

Analiza performanței modelului NowDeepN pentru predicția valorilor produselor radarului meteorologic WSR- 98D Bobohalma

Proiect WeaMyL

Code: RO-NO-2019-0133

Contract: No 26/2020

Proiectul WeaMyL

Parteneriat între ANM, UBB Cluj-Napoca și Met Norway

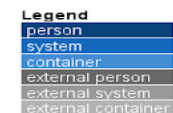
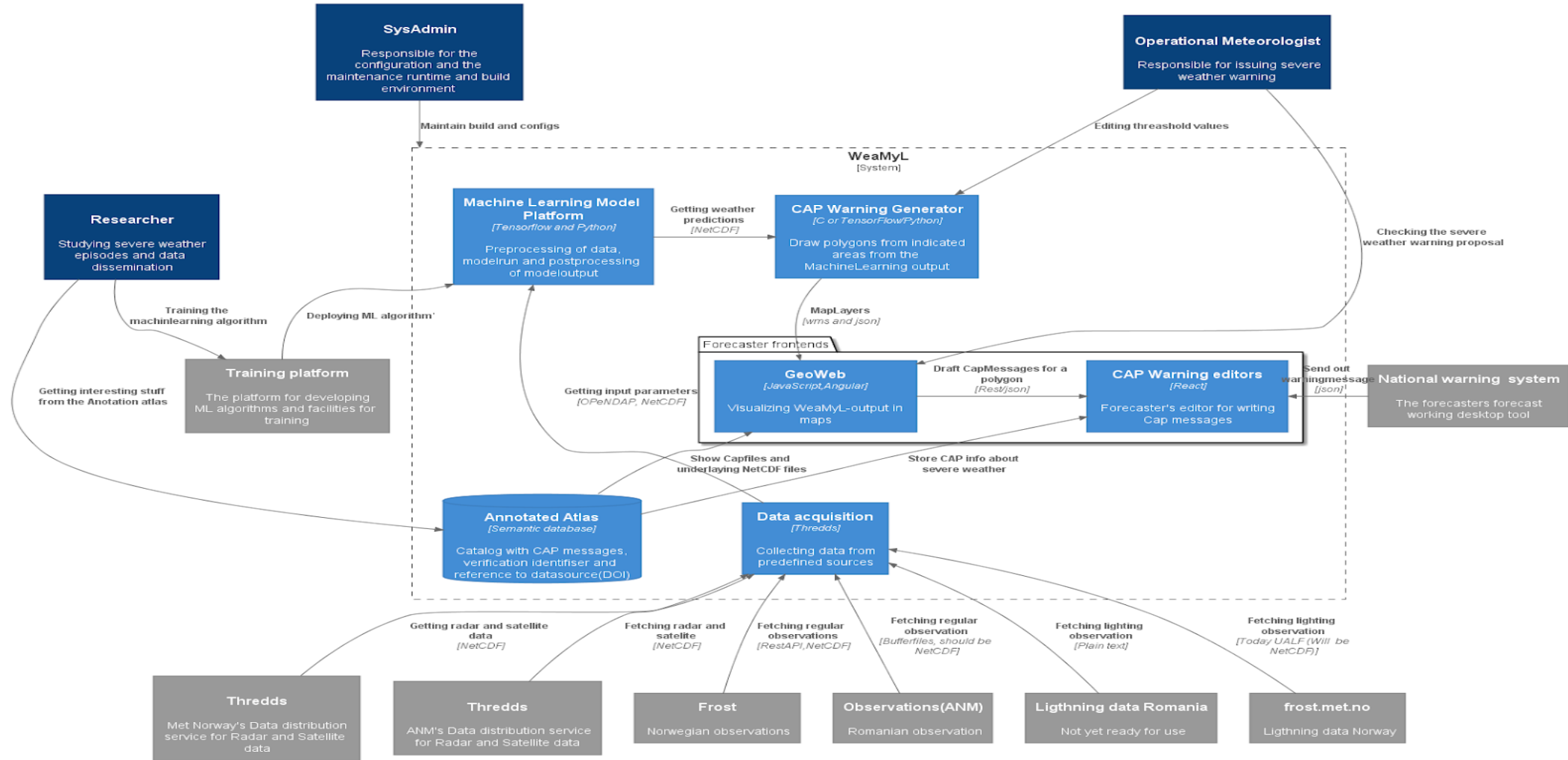
01. Elaborarea și validarea științifică a unor noi modele și tehnici de calcul bazate pe ML, special concepute pentru prognoze precise.

02. Dezvoltarea Atlasului adnotat al observațiilor meteorologice, o bază de date de mari dimensiuni care conține date meteorologice și reprezintă sursa de date pentru modelele de ML dezvoltate în cadrul proiectului.

03. Integrarea platformei WeaMyL în cadrul sistemelor naționale de avertizare de vreme severă din Norvegia și România.

Activități în primul an de proiect

- Stabilirea specificațiilor funcționale ale sistemului
- State-of-the-art
- Stabilirea arhitecturii sistemului
- Identificarea datelor relevante pentru modele de ML
- Dezvoltarea și testarea componentei de achiziție de date
- Preprocesarea datelor meteorologice istorice
- Dezvoltarea și testarea de modele nesupervizate și supervizate pentru analiza datelor
- Activități de diseminare (7 articole publicate, workshop-uri, website)

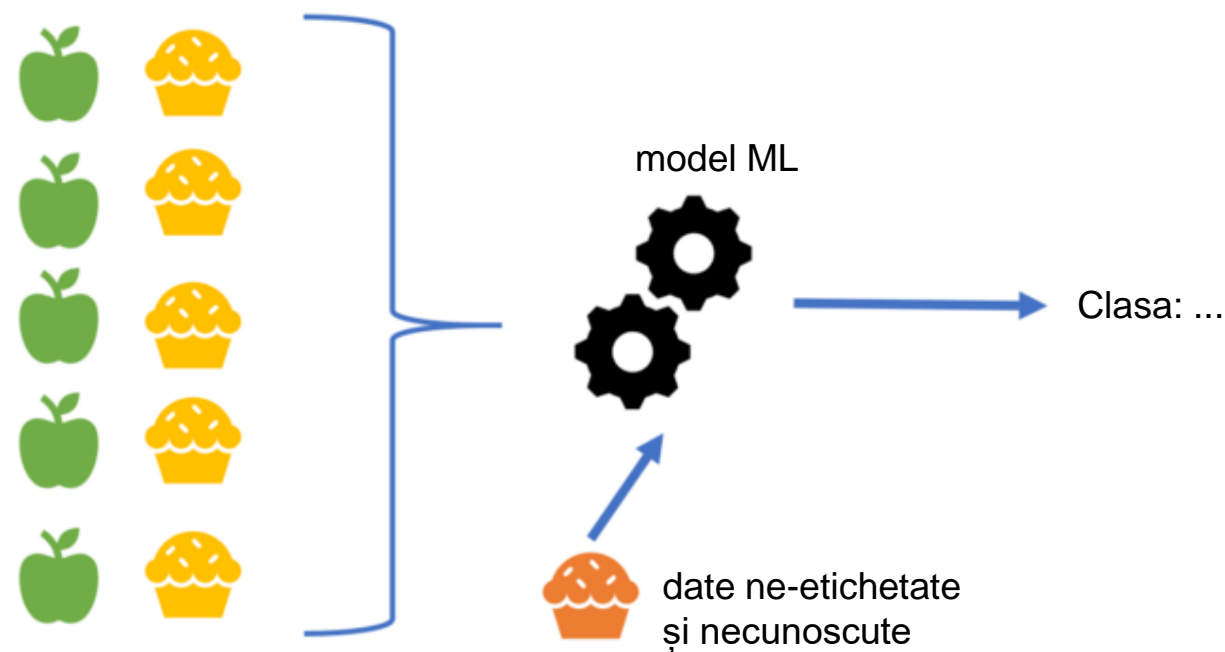


Modelul NowDeepN

Model de învățare supervizată bazat pe un ansamblu de regresori de rețele neuronale profunde (Deep Neural Networks - DNN) pentru prezicerea valorilor pentru produsele radar, care mai departe pot fi utilizate pentru nowcasting.

Modelul NowDeepN

Modele supervizate - un tip de algoritmi ML care au ca scop construirea unei funcții de asociere a unei intrări la o ieșire pe baza unor exemple etichetate de intrări-ieșiri.



Modelul NowDeepN

Modelele DNN

- conțin unități de bază (neuronii artificiali), care primesc ca intrare o serie de numere cu valoare reală și produc o singură ieșire cu valoare reală
- conțin mai multe straturi ascunse și au un număr mare de parametri
- generalizare și asociere de date care nu au fost antrenate în mod explicit și să coreleze aceste date cu o clasă căreia îi aparțin
- problema vitezi mici și a overfitting-ului

Modelul NowDeepN

Scopul analizei este de a obține o dovadă empirică a faptului că, atât în condiții meteorologice normale, cât și în condiții de vreme severă, valorile unui produs radar la un moment dat, într-o anumită locație, sunt previzibile în funcție de valorile din locațiile învecinate din momentele anterioare.

Teme investigate – RQ1

Sunt DNN capabile să prezică valorile unui produs radar la un moment dat într-un anumită locație geografică din valorile locațiilor sale vecine din timpurile momente anterioare? În ce măsură acest lucru este valabil atât pentru condițiile meteorologice normale, cât și pentru cele severe?

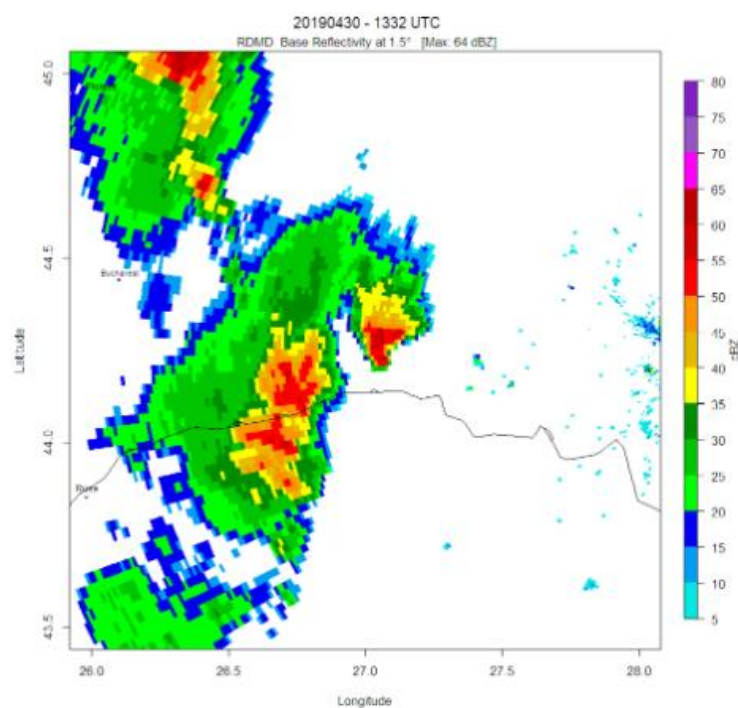
Teme investigate – RQ2

Cum este influențată performanța de predicție a modelului NowDeepN de etapa de curățare a datelor introdusă pentru corectarea datelor de intrare eronate ?

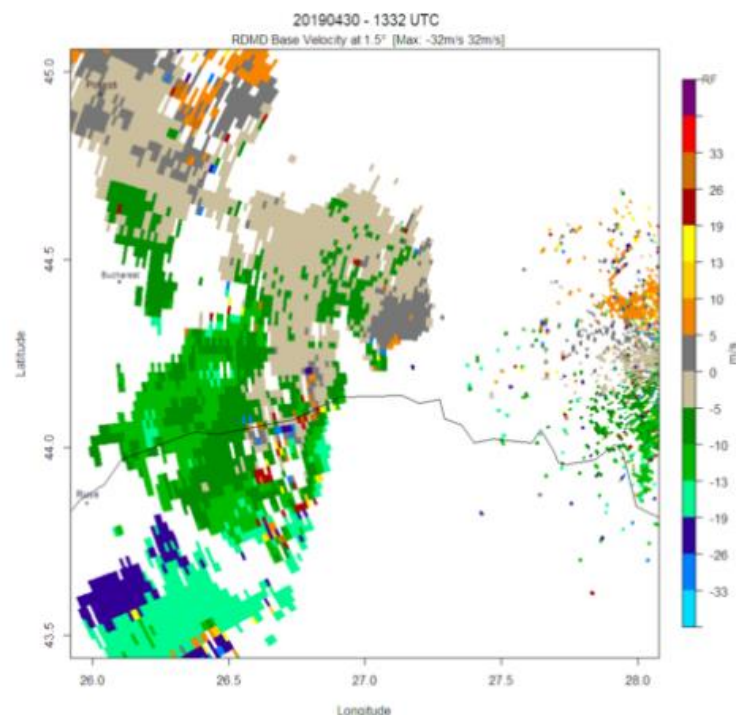
Teme investigate – RQ3

Cât de relevante sunt caracteristicile luate în considerare în sarcina de învățare supravegheată? Mai precis, sunt valorile din zona învecinată a unei anumite locații geografice / la momentul $t-1$ relevante pentru a estima valorile produselor radar din locația / la momentul t ?

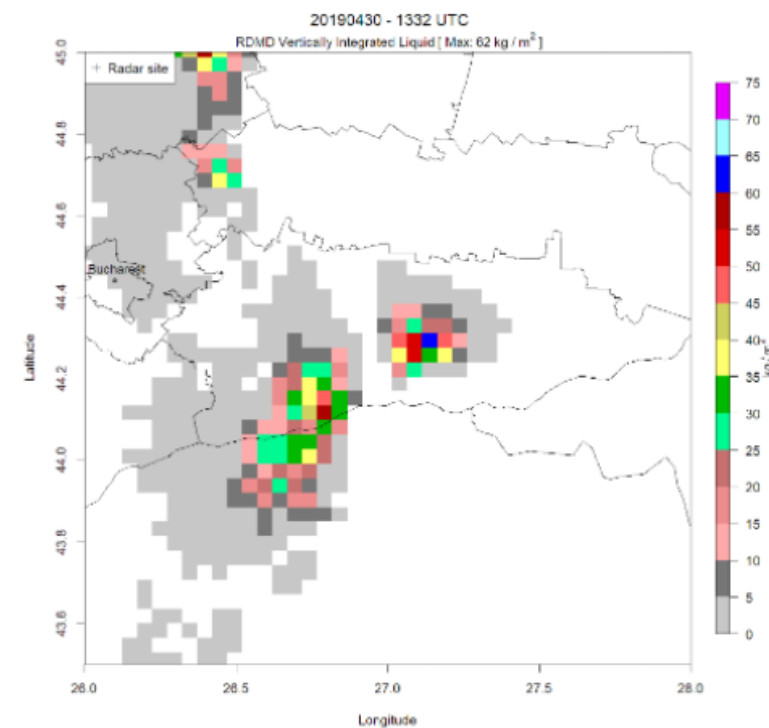
Date radar folosite



Reflectivitate R01-R07



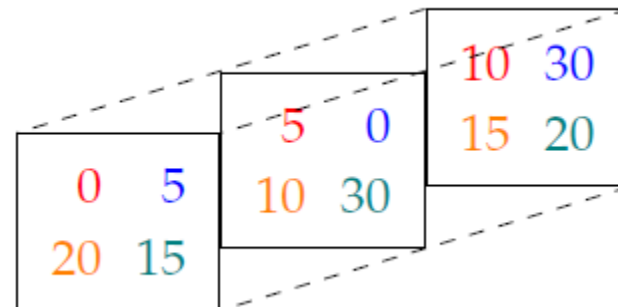
Viteza doppler V01-V07



VIL

Data model

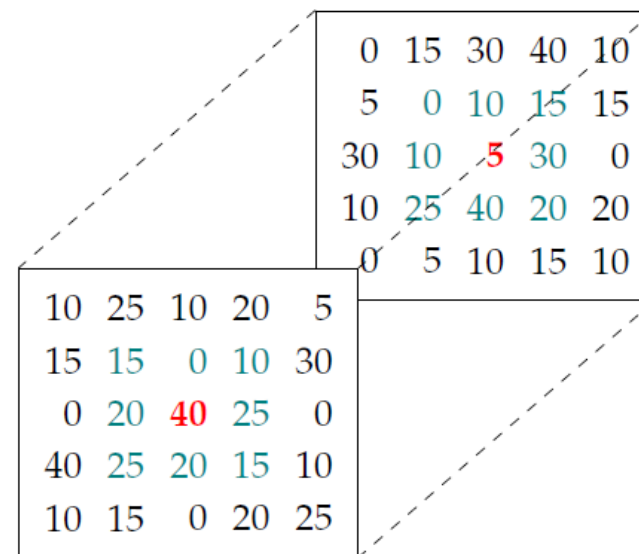
- secvență de matrici de dimensiuni $m \times n$,
fiecare matrice care corespunde unui
anumit moment de timp t și unui
anumit produs meteorologic p
- 240 de matrici zilnice pentru fiecare
produs
- normalizate în preprocesare



0	5
20	15

5	0
10	30

10	30
15	20



0	15	30	40	10
5	0	10	15	15
30	10	5	30	0
10	25	40	20	20
0	5	10	15	10

10	25	10	20	5
15	15	0	10	30
0	20	40	25	0
40	25	20	15	10
10	15	0	20	25

Curățarea datelor

- se înlocuiesc valorile invalide ale lui V într-un anumit punct (i, j) cu media ponderată a valorilor valide ale lui V dintr-o vecinătate de lungime 13 care înconjoară punctul respectiv

0	15	30	40	10
5	-66	10	30	15
30	10	-100	15	0
10	25	-100	20	20
0	5	10	15	10

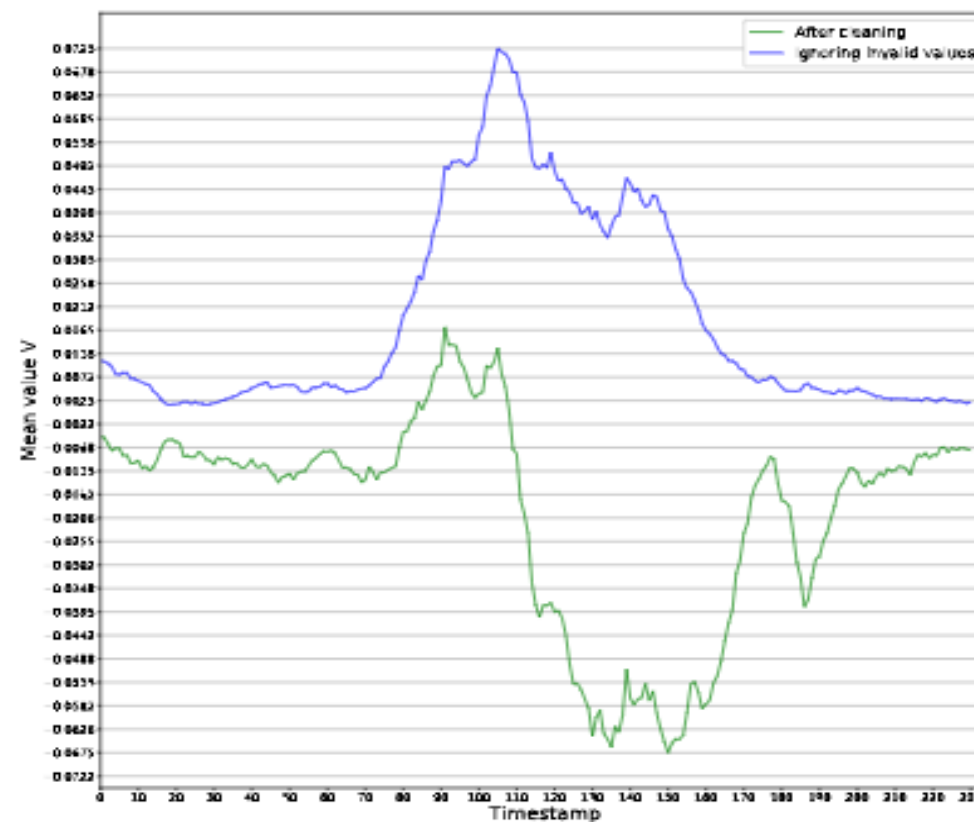
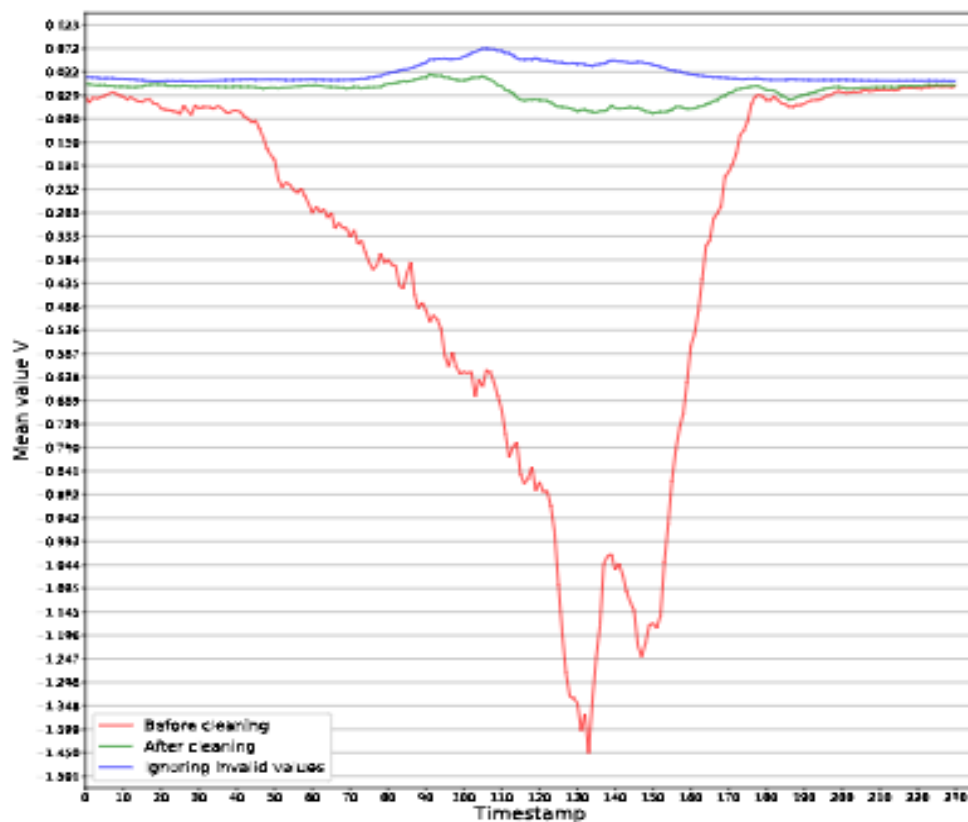


-66	10	30
10	-100	15
25	-100	20

Construirea modelului

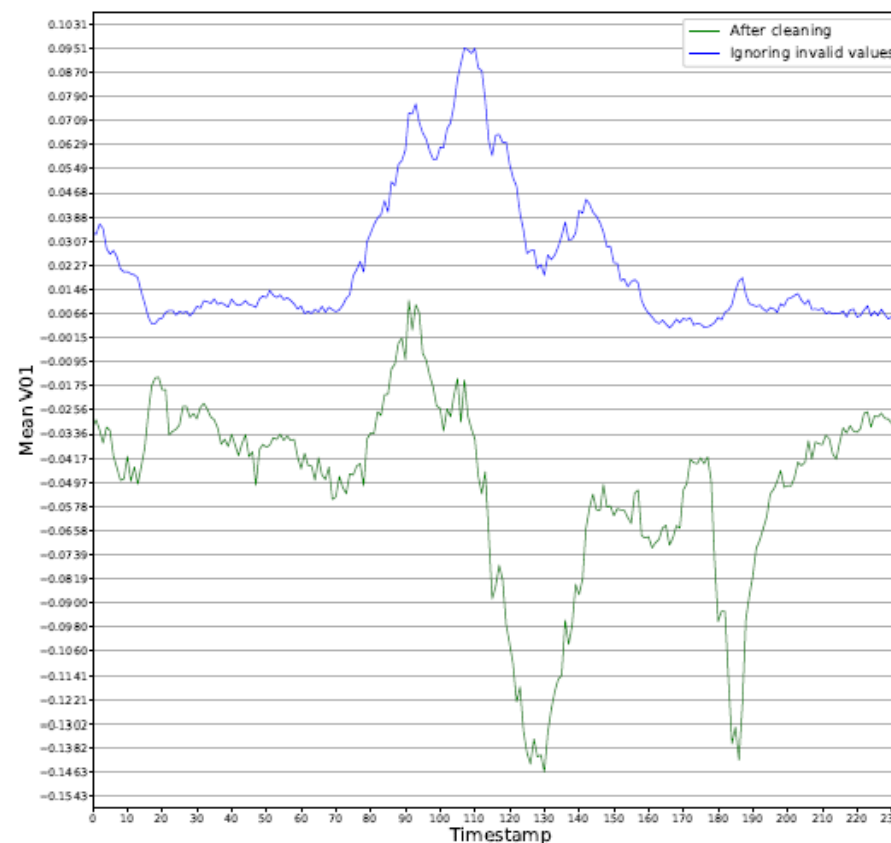
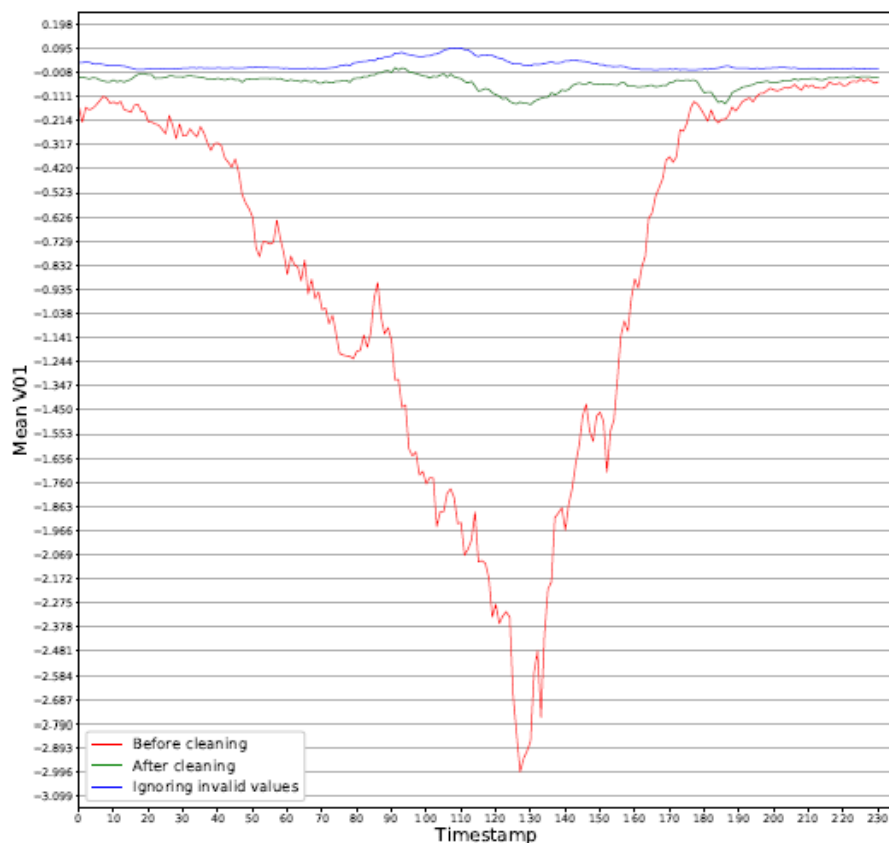
- 12 niveluri ascunse (200 neuroni, 2000 n, 5×500 n și 5×100 n)
- 1 neuron de iesire care conține predicția pentru valoarea produsului
- antrenarea s-a făcut utilizând 30 de epoci și un batch de 1024 de instanțe
- setul de date împărțit în 5 sub-seturi pentru antrenare și testare
- problema biasului

Analiza datelor



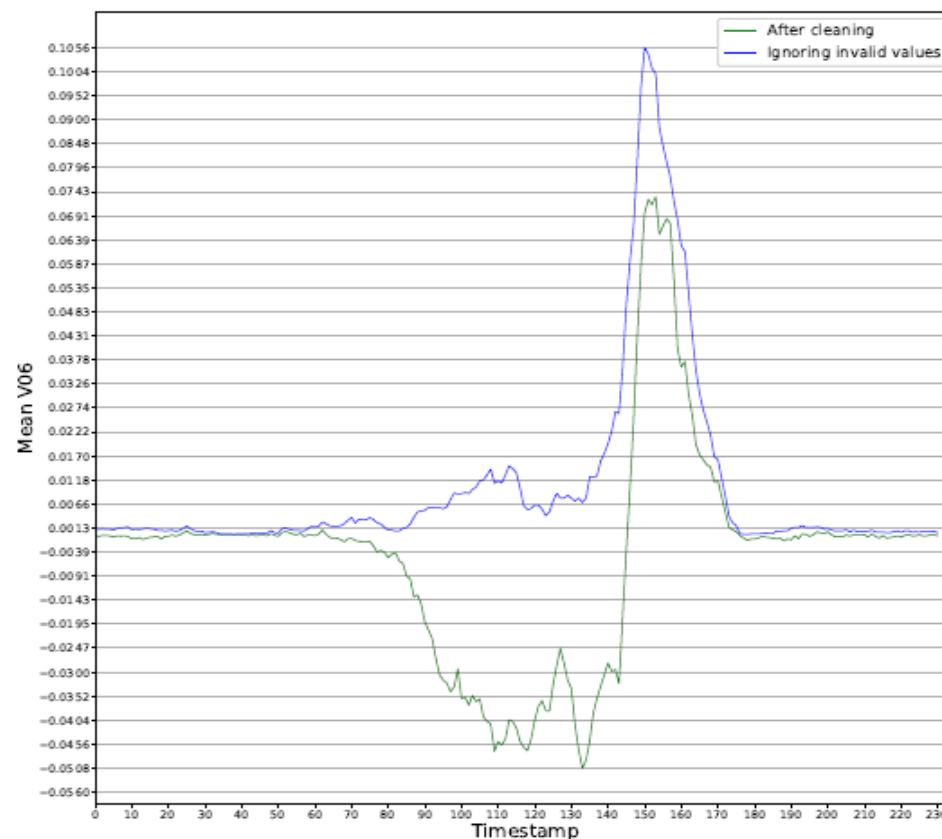
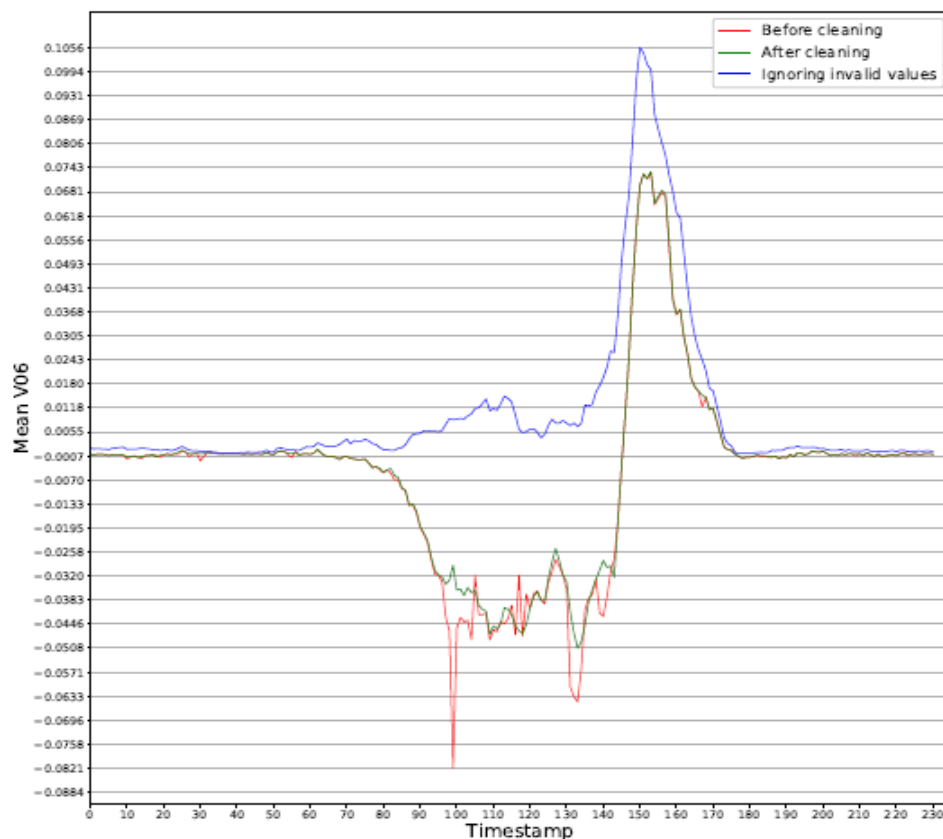
Histograme pentru valorile medii V – înainte și după curățare, ignorând valorile invalide

Analiza datelor



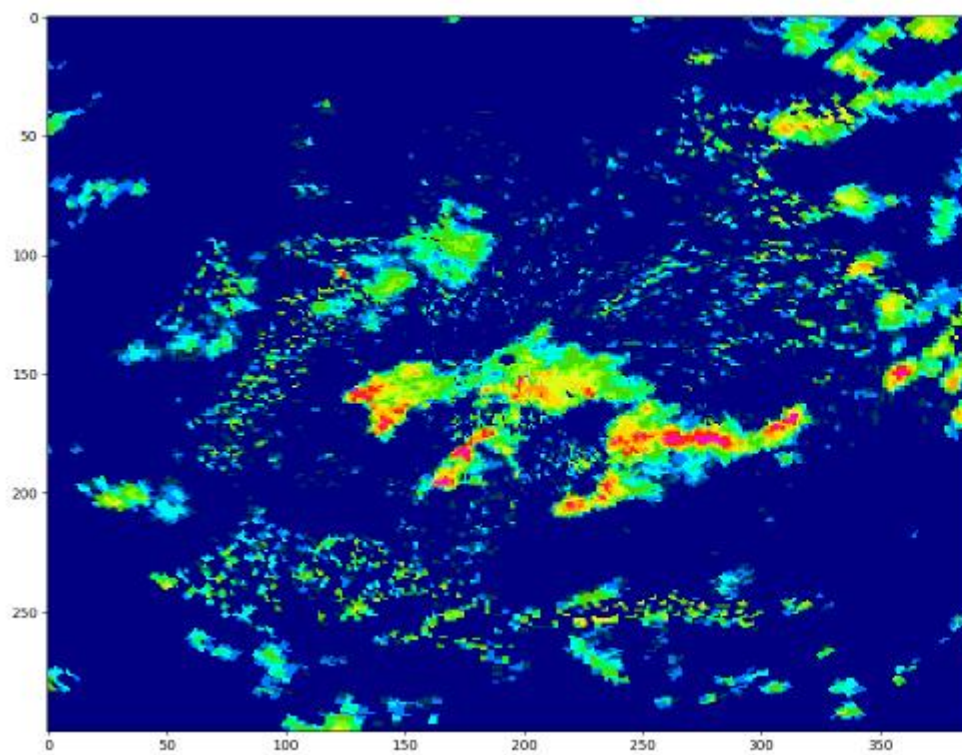
Histograme pentru valorile medii V01 – înainte și după curățare, ignorând valorile invalide

Analiza datelor

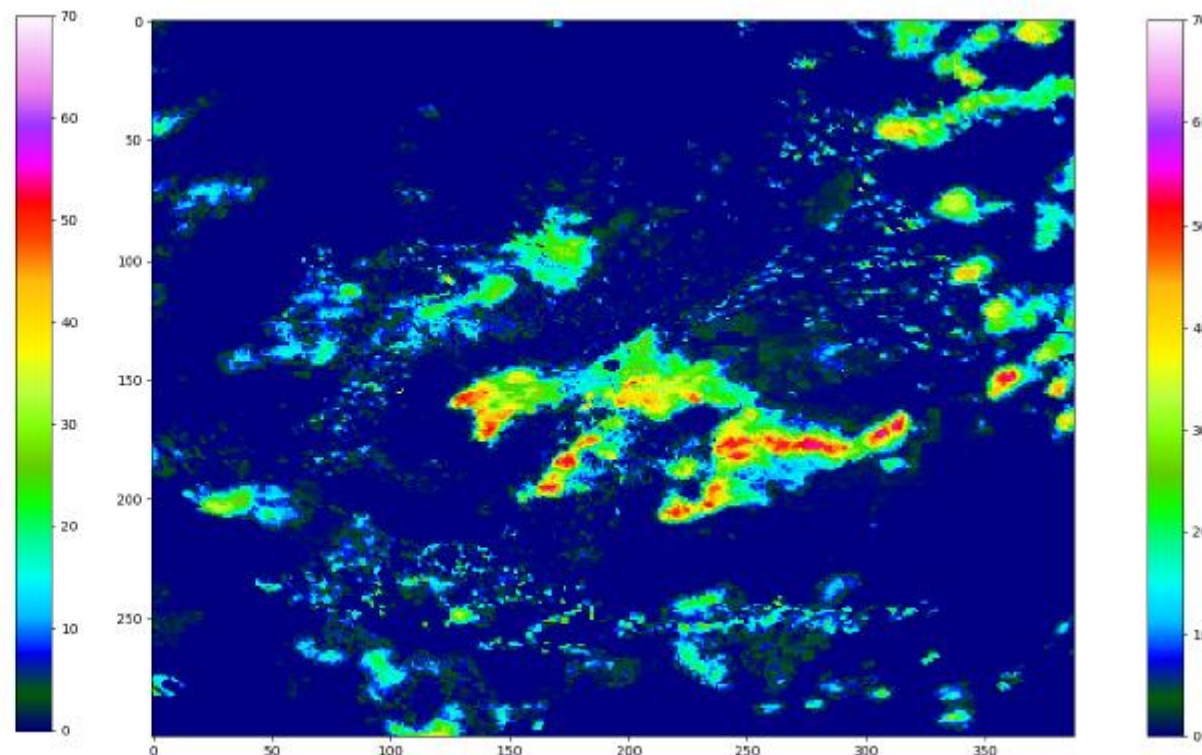


Histograme pentru valorile medii V06 – înainte și după curățare, ignorând valorile invalide

Analiza datelor

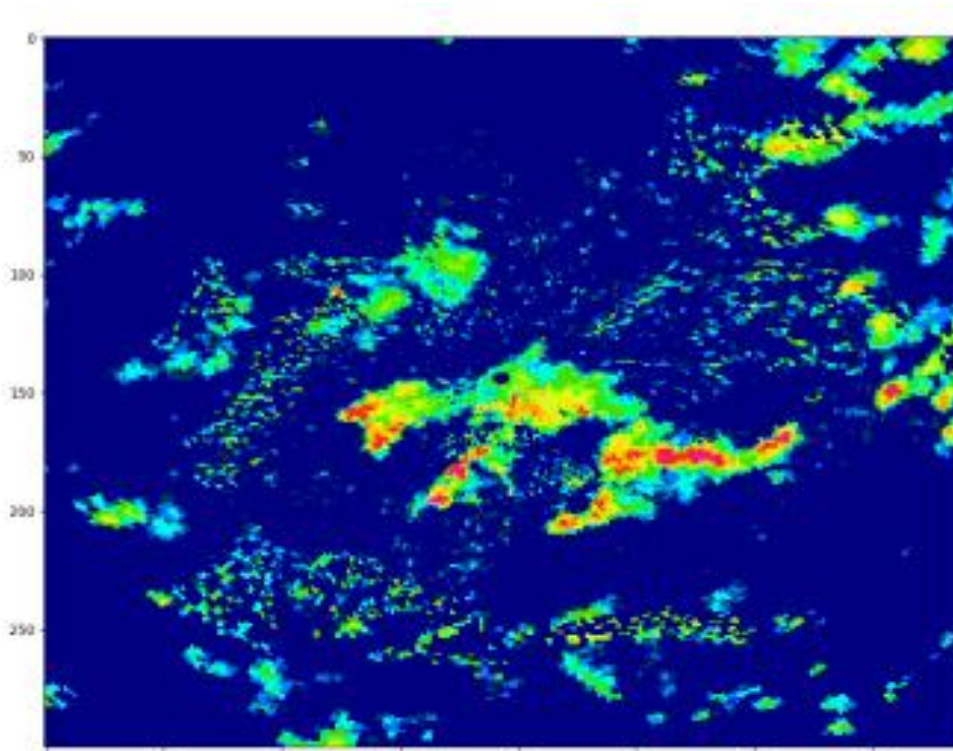


Date radar reale pentru produsul R01 la t+1

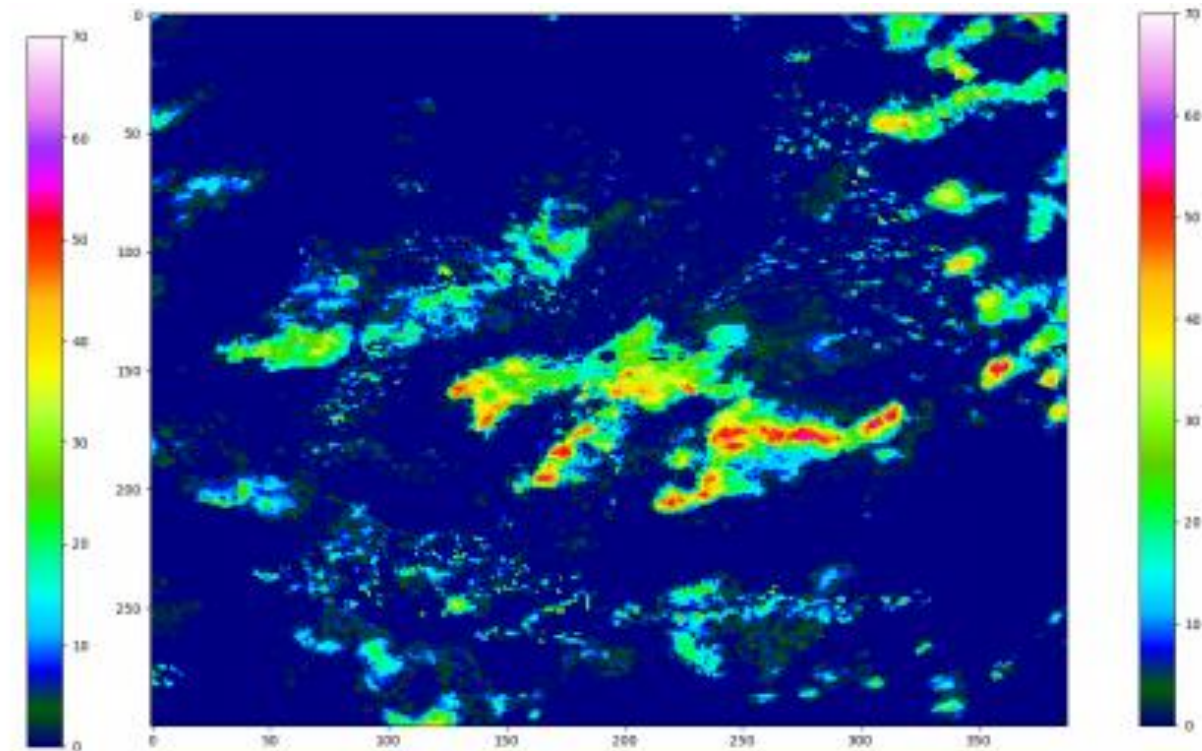


Date radar estimate pentru produsul R01 la t+1

Analiza datelor



Date radar reale pentru produsul R01 la t+1



Date radar estimate pentru produsul R01 la t+1

(date necurățate)

Analiza datelor

Evaluation measure	All 13 products	All R products	All V products	VIL
MAE	0.58 ± 0.02	0.76 ± 0.03	0.41 ± 0.02	0.53 ± 0.02
RMSE	2.25 ± 0.12	2.73 ± 0.17	1.44 ± 0.07	1.62 ± 0.10
NRMSE	$3.27\% \pm 0.17\%$	$3.91\% \pm 0.24\%$	$2.15\% \pm 0.11\%$	$2.32\% \pm 0.14\%$
$MAE_{non-zero}$	4.02 ± 0.12	5.51 ± 0.17	2.73 ± 0.12	2.89 ± 0.04
$RMSE_{non-zero}$	5.93 ± 0.14	7.63 ± 0.15	3.50 ± 0.15	3.9 ± 0.18
$NRMSE_{non-zero}$	$8.60\% \pm 0.21\%$	$10.91\% \pm 0.22\%$	$5.22\% \pm 0.22\%$	$5.63\% \pm 0.26\%$

Evaluation measure	All 13 products	All R products	All V products	VIL	Improvement (%) (cleaning step)
RMSE	4.98 ± 0.06	4.97 ± 0.10	3.99 ± 0.07	5.24 ± 0.17	55%
NRMSE	$7.27\% \pm 0.09\%$	$7.10\% \pm 0.15\%$	$5.95\% \pm 0.10\%$	$7.49\% \pm 0.24\%$	55%
$RMSE_{non-zero}$	10.05 ± 0.40	9.38 ± 0.23	10.88 ± 0.71	9.10 ± 0.31	41%
$NRMSE_{non-zero}$	$14.68\% \pm 0.59\%$	$13.40\% \pm 0.33\%$	$13.24\% \pm 1.06\%$	$13.00\% \pm 0.44\%$	41%

Concluzii

- Demonstrație empirică că, atât în condiții meteorologice normale, cât și în condiții severe, valorile pentru un produs radar la un moment dat, într-o anumită locație, se pot determina în funcție de valorile din locațiile învecinate din momentele de timp anterioare
- NowDeepN va fi extins în continuare prin mărirea setului de date utilizat pentru instruire modelului
- Extinderea caracteristicilor utilizate în procesul de învățare, prin combinarea datelor radar cu alte caracteristici sau date (de exemplu, date satelitare, caracteristici geografice și antropice).