

Unapređenje predikcije snage na izlazu hidroelektrane kroz klaster analizu podataka i kombinovani model veštačkih neuronskih mreža

Stefan Čubonović*, Aleksandar Ranković*, Marko Krstić**

* Fakultet tehničkih nauka u Čačku, Univerziteta u Kragujevcu

** Javno preduzeće Elektroprivreda Srbije

Rezime - Ovo istraživanje nastavlja prethodne napore uložene u primeni veštačkih neuronskih mreža za predikciju snage na izlazu hidroelektrane, sa ciljem daljeg unapređenja preciznosti i robusnosti modela. U radu je fokus na implementaciji i evaluaciji kombinovanog modela koji koristi klaster analizu podataka o protoku vode kroz turbinu sa zadatim pragom i dva modela veštačkih neuronskih mreža: veštačku neuronsku mrežu sa jednosmernim prostiranjem signala (FF-ANN - *feedforward artificial neural network*) i rekurentnu neuronsku mrežu (RNN - *recurrent neural network*). Klaster analiza podataka sa zadatim pragom, koji je definisan na osnovu stabilnosti protoka, primenjena je za odabir jednog od dva predložena modela veštačkih neuronskih mreža. Na taj način osigurano je poverenje u preciznost modela i identifikovali specifični operativni uslovi hidroelektrane. Analiza obuhvata podatke prikupljene tokom različitih hidroloških ciklusa u jednoj godini, testirajući prilagodljivost predloženih modela. Rezultati ovog istraživanja obećavaju značajna poboljšanja efikasnosti prethodno kreiranog modela za predikciju snage hidroelektrane.

Ključne reči - veštačka neuronska mreža, FF-ANN, RNN, hidroelektrana, klaster analiza, protok

1 UVOD

Precizna predikcija snage [1, 2] na izlazu hidroelektrane (HE) ne samo da predstavlja tehnološki izazov već je i važan faktor za optimizaciju resursa i minimizaciju negativnih ekoloških posledica [3]. S obzirom na to, razvoj unapređenih modela veštačkih neuronskih mreža (ANN - eng. *artificial neural network*) za predikciju snage pored naučnog doprinosa predstavlja i praktičan alat koji doprinosi efikasnijem korišćenju hidroenergetskih potencijala.

Inspirisan idejom o optimalnom modelu u radu [1], ovaj rad predstavlja evoluciju u predikciji snage na izlazu HE. U radu [1] su predstavljena dva modela veštačkih neuronskih mreža: veštačka neuronska mreža sa jednosmernim prostiranjem signala (FF-ANN - eng. *feedforward artificial neural network*) [4] i rekurentna neuronska mreža (RNN - eng. *recurrent neural network*) [5]. Rezultati su pokazali da FF-ANN model preciznije predviđa snagu HE pri naglim i brzim promenama protoka vode kroz turbinu. Ovaj model je efikasan u situacijama gde trenutni podaci iz ulaznog sloja direktno utiču na izlaznu vrednost, bez uzimanja u obzir prethodnih stanja sistema. Nasuprot tome, RNN model se bolje pokazao pri postepenim i sporim promenama protoka, pružajući preciznija rešenja zahvaljujući svojoj

sposobnosti da analiziraju dugoročne trendove u podacima.

Uzimajući u obzir promenljivu prirodu protoka, predlaže se korišćenje kombinovanog modela koji u sebi integriše FF-ANN i RNN. Ovaj pristup uključuje klaster analizu podataka o protoku vode kroz turbinu. Odnosno, zadavanjem određenog praga za razdvajanje perioda protoka na one koji su spori i postepeni i na one koji su brze i nagle prirode, osigurava se poverenje u preciznost u svim operativnim uslovima.

Takođe, proširenje perioda validacije i testiranja na vremenski interval od jedne godine će dodatno doprineti preciznosti i prilagodljivosti primene kombinovanog modela u stvarnim operativnim uslovima tokom različitih sezona.

II KLASER ANALIZA PODATAKA O PROTOKU: POBOLJŠANJE KOMBINOVANOG MODELA

U procesu optimizacije modela za predikciju snage na izlazu HE, ključni element predstavlja sposobnost modela da tačno identifikuje i razlikuje periode stabilnog i nestabilnog protoka vode kroz turbinu. Efikasna klaster analiza podataka o protoku vode omogućava bolju adaptaciju ANN različitim operativnim uslovima u kojima se može naći model za predikciju snage. Ovakav pristup se zasniva na matematički definisanom pragu varijanse, koji omogućava jasno kategorisanje podataka.

Proces klaster analize [6] započinje definisanjem intervala analize, pri čemu svaki od njih broji fiksno 100 uzastopnih merenja. Za svaki interval, izračunava se varijansa kako bi se ocenila njegova stabilnost. Varijansa σ^2 za interval sa n merenja definisana je jednačinom (1):

$$\sigma^2 = \frac{1}{(n-1)} \sum_{i=1}^n (Q_i - \mu)^2, \quad (1)$$

gde Q_i predstavlja pojedinačne vrednosti protoka unutar intervala, a μ je aritmetička sredina tih vrednosti koja se računa primenom jednačine (2).

$$\mu = \frac{1}{n} \sum_{i=1}^n Q_i. \quad (2)$$

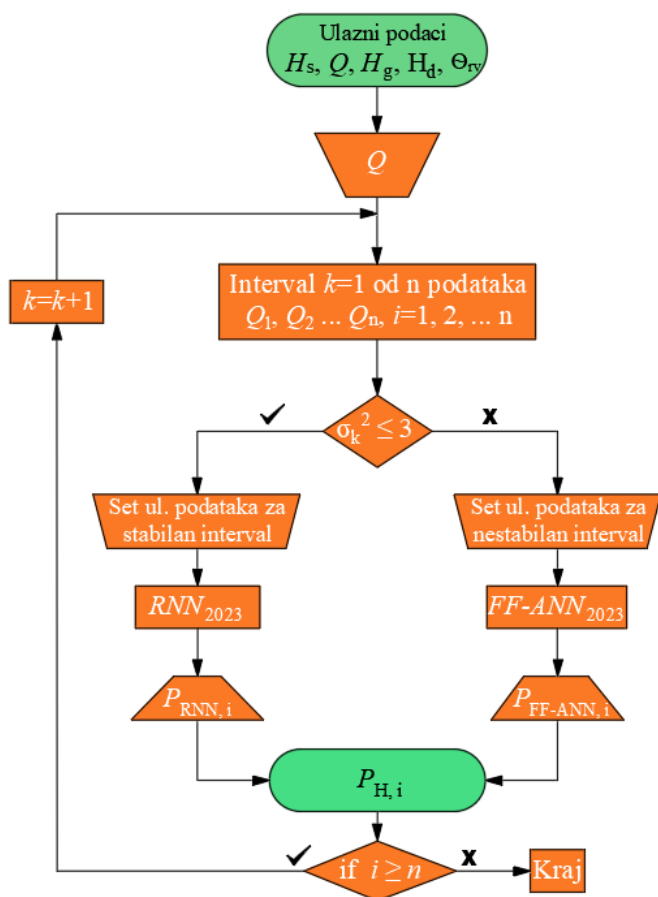
Prag varijanse θ je unapred određen na osnovu analize podataka i iterativnog testiranja. U ovom istraživanju, prag od 3 se pokazao kao optimalan za razlikovanje stabilnih i nestabilnih stanja protoka. Interval T se klasifikuje kao stabilan ako je njegova varijansa σ^2 manja ili jednaka od praga θ i kao nestabilan ako je veća:

$$T = \begin{cases} \text{stabilan}, & \sigma^2 \leq \theta \\ \text{nestabilan}, & \sigma^2 > \theta \end{cases} \quad (3)$$

Kada izdvojimo stabilne i nestabilne intervale, sledi predikcija snage na izlazu HE. Za stabilne intervale koristimo RNN, dok za nestabilne koristimo FF-ANN. Kombinovani model efikasno „pamti“ trenutke svakog predviđanja, čuvajući informacije o tome kada i po kom redosledu su određene snage predviđene. Ove informacije se potom „vraćaju“ u trenutke kada je potrebno izvršiti konačnu predikciju, omogućavajući da se kombinuju predikcije iz pojedinačnih mreža u zavisnosti od njihove relevantnosti u datim uslovima. Na ovaj način, rezultujuća snaga koja se dobija primenom kombinovanog modela ne samo da se efikasno prilagođava trenutnim promenama već i koristi podatke o prošlim stanjima, čime se postiže veća preciznost i pouzdanost krajnjeg rezultata.

III DIJAGRAM TOKA KOMBINOVANOG MODELA ZA PREDIKCIJU SNAGE

Kombinovani model koji integriše FF-ANN i RNN je dizajniran kako bi se unapredila efikasnost i preciznost u predikciji snage na izlazu HE. FF-ANN je izuzetno efikasan u situacijama koje zahtevaju brzo reagovanje na iznenadne promene, dok RNN pruža bolje rezultate u analizi podataka koji pokazuju postepene promene, kako se pokazalo u prethodnom istraživanju [1]. Dijagram toka koji opisuje način rada kombinovanog modela prikazan je na Slici 1.



Slika 1. Dijagram toka kombinovanog modela ANN za predikciju snage na izlazu HE

Kao što je opisano u prethodnom poglavlju, u ovom istraživanju, primenili smo klaster analizu podataka sa zadatim pragom za analizu podataka o protoku vode kroz turbinu, sa ciljem identifikacije stabilnih i nestabilnih intervala. Ovaj pristup nam je omogućio da efikasno grupišemo periode na osnovu varijabilnosti protoka, što je ključno za rad kombinovanog modela. Klaster analiza podataka započinje segmentacijom kontinuiranih podataka o protoku vode u fiksne intervale od 100 merenja (može se desiti da poslednji interval ima manje merenja). Na ovaj način obezbeđujemo uniformnost u analizi, dajući svakom segmentu jednaku težinu u daljem procesu. Nakon segmentacije, izračunavamo standardnu devijaciju za svaki interval i na osnovu toga vršimo procenu stabilnosti protoka unutar svakog segmenta.

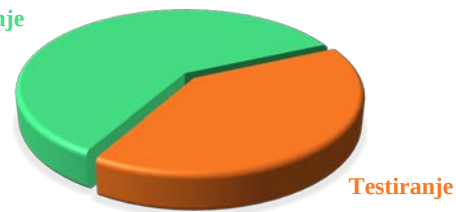
IV PRIKUPLJANJE I PRIPREMA PODATAKA

Ulazni i izlazni sloj su ključni za modelovanje i analizu sistema primenom oba tipa ANN. Ovi slojevi čine određene veličine koje se mogu dobiti iz različitih izvora, u ovom slučaju su to samo podaci dobijeni sa senzora na samoj HE. Veličine u ulaznom sloju se koriste za treniranje neuronske mreže i čine osnovu za njen rad. Za kreiranje i validaciju neuronske mreže u radu [1] korišćen je skup od 60% podataka iz aprila 2023. godine (Slika 2) sa vremenskom diskretizacijom od jednog minuta, odnosno ukupan broj podataka je 25.920. Podaci predstavljaju skup sledećih veličina:

1. statički pad - H_s [m],
2. protok vode kroz turbinu - Q [m³/s],
3. kota gornje vode - H_g [m],
4. kota donje vode - H_d [m],
5. temperatura ulazne rashladne vode - θ_{rv} [°C].

Ove veličine omogućavaju ANN da nauči veze između različitih faktora i njihovog uticaja na izlaznu snagu HE - P_H [MW]. Testiranje ANN je izvršeno na preostalih 40% podataka (Slika 2) prikupljenih u aprilu 2023. godine, čime je demonstrirana prediktivna sposobnost sa visokom tačnošću [1].

Treniranje



Slika 2. Grafički prikaz odnosa podataka za treniranje i testiranje modela razvijenog u radu [1]

Naredni izazov odnosi se na testiranje ANN modela na skupovima podataka za svaki mesec 2024. godine. Ovaj postupak je neophodan za proveru performansi mreže tokom različitih hidroloških perioda i omogućava sveobuhvatnu evaluaciju prediktivnih sposobnosti. Rezultati testiranja će jasno istaknuti poboljšanja u modelu o kojima se diskutuje u ovom radu. Za bolje razumevanje, prvo će biti prikazan grafički odnos podataka za januar 2024. godine u odnosu na april 2023. godine (Slika 3).

Nakon toga, u Tabeli 1 su prikazani skupovi podataka za testiranje ANN, sakupljeni tokom 2024. godine, opet u poređenju sa podacima za treniranje iz aprila 2023. godine.



Slika 3. Grafički prikaz odnosa podataka u jednom mesecu za treniranje i testiranje kombinovanog modela

Tabela 1. Broj i procentualni odnos ulaznih podataka za testiranje kombinovanog modela

Mesec	Broj podataka za testiranje	Procentat u odnosu na ukupan broj podataka [%]
2023		
April	17280	40
2024		
Januar	35.343	57,69
Februar	32.160	55,37
Mart	38.434	59,72
April	41.478	61,54
Maj	44.432	63,16
Jun	42.463	62,10
Jul	42.295	62
Avgust	41.840	61,75
Septembar	43.201	62,5
Oktobar	44.388	63,13
Novembar	42.341	62,03
Decembar	41.165	61,36

U zavisnosti od meseca, odnos između podataka za treniranje i testiranje se razlikuje. Nedostupnost u nekim skupovima podataka je uzrok lošeg merenja ili prekida u telekomunikacionim vodovima.

V EVALUACIJA KOMBINOVANOG MODELA ANN

Evaluacija modela ANN-a zahteva temeljnu analizu njihove efikasnosti sa nepoznatim ulaznim podacima, čime se ispituje njihova sposobnost da primene naučene obrasce na nove ulaze. Ovaj proces obično uključuje različite evaluacione metrike kao što su srednja apsolutna greška (MAE), srednja kvadratna greška (RMSE) i kvadrat koeficijenta korelacije (R^2), koje su takođe primenjene u radu [1]. Nakon detaljne analize načina rada kombinovanog modela u prethodnim poglavljima, izvršeno je poređenje performansi kombinovanog i nekombinovanih ANN modela. Cilj je da istaknemo prednosti kombinovanog pristupa u preciznosti predikcije snage HE i to ćemo prvo uraditi na skupu podataka koji je već analiziran u radu [1], rezultati su prikazani u Tabeli 2. Na osnovu rezultata možemo zaključiti da kombinovani model nije zastao sa tačnošću, u poređenju sa nekombinovanim modelima.

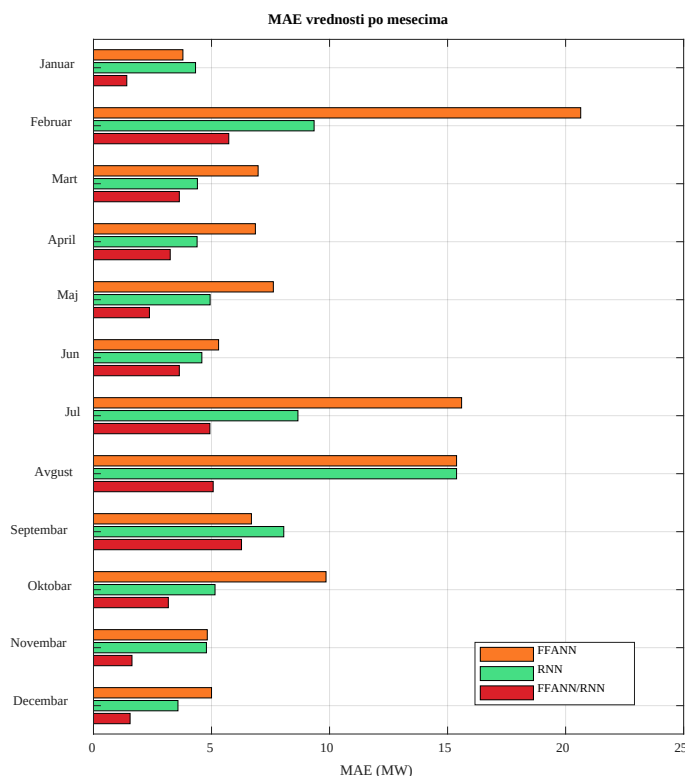
Međutim, prednosti kombinovanog modela postale su očigledne

tek nakon testiranja podataka tokom 2024. godine. Pre prikaza rezultata evaluacije, identifikovani su ključni izazovi s kojima se suočava kombinovani model.

Tabela 2. Evaluacija sva tri modela primenom različitih metrika za april 2023. godine

Mesec (2023)	Metrika evaluacije	Vrednost greške		
		FF-ANN	RNN	FF-ANN/RNN
April	MAE [MW]	4,31	1,14	1,20
	RMSE [%]	4,44	1,44	1,33
	R^2 [%]	91,02	98,79	98,55

Prvi izazov odnosi se na obimnost testiranih podataka [7]. Za razliku od standardne prakse koja podrazumeva podelu podataka na skupove za trening (60% ukupnih podataka), validaciju i testiranje (40% ukupnih podataka), u ovom istraživanju povećan je obim podataka za testiranje. Konkretno evaluacija koja sledi je urađena na mesečnom nivou. Što odgovara odnosima prikazanim u Tabeli 2. Ova promena u raspodeli podataka može imati značajan uticaj na performanse modela i tačnost kreirane neuronske mreže, jer je broj podataka za testiranje daleko veći od broja podataka za treniranje mreže.

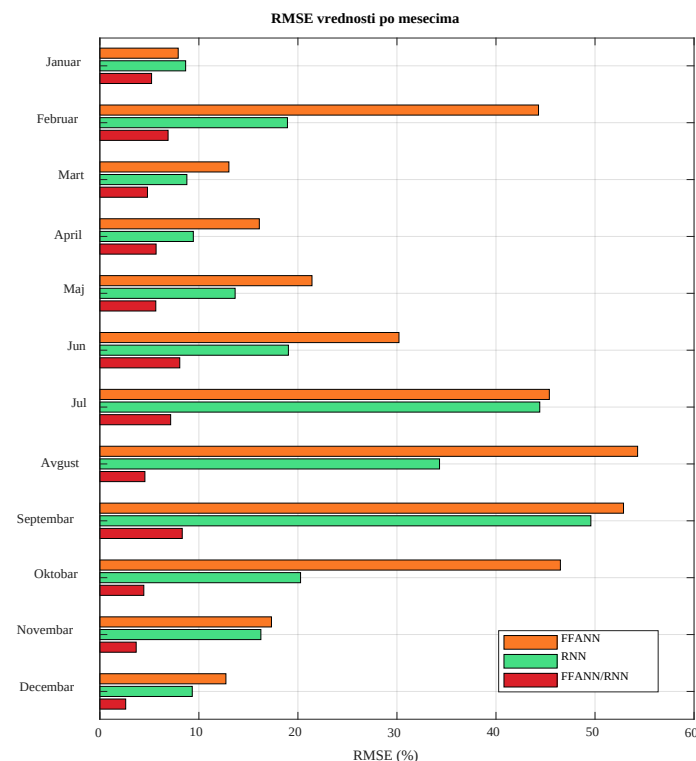


Slika 4. Vrednosti MAE za sva tri modela tokom 2024. godine

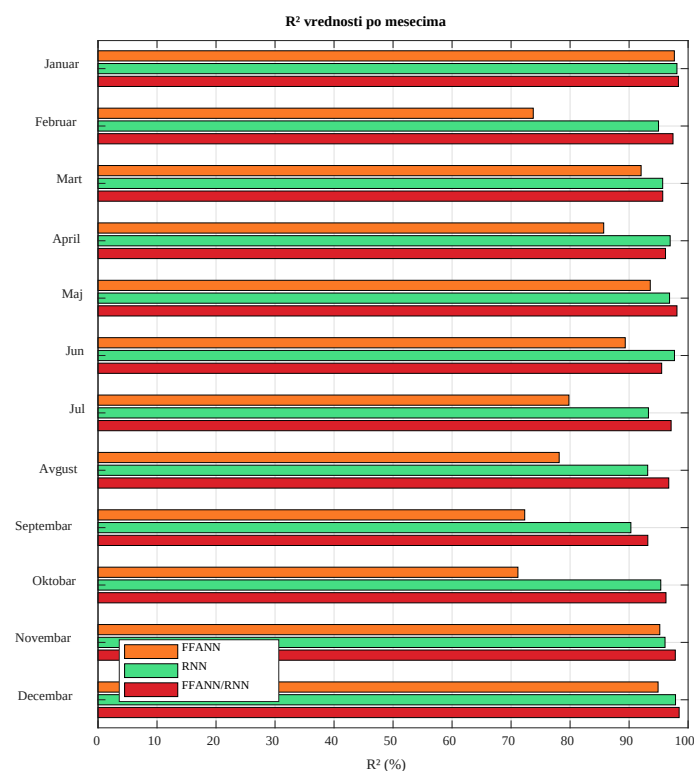
Drugi izazov odnosi se na neskladnost podataka [8], poznatije kao *concept drift* ili *dataset shift* [9, 10], odnosno različita dinamička ponašanja ključnih parametara tokom perioda testiranja neuronske mreže, u odnosu na podatke za koje je mreža trenirana. Konkretno, parametri korišćeni tokom meseca u kojem je mreža trenirana pokazuju odstupanja u odnosu na obrasce prisutne u drugim mesecima. Na primer, protok vode u aprilu može se značajno razlikovati od protoka u avgustu, što može

uticati na sposobnost modela da se prilagođava i precizno predviđa u različitim uslovima.

Rezultati evaluacije prikazani su na Slikama 4, 5 i 6.



Slika 5. Vrednosti RMSE za sva tri modela tokom 2024. godine



Slika 6. Vrednosti R^2 za sva tri modela tokom 2024. godine

Posmatrajući vrednosti MAE (Slika 4), primećuje se da kombinovani, FFANN/RNN, model daje najniže greške u gotovo svim mesecima. Konkretno u februaru, kombinovani model postiže značajno bolje rezultate od pojedinačnih modela, FF-ANN beleži MAE od 23,64 MW, kombinovani model smanjuje ovu grešku na 5,73 MW, što je smanjenje od skoro 76%.

Analizom RMSE (Slika 5), uočava se da RNN u određenim mesecima ostvaruje značajno niže vrednosti u poređenju sa FF-ANN, dok kombinovani model opet dominira u svim mesecima. U februaru i letnjim mesecima, kombinovani model pokazuje RMSE vrednosti koje su mnogo manje od FF-ANN modela.

R^2 ne pokazuje nužno apsolutnu tačnost predikcija, već pokazuje koliko model uspešno prati trend stvarnih vrednosti. Na osnovu Slike 6 može se izvesti zaključak da za kombinovani model postoji izuzetna korelacija između predikcije i stvarnih vrednosti izlazne snage HE.

Rezultati testiranja jasno pokazuju superiornost kombinovanog modela u tačnosti predikcija, sa značajno nižim vrednostima za MAE i RMSE, kao i većim vrednostima R^2 u svim testiranim scenarijima. Uspešnost algoritma se delimično može pripisati sposobnosti kombinovanog modela da obrađuje i analizira veliki broj podataka, čime se ne samo poboljšava njegova adaptacija na varijabilne uslove već i omogućava preciznije predviđanje budućih trendova.

V ZAKLJUČAK

Kombinovani model FF-ANN/RNN za predikciju snage HE, gde je izvršena klaster analiza podataka sa zadatim pragom, koji je definisan na osnovu stabilnosti protoka, u poređenju sa nekombinovanim modelima (FF-ANN i RNN) uspešno je savladao dva značajna izazova: prvi se odnosi na veliki obim testiranih podataka, a drugi na neskladnost podataka tokom različitih perioda testiranja. Rešavanje ovih izazova doprinelo sledećem nizu prednosti:

- visoka preciznost u predikciji snage HE, dokazana različitim metrikama evaluacije,
- fleksibilna predikcije korišćenjem dva modela neuronskih mreža, FF-ANN i RNN, u zavisnosti od njihove relevantnosti u datim uslovima,
- prikupljanje i hronološko organizovanje podatka svakog predviđanja, čime se stvara sveobuhvatna slika o performansama modela tokom vremena, što omogućava detaljniju analizu i potencijalna poboljšanja,

Rezultati prikazani u ovom radu postavljaju temelje za buduća istraživanja u optimizaciji kombinovanog modela, s ciljem poboljšanja njegove efikasnosti i prilagodljivosti u različitim hidrološkim scenarijima. Kao potencijalni pravac budućeg istraživanja, na osnovu preliminarnih testova, autori predlažu uvođenje dinamičkog prilagođavanja intervala i praga stabilnosti, što bi unapredilo trenutni model. Takođe bi bilo korisno analizirati uvođenje novih, nezavisnih, ulaznih parametara za predikciju snage (npr. vodostaj reke ili količina padavina), kao i proširiti analizu na više HE, čime bi se dodatno povećala robusnost rezultata.

ZAHVALNICA

Istraživanja prezentovana u ovom radu su delimično finansirana sredstvima Ministarstva nauke, tehnološkog razvoja i inovacija RS, ugovor br. 451-03-137/2025-03/200132 čiji je realizator Fakultet tehničkih nauka u Čačku - Univerziteta u Kragujevcu.

LITERATURA

- [1] Čubonović, S., Ranković, A., Krstić, M. The Application of Artificial Neural Networks for Predicting the Power Output of a Hydroelectric Power Plant, *Energija, Ekonomija, Ekologija*, Vol. 26, No.2, pp. 47-54, 2024. <https://doi.org/10.46793/EEE24-2.47C>
- [2] Thaeer, H.A., Mohd, H.B.S.N., Abdalla, A. Prediction of small hydropower plant power production in Himreen Lake dam (HLD) using artificial neural network, *Alexandria Engineering Journal*, Vol. 57, No. 1, pp. 211-221, 2018. <https://doi.org/10.1016/j.aej.2016.12.011>
- [3] Renöfält, B.M., Jansson, R., Nilsson, C. Effects of hydropower generation and opportunities for environmental flow management in Swedish riverine ecosystems, *Freshwater Biology*, Vol. 55, No. 1, pp. 49-67, 2009. <https://doi.org/10.1111/j.1365-2427.2009.02241.x>
- [4] Reed, R.D., Marks, R.J. *Neural smithing: Supervised learning in Feedforward Artificial Neural Networks*, IMT Press, 1999. <https://doi.org/10.7551/mitpress/4937.001.0001>
- [5] Mandic, D.P., Chambers, J.A. *Recurrent neural networks for prediction: Learning algorithms, architectures, and stability*, Chichester: John Wiley, 2001. <https://doi.org/10.1002/047084535X>
- [6] Jain, A.K., Murty, M.N., Flynn, P.J. Data Clustering: A Review, *ACM Computing Surveys*, Vol. 31, No. 3, 1999. <https://doi.org/10.1145/331499.331504>
- [7] Zhang, Y., Guo, Q., Wang, J. Big data analysis using neural networks, *Advanced Engineering Sciences*, Vol. 49, No 1, pp. 9-18, 2017. <https://doi.org/10.15961/j.jsuese.2017.01.002>
- [8] Aggarwal, C.C. A framework for clustering evolving data streams, in *Proc. Proceedings 2003 VLDB Conference*, San Francisco, pp. 81-92, 2003. <https://doi.org/10.1016/B978-012722442-8/50016-1>
- [9] Gama, J., Medas, P., Castillo, G., Rodrigues, P. Learning with drift detection, in *Proc. Brazilian Symposium on Artificial Intelligence - SBIA 2004*, pp. 286-295, 2004. https://doi.org/10.1007/978-3-540-28645-5_29
- [10] Baier, L., Kühl, N., Satzger, G. How to Cope with Change? Preserving Validity of Predictive Services over Time, in *Proc. 52nd Hawaii International Conference on System Sciences - HICSS-52*, pp. 1085-1094, 2019. <https://doi.org/10.5445/IR/1000085769>

AUTORI

Stefan Čubonović - asistent, Fakultet tehničkih nauka u Čačku, stefan.cubonovic@ftn.kg.ac.rs, ORCID [0000-0002-0240-5349](https://orcid.org/0000-0002-0240-5349)
Aleksandar Ranković - redovni profesor, Fakultet tehničkih nauka u Čačku, aleksandar.rankovic@gmail.com, ORCID [0000-0002-7205-9584](https://orcid.org/0000-0002-7205-9584)
Marko Krstić, Akcionarsko društvo Elektroprivreda Srbije, marko.krstic996@gmail.com, ORCID [0009-0004-4706-405X](https://orcid.org/0009-0004-4706-405X)

Improvement of Power Output Prediction in Hydroelectric Plants through Cluster Analysis of Data and a Combined Artificial Neural Network Model

Abstract – This research continues previous efforts invested in the application of Artificial Neural Networks (ANN) for predicting the output power of hydroelectric power plants, aiming for further improvements in the accuracy and robustness of the models. The paper focuses on the implementation and evaluation of a combined model that utilizes cluster analysis of water flow through the turbine data with a set threshold and two ANN models: Feedforward Artificial Neural Network (FF-ANN) and Recurrent Neural Network (RNN). Cluster analysis of data with a set threshold, defined based on water flow stability, was applied to select one of the two proposed ANN models. This approach ensured confidence in the model's accuracy and identified specific operational conditions of the hydroelectric power plant. The analysis includes data collected during various hydrological cycles in one year, testing the adaptability of the proposed models. The results of this research promise significant improvements in the efficiency of the previously created model for predicting the power output of hydroelectric power plants.

Index terms – Artificial neural network, FF-ANN, RNN, Hydroelectric power plant, Cluster analysis, Water flow