Klasifikacija ručno pisanih znamenaka i slova korištenjem neuronskih mreža

Tehnička dokumentacija

Verzija 1.0

Studentski tim:

Renato Jurišić

Marko Bagarić

Filip Begović

Filip Weisser

Lovro Glogar

Nastavnik:

Izv. prof. dr. sc. Zoran Kalafatić

Sadržaj

1. Uvod 3

2. Opis skupa podataka 3

3. Pregled ključnih pojmova i arhitektura razvijenih modelae 4

4. Analiza rezultata modela 10

5. Upute za korištenje aplikacije 11

6. Završne napomene 12

7. Zaključak 14

8. Literatura 15

# Uvod

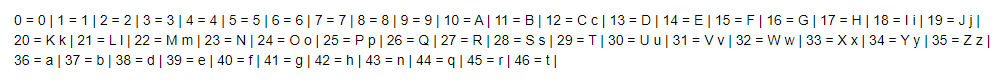
Kao što je već opisano u projektnoj dokumentaciji, tema projekta je klasifikacija ručno pisanih znamenaka i slova korištenjem neuronskih mreža. U tu svrhu razvijena je python aplikacija koja omogućava korisniku da zapiše slovo ili znamenku po izboru i tada se korištenjem treniranog modela pokuša taj zapisan znak klasificirati i rezultat se prikaže korisniku.

U svrhu predviđanja razvijena su dva modela. Prvi, jednostavniji model, sa točnošću od oko 89.46%. Drugi, kompleksniji model, sa točnošću od oko 89.34%. Kako se definira točnost objašnjava se u odjeljku 3.1

# Opis skupa podataka

Skup podataka nad kojim su modeli trenirani je EMNIST Balanced[1] skup podataka. Skup podataka se sastoji od 10 klasa znamenki te 37 klasa veliki i malih slova engleske abecede, dakle sveukupno 47 klasa. Riječ „Balanced“ označava da je broj uzorka za svaku klasu jednak. Sveukupno skup ima 131 600 znakova podijeljenih na skup za treniranje i skup za testiranje koji redom sadrže 112 800 znakova i 18 800 znakova. Svaki znak je slika dimenzija 28×28, tj. matrica u kojoj svaki element predstavlja intenzitet bijele boje u pikselu slike (cijeli broj od 0 do 255). Originalno slike u skupu podataka su zrcaljene i rotirane za 90°, pa ih prije početka treniranja vraćamo u originalan oblik.

Na slici 1 prikazane su klase EMNIST skupa podataka.



Slika 1: Klase EMNIST skupa podataka

Valja uočiti da broj klasa slova nije jednak dvostrukom broju slova engleske abecede. Razlog tomu je to što su neka velika i mala slova nerazlučiva međusobno, na primjer, veliko i malo slovo 'K', veliko i malo slovo 'S' itd. To i vidimo na slici Slika 1.

Također napominjemo da su neka slova i brojevi u nekim slučajevima međusobno slični ili potpuno nerazlučivi. Ovoga ćemo se još dotaknuti u poglavlju 4.

# Pregled ključnih pojmova i arhitektura razvijenih modela

Naši modeli neuronske mreže, njihovu arhitekturu onda čine slojevi te mreže. Svaku mrežu čine ulazni sloj, izlazni sloj te proizvoljan broj skrivenih slojeva između ulaznog i izlaznog.

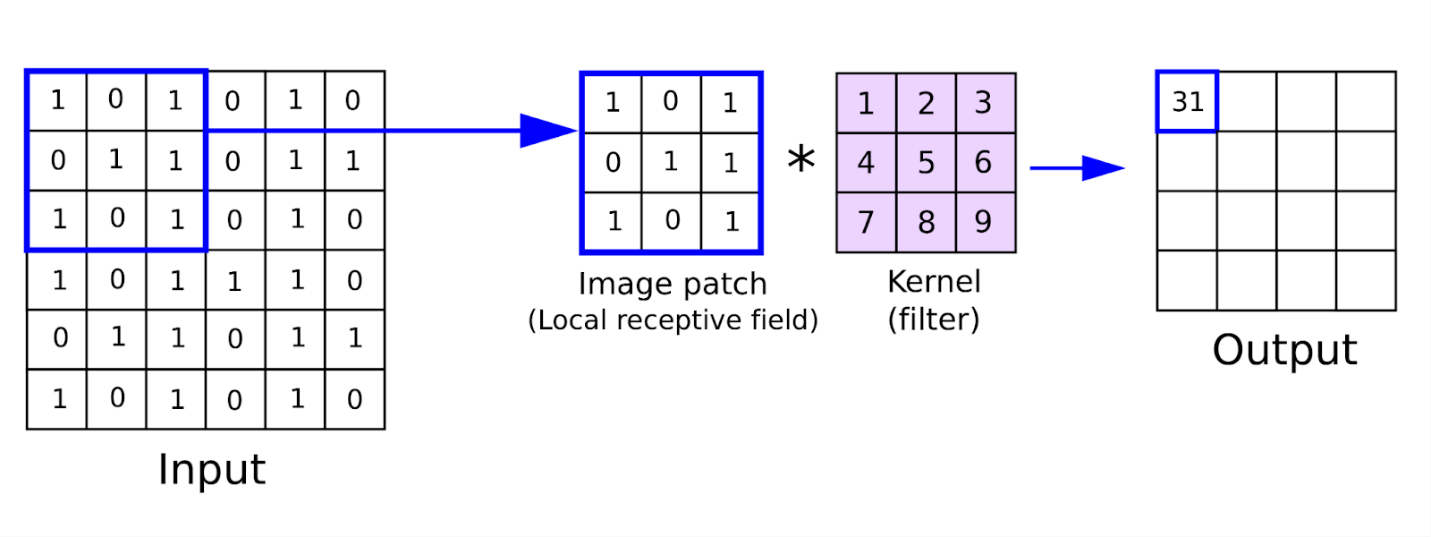
Ulazi neuronskih mreža razvijenih u sklopu ovog projekta čini 28×28 cijelih brojeva koji predstavljaju znak skupa podataka. Kako naši modeli predviđaju znak, tj. klasificiraju ga, izlazni sloj čini 47 izlaznih vrijednosti, gdje vrijednost izlaza *i* predstavlja vjerojatnost da predani znak pripada klasi *i*. Ovdje pod klase mislimo na klase kao što su definirane u skupu podataka. Za vrednovanje razvijenih modela koristit ćemo metriku točnosti.

Svaka mreža razvijena u sklopu projekta čine tri para konvolucijskih i slojeva sažimanja, jednog sloja izravnavanja, jednog sloja nasumičnog izostavljanja neurona(stopa izostavljanja = 0.6) te dva gusta sloja. Sve te slojeve ćemo pobliže objasniti u nastavku.

**3.1 Ključni pojmovi**

*Konvolucijski slojevi* (engl. Convolutional layer)

Konvolucijski slojevi su najčešće korišteni kako bi razaznali pojedine uzorke na slikama kao što su obrub likova, ali koriste se i u ostalim područjima strojnog učenja. Konvolucijski sloj čini filter nekih dimenzija (matrica) koji putuje po slici i izvršava množenje trenutnog dijela slike i filtera što rezultira jednom novom vrijednošću (Slika 2). Taj proces se nastavlja po cijeloj slici sa više različitih filtera i na kraju se dobivaju, takozvane, aktivacijske mape.



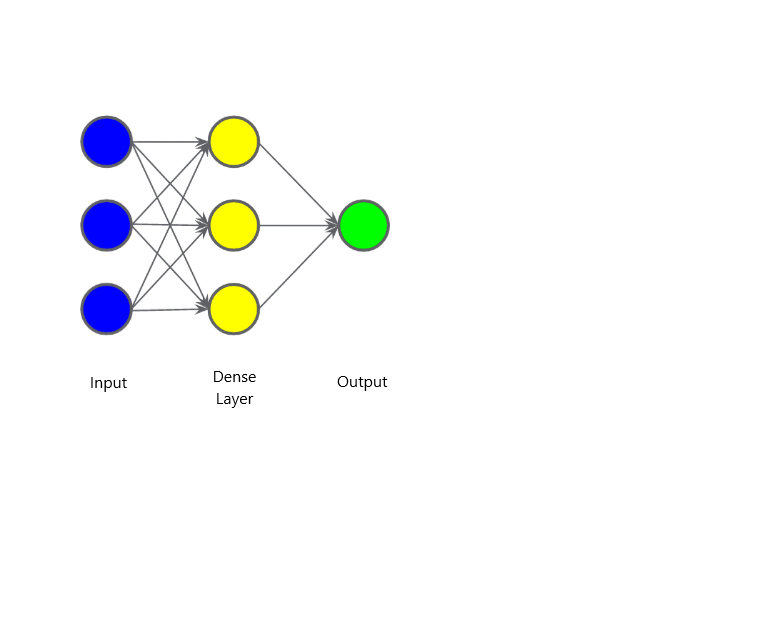
Slika 2: Konvolucija [2]

*Slojevi sažimanja* (engl. Pooling layer)

Slojevi sažimanja se najčešće koriste nakon konvolucijskih slojeva unutar neuronskih mreža. Oni smanjuju rezoluciju mapi dobivenih u konvolucijskim slojevima.

*Gusti slojevi* (engl. Dense layer)

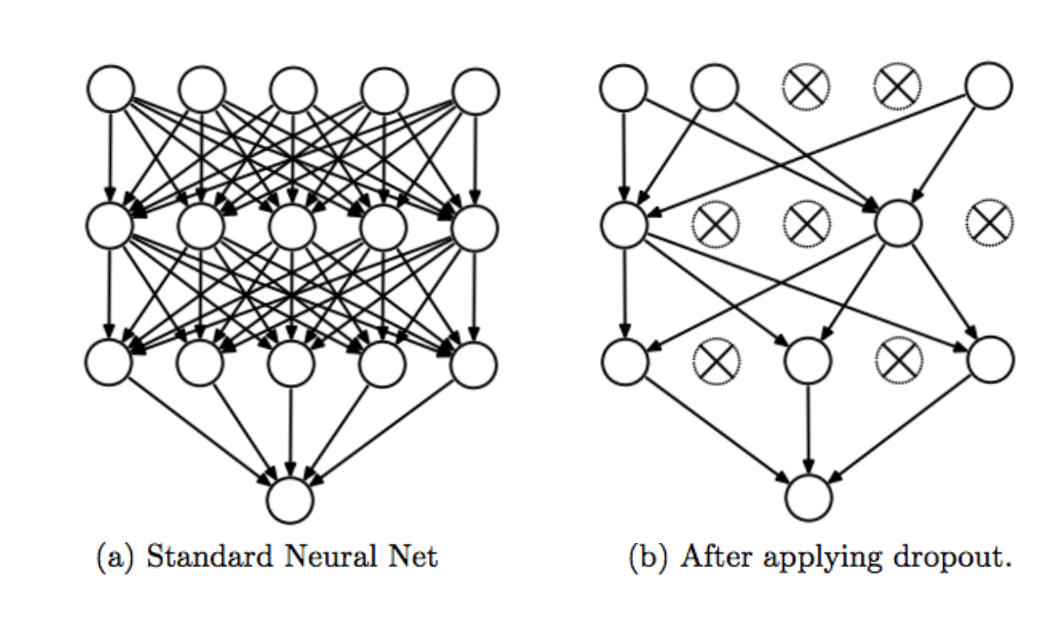
Gusti sloj je sloj u kojem svaki neuron sloja prima izlaz od svakog neurona prijašnjeg sloja. Slika 3 prikazuje jedan takav sloj. Svaki neuron žutog sloja prima izlaz svakog neurona plavog sloja



Slika 3: Gusti sloj

*Slojevi nasumičnog izostavljanja neurona* (engl. Dropout layer)

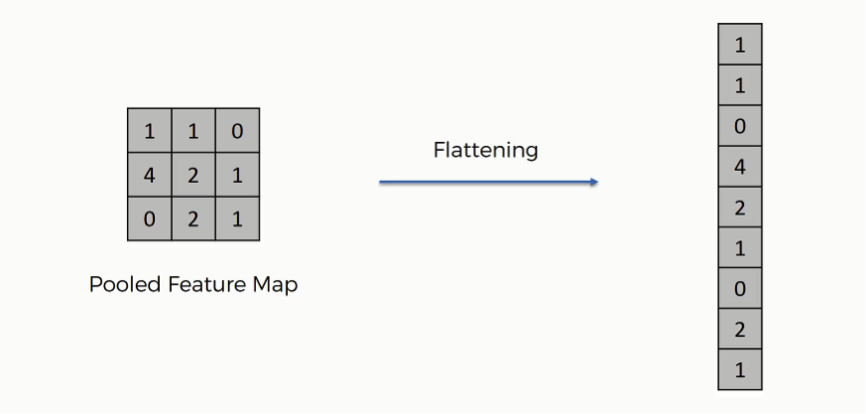
Sloj nasumičnog izostavljanja neurona sloj koji služi za optimiziranje treniranja i pomaže u sprječavanju prenaučenosti. Ideja je vrlo jednostavna, nasumično se odabiru neki neuroni čija se težina postavlja na nulu. Broj odabranih neurona ovisi o stopi izostavljanja.



Slika 4: Izostavljanje neurona [4]

*Slojevi izravnavanja* (engl. Flatten layer)

Sloj izravnavanja izravnava input koji dobiva. Taj sloj izravnava aktivacijske mape dobivene konvolucijskim slojevima i šalje rezultat gustim slojevima.



Slika 5: Izravnavanje [5]

*Mjere evaluacije modela klasifikacije*

Definirajmo prvo neke pojmove koje je najlakše opisati na primjerima.

True Positive (TP) – označava slučaj kada model klasificira sliku kao znamenku '1' dok je na slici uistinu znamenka '1'

False Positive (FP) – označava slučaj kada model klasificira sliku kao znamenku '1' dok na slici nije znamenka '1'

False Negative (FN) – označava slučaj kada model klasificira sliku kao nešto što nije znamenka '1' dok je na slici znamenka '1'

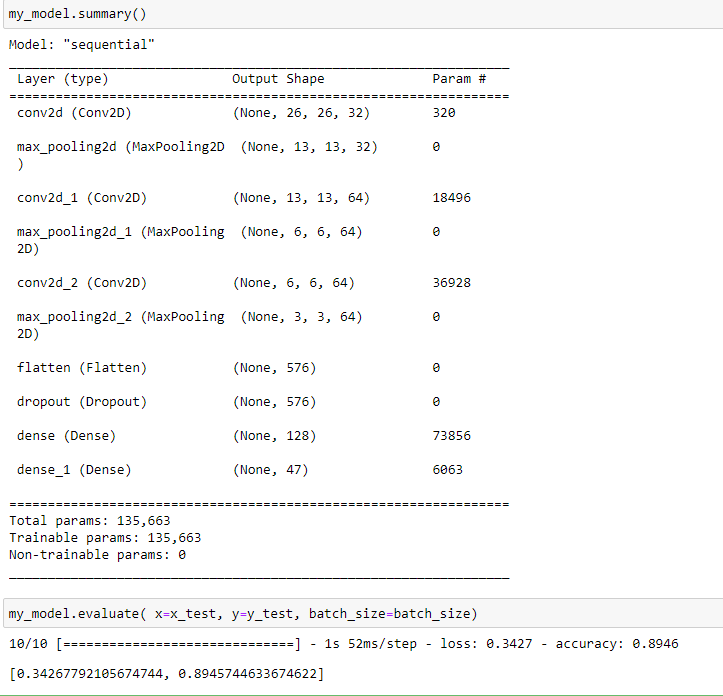
True Negative (TN) – označava slučaj kada model klasificira sliku kao nešto što nije znamenka '1' dok na slici nije znamenka '1'

Neke najčešće korištene mjere evaluacije:

* Točnost (engl. Accuracy) - definira se kao - omjer ispravno klasificiranih primjera i ukupnog broja primjera
* Preciznost (engl. Precision) - definira se kao - omjer broja primjera koje je model označio kao pozitivne i ukupnog broj klasificiranja primjera kao pozitivnog
* Odaziv (engl. Recall) - definira se kao - omjer broja primjera koje je model označio kao pozitivne i ukupnog broj pozitivnih primjera

**3.2 Jednostavniji model**

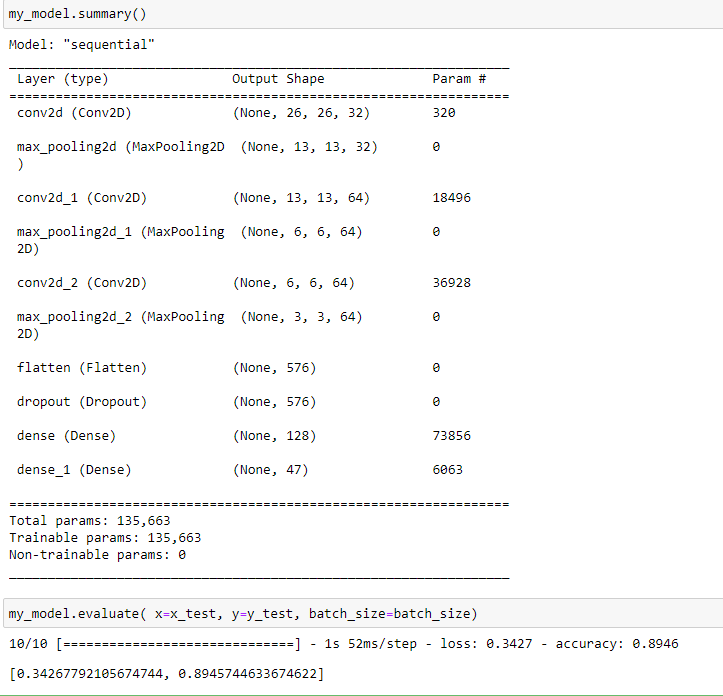
Jednostavniji model ima arhitekturu kako je prikazano na slici 6.



Slika 6: Arhitektura jednostavnijeg modela

Kao što se vidi, mreža se sastoji od 3 para konvolucijskih i slojeva sažimanja, jednog sloja izravnavanja, jednog sloja nasumičnog izostavljanja neurona te dva gusta sloja. Ulazni sloj je prvi konvolucijski sloj koji prima 28×28 sliku, dok je izlazni sloj zadnji gusti sloj i taj ima 47 neurona/izlaza.

Na slici 7 prikazano je testiranje trenirane mreže nad skupom za testiranje.



Slika 7: Rezultat testiranja jednostavnijeg modela

Za jednostavniji model, kao što vidio, dobivamo točnost od 89.46%.

**3.3 Kompleksniji model**

Kompleksniji model ima arhitekturu kako je prikazano na slici 8.

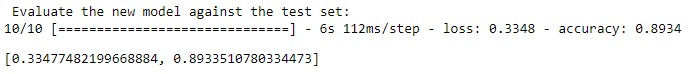
Table

Description automatically generated

Slika 8: Arhitektura kompleksnijeg modela

Kao što se vidi, mreža je vrlo slična jednostavnijem modelu međutim ovaj put konvolucijski slojevi imaju veći broj filtera, sada imaju redom 128, 256 i 256 filtera.

Na slici 9 prikazano je testiranje kompleksnije trenirane mreže nad skupom za testiranje.



Slika 9: Rezultat testiranja kompleksnijeg modela

Vidimo da je točnost kompleksnijeg modela 89.34%

Na slici 10. prikazane su vrijednosti hiperparametara koji su korišteni za treniranje ovih modela.

Text

Description automatically generated

Slika 10: Hiperparametri

# Analiza rezultata modela

Tijekom razvijanja modela u sklopu projekta, isprobali smo mnogo različitih arhitektura, eksperimentirali smo sa različitim vrijednostima hiperparametara te različitim aktivacijskim funkcijama slojeva međutim točnost dobivenih mreža nikada nije prelazila 90%.

Kako bi bolje razumjeli razlog tome, proučimo matricu zabune (Slika 10).

Graphical user interface, histogram

Description automatically generated

Slika 11: Matrica zabune

Matrica je dobivena iz rezultata jednostavnijeg modela. Reci matrice predstavljaju stvarnu vrijednost primjera dok stupci predstavljaju klasifikaciju modela za taj primjer. Intenzitet bijele boje predstavlja vrijednost u tom polju matrice.

Promotrimo primjer, vidimo da postoji neka vrijednost na (21,1). Ako pogledamo kojim znakovima pripadaju te klase vidimo da klasa 21 predstavlja veliko i malo slovo 'L' dok klasa 1 predstavlja znamenku '1'. Pogledajmo primjere tih klasa iz skupa podataka.

A picture containing text

Description automatically generatedText

Description automatically generated with low confidence

Slika 12: Malo slovo 'L' Slika 13: Znamenka '1'

Iz slika 12 i 13 se vidi da gotovo i nema razlike između slova 'L' i znamenke '1'. Naši modeli ne mogu sa visokom sigurnošću odrediti o kojem se znaku stvarno radi bez obzira na arhitekturu ili koliko dugo trenirali. Na to pitanje za ovaj primjer ni čovjek ne bi pouzdano mogao odgovoriti. Zbog tih razloga u matrici zabune se ponegdje vide vrijednosti koje ne leže na glavnoj dijagonali dok bi u savršenom slučaju vidjeli vrijednosti samo na glavnoj dijagonali. To je i jedan od razloga zašto je izrazito teško ili gotovo nemoguće razviti model treniran na ovom skupu podataka koji će predviđati sa točnošću od više od 90%.

# Upute za korištenje aplikacije

Razvijena aplikacija je vrlo jednostavna. Prvo se na platno sa lijeve strane upiše neka znamenka ili slovo. Zatim se pritiskom na gumb „Predict“ s desne strane ekrana ispisuje predviđanje razvijenog jednostavnijeg modela. Pritiskom na gumb „Reset“ se platno i predviđanje brišu i moguće je ponovno crtati. Na slici Slika 14. prikazan je primjer rada aplikacije. Nacrtano je slovo 'S', pritisnut je gumb „Predict“ i ispisale su se predviđanja našeg modela. Vidimo da je dobro predvidio da se na ulazu nalazi slovo 'S' (98.35%), također primjećujemo da ispisuje 1.65% za znamenku '5', to je opet posljedica gore opisanih sličnosti među nekim znamenkama i slovima. U ovome slučaju ta neodlučnost modela je mala ali za neke slučajeve ona može biti značajno veća.

A picture containing graphical user interface

Description automatically generated

Slika : Screenshot rada aplikacije

# Završne napomene

Na samome kraju, pogledajmo koliko dobro naši modeli predviđaju podatke MNIST skupa podataka i „Unbalanced“ EMNIST skupa podataka.

**6.1 MNIST**

MNIST skup podataka sadrži samo znamenke od 0 do 9. Promotrimo matricu zabune dobivenu MNIST skupom podataka i našim jednostavnijim modelom.

Graphical user interface

Description automatically generated

Slika 13: Matrica zabune za MNIST

Ako promotrimo slučajeve gdje je naš model pogriješio, vidimo da je se javlja ista situacija kao i prije. Recimo na mjestu (1, 18) koje predstavlja da je ulaz bila znamenka '1', a naš model je predvidio slovo 'I'. Zbog istih razloga opisanih prije ovakve situacije model teško može sa visokom sigurnošću predvidjeti.

**6.2 Unbalanced EMNIST ByMerge**

Unbalanced EMNIST ByMerge skup podataka sadrži , kao i Balanced EMNIST, 47 klasa, međutim sada broj uzoraka za svaku klasu nije jednak. Promotrimo na slici Slika 14. tu raspodjelu.

Chart, histogram

Description automatically generated

Slika 14: Raspodjela uzoraka po klasama za Unbalanced EMNIST ByMerge

Brojeva (klase od 0 do 10) ima značajnije više nego znamenaka. Pogledajmo matricu zabune za ovaj skup podatak i naš jednostavniji model.

Graphical user interface, application

Description automatically generated

Slika 15: Matrica zabune za Unbalanced EMNIST ByMerge

Vidimo vrlo sličnu situaciju kao i kod Balanced EMNIST skupa podatak. Jedino su neka polja izraženija jer skup ima više uzoraka iz klasa od 0 do 10 nego ostalih.

# Zaključak

Usporedimo li točnost dobivenu na Balanced EMNIST skupu sa točnostima dobivenim na prethodna dva, zaključujemo da naši modeli prihvatljivo dobro klasificiraju podatke. Za MNIST skup ta točnost je malo niža, oko 70%, opet zbog neizbježne sličnosti u nekim uzorcima različitih klasa ( npr. znameka '5' i slovo 'S'). Dok je točnost Unbalanced EMNIST skupa od 86% potpuno prihvatljiva.

Također, možemo usporediti rezultate naših modela ( točnosti 89.46% i 89.34%) sa nekim poznatijim modelima. Na primjer, 'najbolji' model koji smo uspjeli naći , VGG-5 tj. 'SpinalNet: Deep Neural Network with Gradual Input'[7], postiže točnost od 91.05% što također pokazuje prihvatljivost naših modela.

# Literatura

[1] Cohen, G., Afshar, S., Tapson, J., & van Schaik, A. (2017). EMNIST: an extension of MNIST to handwritten letters. Dostupno na: <http://arxiv.org/abs/1702.05373>

[2] Vijawsinh Lendave, „What Is A Convolutional Layer?“, 18. lipnja 2021., <https://analyticsindiamag.com/what-is-a-convolutional-layer> , 1. siječnja 2022.

[3] Kopljar, D. , Konvolucijske neuronske mreže, Završni rad, Sveučilište u Zagrebu, Fakultet Elektrotehnike I Računarstva, 2016.

[4] Amar Budhiraja, „Droput in (Deep)Machine learning)“, 15. prosinca 2016., <https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5> , 1. siječnja 2022.

[5] SuperDataScience Team, „Convolutional Neural Networks (CNN): Step 3 – Flattening“, 18. kolovoza 2018., <https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-3-flattening> , 1.siječnja 2022.

[6] Đuričić, T.; Merćep, A.; Strojno učenje – Bilješke s predavanja, Sveučilište u Zagrebu, Fakultet Elektrotehnike i Računarstva, 2015/2016, <https://www.fer.unizg.hr/_download/repository/SU-2015-Vrednovanje_modela.pdf>

[7] Kabir D., Abdar M., Jalali S., Khosravi A., Atiya A., Nahavandi S., Srinivasan D.; SpinalNet: Deep Neural Network with Gradual Input; <https://arxiv.org/pdf/2007.03347v3.pdf>