

FEHÉRJE KLASSZIFIKÁCIÓ FUNKCIÓSOSZTÁLYOK ALAPJÁN

Szökrön Dorottya
Önálló projekt II.

Témavezető: Lukács András

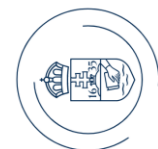
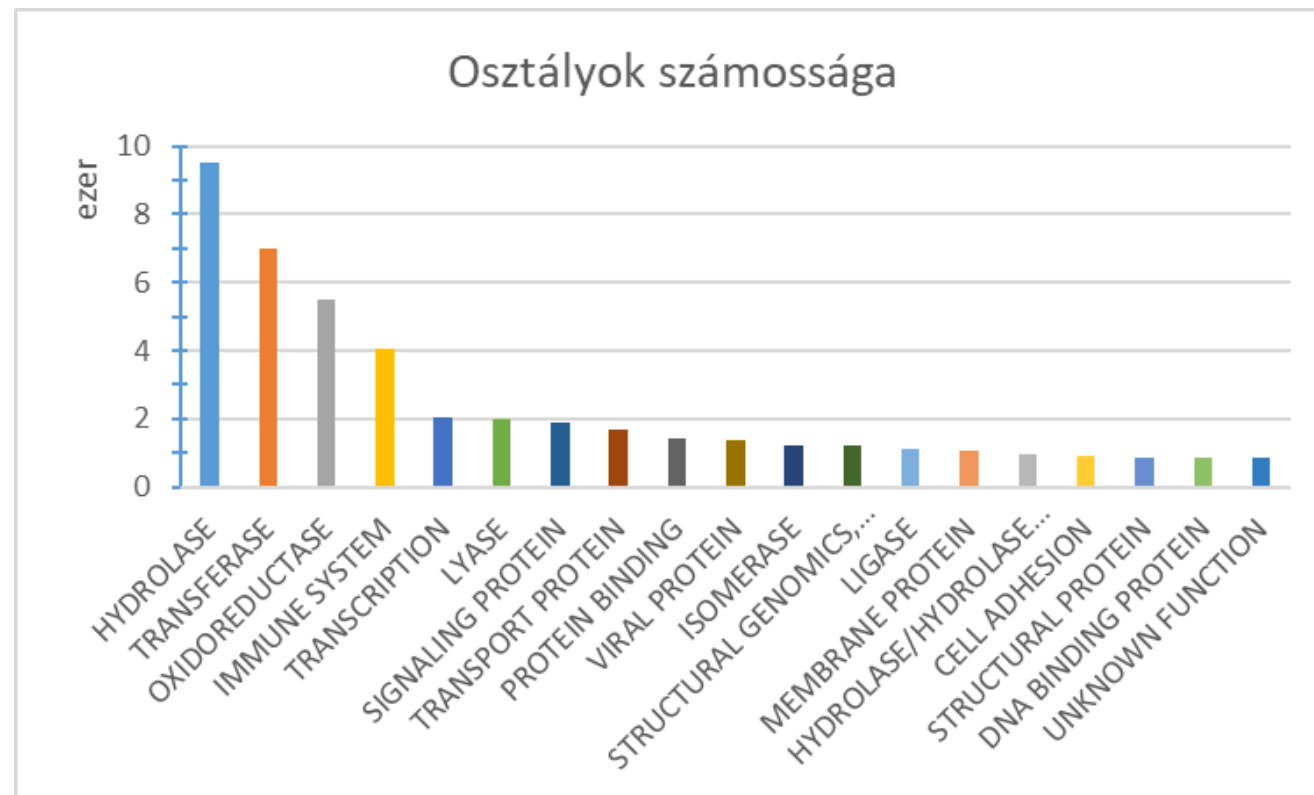


ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Budapest, 2022. május 19.

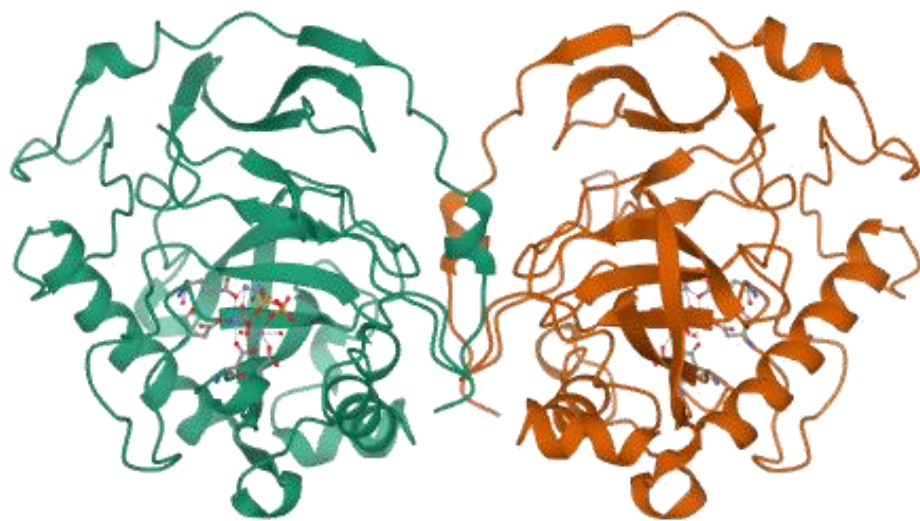
Adathalmaz

HYDROLASE	9508
TRANSFERASE	6974
OXIDOREDUCTASE	5477
IMMUNE SYSTEM	4034
TRANSCRIPTION	2069
LYASE	1975
SIGNALING PROTEIN	1872
TRANSPORT PROTEIN	1689
PROTEIN BINDING	1447
VIRAL PROTEIN	1383
ISOMERASE	1245
STRUCTURAL GENOMICS, UNKNOWN FUNCTION	1198
LIGASE	1122
MEMBRANE PROTEIN	1074
HYDROLASE/HYDROLASE INHIBITOR	937
CELL ADHESION	915
STRUCTURAL PROTEIN	863
DNA BINDING PROTEIN	840
UNKNOWN FUNCTION	835



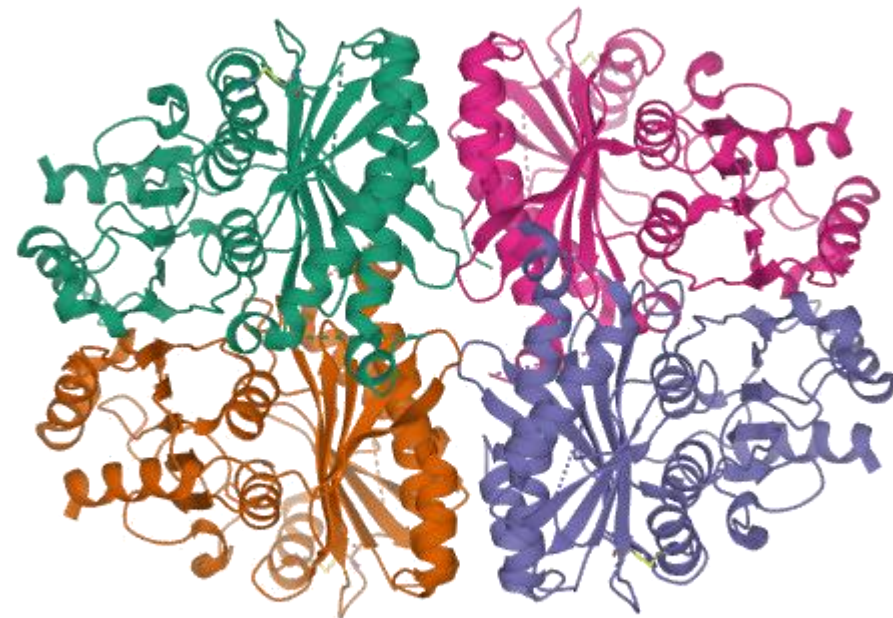
INORGANIC PYROPHOSPHATASE (P00817)

MTYTTRQIGA	KNTLEYKVYI	EKDGPVSAF	HDIPLYADKE	NNIFNMVVEI
PRWTNAKLEI	TKEETLNPII	QDTKKGKLR	VRNCFPHHG	IHNYGAFPQT
WEDPNVSHPE	TKAVGDNDPI	DVLEIGETIA	YTGQVKQVKA	LGIMALLDEG
ETDVKVIAID	INDPLAPKLN	DIEDVEKYFP	GLLRATNEWF	RIYKIPDGKP
ENQFAFSGEA	KNKKYALDII	KETHDSWKQL	IAGKSSDSKG	IDLTNVTLPD
TPTYSKAASD	AIPPASPKAD	APIDKSIDKW	FFISGSV	



FRUCTOSE-1,6-BISPHOSPHATASE, CHLOROPLASTIC (P46275)

MAAATASSQL	IFSKPYSPSR	LCPFQLCVFD	AKSVLSSRR	KHVNGSGVRC
MAVKEATSET	KKRSGYEIIT	LTSWLLQEQ	KGIIDAELTI	VLSSISMACK
QIASLVQRAN	ISNLTGTQGA	VNIQGEDQKK	LDVISNEVFS	NCLRSSGRTG
IIASEEDVA	VAVEESYSGN	YIVVFDPLDG	SSNLDAAVST	GSIFGIYSPN
DECLPDFGDD	SDDNTLGTTE	QRCIVNVCQP	GSNLLAAGYC	MYSSSVAFVL
TIGKGVFVFT	LDPLYGEFVL	TQENLQIPKS	GEIYSFNEGN	YKLWDENLKK
YIDDLKEPGP	SGKPYSARYI	GSLVGDHFRT	LLYGGIYGYP	RDKKSKNGKL
RLLYECAPMS	FIVEQAGGKG	SDGHQRVLDI	QPTEIHRVVP	LYIGSTEEVE KVEKYLA

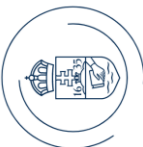
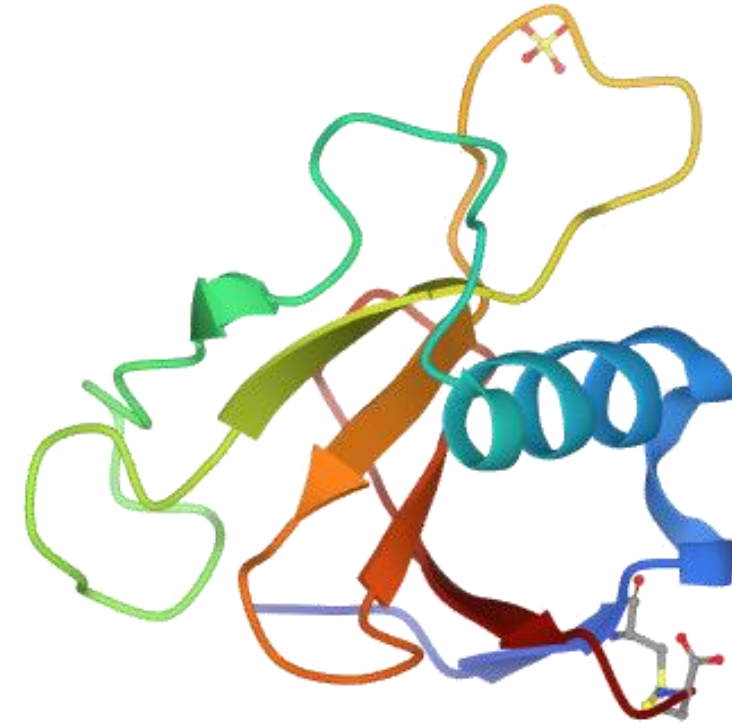
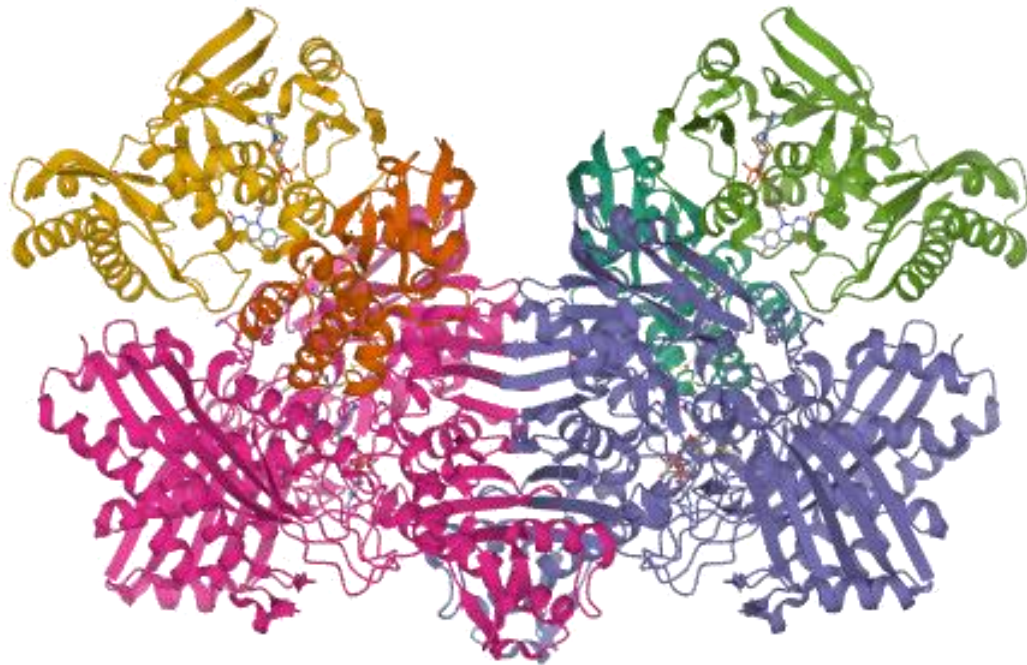


CARBON MONOXIDE DEHYDROGENASE MEDIUM CHAIN (P19914)

MIPPRFEYHA	PKSVGEAVAL	LGQLGSDAKL	LAGGHSLLP	MKLRFAQPEH
LIDINRIPEL	RGIREEGSTV	VIGAMTVEND	LISSPIVQAR	LPLLAEEAAKL
IADPQVRNRG	TIGGDIAHGH	PGNDHPALSI	AVEAHFVLEG	PNGRRTPAD
GFFLGTYMTL	LEENEVMVEI	RVPFAAQGTG	WAYEKLKRT	GDWATAGCAV
VMRKSNTVS	HIRIALTNVA	PTALRREGGR	SRLGKFTK	EAVQAAADAA
IAICEPAEDL	RGDADYKTAM	AGQMVKRALN	AAWARCA	

GUANYL-SPECIFIC RIBONUCLEASE SA (P05798)

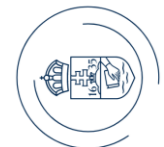
DVSGTVCLSA	LPPEATDTLN	LIASDGPFY	SQDGVVFQNR	ESVLPTQSYG
YYHEYTVITP	GARTRGTRRI	ITGEATQEDY	YTGDIHYATFS	LIDQTC



Feladat

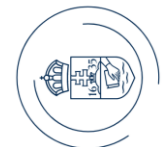
- Fehérje klasszifikáció:

Neurális hálók segítségével vizsgáltam az adathalmazban található aminosavszekvenciákat, hogy kapcsolatot találjak a szekvenciák és az azokhoz rendelt funkciós osztályok között. Ezáltal olyan modellt létrehozva, melynek segítségével besorolatlan fehérjék is a már meglévő osztályokba sorolhatók, így nyerve információt a működésükről.

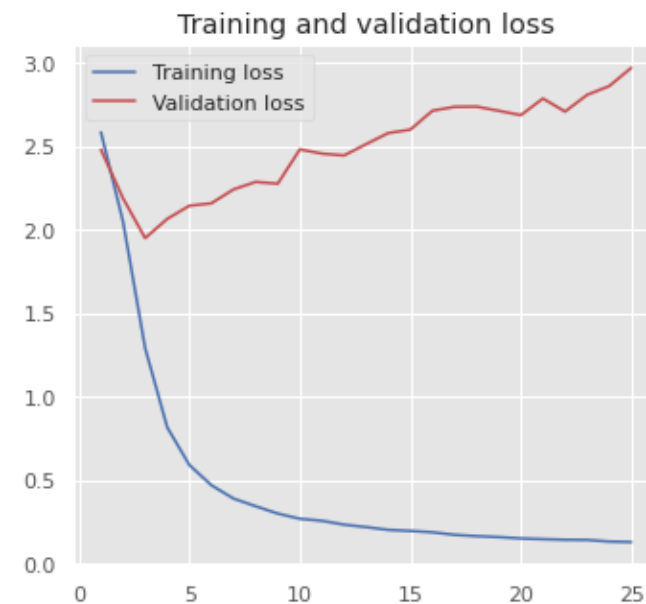
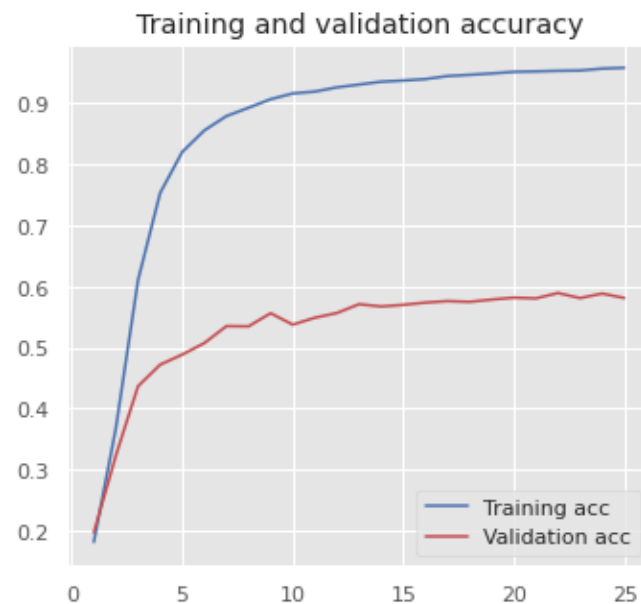
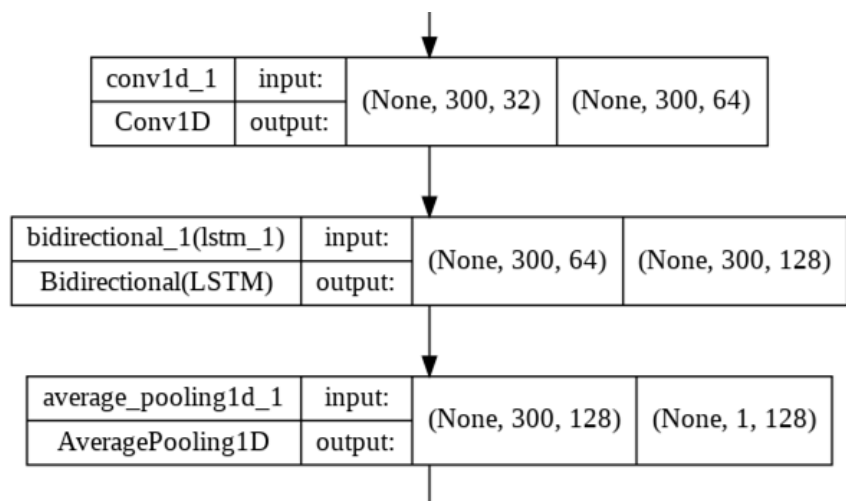


Előző félév összefoglalása

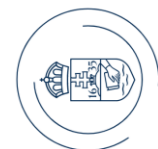
	CONV1D	CONV1D	HYBRID	HYBRID
Database	unbalanced	balanced	unbalanced	balanced
Train acc (%)	98.23	96.71	97.7	96.12
Train loss	0.06	0.15	0.09	0.17
Val acc (%)	55.58	45.18	57.18	38.23
Val loss	3.13	3.22	2.61	3.41
Test acc (%)	56.88	45.98	57.84	39.42
Test loss	2.99	3.19	2.56	3.37

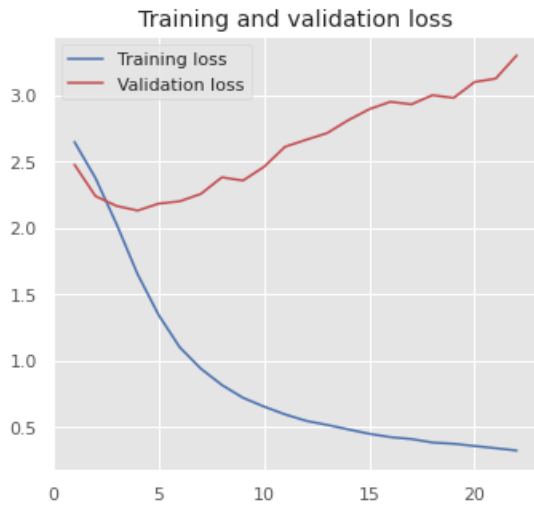
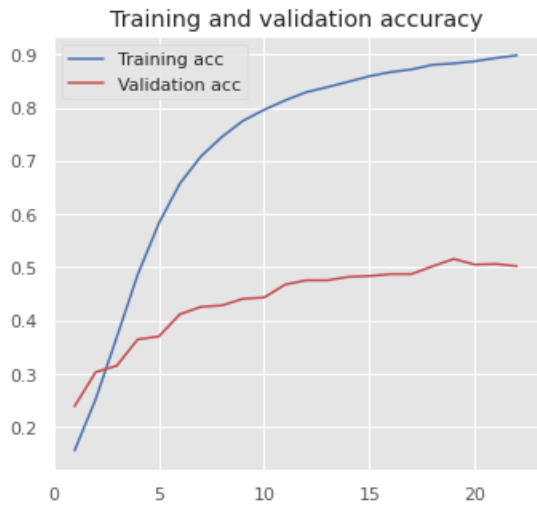


1DCONV & LSTM modell

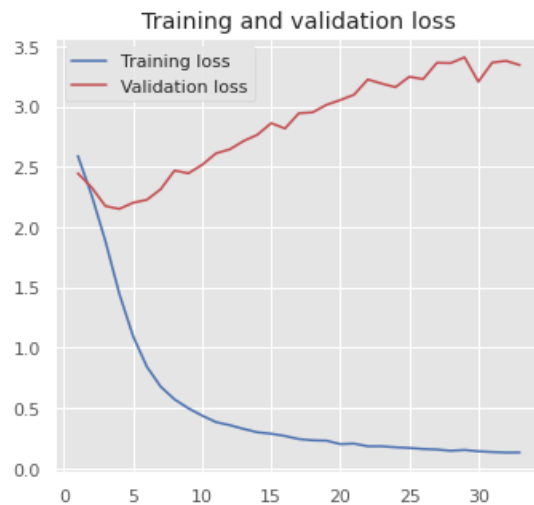
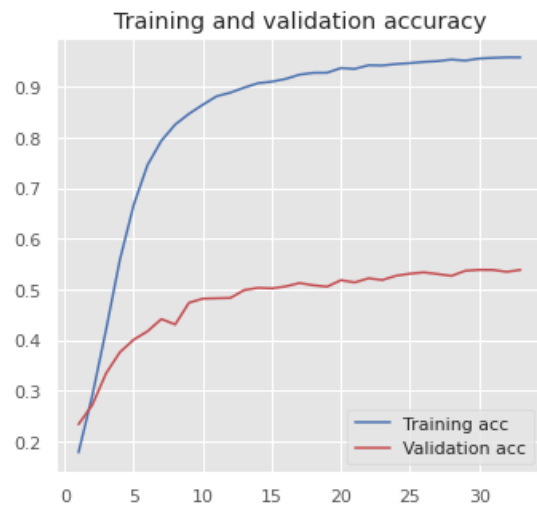


Párhuzamos kutatásban létrejött hálózati felépítés implementálása TensorFlow környezetben alapmodellként. Valamint az alapmodell kezdeti teljesítménye.

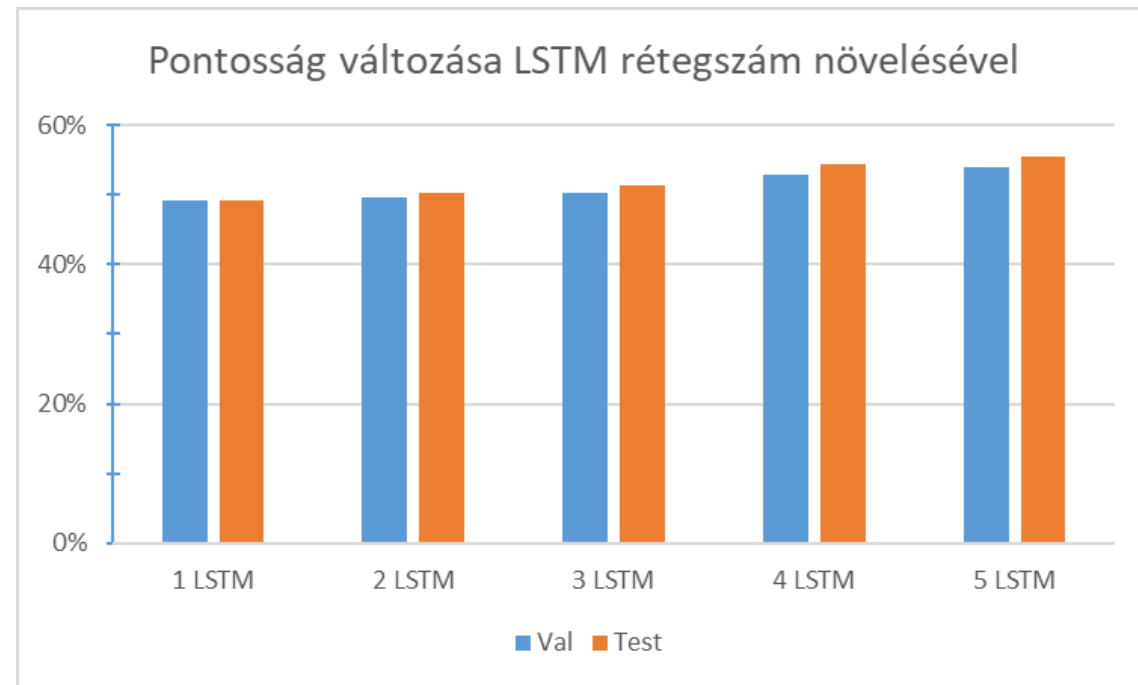




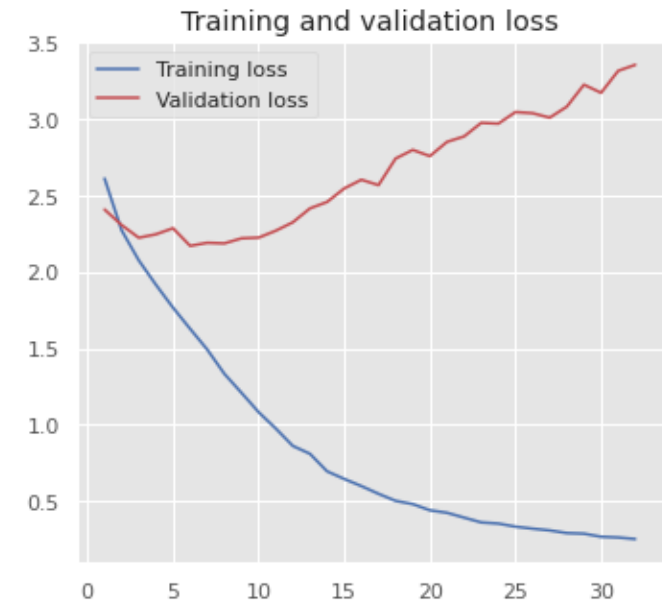
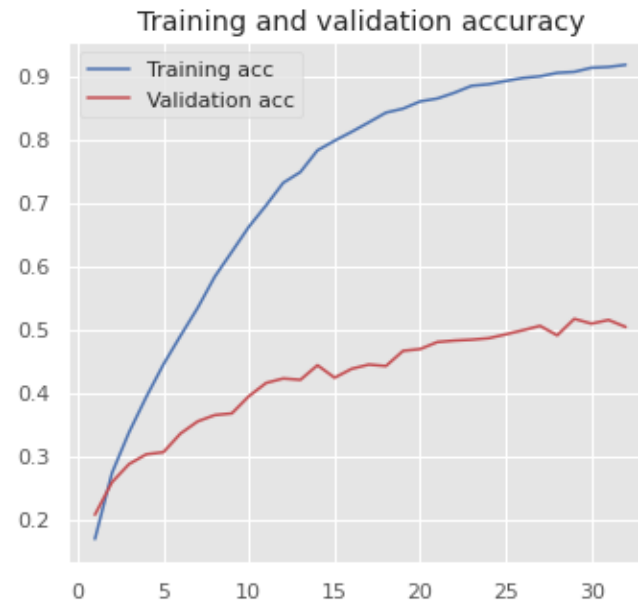
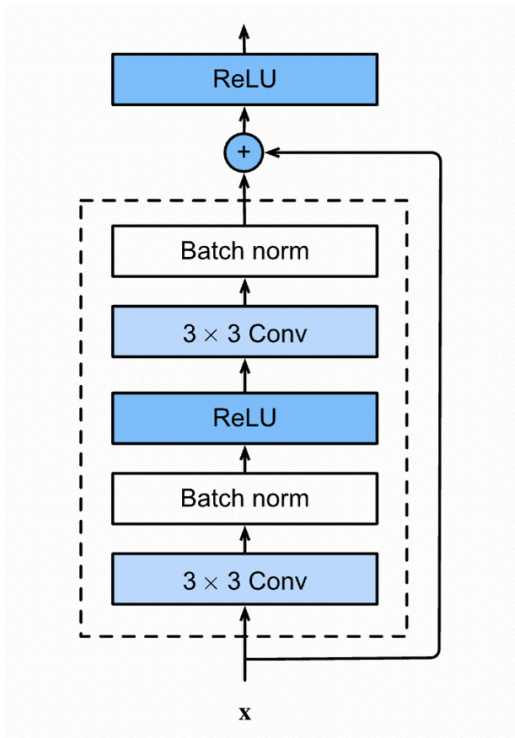
3 LSTM réteggel rendelkező modell teljesítménye.



5 LSTM réteggel rendelkező modell teljesítménye.

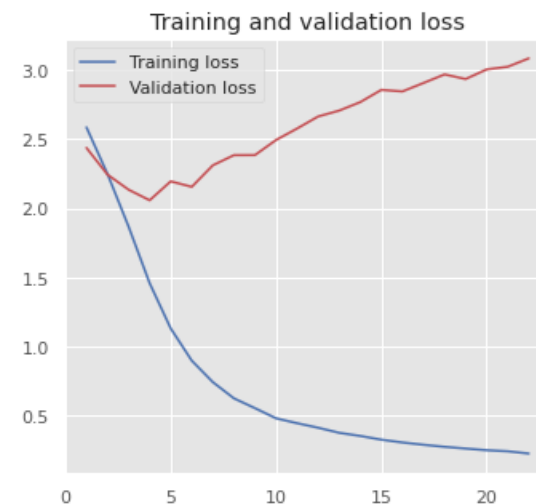
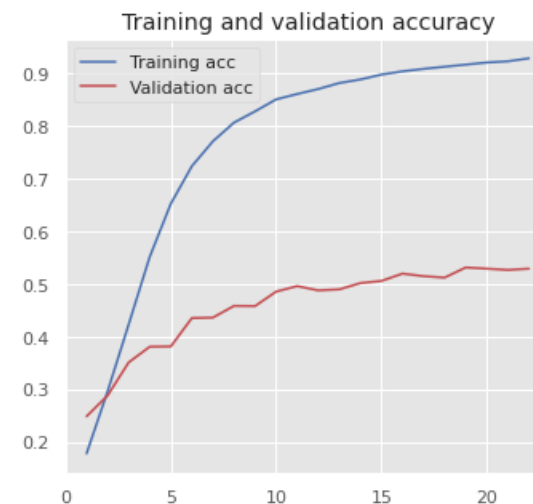
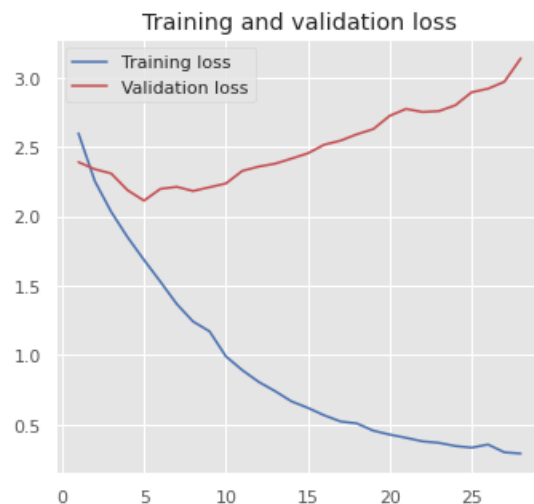
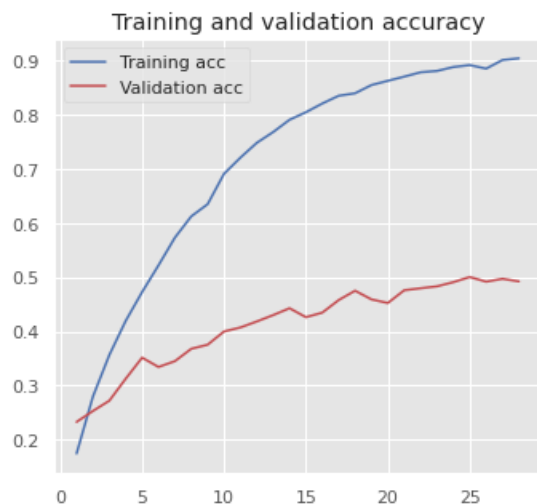


Residual 1D CONV & LSTM modell

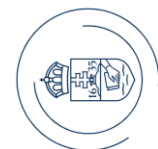


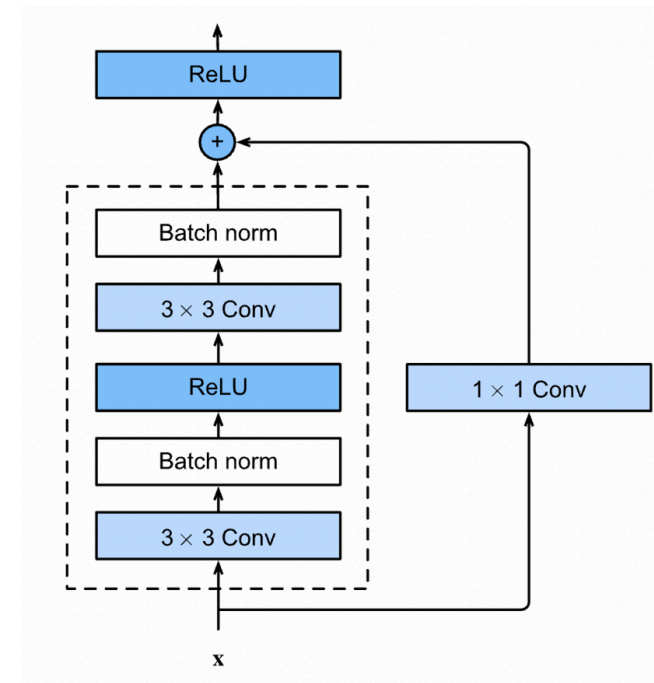
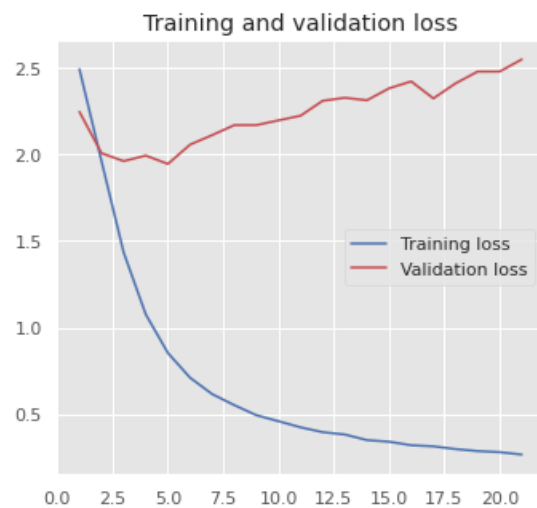
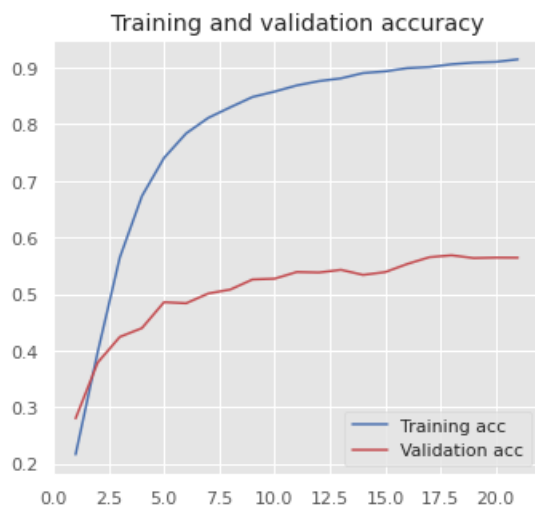
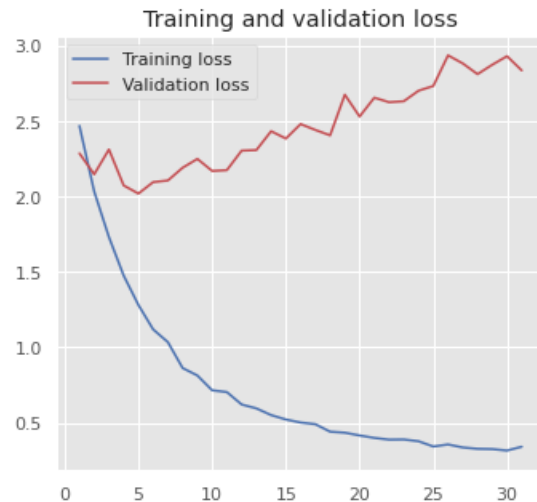
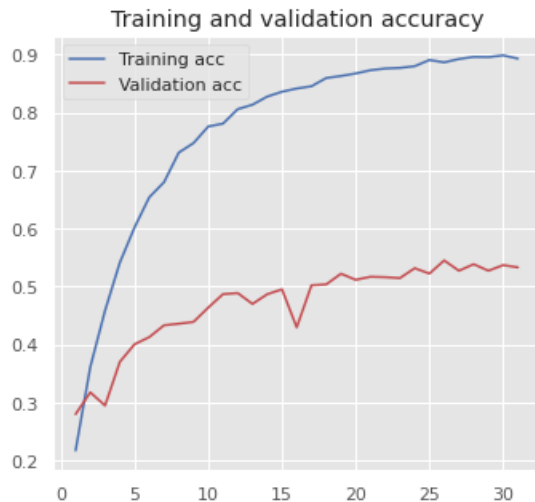
A példában szereplő architektúrához hasonló reziduális blokk alkalmazása LSTM réteggel.

Összehasonlítás

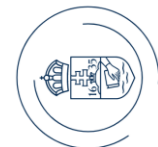


Reziduális blokkot alkalmazó (balra) és nem alkalmazó (jobbra) , 4 LSTM réteggel rendelkező modellek.

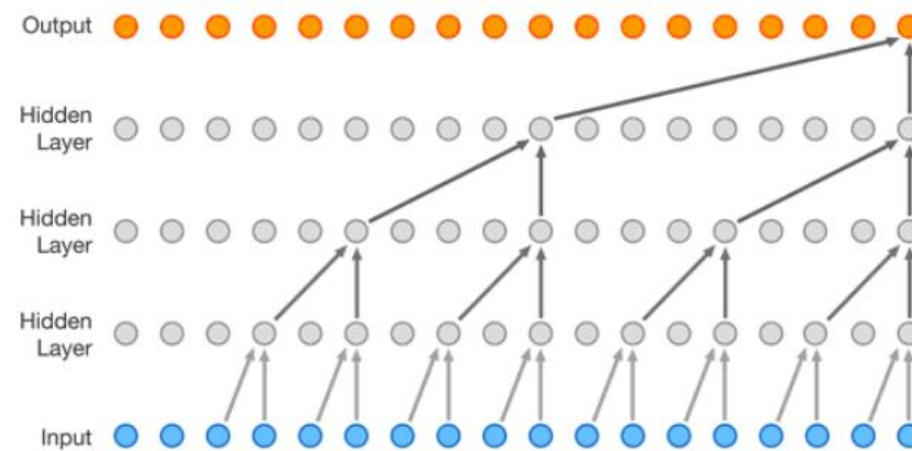
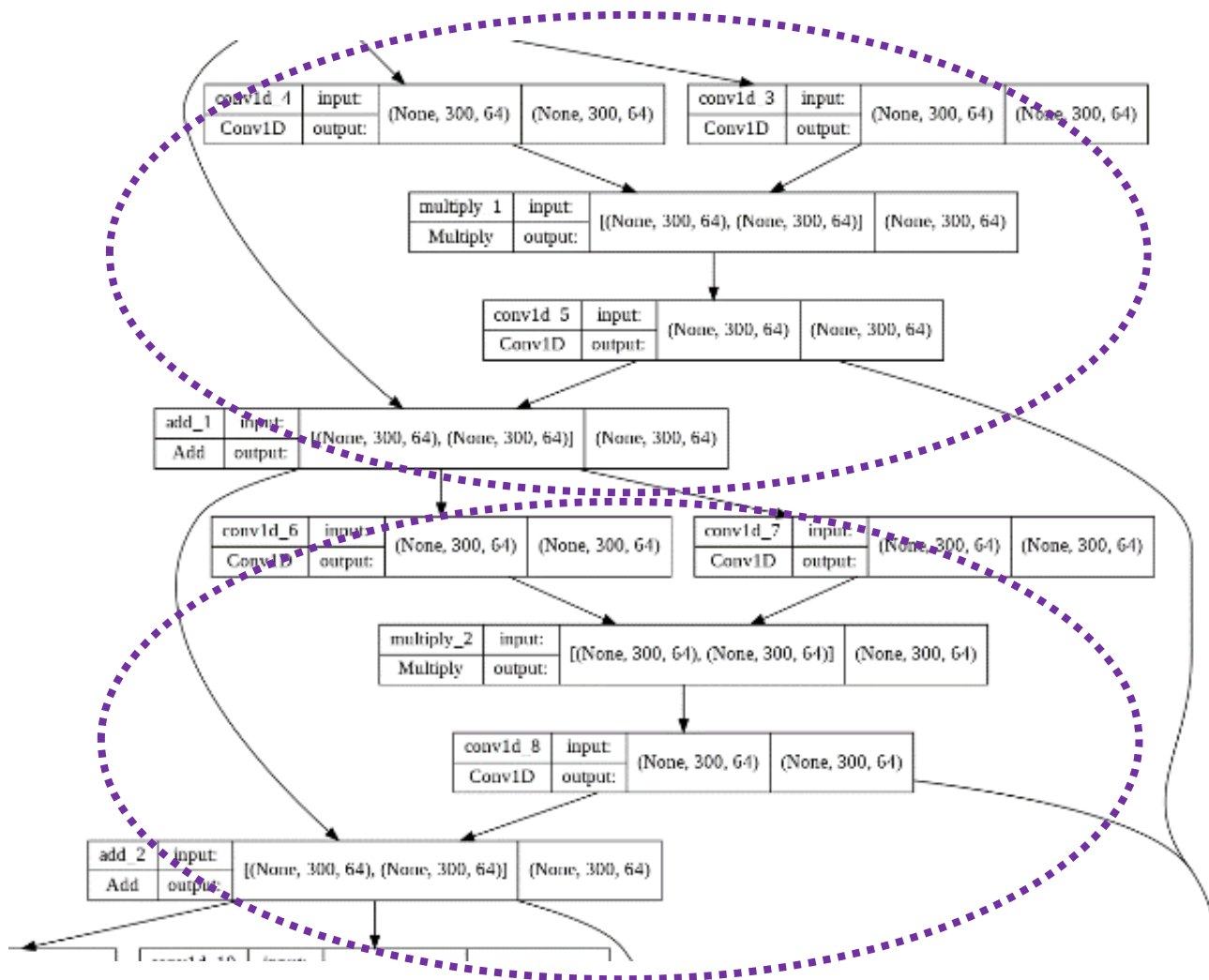




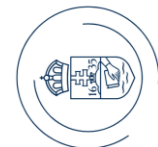
Más típusú reziduális blokkok alkalmazása.



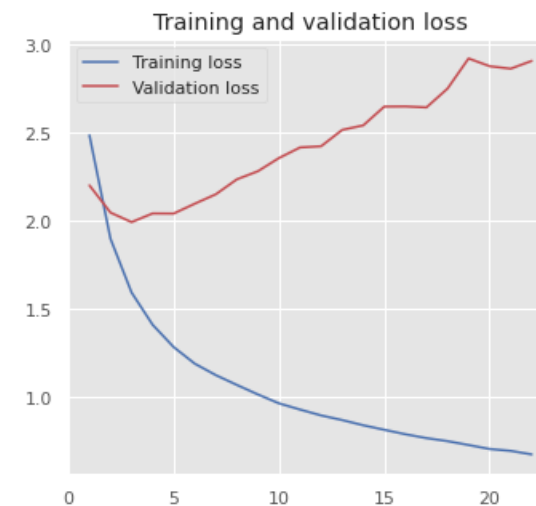
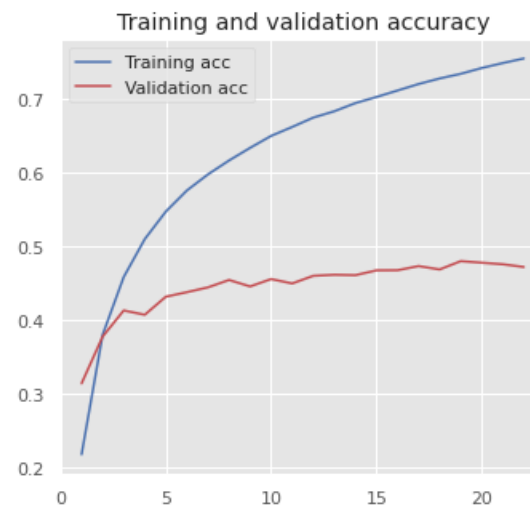
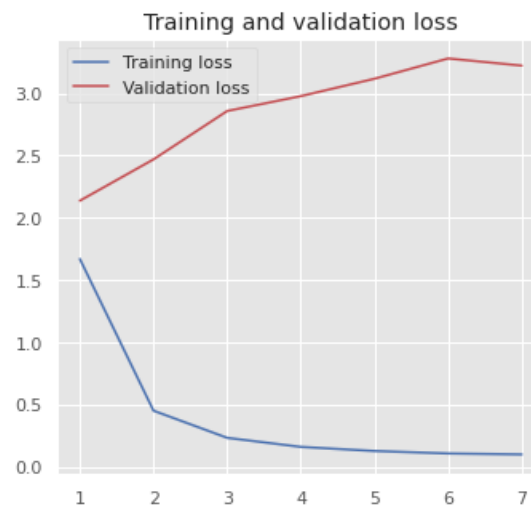
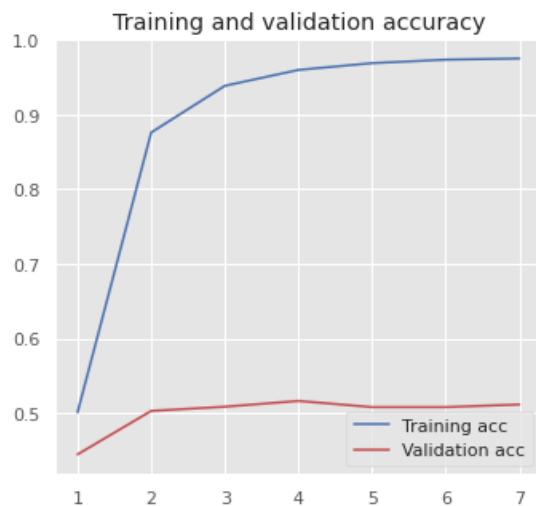
Dilated 1DCONV modell



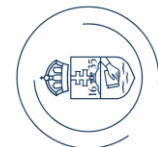
WaveNet modell



Dilated 1D CONV modell



Az előző félévben alkalmazott hagyományos 1D konvolúciós hálózat összehasonlítása a dilatált konvolúciós modellel.



Köszönöm a figyelmüket!



ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM