

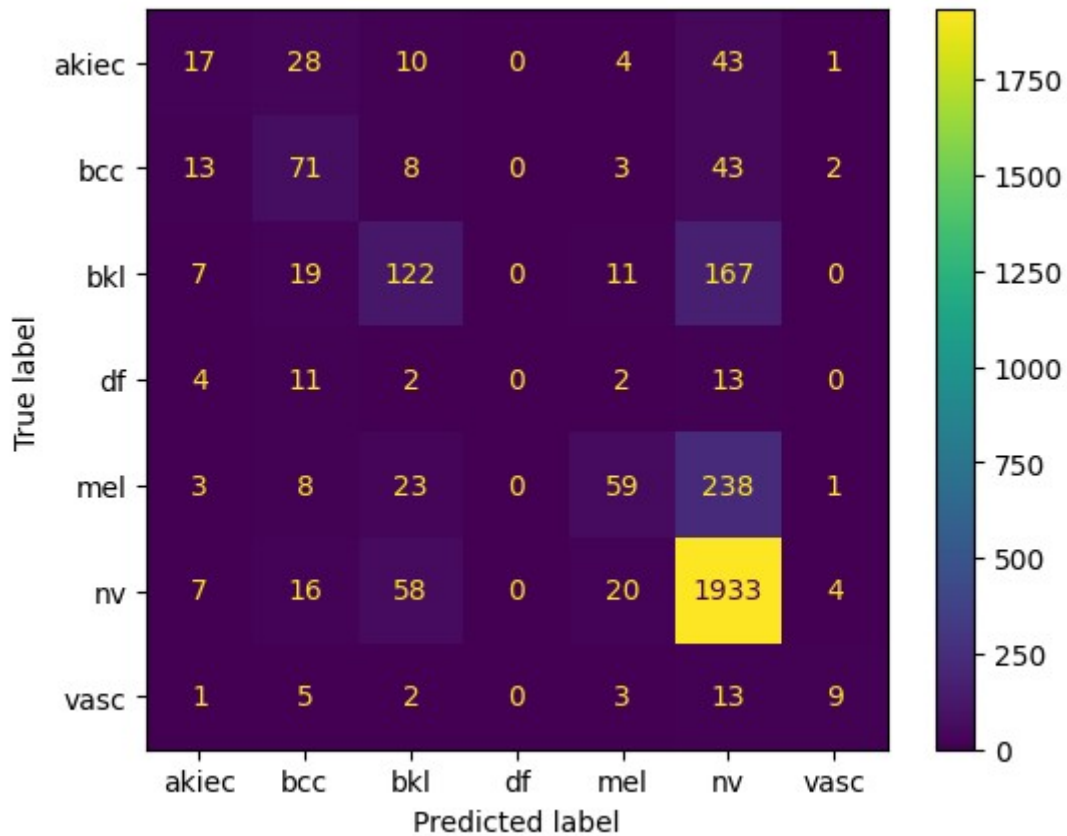
Μηχανική Μάθηση και Εφαρμογές

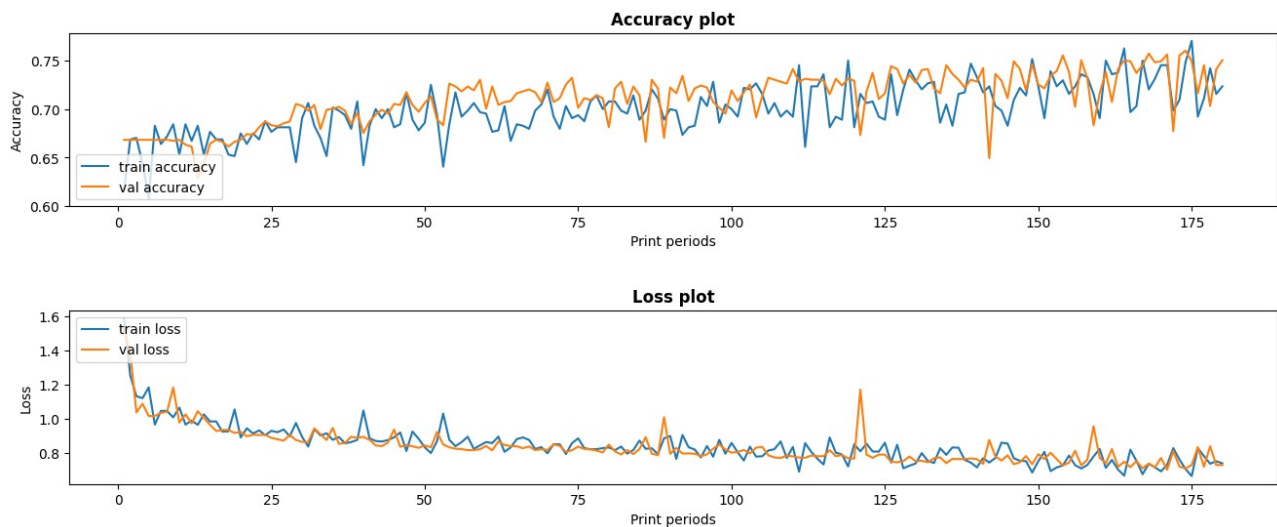
2η Άσκηση: Συνελικτικά Νευρωνικά Δίκτυα

it22039, Καραθανάσης Ανδρέας

Απλό ΣΝΔ:

Στα δεδομένα δοκιμής (testing) το απλό συνελικτικό μοντέλο επιτυγχάνει 0.73% ευστοχία και απώλεια 0.7. Ακολουθεί ο πίνακας σύγχυσης και τα γραφήματα με τις ενδιάμεσες τιμές της απώλειας και της ευστοχίας κατά την εκπαίδευση:

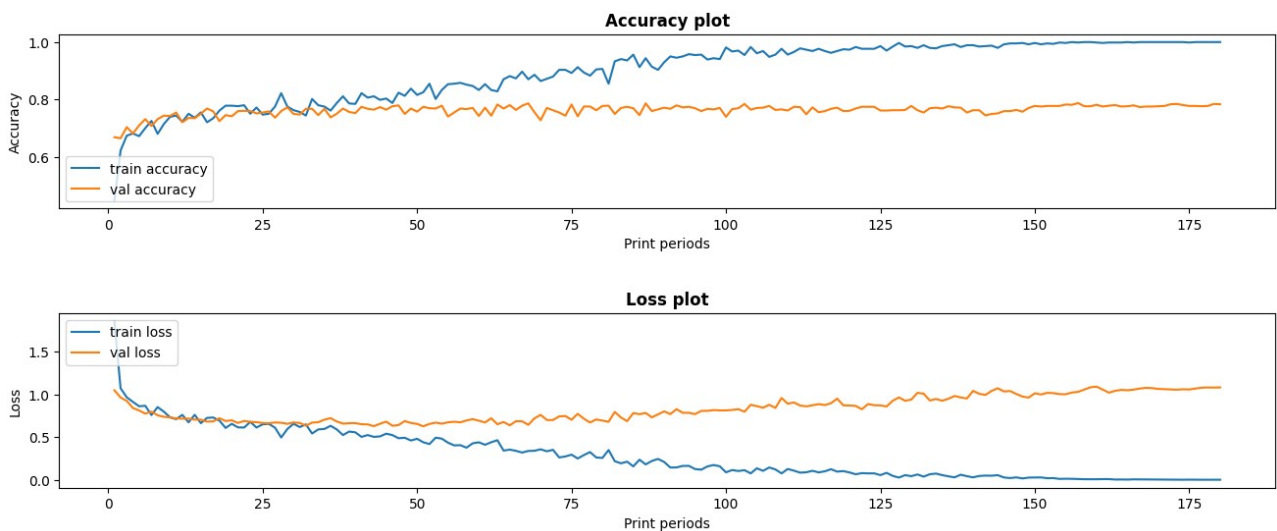
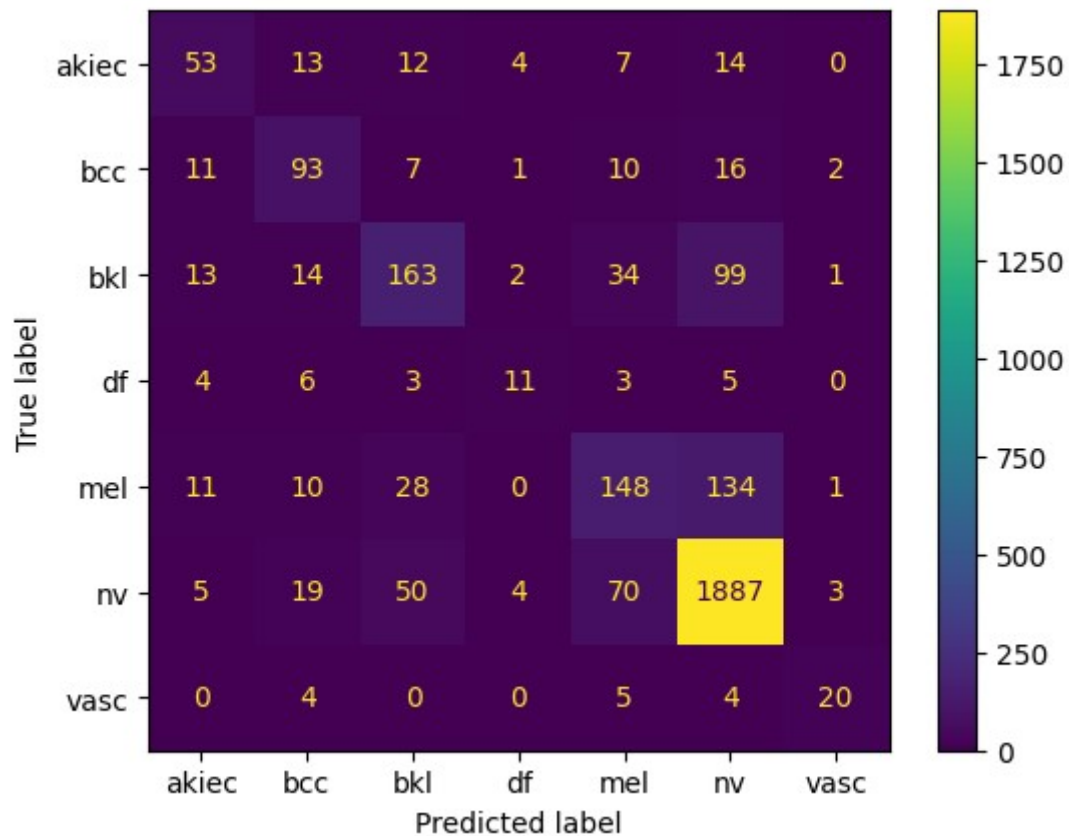




Όπως φαίνεται από τα πάνω διαγράμματα η ευστοχία αυξάνεται και η απώλεια μειώνεται συνεχώς μέχρι το τέλος της εκπαίδευσης του μοντέλου και για δεδομένα εκπαίδευσης και για τα δεδομένα επικύρωσης. Επίσης η ευστοχία και η απώλεια βρίσκονται σε παρόμοιο επίπεδο και για τα δύο κομμάτια δεδομένων (εκπαίδευσης και επικύρωσης). Στα δεδομένα δοκιμής, το ποσοστό ευστοχίας που επιτυγχάνει το μοντέλο, είναι συγκριτικά κοντά στο ποσοστό ευστοχίας για τα δεδομένα επικύρωσης που παρατηρούμε στο διάγραμμα προς το τέλος της εκπαίδευσης. Από αυτά συμπεραίνουμε ότι μοντέλο έχει εκπαιδευτεί σωστά, αφού δεν υπάρχουν εμφανή σημάδια υποεκπαίδευσης ή υπερεκπαίδευσης. Παρ'όλα αυτά υπάρχει το ενδεχόμενο να μπορούσε να εκπαιδευτεί το μοντέλο για λίγες περισσότερες εποχές, χωρίς να υπερεκπαιδευτεί όμως, ώστε να έχει λίγο καλύτερα ακόμα αποτελέσματα.

Σύνθετο ΣΝΔ:

Το σύνθετο συνελικτικό μοντέλο επιτυγχάνει στα δεδομένα δοκιμής ευστοχία 0.79% και απώλεια 0.97. Η ευστοχία είναι λίγο καλύτερη και η απώλεια λίγο χειρότερη από το απλό μοντέλο αλλά δεν υπάρχει κάποια σημαντική διαφορά. Όπως και για το προηγούμενο μοντέλο ακολουθεί ο πίνακας σύγχυσης και τα γραφήματα με τις ενδιάμεσες τιμές της απώλειας και της ευστοχίας κατά την εκπαίδευση:

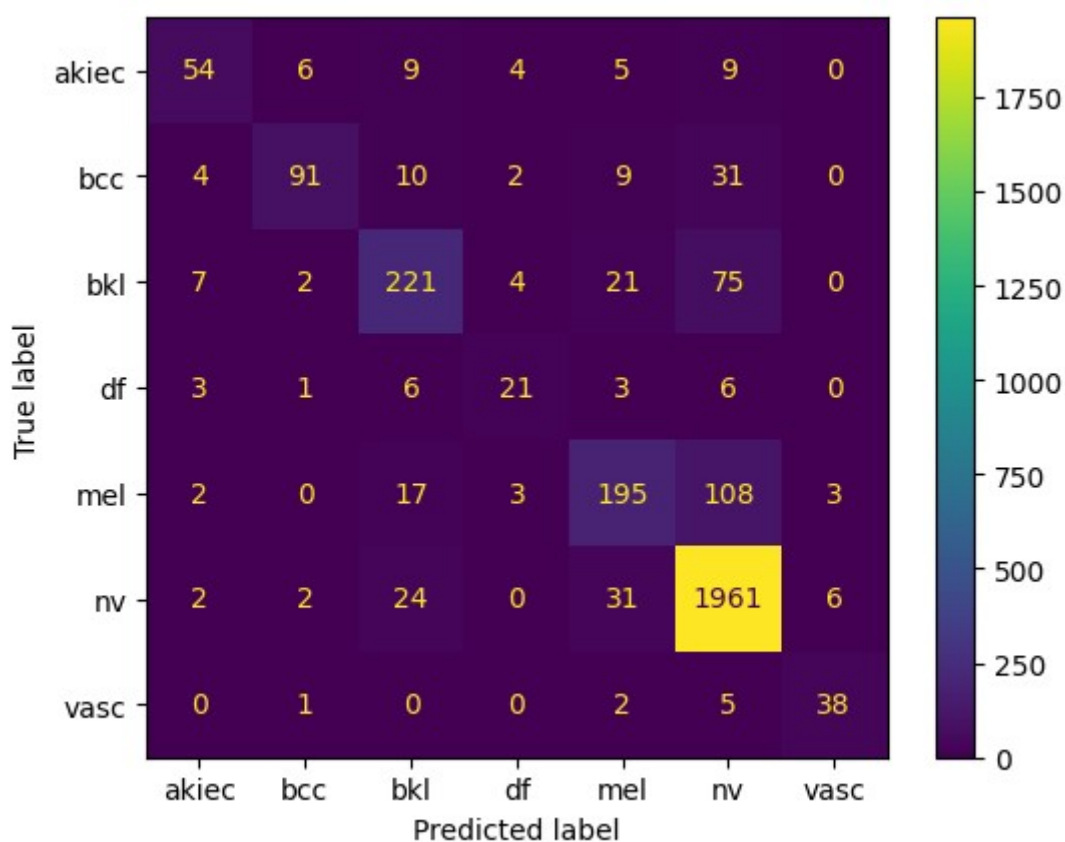


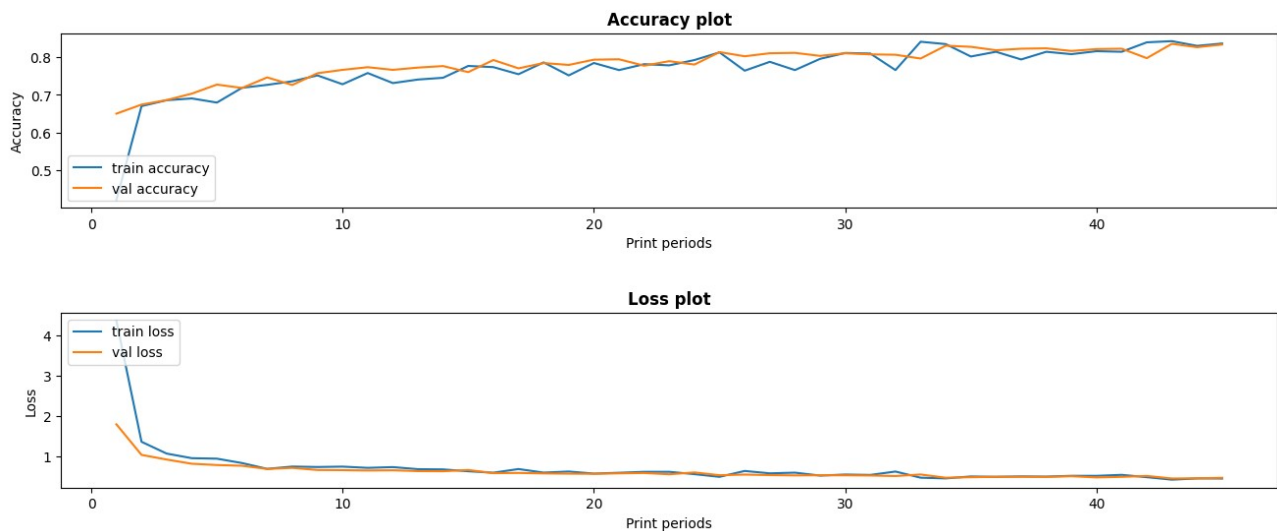
Από το πρώτο διάγραμμα που φαίνεται παραπάνω η ευστοχία ξεκινάει χαμηλά στην αρχή της εκπαίδευσης και αρχίζει να ανεβαίνει σταδιακά και για τα δεδομένα της εκπαίδευσης και για τα δεδομένα της επικύρωσης. Από ένα σημείο και μετά όμως (κοντά στο $\sim 1/4 - 1/3$ της συνολικής εκπαίδευσης) παρατηρείται ότι η ευστοχία στα δεδομένα επικύρωσης αρχίζει να μένει σχεδόν σταθερή και η γραμμή της ευστοχίας να κινείται ασυμπτωτικά κοντά στο $\sim 0.75 - 0.80$ ενώ η ευστοχία στα δεδομένα εκπαίδευσης συνεχίζει να ανεβαίνει και προχωράει προς το τέλος ασυμπτωτικά στο 1 (δηλαδή σχεδόν άριστη ευστοχία για τα δεδομένα εκπαίδευσης). Δίνοντας βάση

στο διάγραμμα της απώλειας βλέπουμε ότι μέχρι περίπου το ίδιο σημείο που αναφέρθηκε πριν, μειώνεται η απώλεια με παρόμοιο και ανάλογο τρόπο και για τα δύο κομμάτια δεδομένων (εκπαίδευσης και επικύρωσης). Από το σημείο αυτό και μετά όμως, η απώλεια στα δεδομένα εκπαίδευσης πέφτει μέχρι που φτάνει κοντά στο 0 ενώ η απώλεια στα δεδομένα επικύρωσης αρχίζει σταδιακά να αυξάνεται αντί να μειώνεται. Σύμφωνα με τις παραπάνω παρατηρήσεις από τα διαγράμματα, συμπεραίνουμε ότι το μοντέλο έχει αρχίσει να υπερεκπαιδεύεται από το ~1/3 περίπου της διάρκειας εκπαίδευσης και μετά. Γι'αυτό κιόλας η μετρική της απώλειας χειροτερεύει (αυξάνεται) και η ευστοχία μένει στάσιμη (δηλαδή δεν βελτιώνεται) για τα δεδομένα επικύρωσης από το σημείο αυτό και ύστερα, ενώ αντίθετα οι μετρικές της ευστοχίας και της απώλειας για τα δεδομένα της εκπαίδευσης αρχίζουν να βελτιώνονται σχεδόν μέχρι την τελειότητα. Αυτό σημαίνει ότι παρόλο που στα δεδομένα δοκιμής έχουμε σχετικά καλά αποτελέσματα (ευστοχία 0.79%), θα έπρεπε προτιμότερα να κάνουμε κάτι για να αποφύγουμε την υπερεκπαίδευση και να βελτιώσουμε το μοντέλο, όπως για παράδειγμα να μειώσουμε τις εποχές εκπαίδευσης, να μειώσουμε το ρυθμό εκμάθησης μετά από κάποιες εποχές, να βάλουμε ίσως momentum ή κάποιο συνδυασμό αυτών των αλλαγών.

Μεταφορά μάθησης - transfer learning:

Το μοντέλο resnet34 μετά από τις 5 εποχές εκπαίδευσης επιτυγχάνει στα δεδομένα δοκιμής 0.85% ευστοχία και 0.39 απώλεια. Από τα τρία μοντέλα που δοκιμάστηκαν με τα ίδια δεδομένα αυτό φαίνεται να είναι το καλύτερο με βάση το ποσοστό ευστοχίας. Ακολουθούν πάλι ο πίνακας σύγχυσης και τα γραφήματα με τις ενδιάμεσες τιμές της απώλειας και της ευστοχίας κατά την εκπαίδευση:





Η εκπαίδευση αυτού του μοντέλου έγινε για 5 εποχές σύμφωνα με τις οδηγίες της εκφώνησης αφού το μοντέλο ήταν προεκπαιδευμένο και χρειαζόταν απλά μία προσαρμογή. Επίσης, το μοντέλο εκπαιδεύτηκε με batch 32 αντί για 128 που εκπαιδεύτηκαν τα άλλα δύο μοντέλα (για λόγους μνήμης), οπότε προκειμένου η αναφορά της ευστοχίας και της απώλειας για τα γραφήματα να γίνεται μετά από το πέρασμα ίδιου αριθμού δειγμάτων των δεδομένων για εκπαίδευση, το print period έγινε 20 από 5 που ήταν για τα άλλα μοντέλα. Από τα γραφήματα βλέπουμε ότι η ευστοχία ανεβαίνει μέχρι το τέλος της εκπαίδευσης με παρόμοιο τρόπο και για τα δεδομένα της εκπαίδευσης και για τα δεδομένα της επικύρωσης και αντίστοιχα η απώλεια μειώνεται ομαλά και με σχεδόν ίδιο τρόπο και για τα δύο κομμάτια δεδομένων. Συγκρίνοντας τα αποτελέσματα στα δεδομένα δοκιμής με τα γραφήματα της εκπαίδευσης δεν φαίνεται ότι υπάρχει υπερεκπαίδευση ή υποεκπαίδευση του μοντέλου καθώς οι μετρικές της ευστοχίας και της απώλειας για τα δεδομένα δοκιμής έχουν παρόμοιο μέγεθος με τις γραμμές της ευστοχίας και της απώλειας στο τέλος των γραφημάτων (τέλος της εκπαίδευσης). Μία διαφορά που φαίνεται αν συγκρίνουμε τα γραφήματα των άλλων δύο μοντέλων, και ιδιαίτερα του πρώτου, με αυτό εδώ είναι ότι οι καμπύλες της ευστοχίας και της απώλειας για το resnet34 είναι αρκετά πιο ομαλές σε σχέση με αυτές των άλλων, δηλαδή δεν υπάρχουν απότομα ανεβάσματα και κατεβάσματα. Αυτό πιστεύω ότι οφείλεται στο momentum που βάλαμε στο resnet34 που σημαίνει ότι οι αλλαγές κατά την εκπαίδευση είναι μικρότερες άρα και η ευστοχία και συνεπώς και η απώλεια δεν θα έχουν απότομες διακυμάνσεις κατά τη διάρκεια της εκπαίδευσης, κάτι που δεν γινόταν στα προηγούμενα δύο μοντέλα.