

KIS SZÁMÍTÁSIGÉNYŰ GÉPJÁRMŰDETEKTÁLÓ ALGORITMUS AKUSZTIKUS FORGALOM MONITOROZÁSHOZ

VEHICLE DETECTION ALGORITHM WITH LOW COMPUTATIONAL REQUIREMENTS FOR ACOUSTIC TRAFFIC MONITORING

Kovács Lóránt⁰⁰⁰⁹⁻⁰⁰⁰⁵⁻¹³⁰⁴⁻⁴⁵⁰⁹¹, Pintér István^{orcid.org/0000-0002-2180-0504}^{1*},

¹ Informatika Tanszék, GAMF Műszaki és Informatikai Kar, Neumann János Egyetem, Magyarország
<https://doi.org/10.47833/2024.3.CSC.003>

Kulcsszavak:

akusztikus forgalom monitorozás
gépjárműdetektálás
burkoló
extrém MLP
sE8, MEMS mikrofon

Keywords:

acoustic traffic monitoring
vehicle detection
envelope
extreme MLP
sE8, MEMS microphones

Cikktörténet:

Beérkezett 2024. október 10.
Átdolgozva 2024. október 31.
Elfogadva 2024. november 5.

Összefoglalás

A cikk az akusztikus gépjárműforgalom monitorozás egyik feladatával, az elhaladó gépjármű detektálásával foglalkozik. Részletesen ismerteti egy kis számításigényű algoritmust és bemutatja az osztályozás eredményét mind az sE8, mind a MEMS mikrofonok esetén.

Abstract

The topic of the paper is detection of the passing vehicle as a problem of acoustic traffic monitoring. The paper describes an algorithm with low computational requirements in details and presents the classification results both in case of sE8 and MEMS microphones.

1. Bevezetés

Az utóbbi évtizedben a városi- és városkörnyéki gépjárműforgalom folyamatos megfigyelése, észlelése, nyomon követése – egyetlen szóban összefoglalva: monitorozása – egyre fontosabb lett, mert ez információforrás a városon belüli forgalom irányításához. A forgalom monitorozásban négy feladat különíthető jól el: 1) az elhaladó gépjármű észlelése (gépjárműdetektálás), 2) a haladási irány meghatározása, 3) a sebesség becslése, 4) a gépjárműfajta felismerése.

A gépjárműforgalom monitorozásához érzékelők szükségesek, melyeket vagy az útba építve telepítenek (ilyenek például az indukciós hurkok), vagy melyeknél ilyesfajta beavatkozásra nincs szükség, hanem vagy az út mellé, vagy afölé rögzíthetők. Ez utóbbi kategóriában sokféle érzékelővel találkozhatunk. Elterjedtek azok, melyek elektromágneses hullámokkal működnek: radar, lidar, kamerák, infrakamerák. Nagyobb terület lefedésére elosztott szenzorhálózat is építhető belőlük. A pontosságuk kiváló, de a telepítési és fenntartási költségük nagy, üzembiztonságuk kedvezőtlen időjárás körülmények között csökken. Az akusztikus forgalom monitorozás érzékelői mikrofonok, melyekkel a) kisebb költségigényű, b) egyszerűen telepíthető és fenntartható, c) kis villamos teljesítmény felvételű, d) nagy terület lefedése érdekében elosztott szenzorhálózatba szervezhető, e) a kedvezőtlen időjárásnak jól ellenálló, f) folyamatos működésre képes eszközök hozhatók létre.

* Kapcsolattartó szerző.
E-mail cím: pinter.istvan@nje.hu

Jelenleg élénk kutatás tárgya annak feltárása, hogy az egyes feladatokban milyen teljesítőképesség érhető el az ilyen eszközökön futó programok alapját jelentő algoritmusokkal.

A cikk az elhaladó gépjárművek detektálására kidolgozott, kis számításigényű algoritmusunkat ismerteti. Az algoritmus fejlesztését nyilvános, az akusztikus forgalom monitorozáshoz létrehozott, IDMT-Traffic nevű hang-adatbázissal végeztük (Fraunhofer Institute for Digital Media Technology, [1] (2021)). Megjegyezzük, hogy időközben elérhetővé vált a Sound of Traffic nevű adatbázis is, amely munkánk folytatásában jut majd szerephez (Bosch Center for Artificial Intelligence, [2] (2024)).

A cikk tárgyát jelentő akusztikus gépjárműdetektálásról számos közlemény látott napvilágot. Ezekből kiolvasható, hogy a 99%-os osztályozási pontosságot meghaladó döntést nagyszámú (tízezres nagyságrendtől egészen a néhány milliós nagyságrendig terjedő) tanítható paraméterű, mély neurális hálózatokkal értek el, melyhez a lényegkiemelő transzformáció is jelentős számításigényű (jellemzően sok diszkrét ortogonális transzformált számításával jár). A rendelkezésünkre álló szakirodalomban azonban olyan közleménnyel nem találkoztunk, ami ezekre alapozva realizált, folyamatosan üzemelő, akusztikus szenzorral mért eredményeket tartalmazott volna. Ugyanakkor a kétmikrofonos hangintenzitás-mérés elvén működő, 24 órán át egy forgalmas közút mellett üzemeltetett szenzorral kapott adatokat részletesen ismerteti [3]. A gépjárműdetektálásban kettő MEMS-mikrofonpárral 95% pontosságot értek el az egyetlen mérési procedúra során. Ha a hangintenzitások meghaladtak egy küszöbértéket, úgy döntöttek, hogy gépjármű haladt el. A küszöbértéket rendre a gépjárműmentes időszakban frissített háttérzaj-intenzitás alapján határozták meg. Egymikrofonos rendszerrel elért eredményeket közöl [4]. A számításigényes lényegkiemelő transzformációt követő, gépi tanulással kifejlesztett (a döntési fák családjába tartozó) osztályozóval elért pontosság 96% volt. Mint lentebb olvasható, a jelen cikkben közölt algoritmusunk is lényegében egymikrofonos felvételekkel készült.

A cikk felépítése a következő. A második részben az algoritmusfejlesztéshez használt hang-adatbázist és a jóság számszerű értékelésének módját ismertetjük. A harmadik rész a mikrofonjel burkolójára alapozott lényegkiemelő eljárást részletezi, a negyedik az osztályozáshoz használt MLP hálózatot mutatja be az osztályozási eredményekkel együtt. A cikk következtetésekkel és az irodalomjegyzékkel zárul.

2. A hang-adatbázis felépítése és a gépjárműdetektáló algoritmus jóságának számszerű értékelése

Az IDMT-Traffic adatbázis 17506 wav-fájlból áll, a kétcsatornás (sztereo) felvételek hossza 2 másodperc, a felvételek teljes időtartama kb. 9,7 óra. A felvételek kétfajta mikrofonpárral készültek, egyidejű mintavételezéssel. Az egyik mikrofonpár hangszermikrofonokból (sE8) állt, a másik MEMS mikrofonokból (IvenSense ICS-43434), ami lehetővé teszi, hogy az sE8 mikrofonos eredményeket referenciának tekinthessük a MEMS mikrofonnal elért értékeléséhez. A MEMS mikrofonok ára alacsony, egyetlen integrált áramkörben helyezkednek el a jeldigitalizálás részegységeivel és az I²S interfésszel, emiatt ígéretes jelöltek az akusztikus gépjárműforgalom monitorozási szenzorfejlesztéshez. A felvételkészítéskor a mintavételi frekvencia 48 kHz volt, a kódszavak 24 bitesek. Az annotált adatbázis a négyféle gépjármű (autóbusz, személygépkocsi, motorkerékpár, teherautó) hangján kívül többféle háttérzajt is tartalmaz. Nemcsak a sebességkorlát volt változó (30, 50, 70, „ismeretlen” km/h), hanem az időjárási körülmények is (nedves/száraz út). A háttérzajt egy hosszabb regisztrátum gépjárműhang-mentes részéből vették át, ami lehetővé teszi ezen hangfájlok sebesség szerinti szétválogatását is, ami a hang-adatbázis értékes tulajdonsága. A háttérzajt fájlok száma az sE8 mikrofon esetén 4071, a MEMS-esetben 4073, a gépjárműhangok esetében számuk rendre 4717, illetve 4645.

A gépjárműdetektálás kétosztályos döntés. Feladatunk annak eldöntése, hogy a mikrofonjel elhaladó gépjármű hangjából származik-e (a továbbiakban ez a „gépjárműhang” címke), vagy nem abból („háttérzajt” címke). Az algoritmusfejlesztéshez az annotált hang-adatbázisból kinyerhető, hogy melyik hangfelvétel kapta az egyik és melyik a másik címkét, amit a továbbiakban GT-vel jelölünk (Ground Truth). Ez teszi lehetővé az algoritmusok teljesítőképességének számszerű értékelését és összehasonlítását a tévesztési mátrix alapján. A részleteket illetően a szakirodalomra utalunk [5]. Esetünkben a tévesztési mátrix szerkezete a következő:

1. Táblázat. A tévesztési mátrix szerkezete

GT bemenet (mikrofontípus)	Döntés	
	Gépjárműhang	Háttérzaj
Gépjárműhang	TP (True Positive)	FN (False Negative)
Háttérzaj	FP (False Positive)	TN (True Negative)

A mátrix elemei értelemszerűen az adott GT címkéjű bemenetek esetén a döntések darabszámai. A tévesztési mátrix elemeiből további jellemzők származtathatók. Ezek közül ebben a cikkben az $FPr = FP/(FP+TN)$, a $TPr = TP/(TP+FN)$ és a $(TP+TN)/(TP+FN+FP+TN)$ értékekkel foglalkozunk. Ezek neve rendre: a téves elfogadás aránya, a helyes elfogadás aránya és pontosság. Az (FPr, TPr) számpár különösen fontos, mert ez az osztályozó jelleggörbéjének, az ROC-görbének egy pontja (Receiver Operating Characteristic). Például ha az osztályozó algoritmus olyan, hogy van egyetlen szabad paramétere, akkor valamely paraméterérték mellett felvéve a tévesztési mátrixot és kiszámítva a számpárt, kapjuk az ROC-görbe egy pontját. A paraméterértékek egy sorozatához az ROC-görbe pontjainak sorozata tartozik, amit grafikon formájában lehet ábrázolni. Az osztályozó jóságát szokás még egyetlen számmal, az ROC-görbe alatti terület nagyságával is jellemezni: ez az AUC érték (Area Under Curve). Optimális paraméterértéknek választható például az, melynél az AUC maximális. Az algoritmusfejlesztés során az ROC-görbét, az AUC értéket és a pontosságot használtuk.

3. Lényegkiemelő eljárás a mikrofonjel burkolója alapján

A hang-adatbázisban lévő gépjárműhang- és háttérzaj felvételeket meghallgatva és a mikrofonjeleken előzetes számításokat végezve döntöttünk úgy, hogy a heurisztikus lényegkiemelő eljárás alapja a két hangcsatorna jelének átlagolásával kapott jelhez rendelt burkoló legyen. Ez utóbbi digitális aluláteresztő szűrés eredménye; a bemenet a jelminták négyzetéből alkotott számsorozat. A burkoló időtartama megegyezik a jellel (2 s). Ezután a burkoló elejéről és végéről 250 ms hosszú tranziens jelrészletet elhagyva adódik a középső burkoló szakasz. Végül azért, hogy a burkolóból számolható jellemzők mindegyike független legyen a mikrofonjelek szintjétől, a kétdimenziós lényegvektor komponensei a következők: 1) a középső burkoló szakasz maximumának és átlagának aránya, 2) a középső burkoló szakasz minimumának és átlagának aránya. A burkolót két, egymást követő digitális szűrővel állítottuk elő: az elsőfokú digitális RC-integrátort követi a kétmenetes mozgóátlag szűrő. A digitális RC-integrátor differencia-egyenletét az analóg RC-integrátor átviteli függvényéből a bilineáris transzformációval származtattuk. Itt nem részletezett számításokkal megmutatható, hogy az eredmény

$$y_n = \frac{2RC - T}{2RC + T} \cdot y_{n-1} + \frac{T}{2RC + T} \cdot (x_n + x_{n-1}) \quad (1)$$

Zérus kezdeti értékeket választva ez közvetlenül programozható. Az RC referencia időállandó és a kétmenetes mozgóátlag-szűrő időállandója 100 ms volt. A következő pontban ismertetett osztályozókkal az így kapott kétdimenziós lényegtérben végeztük el a gépjárműhang/háttérzaj döntést.

4. Osztályozás előre-csatolt neurális hálózattal

Az osztályozókat tanuló algoritmusokkal hoztuk létre, amihez a többrétegű perceptron hálózatot (MLP) és a szupport vektor gépet (SVM) használtuk. Ezek működési elvét részletesen tárgyalja a szakirodalom [6]. A tanítást, validálást, tesztelést MATLAB-bal végeztük. Terjedelmi okok miatt cikkünkben az MLP-s változatot ismertetjük, megjegyezve azt, hogy a lineáris SVM osztályozó pontossága is hasonló (az eltérés 0,1%-on belüli volt).

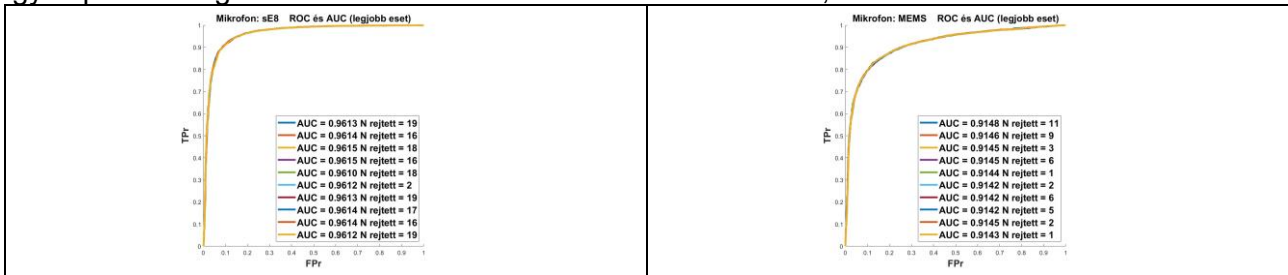
4.1. A sekély neurális hálózat kiválasztása

A kis számításigény érdekében eleve sekély neurális hálózat mellett döntöttünk, melyet a patternnet függvénnyel hoztunk létre. A hálózat egyetlen rejtett réteget tartalmazott, a rejtett

rétegben a feldolgozó egységek kimeneti nemlinearitása a hiperbolikus tangens függvény volt, a kimeneti réteg aktivációs függvénye pedig a szigmoid függvény.

A tanító adatrendszer az sE8 mikrofon esetén rendre a 2x4717 méretű gépjárműhang és a 2x4071 méretű háttérzaj lényegvektorokból álló mátrixok konkatenációja volt. Az 1x8788 méretű célértékvektor első 4717 eleme 1, a többi 0. A MEMS mikrofon esetén a gépjárműhangnál a mátrix mérete 2x4645, a háttérzajé 2x4073, a célértékvektoré 1x8718. A hálózatot a train függvénnyel tanítottuk. Már itt fontos megemlíteni, hogy a tanító vektorok skálázott változata kerül a hálózat bemenetére. A skálázás komponensenként történik, és eredménye az, hogy a [minimum;maximum] intervallum „képe” a $[-1;1]$ intervallum lesz.

Az osztályozást a sim függvénnyel végeztük. Az osztályozás jóságának számszerű jellemzéséhez az ROC-görbét a roc függvénnyel állítottuk elő, és ennek tpr, fpr eredményéből külön számoltuk ki az AUC értéket numerikus integrálással, trapéz-módszerrel. Több hálózatot hoztunk létre egyre növekvő számú (1..20) rejtett rétegbeli feldolgozó egységgel, majd a legnagyobb AUC értéket tekintettük a legjobbnak. Továbbá mivel a train függvény véletlenszerűen választ tanító (70%)/validáló (15%)/tesztelő (15%) diszjunkt részhalmazokat a teljes bemeneti adathalmazból, a rejtett rétegek számára vonatkozó iménti procedúrát megismételtük 10-szer (a kísérletek száma). Összesen 200 neurális hálózat vizsgálatát végeztük el, ezek közül választottuk ki a legjobbakat. Az így kapott ROC-görbék az AUC értékekkel az 1. ábrán láthatók, mikrofononként.



1. ábra. A legjobb MLP hálózatok ROC-görbéi az AUC értékekkel, kísérletenként

A legjobb hálózatok AUC értékeit megvizsgálva kaptuk, hogy az sE8 mikrofonnal kicsit pontosabb döntésre lehet számítani (0,96 körüli AUC-k), mint a MEMS esetén (0,91), de nem ugrásszerű a különbség.

A rejtett rétegek számát nézve meglepő eredményre jutottunk: kiderült, hogy az egy rejtett rétegű, ebben egyetlen neuronnal rendelkező hálózatok is rendre jól teljesítenek. A továbbiakban ezeket extrém MLP-nek nevezzük. Velük szintén 10, a fentihez hasonló kísérletet végeztünk. A kapott eredményeket a 2. ábra szemlélteti.



2. ábra. ROC és AUC eredmények az egy rejtett rétegben egyetlen neuronos hálózattal

Az egyes kísérleteknél a tévesztési mátrixból kiszámoltuk az osztályozás pontosságát is, és a legpontosabbhoz tartozó neurális hálózat súlyvektorát és eltolását feljegyeztük mindkét mikrofonnál. Így kaptuk meg a legkisebb számításigényű neurális hálózatot. A kapott súlyvektorokat a 2. táblázat tartalmazza sorvektorként felírva. A súlyvektoroknál *kiemeltük* azt a komponenst, amit szoktak eltolásnak (*bias*) is nevezni.

2. Táblázat. A legjobb extrém MLP súlyvektorai mikrofontípusonként

Az extrém MLP súlyvektorai	sE8	MEMS
Rejtett réteg	(8,2690 8,7611 0,7925)	(11,3060 11,2444 0,7300)
Kimeneti réteg	(-1,8159 -11,9981)	(4,4968 -11,2394)

4.2. Az extrém MLP, mint lineáris osztályozó

Az extrém MLP előrehaladó számításait részletesen is felírva megmutatható, hogy az valójában lineáris osztályozó. Legyen a hálózat bemenetére jutó lényegvektor $\mathbf{x} = (x_1 \ x_2)$, az egyetlen rejtett réteg egyetlen neuronjának súlyvektora $\mathbf{w}_R = (b \ u \ v)$, kimenete h , a kimeneti réteg egyetlen neuronjának súlyvektora $\mathbf{w}_{KI} = (c \ w)$, kimenete y (ez egyben a hálózat kimenete is). A rejtett rétegben az aktivációs függvény a hiperbolikus tangens függvény (\tanh), a kimeneti rétegben a szigmoid függvény (sigm). A fenti jelölésekkel a kimenet: $y = \text{sigm}(w \cdot h + c)$. Mivel a rejtett rétegben a kimenet $h = \tanh(u \cdot x_1 + v \cdot x_2 + b)$, adódik $y = \text{sigm}[w \cdot \tanh(u \cdot x_1 + v \cdot x_2 + b) + c]$. Amikor éppen nem lehet eldönteni, hogy a bemenet melyik osztályba tartozik, y értéke (az egyenlet bal oldala) 0,5. A szigmoid függvény inverzének értéke a 0,5 helyen zérus, így ekkor $w \cdot \tanh(u \cdot x_1 + v \cdot x_2 + b) + c = 0$. Írható, hogy $\tanh(u \cdot x_1 + v \cdot x_2 + b) = -c/w$, amiből $u \cdot x_1 + v \cdot x_2 + b = \text{atanh}(-c/w)$, ahol atanh a hiperbolikus tangens függvény inverze. Az egyenletből x_2 -t kifejezve adódik az osztályozásra felhasználható egyenes egyenlete:

$$x_2 = -\frac{u}{v}x_1 + \frac{\text{atanh}\left(-\frac{c}{w}\right) - b}{v} \quad (2)$$

A (2) egyenletből látható, hogy az egyenes meredeksége és tengelymetszéke a hálózat súlyvektorai komponenseiből számolhatók, ami azt jelenti, hogy az osztályozáshoz végül kettő paraméter is elegendő, nem szükséges mind az öt, sőt, a számítások a négy alapművelettel elvégezhetők. A 4.1 pontban már említettük, hogy a hálózatra a tanító adatok skálázott változata kerül, emiatt az egyenes iménti paraméterei ebben a koordináta-rendszerben értendők. Tehát ha a burkolóból számolt lényegvektorokkal szeretnénk közvetlenül számolni (kihagyva az említett skálázást), akkor ezt még figyelembe kell venni az osztályozást végző egyenes meredekségének és tengelymetszékének meghatározásakor. A 3. táblázat már az így kapott adatokat tartalmazza.

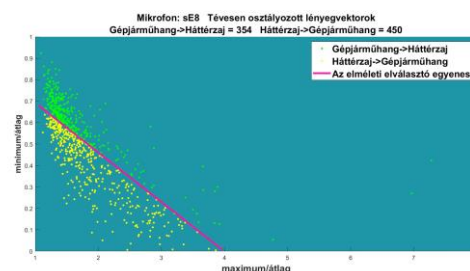
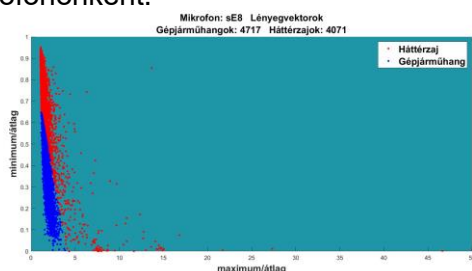
3. Táblázat. Az extrém MLP, mint lineáris osztályozó

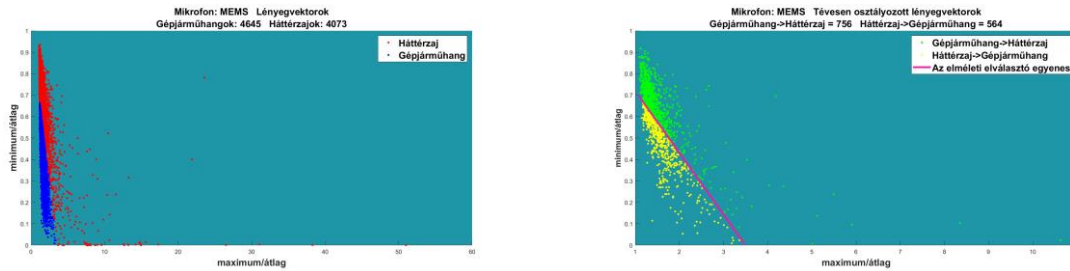
Az egyenesek adatai mikrofontípusonként	sE8	MEMS
Meredekség (m)	-0,2308	-0,2883
Tengelymetszék (b)	0,9218	1,0076

Az egyenes ismeretében az osztályozó algoritmus lépései a következők:

1. Az elsőfokú differencia-egyenlettel és a mozgóátlag-szűrővel állítsuk elő a burkolót.
2. Határozzuk meg a burkoló átlagát, maximumát és minimumát.
3. Számítsuk ki az $x_1 = \text{maximum}/\text{átlag}$ és $x_2 = \text{minimum}/\text{átlag}$ értékeket.
4. Az egyenes 3. táblázatbeli adataival számítsuk ki az $y = m \cdot x_1 + b$ értéket.
5. Ha $y < x_2$, akkor a döntés „gépjárműhang”, egyébként a döntés „háttérzaj”

A lényegvektorok és a tévesztés látható a 3. ábrán a döntéshez használt egyenessel együtt, mikrofononként.





3. ábra. A lényegtér és a tévesztések az extrém MLP-vel

A fenti algoritmussal az sE8 mikrofon esetében 0,9085 (91%), a MEMS mikrofonokkal 0,8486 (85%) pontosságot tudtunk elérni.

A döntés 1 szorzással, 1 összeadással és 1 összehasonlítással meghozható. A lényegvektor számításánál a differencia-egyenlettel az új kimenet az előzőből számolható 2 szorzással és 2 összeadással. Az új mozgóátlag-érték is számítható az előzőből rekurzív eljárással, ami 2 szorzást és 1 összeadást igényel. Az új maximum és minimum is folyamatosan képezhető az előzőből 2 összehasonlítással, majd a lényegvektor komponensei 2 osztás után adódnak. Emiatt nevezzük az algoritmust kis számításigényűnek. Ígéretesnek látszik, de külön megfontolás tárgya kell legyen, hogy a számítások elvégezhetők fixpontos aritmetikával is az osztályozási pontosság megtartása mellett.

5. Következtetések

A cikkben az IDMT-Traffic hang-adatbázissal kidolgozott, kis számításigényű algoritmust ismertettük az akusztikus gépjárműdetektálási feladathoz. Az sE8 mikrofonnal 91%, a MEMS mikrofonnal 85% osztályozási pontosságot értünk el a burkolóra alapozott lényegkiemelő eljárásunkkal és a többretegű perceptron hálózattól származtatott lineáris osztályozóval. Mivel a kis számításigényű kis villamos teljesítményfelvétellel is jár, az algoritmus alapja lehet extrém kis fogyasztású egymikrofonos gépjárműdetektáló szenzorok fejlesztésének. Bár az adatbázis kétmikrofonos méréseket tartalmaz, de a burkolódetekció művelete miatt lényegében egyetlen mikrofon is elegendő abban az esetben, ha gyakorlati megvalósításra kerül a bemutatott algoritmus. A munka során szerzett tapasztalatok alapján a kutatás további céljai a következők: 1) olyan lényegkiemelő transzformáció és mély neurális hálózat struktúra kidolgozása, mellyel elérhető a 99%-ot meghaladó gépjárműdetektálási pontosság, 2) a haladási irány meghatározására szolgáló osztályozók létrehozása mind a kis számításigényű, mind a nagy pontosság-igényű feladatokra.

Irodalomjegyzék

- [1] Abeßer, J., Gourishetti, J., S., Kátai, S., A., Clauß, A., Sharma, T.P., & Liebetrau, J.: IDMT-Traffic: An Open Benchmark Dataset for Acoustic Traffic Monitoring Research. Proceedings of 29th EUSIPCO European Signal Processing Conference, Dublin, Ireland, 23-27, August 2021. pp. 551-555-, DOI:10.23919/EUSIPCO54536.2021.9616080
- [2] Ghaffarzadegan, S., Bondi, L., Lin, W., Kumar, A., Wu, Ho., Horst, H., Das S.: Sound of Traffic: A Dataset for Acoustic Traffic Identification and Counting. Interspeech 2024, 1-5 September 2024, Kos, Greece. pp. 117-121. DOI:10.21437/Interspeech.2024-1205
- [3] Grzegorz Szwoch and Józef Kotus: Acoustic Detector of Road Vehicles Based on Sound Intensity. Sensors 2021, 21, 7781. DOI:10.3390/s21237781
- [4] Gatto, R.C.; Forster, C.H.Q. Audio-based machine learning model for traffic congestion detection. IEEE Trans. Intell. Trans. Syst. 2020, 22, 7200–7207. DOI:10.1109/TITS.2020.3003111
- [5] T. Fawcett: An introduction to ROC analysis. Pattern Recognition Letters. 27 (2006) pp. 861-874. DOI:10.1016/j.patrec.2005.10.010
- [6] S. Haykin.: Neural Networks and Learning Machines. Third Edition. Pearson 2009.