

Uticaj mehanizma samopažnje na model dubokog učenja sa ConvLSTM2D slojevima

Uzahir Ramadani, *član ETRAN*, Ivan Lazović, Dušan Nikezić, *član ETRAN*, Nikola Mirkov i Tatjana Miljojić

Apstrakt—U ovom radu smo prethodno razvijeni model za duboko učenje koji se zasniva na ConvLSTM2D slojevima nadogradili mehanizmom samopažnje (eng. *self-attention mechanism*). Novi model smo optimizovali kombinacijom metode pretrage i inteligencijom roja (eng. *Particle Swarm Optimization - PSO*). Model sa mehanizmom samopažnje vrši predikciju optičke debljine aerosola dok su ulazni podaci satelitske slike od 2000. godine na svakih 8 dana. Novi optimizovani model sa mehanizmom samopažnje pokazao je bolje performanse od prethodno razvijenog modela. Izvršili smo ablaciju sloja samopažnje i uporedili sa novorazvijenim optimizovanim modelom. Dobijeni rezultati su pokazali slabije performanse modela sa mehanizmom samopažnje.

Ključne reči—inteligencija roja, mehanizam samopažnje, ConvLSTM2D, aerosol.

I. UVOD

Aerosoli su čvrste čestice ili male tečne kapi nastale iz dima, prašine, letećeg pepela i kondenzovanih gasovitih supstanci koje mogu da se nađu u atmosferskom vazduhu. Aerosoli mogu ostati u atmosferi nekoliko dana. Identifikovanjem procesa koji kontrolišu kako se aerosoli transformišu u atmosferi, naučnici će moći bolje da objasne i predvide njihov uticaj na životnu sredinu i klimu.

U prethodnom radu smo razvili model dubokog učenja uz pomoć ConvLSTM2D slojeva za predviđanje aerosola na 8 dana [1]. Ulazni podaci su bile satelitske slike optičke debljine aerosola (eng. *Aerosol Optical Thickness - AOT*) od 2000. godine čije je merenje vršeno svaki 8 dan što je srednje vreme aerosola provedenog u atmosferi.

Cilj ovog rada je bio da sa istim ulaznim podacima dogradimo model pomoću mehanizma samopažnje. Za bolju

optimizaciju hiperparametara koristili smo osim standardne metode pretrage i optimizaciju inteligencijom roja.

II. METOD

Kao ulazne podatke koristili smo bazu optičke debljine aerosola MODAL2_E_AER_OD u vremenskom razdoblju od 18.02.2000. do 30.10.2022. što je dalo ukupno 1041 slika u PNG formatu rezolucije 3600x1800 piksela. Podatke smo prebacili u JPG format, a zatim napravili ulaznu sekvencu 10 uzastopnih slika sa izmenjenom rezolucijom 144x288 piksela.

Za razliku od prethodnog rada tehnika preklapanja slika nije vršena za 10 slika već za 1 sliku. Time smo dobili 10 puta povećanje ulazne baze podataka.

Takođe smo neočitane piksele konvertovali u belu boju čime su dobijeni rezultati komparativniji. Tamno smeđi pikseli pokazuju visoke koncentracije aerosola, dok žuti pikseli pokazuju srednje koncentracije, a svetložute oblasti pokazuju male količine aerosola.

U veštačkim neuronskim mrežama, mehanizam pažnje je tehnika koja treba da oponaša kognitivnu pažnju. Efekat pažnje poboljšava neke delove ulaznih podataka dok umanjuje druge delove sa ciljem da se posveti više pažnje malim, ali važnim delovima podataka.

A. Pregled literature

Transformatori su zasnovani na enkoder-dekoder arhitekturi i sposobni su da obrađuju čitav niz sekvenci sa mehanizmom pažnje. Vrsta pažnje koja se koristi u transformatoru naziva se samopažnja jer je pažnja usmerena na jednu sekvencu. Ovaj mehanizam povezuje različite pozicije jedne sekvence da bi se izračunala reprezentacija iste sekvence. Mehanizam samopažnje je u stanju da uhvati dugoročne prostorno-vremenske zavisnosti izračunavanjem parnih odnosa između različitih pozicija mapa obeležja tokom funkcije binarne relacije [2].

Transformator nema rekurentnu ili konvolucionu strukturu. Transformator vida (eng. *Vision Transformer - ViT*) deli sliku na niz ne-preklapajućih delića, a zatim snima prostorne karakteristike u globalnom kontekstu pomoću mehanizma samopažnje[3]. Delići slike se tretiraju na isti način kao i reči u obrada prirodnih jezika (eng. *natural language processing - NLP*). Kao alternativa delićima slike, hibridna arhitektura je kada se ulazna sekvencu može formirati iz mapa karakteristika konvolucionog sloja. Svaka lokacija konvolucionih slojeva odgovara nekoj lokaciji slike. Ugrađivanje klasifikacijskog ulaza i ugrađivanje položaja se dodaje (ugrađivanje sekvence) [3].

Uzahir R. Ramadani - Institut za nuklearne nauke „Vinča“ - Institut od nacionalnog značaja za Republiku Srbiju, Univerzitet u Beogradu, Mike Petrovića Alasa bb., 1100 Beograd, Srbija (e-mail: uzahir@vin.bg.ac.rs).

Ivan Lazović - Institut za nuklearne nauke „Vinča“ - Institut od nacionalnog značaja za Republiku Srbiju, Univerzitet u Beogradu, Mike Petrovića Alasa bb., 1100 Beograd, Srbija (e-mail: ivan.lazovic@vin.bg.ac.rs).

Dušan P. Nikezić - Institut za nuklearne nauke „Vinča“ - Institut od nacionalnog značaja za Republiku Srbiju, Univerzitet u Beogradu, Mike Petrovića Alasa bb., 1100 Beograd, Srbija (e-mail: dusan@vin.bg.ac.rs).

Nikola Mirkov - Institut za nuklearne nauke „Vinča“ - Institut od nacionalnog značaja za Republiku Srbiju, Univerzitet u Beogradu, Mike Petrovića Alasa bb., 1100 Beograd, Srbija (e-mail: nmirkov@vin.bg.ac.rs).

Tatjana Miljojić - Institut za nuklearne nauke „Vinča“ - Institut od nacionalnog značaja za Republiku Srbiju, Univerzitet u Beogradu, Mike Petrovića Alasa bb., 1100 Beograd, Srbija (e-mail: tatjana.miljojcic@vin.bg.ac.rs).

Osnovni mehanizam pažnje je skalirani upit-ključ tačka-proizvod (eng. *scaled query-key dot-product*) koji se još naziva i Lungova pažnja [4, 5] i izračunava se preko jednačine (1).

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

Ulaz se sastoji od upita i ključeva dimenzije d_k i vrednosti dimenzije d_v . Za velike vrednosti d_k , skalarni proizvod vektora raste ka velikim vrednostima, gurajući Softmax funkciju u regione gde ima izuzetno male gradijente. Da bi se suprotstavio ovom efektu, skalarni proizvod vektora se skalira za vrednost $(d_k)^{-1/2}$ [2].

Druga vrsta pažnje se sprovodi kroz aditivni sloj pažnje (pažnja u stilu Bahdanau [6]). Lungova pažnja koristi stanja gornjeg skrivenog sloja i u enkoderu i u dekoderu, dok Bahdanau pažnja se fokusira na konkatenciju prethodnog skrivenog stanja i sledećeg skrivenog stanja (eng. *hidden state*).

Za razliku od mehanizma pažnje mehanizam samopažnje ažurira svaku komponentu sekvence agregirajući globalne informacije iz kompletne ulazne sekvence [7]. Funkcija Softmax se često koristi kao poslednja aktivaciona funkcija neuronske mreže za normalizaciju izlaza mreže i pretvaranje linearnog izlaza u verovatnoća. Jednačina (1) je zasnovana na jednačini za kosinusnu sličnost [8].

Sloj pažnje može pomoći neuronskoj mreži u pamćenju velikih nizova podataka. Pažnja pomaže da se pronađe korelacija između stanja kada mreža čita različite tokene u nizu. Stoga, rekurentna neuronska mreža sa pažnjom može da se nosi sa dužom sekvencom i fokusira na različite delove dok proizvodi tokene. Rekurentna neuronska mreža uzima dva parametra u svakom vremenskom koraku a to su ulazno i skriveno stanje [9].

Dakle, memorijska ćelija prethodnog vremenskog koraka je u stanju da sadrži globalne prošle prostorno-vremenske informacije što je prikazano u radu [10] sa mehanizmom samopažnje.

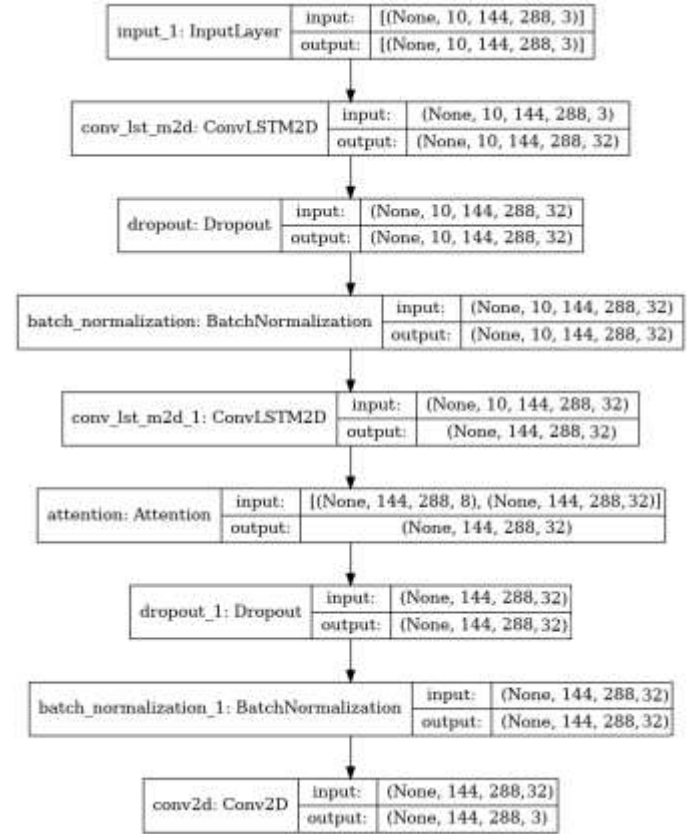
III. MODEL SA MEHANIZMOM SAMOPAŽNJE

U prethodnom radu smo poredili četiri modela za predviđanje slike. U osnovi su bila 2 modela koja su mogla da predviđaju sekvencu iste dužine kao ulaznu (u našem slučaju to je bilo 10 slika) i da predikuju samo jednu sliku. Model gde je na ulazu bilo 10 slika a na izlazu samo jedna da je najbolje rezultate. Stoga smo taj model uzeli kao osnovni model u koji želimo da implementiramo mehanizam samopažnje kako bismo dobili novi model sa boljim performansama.

Novi model smo razvijali korišćenjem softverske biblioteke otvorenog koda koja obezbeđuje interfejs za veštačke neuronske mreže zvane Keras. Keras sadrži sloj pažnje ali ne i sloj samopažnje. Sloj samopažnje smo dobili tako što smo koristili sloj pažnje dva puta, tj. sloj samopažnje je multiplikativne vrste gde napajamo isti tenzor.

Na slici 1 prikazan je dijagram razvijenog modela sa mehanizmom samopažnje kao i sa svim slojevima i

hiperparametrima.



Sl. 1. Dijagram modela sa mehanizmom samopažnje za predviđanje aerosola na 8 dana.

Optimalne hiperparametre smo prvo dobili metodom pretrage za oba ConvLSTM2D sloja. Slojevi prikazani na slici 1 Dropout i BatchNormalization su dodati posle ConvLSTM2D slojeva da bi se izbeglo preterano preklapanje (eng. *overfitting*), dok su vrednosti Dropout sloja dobijene pomoću inteligencijom roja algoritma [11].

IV. REZULTATI I DISKUSIJA

Razvijeni model uz pomoć mehanizma samopažnje smo poredili sa referentnim modelom koji smo razvili u prethodnom radu. Dobijeni rezultati poređenja prikazani su u Tabeli I i II.

Poređenje dva modela smo radili sa novom bazom podataka (korak jedan slika) i sa tehnikom unakrsne validacije (eng. *cross validation*). Unakrsna validacija je tehnika za procenu modela mašinskog učenja obučavanjem nekoliko modela na podskupovima dostupnih ulaznih podataka i njihovom evaluacijom na komplementarnom podskupu podataka. Koristi se da bi se otkrilo preterano poklapanje, tj. neuspeh da se generalizuje obrazac.

Unakrsna validacija je rađena na osnovu definisanog odnosa podskupova za treniranje, validaciju i testiranje baze podataka gde je osnovni odnos bio 70% : 20% : 10%, a ostalo njegove permutacije. Ulazni podaci se kod mašinskog

učenja dele na 3 podskupa i na taj način se vrši evaluacija modela.

Evaluacija novonastalog modela je moguća pomoću nekoliko metrika evaluacije kao što su kosinusna sličnost (eng. *cosine similarity* – CS), srednja kvadratna greška (eng. *root mean square error* - RMSE kao i Euklidsko rastojanje (eng. *Euclidean distance* – EUCD).

U analizi podataka, kosinusna sličnost je mera sličnosti između dva vektora različita od nule definisana u unutrašnjem prostoru proizvoda. Kosinusna sličnost je kosinus ugla između vektora; odnosno to je skalarni proizvod vektora podeljen proizvodom njihovih dužina. Dva vektora se poklapaju ako je kosinusna sličnost jednaka 1 i što je bliža 1 vektori su sličniji. Srednje kvadratno odstupanje je često korišćena mera razlika između vrednosti koje predviđa model i posmatranih vrednosti. RMSE predstavlja kvadratni koren drugog uzorka momenta razlika između predviđenih vrednosti i posmatranih vrednosti ili kvadratnu sredinu ovih razlika i što je RMSE manja manja je i greška. Euklidsko rastojanje predstavlja najkraću udaljenost između dva vektora i što je manje manja je i greška. To je kvadratni koren zbira kvadrata razlika između odgovarajućih elemenata. Euklidsko rastojanje pokazuje visok nivo osetljivosti pa je prikladan za poređenje. U našem slučaju vektor se sastoji od piksela slike koja se predviđa gde se prvo pikseli slike iz dvodimenzionalnog prostora konvertuju u jednodimenzioni prostor.

Iz Tabele I smo zaključili da je najbolja vrednost (najbliža 1) za CS = 0,9986 na podskupu podataka za validaciju. To znači da su dva vektora tj. u našem slučaju dve slike najbližije. Najbolje RMSE i EUCD metrike su 0,0656 odnosno 22,8610 iz podskupa podataka za testiranje.

Dobijeni rezultati pokazuju da je referentni model u stanju da generiše globalnu sliku optičke debljine aerosola, a delimično i da pronade obrasce u vremenskom domenu.

Iz Tabele II smo zaključili da je najbolja vrednost (najbliža

1) za CS=0,9988 za trening podskup podataka kao i za podskup podataka za testiranje. Najbolje vrednosti koje smo dobili za sledeće dve metrike su za RMSE = 0,0613 i EUCD = 21,1876 za podskup podataka za testiranje. Dobijeni rezultati pokazuju da je referentni model u stanju da generiše globalnu sliku optičke debljine aerosola, a delimično i da pronade obrasce u vremenskom domenu.

Optimizacija inteligencijom roja je računarski metod koji optimizuje problem iterativnim pokušajem da se poboljša rešenje kandidata s obzirom na datu meru kvaliteta. Problem se rešava tako što ima populaciju kandidata rešenja, tzv. čestica, i pomera ove čestice u prostoru za pretragu prema jednostavnoj matematičkoj formuli preko položaja i brzine čestice:

$$v(t+1) = wv_i(t) + c_1r_1(p_i - x_i(t)) + c_2r_2(p_g - x_i(t)) \quad (2)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (3)$$

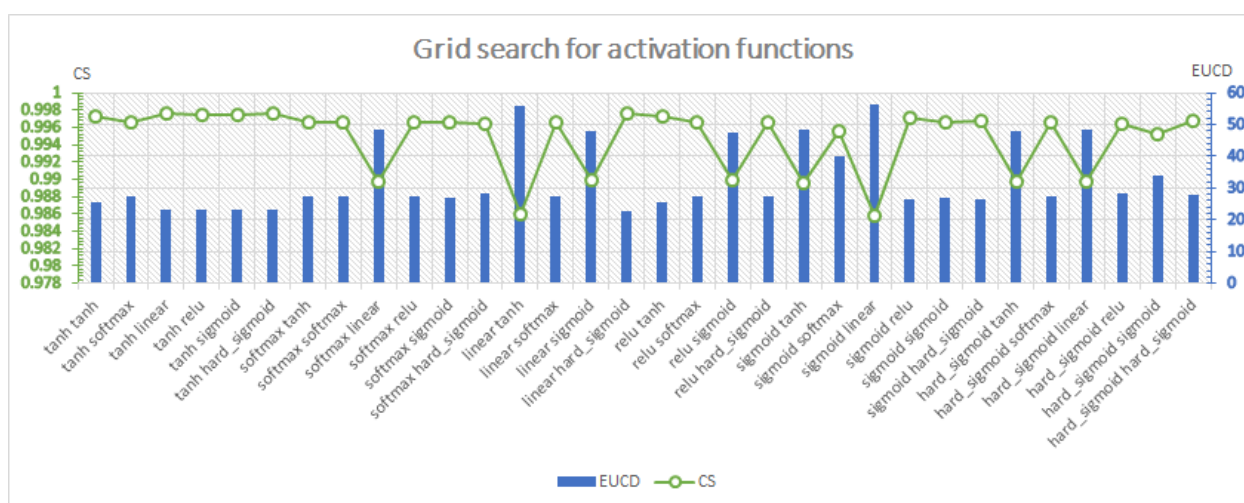
Na kretanje svake čestice utiče njen lokalni najpoznatiji položaj, ali se takođe vodi ka najpoznatijim pozicijama u prostoru za pretragu, koje se ažuriraju kako druge čestice pronalaze bolje pozicije. Očekuje se da će ovo pomeriti roj ka globalnom minimumu odnosno optimalnom rešenju.

TABELA I UNAKRSNA VALIDACIJA ZA MODEL REFERENTNI MODEL

Odnos podskupova za treniranje, validaciju i testiranje	Trening podskup			Validacioni podskup			Podskup za testiranje		
	CS	RMSE	EUCD	CS	RMSE	EUCD	CS	RMSE	EUCD
70:20:10	0.9982	0.0736	23.1746	0.9985	0.0664	23.1272	0.9975	0.0665	23.0200
70:10:20	0.9982	0.0727	23.2243	0.9985	0.0670	23.2473	0.9976	0.0660	22.9765
20:70:10	0.9982	0.0741	23.3053	0.9986	0.0679	23.5969	0.9975	0.0668	23.1850
20:10:70	0.9982	0.0682	23.4433	0.9984	0.0682	23.6810	0.9976	0.0656	22.8610
10:70:20	0.9981	0.0752	23.6591	0.9986	0.0683	23.7524	0.9974	0.0680	23.5864
10:20:70	0.9984	0.0711	23.1147	0.9986	0.0665	23.0807	0.9976	0.0667	23.0897

TABELA II UNAKRSNA VALIDACIJA ZA MODEL SA MEHANIZMOM SAMOPAŽNJE

Odnos podskupova za treniranje, validaciju i testiranje	Trening podskup			Validacioni podskup			Podskup za testiranje		
	CS	RMSE	EUCD	CS	RMSE	EUCD	CS	RMSE	EUCD
70:20:10	0.9987	0.0624	21.7002	0.9987	0.0625	21.6925	0.9978	0.0624	21.5246
70:10:20	0.9988	0.0619	21.2434	0.9988	0.0614	21.2196	0.9979	0.0613	21.2229
20:70:10	0.9987	0.0621	21.308	0.9987	0.0633	21.8114	0.9979	0.0615	21.1876
20:10:70	0.9987	0.0637	22.2504	0.9986	0.0652	22.5577	0.9979	0.0624	21.6773
10:70:20	0.9987	0.0623	21.5556	0.9988	0.0623	21.5717	0.9978	0.0633	21.7486
10:20:70	0.9987	0.063	22.045	0.9987	0.0636	22.0764	0.9977	0.0645	22.2383



Sl. 2. Metoda pretraga za aktivaciju kroz oba ConvLSTM2D sloja.

TABELA III STATISTIČKO POREĐENJE IZMEĐU MODELA SA MEHANIZMOM SAMOPAŽNJE I REFERENTNOG MODELA

	CS		RMSE		EUCD	
	AVG	DIFF [%]	AVG	DIFF [%]	AVG	DIFF [%]
Trening podskup	0.99872 ±4E-3 %	5E-2 ±1E-2	6.25E-2 ±1.07 %	-13.86 ±3.63	21.68 ±1.85 %	-7.01 ±2.04
Validacioni podskup	0.99872 ±7.5E-3 %	1.8E-2 ±1.1E-2	6.3E-2 ±2.08 %	-6.43 ±2.44	21.82 ±2.10 %	-6.80 ±2.45
Podskup za testiranje	0.99783 ±8.2E-3 %	3E-2 ±1.2E-2	6.3E-2 ±1.9 %	-6.06 ±2.27	21.60 1.80 %	-6.57 ±2.10

TABELA IV UNAKRSNA VALIDACIJA ZA MODEL SA ABLACIJOM SLOJA SAMOPAŽNJE

Odnos podskupova za treniranje, validaciju i testiranje	Trening podskup			Validacioni podskup			Podskup za testiranje		
	CS	RMSE	EUCD	CS	RMSE	EUCD	CS	RMSE	EUCD
70:20:10	0.9987	0.0634	21.4830	0.9988	0.0619	21.4671	0.9979	0.0617	21.2680
70:10:20	0.9987	0.0627	21.2872	0.9988	0.0617	21.3075	0.9979	0.0610	21.1572
20:70:10	0.9988	0.0618	21.1337	0.9988	0.0625	21.5810	0.9979	0.0611	21.0578
20:10:70	0.9987	0.0624	21.3761	0.9987	0.0629	21.7307	0.9980	0.0597	20.7411
10:70:20	0.9987	0.0626	21.2960	0.9988	0.0617	21.3259	0.9979	0.0622	21.3959
10:20:70	0.9987	0.0622	21.2285	0.9988	0.0613	21.2381	0.9979	0.0622	21.3617

Hiperparametarsku optimizaciju modela sa mehanizmom samopažnje smo radili u 3 etape. Prve 2 etape su bile zasnovane na metodi pretrage sa dva tipa parametara. Prvi tip je niska tip podataka (eng. *string*) koji smo koristili za najbolju kombinaciju funkcija za aktivaciju ConvLSTM2D sloja. Drugi tip parametra je bio celobrojna vrednost (eng. *integer*) koji smo koristili za najbolju kombinaciju brojeva i veličine filtera u ConvLSTM2D slojevima. Treći tip su brojevi sa pokretnim zarezom (eng. *float*) koje smo koristili sa inteligencijom roja za pretraživanje u kontinuiranom prostoru vrednosti za slojeve dropout i dropout_1 sa slike 1, kao i za vrednost parametra brzina učenja (eng. *learning rate*).

Radili smo dve metode pretrage. Prvu pretragu smo radili da bi dobili najbolju funkciju aktivacije za ConvLSTM2D slojeve. Dobijeni rezultati pokazuju da je najbolja kombinacija za model linearna funkcija za aktivaciju izlaza i kao i tvrda sigmoidna funkcija (eng. *hard_sigmoid*) za sloj recurrent_activation. Te vrednosti su: CS = 0,9975, EUCD = 22,8787 i RMSE = 0,0662, slika 2.

Drugu metodu pretrage smo radili kako bi odabrali adekvatan broj filtera i veličinu filtera za oba ConvLSTM2D sloja. Optimalna kombinacija koju smo dobili bila je 32,32,3 sa metrikom CS = 0,9977, EUCD = 22,3570 i RMSE = 0,0648 za podskup podataka za testiranje.

Poslednju fazu optimizacije hiperparametara radili smo inteligencijom roja za slojeve dropout i dropout_1 sa slike 1 kao i za dobijanje vrednosti parametra brzina učenja. Optimalne vrednosti koje smo dobili bile su 0,15649 za dropout sloj, 0,43055 za dropout_1 sloj kao i 0,00436 za vrednost parametra brzina učenja. Za ove parametre metrike koje smo dobili su bile CS = 0,9978, EUCD = 21,6647 i RMSE = 0,0632 za podskup podataka za testiranje.

Nakon što smo optimizacijom hiperparametara dobili novi model sa mehanizmom samopažnje izvršili smo statističko poređenje između tog modela i referentnog modela. Dobijeni rezultati su prikazani u Tabeli III.

Iz Tabele III se vidi da metrika kosinusne sličnosti nije značajno promenjena, ali druge dve metrike RMSE i EUCD

pokazuju najbolja poboljšanja od -13,86% i -7,01%, respektivno. Rezultate za skup testova treba smatrati najobjektivnijim, pa su dobijena poboljšanja komparativno 0,03%, -6,06% i -6,57% za CS, RMSE i EUCD metrike respektivno. Ove vrednosti nam govore da je novi model razvijen sa mehanizmom samopažnje pokazao bolje performanse od referentnog modela.

Da bismo procenili uticaj i doprinos mehanizma samopažnje, izvršili smo ablaciju sloja samopažnje novog modela koji smo dobili optimizacijom hiperparametara koju smo prethodno opisali. Ispitivanje smo izvršili sa istim vrednostima hiperparametrima. Rezultate unakrsne validacije za model sa ablacijom prikazali smo u Tabeli IV.

Iz Tabele IV smo zaključili da je najbolja vrednost (najbliža 1) za CS = 0,9988 što je kao i za model sa mehanizmom samopažnje ali za druge dve metrike smo dobili bolje rezultate RMSE = 0,0597 i EUCD = 20,7411 i iz toga zaključujemo da mehanizam samopažnje nije poboljšao novi model koji smo dobili optimizacijom metode pretrage i inteligencijom roja.

Slične rezultate i zaključak su dobili autori u radu [12] gde je njihov razvijen model davao bolje rezultate bez sloja pažnje.

V. ZAKLJUČAK

U cilju poboljšanja prethodno razvijenog modela implementirali smo mehanizam samopažnje gde smo prethodno izvršili optimizaciju hiperparametara novom metodom koju smo razvili i koja se sastojala u kombinaciji metode pretrage i optimizacije inteligencijom roja. Novi optimizovani model sa mehanizmom samopažnje je pokazao bolje performanse od prethodno razvijenog modela tj. referentnog modela iz prethodne studije. Studija ablacije sa slojem samopažnje pokazala je niže vrednosti metrika RMSE i EUCD (u proseku za $\approx 1,9\%$) u poređenju sa novorazvijenim modelom sa mehanizmom samopažnje. Poređenje i evaluacija modela nam je pokazala da mehanizam samopažnje nije ispunio naša očekivanja.

ZAHVALNICA

Finansirano od strane Ministarstva nauke, tehnološkog razvoja i inovacija Republike Srbije 451-03-47/2023-01/200017 (istraživačka tema 1002305).

LITERATURA

- [1] Nikezić, D.P.; Ramadani, U.R.; Radivojević, D.S.; Lazović, I.M.; Mirkov, N.S. Deep Learning Model for Global Spatio-Temporal Image Prediction. *Mathematics* 2022, 10, 3392.
- [2] Vaswani A., et al., Attention Is All You Need, 2017, arXiv:1706.03762; <https://doi.org/10.48550/arXiv.1706.03762>.
- [3] Dosovitskiy A., et al., An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, 2021, arXiv:2010.11929;
- [4] Minh-Thang Luong, Hieu Pham, Christopher D. Manning, Effective Approaches to Attention-based Neural Machine Translation, 2015, arXiv:1508.04025.
- [5] Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, Mubarak Shah, Transformers in Vision: A Survey, 2022, arXiv:2101.01169;
- [6] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio, Neural Machine Translation by Jointly Learning to Align and Translate, 2016, arXiv:1409.0473.
- [7] Zhihui Lin et al., Self-Attention ConvLSTM for Spatiotemporal Prediction, The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20); <https://doi.org/10.1609/aaai.v34i07.6819>
- [8] James Wensel, Hayat Ullah, Arslan Munir, ViT-ReT: Vision and Recurrent Transformer Neural Networks for Human Activity Recognition in Videos, 2022, arXiv:2208.07929
- [9] Łukasz Kaiser, Samy Bengio, Can Active Memory Replace Attention?, 2017, arXiv:1610.08613.
- [10] Ge, H.; Li, S.; Cheng, R.; Chen, Z. Self-Attention ConvLSTM for Spatiotemporal Forecasting of Short-Term Online Car-Hailing Demand. *Sustainability* 2022, 14, 7371.; <https://doi.org/10.3390/su14127371>
- [11] Zivkovic, Miodrag & Jovanovic, Luka & Ivanovic, Milica & Bacanin, Nebojsa & Strumberger, Ivana & pm, Joseph. (2022). XGBoost Hyperparameters Tuning by Fitness-Dependent Optimizer for Network Intrusion Detection. 10.1007/978-981-19-2130-8_74.
- [12] Maximilian Spliethöver, Jonas Klaff, Hendrik Heuer, Is It Worth the Attention? A Comparative Evaluation of Attention Layers for Argument Unit Segmentation, 2019, arXiv:1906.10068 <https://doi.org/10.48550/arXiv.1906.10068>

ABSTRACT

In this study, we upgraded with a self-attention mechanism the previously developed deep learning model based on ConvLSTM2D layers. We optimized the new model using a combination of the grid search method and particle swarm optimization. The model with a self-attention mechanism predicts the aerosol optical thickness while the input data are satellite images from the year 2000 every 8 days. The new optimized model with the self-attention mechanism showed better performance than the previously developed deep learning model. We performed an ablation study of the self-attention mechanism and compared it with the new developed optimized model. The obtained results showed a weaker performance of the model with the self-attention mechanism.

Impact of self-attention mechanism on deep learning model with ConvLSTM2D layers

Uzahir Ramadani, Ivan Lazović, Dušan Nikezić i Nikola Mirkov