

**ДЕТЕКЦИЈА ОМЕТЕНОСТИ ВОЗАЧА НА ОСНОВУ АНАЛИЗЕ СНИМКА ВОЗАЧА ПРИМЕНОМ МЕТОДА ДУБОКОГ УЧЕЊА****DISTRACTED DRIVER DETECTION BASED ON DRIVER VIDEO ANALYSIS USING DEEP LEARNING METHODS**

Ален Мујо, Факултет техничких наука, Нови Сад

**Област – СОФТВЕРСКО ИНЖЕЊЕРСТВО И ИНФОРМАЦИОНЕ ТЕХНОЛОГИЈЕ**

**Кратак садржај** – Људска грешка и расејана возња представљају главни узрок тешких саобраћајних несрећа. Један од начина за решавање овог проблема је уградња камера на инструмент табли унутар возила које би могле да алармирају возача када се упусти у било коју активност која му одвлачи пажњу од возње. У овом раду представљамо систем заснован на дубоком учењу за откривање ометаних возача користећи 2D слике контролне табле. Скуп података је сачињен од 22.388 RGB слика 26 јединствених возача који су класификовани у десет класа, од којих девет представља ометено понашање возача. Модел се састоји од два дела: VGG19 конволуциона неуронска мрежа за издвајање карактеристика из улазне слике и OpenPose модел за процену људске позе. Коначни модел је постигао тачност од 84% на тестном скупу података.

**Кључне ријечи:** дубоко учење, детекција ометености возача, VGG19, OpenPose, естимација позе човека.

**Abstract** – Human error and distracted driving is the main cause of severe car accidents. One way of solving this problem is installing dashboard cameras inside the vehicle, which would be able to alarm the driver once he engages in any activity that distracts him from driving. This paper presents a deep learning-based system for detecting distracted drivers using 2D dashboard images. The dataset comprises 22,388 RGB images of 26 unique drivers whose behavior is classified into ten classes, nine representing distracted driver behavior. The model consists of the VGG19 convolutional neural network for extracting features from the input image and the OpenPose model for human pose estimation. The final model achieved an accuracy of 84% on the test dataset

**Keywords:** deep learning, distracted driver detection, VGG19, OpenPose, human pose estimation

**1. УВОД**

Према Националној администрацији за безбедност саобраћаја на аутопутевима (NHTSA – National Traffic Safety Administration), 2015. године, 3.477 људи

**НАПОМЕНА:**

Овај рад проистекао је из мастер рада чији ментор је била др Јелена Сливка, ванр. проф.

је погинуло, а 391.000 је повређено у сударима моторних возила у којима су учествовали расејани возачи [1]. Извештај Светске здравствене организације (СЗО) *Global Status Report* за 2017. за 2017. процењује 1.25 милиона смртних случајева годишње због саобраћајних несрећа широм света, при чему је до 50 милиона људи задобило повреде без смртног исхода као резултат саобраћајних несрећа [2]. Према NHTSA, петину саобраћајних несрећа изазове ометани возач, а отприлике 90% саобраћајних несрећа је последица људских грешака у Сједињеним Државама [3].

Центар за контролу и превенцију болести (CDC - *The Centers for Disease Control and Prevention*) дефинише понашање ометане возње узимајући у обзир визуелне (нпр., скретање очију са пута), мануелне (нпр., скидање руку са волана) и когнитивне (нпр., скретање ума са пута) узроке [4]. Да би боље осигурали своје клијенте, *State Farm* је објавио скуп података 2D слика камере на контролној табли са различитим врстама ометаног понашања. Овај рад се фокусира на креирање система који је у стању да детектује девет типова понашања ометених возача како би их алармирао и спречио било какве потенцијалне незгоде.

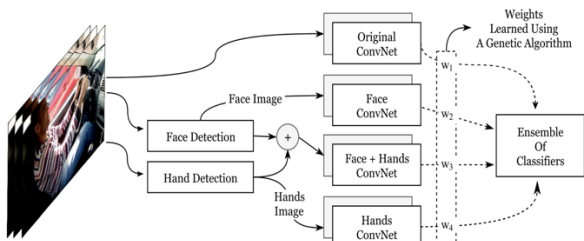
У другом поглављу су приказана два релевантна рада који се баве истим проблемом. Треће поглавље представља скуп података који се користи у овом раду док четврто поглавље приказује коришћену методологију. Методе евалуације и постигнути резултати приказани су у петом поглављу. Последње резимира рад и објашњава потенцијална побољшања.

**2. ПРЕТХОДНА РЕШЕЊА**

У овом поглављу ће укратко бити објашњени сродни радови који решавају проблем детекције ометаних возача користећи исти скуп података који је описан у трећем поглављу.

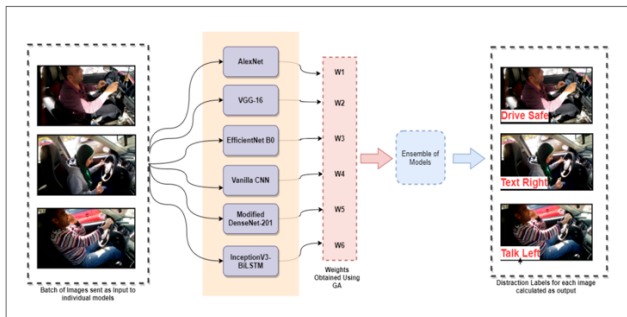
У раду [5] аутори су предложили метод који се састоји од *genetically-weighted* ансамбла конволуционих неуронских мрежа (*Convolutional Neural Network*, CNN). Коришћени CNN модели су *AlexNet* и *InceptionV3*. Ови модели обучени су на четири извора слика: необрађене слике, слике лица, слике руку и слике "лице+руке". Резултат обучавања представља укупно осам модела који производе вектор вероватноће од 10 бројева који одговарају вероватноћама постојећих класа. Коначно предвиђање је израчунато коришћењем отежињеног (*weighted*) система гласања,

наспрот већинског система гласања који би претпоставио да сваки класификатор подједнако доприноси коначном предвиђању. Да би проценили тежину доприноса за сваки класификатор, аутори су користили генетски алгоритам. Скуп података је подељен у размери од 75% за обуку и 25% за тестирање и решење је евалуирано коришћењем тачности као метрике. Предложени метод постигао је тачност класификације од 95.98%. Једно од запажања које су аутори извели је да руке имају већу тежину у предикцији тачне класе него лице. Предложени метод је приказан на слици 1.



Слика 1 - Метода предложена у раду [5].

У раду [6] аутори су користили сличан приступ са *weighted* ансамблима неколико CNN модела приказаних на слици 2. Аутори су користили два скупа података за своје истраживање: AUC (*American University of Cairo*) скуп података о ометаном возачу и *State Farm* скуп података.



Слика 2 - Коначна предикција се даје на основу више различитих модела.

Оба скупа садрже слике које су сврстане у 10 различитих категорија, од безбедне вожње до разних понашања ометања. Предложена методологија обухвата одвојено тренирање неколико различитих модела неуронских мрежа као што су *AlexNet*, *VGG-16*, *EfficientNet B0*, *Vanilla CNN*, модификовани *DenseNet-201* хијерархијски модел и *InceptionV3* интегрисан са слојевима двосмерне дуготрајне меморије (*BiLSTM*).

Сваки од ових модела обучен је на сликама димензије од 224 x 224. Аутори затим израчунавају отежињени збир резултата свих класификатора користећи генетски алгоритам, дајући коначни вектор вероватноће сваке класе.

Овај приступ користи предност идеје да ансамбли могу послужити као јачи класификатори или прецизнији механизам за предвиђање од појединачних слабих класификатора од којих се састоје. Резултати

на тест скупу показали су да је 6-модел *GA-weighted* ансамбл постигао укупну тачност од 96.37% на *AUC* скупу података и 99.57% на *State Farm* скупу података, надмашујући тачност примене појединачних модела над истим скуповима.

### 3. СКУП ПОДАТАКА

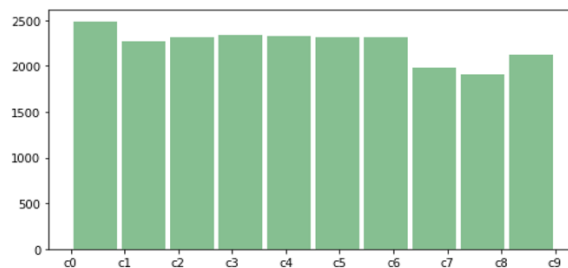
Скуп података који се користи за овај задатак је јавно доступан скуп података који је пре неколико година коришћен за *State Farm Distracted Driver Detection* такмичење на сајту *Kaggle* [7].

Слике су добијене у контролисаном окружењу где је од возача тражено да изведу неколико различитих активности. Садржи 22.388 слика у боји са димензијама 640 x 480 пиксела. Постоји укупно 10 класа у скупу података, које су приказане на слици 3: (c0) безбедна вожња; (c1/c3) слање порука десном/левом руком; (c2/c4) разговара телефоном са десна/лева рука; (c5) управљање радиом; (c6) пијење; (c7) посезање позади; (c8) фризура и шминка; (c9) причање са сапутником (Слика 3).



Слика 3 - Примери 10 класа присутних у скупу података

Свака класа садржи око 2.300 слика, што ово чини уравнотеженим скупом података (Слика 4). Скуп података такође садржи датотеку у којој је свака слика повезана са одређеним идентификационим бројем возача. Постоји 26 јединствених возача на сликама. Тренинг и тест скуп су подељени тако да сваки возач који се налази у тренинг скупу података не буде део и тест скупа података, што се разликује од других радова који су поделили своје податке насумично.



Слика 4 - Дистрибуција броја слика по класама.

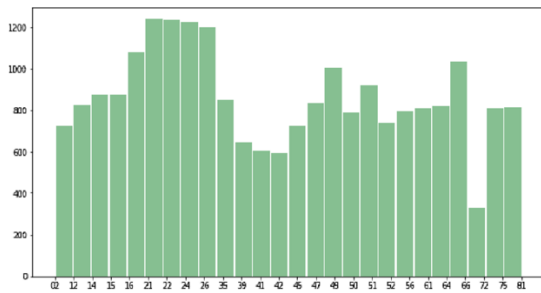
На тај начин циљ модела јесте да научи да препозна активности и да не буде под утицајем самих возача. Тренинг слике се даље деле на основу возача који је присутан на слици насумичним одабиром пет возача и коришћењем њихових слика као скуп података за

валидацију, а преостале слике као скуп података за обуку.

Свака слика је у *JPG* формату и означена је ознаком која указује на активности коју возач ради. Ознаке су нумеричке, одговарају горе наведеним активностима и дате су у засебној *CSV* датотеци.

Поред тога, у *CSV* датотеци постоји колона “*subject*”, која представља јединствени идентификациони број возача са слике.

На слици 5 приказан је број слика за сваког јединственог возача у скупу података.



Слика 5 - Број слика за сваког јединственог возача.

#### 4. МЕТОД

Предложено решење се може поделити у два корака: екстракција особина улазне слике коришћењем *CNN* модела и детекција кључних тачака тела коришћењем модела за естимацију поза људи.

За издвајање особина са улазне слике је коришћен *VGG19* модел са 19 конволуционих слојева који је оригинално трениран на задатку класификације слика на *ImageNet* скупу података од 1000 различитих категорија. Због ограничених *GPU* ресурса, улаз у *VGG19* је слика у боји димензије 84 x 84 x 3.

Уместо коришћења потпуно повезаних слојева димензија 4096 и 1000 које оригинална имплементација модела има, у овом решењу су ти слојеви замењени једним потпуно повезаним слојем од 1024 неурона. За тај вектор од 1024 кажемо да представља издвојене особине са улазне слике.

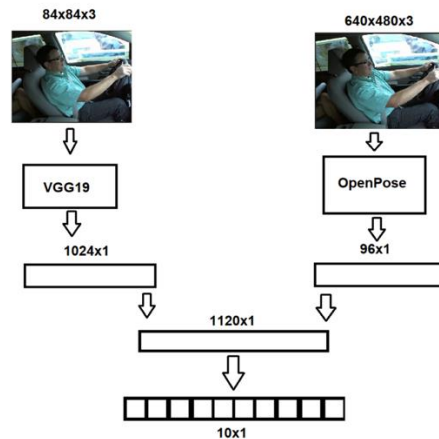
За детекцију кључних тачака тела је коришћен *OpenPose* модел, који је један од првих *real-time multi-person* модела за детекцију људи на слици. За улазну слику модел даје укупно 135 предикција кључних тачака. Улаз у модел је оригинална слика димензија 640 x 480, а излаз је вектор бројева који представља *x* и *y* координате сваке кључне тачке, као и један број који представља поузданост те предикције - *confidence score*.

Кључне тачке које нису детектоване у бар 60% улазних слика, као што су леви или десни чланак, лево ухо, лева или десна пета, су уклоњене. Свака слика је описана са 25 одабраних кључних тачака, тј., вектором од 75 бројева где је свака кључна тачка представљена *x* и *y* координатом и својим *confidence score*-ом.

Како би из овога издвојили додатне карактеристике, на овај вектор су додате још 21 додатне карактеристике које представљају углове између одређених кључних тачака. На пример: угао између кључних тачака зглоба десне руке, десног кука и десног рамена, или

угао између десног зглоба руке, носа и левог зглоба руке. Коначни излаз представља вектор од 96 бројева.

На слици 6 је приказан коначан модел. Излаз из *VGG19* модела за екстракцију особина од 1024 x 1 и излаз *OpenPose* модела за детекцију кључних тачака тела од 96 x 1 су комбиновани у један вектор димензије 1120 x 1. На овај вектор се примењује *softmax* слој који рачуна вероватноћу за сваку од 10 присутних класа. Модел је трениран *cross-entropy* функцијом грешке и *Adam* оптимизатором укупно 10 епоха и вредношћу параметра *batch-size* 40.

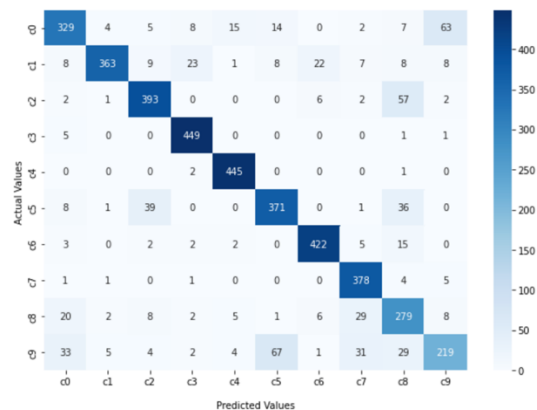


Слика 6 - Предложено решење уључује комбинацију излаза *VGG19* и *OpenPose* модела.

#### 5. РЕЗУЛТАТИ И ДИСКУСИЈА

Скуп података је подељен на основу јединствених возача који су присутни на сликама. Тренинг скуп се састоји од 18.065 слика на којима је присутан 21 возач, од којих је пет возача коришћено као валидациони скуп приликом тренирања. Тест скуп садржи 4.323 слика на којима је присутно пет возача. Оваква подела је извршена како би се модел приликом тренирања фокусирао на саме активности које возачи раде и да не би *overfit*-овао на самим возачима и научио да ради препознавање активности на основу возача.

На слици 7 је приказана матрица конфузије за остварене предикције.



Слика 7 - Матрица конфузије за остварене предикције.

Модел представљен у овом раду је остварио тачност од 84% на тренинг скупу података. У табели 1 су приказане вредности за *precision*, *recall* и *F1-score* за



сваку од 10 класа. Најмања вредност *F1-score* (испод 0.85) је постигнута за класу c0 (безбедна возња), c5 (управљање радиом), c8 (коса и шминка) и c9 (разговор са сувозачем).

<i>class</i>	<i>precision</i>	<i>recall</i>	<i>f1-score</i>
C0	0.80	0.74	0.77
C1	0.96	0.79	<b>0.87</b>
C2	0.85	0.85	<b>0.85</b>
C3	0.92	0.98	<b>0.95</b>
C4	0.94	0.99	<b>0.97</b>
C5	0.80	0.81	0.81
C6	0.92	0.94	<b>0.93</b>
C7	0.83	0.97	<b>0.89</b>
C8	0.64	0.78	0.70
C9	0.72	0.55	0.62

Табела 1 – Прецизност, одзив и *F1-score* за сваку од класа.

Нормална возња (c0) је најчешће погрешно класификована као разговором са сувозачем (c9). Управљање радиом (c5) је најчешће погрешно класификовано као разговор телефоном десном руком (c2) и фризуром и шминком (c8). Коса и шминка (c8) је најчешће класификована као нормална возња (c0) и као посезање позади (c7). Разговор са путницима (c9) најчешће се класификовао као управљање радиом (c5) и нормална возња (c0).

## 6. ЗАКЉУЧАК

У овом раду је представљен модел за детекцију понашања возача који може да детектује 10 различитих активности. Примена решења може потенцијално смањити број саобраћајних несрећа које су углавном узроковане људским грешкама.

За решавање овог проблема је коришћен јавно доступан скуп података који се састоји од 22.388 слика са девет типова понашања која указују на ометеност возача. Проблем је решаван комбиновањем два модела дубоког учења, *VGG19* и *OpenPose* модела. Задатак првог је био да из улазне слике димензија 84 x 84 издвоји карактеристике које су представљене вектором димензије 1024 x 1. Задатак *OpenPose* модела је био да из улазне слике детектује кључне тачке возача присутних на сликама. Улаз у модел је слика оригиналних димензија, а излаз су координате кључних тачака делова тела. На добијених 25 кључних тачака које су описане својим координатама и *confidence score*-ом је додато још 21 карактеристика које представљају углове између ручно одабраних делова тела. Коначан модел примењује *softmax* слој за класификацију на вектор који је добијен комбиновањем излаза *VGG19* и *OpenPose* модела.

Коначни модел је остварио тачност од 84% која је мања од резултата пријављених у релевантној литератури обзиром на ограничене ресурсе приликом тренирања модела. Циљ овог рада није био представљање модела који постиже најбољу тачност,

јер и најтачнији модел који је спор неће наћи примену у решавању почетног проблема, спречавању саобраћајних несрећа упозоравањем возача. Модел је направио највише грешака класификујући безбедну возњу, руковање радиом, фризуру и шминку и разговоре са путницима.

Главни изазов су били ограничени рачунарски ресурси. Једно од усавршавања овог решења би било могуће постојањем више рачунарских ресурса, тиме би могли да повећамо величину улазних слика за *VGG19* мрежу, што би побољшало екстракцију карактеристика.

## 7. ЛИТЕРАТУРА

- [1] Timothy M. Pickrell, Hongying (Ruby) Li, and Shova KC. TRAFFIC SAFETY FACTS, 2016. URL <https://www.nhtsa.gov/risky-driving/distracted-driving>.
- [2] W. H. Organization. World health statistics 2017: monitoring health for the SDGs, sustainable development goals, 2017.
- [3] Strickland. How autonomous vehicles will shape the future of surface transportation. Testimony of The Honorable David L. Strickland Administrator NHTSA, 2013
- [4] US Department of Health & Human Services. Distracted Driving, 2016. URL <https://www.cdc.gov/transportationsafety/distracted-driving/index.html>
- [5] Abouelnaga, Y., Eraqi, H.M. and Moustafa, M.N., 2017. Real-time distracted driver posture classification. arXiv preprint arXiv:1706.09498
- [6] Kumar, A., Sangwan, K.S. and Dhiraj, 2021. A computer vision based approach for driver distraction recognition using deep learning and genetic algorithm based ensemble. ICAISC 2021, Proceedings, Part II 20 (pp. 44-56).
- [7] "State farm distracted driver detection" [www.kaggle.com/c/state-farm-distracted-driver-detection](http://www.kaggle.com/c/state-farm-distracted-driver-detection)

## Кратка биографија:



Ален Мујо рођен је 1999. године. Основне академске студије завршио је 2021. године на Факултету техничких наука, на ком брани и мастер рад 2023. године из области Електротехнике и рачунарства – Софтверско инжењерство и информационе технологије.  
контакт: alen.mujo@gmail.com