydis,  $y(n)$  – laukiamas išėjimas  $n$ -tajam duomenų taškui. Jei pasirinktas atsitiktinis modelis  $k$ , tai VKP lygi

$$\langle VKP \rangle = \frac{1}{K} \sum_{k=1}^K VKP_k, \quad (3)$$

o kolektyvo VKP lygi

$$VKP_{kol} = \frac{1}{N} \sum_{n=1}^N (y(n) - \hat{y}_{kol}(n))^2. \quad (4)$$

Pasinaudojame Cauchy-Schwartz nelygybe vektoriams  $\vec{a}$  ir  $\vec{b}$

$$(\vec{a}^T \vec{b})^2 \leq \|\vec{a}\|^2 \|\vec{b}\|^2. \quad (5)$$

Jei naudojami  $K$ -mačiai vektoriai

$$\vec{a} = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix} \text{ ir } \vec{b} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \dots \\ \hat{y}_K \end{pmatrix}, \quad (6)$$

tuomet pagal Cauchy-Schwartz nelygybę

$$\left( \sum_{k=1}^K \hat{y}_k \right)^2 = (\vec{a}^T \vec{b})^2 \leq \|\vec{a}\|^2 \|\vec{b}\|^2 = K \sum_{k=1}^K \hat{y}_k^2. \quad (7)$$

Išskleidus (3) ir (4) lygybes gaunama

$$\langle VKP \rangle = \frac{1}{N} \sum_{n=1}^N y^2(n) - \frac{2}{NK} \sum_{n=1}^N \sum_{k=1}^K y(n) \hat{y}_k(n) + \frac{1}{NK} \sum_{n=1}^N \sum_{k=1}^K \hat{y}_k^2(n), \quad (8)$$

$$VKP_{kol} = \frac{1}{N} \sum_{n=1}^N y^2(n) - \frac{2}{NK} \sum_{n=1}^N \sum_{k=1}^K y(n) \hat{y}_k(n) + \frac{1}{NK^2} \sum_{n=1}^N \left( \sum_{k=1}^K \hat{y}_k(n) \right)^2. \quad (9)$$

Jas suprastinus gaunama (7) nelygybė, tai reiškia, kad

$$VKP_{kol} \leq \langle VKP \rangle. \quad (10)$$

Šiame darbe neuroninių tinklų kolektyvai kaip tik ir bus apjungiami pagal (1) formulę, t. y. bus formuojamas vidurkio kolektyvas.

## 1.2. Skirtingų tinklų kūrimas

Krogh ir Vedelsby [8] parodė, kad kolektyvo apibendrinimo klaida  $E$ , mokymo metu nematytiems duomenims, yra apskaičiuojama taip:

$$E = \bar{E} - \bar{A} \quad (11)$$

kur  $\bar{E}$  yra kolektyvą sudarančių tinklų apibendrinimo klaidų svorinis vidurkis ir  $\bar{A}$  yra kolektyvo narių skirtingumas.

Atskiram neuroniniam tinklui bazinė klaida matuoja dydį, kuriuo neuroninio tinklo formuojamas išėjimo  $\hat{y}(n)$  (visiems duomenų taškams) skiriasi nuo laukiamo išėjimo  $y(n)$ . Sklaidos klaidos dydis rodo, kaip neuroninio tinklo funkcija jautri duomenų imčiai ir jos pasirinkimui.

(11) lygybė parodo santykį tarp bazinės klaidos ir sklaidos klaidos kolektyve. Jei kolektyvas yra smarkiai šališkas, tai skirtingumas yra mažas, todėl, kad tinklai realizuoja labai panašias funkcijas ir sutaria dėl įėjimų priklausomybės tai pačiai klasei, net ir tiems duomenų taškams, kurie nebuvo naudoti mokymui. Taigi apibendrinimo klaida iš esmės bus lygi atskirų tinklų apibendrinimo klaidų svoriniam vidurkiui. Tačiau, jei sklaida yra didelė, skirtingumas irgi yra didelis, tuomet apibendrinimo klaida bus mažesnė nei atskirų tinklų vidutinė apibendrinimo klaida.

Manipuliacijos mokymo imtimi [2], skirtingų įėjimo kintamųjų naudojimas ir skirtingos tinklo architektūros parinkimas yra trys populiariausi būdai kurti skirtingus neuroninius tinklus. Plėtros (bootstrapping) [9], Ada skatinimo (AdaBoosting) [10], balsų siuntimo (Pasting Votes) [11] algoritmai, yra žinomiausi, tarp algoritmų skirtų duomenų imčių sudarymui, siekiant sukurti skirtingus kolektyvo narius. Visi šie algoritmai kolektyvo narius moko atskirai. Skirtingai nei kiti algoritmai, NKM [3] moko individualius kolektyvo tinklus lygiagrečiai, naudodamas tą pačią mokymo imtį ir tuos pačius įėjimo kintamuosius, visiems tinklams. Šiame darbe tiriamos NKM išplėtimo galimybės šį algoritmą apjungiant su skirtingus įėjimo kintamuosius naudojančiais algoritmais.

### 1.2.1. Požymių atrinkimas

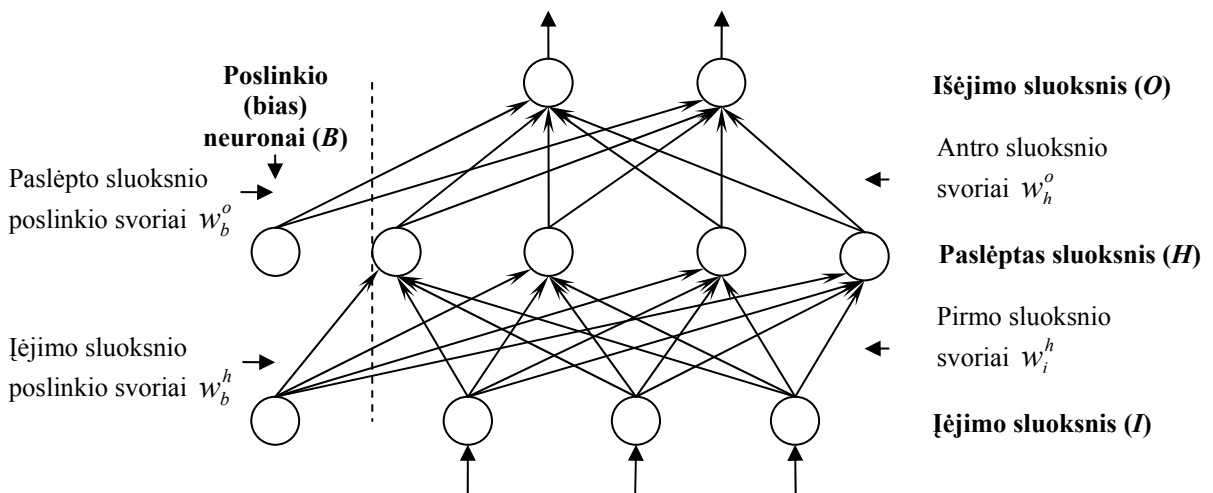
Klasifikavimo problema gali būti suskirstyta į du etapus: požymių išskyrimo ir klasifikavimo [12]. Požymių išskyrimas – tai procesas, kurio metu ieškomas atvaizdavimas padedantis sumažinti duomenų matų skaičių. Atskiras požymių išskyrimo atvejis yra požymių atrinkimas. Jei duomenis apibūdina  $I$  atributų aibė, naudojantis požymių atrinkimu, yra bandoma išskirti tokį šių atributų poaibį, kuris yra svarbus klasifikavimui. Požymių atrinkimo svarba yra labai gerai žinoma. Iš klasifikavimo proceso pašalinus perteklinius/nereikšmingus

atributus galima gauti klasifikatorių su geresnėmis apibendrinimo savybėmis, pvz. geriau nuspėjanti naujų/nematytų duomenų taškų klases. Duomenims, su stipriai koreliuotais atributais, matų skaičius gali būti sumažintas prarandant mažai ar visiškai neprarandant informacijos [13].

Yra daug būdų požymiams atrinkti. Du iš populiariausių, tai svarbių atributų įtraukimas ir nereikšmingų atributų atmetimas. ID3 [15] ir CART [16] yra du skirtingi pirmojo būdo algoritmai. Šiame darbe pasirinktas nereikšmingų atributų atmetimo algoritmas aprašytas [13] straipsnyje.

### 1.2.1.1. Požymių atmetimo klaidos funkcija

Požymių atrinkimas taikomas standartiniam, pilnai sujungtam trijų sluoksnių neuroniniam tinklui (4 pav.).



4 pav. Standartinis trijų sluoksnių neuroninis tinklas

Standartinė klaidos funkcija, kuri yra minimizuojama apmokymo metu skaičiuoja vidutinę kvadratinę, tinklo formuojamo atsakymo, paklaidą

$$VKP = \frac{1}{N} \sum_{n=1}^N (y(n) - \hat{y}(n))^2 = \frac{1}{N} \sum_{n=1}^N (e(n))^2, \quad (12)$$

čia  $N$  – duomenų taškų skaičius,  $y(n)$  – laukiamas išėjimas  $n$ -tajam duomenų taškui,  $\hat{y}(n)$  – tinklo formuojamas išėjimas  $n$ -tajam duomenų taškui,  $e(n)$  – tinklo daroma klaida  $n$ -tajam duomenų taškui.

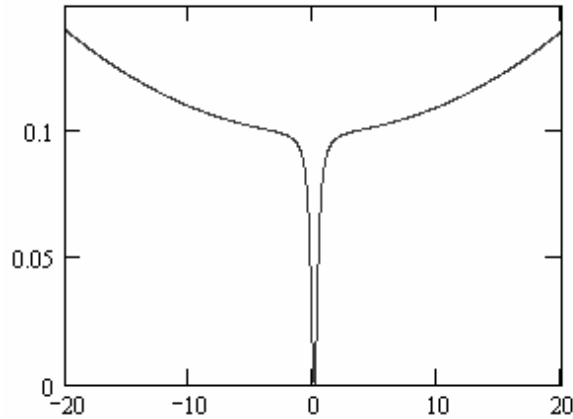
Svarbūs ir nesvarbūs įėjimai atskiriami pagal ryšių, tarp įėjimo sluoksniu ir paslėpto sluoksniu neuronų, stiprumą. Tinklas mokomas taip, kad svoriai, jungiantys nereikšmingus įėjimus ir paslėpto sluoksniu neuronus, būtų nedideli. Šie svoriai gali būti pašalinti iš tinklo smarkiai neįtakojant jo tikslumo. Kadangi bandoma surasti mažiausią įėjimų poaibį, kuris vis dar reprezentuotų duomenų charakteristikas, svarbu, kad tinklas būtų apmokytas taip, kad tik tų svorių, kurie jungia svarbius įėjimo ir paslėptojo sluoksniu neuronus, reikšmės būtų

didelės. Norint tai pasiekti, prie klaidos funkcijos pridedamas papildomas narys, skaičiuojamas svoriams, jungiantiems įėjimo sluoksnio neuronus su paslėptojo sluoksnio neuronais.

Šis papildomas narys skaičiuojamas pagal formulę [13]:

$$f(w) = \varepsilon_1 \frac{\beta w^2}{(1 + \beta w^2)} + \varepsilon_2 w^2, \quad (13)$$

Funkcijos  $f(w)$  grafikas, kai  $\varepsilon_1 = 10^{-1}$ ,  $\varepsilon_2 = 10^{-4}$  ir  $\beta = 10$  yra parabolė su šuolinių sumažėjimu apatinėje jos dalyje (5 pav.). Maži svoriai yra skatinami mažėti iki nulio, tai atspindi šuolinis funkcijos reikšmės sumažėjimas prie nulio. Kita vertus svoriai negali labai didėti, nes tuo atveju kvadratinė funkcijos komponentė ( $w^2$ ) tampa dominuojančia.



5 pav. Funkcijos  $f(w)$  grafikas, kai  $\varepsilon_1 = 10^{-1}$ ,  $\varepsilon_2 = 10^{-4}$  ir  $\beta = 10$

(13) funkciją pridėjus prie VKP klaidos funkcijos gaunasi:

$$E(w, e) = VKP(e) + PA(w), \quad (14)$$

kur  $E(w, e)$  – minimizuojama klaidos funkcija, o  $PA(w)$  lygi

$$PA(w) = \varepsilon_1 \left( \sum_{i=1}^I \sum_{h=1}^H \left( \frac{\beta (w_i^h)^2}{(1 + \beta (w_i^h)^2)} \right) \right) + \varepsilon_2 \left( \sum_{i=1}^I \sum_{h=1}^H (w_i^h)^2 \right), \quad (15)$$

čia  $w_i^h$  – neuroninio tinklo svoris jungiantis i-tąjį įėjimo sluoksnio neuroną su h-tuoju paslėpto sluoksnio neuronu. Vėliau bus aptarti šio papildomo nario trūkumai ir pateikta jo modifikacija.

### 1.2.1.2. Požymių atmetimas

Svarbių tinklo įėjimų atrinkimui, neuroninis tinklas mokomas minimizuojant (14) formulėje pateiktą klaidos funkciją. Naudojant klaidos funkciją su papildomu nariu tikimasi, kad po apmokymo didesni bus tie tinklo svoriai, kurie jungiasi su klasifikavimui svarbiais

požymiais. Sujungimai su mažais svoriais gali būti išmesti iš tinklo, beveik nepaveikiant tinko tikslumo, t. y., arba tikslumas nepakinta, arba jei tikslumas sumažėja, jį vėl galima pasiekti papildomai apmokant tinklą. Norint nustatyti, kuris įėjimo požymis yra nesvarbus naudojami labai paprastą kriterijų. Šis kriterijus paremtas triukšmingo įėjimo kintamojo pridėjimu prie mokymui naudojamų požymių. Kadangi visų požymių reikšmės prieš apmokymą yra normalizuojamos ir tikimasi, kad po normalizavimo įgyja artima Gausiniam pasiskirstymą, tai ir triukšmingas įėjimas generuojamas pagal Gausinį pasiskirstymo dėsnį. Apmokius neuroninį tinklą kiekvieno įėjimo, tarp jų ir triukšmingo, svarba yra skaičiuojama pagal tokią formulę:

$$IS(i) = \sum_{h=1}^H |w_i^h|, \quad (16)$$

čia  $w_i^h$  – neuroninio tinklo svoris jungiantis  $i$ -tajį įėjimo sluoksnio neuroną su  $h$ -tuoju paslėpto sluoksnio neuronu.

Visi įėjimai, kurių svarba  $IS$  yra mažesnė už triukšmingo įėjimo svarbą, yra atmetami kaip nesvarbūs klasifikavimui. Atmetus nesvarbius požymius neuroninis tinklas yra vėl mokomas. Po apmokymo vėl yra ieškoma nesvarbių požymių ir jie atmetami. Mokymo ir atmetimo žingsniai kartojami tol, kol lieka tik svarbūs požymiai, t. y. visų likusių požymių svarba ( $IS$ ) yra didesnė arba lygi triukšmingo įėjimo svarbai. Kuomet nebelieka nesvarbių požymių, atmetamas triukšmingas požymis ir neuroninis tinklas apmokomas paskutini kartą.

### 1.2.1.3. Sviurių reguliavimo narys

Praktikoje labai dažnai naudojamas paprastas neuroninio tinklo sviurių reguliavimo narys, kuris neleidžia sviuriam labai padidėti. Tai visų sviurių kvadratų suma:

$$SR(w) = \frac{1}{W} \sum_{i=1}^W (w_i)^2. \quad (17)$$

čia  $W$  sviurių kiekis lygus

$$W = I \cdot H + H \cdot O + H + O, \quad (18)$$

Kadangi į šį narį galima žiūrėti, kaip į sviurius mažinantį narį, tai šiame darbe bus tiriamos galimybės panaudoti jį nereikšmingų požymių atmetimui. Klaidos formulė pridėjus šį papildomą narį gaunasi

$$E(w, e) = VKP(e) + \alpha \cdot SR(w), \quad (19)$$

čia  $\alpha$  ( $0 \leq \alpha \leq 1$ ) – koeficientas nusakantis  $SR(w)$  nario stiprumą.

#### 1.2.1.4. Požymių atmetimo klaidos funkcijos modifikacijos

1.2.1.1 skyrelyje aprašytas klaidos funkcijos papildomas narys kiekvieną svorį mažina atskirai, t. y. prie to paties požymio gali būti ir didelių svorių, kurie yra svarbūs klasifikavimo procesui, ir mažų, kurie yra ne tokie svarbūs. Nors pagal (16) formulę apskaičiuota įėjimo požymio svarba  $IS$  bus pakankamai maža, kad šis požymis būtų atmestas, tačiau tas vienas didelis svoris gali turėti labai didelį poveikį neuroninio tinko klasifikavimo klaidai. Todėl buvo pasiūlyta tokia (15) formulės modifikacija:

$$PAI(w) = \varepsilon_1 \left( \sum_{i=1}^I \frac{\beta \left( \sum_{h=1}^H |w_i^h| \right)^2}{1 + \beta \left( \sum_{h=1}^H |w_i^h| \right)^2} \right) + \varepsilon_2 \left( \sum_{i=1}^I \sum_{h=1}^H (w_i^h)^2 \right). \quad (20)$$

Šioje formulėje vieno įėjimo svoriai yra sudedami ir minimizuoti stengiamasi jų sumą, o ne atskirus svorius. Tokiu būdu tiesiogiai yra minimizuojamas įėjimo svarbos parametras, kurio pagalba vėliau nustatoma kuriuos svorius reikia atmesti.

Dar vienas pakeitimas, naudojamas šiame darbe, tai svorių didėjimą reguliuojantis, narys, aprašytas skyrelyje 1.2.1.3

$$SR1(w) = \frac{1}{W} \sum_{i=1}^W (w_i)^2. \quad (21)$$

tačiau čia  $W$  lygus

$$W = H \cdot O + H + O, \quad (22)$$

Šis narys yra taikomas tik tiems svoriams, kuriems netaikomas požymių atrinkimo narys, t. y. antrojo sluoksnio svoriams ir poslinkio (bias) svoriams. Bendra klaidos formulė pridėjus visus papildomus narius gaunasi

$$E(w, e) = VKP(e) + PAI(w) + \alpha \cdot SR1(w), \quad (23)$$

čia  $\alpha$  ( $0 \leq \alpha \leq 1$ ) – koeficientas nusakantis  $SR(w)$  nario stiprumą.

#### 1.2.2. Neigiamos koreliacijos mokymas

Neigiamos koreliacijos mokymo tikslas – neuroninių tinklų kolektyve sukurti skirtingus individualius tinklus. NKM yra taikomas vidurkio kolektyvui, t. y. individualių neuroninių tinklų įėjimai apjungiami pagal (1) formulę. Šis algoritmas iš kitų išsiskiria tuo, kad moko individualius kolektyvo tinklus lygiagrečiai, naudodamas tą pačią mokymo imtį ir tuos pačius įėjimo kintamuosius, visiems tinklams. Be to, mokoma tik tiek tinklų kiek bus apjungta į kolektyvą, taip sumažinant apmokymui reikalingų skaičiavimų kiekį.



Siekiant skatinti skirtingumą prie neuroninio tinklo klaidos funkcijos pridedamas koreliacijos baudos narys. Klaidos funkcija  $E_k$  kiekvienam  $k$ -tajam tinklui neigiamos koreliacijos mokyme įgauna tokią išraišką:

$$E_k = VKP_k + \lambda \sum_{n=1}^N NK_k(n), \quad (24)$$

kur  $N$  yra mokymo imties dydis,  $\lambda$  ( $0 \leq \lambda \leq 1$ ) – parametras nustatantis baudos nario stiprumą ir  $NK_k$  – neigiamos koreliacijos baudos narys apskaičiuojamas pagal formulę:

$$NK_k(n) = (\hat{y}_k(n) - \hat{y}_{kol}(n)) \sum_{j=1, j \neq k}^K (\hat{y}_j(n) - \hat{y}_{kol}(n)), \quad (25)$$

kur  $K$  yra tinklų skaičius kolektyve,  $\hat{y}_k$  –  $k$ -tojo kolektyvo nario apskaičiuotas išėjimas, o  $\hat{y}_{kol}$  – kolektyvo apskaičiuotas išėjimas. Neigiamos koreliacijos baudos nario minimizavimo tikslas yra sumažinti koreliaciją tarp apmokomo tinklo klaidos ir kitų kolektyvo narių klaidų.

### 1.2.3. Mokymo algoritmai

#### 1.2.3.1. Vieno tinklo mokymas

##### 1.2.3.1.1. Vieno tinklo mokymas su papildomais nariais atmetant nereikšmingus požymius

1. Pasirenkame mokymo parametrus:
  - a.  $\varepsilon_1$  – PA (PAI) nario pirmos dalies įtaka klaidai,
  - b.  $\varepsilon_2$  – PA (PAI) nario antros dalies įtaka klaidai,
  - c.  $\beta$  – reguliuoja nuo kokios reikšmės įsijungia pirmoji požymių atrinkimo nario dalis.
  - d.  $\alpha$  – reguliuoja SR (SRI) nario stiprumą.
2. Į įėjimo kintamųjų aibę įtraukiamas triukšmo kintamasis. Duomenų imtis atsitiktiniu būdu sudalinama į mokymo, tikrinimo ir testavimo imtis.
3. Prie klaidos funkcijos prijungiami papildomi nariai ir tinklas apmokomas.
4. Tinklui šalinami įėjimai, kurių IS reikšmė yra mažesnė už triukšmo įėjimo IS.
5. Tinklas vėl apmokomas.
6. Ketvirtas ir penktas žingsniai kartojami, kol tinklui nebegalima atmesti jokio įėjimo kintamojo.
7. Tinklui šalinamas triukšmo kintamasis.
8. Tinklas apmokomas priskyrus  $\varepsilon_1=0$ ,  $\varepsilon_2=\alpha$ .

## 1.2.3.2. Tinklų kolektyvo mokymas

### 1.2.3.2.1. Tinklų kolektyvo mokymas su papildomais nariais neatmetant nereikšmingų požymių

1. Pasirenkame mokymo parametrus:
  - a.  $\lambda$  – neigiamos koreliacijos nario įtaka klaidai,
  - b.  $\varepsilon_1 - PA (PAI)$  nario pirmos dalies įtaka klaidai,
  - c.  $\varepsilon_2 - PA (PAI)$  nario antros dalies įtaka klaidai,
  - d.  $\beta$  – reguliuoja nuo kokios reikšmės įsijungia pirmoji požymių atrinkimo nario dalis.
  - e.  $\alpha$  – reguliuoja  $SR (SRI)$  nario stiprumą.
2. Duomenų imtis atsitiktiniu būdu sudalinama į mokymo, tikrinimo ir testavimo imtis.
3. Sukuriamas kolektyvas.
4. Prie klaidos funkcijos prijungiami papildomi nariai ir kolektyvas apmokomas.

### 1.2.3.2.2. Tinklų kolektyvo mokymas su papildomais nariais atmetant nereikšmingus požymius

1. Pasirenkame mokymo parametrus:
  - a.  $\lambda$  – neigiamos koreliacijos nario įtaka klaidai,
  - b.  $\varepsilon_1 - PA (PAI)$  nario pirmos dalies įtaka klaidai,
  - c.  $\varepsilon_2 - PA (PAI)$  nario antros dalies įtaka klaidai,
  - d.  $\beta$  – reguliuoja nuo kokios reikšmės įsijungia pirmoji požymių atrinkimo nario dalis.
  - e.  $\alpha$  – reguliuoja  $SR (SRI)$  nario stiprumą.
2. Į įėjimo kintamųjų aibę įtraukiamas triukšmo kintamasis. Duomenų imtis atsitiktiniu būdu sudalinama į mokymo, tikrinimo ir testavimo imtis.
3. Sukuriamas kolektyvas.
4. Prie klaidos funkcijos prijungiami papildomi nariai ir kolektyvas apmokomas.
5. Kiekvienam kolektyvo nariui šalinami įėjimai, kurių  $IS$  reikšmė yra mažesnė už triukšmo įėjimo  $IS$ .
6. Kolektyvas vėl apmokomas.
7. Penktas ir šeštas žingsniai kartojami, kol nei vienam kolektyvo nariui nebegalima atmesti jokio įėjimo kintamojo.
8. Kiekvienam kolektyvo nariui šalinamas triukšmo kintamasis.
9. Kolektyvas apmokomas priskyrus  $\varepsilon_1=0$ ,  $\varepsilon_2=\alpha$ .

### **1.2.3.2.3. Tinklų kolektyvo mokymas su papildomais nariais atmetant nereikšmingus požymius ir taikant pradini apmokymą**

1. Pasirenkame mokymo parametrus:
  - a.  $\lambda$  – neigiamos koreliacijos nario įtaka klaidai,
  - b.  $\varepsilon_1 - PA (PAI)$  nario pirmos dalies įtaka klaidai,
  - c.  $\varepsilon_2 - PA (PAI)$  nario antros dalies įtaka klaidai,
  - d.  $\beta$  – reguliuoja nuo kokios reikšmės įsijungia pirmoji požymių atrinkimo nario dalis.
  - e.  $\alpha$  – reguliuoja  $SR (SRI)$  nario stiprumą.
  - f.  $pr_e$  – pradinio apmokymo epochų skaičius.
2. Į įėjimo kintamųjų aibę įtraukiamas triukšmo kintamasis. Duomenų imtis atsitiktiniu būdu sudalinama į mokymo, tikrinimo ir testavimo imtis.
3. Sukuriamas kolektyvas ir jo nariai  $pr_e$  epochų apmokomi individualiai, naudojant klaidos funkciją su  $SR$  papildomu nariu.
4. Prie klaidos funkcijos prijungiami papildomi nariai ir kolektyvas apmokomas.
5. Kiekvienam kolektyvo nariui šalinami įėjimai, kurių  $IS$  reikšmė yra mažesnė už triukšmo įėjimo  $IS$ .
6. Kolektyvas vėl apmokomas.
7. Penktas ir šeštas žingsniai kartojami, kol nei vienam kolektyvo nariui nebegalima atmesti jokio įėjimo kintamojo.
8. Kiekvienam kolektyvo nariui šalinamas triukšmo kintamasis.
9. Kolektyvas apmokomas priskyrus  $\varepsilon_1=0$ ,  $\varepsilon_2=\alpha$ .

### **1.3. Kolektyvo narių skirtingumo įvertinimas**

Ryšio tarp neuroninių tinklų kolektyvo narių skirtingumo ir jų tikslumo įvertinimui naudojamos tikslumo-skirtingumo (accuracy-diversity) diagramos. Diagramose grafiškai vaizduojamas kolektyvą sudarančių tinklų tikslumas ir jų skirtingumas. Tikslumas matuojamas VKP, o skirtingumas įvertinamas skaičiuojant skirtingumo matą. Skirtingumo įvertinimui tyrinėtojai naudoja įvairius matus. Yule Q [4] statistika ir  $\kappa$ -statistika [5] yra du populiariausi matai skirtingumui įvertinti. Išsamią skirtingumo matų analizę galima rasti [4]. Šiame darbe bus naudojami abu paminėti matai ir dar bus pasiūlytas apibendrinimo klaida pagrįstas matas  $\rho$ . Yule Q statistikai ir  $\kappa$ -statistikai skaičiavimai atliekami su testine duomenų imtimi, o pasiūlytam  $\rho$  matui – su mokymo duomenų imtimi.

### 1.3.1. $\kappa$ -statistika

Sutapimo laipsnio (degree-of-agreement) statistika  $\kappa$  dviem klasifikatoriams  $d$  ir  $e$  yra apskaičiuojama taip:

$$\kappa_{de} = \frac{\theta_1 - \theta_2}{1 - \theta_2}, \quad (26)$$

kur

$$\theta_1 = \sum_{i=1}^c \frac{f_{ii}}{N}, \quad \theta_2 = \sum_{i=1}^c \left( \sum_{j=1}^c \frac{f_{ij}}{N} \sum_{j=1}^c \frac{f_{ji}}{N} \right), \quad (27)$$

čia  $C$  yra klasių skaičius,  $N$  yra duomenų taškų skaičius,  $F$  yra  $C \times C$  kvadratinė matrica, kurios elementuose  $f_{ij}$  saugomas duomenų taškų, priskirtų į klasę  $i$   $d$ -tojo klasifikatoriaus ir į klasę  $j$   $e$ -tojo klasifikatoriaus, skaičius.  $\kappa = 1$ , kai abu tinklai sutaria dėl kiekvieno duomenų taško, ir  $\kappa = 0$ , kai sutarimas lygus tokiam, kurio galima būtų tikėtis atsitiktinai.

### 1.3.2. Q statistika

Yule's Q statistika dviem klasifikatoriams  $d$  ir  $e$  apibrėžiama taip:

$$Q_{de} = \frac{p_{11}p_{00} - p_{01}p_{10}}{p_{11}p_{00} + p_{01}p_{10}}, \quad (28)$$

kur  $P$  yra  $2 \times 2$  kvadratinė matrica, kurios elementuose  $p_{ij}$ , saugomas duomenų taškų, priskirtų į eilutę  $i = 1$ , jei duomenų taškas buvo klasifikuotas teisingai, ir į  $i = 0$ , jei duomenų taškas buvo klasifikuotas neteisingai  $d$ -tojo klasifikatoriaus, ir priskirtų į stulpelį  $j = 1$ , jei duomenų taškas buvo klasifikuotas teisingai, ir į  $j = 0$ , jei duomenų taškas buvo klasifikuotas neteisingai  $e$ -tojo klasifikatoriaus, skaičius.  $Q$  gali įgyti reikšmes tarp  $-1$  ir  $1$ . Statistiškai nepriklausomiems klasifikatoriams  $Q$  reikšmė yra lygi  $0$ .

### 1.3.3. Apibendrinimo klaida pagrįsta statistika

Tarkime, kad  $\mathbf{X} = [\vec{x}_1 \vec{x}_2 \dots \vec{x}_N]$  yra mokymo duomenų imtis sudaryta iš  $N$  duomenų taškų,  $\vec{w}$  yra tinklo svorių vektorius, ir  $\mathbf{Z} = [\vec{z}_1 \vec{z}_2 \dots \vec{z}_N]$  yra Jakobian'o matrica su elementais  $\vec{z}_n = \left. \frac{\partial f(\vec{x}_n, \vec{w})}{\partial \vec{w}} \right|_{\vec{w}=\vec{w}_{LS}}$ , kur  $f(\vec{x}_n, \vec{w})$  reiškia neuroninio tinklo išėjimą, o  $\vec{w}_{LS}$  reiškia parametrų vektorius, gautą mažiausio kvadratinio nuokrypio sprendimo būdu.

Ortogonalios projekcijų matricos  $\mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T$  įstrižainės elementai  $h_{nn} = \vec{z}_n^T (\mathbf{Z}^T \mathbf{Z})^{-1} \vec{z}_n$ ,  $n = 1, \dots, N$ , vadinami svertine mokymo imties įėjimų įtaka (leverages),

tenkina tokias sąlygas:  $\sum_{n=1}^N h_{nn} = q, 0 \leq h_{nn} \leq 1$  [2].  $h_{nn}$  reikšmė rodo kokią įtaką  $n$ -tasis taškas turi modeliui. Kuo  $h_{nn}$  artimesnis vienetui, tuo didesnę įtaką  $n$ -tasis taškas turi modeliui. Ši savybė naudojama pasiūlytame mate  $\rho$  kolektyvo narių skirtingumui nustatyti.

Yra žinoma, kad apytikslė *palikti-vieną* (leave-one-out) klaida  $e(n)$  yra lygi [17]:

$$e(n) = \frac{r(n)}{1 - h_{nn}}, \quad (29)$$

kur  $r(n) = y(n) - f(\bar{x}_n, \bar{w}_{LS})$  yra  $n$ -tasis likutis, o  $y(n)$  yra laukiama reikšmė. Tuomet apskaičiavus  $e_k(n)$  kiekvienam duomenų taškui  $n=1, 2, \dots, N$  ir kiekvienam tinklui  $k=1, \dots, K$ , tinklų  $d$  ir  $e$  skirtingumų palyginimui skaičiuojamas skirtingumo matas  $\rho$ , kuris grindžiamas apytiksliais *palikti-vieną* (leave-one-out) klaidos vektoriais  $\bar{e}_d$  ir  $\bar{e}_e$ :

$$\rho_{de} = 0.5 - \frac{0.5 \bar{e}_d^T \bar{e}_e}{\|\bar{e}_d\|_2 \|\bar{e}_e\|_2}, \quad (30)$$

kur  $\|\bullet\|_2$  reiškia Euklido normą.  $\rho$  gali įgyti reikšmes tarp 0 ir 1. Statistiškai nepriklausomiems klasifikatoriams  $\rho$  reikšmė yra lygi 1.

## 2. Eksperimentinė dalis

Visuose testuose eksperimentas vykdomas 30 kartų su skirtingomis pradinėmis tinklų inicializacijomis ir skirtingu duomenų imties skirstymu į mokymo, tikrinimo ir testinę imtis. Vidutinės reikšmės ir standartiniai nuokrypiai skaičiuojami iš šių 30 bandymų. Visi neuroniniai tinklai turi 12 neuronų paslėptam sluoksnyje, visi kolektyvai apmokymų pradžioje sudaryti iš 7 neuroninių tinklų. Apmokymui naudojamas kvazi-Niutono algoritmo variantas – BFGS (Broyden-Fletcher-Goldfarb-Shanno) [14].

### 2.1. Naudoti duomenys

Visi algoritmai buvo testuojami naudotojant 3 realaus pasaulio problemas. Duomenis naudotus eksperimentuose galima parsisiųsti iš interneto adresu: <http://www.ics.uci.edu/~mlearn/>.

*JAV kongreso balsavimo įrašų problema.* JAV kongreso balsavimo įrašų duomenų imtis yra sudaryta iš 435 kongreso narių balsų sprendžiant 16 svarbių įstatymų 1998-tais metais. Balsai yra suskirstyti į tris balsų tipus: (1) Taip, (2) Ne ir (3) Nežinoma. Tikslas yra nustatyti kuriai politinei partijai priklauso kongreso narys. Duomenys yra skirstomi taip: 197 pavyzdžių atsitiktinai išrinkta į mokymo imtį, 21 į tikrinimo ir likę 217 į testavimo.

*Diabeto diagnozavimo problema.* Pima indėnų diabeto duomenų imtis sudaryta iš 768 mėginių paimtų iš pacientų, kurie galėjo sirgti diabetu. Kiekvienas mėginys sudarytas iš 8 požymių. 500 mėginių paimti iš žmonių nesergančių diabetu ir 268 iš sergančių. Duomenys yra skirstomi taip: 345 pavyzdžių atsitiktinai išrinkta į mokymo imtį, 39 į tikrinimo ir likę 384 į testavimo.

*Viskonsino diagnostinė krūtų vėžio problema.* Imtį sudaro 569 duomenų taškai, kiekvienas iš jų sudarytas iš 30 požymių. Požymiai yra apskaičiuoti iš skaitmeninio vaizdo, gauto plona adata paėmus krūtų masės, ir apibudina ląstelės branduolio charakteristikas. 357 duomenų taškai apibudina nepiktybinį ir 212 piktybinį auglį. Duomenys yra skirstomi taip: 269 pavyzdžių atsitiktinai išrinkta į mokymo imtį, 30 į tikrinimo ir likę 270 į testavimo.

Šios duomenų imtys buvo pasirinktos, todėl, kad darbo rezultatus būtų galima palyginti su kitais panašaus tipo darbais.

### 2.2. Mokymo parametrai

Šiame darbe yra šeši pasirenkamieji parametrai, t.y. klaidos funkcijos narių stiprumą nusakantys parametrai  $\lambda$ ,  $\epsilon_1$ ,  $\epsilon_2$  ir  $\alpha$ , parametras nusakantis mažų svorių mažinimą iki nulio  $\beta$  ir parametras  $pr_e$  nusakantis kiek reikia mokyti tinklus atskirai. Parametro  $\lambda$  reikšmė priklauso intervalui  $[0, 1]$  ir šiame darbe parinkta 0,9. Parametrų reikšmės  $\epsilon_1=0,1$  ir  $\epsilon_2=0,0001$  parinktos

pagal rekomendacijas pateiktas [13] straipsnyje. Šiems parametrams turėtų galioti nelygybė  $0.1\varepsilon_1 \leq 1000\varepsilon_2 \leq 10\varepsilon_1$  norint, kad įėjimai būtų priverstinai mažinami. Naudojant modifikuotą požymių atrinkimo narį, parametrai  $\varepsilon_1$  ir  $\varepsilon_2$  dauginami iš paslėptų neuronų skaičiaus  $H$ . Parametro  $\beta$  pasirinkta reikšmė 10. Parametro  $pr_e$  reikšmė nustatoma eksperimentiškai, kad tinklas per daug neapsimokytų, tačiau ir nebūtų visiškai naujai inicializuotas. Šiame darbe parinkta reikšmė 10. Parametras  $\alpha$  parinktas 0,9 atsižvelgiant į programinio paketo MATLAB naudojamą standartinę (default) reikšmę.

## 2.3. Eksperimento rezultatai

### 2.3.1. Požymių atrinkimas vienam tinklui

#### 2.3.1.1. Požymiai neatrenkami

Tyrimas pradedamas nuo patikrinimo kokie rezultatai gaunami pasirinktoms duomenų imtims apmokant vieną tinklą. Pirmiausiai tinklas buvo mokomas 100 epochų, nes vidutiniškai tiek epochų pakanka, kad tinklas būtų apmokytas ir jo svoriai tolimesnio mokymo metu nebekinta. Toliau, remiantis tuo, kad neuroninį tinklą mokant be jokių apribojimų, jis dažnai per daug prisitaiko prie mokymo duomenų ir dėl to blogiau klasifikuoja testinėje duomenų imtyje esančius duomenų taškus, buvo panaudotas ankstyvasis tinklo mokymo stabdymas. Ankstyvojo mokymo stabdymo metu, tinklas mokomas su mažiau epochų, nei jam reikia pilnai prisitaikyti prie mokymo imties, tačiau epochų yra pakankamai, kad tinklas išmoktų pagrindinę duomenų taškais nusakytą, tačiau nežinomą funkciją. Taip apmokytas tinklas dažnai demonstruoja geresnius klasifikavimo rezultatus testiniai imčiai. 1 lentelėje pateikiama klasifikavimo klaidos vidutinės reikšmės (procentais) gautos iš 30 bandymų, testinei duomenų imčiai. Skliausteliuose pateiktas standartinis nuokrypis. Stulpelyje „Šaltinis“ nurodytas prieduose esančios lentelės numeris arba literatūra iš kurios buvo paimti skaičiai.

Lentelė 1. Vieno tinklo vidutinė klasifikavimo klaida testinei imčiai

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnozavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>	<i>Šaltinis</i>
100 epochų	5,91(1,49)	28,47(2,30)	4,11(1,10)	Lent. 1
50 epochų	5,24(1,20)	26,12(1,95)	3,61(0,85)	Lent. 2
100 epochų su <i>SR</i>	5,25(1,15)	24,08(1,70)	3,56(1,06)	Lent. 3

Kaip matyti iš lentelės rezultatų mokant neuroninį tinklą su *SR* nariu gaunami rezultatai yra panašūs ir netgi geresni, už gautus naudojant ankstyvą mokymo stabdymą. Būtent todėl ir buvo nuspręsta prie modifikuoto požymių atrinkimo nario pridėti svorių

reguliuojami narių tiems svoriams, kurie neįeina į požymių atrinkimo narių. Taip pat galutiniam tinklo ar kolektyvo apmokymui irgi naudojamas *SR* narys.

### 2.3.1.2. Požymiai atrenkami

Šiame etape tyrinėjamas požymių atrinkimas. Tarpusavyje lyginami įvairūs požymių atrinkimo nariai, t. y. *SR*, *PA+SR1* ir *PAI+SR1*. Mokoma pagal algoritmą aprašytą 1.2.3.1.1 skyrelyje. Apibendrinti tyrimo rezultatai pateikiami 2 lentelėje.

Lentelė 2. Vieno tinklo vidutinė klasifikavimo klaida testinei imčiai, mokymui naudojant įvairius požymių atrinkimo narius

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnozavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>	Šaltinis
[13] straipsnis	5,21(0,29)	25,71(0,59)	3,67(0,81)	Str. [13]
<i>PA</i> narys	5,93(1,47)	26,74(1,90)	4,54(0,56)	Lent. 4
<i>PAI+SR1</i> narys	5,04(0,17)	23,77(1,43)	3,53(0,35)	Lent. 5
<i>SR</i> narys	5,42(1,05)	23,96(0,69)	2,73(0,74)	Lent. 6

Kaip matyti iš lentelės, [13] straipsnyje pateikta klaida yra mažesnė, nei mūsų gauta, taikant [13] straipsnyje pasiūlytą papildomą narių. Šis skirtumas gaunamas dėl skirtingų požymių atmetimo metodų, nors klaidos funkcija išplėsta tuo pačiu papildomu nariu *PA*. Tačiau pasinaudojus modifikuotu požymių atrinkimo nariu *PAI+SR1* klaida ne tik kad pasiekia buvusią (ar panašų) lygį, bet *Diabeto diagnozavimo problemai* gaunami netgi geresni rezultatai.

Svorių reguliavimo nario *SR* panaudojimas požymių atrinkimui duoda labai prieštarigus rezultatus, lyginant su *PAI+SR1* nario panaudojimu. Vienai duomenų imčiai (*JAV kongreso balsavimo įrašų problema*) klaida gavosi didesnė, vienai (*Viskonsino diagnostinė krūtų vėžio problema*) – mažesnė, o dar kitai (*Diabeto diagnozavimo problema*) – beveik tokia pat. Iš to galima spręsti, kad šis metodas tinka tik atskiroms problemoms spręsti.

Toliau tiriama kaip yra atmetami požymiai. Tyrimo rezultatų pavyzdys *Diabeto diagnozavimo problemai* parodytas 3 lentelėje. Lentelėje matosi, kaip mokymo eigoje – skirtingose iteracijose neuroniniam tinklui yra atmetami nereikšmingi požymiai. Lentelėje pateikti skaičiai yra įėjimo svarba (*IS*) apskaičiuota pagal (16) formulę.

Lentelė 3. Požymių atrinkimo eiga *Diabeto diagnozavimo problemai*

		Požymis								
		1	2	3	4	5	6	7	8	Triukšmas
Iteracija	1	1,42	4,28	0,89	1,14	2,02	2,61	2,27	1,80	1,06
	2	0,82	4,18	—	1,11	2,13	2,55	1,74	2,01	1,20
	3	—	4,24	—	—	1,56	2,42	1,47	1,37	1,18
	4	—	4,52	—	—	1,23	2,54	1,59	0,87	—



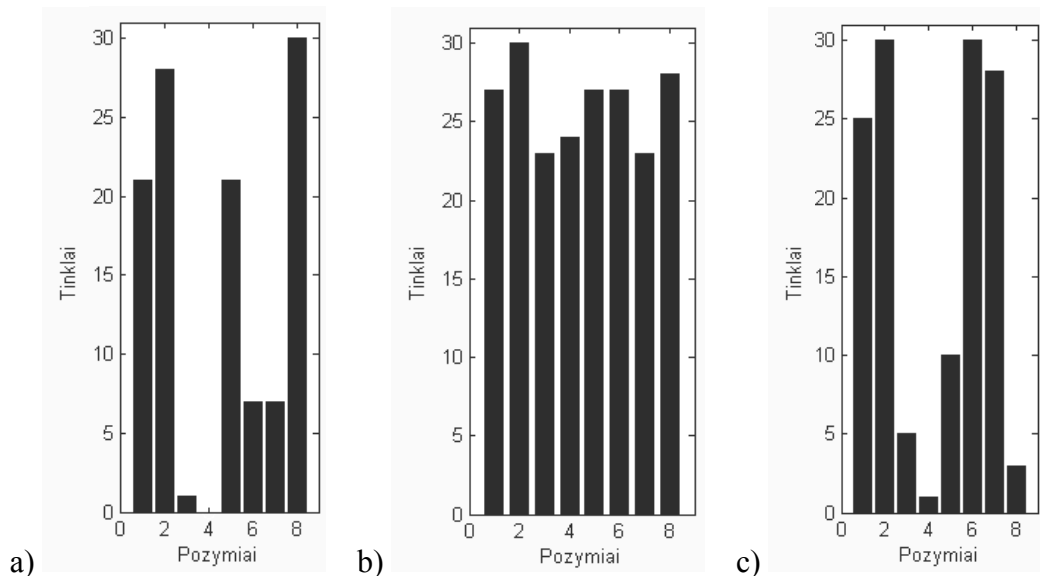
4 lentelėje pateikti skaičiai rodantys, kiek vidutiniškai požymių liko po požymių atmetimo. Skliausteliuose pateiktas standartinis nuokrypis.

Lentelė 4. Vidutinis atrinktų požymių skaičius vienam tinklui, mokymui naudojant įvairius požymių atrinkimo narius

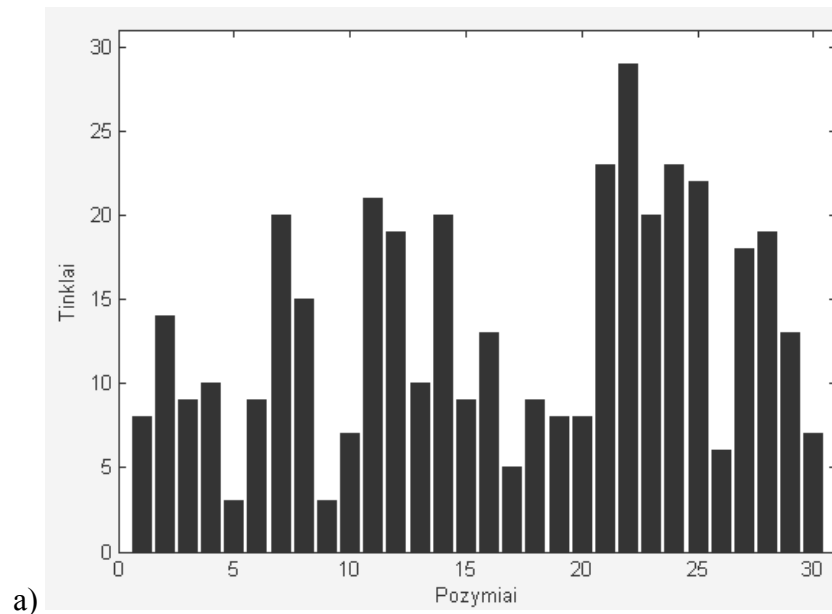
	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnozavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
Viso įėjimų	16,00(0,00)	8,00(0,00)	30,00(0,00)
<i>SR</i> narys	13,03(3,31)	4,40(0,93)	17,27(7,52)
<i>PA+ SRI</i> narys	10,63(5,16)	6,97(1,62)	21,80(9,25)
<i>PAI+ SRI</i> narys	11,87(3,63)	3,83(0,65)	13,33(7,65)
[13] straipsnis	2,03(0,18)	2,03(0,18)	3,52(1,02)

Šiuo atveju geriausias rezultatus pademonstravo [13] straipsnyje aprašytas algoritmas. Jo surastų svarbių požymių skaičius yra mažiausias ir pagal tai, kad klaidos pablogėjimas nėra labai didelis, galima spręsti, kad surasti požymiai ir yra patys svarbiausi.

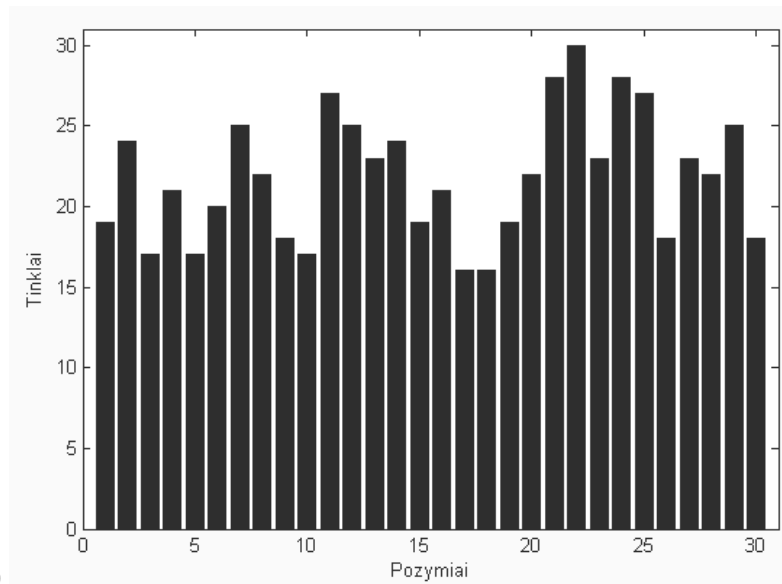
Toliau pateiktuose paveikslėliuose pavaizduota požymių svarba. X ašyje parodyta kiek požymių yra pradinuose duomenyse atitinkamai problemai spręsti, o Y – kiek tinklų po apmokymo turėjo šį požymį. Logiška būtų manyti, kad kuo daugiau tinklų turi atitinkamą požymį tuo svarbesnis jis yra.



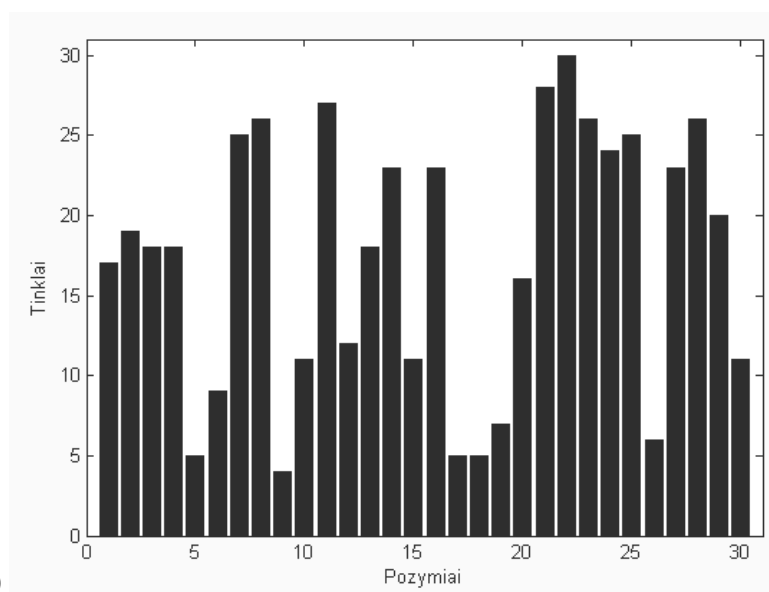
6 pav. Požymių naudojimas *Diabeto diagnozavimo problemai spręsti* su nariu: a)*PAI+SRI*, b)*PA+SRI*, c)*SR*



a)

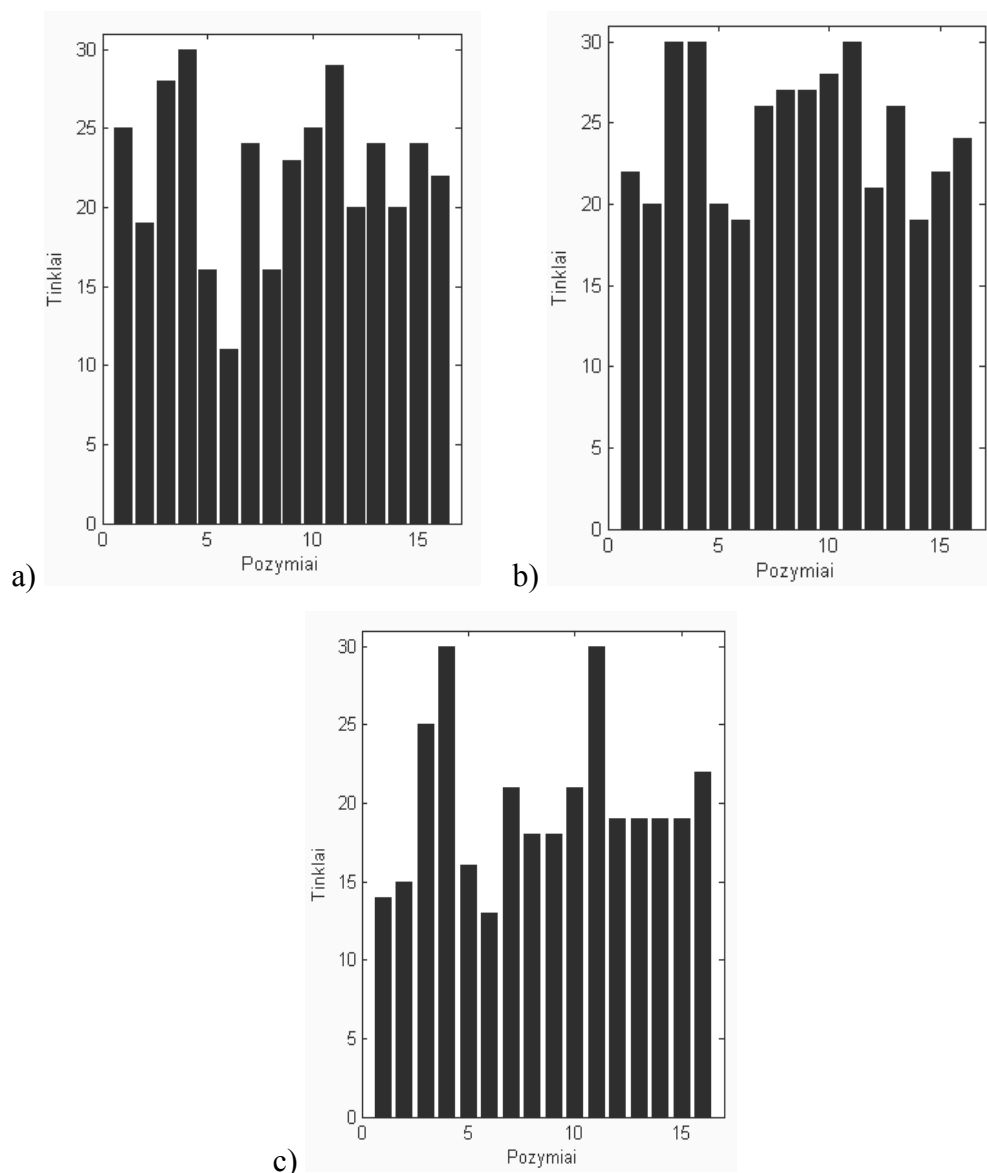


b)



c)

7 pav. Požymių naudojimas *Viskonsino diagnostinei krūtų vėžio problemai spęsti su nariu a)PA1+ SRI, b)PA+SRI, c)SR*



**8 pav. Požymių naudojimas JAV kongreso balsavimo įrašų problemai spręsti su nariu: a) PAI+SRI, b) PA+SRI, c) SR**

Kaip matyti iš pateiktų paveikslų visi metodai išskiria tuos pačius požymius kaip svarbiausius: *JAV kongreso balsavimo įrašų problema* – 3, 4, 11 požymiai, *Viskonsino diagnostinė krūtų vėžio problema* – 21, 22, 24, 25 požymiai, tik *Diabeto diagnostavimo problemai* naudojant modifikuotą požymių atrinkimą su svorių reguliavimu (*PAI+SRI*) ir vien svorių reguliavimą (*SR*) gaunami prieštaringi rezultatai. *PAI+SRI* ir *PA* kaip svarbiausius atrenka – 2 ir 8 požymius, o *SR* – 2, 6 ir 7 požymius. Tačiau dėl to, kad svarbiausias požymis yra antras, net ir šioje problemoje sutaria visi metodai.

Neuroninių tinklų kolektyvų formavimui bus naudojami *PAI+SRI*, *PA+SRI* ir *SR* papildomi nariai bei supaprastintas požymių atrinkimo metodas. [13] straipsnyje aprašytas požymių algoritmas nebus naudojamas, nes yra labai reiklus skaičiavimams (norint atmesti vieną požymį reikia apmokėti tiek tinklų kiek yra požymių), o formuojant kolektyvą

skaičiavimų kiekis dar labiau išauga. Be to, tyrimo tikslas nėra atmesti kaip galima daugiau požymių, bet sukurti tokius tinklus, kurie pagerintų viso kolektyvo klasifikavimo savybės.

### 2.3.2. Požymių atrinkimas tinklų kolektyvui

Toliau tikrinama prielaida, kad skirtingi kolektyvo tinklai, kaip nereikalingus, atmes skirtingus įėjimo kintamuosius. Žemiau pateiktose lentelėse + žymi atvejį, kai tinklas naudoja atitinkamą požymį, o - jei tinklas šio požymio nenaudoja. Jeigu tinklas atmeta visus požymius, tai jis yra šalinamas iš kolektyvo ir į lentelę neįtraukiamas.

Lentelė 5. Reikšmingais pripažintų požymių pasiskirstymas tarp kolektyvo narių *Diabeto diagnostavimo problemai su SR nariu*

		Požymis							
		1	2	3	4	5	6	7	8
Tinklas	1	-	-	-	-	+	-	-	+
	2	-	+	+	+	+	+	+	-
	3	+	+	-	-	-	-	-	+
	4	+	+	+	-	+	+	+	+
	5	+	+	+	+	+	+	+	+
	6	-	-	-	-	-	+	+	+

Lentelė 6. Reikšmingais pripažintų požymių pasiskirstymas tarp kolektyvo narių *Diabeto diagnostavimo problemai su PA+SR1 nariu*

		Požymis							
		1	2	3	4	5	6	7	8
Tinklas	1	+	+	+	+	+	+	-	+
	2	+	+	+	-	-	-	+	+
	3	-	+	-	-	+	-	-	+
	4	-	+	-	-	-	+	-	-
	5	+	+	-	+	+	+	+	-
	6	-	+	+	-	+	+	+	-

Lentelė 7. Reikšmingais pripažintų požymių pasiskirstymas tarp kolektyvo narių *Diabeto diagnostavimo problemai su PA1+SR1 nariu*

		Požymis							
		1	2	3	4	5	6	7	8
Tinklas	1	-	+	-	-	+	+	+	+
	2	-	+	+	+	+	-	-	-
	3	+	+	-	-	+	+	-	+
	4	+	+	+	+	+	+	+	+
	5	+	+	+	+	+	+	-	+
	6	-	+	-	-	+	+	-	-
	7	-	+	-	+	+	-	-	+

Matyti, kad visiems algoritmams skirtingi kolektyvo nariai pasiliko skirtingus įėjimo kintamuosius, todėl galima tikėtis, kad taip bus suformuoti labiau besiskiriantys kolektyvo nariai ir bendras kolektyvo sprendimas bus tikslesnis. Be to, galima pastebėti, kad anksčiau nustatytas svarbiausias (antras) požymis paliekamas dažniausiai, tačiau ne visada.

### 2.3.3. Tinklų kolektyvo klasifikavimo klaida

#### 2.3.3.1. Kolektyvas be papildomų narių

Pirmiausiai neuroninių tinklų kolektyvas buvo apmokytas be jokių papildomų narių, t. y. naudojant paprastą NKM mokymą. 8 lentelėje pateiktas šio neuroninio tinklo kolektyvo palyginimas su vieno neuroninio tinklo apmokymo algoritmais.

Lentelė 8. Tinklų kolektyvo ir vieno tinklo vidutinė klasifikavimo klaida testinei imčiai

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnozavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>	<i>Šaltinis</i>
100 epochų	5,91(1,49)	28,47(2,30)	4,11(1,10)	Lent. 1
100 epochų su SR	5,25(1,15)	24,08(1,70)	3,56(1,06)	Lent. 3
[13] straipsnis	5,21(0,29)	25,71(0,59)	3,67(0,81)	Str. [13]
NKM	5,28(1,31)	24,23(1,48)	3,70(1,14)	Lent. 7

Kaip matyti paprastas NKM mokymas duoda panašius rezultatus kaip vieno neuroninio tinklo apmokymas su papildomais nariais. Taigi yra didelė tikimybė, kad pridėjus papildomus narius prie NKM mokymo, rezultatas pagerės, taip kaip atsitiko su vienu neuroniniu tinklu.

#### 2.3.3.2. Kolektyvas su papildomais nariais

Šiame tyrimo etape, neuroninių tinklų kolektyvai buvo apmokyti su įvairiais papildomais nariais. Jų mokymui buvo naudojami trys skirtingi algoritmai:

1. Mokymas pridedant papildomą narį, bet neatmetant nereikšmingų požymių (žymimas: be PA), mokomas pagal algoritmą aprašyta 1.2.3.2.1 skyrelyje.
2. Mokymas pridedant papildomą narį ir atmetant nereikšmingus požymius (žymimas: su PA), mokomas pagal algoritmą aprašyta 1.2.3.2.2 skyrelyje.
3. Mokymas pridedant papildomą narį, atmetant nereikšmingus požymius, bei naudojant pradinį apmokymą (žymimas: su PA+PrA), mokomas pagal algoritmą aprašyta 1.2.3.2.3 skyrelyje.

Apibendrinti tyrimo rezultatai pateikiami 9 lentelėje.

**Lentelė 9. Tinklų kolektyvo vidutinė klasifikavimo klaida testinei imčiai, mokymui naudojant įvairius papildomus narius**

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnozavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>	<i>Šaltinis</i>
<i>SR</i> be <i>PA</i>	5,36(1,16)	24,31(1,72)	3,47(1,12)	Lent. 8
<i>SR</i> su <i>PA</i>	4,18(0,12)	23,61(1,52)	2,89(0,92)	Lent. 9
<i>SR</i> su <i>PA+PrA</i>	4,53(0,52)	24,01(1,64)	3,15(0,82)	Lent. 10
<i>PA+SRI</i> be <i>PA</i>	4,92(1,49)	23,62(1,52)	3,21(1,05)	Lent. 11
<i>PA+SRI</i> su <i>PA</i>	3,78(1,19)	22,99(1,09)	2,98(0,67)	Lent. 12
<i>PA+SRI</i> su <i>PA+PrA</i>	4,58(1,36)	23,21(1,23)	3,69(1,08)	Lent. 13
<i>PAI+SRI</i> be <i>PA</i>	5,10(1,32)	23,40(1,78)	3,48(1,12)	Lent. 14
<i>PAI+SRI</i> su <i>PA</i>	4,87(1,31)	25,16(0,71)	3,42(1,06)	Lent. 15
<i>PAI+SRI</i> su <i>PA+PrA</i>	3,82(0,61)	21,94(1,11)	2,62(0,85)	Lent. 16

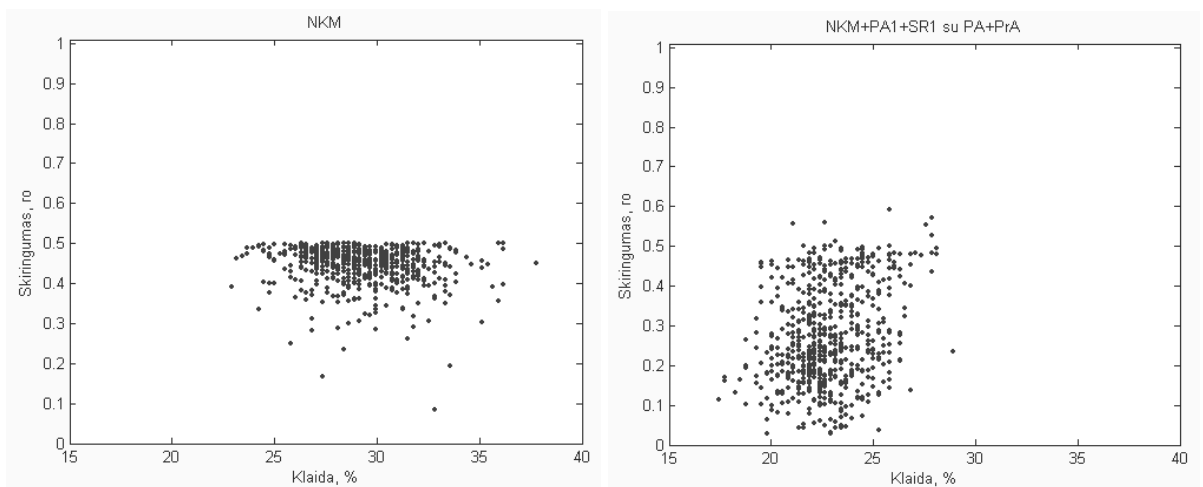
Kaip matyti iš lentelės geriausi rezultatai gaunami naudojant modifikuotą požymių atrinkimo narį (*PAI+SRI*) su požymių atmetimu ir pradiniu apmokymu. Kitas labai geras rezultatas yra paprasto požymių atrinkimo nario (*PA+SRI*) su požymių atmetimu, bet be pradinio apmokymo. Pradinis neuroninių tinklų apmokymas daugiausiai naudos davė naudojant modifikuotą požymių atrinkimo narį. Naudojant jį su kitais papildomais nariais rezultatas ne tik kad nepagerėja, bet netgi pablogėja. Galima pastebėti, kad modifikuotas požymių atrinkimo narys be pradinio apmokymo duoda daug prastesnius rezultatus, nei kiti papildomi nariai.

Lyginant neigiamos koreliacijos mokymą su papildomais nariais ir be jų, akivaizdžiai matosi, kad papildomi nariai pagerina klasifikavimo rezultatus. Be to, naudojant papildomus narius yra atmetama dalis įėjimo kintamųjų ir taip supaprastinamas modelis.

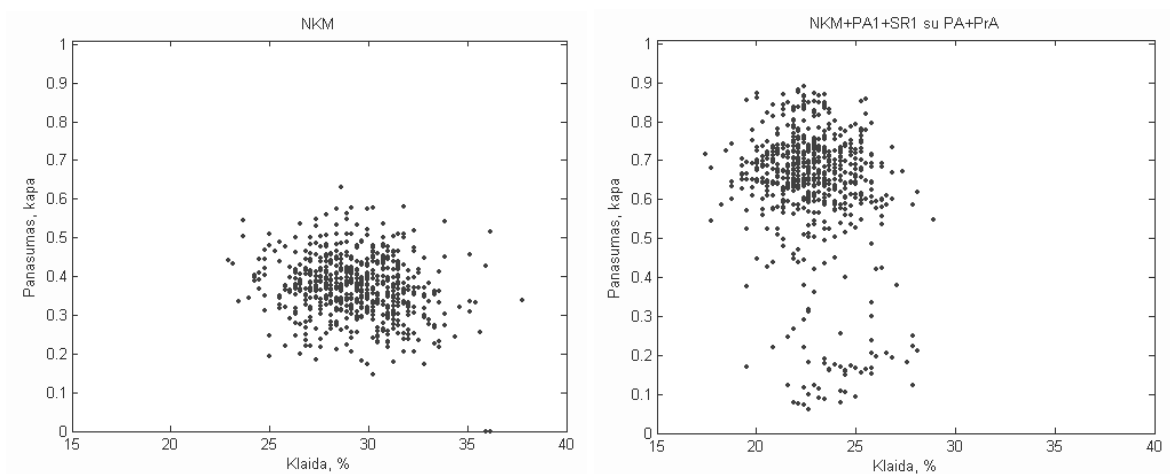
### **2.3.4. Skirtingumo palyginimas**

Žemiau pateiktuose paveikslėliuose pateiktas įvairių kriterijų palyginimas NKM ir NKM+*PAI+SRI* algoritams sprendžiant *Diabeto diagnozavimo problemą*. Lyginami visų 30 neuroninių tinklų kolektyvų nariai. Kiekvienas taškas paveikslėlyje nusako dviejų kolektyvų narių skirtingumą ir klasifikavimo klaidą, jei kolektyve būtų tik šie du kolektyvo nariai. Kuo  $\rho$  reikšmė artimesnė 1 tuo tinklai yra skirtingesni. Kaip matyti paveikslėlyje, skirtingesni tinklai gaunami paprastam NKM kolektyve. Tokie patys rezultatai gaunami ir su  $\kappa$  statistika, bei Q statistika. Kuo  $\kappa$  statistikos reikšmė artimesnė 1, tuo tinklai yra panašesni. Matuojant pagal Q statistiką tinklai taip pat laikomi panašiais kai Q statistikos reikšmė yra artima 1. Ir  $\kappa$  ir Q statistikos rezultatai yra panašūs – yra reikšmių artimų 0, tačiau didžioji dauguma reikšmių yra artimesnės 1. Šie rezultatai buvo netikėti. Tačiau tinklų skirtingumo ir klasifikavimo tikslumo diagramos viską paaiškina. NKM sukuria skirtingesnius atskirus tinklus, tačiau

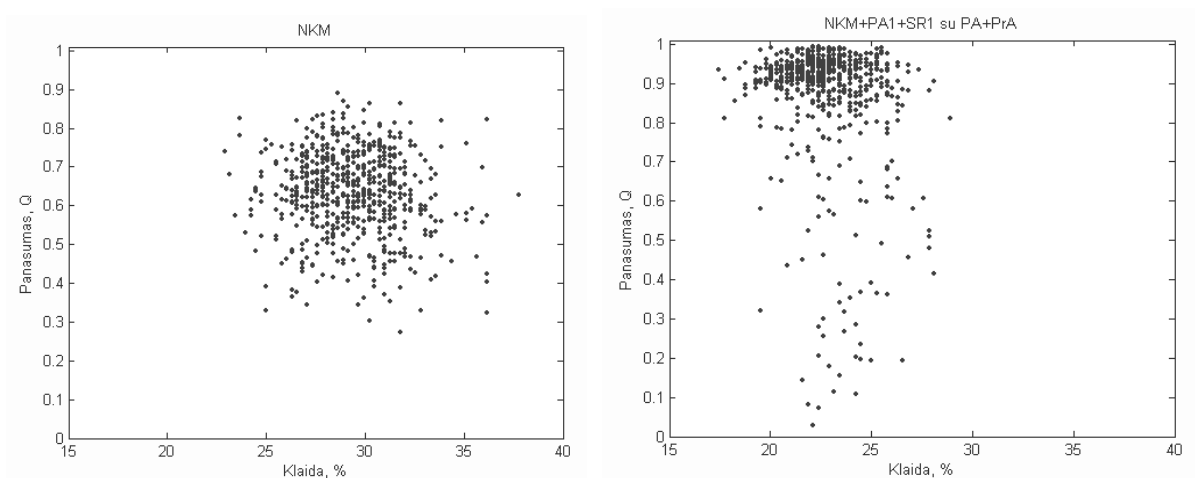
klasifikavimo tikslumo sąskaita. Gautas skirtingumo padidėjimas nekompensuoja tikslumo sumažėjimo. Mokant kolektyvus su papildomais nariais, jų individualių narių klaidos stipriai sumažėjo, dėl ko, padidėjo ir kolektyvo klasifikavimo tikslumas.



9 pav.  $\rho$  kriterijaus palyginimas skirtingiems algoritams sprendžiant *Diabeto diagnostavimo problemą*



10 pav.  $\kappa$  statistikos palyginimas skirtingiems algoritams sprendžiant *Diabeto diagnostavimo problemą*



11 pav.  $Q$  statistikos palyginimas skirtingiems algoritams sprendžiant *Diabeto diagnostavimo problemą*

### 3. Išvados

1. Darbe buvo pristatyti požymių atrinkimo algoritmai neuroniniams tinklams. Pasiūlyta neuroninio tinklo klaidos funkcijos modifikacija, supaprastino nereikšmingų įėjimo kintamųjų atmetimo procedūrą. Nors atmetamų požymių skaičius ir sumažėjo, tačiau klasifikavimo klaida išliko beveik tokia pati, o, be to, buvo žymiai sumažintas skaičiavimų kiekis, kas įgalino naudoti šį algoritmą neuroninių tinklų kolektyvų apmokymui.
2. Kolektyvo mokymui naudotas neigiamos koreliacijos mokymo algoritmas su papildomais požymių atrinkimo nariais. Su visais papildomais nariais algoritmas sėkmingai atmetė skirtingus požymius skirtingiems kolektyvo nariams.
3. Algoritmų testavimui naudotos trys realaus pasaulio problemos. Visais atvejais algoritmai pademonstravo geresnius rezultatus, nei paprastas neigiamos koreliacijos mokymo algoritmas, be to atmetė dalį požymių ir taip supaprastino modelį.
4. Geriausi rezultatai buvo pasiekti naudojant modifikuotą svorių atrinkimo narį ir neuroninių tinklų kolektyvą mokant su pradiniu apmokymu, t.y. pradžioje tinklai apmokomi atskirai, o vėliau visi kartu.
5. Pasiektas pagerėjimas: *JAV kongreso balsavimo įrašų problema* 1,46%, *Diabeto diagnozavimo problema* 2,29%, *Viskonsino diagnostinė krūtų vėžio problema* 1,08%.
6. Darbe buvo ištirta galimybė paprastą svorių reguliavimo narį naudoti požymių atrinkimui viename tinkle ir tinklų kolektyve. Abiem atvejais gauti rezultatai buvo patenkinami. Algoritmas nebuvo pats geriausias, tačiau ieškant paprasto požymių atrinkimo sprendimo visiškai priimtinas.
7. Šiame darbe pasiūlyto skirtingumo mato, pagrįsto apibendrinimo klaida, rezultatai sutapo su populiariaisiais  $Q$  ir  $\kappa$  statistikų, skirtingumo matais, todėl galima sakyti, kad šis matas yra tinkamas skirtingumui tarp tinklų matuoti. Be to, minėtieji matai skaičiuoja klaidų skaičių, kai tuo tarpu pasiūlytasis matas grindžiamas apibendrinimo klaidos įverčiu ir tokiu būdu naudoja išsamesnę informaciją tinklų skirtingumo įvertinimui.
8. Vidutiniškai paprastu NKM algoritmu apmokyti tinklai pademonstravo didesnę kolektyvo narių skirtingumą, nei naudojant NKM+PAI+SRI algoritmą, tačiau individualių tinklų klasifikavimo savybių pagerėjimas vis tiek pagerino ir bendrą kolektyvo sprendimą.



## 4. Literatūra

- [1] **Verikas A., Lipnickas A., Bačauskienė M., Malmqvist K.** Fusing neural networks through fuzzy integration// World Scientific, 2002, p. 227-252.
- [2] **Verikas A., Lipnickas A.** Fusing neural networks through space partitioning and fuzzy integration// Neural Processing Letters, 2002, Nr. 16, p. 53-65.
- [3] **Liu Y., Yao X.** Ensemble learning via negative correlation// Neural Networks, 1999, Nr. 12, p. 1399-1404.
- [4] **Kuncheva L. I. Whitaker C. J.** Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy// Machine Learning. 2003, Nr. 51, p. 181-207.
- [5] **Fleiss J. I., Levin B., Paik M. C.** Statistical Methods for Rates and Proportions. – NY.: John Wiley, 1981. – 500-510 p.
- [6] **Verikas A., Gelžinis A.** Neuroniniai tinklai ir neuroniniai skaičiavimai. – K.: Technologija, p. 7-10
- [7] **Bishop C.** Neural networks for pattern recognition// Oxford University Press, 1995.
- [8] **Krogh A., Vedelsby J.** Neural network ensembles, cross validation, and active learning// Advances in Neural Information Processing Systems. 1995, Nr. 7, p. 231-238.
- [9] **Breiman L.** Bagging predictors// Machine Learning, 1996, Nr. 24, p. 123-140.
- [10] **Freund Y., Schapire R.E.** A decision-theoretic generalization of on-line learning and an application to boosting// Journal of Computer and System Sciences, 1997, Nr. 55, p. 119-139.
- [11] **Breiman L.** Pasting small votes for classification in large databases and on-line// Machine Learning, 1999, Nr. 36, p. 85-103.
- [12] **Young T. Y., Calvert T. W.** Classification, Estimation, and Pattern Recognition. New York: Elsevier, 1974.
- [13] **Setiono R., Liu H.** Neural-network feature selector// Neural Networks. 1997, Nr. 8, p. 654-662.
- [14] **Gill P. E., Murray W., Wright M. H.** Practical Optimization. New York: Academic Press, 1981.
- [15] **Quinlan J. R.** Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann, 1993.
- [16] **Breiman L., Friedman J. H., Olshen R. A., Stone C. J.** Classification and Regression Trees. Belmont, CA: Wadsworth and Rooks, 1984.
- [17] **Monari G., Dreyfus G.** Local overfitting control via leverages// Neural Computation, 2002, Nr. 14, p. 1481-1506.

# 1 Priedas. Neuroninių tinklų ir kolektyvų mokymo rezultatai

Lentelė 1. Vieno tinklo mokymas, 100 epochų

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnostavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	5,91(1,49)	28,47(2,30)	4,11(1,10)
<i>Mokymo imtis</i>	0,00(0,00)	8,71(2,66)	0,03(0,11)
<i>Tikrinimo imtis</i>	5,40(4,96)	27,09(6,27)	4,22(3,21)

Lentelė 2. Vieno tinklo mokymas, 50 epochų

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnostavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	5,24(1,20)	26,12(1,95)	3,61(0,85)
<i>Mokymo imtis</i>	0,19(0,28)	15,62(1,88)	0,38(0,32)
<i>Tikrinimo imtis</i>	6,19(4,53)	24,87(6,87)	3,78(3,04)

Lentelė 3. Vieno tinklo mokymas 100 epochų su SR nariu

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnostavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	5,25(1,15)	24,08(1,70)	3,56(1,06)
<i>Mokymo imtis</i>	1,68(0,79)	18,65(2,04)	0,24(0,41)
<i>Tikrinimo imtis</i>	5,56(5,17)	24,79(7,36)	3,89(4,02)

Lentelė 4. Vieno tinklo mokymas su PA+SR1 nariu atmetant požymius

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnostavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	5,93(1,47)	26,74(1,90)	4,54(0,56)
<i>Mokymo imtis</i>	2,17(1,93)	19,02(3,75)	1,46(0,32)
<i>Tikrinimo imtis</i>	3,33(4,18)	21,03(8,26)	2,67(1,36)

Lentelė 5. Vieno tinklo mokymas su PAI+SR1 nariu atmetant požymius

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnostavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	5,04(0,17)	23,77(1,43)	3,53(0,35)
<i>Mokymo imtis</i>	2,99(0,28)	22,92(1,82)	1,09(0,35)
<i>Tikrinimo imtis</i>	9,21(1,74)	20,34(4,95)	3,67(2,68)

**Lentelė 6. Vieno tinklo mokymas su SR nariu atmetant požymius**

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnostavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	5,42(1,05)	23,96(0,69)	2,73(0,74)
<i>Mokymo imtis</i>	2,12(0,69)	21,04(0,76)	1,43(0,51)
<i>Tikrinimo imtis</i>	5,87(5,11)	21,03(6,33)	2,33(2,65)

**Lentelė 7. Tinklų kolektyvo NKM mokymas**

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnostavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	5,28(1,31)	24,23(1,48)	3,70(1,14)
<i>Mokymo imtis</i>	0,02(0,09)	0,70(0,41)	0,02(0,09)
<i>Tikrinimo imtis</i>	5,87(3,69)	24,18(6,92)	4,33(3,83)

**Lentelė 8. Tinklų kolektyvo NKM mokymas su SR papildomu nariu neatmetant nereikšmingų įėjimų**

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnostavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	5,36(1,16)	24,31(1,72)	3,47(1,12)
<i>Mokymo imtis</i>	0,03(0,13)	17,26(1,90)	0,17(0,23)
<i>Tikrinimo imtis</i>	3,65(3,89)	23,16(6,30)	3,78(3,69)

**Lentelė 9. Tinklų kolektyvo NKM mokymas su SR papildomu nariu atmetant nereikšmingus įėjimus**

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnostavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	4,18(0,12)	23,61(1,52)	2,89(0,92)
<i>Mokymo imtis</i>	0,95(0,26)	18,28(1,58)	0,90(0,48)
<i>Tikrinimo imtis</i>	4,67(2,16)	25,64(7,41)	3,56(1,94)

**Lentelė 10. Tinklų kolektyvo NKM mokymas su SR papildomu nariu atmetant nereikšmingus įėjimus, bei pradinį apmokymą**

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnostavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	4,53(0,52)	24,01(1,64)	3,15(0,82)
<i>Mokymo imtis</i>	0,35(0,14)	17,52(1,56)	0,59(0,32)
<i>Tikrinimo imtis</i>	4,34(3,08)	25,73(6,94)	3,24(2,36)

Lentelė 11. Tinklų kolektyvo NKM mokymas su PA+SRI papildomu nariu neatmetant nereikšmingų  
įėjimų

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnozavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	4,92(1,49)	23,62(1,52)	3,21(1,05)
<i>Mokymo imtis</i>	0,29(0,32)	19,04(1,79)	0,62(0,44)
<i>Tikrinimo imtis</i>	5,71(4,91)	23,59(6,61)	3,78(3,24)

Lentelė 12. Tinklų kolektyvo NKM mokymas su PA+SRI papildomu nariu atmetant nereikšmingus  
įėjimus

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnozavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	3,78(1,19)	22,99(1,09)	2,98(0,67)
<i>Mokymo imtis</i>	0,25(0,66)	17,74(0,54)	0,36(0,36)
<i>Tikrinimo imtis</i>	10,63(4,95)	17,44(4,78)	1,44(2,26)

Lentelė 13. Tinklų kolektyvo NKM mokymas su PA+SRI papildomu nariu atmetant nereikšmingus  
įėjimus, bei pradiniu apmokymu

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnozavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	4,58(1,36)	23,21(1,23)	3,69(1,08)
<i>Mokymo imtis</i>	0,61(0,43)	20,65(2,01)	0,81(1,07)
<i>Tikrinimo imtis</i>	3,33(3,57)	22,99(10,72)	4,11(2,86)

Lentelė 14. Tinklų kolektyvo NKM mokymas su PAI+SRI papildomu nariu neatmetant nereikšmingų  
įėjimų

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnozavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	5,10(1,32)	23,40(1,78)	3,48(1,12)
<i>Mokymo imtis</i>	0,27(0,37)	19,53(2,22)	0,38(0,33)
<i>Tikrinimo imtis</i>	6,35(6,04)	23,68(5,96)	3,56(3,02)

Lentelė 15. Tinklų kolektyvo NKM mokymas su PAI+SRI papildomu nariu atmetant nereikšmingus  
įėjimus

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnozavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	4,87(1,31)	25,16(0,71)	3,42(1,06)
<i>Mokymo imtis</i>	0,34(0,34)	19,16(0,70)	0,51(0,64)
<i>Tikrinimo imtis</i>	5,08(4,13)	23,08(3,87)	4,11(4,26)

Lentelė 16. Tinklų kolektyvo NKM mokymas su *PAI+SR1* papildomu nariu atmetant nereikšmingus įėjimus, bei pradiniu apmokymu

	<i>JAV kongreso balsavimo įrašų problema</i>	<i>Diabeto diagnozavimo problema</i>	<i>Viskonsino diagnostinė krūtų vėžio problema</i>
<i>Testavimo imtis</i>	3,82(0,61)	21,94(1,11)	2,62(0,85)
<i>Mokymo imtis</i>	1,54(0,34)	20,53(0,41)	0,38(0,37)
<i>Tikrinimo imtis</i>	0,63(2,07)	26,84(4,64)	2,67(3,65)