

## **PRIMENA NEURALNIH MREŽA ZA KOMPENZACIJU NELINEARNE DISTORZIJE U MOBILNIM SISTEMIMA NOVE GENERACIJE**

Tamara Muškatirović-Zekić<sup>1,2</sup>, Nataša Nešković<sup>3</sup>, Đurađ Budimir<sup>4</sup>

<sup>1</sup>Univerzitet u Beogradu – Elektrotehnički fakultet, muskatirovic@gmail.com

<sup>2</sup>Regulatorna agencija za elektronske komunikacije i poštanske usluge,  
tamara.muskatirovic@ratel.rs

<sup>3</sup>Univerzitet u Beogradu – Elektrotehnički fakultet, natasha@etf.rs

<sup>4</sup>University of Westminster – Wireless Communications Research Group,  
d.budimir@wmin.ac.uk

**Rezime:** *Kompensacija nelinearne distorzije u mobilnim sistemima nove generacije je od izuzetnog značaja, kako bi se ispunili novi zahtevi korisnika za sve bržim i pouzdanijim servisima i uslugama. Mašinski modeli učenja zasnovani na veštačkim neuralnim mrežama (ANN - Artificial Neural Network) imaju sve rasprostranjeniju primenu u svim oblastima savremenog života, posebno u obradi signala i bežičnim komunikacijama, a u poslednje vreme raste i njihova primena u modelima digitalne predistorzije (DPD) koji se koriste za kompenzaciju nelinearne distorzije signala u mobilnim sistemima. Stoga je u ovom radu izložen pregled postojećih DPD rešenja zasnovanih na primeni neuralnih mreža. Pored opisa samih modela, izvedena je i njihova komparativna analiza sa posebnim osvrtom na kompleksnost implementacije predloženih modela. Takođe su predloženi neki od mogućih problema pri primeni predloženih modela u budućim širokopojasnim mobilnim sistemima.*

**Ključne reči:** *neuralne mreže, pojačavači snage, digitalna predistorzija, MIMO, 5G*

### **1. Uvod**

U savremenom i modernom društvu, potreba za bežičnim povezivanjem pametnih uređaja, poput telefona, tableta, ali i različitih senzora i mašina, je u konstantnom porastu, što dovodi do ubrzanog razvoja mobilnih sistema nove generacije (5G i budući 6G). Osnovni ciljevi mobilnih sistema nove generacije su ultra brz širokopojasni prenos podataka (brzinama do 20 Gb/s), veoma pouzdana komunikacija sa ultra malim kašnjenjem (ispod 10 ms), masovna komunikacija, tj. povezivanje velikog broja uređaja, visoka pouzdanost i efikasna upotreba energije [1]. Kako bi se ispunili ovi zahtevi potrebno je povećati propusni opseg (*bandwidth*) i spektralnu efikasnost, što se postiže korišćenjem milimetarskih talasa (*mmWave*), širokopojasnih (*wideband*)

pojačavača (PA), mMIMO (*massive multiple-input multiple-output*) predajnika i usmeravanjem snopa (*beamforming*).

Pojačavač snage predstavlja glavni potrošač snage u mobilnim sistemima nove generacije i potrebno je da energetski bude što efikasniji [2]. Kako bi se postigla maksimalna energetska efikasnost pojačavači rade u oblasti blizu zasićenja, pri čemu dolazi do pojave nelinearne distorzije signala, kako unutar, tako i van opsega signala. Kako bi se izbeglo narušavanje ukupnih performansi sistema usled nelinearne distorzije, potrebno je što efikasnije linearizovati pojačavač koji radi u oblasti blizu zasićenja.

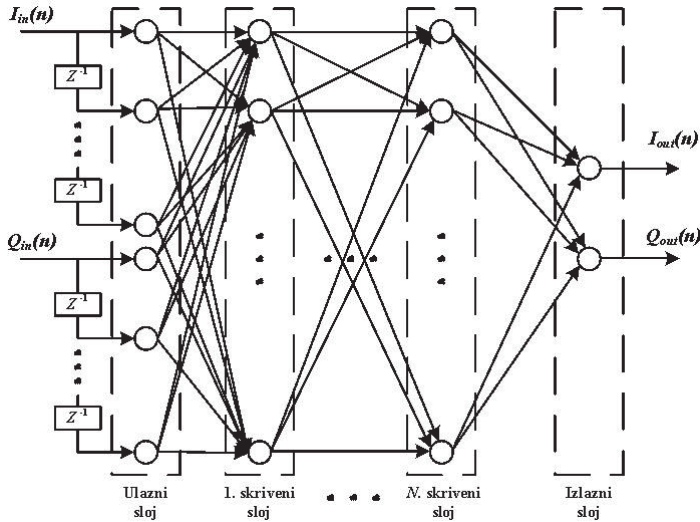
Jedno od mogućih rešenja problema nelinearne distorzije jeste primena digitalne predistorzije (DPD). Osnovni princip digitalne predistorzije je uvođenje dodatnog nelinearnog DPD bloka pre pojačavača, čija je funkcija prenosa inverzna funkciji prenosa pojačavača, tako da se množenjem nelinearne funkcije pojačavača i nelinearne funkcije DPD bloka dobija linearan izlaz. Iako je sam princip prilično jednostavan, razvijanje jeftinog i visoko efikasnog rešenja je kompleksan problem, kojim se bave mnogobrojni istraživački radovi u ovoj oblasti. Da bi se kompenzovala nelinearna distorzija usled širenja spektra, klasičan DPD treba da ima 5 puta veći propusni opseg od širine signala, što sa porastom širine signala na nekoliko GHz postaje skupo i energetski neisplativo rešenje, zbog potrebe za postojanjem konevrtora sa izuzetno velikim brzinama odabiranja. U literaturi su predložena izvesna rešenja za realizaciju širokopojasnog DPD sa ograničenom širinom propusnog opsega [3]. Drugi problem koji se javlja kod mMIMO predajnika jeste potreba za više DPD blokova, obično onoliko koliko ima i pojačavača u predajniku, čime se dodatno poskupljuje sistem. Određena istraživanja su odrađena i na ovom polju, i postoje izvesna rešenja koja efikasno linearizuju više pojačavača u nizu [4]. Međutim i dalje ostaje problem kompleksnosti postojećih rešenja.

Neuralne mreže veoma dobro aproksimiraju nelinearne funkcije, i stoga sve više privlače pažnju istraživača u oblasti linearizacije pojačavača. U poređenju sa konvencionalnim DPD modelima zasnovanim na polinomima, poput *memory polynomial* (MP), *parallel Hammerstein* (PH), *Volterra* i drugima, DPD modeli realizovani primenom neuralnih mreža, značajno poboljšavaju performanse modelovanja u smislu nelinearne procene izlaznog signala pojačavača. Osnovna podela neuralnih mreža je na "plitke" (*shallow neural network* - SNN) i "duboke" (*deep neural network* - DNN), u zavisnosti od broja skrivenih slojeva neuralne mreže.

Opisi predloženih modela neuralnih mreža dati su u drugom poglavlju, poređenje DPD modela zasnovanih na opisanim neuralnim mrežama dato je u trećem poglavlju, dok je u četvrtom poglavlju dat zaključak.

## 2. Neuralne mreže

Arhitektura tipične neuralne mreže sastoji se od više nivoa, jednog ulaznog sloja, jednog ili više skrivenih slojeva i jednog izlaznog sloja, kao što je prikazano na slici 1. U zavisnosti od broja skrivenih slojeva neuralne mreže delimo na „plitke“ neuralne mreže (SNN - *Shallow Neural Network*), koje imaju jedan ili dva skrivena sloja, i „duboke“ neuralne mreže (DNN - *Deep Neural Network*), koje imaju tri ili više skrivenih slojeva. U ovom radu su analizirane RVTDDNN (*Real-Valued Time-Delay Neural Network*) i ARVTDDNN (*Augmented Real-Valued Time-Delay Neural Network*), koje spadaju u SNN mreže, kao i LSTM (*Long Short-Term Memory*) i RVTDCNN (*Real-Valued Time-Delay Convolutional Neural Network*), koje spadaju u DNN mreže.



Slika 1. Arhitektura NN mreže.

Ovi modeli prate tzv. propagaciju unapred (*feedforward*), tj. podaci putuju od neurona nižeg sloja ka neuronima višeg sloja, pri čemu se ulazni podaci u  $(l+1)$ -i sloj računaju po sledećoj formuli [5]:

$$net_j^{l+1} = \sum_{i=1}^q \omega_{ji}^{l+1} \sigma_i^l + b_j^{l+1} \quad (1)$$

gde:  $b_j^{l+1}$  označava pomeraj  $j$ -tog neurona u  $(l+1)$ -om sloju,  $q$  označava ukupan broj neurona u prethodnom sloju,  $\omega_{ji}^{l+1}$  označava sinaptičku težinu između  $i$ -tog ulaza prethodnog sloja i  $j$ -tog neurona trenutnog sloja,  $\sigma_i^l$  označava izlaz  $j$ -tog neurona u  $(l+1)$ -om sloju i zavisi od aktivacione funkcije  $f$ , tj.  $\sigma_i^l = f(net_j^{l+1})$ . Najčešće korišćene aktivacione funkcije su tangens hiperbolički (*tanh*), sigmoid (*logsig*) i ispravljena linearna f-ja (*ReLU*), i date su sledećim matematičkim izrazima:

$$\tanh(x) = \frac{\exp(2x)-1}{\exp(2x)+1} \quad (2)$$

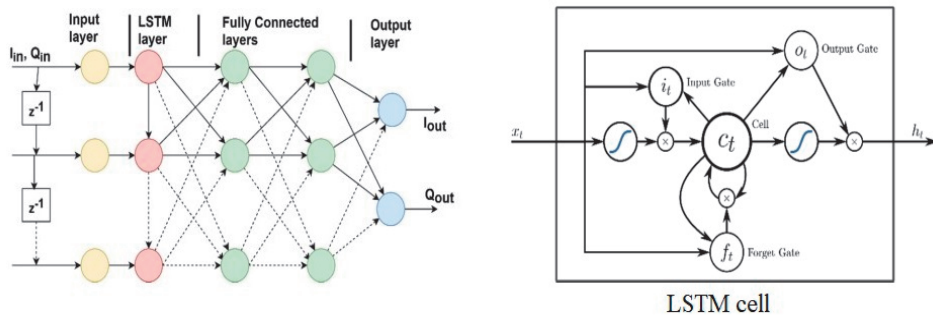
$$\log sig(x) = \frac{1}{1+e^{-x}} \quad (3)$$

$$ReLU = \max(0, x) \quad (4)$$

gde je  $x$  ulazni signal.

Ulazni signali u RVTDDN mrežu su  $I$  (*inphase*) i  $Q$  (*quadrature*) komponente ulaznog signala. Dok su ulazni signali u ARVTDDN mrežu, pored  $I$  i  $Q$  komponenti i  $|x(n)|$ ,  $|x(n)|^2$  i  $|x(n)|^3$ , gde je  $|x(n)|$  amplituda ulaznog signala [6].

LSTM neuralna mreža prikazana je na slici 2 [7]. Sastoji se od ulaznog sloja, LSTM sloja, više skrivenih slojeva i izlaznog sloja. Ulazne signale čine trenutni i zakašnjeni odbirci ulaznog signala. LSTM sloj sadrži ulaznu kapiju (*unput gate*), izlaznu kapiju (*output gate*) i tzv. kapiju „za zaborav“ (*forget gate*). Zahvaljujući ovakvoj svojoj konstrukciji LSTM mreže mogu da zapamte težinske koeficijente u toku dužeg vremenskog perioda, čime se postiže bolja predikcija vremenskog niza signala koji zavisi od prethodnih stanja, tj. u slučaju predikcije signala pojačavača koji ispoljavaju memorijske efekte.



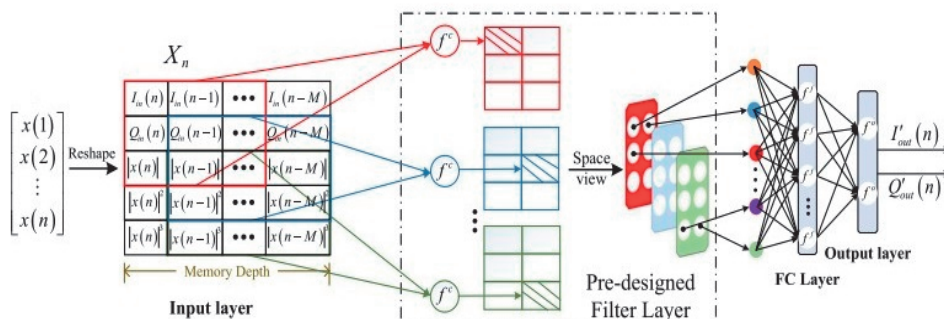
Slika 2. Arhitektura LSTM DNN mreže [7].

Na slici 3 prikazana je arhitektura RVTDCNN mreže [8]. Ovaj model se sastoji od četiri sloja: ulaznog sloja, predizajniranog filterskog sloja, potpuno povezanog skrivenog sloja i izlaznog sloja. Filterski sloj je konstruisan korišćenjem konvolucionih blokova, pomoću kojih se izdvajaju bitne karakteristike ulaznog signala. Dimenzije konvolucionog bloka se dizajniraju tako da se smanjuje računarska kompleksnost, dok se istovremeno zadržavaju dobre performanse predviđanja modela.

Treba napomenuti da su ulazni podaci u konvolucione mreže obično u 2D formatu, dok su podaci koji se koriste za modelovanje pojačavača i digitalnu predistorziju 1D, tj. zavise samo od vremena. Stoga je potrebno pažljivo mapirati ulazne 1D podatke u 2D podatke pogodne za konvolucionu obradu. U analiziranoj RVTDCNN mreži ulazni podaci koji odgovaraju susednim zakašnjenim signalima su raspoređeni jedan do drugog kako bi se obezbedilo da 2-D konvolucioni blok izdvaja unakrsne članove različito zakašnjenih signala, a mapiranje je izvršeno formiranjem 2D matrice, prema sledećoj formuli:

$$X_n = \begin{bmatrix} I_{in}(n), I_{in}(n-1), \dots, I_{in}(n-M); \\ Q_{in}(n), Q_{in}(n-1), \dots, Q_{in}(n-M); \\ |x(n)|, |x(n-1)|, \dots, |x(n-M)|; \\ |x(n)|^2, |x(n-1)|^2, \dots, |x(n-M)|^2; \\ |x(n)|^3, |x(n-1)|^3, \dots, |x(n-M)|^3 \end{bmatrix} \quad (5)$$

gde:  $(n-i)$ ,  $i=1,2,\dots,M$ , označava  $i$ -ti zakašnjeni odbirak, a  $M$  dubinu memorije.



Slika 3. Arhitektura RVTDCNN mreže [8].

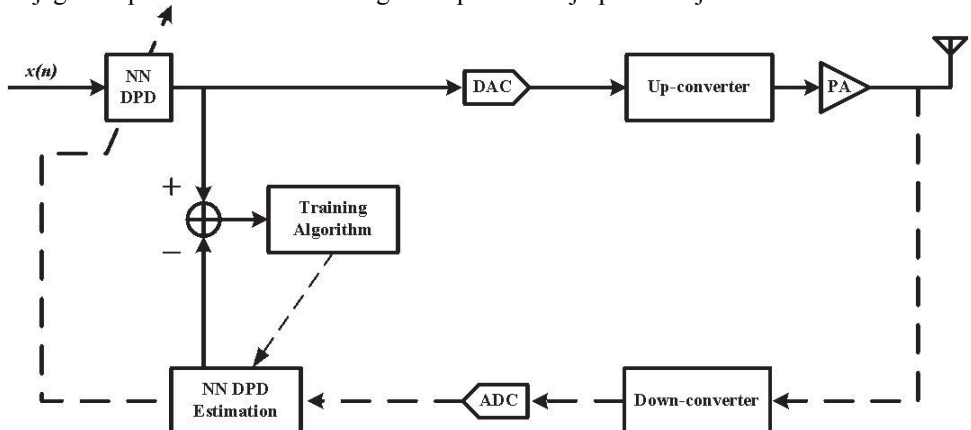
Struktura različitih modela neuralnih mreža predloženih u literaturi [6]- [10], data je u tabeli 1. Određivanje optimalnog broja neurona u skrivenom sloju, zatim broja neurona u LSTM sloju kod LSTM mreže, kao i veličine i broja konvolucionih blokova kod RVTDCNN mreže se vrši eksperimentalno u toku treniranja mreže, tako da se postignu najbolje moguće performanse same mreže. Tokom treniranja NN mreže koristi se postupak unakrsne validacije (*cross-validation*), koji se sastoji od tri faze: faze treniranja, faze validacije i faze testiranja. Prve dve faze, treniranje i validacija, vrše se korišćenjem različitih skupova izmerenih ulaznih i izlaznih signala, dok se za testiranje NN mreže koristi treći skup ulazno/izlaznih podataka, različit od prethodna dva. Faza testiranja se vrši u cilju proveravanja sposobnosti predloženog modela neuralne mreže da isprati izmerene podatke.

Tabela 1. Struktura različitih NN mreža

| Tip NN mreže | Ulazni podaci                      | Broj neurona u skrivenom sloju       | Aktivaciona funkcija |
|--------------|------------------------------------|--------------------------------------|----------------------|
| RVTDNN [9]   | $I, Q$                             | 35                                   | $\tanh$              |
| ARVTDNN [6]  | $I, Q,  x(n) ,  x(n) ^2,  x(n) ^3$ | 17                                   | $\tanh$              |
| DNN [10]     | $I, Q$                             | [17 17 17]                           | $\text{logsig}$      |
| LSTM [7]     | $I, Q$                             | 10 u LSTM sloju<br>[7 5] u skrivenim | $\text{ReLU}$        |
| RVTDCNN [8]  | $I, Q,  x(n) ,  x(n) ^2,  x(n) ^3$ | 6                                    | $\tanh$              |

### 3. DPD modeli zasnovani na neuralnim mrežama

Za formiranje DPD modela zasnovanih na neuralnim mrežama (NN DPD modela) korišćena je arhitektura indirektnog učenja (ILA - *Indirect Learning Architecture*), kod koje se prvo određuju koeficijenti modela, tj. formira se i trenira neuralna mreža na osnovu ulaznih i izlaznih signala pojačavača. Zatim se tako formirana neuralna mreža koristi ispred pojačavača, kako bi se izvršila predistorzija signala. Dijagram opisane arhitekture za digitalnu predistorziju prikazan je na slici 4.



Slika 4. Dijagram DPD arhitekture primenom neuralne mreže.

Pri izboru tipa neuralne mreže za DPD, veoma bitan parametar je kompleksnost modela, koja se može uporediti na osnovu broja koeficijenata modela  $N_{coef}$ , kao i na osnovu broja FLOP-ova (*floating point operations*). Broj koeficijenata modela je jednak zbiru koeficijanata u svakom sloju NN modela, i može se izračunati korišćenjem sledeće formule:

$$N_{coef} = (N_i + 1)N_1 + \sum_{f=2}^F (N_{f-1} + 1)N_f + (N_F + 1) N_o \quad (6)$$

gde je:  $N_i$  broj neurona u ulaznom sloju,  $F$  broj skrivenih slojeva,  $N_f$  broj neurona u  $f$ -tom skrivenom sloju i  $N_o$  broj neurona u izlaznom sloju.

Broj FLOP-ova zavisi od korišćene aktivacione funkcije, tj. od broja i vrste korišćenih aritmetičkih operacija [11]. Za NN model koji koristi *tanh* aktivacionu funkciju, broj FLOP-ova se može izračunati korišćenjem sledeće formule:

$$FLOP = 2N_iN_1 + 2N_FN_o + 2 \sum_{f=2}^F N_{f-1}N_f + 15 \sum_{f=1}^F N_f \quad (7)$$

Za NN model koji koristi *logsig* aktivacionu funkciju, broj FLOP-ova se može izračunati korišćenjem sledeće formule:

$$FLOP = 2N_iN_1 + 2N_FN_o + 2 \sum_{f=2}^F N_{f-1}N_f + 13 \sum_{f=1}^F N_f \quad (8)$$

Za procenu performansi NN DPD modela može se koristiti normalizovana srednja kvadratna greška NMSE (*Normalized Mean-Squared Error*), koja se računa po sledećoj formuli:

$$NMSE = 10 \log_{10} \left( \frac{\sum_{n=1}^K |y_{meas}(n) - y_{est}(n)|^2}{\sum_{n=1}^K |y_{meas}(n)|^2} \right) \quad (9)$$

gde su:  $y_{meas}$  i  $y_{est}$  - izmereni i procenjeni talasni oblici izlaznog signala, respektivno.

U tabeli 2 je dat pregled uporednih karakteristika različitih analiziranih NN DPD modela. Date su izračunate vrednosti broja koeficijenata i broja FLOP-ova. Takođe su date i eksperimentalno dobijene vrednosti za NMSE iz [8], gde je izvršeno poređenje modela na LTE signalu širine 100 MHz. Korišćen je *Doherty* pojačavač snage saturacije 43 dBm, na centralnoj frekvenciji 2.14 GHz.

Tabela 2. Performanse i kompleksnost različitih NN DPD modela

| Tip NN mreže | Broj koeficijenata | Broj FLOP-ova | NMSE [dB] |
|--------------|--------------------|---------------|-----------|
| RVTDNN [9]   | 387                | 1155          | -35.09    |
| ARVTDNN [6]  | 393                | 1008          | -36.47    |
| DNN [10]     | 801                | 2306          | -36.42    |
| LSTM [7]     | 467                | 5034          | -36.27    |
| RVTDCNN [8]  | 158                | 876           | -36.44    |

Može se primetiti da svi predloženi NN DPD modeli, veoma dobro kompenzuju nelinearnu distorziju pojačavača, dok se u pogledu kompleksnosti modela, na osnovu broja koeficijenata i FLOP-ova, izdvaja DPD model zasnovan na RVTDCNN mreži. Važno je napomenuti da sa povećanjem propusnog opsega signala, raste i dubina memorije, što rezultira potrebom za kompleksnijom mrežnom strukturom, koja će biti u stanju da isprati memorijske efekte pojačavača.

#### 4. Zaključak

Na osnovu izloženog, može se zaključiti da se neuralne mreže mogu efikasno primeniti za kompenzaciju nelinearne distorzije u mobilnim sistemima nove generacije. Njihova osnovna prednost u odnosu na klasične modele za digitalni predistorziju zasnovane na polinomima, jeste u smanjenoj kompleksnosti, kao i u njihovoj adaptibilnosti, tj. mogućnosti za lakim i brzim ponovnim treniranjem mreže u slučaju bilo kakvih promena okruženja. Doprinos ovog rada je u kompaktnom prikazu i analizi postojećih modela za digitalnu predistorziju signala zasnovanih na neuralnim mrežama. Budući rad ide u pravcu ispitivanja mogućnosti optimizacije postojećih NN DPD modela, kao i mogućnosti primene neke druge neuralne mreže za kompenzaciju nelinearne distorzije u mobilnim sistemima nove generacije.

#### Literatura

- [1] Harri Holma, Antti Toskala, Takehiro Nakamura, *5G Technology: 3GPP New Radio*, JohnWiley & Sons Ltd. 2020.
- [2] L. Guan and A. Zhu, "Green Communications: Digital Predistortion for Wideband RF Power Amplifiers," in *IEEE Microwave Magazine*, vol. 15, no. 7, pp. 84-99, November-December 2014, DOI: 10.1109/MMM.2014.2356037
- [3] T. Muškatirović-Zekić, M. Čabarkapa, N. Nešković, Đ. Budimir, „Pregled postojećih DPD modela sa ograničenom širinom propusnog opsega”, 65. Konferencija za elektroniku, telekomunikacije, računarstvo, automatiku i nuklearnu tehniku ETRAN 2021, pp. 604-609.
- [4] T. Muškatirović-Zekić, M. Čabarkapa, N. Nešković, Đ. Budimir, „An Overview of Digital Predistortion Solutions for Hybrid Beamforming Mimo Transmitters”, 29th Telecommunications Forum (TELFOR), 2021, pp. 1-4, DOI: 10.1109/TELFOR52709.2021.9653422
- [5] P. Jaraut, M. Helaloui, W. Chen, M. Rawat, N. Boulejfen and F. M. Ghannouchi, "Review of the Neural Network based Digital Predistortion Linearization of Multi-Band/MIMO Transmitters," 2021 IEEE MTT-S International Wireless Symposium (IWS), 2021, pp. 1-3, DOI: 10.1109/IWS52775.2021.9499466.
- [6] D. Wang, M. Aziz, M. Helaloui and F. M. Ghannouchi, "Augmented Real-Valued Time-Delay Neural Network for Compensation of Distortions and Impairments in Wireless Transmitters," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 242-254, Jan. 2019, DOI: 10.1109/TNNLS.2018.2838039.
- [7] D. Phartiyal and M. Rawat, "LSTM-Deep Neural Networks based Predistortion Linearizer for High Power Amplifiers," 2019 National Conference on Communications (NCC), 2019, pp. 1-5, DOI: 10.1109/NCC.2019.8732178.
- [8] X. Hu et al., "Convolutional Neural Network for Behavioral Modeling and Predistortion of Wideband Power Amplifiers," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 8, pp. 3923-3937, August 2022, DOI: 10.1109/TNNLS.2021.3054867.
- [9] Taijun Liu, S. Boumaiza and F. M. Ghannouchi, "Dynamic behavioral modeling of 3G power amplifiers using real-valued time-delay neural networks," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 3, pp. 1025-1033, March 2004, DOI: 10.1109/TMTT.2004.823583.

- [10] R. Hongyo, Y. Egashira, T. M. Hone and K. Yamaguchi, "Deep Neural Network-Based Digital Predistorter for Doherty Power Amplifiers," in *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 2, pp. 146-148, Februar 2019, DOI: 10.1109/LMWC.2018.2888955.
- [11] A.S. Tehrani, C. Haiying, S. Afsardoost, T. Eriksson, M. Isaksson and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 6, pp. 1510-1520, June 2010, DOI: 10.1109/TMTT.2010.2047920.

**Abstract:** *The compensation of nonlinear distortion in mobile systems of the new generation is extremely important, in order to meet the new demands of users for faster and more reliable services. Machine learning models based on artificial neural networks (ANN - Artificial Neural Network) are increasingly used in all areas of modern life, especially in signal processing and wireless communications, and recently their application in digital predistortion (DPD) models that are used for compensation of nonlinear signal distortion in mobile systems is growing. Therefore, this paper presents an overview of existing DPD solutions based on the application of neural networks. In addition to the description of the models themselves, their comparative analysis is performed with special attention to the complexity of the implementation of the proposed models. Also, some of the possible problems when applying the proposed models in future broadband mobile systems are pointed out.*

**Keywords:** *neural networks, power amplifiers, digital predistortion, MIMO, 5G*

## **APPLICATION OF NEURAL NETWORKS FOR COMPENSATION OF NONLINEAR DISTORTION IN NEW GENERATION MOBILE SYSTEMS**

Tamara Muškatirović-Zekić, Nataša Nešković, Đurađ Budimir