



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΙΩΑΝΝΙΝΩΝ

ΣΧΟΛΗ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ Τ.Ε.

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**ΑΝΑΠΤΥΞΗ ΕΦΑΡΜΟΓΗΣ ΟΠΤΙΚΗΣ ΑΝΑΓΝΩΡΙΣΗΣ
ΧΑΡΑΚΤΗΡΩΝ ΓΙΑ ΑΛΦΑΡΙΘΜΗΤΙΚΑ ΔΕΔΟΜΕΝΑ ΜΕ ΧΡΗΣΗ
ΤΕΧΝΙΚΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ**

Ηλίας – Ιωάννης Σαδικάκης

Επιβλέπων: Σταύρος Αδάμ

Τίτλος, βαθμίδα

Άρτα, **μήνας έκδοσης**, 2021

**DEVELOPMENT OF AN OPTICAL CHARACTER RECOGNITION
APPLICATION FOR ALPHANUMERIC DATA USING MACHINE
LEARNING TECHNIQUES**

Εγκρίθηκε από τριμελή εξεταστική επιτροπή

Τόπος, Ημερομηνία

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Επιβλέπον καθηγητής

Όνομα Επίθετο,

2. Μέλος επιτροπής

Όνομα Επίθετο,

3. Μέλος επιτροπής

Όνομα Επίθετο,

© Σαρικάκης Ηλίας – Ιωάννης, 2021.

Με επιφύλαξη παντός δικαιώματος, All rights reserved.

Δήλωση μη λογοκλοπής

Δηλώνω υπεύθυνα και γνωρίζοντας τις κυρώσεις του Ν. 2121/1993 περί Πνευματικής Ιδιοκτησίας, ότι η παρούσα πτυχιακή εργασία είναι εξ ολοκλήρου αποτέλεσμα δικής μου ερευνητικής εργασίας, δεν αποτελεί προϊόν αντιγραφής ούτε προέρχεται από ανάθεση σε τρίτους. Όλες οι πηγές που χρησιμοποιήθηκαν (κάθε είδους, μορφή και προέλευσης) για τη συγγραφή της περιλαμβάνονται στη βιβλιογραφία.

Σαρικάκης Ηλίας – Ιωάννης

ΕΙΚΟΝΑ ΥΠΟΓΡΑΦΗΣ

ΕΥΧΑΡΙΣΤΙΕΣ

Κείμενο Ευχαριστιών

ΠΕΡΙΛΗΨΗ

ΚΕΙΜΕΝΟ ΠΕΡΙΛΗΨΗΣ

Λέξεις-κλειδιά: 3 – 5 λέξεις ή φράσεις κλειδιά στα ελληνικά που περιγράφουν το θέμα

ABSTRACT

Κείμενο περίληψης στα Αγγλικά

Keywords: 3-5 λέξεις ή φράσεις στα Αγγλικά που περιγράφουν το θέμα

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΧΑΡΙΣΤΙΕΣ.....	6
ΠΕΡΙΛΗΨΗ.....	7
ABSTRACT	8
ΠΕΡΙΕΧΟΜΕΝΑ	9
ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ.....	11
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ.....	12
ΕΙΣΑΓΩΓΗ.....	13
1. ΔΕΔΟΜΕΝΑ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑ.....	14
1.1 ΟΡΙΣΜΟΣ	14
1.2 ΕΙΔΗ ΔΕΔΟΜΕΝΩΝ	15
1.3 ΤΡΟΠΟΙ ΣΥΛΛΟΓΗΣ ΔΕΔΟΜΕΝΩΝ.....	16
1.4 ΕΞΑΓΩΓΗ ΠΛΗΡΟΦΟΡΙΑΣ	16
1.5 ΑΞΙΟΠΟΙΗΣΗ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ.....	18
2. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	19
2.1 ΤΙ ΕΙΝΑΙ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ.....	19
2.2 ΤΡΟΠΟΣ ΛΕΙΤΟΥΡΓΙΑΣ.....	19
2.2.1 Κρυφές Στρώσεις	20
2.2.2 Βάρη και Bias	21
2.2.3 Συνάρτηση Ενεργοποίησης	22
2.3 ΠΩΣ ΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΕΞΑΡΤΩΝΤΑΙ ΑΠΟ ΤΑ ΔΕΔΟΜΕΝΑ.....	26
2.4 ΕΙΔΗ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ	28
2.5 ΤΡΟΠΟΙ ΑΞΙΟΠΟΙΗΣΗΣ	32
3. ΟΠΤΙΚΗ ΑΝΑΓΝΩΡΙΣΗ ΧΑΡΑΚΤΗΡΩΝ	33
3.1 ΙΣΤΟΡΙΑ.....	33
3.2 ΠΛΕΟΝΕΚΤΗΜΑΤΑ	36

3.3 ΕΦΑΡΜΟΓΕΣ	38
3.4 ΚΑΤΗΓΟΡΙΕΣ.....	40
3.5 ΤΕΧΝΙΚΕΣ	42
4. ΥΛΟΠΟΙΗΣΗ ΕΦΑΡΜΟΓΗΣ	45
4.1 Η ΣΥΛΛΗΨΗ ΤΗΣ ΙΔΕΑΣ	45
4.2 DATASETS	46
4.2.1 Γεννήτρια παραγωγής εικόνων	48
4.2.2 Ενίσχυση εικόνων	51
4.3 ΚΑΤΑΣΚΕΥΗ ΜΟΝΤΕΛΟΥ.....	53
4.3.1 Το μοντέλο.....	54
4.3.2 Datagen, Ανακλητές, Μεταγλώττιση	57
4.3.3 Αξιολόγηση του μοντέλου.....	60
4.4 ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ	62
4.5 ΓΡΑΦΙΚΗ ΔΙΕΠΑΦΗ	62
4.6 ΑΠΟΤΕΛΕΣΜΑΤΑ.....	63
4.7 ΒΕΛΤΙΩΣΕΙΣ	63
ΠΑΡΑΡΤΗΜΑ Α – ΠΗΓΑΙΟΣ ΚΩΔΙΚΑΣ ΕΦΑΡΜΟΓΗΣ.....	63
ΑΝΑΦΟΡΕΣ	64

ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ

Εικόνα 1: Βαθύ νευρωνικό δίκτυο.	20
Εικόνα 2: Βηματική συνάρτηση.....	23
Εικόνα 3: Σιγμοειδής συνάρτηση.....	24
Εικόνα 4: Συνάρτηση ReLU.....	25
Εικόνα 5: Συναρτήσεις Leaky ReLU και Parametric ReLU.....	26
Εικόνα 6: Ένα σχεδόν ολοκληρωμένο διάγραμμα των νευρωνικών δικτύων.....	29
Εικόνα 7: Perceptron και δίκτυο εμπρόσθιας τροφοδοσίας.....	30
Εικόνα 8: Επαναλαμβανόμενο νευρωνικό δίκτυο.	31
Εικόνα 9: Συνελικτικό νευρωνικό δίκτυο.	32
Εικόνα 10: Ένα μικρό δείγμα από το Mnist dataset.....	46
Εικόνα 11: Ένα μικρό δείγμα από την γεννήτρια εικόνων.	47
Εικόνα 12: Πριν και μετά την συνάρτηση Augment_Image	53
Εικόνα 13: Τρόπος λειτουργίας του Conv2D.....	56
Εικόνα 14: Γραφική αναπαράσταση της λειτουργίας MaxPooling2D.	56
Εικόνα 15: Γραφική αναπαράσταση αποτελεσμάτων Πίνακα 9.	62

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1: Ορισμός γραμματοσειρών.....	48
Πίνακας 2: Σύλλογή χαρακτήρων.....	48
Πίνακας 3: Γεννήτρια εικόνων.....	50
Πίνακας 4: Συνάρτηση Augment_Image.....	52
Πίνακας 5: Το διαδοχικό μοντέλο βαθιάς μάθησης της εφαρμογής.....	55
Πίνακας 6: Datagen, train_set και validation_set.....	58
Πίνακας 7: Οι ανακλητές του προγράμματος.....	59
Πίνακας 8: Μεταγλώττιση του μοντέλου.....	60
Πίνακας 9: Αποτελέσματα εκπαίδευσης αριθμητικών χαρακτήρων για 10 εποχές	61

ΕΙΣΑΓΩΓΗ

WIP

1. ΔΕΔΟΜΕΝΑ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑ

Βρισκόμαστε σε μια εποχή που θα μπορούσε να χαρακτηριστεί εύκολα ως η χρυσή εποχή της πληροφορίας. Με 5,27 δισεκατομμύρια ανθρώπους (67,1% του συνολικού πληθυσμού) να κατέχουν ένα κινητό τηλέφωνο, 4,72 δισεκατομμύρια χρήστες του διαδικτύου, και 4,33 δισεκατομμύρια ενεργούς χρήστες στα μέσα κοινωνικής δικτύωσης (datareportal, 2021), θα μπορούσε κανείς να συμπεράνει πως υπάρχει ένας άπειρος όγκος από δεδομένα και πληροφορίες που περιμένουν να συλλεχθούν και να αναλυθούν. Στο τρέχον κεφάλαιο γίνεται μια απόπειρα εξήγησης του τρόπου με τον οποίο τα δεδομένα συλλέγονται, αναλύονται και μετατρέπονται σε πληροφορίες.

1.1 ΟΡΙΣΜΟΣ

Δεδομένα και πληροφορίες. Δύο λέξεις που θα μπορούσαν να χαρακτηριστούν ως όμοιες ή ακόμα και συνώνυμες, παρόλα αυτά υπάρχει μια λεπτή διαφορά μεταξύ των δύο όρων.

Σύμφωνα με το λεξικό της Οξφόρδης, η λέξη «Δεδομένο-α» ('Data' in Lexico Dictionaries, n.d) ορίζεται ως γεγονότα και στατιστικά στοιχεία, τα οποία έχουν συλλεχθεί συνολικά για αναφορά ή ανάλυση. Συγκεκριμένα στον τομέα της πληροφορικής, τα δεδομένα αναφέρονται σε ποσότητες, χαρακτήρες ή σύμβολα, στα οποία πραγματοποιούνται λειτουργίες από έναν υπολογιστή, ενώ ταυτόχρονα είναι δυνατή η αποθήκευση και η μετάδοση αυτών των δεδομένων με την μορφή ηλεκτρικών σημάτων.

Κατά κύριο λόγο, τα δεδομένα βρίσκονται σε μια ακατέργαστη (raw), και ανοργάνωτη (unorganized) μορφή. Μπορεί να περιέχουν απλά γεγονότα ή στοιχεία, τα οποία μπορεί να φανούν τυχαία, ή ακόμα και άχρηστα μέχρι να υποστούν μια οργάνωση ή μια επεξεργασία.

Όταν η επεξεργασία των δεδομένων έχει ολοκληρωθεί και τα δεδομένα βρίσκονται σε μια σωστή δομή, ή παρουσιάζονται επάνω σε κάποιο συγκεκριμένο θέμα, τότε τα δεδομένα μετατρέπονται σε πληροφορία. (Diffen, n.d)

1.2 ΕΙΔΗ ΔΕΔΟΜΕΝΩΝ

Υπάρχουν διάφορα είδη δεδομένων, από γραπτά κείμενα ή αριθμούς σε χαρτί, μέχρι και bits δεδομένων μέσα σε μνήμες ηλεκτρονικών συσκευών. Δεδομένα όμως μπορούν να χαρακτηριστούν ακόμα και οι σκέψεις που βρίσκονται βαθιά μέσα στο μυαλό ενός ανθρώπου. Ορισμένες κύριες κατηγορίες δεδομένων θα μπορούσαν να χαρακτηριστούν οι ακόλουθες: (Bridgwater, 2018)

Μεγάλα Δεδομένα (Big Data):

Τα Big Data ορίζονται ως ένας τεράστιος όγκος δεδομένων, τα οποία συχνά έχουν μέγεθος από μερικά terabytes σε εκατοντάδες zettabytes, ενώ σε πολλές περιπτώσεις η παραγωγή τέτοιου μεγάλου όγκου δεδομένων μπορεί να πραγματοποιηθεί σε πραγματικό χρόνο. Για να μπορέσουμε να κατανοήσουμε το πραγματικό μέγεθος, και την σημασία που έχουν τα μεγάλα δεδομένα, ακολουθούν ορισμένα στατιστικά στοιχεία:

1. Σε χρονικό διάστημα μιας ημέρας πραγματοποιούνται πάνω από 3,5 δισεκατομμύρια αναζητήσεις μέσω της μηχανής αναζήτησης Google. (Internet Live Stats, n.d)
2. Κάθε λεπτό της ημέρας, οι καταναλωτές ξοδεύουν ένα εκατομμύριο δολάρια ηλεκτρονικά, πραγματοποιούν 1,4 εκατομμύρια βιντεοκλήσεις, στέλνουν 150 χιλιάδες μηνύματα στο Facebook και στέλνουν 69 χιλιάδες αιτήσεις εργασίας στο LinkedIn. (Domo, 2020)

Η συγκεκριμένη κατηγορία δεδομένων αποτελεί το «καύσιμο» που τροφοδοτεί την μηχανική μάθηση, και κατά συνέπεια την τεχνητή νοημοσύνη, καθώς οι πολυεθνικές εταιρίες και οι μεγάλες διαδικτυακές υπηρεσίες προσπαθούν να επεξεργαστούν τα μεγάλα αυτά δεδομένα που συλλέγουν από τους χρήστες και πελάτες τους για να μπορέσουν να ανταπεξέλθουν στις ανάγκες και προτιμήσεις των καταναλωτών.

Ανοιχτά Δεδομένα (Open Data):

Τα Ανοιχτά Δεδομένα αποτελούν μια κατηγορία δεδομένων που είναι ελεύθερα σε οποιονδήποτε ενδιαφερόμενο όσον αφορά τα πνευματικά δικαιώματα. Θα μπορούσε δηλαδή ο οποιοσδήποτε να αποκτήσει πρόσβαση στα συγκεκριμένα δεδομένα, να τα επεξεργαστεί, ή ακόμα και να εξάγει συμπεράσματα έπειτα από ανάλυση επάνω σε αυτά.

Δεδομένα Πραγματικού Χρόνου (Real Time Data):

Αποτελούν δεδομένα τα οποία δέχονται επεξεργασία και ανάλυση σε πραγματικό χρόνο από υπολογιστικά συστήματα. Κατά κύριο λόγο, τέτοια δεδομένα μπορούν να αξιοποιηθούν στην ανάλυση και πρόβλεψη της αύξησης ή πτώσης μετοχών στο χρηματιστήριο, ή ακόμα και στην παρακολούθηση της κίνησης σε κάποιον πολυσύχναστο δρόμο.

Επιχειρησιακά Δεδομένα (Operational Data):

Έσος τα σημαντικότερα δεδομένα που θα μπορούσε να έχει μια επιχείρηση. Συχνά περιλαμβάνουν στατιστικά στοιχεία της εταιρίας, πληροφορίες προμηθευτών, δεδομένα λογιστηρίου, ακόμα και πληροφορίες για τις κινήσεις των ανταγωνιστών τους.

1.3 ΤΡΟΠΟΙ ΣΥΛΛΟΓΗΣ ΔΕΔΟΜΕΝΩΝ

Προτού το διαδίκτυο εισαχθεί για τα καλά τόσο στην καθημερινή ζωή μας, όσο και στην ζωή των επιχειρήσεων, διάφορες εταιρίες συνήθιζαν να αξιοποιούν τεχνικές συλλογής δεδομένων που δύσκολα θα συναντούσε κάποιος στις μέρες μας. Συγκεκριμένα, μια επιχείρηση μπορούσε να δημιουργήσει ερωτηματολόγια ή ακόμα και να πάρει συνεντεύξεις από άτομα που βρίσκονται σε μια συγκεκριμένη ομάδα, για την οποία υπάρχει κάποιο ενδιαφέρον από την πλευρά μιας εταιρίας.

Στις μέρες μας μια επιχείρηση μπορεί κυριολεκτικά να συλλέξει τεράστιους όγκους δεδομένων από οπουδήποτε. Χρησιμοποιώντας το διαδίκτυο και εργαλεία ανάλυσης, μία κινητή συσκευή, μια ιστοσελίδα, ακόμα και η δραστηριότητα σε έναν server θα μπορούσαν να χαρακτηριστούν αξιόπιστες πηγές συνεχούς συλλογής πληροφορίας. Φτάνοντας, δηλαδή, στο σημείο συλλογής των Big Data. (McLaughlin, 2020)

1.4 ΕΞΑΓΩΓΗ ΠΛΗΡΟΦΟΡΙΑΣ

Με την ολοκλήρωση της συλλογής δεδομένων από κάποια, ή κάποιες πηγές, τα δεδομένα βρίσκονται σε μια ακατέργαστη μορφή, από την οποία δεν είναι εύκολο να εξάγουμε μια πληροφορία. Για να επιτευχθεί η εξαγωγή της πληροφορίας από ένα σύνολο δεδομένων, θα πρέπει πρώτα να ολοκληρωθούν μια σειρά από βήματα: (Talend, n.d)

1. **Προετοιμασία Δεδομένων (Data Preparation).** Η προετοιμασία δεδομένων είναι το αμέσως επόμενο στάδιο από την συλλογή δεδομένων. Συχνά αποκαλείται ως το στάδιο προ-επεξεργασίας, και αναφέρεται κυρίως στον καθαρισμό και την οργάνωση που δέχονται τα δεδομένα για να εισέλθουν στο επόμενο βήμα. Συγκεκριμένα, τα δεδομένα ελέγχονται εξονυχιστικά για τυχόν σφάλματα, με σκοπό να αφαιρεθούν αχρείαστα ή ελλιπή δεδομένα και να παραμείνουν μόνο τα δεδομένα υψηλής ποιότητας.
2. **Είσοδος Δεδομένων (Data Input).** Τα καθαρισμένα δεδομένα εισάγονται σε ένα σύστημα Διαχείρισης Πελατειακών Σχέσεων (CRM – Customer Relationship Model) ή μια αποθήκη δεδομένων (Data Warehouse), όπου και αρχίζουν να παίρνουν μια πιο ξεκάθαρη μορφή από την οποία μπορεί να ξεκινήσει η εξαγωγή χρήσιμης πληροφορίας.
3. **Επεξεργασία (Processing).** Τα δεδομένα, έχοντας πλέον εισαχθεί σε κάποιο υπολογιστικό σύστημα, βρίσκονται πλέον στο βήμα της επεξεργασίας. Η πραγματοποιείται κυρίως χρησιμοποιώντας αλγορίθμους μηχανικής μάθησης, αν και η ίδια η επεξεργασία μπορεί να διαφέρει ελαφρώς ανάλογα με το είδος, την πηγή των δεδομένων που δέχονται την επεξεργασία, αλλά και από τον τρόπο που επιθυμούμε να αξιοποιήσουμε την εξαγόμενη πληροφορία.
4. **Έξοδος / Διερμηνεύση Δεδομένων (Data Output / Interpretation).** Στο τρέχον στάδιο, τα αρχικά δεδομένα βρίσκονται σε μια ιδιαίτερα απλή μορφή, και η πληροφορία αυτών των δεδομένων μπορεί να γίνει εύκολα κατανοητή από οποιονδήποτε. Συχνά βρίσκονται σε μορφή γραφήματος, εικόνας, κειμένου, ακόμα και βίντεο.
5. **Αποθήκευση Δεδομένων (Data Storage).** Το τελικό στάδιο της επεξεργασίας των δεδομένων είναι η αποθήκευση. Παρόλο που ένα ποσοστό της εξαγόμενης πληροφορίας μπορεί να αξιοποιηθεί άμεσα, ένα σημαντικό ποσοστό ενδέχεται να χρησιμοποιηθεί μελλοντικά. Η σωστή αποθήκευση της πληροφορίας μπορεί να παρέχει άμεση και εύκολη πρόσβαση σε μέλη ενός οργανισμού όποτε αυτό χρειαστεί.

1.5 ΑΞΙΟΠΟΙΗΣΗ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

Για μια εταιρία, μια επιχείρηση ή έναν οργανισμό, η πληροφορία ίσως είναι το πολυτιμότερο στοιχείο που θα μπορούσε να κατέχει. Η πληροφορία θεωρείται ως γνώση, και η γνώση με τη σειρά της θεωρείται δύναμη. Κάθε επιχείρηση έχει έναν στόχο. Το κέρδος. Σε μια εποχή, όπου υπάρχουν εκατομμύρια μικρές και μεγάλες επιχειρήσεις, ο ανταγωνισμός είναι τεράστιος. Για να μπορέσει μια μικρομεσαία επιχείρηση να ανταπεξέλθει, χρειάζεται την δύναμη που της παρέχει η πληροφορία.

Οι εταιρίες χρησιμοποιούν την πληροφορία για να λάβουν ένα σημαντικό πλεονέκτημα απέναντι στους ανταγωνιστές τους, ενώ ταυτόχρονα προσπαθούν να γίνουν πιο ελκυστικοί προς τους πελάτες τους. Παρόλα αυτά, η πληροφορία είναι εξίσου σημαντική για την σωστή και αποτελεσματική λειτουργία της ίδιας της εταιρίας, αξιοποιώντας αυτή τη δύναμη μέσω των ακόλουθων ενεργειών: (Williams, 2019)

1. **Εισαγωγές (Inbound Logistics):** Η διαδικασία της εισόδου πρώτων υλών. Εδώ η πληροφορία μπορεί να αυξήσει την αποτελεσματικότητα των εισαγωγών με συστήματα διαχείρισης εφοδιαστικής αλυσίδας (supply-chain management systems), τα οποία εξυπηρετούν στην σωστή διαχείριση του αποθέματος.
2. **Διεργασίες (Operations):** Οποιοδήποτε κομμάτι της επιχείρησης που συμμετέχει στην μετατροπή πρώτων υλών σε ένα τελικό προϊόν, ή υπηρεσία. Η πληροφορία μπορεί να παρέχει ενίσχυση στην αποτελεσματικότητα των διεργασιών, καθώς επίσης και στην καινοτομία.
3. **Εξαγωγές (Outbound Logistics):** Όπως και στις εισαγωγές, έτσι και στις εξαγωγές, η πληροφορία είναι σε θέση να βελτιώσει την αποτελεσματικότητα της μεταφοράς του τελικού προϊόντος προς τον πελάτη, χρησιμοποιώντας τεχνικές όπως τον έλεγχο αποθέματος σε πραγματικό χρόνο.
4. **Πωλήσεις / Μάρκετινγκ (Sales / Marketing):** Οι λειτουργίες που θα προσελκύσουν τους καταναλωτές στο να αγοράσουν τα προϊόντα μιας επιχείρησης. Η πληροφορία θεωρείται κρίσιμη σε ολόκληρο το κομμάτι των πωλήσεων και του μάρκετινγκ. Από ηλεκτρονικές διαφημίσεις σε ηλεκτρονικές δημοσκοπήσεις, η πληροφορία μπορεί να αξιοποιηθεί για να καινοτομία στον τομέα του σχεδιασμού προϊόντων, και παράλληλα να επιτρέψει στην επιχείρηση να βρίσκεται πιο κοντά στους καταναλωτές τους.

2. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Στον κόσμο της πληροφορικής, τα νευρωνικά δίκτυα, και γενικώς η μηχανική μάθηση, χαρακτηρίζονται ως το αποτέλεσμα της σύνδεσης μεταξύ του τρόπου σκέψης, κατανόησης και μάθησης ενός ανθρώπινου εγκεφάλου με ένα υπολογιστικό σύστημα. Σε αυτό το κεφάλαιο πραγματοποιείται μια απόπειρα εξήγησης του κόσμου των νευρωνικών δικτύων, από την γενική περιγραφή του τρόπου λειτουργίας τους, μέχρι και ορισμένους από τους τρόπους αξιοποίησης τους.

2.1 ΤΙ ΕΙΝΑΙ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ένα νευρωνικό δίκτυο ορίζεται ως μια σειρά από αλγορίθμους μηχανικής μάθησης, οι οποίοι αποσκοπούν να αναγνωρίσουν κρυμμένες σχέσεις σε ένα σετ δεδομένων μέσω μιας διαδικασίας που μιμείται τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου. Συγκεκριμένα, τα νευρωνικά δίκτυα αναφέρονται σε συστήματα νευρώνων, είτε οργανικής ή τεχνητής φύσεως, τα οποία είναι σε θέση να επιλύσουν προβλήματα χρησιμοποιώντας την μέθοδο δοκιμής και σφάλματος (trial and error), δίχως αυτά να έχουν προγραμματιστεί να ακολουθούν συγκεκριμένες οδηγίες. (Chen, 2020)

2.2 ΤΡΟΠΟΣ ΛΕΙΤΟΥΡΓΙΑΣ

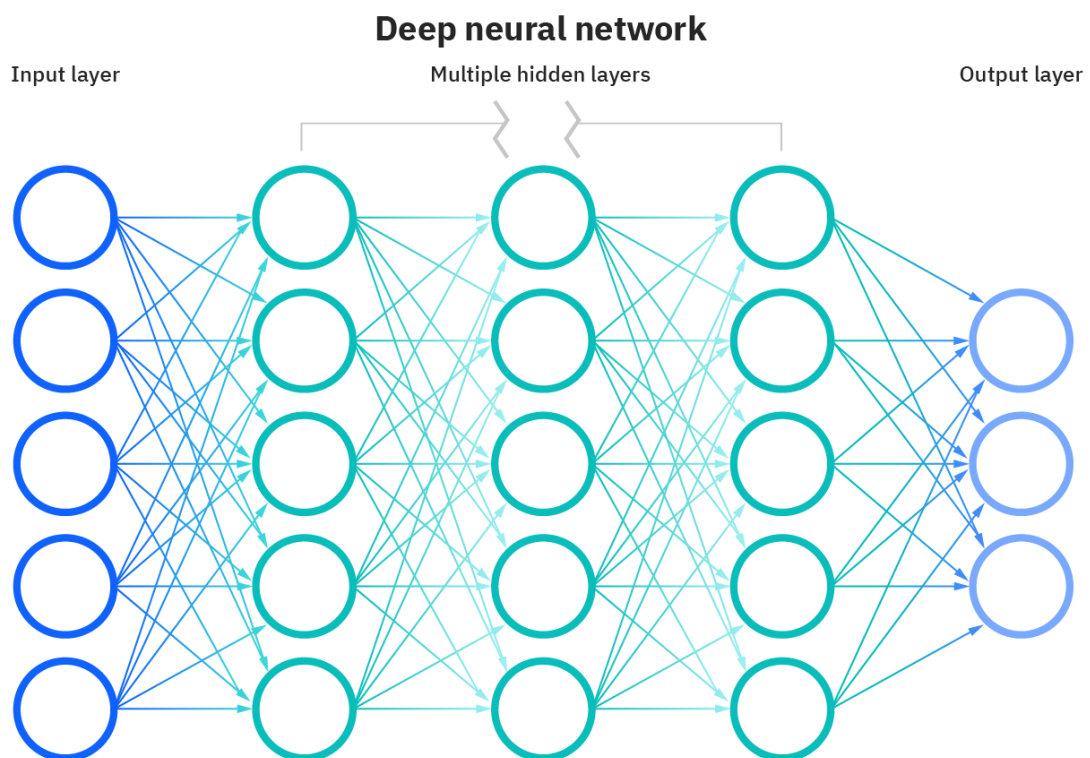
Τα νευρωνικά δίκτυα είναι ένας τρόπος πραγματοποίησης της μηχανικής μάθησης, δηλαδή της διαδικασίας στην οποία ένα υπολογιστικό σύστημα εκπαιδεύεται αναλύοντας εκπαιδευτικά παραδείγματα για να είναι σε θέση να πραγματοποιεί μια εργασία. Συνήθως αυτά τα παραδείγματα έχουν ενταχθεί σε κάποια κατηγορία. Για παράδειγμα, ένα σύστημα αναγνώρισης αντικειμένων μπορεί να έχει αναλύσει χιλιάδες φωτογραφίες από αυτοκίνητα, σπίτια, έπιπλα, κλπ. με αποτέλεσμα να μπορεί να αναγνωρίσει μοτίβα και ακολουθίες σε αυτές τις εικόνες, και στη συνέχεια να είναι σε θέση να τις συσχετίσει με τις κατηγορίες στις οποίες ανήκουν.

Μιας και τα νευρωνικά δίκτυα είναι σχεδιασμένα να μοιάζουν με τον ανθρώπινο εγκέφαλο, ένα δίκτυο μπορεί να αποτελείται από χιλιάδες, ή ακόμα και εκατομμύρια κόμβους, οι οποίοι είναι πυκνά συνδεδεμένοι μεταξύ τους. Στις μέρες μας, τα περισσότερα νευρωνικά δίκτυα είναι οργανωμένα σε στρώσεις (layers) από κόμβους, και

χρησιμοποιούν το μοντέλο της εμπρόσθιας τροφοδοσίας (feed-forward). Αυτό σημαίνει πως τα δεδομένα κινούνται προς μια μόνο κατεύθυνση. Δηλαδή, ένας κόμβος μπορεί να είναι συνδεδεμένος με αρκετούς άλλους κόμβους από μια προηγούμενη στρώση κόμβων για να λαμβάνει δεδομένα, και ταυτόχρονα να συνδέεται με άλλους κόμβους σε μια επόμενη στρώση για να στέλνει δεδομένα. (Hardesty, 2017)

2.2.1 Κρυφές Στρώσεις

Στα νευρωνικά δίκτυα, οι προαναφερόμενες στρώσεις ονομάζονται κρυφές (Hidden Layers) και βρίσκονται μεταξύ της εισόδου και της εξόδου του αλγορίθμου, στις οποίες το μοντέλο εισάγει βάρη (weights) στις εισόδους των κόμβων, και τις προωθεί σε μια συνάρτηση ενεργοποίησης (activation function) ως έξοδο. Με λίγα λόγια, οι κρυφές στρώσεις πραγματοποιούν μη-γραμμικούς μετασχηματισμούς (nonlinear transformations) των εισόδων που έχουν εισαχθεί στο δίκτυο. Αυτές οι στρώσεις ποικίλουν ανάλογα με το μοντέλο του νευρωνικού δικτύου, και ομοίως, μπορούν να ποικίλουν ανάλογα με τα συσχετισμένα βάρη τους.



Εικόνα 1: Βαθύ νευρωνικό δίκτυο.

Θα μπορούσαμε να χαρακτηρίσουμε τις κρυφές στρώσεις ως μαθηματικές συναρτήσεις, οι οποίες είναι σχεδιασμένες να παράγουν συγκεκριμένες εξόδους ανάλογα με το επιθυμητό αποτέλεσμα. Για παράδειγμα, ορισμένες μορφές κρυφών στρώσεων είναι γνωστές ως συναρτήσεις σύνθλιψης (squashing functions). Αυτές οι συναρτήσεις είναι ιδιαίτερα χρήσιμες όταν η επιθυμητή έξοδος του αλγορίθμου είναι πιθανότητα, διότι είναι σε θέση να επεξεργαστούν την είσοδο και να την εξάγουν ως μια τιμή μεταξύ 0 και 1, δηλαδή το εύρος στο οποίο ορίζεται η πιθανότητα.

Κάθε στρώση μπορεί να εξειδικεύεται στην παραγωγή μιας συγκεκριμένης εξόδου. Για παράδειγμα, οι συναρτήσεις μιας στρώσης που χρησιμοποιούνται για να αναγνωρίζουν ανθρώπινα μάτια και αφτιά, μπορούν να χρησιμοποιηθούν σε συνδυασμό με επόμενες στρώσεις για να αναγνωρίσουν πρόσωπα σε εικόνες. (DeerAI a, n.d)

2.2.2 Βάρη και Bias

Στα νευρωνικά δίκτυα τα βάρη (weights) λειτουργούν σαν παράμετροι που τροποποιούν τα δεδομένα εισόδου μέσα στις κρυφές στρώσεις του δικτύου. Κάθε κόμβος περιλαμβάνει την είσοδο, το βάρος του κόμβου, και μια τιμή bias. Η είσοδος που εισέρχεται στον νευρώνα πολλαπλασιάζεται με την τρέχουσα τιμή του βάρους που έχει ο κόμβος. Στη συνέχεια προστίθεται μια τιμή bias, και το τελικό σύνολο αποτελεί την έξοδο του κόμβου. Αυτή η έξοδος μπορεί είτε να παρατηρηθεί από το μοντέλο, ή να προχωρήσει στον επόμενο κόμβο του δικτύου. (DeerAI b, n.d)

Το bias λειτουργεί ως μια έξτρα ώθηση επάνω σε κάθε κόμβο. Μπορεί να έχει μια συγκεκριμένη, σταθερή τιμή σε ολόκληρο το δίκτυο, ή μπορεί να μεταβάλλεται κατά την διάρκεια της εκπαίδευσης, αναλόγως το μοντέλο. Το bias αντιπροσωπεύει πόσο μακριά είναι η πρόβλεψη του κόμβου για μια είσοδο από την πραγματική της τιμή. Ένα χαμηλό bias δηλώνει πως το δίκτυο βρίσκεται πολύ κοντά στην αναμενόμενη έξοδο, ενώ μια υψηλή τιμή δηλώνει πως πιθανώς το δίκτυο δεν πραγματοποιεί τις αναμενόμενες προβλέψεις.

Τα βάρη μπορούν να χαρακτηριστούν ως την δύναμη που έχει ένας κόμβος, μιας και είναι σε θέση να αλλάζουν ριζικά το αποτέλεσμα της εισόδου, και στη συνέχεια να το προωθήσουν προς την έξοδο. Όταν το βάρος είναι χαμηλό, τότε η τιμή της εξόδου θα

παραμένει παρόμοια, ενώ ταυτόχρονα ένα υψηλό βάρος μπορεί ακόμη και να καθορίσει την τελική έξοδο του δικτύου.

2.2.3 Συνάρτηση Ενεργοποίησης

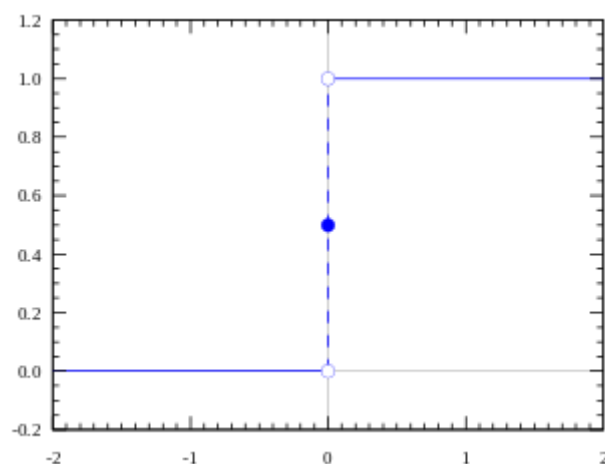
Η συνάρτηση ενεργοποίησης (Activation Function) είναι μια μη-γραμμική μαθηματική συνάρτηση, η οποία ενεργοποιείται επάνω στην είσοδο ενός νευρώνα όταν η τιμή αυτής της εισόδου είναι μεγαλύτερη ενός συγκεκριμένου ορίου, και στη συνέχεια προωθείται στην έξοδο του νευρώνα, και κατά συνέπεια, στην επόμενη στρώση νευρώνων.

Αυτές οι συναρτήσεις είναι ιδιαίτερα χρήσιμες στα νευρωνικά δίκτυα, μιας και η μη-γραμμικότητά τους, επιτρέπει στο δίκτυο να εκπαιδευτεί επάνω σε ισχυρότερες λειτουργίες. Αν αφαιρούσαμε τις συναρτήσεις ενεργοποίησης από ένα νευρωνικό δίκτυο, τότε το δίκτυο θα μετατρεπόταν σε ένα δίκτυο απλής, γραμμικής λειτουργίας. Ως αποτέλεσμα, το δίκτυο δεν θα ήταν πλέον ικανό να πραγματοποιήσει περίπλοκες λειτουργίες, όπως για παράδειγμα η αναγνώριση εικόνων. (Wood, n.d)

Ορισμένες από τις δημοφιλέστερες συναρτήσεις ενεργοποίησης μπορούν να χαρακτηριστούν οι επόμενες, παρόλα μια συνάρτηση μπορεί να φανεί χρησιμότερη από κάποια άλλη, αναλόγως πάντα με το πρόβλημα που επιχειρούμε να λύσουμε χρησιμοποιώντας ένα νευρωνικό δίκτυο.

Βηματική Συνάρτηση (Step Function)

Έσος η πιο απλή μορφή συνάρτησης όταν αναφερόμαστε σε κάποιο όριο. Όταν η τιμή της εισόδου είναι μεγαλύτερη από κάποια συγκεκριμένη τιμή ορίου, τότε η συνάρτηση ενεργοποιείται, δηλαδή βρίσκεται σε τιμή 1. Αντιθέτως, όταν η είσοδος είναι μικρότερη του ορίου, η συνάρτηση παραμένει ανενεργή, δηλαδή 0.

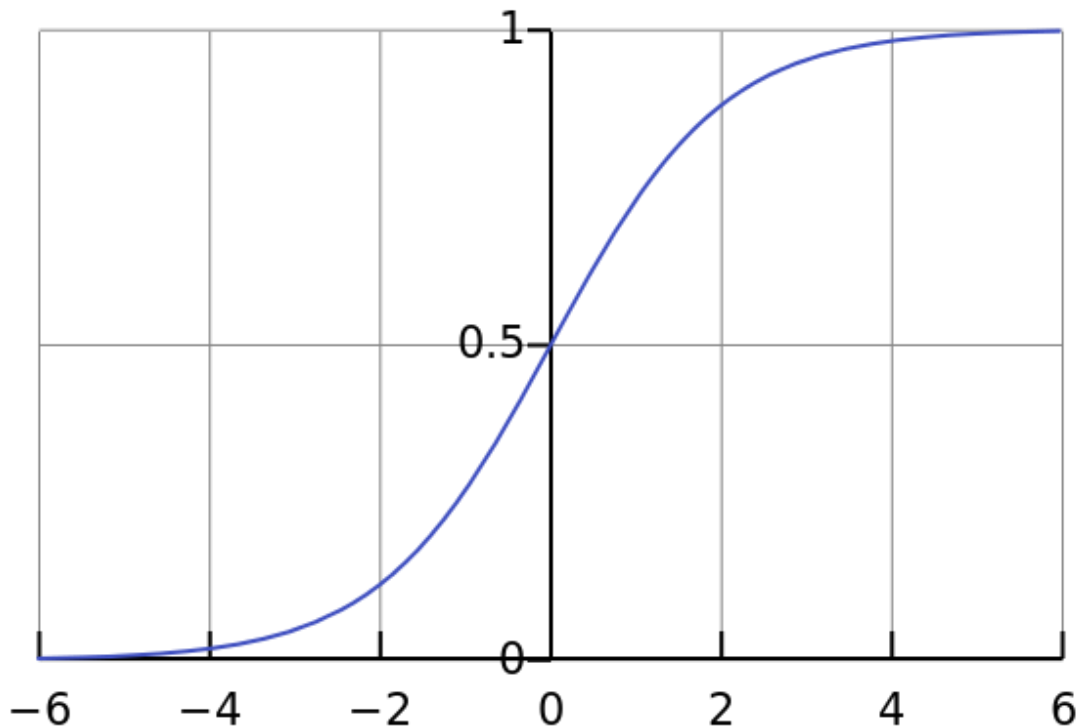


Εικόνα 2: Βηματική συνάρτηση.

Η βηματική συνάρτηση μπορεί να είναι απλή στην χρήση της, παρόλα αυτά είναι αρκετά περιορισμένη ως προς τις δυνατότητές της. Συγκεκριμένα, μια τέτοιου είδους συνάρτηση μπορεί να λειτουργήσει εξαιρετικά σε ένα σύστημα δυαδικής κατηγοριοποίησης (binary classification), όπου η έξοδος μπορεί να είναι μόνο μια από δύο πιθανές κατηγορίες (κατηγορία Α ή κατηγορία Β). Σε συστήματα που επιθυμούμε να κατηγοριοποιήσουμε μια είσοδο σε παραπάνω από δύο κατηγορίες (κατηγορία Α, κατηγορία Β, κατηγορία Γ, κλπ.), τότε τα πράγματα περιπλέκονται όσον αφορά την χρήση της βηματικής συνάρτησης, μιας και υπάρχει η πιθανότητα να ενεργοποιηθούν περισσότεροι από έναν μόνο νευρώνα. Μια είσοδος θα κατηγοριοποιούνταν σε δύο ή ακόμα και περισσότερες κατηγορίες ταυτόχρονα, κάτι που θα έκανε την τελική απόφαση ιδιαίτερα δύσκολη.

Σιγμοειδής Συνάρτηση (Sigmoid Function)

Για να αντιμετωπιστεί το πρόβλημα που παρουσιάζει η βηματική συνάρτηση, δηλαδή την κατηγοριοποίηση σε περισσότερες από δύο κατηγορίες, μπορούμε να χρησιμοποιήσουμε την σιγμοειδή συνάρτηση.



Εικόνα 3: Σιγμοειδής συνάρτηση.

Όντας μια από τις πιο διαδεδομένες συναρτήσεις ενεργοποίησης, η σιγμοειδής συνάρτηση περιλαμβάνει ορισμένες ιδιαιτερότητες που την ξεχωρίζουν από άλλες συναρτήσεις. Μεταξύ των σημείων $X = -2$ και $X = 2$, παρατηρούμε πως η συνάρτηση είναι αρκετά απότομη. Αν μια τιμή X που βρίσκεται μεταξύ αυτών των δύο σημείων, αλλάξει ελαφρώς προς τα δεξιά ή τα αριστερά, θα υπάρξει ιδιαίτερα μεγάλη αλλαγή στο Y της συνάρτησης. Θα μπορούσαμε να πούμε πως η σιγμοειδής συνάρτηση έχει μια τάση να φέρνει τις Y τιμές πιο κοντά στα δύο άκρα της καμπύλης, κάτι που μπορεί να ξεκαθαρίσει την πρόβλεψη της κατηγοριοποίησης.

Ταυτόχρονα, η τιμή της συνάρτησης θα βρίσκεται πάντα μεταξύ των τιμών 0 και 1, ως αποτέλεσμα, στην έξοδο ενός δικτύου μπορούμε να λαμβάνουμε αποτελέσματα της τάξεως 0,21 κατηγορία Α | 0,95 κατηγορία Β | 0,33 κατηγορία Γ για μια είσοδο X (πχ. μια εικόνα), τα οποία αποτελέσματα μπορούν πολύ εύκολα να μετατραπούν σε ποσοστά επί τις εκατό (21% κατηγορία Α, 95% κατηγορία Β, 33% κατηγορία Γ, κλπ.)

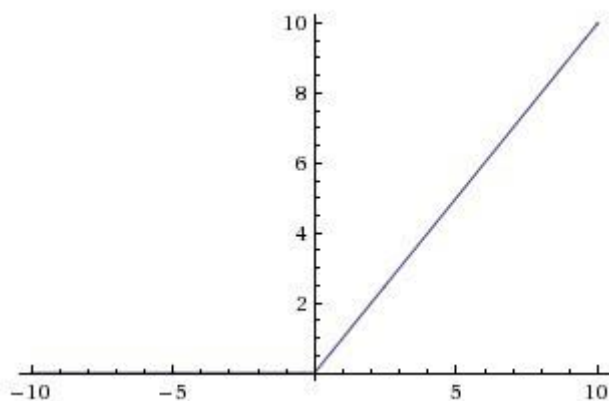
Αν και η σιγμοειδής συνάρτηση είναι αρκετά εύχρηστη, όταν μια τιμή Y βρίσκεται πολύ κοντά στο 0 ή το 1, τείνει να επηρεάζεται ελάχιστα από κάποια αλλαγή στο X , ως

αποτέλεσμα το νευρωνικό δίκτυο να δυσκολεύεται αρκετά κατά την διάρκεια της εκπαίδευσης.

Συνάρτηση ReLU (Rectified Linear Unit)

Σε ένα νευρωνικό δίκτυο εκατοντάδων χιλιάδων ή ακόμα και εκατομμυρίων νευρώνων, αν χρησιμοποιούσαμε μια συνάρτηση ενεργοποίησης όπως την σιγμοειδή, τότε κατά πάσα πιθανότητα θα υπήρχε ενεργοποίηση σχεδόν όλων των νευρώνων. Αυτό σημαίνει πως σχεδόν όλοι οι ενεργοποιημένοι νευρώνες θα δεχόντουσαν επεξεργασία για να υπολογιστεί η έξοδος του δικτύου, κάτι που θα χαρακτήριζε το δίκτυο αρκετά πυκνό.

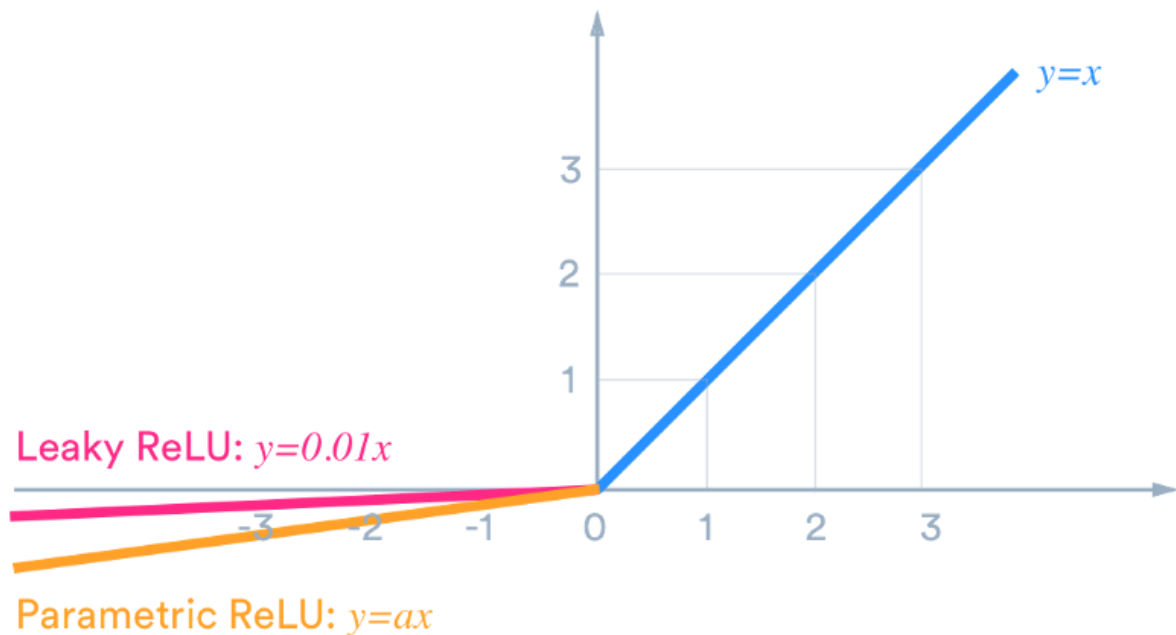
Η μεγάλη πυκνότητα του δικτύου μπορεί να φέρει αρνητικά αποτελέσματα τόσο στην ταχύτητα αλλά και στην αποδοτικότητα του. Ιδανικά επιθυμούμε ένα μέρος των νευρώνων να μην ενεργοποιούνται. Για να επιτευχθεί αυτό, μπορούν να χρησιμοποιηθούν συναρτήσεις όπως η ReLU.



Εικόνα 4: Συνάρτηση ReLU.

Χρησιμοποιώντας την ReLU, όλοι οι νευρώνες που έχουν αρνητική τιμή δεν θα ενεργοποιηθούν, κάτι που σημαίνει πως το δίκτυο γίνεται ελαφρύτερο. Αυτό ακριβώς είναι και το μεγαλύτερο μειονέκτημα της ReLU. Όταν ένας νευρώνας βρίσκεται στην αρνητική πλευρά της συνάρτησης, θεωρείται εξαιρετικά απίθανη η ανάκαμψη του, κάτι που σημαίνει πως ο νευρώνας θεωρείται «νεκρός». Τέτοιοι νευρώνες δεν βοηθούν στην κατηγοριοποίηση μιας εισόδου, και στην ουσία αχρηστεύονται. Με το πέρασμα του χρόνου, ένα μεγάλο ποσοστό του δικτύου θεωρείται πλέον ανενεργό. Συνήθως αυτό το

φαινόμενο παρατηρείται όταν ο ρυθμός εκπαίδευσης (learning rate) του μοντέλου είναι ιδιαίτερα υψηλός ή υπάρχει ένα υψηλό αρνητικό bias. (Sharma, 2017)



Εικόνα 5: Συναρτήσεις Leaky ReLU και Parametric ReLU.

Για να επιλυθεί το πρόβλημα των «νεκρών» νευρώνων, υπάρχουν ορισμένες τροποποιήσεις στην συνάρτηση της ReLU, οι οποίες προσπαθούν να κρατήσουν τις τιμές των νευρώνων μακριά από την τιμή 0. Για παράδειγμα, οι συναρτήσεις Leaky ReLU και Parametric ReLU μπορούν να θεωρηθούν καλές εναλλακτικές της ReLU. (Liu, 2017)

2.3 ΠΩΣ ΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΕΞΑΡΤΩΝΤΑΙ ΑΠΟ ΤΑ ΔΕΔΟΜΕΝΑ

Όπως αναφέραμε στο κεφάλαιο 1. της παρούσας πτυχιακής εργασίας, ο κόσμος μας είναι κυριολεκτικά καταιγισμένος από εξωπραγματικούς όγκους δεδομένων, εκ των οποίων μόνο ένα μικρό ποσοστό είναι σωστά δομημένο και κατηγοριοποιημένο. Αυτό σημαίνει πως μεγάλο μέρος του συνόλου των διαθέσιμων δεδομένων δεν μπορεί να αξιοποιηθεί από τον ένα εκ των δύο κυριότερων μεθόδων εκπαίδευσης νευρωνικών δικτύων.

Αυτές οι δύο μέθοδοι είναι η εκπαίδευση με επίβλεψη (supervised learning) και η εκπαίδευση χωρίς επίβλεψη (unsupervised learning).

Στην εκπαίδευση με επίβλεψη, όλα τα δεδομένα που εισάγονται σε ένα νευρωνικό δίκτυο έχουν ελεγχθεί, οργανωθεί, και κατηγοριοποιηθεί εκ των προτέρων, προτού ξεκινήσει η διαδικασία της εκπαίδευσης. Η κεντρική ιδέα είναι ότι τα υπολογιστικά συστήματα μπορούν να φτάσουν στο σημείο όπου είναι σε θέση να κατανοήσουν τα δεδομένα εντοπίζοντας κοινά στοιχεία σε μέλη μιας συγκεκριμένης κατηγορίας εικόνων, τα οποία δεν εντοπίζονται σε άλλες κατηγορίες. Ο απώτερος σκοπός αυτού του είδους εκπαίδευσης είναι η χρήση αυτών των στοιχείων που έχει εντοπίσει το δίκτυο, έτσι ώστε να ενισχυθεί σημαντικά η ακρίβεια αναγνώρισης και κατηγοριοποίησης εικόνων, τις οποίες το δίκτυο δεν έχει ξαναδεί. Όσο περισσότερες εικόνες επεξεργάζεται ένα δίκτυο κατά την εκπαίδευση του, και όσο μεγαλύτερο και ποικιλόμορφο είναι το σετ δεδομένων (dataset) που αξιοποιείται, τόσο καλύτερη ακρίβεια αναγνώρισης μπορεί να αποκτήσει το νευρωνικό δίκτυο.

Αυτή η τεχνική είναι ιδιαίτερα χρήσιμη στην εκπαίδευση συστημάτων για αναγνώριση οπτικών μέσων, όπως εικόνες, βίντεο, γραφικούς χαρακτήρες, κλπ. Επιπλέον, βρισκόμαστε σε μια εποχή, στην οποία ένα δίκτυο είναι σε θέση να αναγνωρίσει εικόνες εξίσου καλά, ή ακόμα και καλύτερα από ανθρώπους.

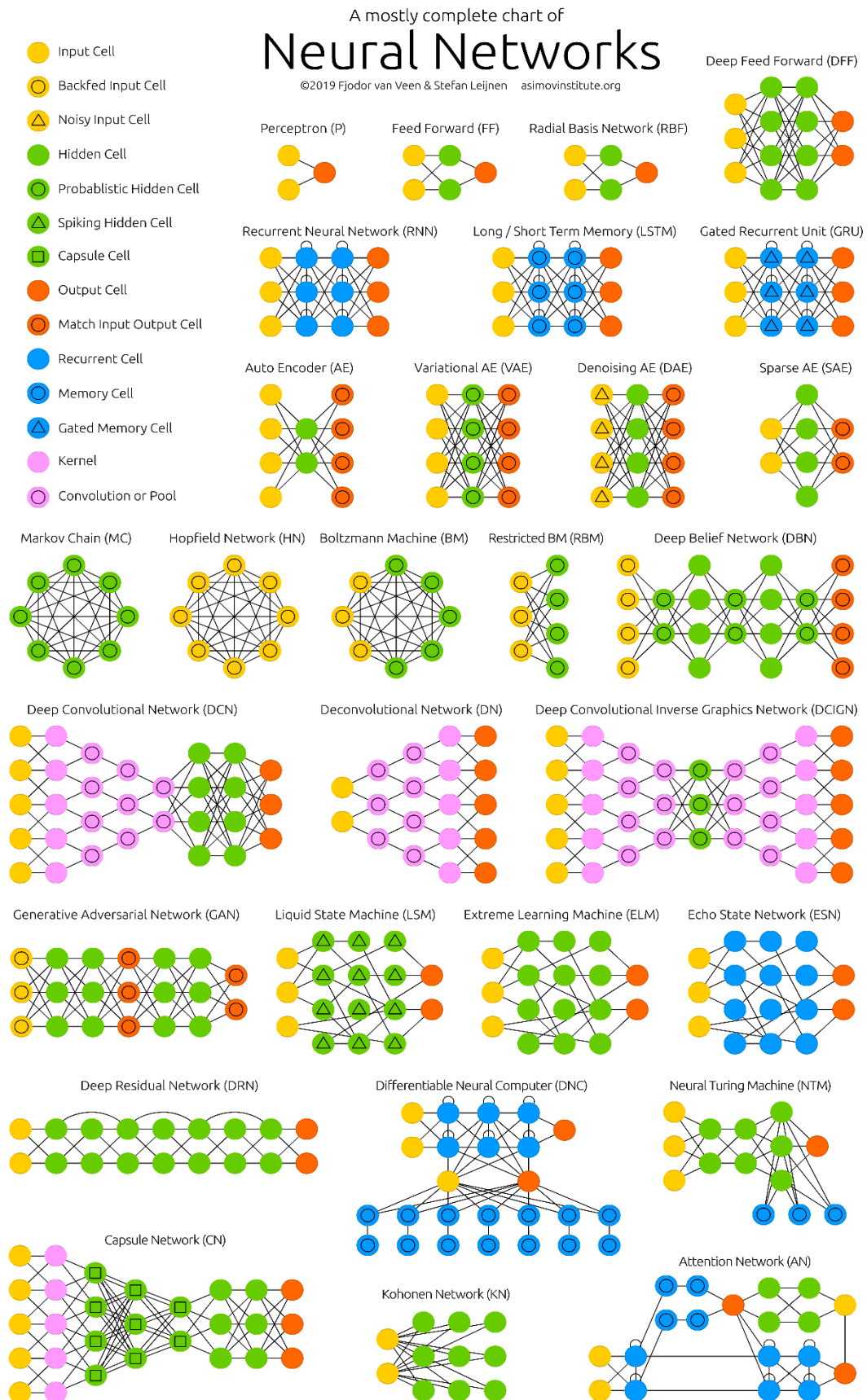
Από την αντίπερα όχθη, η ιδέα της εκπαίδευσης χωρίς επίβλεψη θα μπορούσε να χαρακτηριστεί ως μια πιο αναρχική προσέγγιση στον τρόπο με τον οποίο ένα δίκτυο εκπαιδεύεται. Συγκεκριμένα, η ιδέα είναι ότι τα δεδομένα που εισάγονται στο μοντέλο δεν περιέχουν κάποια ένδειξη κατηγορίας. Ένα μοντέλο, δηλαδή, αφήνεται ελεύθερο να εξερευνήσει το δεδομένα που του έχουν δοθεί, και στη συνέχεια να εντοπίσει από μόνο του πρότυπα και συνδέσεις μεταξύ των δεδομένων. (Tanz & Cambren, 2017)

Έχοντας λοιπόν υπόψιν τους δύο κυριότερους τρόπους εκμάθησης νευρωνικών δικτύων, θα μπορούσαμε πλέον να συμπεράνουμε πως τα νευρωνικά δίκτυα, και ιδιαίτερα τα δίκτυα βαθιάς μάθησης που χρησιμοποιούν την τεχνική της εκπαίδευσης με επίβλεψη, εξαρτώνται σε τεράστιο βαθμό από σωστά κατηγοριοποιημένα δεδομένα για την σωστή τους εκπαίδευση. Το πρόβλημα όμως βρίσκεται στο γεγονός πως το πλήθος αυτών των συγκεκριμένων δεδομένων είναι αρκετά μικρό σε σύγκριση με τον συνολικό όγκος δεδομένων που δημιουργούνται συνεχώς. Ενδέχεται το μέλλον να επιφυλάσσει την αυξημένη χρήση τεχνικών εκπαίδευσης χωρίς επίβλεψη, με αποτέλεσμα να αυξηθεί δραματικά το ποσοστό χρήσης των συνολικών παραγόμενων δεδομένων.

2.4 ΕΙΔΗ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Στις μέρες μας υπάρχουν πολλά, και διαφορετικά είδη νευρωνικών δικτύων που είτε είναι ήδη διαθέσιμα, ή βρίσκονται σε κατάσταση υλοποίησης. Αυτά τα δίκτυα μπορούν να κατηγοριοποιηθούν ανάλογα με την δομή τους, την ροή των δεδομένων, το πλήθος και την πυκνότητα των νευρώνων που χρησιμοποιούνται, τον συνολικό αριθμό κρυφών στρώσεων, κλπ.

Φυσικά, κάθε είδος ή ακόμα και κατηγορία νευρωνικών δικτύων χρησιμοποιείται κάτω από συγκεκριμένες περιπτώσεις, ανάλογα πάντα με το πρόβλημα που επιθυμούμε να επιλύσουμε. Λαμβάνοντας τόσο τα θετικά, όσο και τα αρνητικά στοιχεία κάθε ενός δικτύου, μπορούμε να πάρουμε μια αξιοπρεπής απόφαση όσον αφορά το κατάλληλο είδος δικτύου που θα καλύψει τις ανάγκες μας.



Εικόνα 6: Ένα σχεδόν ολοκληρωμένο διάγραμμα των νευρωνικών δικτύων.

Ορισμένα χαρακτηριστικά είδη νευρωνικών δικτύων θα μπορούσαν να θεωρηθούν τα επόμενα: (Great Learning Team, 2020)

Perceptron και δίκτυα εμπρόσθιας τροφοδοσίας (feed-forward networks)

Το μοντέλο του Perceptron αποτελεί ένα από τα παλαιότερα, και ταυτόχρονα απλούστερα μοντέλα νευρωνικών δικτύων. Το Perceptron μπορεί να χαρακτηριστεί ως ένα απλό είδος δικτύου εμπρόσθιας τροφοδοσίας, και συγκεκριμένα εντάσσεται στην κατηγορία των γραμμικών ταξινομητών. Αυτό σημαίνει πως είναι σε θέση να κατηγοριοποιήσει μια είσοδο σε μια εκ των δύο πιθανών κατηγοριών, διαχωρίζοντας αυτές τις δύο κατηγορίες με μια ευθεία γραμμή. Επιπλέον, έχει την δυνατότητα να υλοποιεί λογικές πύλες (Logic Gates), όπως AND, OR ή NAND



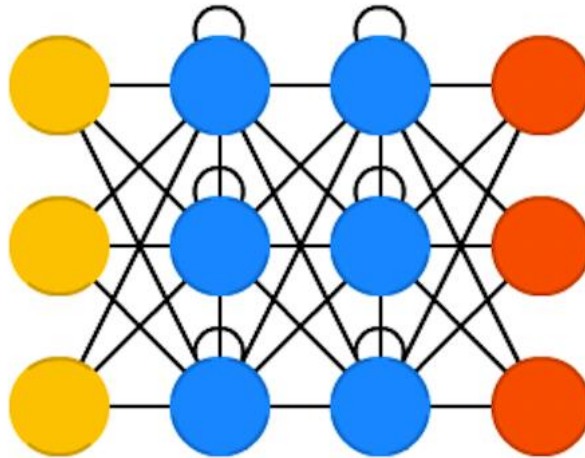
Εικόνα 7: Perceptron και δίκτυο εμπρόσθιας τροφοδοσίας.

Σε ένα δίκτυο εμπρόσθιας τροφοδοσίας, τα δεδομένα κινούνται προς μια μόνο κατεύθυνση, περνώντας μέσα από τεχνικούς νευρωνικούς κόμβους, και στην συνέχεια εξάγονται από κόμβους εξόδου. Ανάλογα με τον αριθμό των κρυφών στρώσεων, μπορούν να χαρακτηριστούν ως δίκτυα μονής στρώσεως ή πολλαπλών στρώσεων. Είναι σε θέση να πραγματοποιεί απλές κατηγοριοποιήσεις, όπου πολλοί παραδοσιακοί αλγόριθμοι ταξινόμησης αποτυγχάνουν, αναγνώριση προσώπων, ακόμα και αναγνώριση φωνής. Παρόλα αυτά δεν είναι έχει την δυνατότητα να χρησιμοποιηθεί σε περιπτώσεις βαθιάς μάθησης (deep learning).

Επαναλαμβανόμενα Νευρωνικά Δίκτυα (Recurrent Neural Networks)

Τα επαναλαμβανόμενα νευρωνικά δίκτυα είναι σχεδιασμένα να αποθηκεύουν την έξοδο του κόμβου, και στην συνέχεια να την ανατροφοδοτούν στην είσοδο για να βοηθήσουν στην σωστή πρόβλεψη της επόμενης εξόδου. Συγκεκριμένα, αποθηκεύει πληροφορίες

που μπορούν να αξιοποιηθούν μελλοντικά, σε μια επόμενη είσοδο δεδομένων. Αν η πρόβλεψη του δικτύου αποδειχθεί πως είναι λανθασμένη, πραγματοποιούνται μικρές αλλαγές στο μοντέλο ανάλογα με το ρυθμό εκπαίδευσης (learning rate), το οποίο κατά συνέπεια ωθεί το δίκτυο να πραγματοποιεί σωστές προβλέψεις.



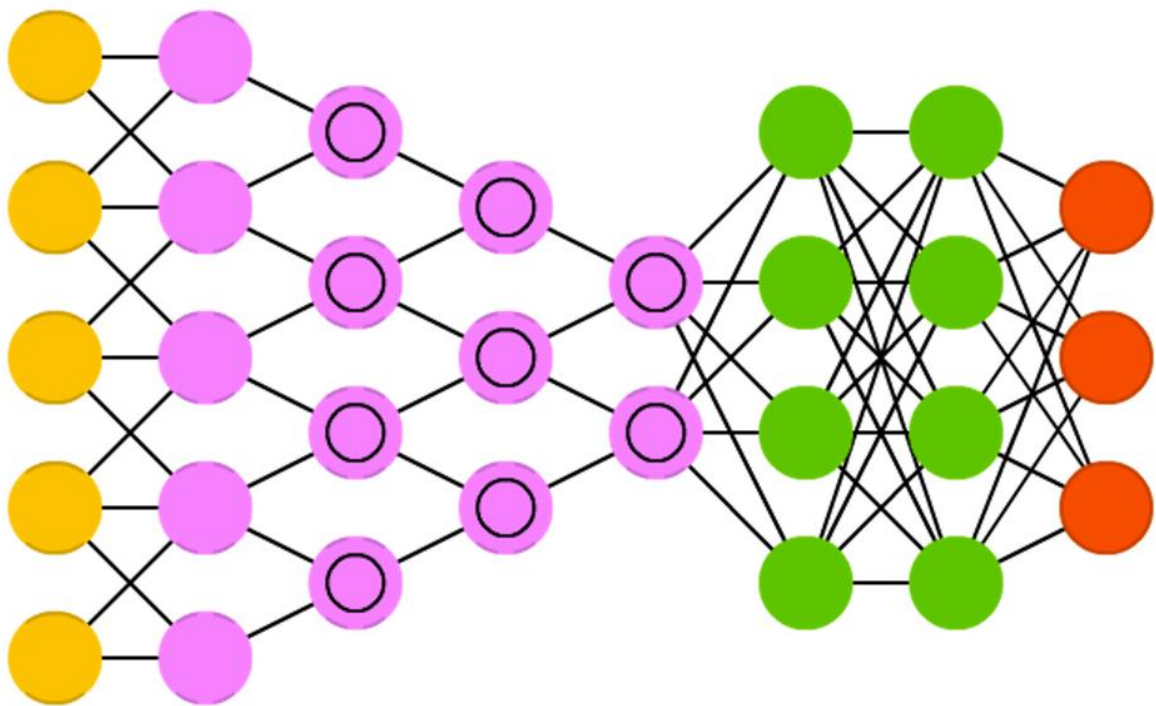
Εικόνα 8: Επαναλαμβανόμενο νευρωνικό δίκτυο.

Τα επαναλαμβανόμενα δίκτυα συχνά χρησιμοποιούνται σε περιπτώσεις όπως η τεχνολογία κειμένου σε φωνή (text-to-speech), μεταφράσεις, και επεξεργασία κειμένου σε πραγματικό χρόνο, όπως οι γραμματικοί έλεγχοι και οι συμβουλές λέξεων.

Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks)

Τα συνελικτικά νευρωνικά δίκτυα διαφέρουν αρκετά σε σύγκριση με δίκτυα άλλων ειδών. Χρησιμοποιούνται κυρίως για αναγνώριση εικόνων, αλλά μπορούν επίσης να χρησιμοποιηθούν για αναγνώριση ήχου, και μηχανική όραση (computer vision). Όταν η είσοδος του δικτύου είναι μια εικόνα, τότε κάθε νευρώνας στο συνελικτικό κομμάτι του μοντέλου υπολογίζει την είσοδο σε τμήματα μικρά τμήματα της συνολικής εισόδου, μέχρι την ολική επεξεργασία της αρχικής εικόνας.

Η επεξεργασία συνήθως περιλαμβάνει την μετατροπή της αρχικής εικόνας από μορφή RGB ή HSI σε κλίμακα του γκρι (gray-scale), μια κίνηση που βοηθάει στον εντοπισμό των άκρων μέσα στην εικόνα, και κατά συνέπεια στην καλύτερη κατηγοριοποίηση της εισόδου.



Εικόνα 9: Συνελικτικό νευρωνικό δίκτυο.

2.5 ΤΡΟΠΟΙ ΑΞΙΟΠΟΙΗΣΗΣ

Όπως έγινε αναφορά στο κεφάλαιο 2.4, υπάρχουν πολλών και διαφορετικών ειδών νευρωνικών δικτύων. Κάθε ξεχωριστό δίκτυο, όχι μόνο περιέχει τα θετικά και τα αρνητικά του στοιχεία, αλλά μπορεί ταυτόχρονα να αξιοποιηθεί σε διαφορετικές καταστάσεις, ανάλογα με τις ανάγκες και τα προβλήματα που επιθυμούμε να επιλύσουμε.

Στον χώρο του μάρκετινγκ, τα νευρωνικά δίκτυα μπορούν να χρησιμοποιηθούν για την βελτίωση των ήδη υπάρχων εργαλείων, παρέχοντας την δυνατότητα της πρόβλεψης της συμπεριφοράς των καταναλωτών, την αυτοματοποίηση του χώρου, και την πρόγνωση πωλήσεων. Ιδιαίτερη χρήση αυτών των δικτύων παρατηρείται στον τομέα της αναλυτικής πρόβλεψης (predictive analytics), καθώς τα νευρωνικά δίκτυα μπορούν να βοηθήσουν στην πρόβλεψη των αποτελεσμάτων μιας επιχειρηματικής εκστρατείας εξετάζοντας τα αποτελέσματα προηγούμενων εκστρατειών. (Brenner, 2018)

Παρόμοια χρήση των νευρωνικών δικτύων παρατηρείται και στον οικονομικό και τραπεζικό τομέα. Δηλαδή χρησιμοποιούνται ιστορικά δεδομένα για την πρόβλεψη της αύξησης ή πτώσης της τιμής των χρηματιστηριακών μετοχών. Ιδιαίτερη χρήση αυτών των δικτύων πραγματοποιείται επίσης στην διαδικασία του ελέγχου των αιτήσεων

δανείων, όπου τα νευρωνικά δίκτυα βοηθούν σε έναν εξαιρετικά μεγάλο βαθμό στην λήψη της τελικής απόφασης όσον αφορά την έγκριση ή απόρριψη ενός δανείου. (Seetharaman, 2018)

Άλλοι εξίσου σημαντικοί τρόποι αξιοποίησης αποτελούν την χρήση των νευρωνικών δικτύων στην ιατρική για την ανίχνευση κρίσιμων νευρολογικών παθήσεων, όπως τα εγκεφαλικά, ή ακόμα και τον εντοπισμό καρκινικών κυττάρων (Bresnick, 2018), στην αναγνώριση χαρακτήρων, εικόνων, φωνής, όπως ακόμα και σε συστήματα αυτόματης οδήγησης.

3. ΟΠΤΙΚΗ ΑΝΑΓΝΩΡΙΣΗ ΧΑΡΑΚΤΗΡΩΝ

Όπως τα νευρωνικά δίκτυα, έτσι και ο κλάδος της οπτικής αναγνώρισης χαρακτήρων είναι εξαιρετικά ευρείς, τόσο από ιστορικής πλευράς, όσο και στην χρήση του στις μέρες μας. Από τις πρώτες μηχανές αναγνώρισης μέχρι και σήμερα έχουν υπάρξει πολλές και διάφορες εφευρέσεις, τεχνικές και αξιοποιήσεις της OCR, κάθε μια με τα δικά της πλεονεκτήματα και μειονεκτήματα, αλλά κάθε μια από αυτές βοήθησε να στρωθεί ένα μονοπάτι για να φτάσει η συγκεκριμένη τεχνολογία στο σημείο όπου βρίσκεται σήμερα. Ως τρίτο κεφάλαιο της παρούσας πτυχιακής εργασίας, πραγματοποιείται μια απόπειρα εξήγησης αυτής της ιδιαίτερα σημαντικής ιστορίας της οπτικής αναγνώρισης με ιδιαίτερη αναφορά στα πλεονεκτήματα και στον τρόπο με τον οποίο αυτή η τεχνολογία αξιοποιεί όλες αυτές τις διάφορες υποκατηγορίες OCR μαζί με ορισμένες εξίσου σημαντικές τεχνικές βελτιστοποίησης.

3.1 ΙΣΤΟΡΙΑ

Η ιστορία της οπτικής αναγνώρισης χαρακτήρων (OCR) πηγάζει από την περίοδο της τηλεγραφίας, και συγκεκριμένα στην ακμή του πρώτου παγκοσμίου πολέμου. Από το 1850 μέχρι και εκείνη την χρονική περίοδο, υπήρχε ένα ιδιαίτερα μεγάλο ενδιαφέρον για την συγκεκριμένη τεχνολογία, μιας και πολύ εφευρέτες έστελναν σε γραφεία πατεντών σχέδια με εφευρέσεις τις εφευρέσεις τους που μπορούσαν να διαβάσουν κείμενα. Παρόλα αυτά, η τεχνολογία της οπτικής αναγνώρισης χαρακτήρων άνθησε ταυτόχρονα με την

τεχνολογία των υπολογιστικών συστημάτων, και απογειώθηκε μέσα από την χρήση της στην επεξεργασία δεδομένων.

Το 1914 περίπου, ο Ισραηλινός φυσικός Emanuel Goldberg κατασκεύασε μια μηχανή, η οποία μπορούσε να διαβάσει χαρακτήρες, και στη συνέχεια τους μετέτρεπε σε κώδικα τηλεγράφου, γνωστός και ως κώδικας μορς, και μετέπειτα μετέδιδε τηλεγραφικά μηνύματα μέσω καλωδίων χωρίς να υπάρχει ανθρώπινη παρέμβαση. Εκείνη την περίοδο, τόσο οι εταιρίες, όσο και οι επιχειρήσεις συνήθιζαν να αποθηκεύουν τα οικονομικά τους στοιχεία σε φιλμ, μια τεχνική που είχε ένα σημαντικό μειονέκτημα στην ανάκτηση αυτών των δεδομένων, καθώς θα έπρεπε πρώτα να γίνει μια χρονοβόρα διαδικασία επάνω στο φιλμ. Για να αντιμετωπίσει αυτό το πρόβλημα, ο Goldberg χρησιμοποίησε φωτοκύτταρα, δηλαδή συσκευές που μετατρέπουν το φως σε ενέργεια, για να πραγματοποιήσει αναγνώριση προτύπων με την βοήθεια ενός προβολέα ταινιών. Βρίσκοντας νέες χρήσεις για ήδη υπάρχοντες τεχνολογίες, ο Goldberg έκανε τα πρώτα βήματα προς την αυτοματοποίηση επιχειρησιακών εγγράφων. (Britton, 2019)

Ταυτόχρονα, ο Dr. Edmund Fournier d'Albe του πανεπιστημίου του Birmingham δημιούργησε και εκείνος μια συσκευή αναγνώρισης χαρακτήρων, η οποία παρήγαγε ήχους. Η συσκευή αυτή ονομάστηκε Optophone, ένα σκάνερ χειρός, που όταν κινούταν επάνω από μία τυπωμένη σελίδα παρήγαγε συγκεκριμένους ηχητικούς τόνους, οι οποίοι συνδέονταν με ένα συγκεκριμένο γράμμα ή σύμβολο. Κατά συνέπεια, άνθρωποι με προβλήματα όρασης μπορούσαν να κατανοήσουν το περιεχόμενο ενός εγγράφου, μαθαίνοντας πρώτα τους χαρακτηριστικούς ήχους των κάθε γραμμάτων.

Το 1951 ο Αμερικανός εφευρέτης David Shepard, μαζί με τον συνάδελφο του Harvey Cook Jr., δημιούργησαν το πρώτο σύστημα οπτικής αναγνώρισης χαρακτήρων, επονομαζόμενο ως Gismo, το οποίο αργότερα μετονομάστηκε σε Analyzing Reader. Το Gismo ήταν μια μηχανή, σχεδιασμένη να μετατρέπει τυπωμένα μηνύματα σε γλώσσα μηχανής για μετέπειτα επεξεργασία από υπολογιστικά συστήματα. Μπορούσε να αναγνωρίσει 23 γράμματα του λατινικού αλφάβητου, όπως τυπώνονταν επάνω στο χαρτί από μια απλή γραφομηχανή. Έπειτα από έναν χρόνο αναβαθμίσεων, η μηχανή μπορούσε πλέον να αναγνωρίζει όλους τους 26 λατινικούς χαρακτήρες, και την ίδια περίοδο ο Shepard έκανε αίτηση πατέντας, η οποία και εγκρίθηκε το 1953 υπ' αριθμόν US 2663758. (Shepard, 1953) Αργότερα, ο Shepard δημιούργησε την δική του επιχείρηση με όνομα Intelligent Machines Research Co. (IMR), όπου και προχώρησε στην

εμπορευματοποίηση του Gismo σε εταιρίες όπως η AT&T, First National City Bank, Reader's Digest, καθώς και σε μεγάλες τράπεζες, και πετρελαϊκές επιχειρήσεις. (History Computer, n.d)

Μια παρόμοια συσκευή με το Optophone δημιουργήθηκε το 1962, όταν ο John Linvill, καθηγητής του πανεπιστημίου του Stanford, κατασκεύασε μια συσκευή με το όνομα Optacon για την κόρη του που είχε χάσει την όραση της από την ηλικία των τριών. Το Optacon αποτελούνταν από την κύρια συσκευή που περιείχε μια ειδική υποδοχή, στην οποία μπορούσε ο χρήστης να εισάγει τα δάχτυλα του ενός χεριού του. Η συσκευή συνδεόταν με ένα σκάνερ χειρός, με το οποίο μπορούσε ο χρήστης να κινηθεί επάνω από ένα έντυπο. Το σκάνερ αναγνώριζε τους χαρακτήρες από το έντυπο και στη συνέχεια έστελνε παλμούς στις άκρες των δακτύλων του χρήστη, δημιουργώντας μια «εικόνα» επάνω στο δάκτυλο, με την οποία ο χρήστης μπορούσε να κατανοήσει το περιεχόμενο ενός κειμένου. (How We Read, n.d)

Ορισμένες άλλες σημαντικές εφευρέσεις και αξιοποιήσεις της οπτικής αναγνώρισης χαρακτήρων μπορούν να χαρακτηριστούν οι ακόλουθες:

- Το 1974 ο Αμερικανός εφευρέτης Ray Kurzweil δημιουργεί το πρώτο λογισμικό που μπορεί να αναγνωρίσει τυπωμένο κείμενο γραμμένο σε οποιαδήποτε γραμματοσειρά.
- Το 1984 η Caere Corporation κατασκευάζει το πρώτο σκάνερ διαβατηρίων για το Αμερικάνικο State Department.
- Το 1992 η Ρωσική εταιρεία Okrus δημιουργεί το πρώτο πρόγραμμα αναγνώρισης του κυριλλικού αλφάβητου.
- Το 2011 δημιουργείται το Google Ngram Viewer για την χαρτογράφηση της συχνότητας των λέξεων από οποιαδήποτε τυπωμένη πηγή μεταξύ του 1950 και του 2008.
- Το 2013 δημιουργείται το dataset MNIST για την εκπαίδευση μοντέλων μηχανικής μάθησης στην αναγνώριση προτύπων.

Στις μέρες μας, υπάρχει μια ιδιαίτερα ευρεία πρόσβαση σε αλγορίθμους και συστήματα οπτική αναγνώρισης χαρακτήρων με την μορφή των APIs, ή ακόμα και στο διαδίκτυο, τα οποία είναι σε θέση να αναγνωρίσουν τους περισσότερους χαρακτήρες και τις περισσότερες γραμματοσειρές με αρκετά μεγάλο επίπεδο ακρίβειας. Αν και η

συγκεκριμένη τεχνολογία εξελίσσεται συνεχώς, το περιθώριο λαθών συνεχίζει να υπάρχει, ένα γεγονός που κρίνει τόσο τον ανθρώπινο έλεγχο, όσο και την ανθρώπινη παρέμβαση άκρως απαραίτητη για την ομαλή λειτουργία τέτοιων συστημάτων.

3.2 ΠΛΕΟΝΕΚΤΗΜΑΤΑ

Οι επιχειρήσεις και οι οργανισμοί μπορούν να χαρακτηριστούν ως οι καλύτεροι πελάτες όσον αφορά τα συστήματα οπτικής αναγνώρισης χαρακτήρων. Με την εξέλιξη της τεχνολογίας των ηλεκτρονικών υπολογιστών, οι εταιρίες ξεκίνησαν να χρησιμοποιούν αυτά τα νέα και προηγμένα συστήματα για να πραγματοποιούν τις καθημερινές τους δραστηριότητες. Σταδιακά, οι επιχειρήσεις ανακάλυψαν τις δυνατότητες που τους παρέχει η ψηφιοποίηση εγγράφων, και ως αποτέλεσμα, ένα σημαντικό ποσοστό των συνολικών δραστηριοτήτων τους ξεκίνησε να πραγματοποιείται με την χρήση ψηφιοποιημένων εντύπων. Παρόλα αυτά, η χρήση των φυσικών εγγράφων παρέμεινε, έως ότου οι ίδιες οι επιχειρήσεις ανακάλυψαν τις δυνατότητες που τους προσέφερε η τεχνολογία της οπτικής αναγνώρισης χαρακτήρων. Ορισμένες από αυτές τις δυνατότητες μπορούν να χαρακτηριστούν οι ακόλουθες:

Μείωση εξόδων λειτουργίας

Η χρήση της οπτικής αναγνώρισης χαρακτήρων για εύκολη και άμεση ψηφιοποίηση των εγγράφων επέτρεψε στις επιχειρήσεις να μετατρέψουν ένα ιδιαίτερα σημαντικό μέρος των συνολικών επιχειρησιακών εντύπων τους σε ψηφιακή μορφή. Δίχως αυτήν την τεχνολογία, ο μόνος τρόπος που θα μπορούσε να πραγματοποιηθεί μια τέτοια ενέργεια θα ήταν μέσα από την μίσθωση ατόμων, ενώ ταυτόχρονα η ίδια η ενέργεια θα έπαιρνε ιδιαίτερα μεγάλο και πολύτιμο χρόνο για να ολοκληρωθεί. Για τις επιχειρήσεις, ο χρόνος είναι χρήμα, όσο χρονοβόρα είναι μια ενέργεια, τόσο λιγότερο κέρδος μπορεί να βγάλει μια εταιρία. Επιπλέον, η ανθρώπινη προσπάθεια που θα αξιοποιούνταν για ένα τόσο μεγάλο εγχείρημα θα μπορούσε κάλλιστα να αξιοποιηθεί σε άλλους, πιο κερδοφόρους τομείς.

Αύξηση αποδοτικότητας και παραγωγικότητας

Για έναν εργαζόμενο, μια θεωρητικά απλή αναζήτηση ενός εντύπου μπορεί να διαρκέσει από μερικά λεπτά, έως και μερικές ώρες, ανάλογα με την ευκολία του εντοπισμού αυτού

του εγγράφου, και την ευκολία πρόσβασης σε αυτό. Αυτός ο χρόνος που αξιοποιείται άσκοπα για ένα μόνο έγγραφο κοστίζει όχι μόνο στην επιχείρηση από πλευράς παραγωγικότητας, αλλά και από τον ίδιο τον εργαζόμενο, καθώς του δημιουργεί ψυχολογική κούραση, και συχνά, ακόμα και εκνευρισμό.

Με την ψηφιοποίηση των επιχειρησιακών εγγράφων, την σωστή αποθήκευση και αρχειοθέτηση τους, ο ίδιος ο προαναφερόμενος εργαζόμενος μπορεί να αποκτήσει εύκολη, γρήγορη και άμεση πρόσβαση στο απαιτούμενο έντυπο μέσα σε λίγα μόνο δευτερόλεπτα, δίχως να υπάρχει η ανάγκη της φυσικής αναζήτησης και εντοπισμού του. Επιπλέον, με την σωστή ψηφιοποίηση, συχνά δίνετε η δυνατότητα επεξεργασίας του ψηφιοποιημένου αρχείου, ανάλογα με την απαιτούμενη περίσταση. Αν δεν υπήρχε η ψηφιοποίηση, μια τέτοια ενέργεια ίσως απαιτούσε την εξ' ολοκλήρου δημιουργία ενός νέου φυσικού εγγράφου, αποσπώντας σημαντικό χρόνο από τον εργαζόμενο και σημαντικό κέρδος από την επιχείρηση.

Βελτίωση προσβασιμότητας

Η ψηφιοποίηση, η σωστή αποθήκευση και αρχειοθέτηση των εγγράφων, μπορεί να βοηθήσει σε έναν μεγάλο βαθμό την αναζήτηση κάποιου εντύπου, ιδιαίτερα σε μέρη όπου η αναζήτηση εγγράφων είναι μια συχνή διαδικασία. Η δυνατότητα πρόσβασης σε ένα έγγραφο από πολλούς εργαζόμενους, η ικανότητα τροποποίησης και αντιγραφής τμημάτων από το περιεχόμενο ενός αρχείου, είναι ένα αποτέλεσμα της δύναμης που παρέχει η ψηφιοποίηση.

Μεγαλύτερη ασφάλεια

Η εύκολη δημιουργία αρχείων ανάκτησης (back up) αποτελεί ένα εξίσου σημαντικό πλεονέκτημα που παρέχεται από την οπτική αναγνώριση χαρακτήρων, και κατά συνέπεια, από την ψηφιοποίηση. Σε περίπτωση που υπάρξει κάποια καταστροφή στον χώρο εργασίας, ένα μεγάλο, ή ακόμα και ολικό, μέρος των φυσικών εγγράφων μπορούν να χαθούν οριστικά, δίχως την επιλογή της ανάκτησης μέρους τους. Ευαίσθητα και άκρως σημαντικά έγγραφα για την ορθή λειτουργία της επιχείρησης μπορούν να αποθηκευτούν σε ηλεκτρονικές βάσεις δεδομένων, και σε συστήματα cloud, μακριά δηλαδή από τον φυσικό χώρο εργασίας, κάνοντας τα άφθαρτα σε οποιαδήποτε καταστροφή.

3.3 ΕΦΑΡΜΟΓΕΣ

Είναι προφανές πως από την πρώτη στιγμή που εφευρέθηκαν οι πρώτες μηχανές οπτικής αναγνώρισης χαρακτήρων, εμφανίστηκε μια ιδιαίτερα σημαντική ανάγκη της αξιοποίησης μιας τέτοιας τεχνολογίας από άτομα με προβλήματα όρασης. Με την πάροδο του χρόνου και της εξέλιξης της τεχνολογίας των υπολογιστικών συστημάτων, η αναγνώριση χαρακτήρων βρέθηκε να αξιοποιείται από πολλούς διαφορετικούς τομείς, οι οποίοι φάνταζαν απίστευτοι από τους πρώτους εφευρέτες τέτοιων μηχανών. Ορισμένοι από αυτούς τους τομείς που εφαρμόζουν τα πλεονεκτήματα της οπτικής αναγνώρισης χαρακτήρων είναι οι ακόλουθοι: (Khurana, et al., 2018)

Τραπεζικός τομέας

Τόσο ο τραπεζικός τομέας, καθώς και άλλα οικονομικά τμήματα, όπως ο χώρος των ασφαλίσεων, είναι ιδιαίτερα σημαντικοί καταναλωτές της οπτικής αναγνώρισης χαρακτήρων. Η συνηθέστερη χρήση αυτής της τεχνολογίας συναντάται στον έλεγχο επιταγών. Το περιεχόμενο μιας χειρόγραφης επιταγής μετατρέπεται σε ψηφιακή μορφή έπειτα από το πέρασμα της από ένα σκάνερ, η εγκυρότητα της υπογραφής επαληθεύεται έπειτα από ταυτοποίηση της μέσω μιας βάσης δεδομένων, και στην συνέχεια επικυρώνεται ή ακυρώνεται σε πραγματικό χρόνο.

Νομικός τομέας

Ο κλάδος της νομικής θεωρείται από τους κλάδους που παράγουν τον μεγαλύτερο όγκο εγγράφων παγκοσμίως. Τεράστιες στοίβες από καταθέσεις ενόρκων, αποφάσεις, διαθήκες, και πολλά άλλα είδη έντυπων νομικών εγγράφων μπορούν να ψηφιοποιηθούν, αποθηκευτούν και αρχειοθετηθούν χρησιμοποιώντας συστήματα OCR. Κάτι που επιτρέπει την απλή και ταχύρρυθμη αναζήτηση και πρόσβαση σε εκατοντάδες χιλιάδες έγγραφα που πηγάζουν από δεκάδες ή ακόμα και εκατοντάδες χρόνια πίσω.

Ιατρική

Σε παρόμοια κατάσταση με τον κλάδο της νομικής βρίσκεται και η ιατρική κοινότητα. Ο όγκος εγγράφων είναι ιδιαίτερα μεγάλος, και ταυτόχρονα γεμάτος με άκρως σημαντικά στοιχεία, όπως το ιατρικό ιστορικό ασθενών, αποτελέσματα διαγνώσεων, νοσοκομειακά στοιχεία, καθώς και πολλά άλλα. Η αποθήκευση όλων αυτών των εγγράφων σε μια ψηφιοποιημένη μορφή, όχι μόνο διευκολύνει στην αναζήτηση και την πρόσβαση τέτοιων

δεδομένων, αλλά ταυτόχρονα βελτιώνει την λογιστική διαδικασία της λειτουργίας του ίδιου του νοσοκομείου, όπως την παραγγελία φαρμακευτικών ειδών, εξοπλισμού, και άλλων προμηθειών.

Διατήρηση ιστορικών εγγράφων

Παλαιές βιβλιοθήκες, ιστορικά πολιτιστικά κέντρα, ακόμα και μουσεία είναι σημεία όπου αποθηκεύονται χιλιάδες ιστορικά χειρόγραφα έντυπα και απομνημονεύματα. Ντοκουμέντα μεγάλης ηλικίας είναι ιδιαίτερα ευάλωτα σε εξωτερικές συνθήκες, και ταυτόχρονα χρειάζονται ιδιαίτερη προσοχή τόσο στην αποθήκευση, όσο και στην διατήρησή τους. Τεχνικές όπως η OCR χρησιμοποιούνται ευρέως για την μετατροπή αυτών των εύθραυστων εγγράφων σε μια ψηφιακή μορφή, αναλλοίωτη στον χρόνο.

Ταυτοποίηση

Στις μέρες μας η ταυτοποίηση αποτελεί μια άκρως σημαντική διαδικασία που πρέπει να πραγματοποιηθεί σε καταστάσεις όπου απαιτούνται τα προσωπικά δεδομένα ενός ατόμου. Πολλές φορές όμως αυτή η διαδικασία μπορεί να αποβεί ιδιαίτερα χρονοβόρα, ειδικά όταν υπάρχει ένα μεγάλο πλήθος ατόμων που πρέπει να περάσουν από την συγκεκριμένη διαδικασία. Τεχνικές οπτικής αναγνώρισης χαρακτήρων μπορούν να αξιοποιηθούν επάνω σε μηχανήματα σκάνερ για τον έλεγχο ταυτοτήτων, διαβατηρίων, διπλωμάτων οδήγησης, καθώς και άλλων εγγράφων ταυτοποίησης που πρέπει να ελεγχθούν με έναν αρκετά γρήγορο ρυθμό από αστυνομικές διευθύνσεις, αεροδρόμια και λοιπούς οργανισμούς για την ορθότητα και εγκυρότητα των προσωπικών δεδομένων.

Άτομα με προβλήματα όρασης

Ιστορικά, η πρώτη χρήση της οπτικής αναγνώρισης χαρακτήρων πραγματοποιήθηκε για την βοήθεια στην ανάγνωση και κατανόηση κειμένων σε άτομα με προβλήματα όρασης. Ακόμα και σήμερα, αυτή η τεχνολογία μπορεί να χρησιμοποιηθεί μαζί με ένα σύστημα κειμένου σε ομιλία (text-to-speech) για τον εντοπισμό κειμένου μέσα από έντυπα βιβλία, έγγραφα, κλπ. Εναλλακτικά, η τεχνολογία της OCR θα μπορούσε να αποτελέσει ένα άκρως σημαντικό κομμάτι σε ένα σύστημα από εκτυπωτές, σχεδιασμένοι να τυπώνουν βιβλία σε μορφή Braille.

3.4 ΚΑΤΗΓΟΡΙΕΣ

Η τεχνολογία της οπτικής αναγνώρισης αποτελείται από διάφορες κατηγορίες, και κατ' επέκταση από διαφορετικές τεχνολογίες. Εκ πρώτης όψεως μπορεί αυτές οι κατηγορίες να φαίνονται αρκετά πανομοιότυπες, παρόλα αυτά κάθε μια από αυτές περιέχει τα δικά της μοναδικά χαρακτηριστικά, εφαρμόζεται κάτω από συγκεκριμένες περιπτώσεις, και τέλος κατέχει τα δικά της πλεονεκτήματα και μειονεκτήματα. Στο δια ταύτα όμως θα μπορούσαμε να συμπεράνουμε πως όλες αυτές οι κατηγορίες εντάσσονται κάτω από την ίδια τεχνολογική οικογένεια. Ορισμένες από τις σημαντικότερες κατηγορίες μπορούν να χαρακτηριστούν οι εξής.

Οπτική αναγνώριση χαρακτήρων (OCR)

Η οπτική αναγνώριση χαρακτήρων (Optical Character Recognition) ή αλλιώς οπτική ανάγνωση χαρακτήρων, χαρακτηρίζεται ως την πιο διαδεδομένη μορφή της οπτικής αναγνώρισης. Είναι μια τεχνολογία που επιτρέπει την μετατροπή εικόνων που περιέχουν κάποια μορφή (συνήθως) τυπωμένου κειμένου σε ηλεκτρονική μορφή, και την αποθήκευση των αποτελεσμάτων μέσα σε ένα υπολογιστικό σύστημα. Αυτή η μετατροπή πραγματοποιείται σχεδόν άμεσα, και μπορεί να παρέχει δυνατότητες όπως την εύκολη και γρήγορη επεξεργασία του αναγνωρισμένου κειμένου, την εύκολη αναζήτηση λέξεων εντός του περιεχομένου του κειμένου, καθώς επίσης και την άμεση εισαγωγή δεδομένων σε φόρμες ή ακόμα και βάσεις δεδομένων.

Ένα σύστημα OCR εκπαιδεύεται επάνω σε ξεχωριστές εικόνες ενός μόνο χαρακτήρα ή συμβόλου κάθε φορά, ως αποτέλεσμα, η αναγνώριση μιας λέξης πραγματοποιείται με την αναγνώριση κάθε γράμματος ξεχωριστά. Αυτό αποτελεί και ένα από τα κυριότερα αρνητικά στοιχεία της παρούσας τεχνολογίας, μιας και σε περίπτωση που υπάρξει αλλοίωση της εικόνας προς αναγνώριση, και δύο χαρακτήρες εμφανίζονται πολύ κοντά ο ένας με τον άλλον, το σύστημα ενδέχεται να αναγνωρίσει αυτούς τους δύο χαρακτήρες ως έναν, με αποτέλεσμα να πραγματοποιηθεί εσφαλμένη έξοδος από το σύστημα.

Μπορούμε λοιπόν να συμπεράνουμε πως η συγκεκριμένη τεχνολογία δεν είναι 100% ακριβής, μιας και μπορούν να υπάρξουν λανθασμένα αποτελέσματα λόγω μη ξεκάθαρων γραμματοσειρών, ή αλλοίωσης της εισόδου. Θεωρείται λοιπόν ιδιαίτερα σημαντικός ο έλεγχος των αποτελεσμάτων για τον εντοπισμό τυχόν λαθών κατά την διαδικασία της

αναγνώρισης, κάτι που μπορεί να αποτελέσει μια ιδιαίτερα χρονοβόρα διαδικασία, ανάλογα με το μέγεθος της εξόδου.

Οπτική αναγνώριση λέξεων (OWR)

Η οπτική αναγνώριση λέξεων (Optical Word Recognition) αποτελεί μια δεύτερη τεχνική αναγνώρισης στην ίδια τεχνολογία της OCR. Στην συγκεκριμένη περίπτωση, η αναγνώριση πραγματοποιείται σε ολόκληρες λέξεις σε γλώσσες όπου υπάρχει ένα εμφανές κενό ενδιάμεσα από κάθε λέξη, έναντι της αναγνώρισης επάνω σε κάθε χαρακτήρα ξεχωριστά. Τις περισσότερες φορές, όταν γίνεται αναφορά σε συστήματα OCR, θεωρείται πως η αναγνώριση πραγματοποιείται σε ολόκληρες λέξεις, παρόλα αυτά οι δύο κατηγορίες διαφέρουν ελαφρώς μεταξύ τους.

Οπτική αναγνώριση σημείων (OMR)

Σε αντίθεση με τις τεχνολογίες OCR, η οπτική αναγνώριση σημείων (Optical Mark Recognition) βασίζεται κυρίως στον εντοπισμό συγκεκριμένων σημείων ή προτύπων σε ειδικά έντυπα. Η συγκεκριμένη τεχνολογία σχεδιάστηκε κατά κύριο λόγο για χρήση στην διόρθωση απαντήσεων σε τεστ εκπαιδευτικών οργανισμών, όπου και χρησιμοποιείται ακόμα και σήμερα. Επιπλέον η ίδια τεχνολογία μπορεί να συναντηθεί και σε πρακτορεία στοιχηματισμού για την επικύρωση στοιχηματικών δελτίων από ειδικά σχεδιασμένα μηχανήματα, για την άμεση εξαγωγή δεδομένων από φόρμες, ακόμα και για την καταμέτρηση ψήφων σε εκλογικές διαδικασίες.

Οι συγκεκριμένες συσκευές λειτουργούν σαν σκάνερ, εκπέμποντας δηλαδή φως επάνω στο εισερχόμενο έγγραφο, το οποίο βοηθάει στον εντοπισμό των συγκεκριμένων πεδίων που προορίζονται προς αναγνώριση. Τα ίδια τα έγγραφα που χρησιμοποιούνται είναι σχεδιασμένα να περιέχουν μικρούς κύκλους ή τετράγωνα που λειτουργούν ως πεδία, μερικά από τα οποία μπορεί να είναι μαρκαρισμένα ανάλογα με τις επιλογές του κάθε χρήστη. (Ask Any Difference, n.d)

Έξυπνη αναγνώριση χαρακτήρων (ICR)

Η έξυπνη αναγνώριση χαρακτήρων (Intelligent Character Recognition) θα μπορούσε πολύ εύκολα να χαρακτηριστεί ως μια εξελιγμένη έκδοση της τεχνολογίας OCR. Συγκεκριμένα, η ICR είναι σχεδιασμένη να εντοπίζει, να αναγνωρίζει πολλούς διαφορετικούς τρόπους γραφικών χαρακτήρων, και στη συνέχεια να τους μετατρέπει σε

ψηφιακούς χαρακτήρες. Αυτή είναι και η κυριότερη διαφορά μεταξύ των τεχνολογιών OCR και ICR. Για την αναγνώριση τυπωμένων εγγράφων λαμβάνουμε καλύτερα αποτελέσματα με την OCR, ενώ για την αναγνώριση σημειώσεων ή χειρόγραφων εντύπων, η τεχνολογία της ICR αποτελεί μονόδρομος.

Μια εξίσου σημαντική διαφορά παρατηρείται επίσης και στον τρόπο με τον οποίο οι δύο αυτές τεχνολογίες εκπαιδεύονται. Η ICR περιέχει ένα νευρωνικό σύστημα, το οποίο επιτρέπει στο λογισμικό να ενημερώνει αυτόματα μια εσωτερική βάση δεδομένων με νέους γραφικούς χαρακτήρες. Ως αποτέλεσμα, ένα πρόγραμμα ICR μπορεί να κατανοήσει οποιονδήποτε χειρόγραφο χαρακτήρα, και κατά συνέπεια αυξάνεται η ακρίβεια και η αποτελεσματικότητα της αναγνώρισης ανάλογα με τον χρόνο και την εμπειρία που λαμβάνει. (Taylor, n.d)

Έξυπνη αναγνώριση λέξεων (IWR)

Όπως και με την τεχνολογία της OCR, έτσι και η έξυπνη αναγνώριση λέξεων (Intelligent Word Recognition) αποτελεί μια τροποποιημένη έκδοση της ICR, η οποία είναι σχεδιασμένη για την αναγνώριση επάνω σε ολόκληρες λέξεις, έναντι της ICR που η αναγνώριση γίνεται σε κάθε χαρακτήρα ξεχωριστά.

3.5 ΤΕΧΝΙΚΕΣ

Μιας και η τεχνολογία της οπτικής αναγνώρισης χαρακτήρων εξαρτάται σε μεγάλο βαθμό από την ποιότητα των δεδομένων εισόδου που εισάγονται σε ένα τέτοιου είδους μοντέλο. Αυτά τα δεδομένα, ή αλλιώς εικόνες εισόδου, μπορεί να βρίσκονται σε μια καλή, ή ακόμα και εξαιρετική ποιότητα, κάτι που διευκολύνει την διαδικασία της αναγνώρισης και κατά συνέπεια παράγει μεγάλα ποσοστά ακρίβειας κατά την ολοκλήρωση της. Αντιθέτως όμως, τα δεδομένα μπορούν επίσης να έχουν μια αρκετά κακή ποιότητα, όπως θόρυβος, κακή ανάλυση της εικόνας, κακός φωτισμός, θόλωση της εικόνας, κλπ. Σε αυτή την περίπτωση, συχνά απαιτείται να πραγματοποιηθούν ορισμένες τεχνικές που εξυπηρετούν στην βελτίωση της ακρίβειας αναγνώρισης του συστήματος. Ορισμένες κατηγορίες βελτιστοποίησης μπορούν να χαρακτηριστούν οι ακόλουθες: (Karandish, n.d), (Magesh, n.d)

Προ-Επεξεργασία (Pre-Processing)

Το πρώτο βήμα που πραγματοποιείται προτού οι εικόνες εισόδου εισέλθουν εντός του μοντέλου οπτικής αναγνώρισης είναι η προ-επεξεργασία αυτών των εικόνων. Για να μπορέσει το δίκτυο να εξάγει τα καλύτερα δυνατά αποτελέσματα, το περιεχόμενο των εικόνων που χρησιμοποιούνται πρέπει να είναι όσο το δυνατόν ξεκάθαρο. Η διαδικασία αυτή πραγματοποιείται με διάφορους τρόπους, αναλόγως της κατάστασης κάθε εικόνας. Γενικότερα, ορισμένες κύριες μέθοδοι προ-επεξεργασίας θεωρούνται οι εξής:

- **Διόρθωση στρέβλωσης (De-skew):** Η διαδικασία της περιστροφής μιας εικόνας σε τέτοιο βαθμό όπου το περιεχόμενο της εικόνας είναι απόλυτα ίσιο τόσο οριζόντια όσο και κατακόρυφα. Συχνά η περιστροφή πραγματοποιείται κατά λίγες μοίρες ακολουθώντας την φορά των δεικτών του ρολογιού ή και αντίστροφα. Αυτή η διόρθωση συνήθως συναντάται όταν ένα έγγραφο δεν έχει σωστά.
- **Διαδική μετατροπή (Binarization):** Η μετατροπή μιας έγχρωμης εικόνας σε ασπρόμαυρη. Θεωρείται ως μια εύκολη και ιδιαίτερα ακριβής τεχνική, καθώς μπορεί να ξεκαθαρίσει τα σημεία στα οποία υπάρχει κείμενο, ή ακόμα και άλλα απαιτούμενα στοιχεία εντός της εικόνας.
- **Ανάλυση διάταξης (Layout analysis):** Αποτελεί τον εντοπισμό διαφόρων τμημάτων σε μια εικόνα ή ένα έγγραφο. Συγκεκριμένα μπορεί να πραγματοποιηθεί ο διαχωρισμός πινάκων, παραγράφων, κλπ. σε ξεχωριστά κομμάτια για την αναγνώριση ή ακόμα και την αποφυγή αναγνώρισης επάνω σε αυτά τα συγκεκριμένα τμήματα.
- **Αναγνώριση γραφής (Script recognition):** Σε έγγραφα που περιέχουν γλώσσες πολλών διαφορετικών αλφάβητων και μπορούν να εναλλάσσονται τακτικά εντός του κειμένου, θεωρείται απαραίτητος ο εντοπισμός των σημείων όπου η γλώσσα αλλάζει, έτσι ώστε να μπορεί να αξιοποιηθεί το καλύτερο δυνατό σύστημα για κάθε σημείο του κειμένου.
- **Απομόνωση χαρακτήρων (Character isolation / Segmentation):** Πολλές φορές η κακή ανάλυση μιας εικόνας μπορεί να ενώσει δύο ξεχωριστούς χαρακτήρες σε σημείο όπου εμφανίζονται σαν ένας. Εναλλακτικά υπάρχει και η περίπτωση όπου ένας χαρακτήρας έχει χωριστεί σε δύο ή περισσότερα τμήματα. Η τεχνική της απομόνωσης χρησιμοποιείται σε τέτοιες περιπτώσεις για να διαχωρίσει ή ακόμα

και για να ενώσει αυτούς τους χαρακτήρες, και κατά συνέπεια να μειώσει πιθανά λάθη που μπορεί να συμβούν κατά την διάρκεια της αναγνώρισης.

Αναγνώριση κειμένου (Text Recognition)

Η διαδικασία της αναγνώρισης κειμένου αποτελεί τον κορμό ολόκληρης της τεχνολογίας OCR, καθώς είναι υπεύθυνο για ολόκληρο το κομμάτι της αναγνώρισης χαρακτήρων ή λέξεων σε μια εικόνα ή ένα έγγραφο. Για να επιτευχθεί η αναγνώριση μπορούν να χρησιμοποιηθούν δύο κύριες τεχνικές, οι οποίες είναι οι ακόλουθες:

- **Ταύτιση εικόνων (Image correlation):** Η ταύτιση εικόνων θεωρείται η πιο απλή μορφή αναγνώρισης, καθώς το σύστημα οπτικής αναγνώρισης προσπαθεί να αντιστοιχήσει έναν χαρακτήρα από μια εικόνα με έναν από τους χαρακτήρες που γνωρίζει από την διαδικασία της εκπαίδευσης του. Η ταύτιση επιτυγχάνεται απομονώνοντας κάθε χαρακτήρα ξεχωριστά, και στη συνέχεια αναλύοντας κάθε pixel αυτού του χαρακτήρα, με σκοπό την ταυτοποίηση του με κάποιον χαρακτήρα που υπάρχει μέσα στο εκπαιδευμένο μοντέλο. Η συγκεκριμένη τεχνική λειτουργεί καλύτερα σε τυπωμένους χαρακτήρες, αλλά υστερεί σημαντικά όταν καλείται να αναγνωρίσει χαρακτήρες από γραμματοσειρές τις οποίες το σύστημα δεν γνωρίζει.
- **Εξαγωγή χαρακτηριστικών (Feature extraction):** Ο δεύτερος τρόπος αναγνώρισης είναι η εξαγωγή χαρακτηριστικών. Το μοντέλο προσπαθεί να εντοπίσει και να εξάγει πρότυπα που υπάρχουν σε κάθε χαρακτήρα κατά την διαδικασία της εκπαίδευσης, όπως για παράδειγμα γραμμές, κύκλους που δημιουργούνται μέσα σε χαρακτήρες, γραμμές οι οποίες ενώνονται με άλλες, κλπ. Αποτελεί μια ιδιαίτερα αποτελεσματική τεχνική τόσο από πλευράς ακρίβειας, όσο και υπολογιστικής απόδοσης, καθώς δεν περιορίζεται σε συγκεκριμένες γραμματοσειρές όπως στην τεχνική της ταύτισης εικόνων. Συχνά συναντάται σε συστήματα ICR, καθώς επίσης και στα πιο προηγμένα συστήματα OCR.

Μετα-Επεξεργασία (Post-Processing)

Κατά την ολοκλήρωση της αναγνώρισης, η έξοδος του συστήματος μπορεί επίσης να περάσει από μια διαδικασία μετα-επεξεργασίας για περεταίρω ενίσχυση της ακρίβειας αναγνώρισης. Μια ιδιαίτερα χρήσιμη τεχνική τέτοιου είδους περιλαμβάνει την χρήση λεξικών, δηλαδή μια τεράστια λίστα λέξεων που θα μπορούσαν να εντοπιστούν εντός της

εικόνας προς αναγνώριση. Το αποτέλεσμα που εξάγεται από το στάδιο της αναγνώρισης ελέγχεται με το λεξικό για τον εντοπισμό πιθανών λέξεων. Η παρούσα τεχνική μπορεί να εντοπίσει πιθανές λάθος αναγνώσεις που έχουν πραγματοποιηθεί, και στη συνέχεια να διορθώσει αυτά τα λάθη σύμφωνα με τις ήδη υπάρχουσες λέξεις εντός του λεξικού.

Βελτιστοποιήσεις για συγκεκριμένες εφαρμογές (Application-specific optimization)

Μια ακόμα τεχνική που αξιοποιείται για την ενίσχυση της ακρίβειας των μοντέλων OCR αποτελεί την τροποποίηση ήδη υπάρχων μοντέλων για την εξειδίκευσή τους επάνω σε συγκεκριμένους τύπους εισόδων, όπως για τον καλύτερο εντοπισμό πινακίδων αυτοκινήτων, ανάλυση ταυτοτήτων, διαβατηρίων, κλπ.

4. ΥΛΟΠΟΙΗΣΗ ΕΦΑΡΜΟΓΗΣ

4.1 Η ΣΥΛΛΗΨΗ ΤΗΣ ΙΔΕΑΣ

Ένα διάστημα πριν την εκδήλωση ενδιαφέροντος για την υλοποίηση της παρούσας πτυχιακής εργασίας, βρήκα τον εαυτό μου σε μια κατάσταση όπου χρειαζόταν να δακτυλογραφώ πληροφορίες σε λογιστικά φύλλα τύπου excel, τις οποίες πρώτα εξήγαγα από σαρωμένα έντυπα μορφής pdf.

Εκείνη την περίοδο, καθώς υλοποιούσα την προαναφερόμενη διαδικασία, ξεκίνησα να σκέφτομαι διάφορους τρόπους με τους οποίους θα μπορούσε αυτή η πράξη να αυτοματοποιηθεί, μιας και ήταν κάπως χρονοβόρα για την ολοκλήρωση της. Με αυτό το σκεπτικό, μετά από ένα απόγευμα έρευνας ανακάλυψα την τεχνολογία της οπτικής αναγνώρισης χαρακτήρων, και αμέσως σκέφτηκα πως θα μπορούσα να πραγματοποιήσω μια περεταίρω έρευνα επάνω στον συγκεκριμένο τομέα, ικανή για να μπορέσει να αποτελέσει την πτυχιακή μου εργασία.

Το παρόν έντυπο, καθώς επίσης και η υλοποίηση της εφαρμογής που το συνοδεύει, αποτελεί το αποτέλεσμα αυτής της έρευνας διάρκειας, περίπου, έξι μηνών.

Ως γλώσσα προγραμματισμού για την κατασκευή της συγκεκριμένης εφαρμογής, επέλεξα να χρησιμοποιήσω την Python, και συγκεκριμένα την έκδοση 3.8 για να μην υπάρχουν τυχόν προβλήματα συμβατότητας κατά την διάρκεια της υλοποίησης.

4.2 DATASETS

Από την πρώτη στιγμή που ξεκίνησα την έρευνα μου επάνω στο κομμάτι της οπτικής αναγνώρισης, ανακάλυψα την ιδιαίτερη σημασία που τα σετ δεδομένων, λεγόμενα ως datasets, για την σωστή εκπαίδευση και πρόβλεψη των νευρωνικών μοντέλων.

Πολύ γρήγορα βρέθηκα γρήγορα αντιμέτωπος με την βάση δεδομένων Mnist, του εθνικού ινστιτούτου τεχνολογίας και προτύπων (NIST), των Ηνωμένων Πολιτειών. Η συγκεκριμένη βάση περιέχει περίπου 70.000 ασπρόμαυρες, χειρόγραφες εικόνες αριθμητικών χαρακτήρων, μεγέθους 28x28 pixels έκαστος, εκ των οποίων οι 60.000 προορίζονται προς εκπαίδευση μοντέλων και οι υπόλοιπες 10.000 προς επαλήθευση των αποτελεσμάτων του μοντέλου.

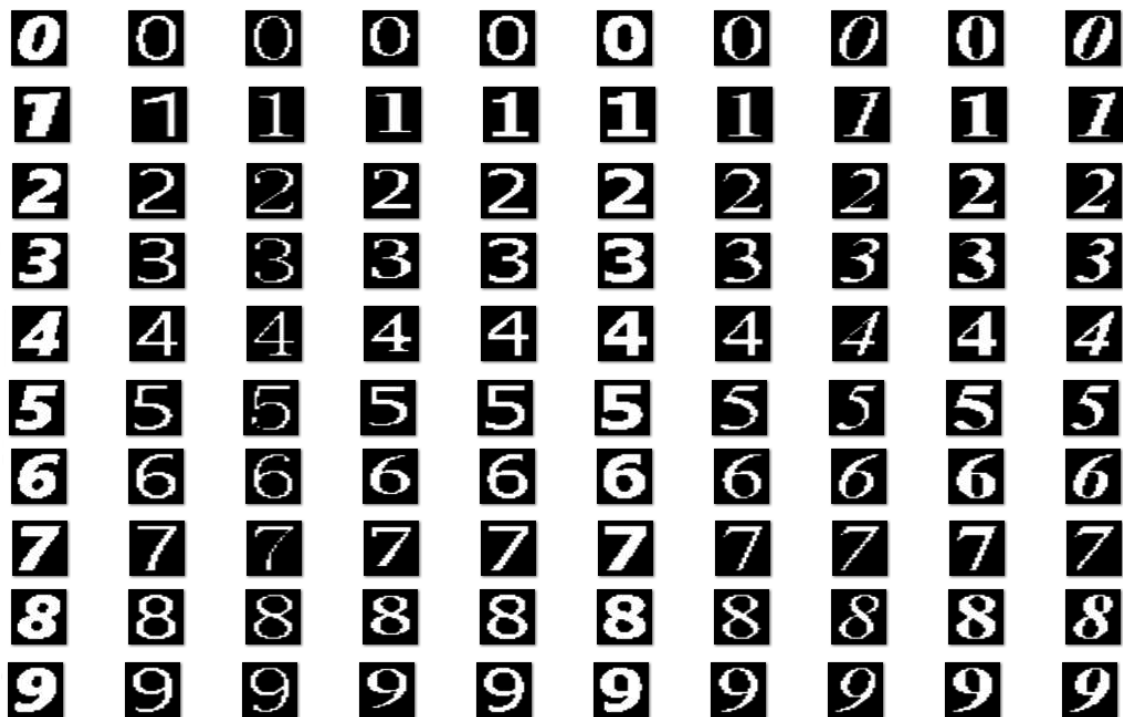


Εικόνα 10: Ένα μικρό δείγμα από το Mnist dataset.

Παίρνοντας ένα μικρό μέρος του Mnist, ξεκίνησα να πραγματοποιώ μια πρόχειρη ανάλυση στον τρόπο με τον οποίο το συγκεκριμένο dataset αλληλοεπιδρά με νευρωνικά μοντέλα στο κομμάτι της εκπαίδευσης και της πρόβλεψης. Τα αποτελέσματα ήταν ιδιαίτερα ενθαρρυντικά, καθώς το εκπαιδευμένο μοντέλο ήταν σε θέση να αναγνωρίσει παρόμοιους χαρακτήρες σε έναν αρκετά ικανοποιητικό βαθμό.

Ταυτόχρονα παρατήρησα πως μεγάλο μέρος της ακρίβειας των μοντέλων που έχουν εκπαιδευτεί με το Mnist έρχεται από το γεγονός πως υπάρχει μια πληθώρα διαφορετικών, και μοναδικών εικόνων για κάθε ξεχωριστό χαρακτήρα. Αυτό σημαίνει πως κατά την εκπαίδευση, το νευρωνικό μοντέλο προσπαθεί να εξάγει ορισμένα κοινά σημεία μεταξύ των διαφορετικών εικόνων που ανήκουν στον ίδιο χαρακτήρα, όπως για παράδειγμα συγκεκριμένες ακμές, κυκλικά σημεία, κλπ. Έτσι λοιπόν, για να μπορέσει ένα dataset να χαρακτηριστεί ως πετυχημένο, χρειάζεται να περιέχει πολλές διαφοροποιήσεις ενός κοινού χαρακτήρα ή συμβόλου.

Για την υλοποίηση του προγράμματος θεώρησα πως θα ήταν μια καλή ιδέα η δημιουργία μιας αυτόματης γεννήτριας παρόμοιων εικόνων με το Mnist, η οποία θα ήταν σε θέση να κατασκευάσει ασπρόμαυρες εικόνες, κάθε μια με έναν ξεχωριστό χαρακτήρα σε διαφορετική γραμματοσειρά.



Εικόνα 11: Ένα μικρό δείγμα από την γεννήτρια εικόνων.

Η εικόνα 11 παρουσιάζει το αποτέλεσμα αυτής της αυτόματης γεννήτριας εικόνων. Κάθε εικόνα αποτελείται από διαστάσεις 50x50 pixel, τριών χρωματικών καναλιών (rgb), μαύρου φόντο με έναν μόνο λευκό αλφαριθμητικό χαρακτήρα διαφορετικών γραμματοσειρών σε κάθε εικόνα.

4.2.1 Γεννήτρια παραγωγής εικόνων

Κατά την εκτέλεση της εφαρμογής, η συνάρτηση `generate_store_dataset` καλείται όταν ο χρήστης επιθυμεί να ξεκινήσει την διαδικασία δημιουργίας ενός νέου μοντέλου. Η συγκεκριμένη συνάρτηση αποτελεί και την κύρια δομή της γεννήτριας εικόνων. Στην προκειμένη περίπτωση, ο χρήστης καλείται να επιλέξει ανάμεσα από ορισμένες επιλογές που του δίνονται από το γραφικό περιβάλλον της εφαρμογής, με σκοπό να δημιουργηθεί το κατάλληλο dataset για τις ανάγκες του χρήστη. Περισσότερα στο κεφάλαιο 4.5 – Γραφική Διεπαφή.

```
font_name_array = ['arial', 'times', 'timesi', 'timesbd',  
'timesbi', 'bahnschrift', 'cambria', 'constan', 'lucon',  
'calibri', 'impact', 'segoepr', 'segoesc', 'comic',  
'pala', 'verdana', 'verdanai', 'verdanab', 'verdanaz',  
'trebuc', 'trebucit', 'trebucbd', 'trebucbi']  
  
font_size = 35 # Font size of each character
```

Πίνακας 1: Ορισμός γραμματοσειρών.

Αρχικά ορίζονται στην λίστα `font_name_array` μια σειρά από ονόματα διαφορετικών γραμματοσειρών, τα οποία χρησιμοποιούνται στην συνέχεια από την γεννήτρια για την εμφάνιση των χαρακτήρων σε κάθε δοθούμενη γραμματοσειρά. Επιπλέον εντάσσεται στην μεταβλητή `font_size` το μέγεθος με το οποίο θα εμφανιστούν οι χαρακτήρες επάνω στην εικόνα.

```
if(selected_dataset == 'Latin'):  
    for l in range(65, 91):  
        letters.append(chr(l))  
        directory_letters.append(chr(l))  
        letters_in_image.append(chr(l))
```

Πίνακας 2: Συλλογή χαρακτήρων.

Η εφαρμογή λαμβάνει την επιλογή του χρήστη για την δημιουργία ενός dataset, και στην συνέχεια αντιστοιχεί αυτήν την επιλογή με τους απαραίτητους χαρακτήρες. Συγκεκριμένα, αν για παράδειγμα ο χρήστης έχει επιλέξει την επιλογή για λατινικούς χαρακτήρες, θα χρησιμοποιηθούν μόνο οι χαρακτήρες που ανήκουν στο λατινικό αλφάβητο, με κωδικούς αριθμούς από το 65 έως το 90 στο σύστημα Unicode. Αυτοί οι κωδικοί εισάγονται σε τρεις ξεχωριστές λίστες `letters`, `directory_letters`,

letters_in_image ως χαρακτήρες και στην συνέχεια χρησιμοποιούνται για την δημιουργία των εικόνων.

```
#Creating the database folder.
os.mkdir(dataset_directory_root)

for l in range (0, len(letters)):

    counter = 1

    # Setting up the directory path for each character
    letter_directory = dataset_directory_root + '/'
                    + directory_letters[l]

    # Creating the directories
    os.mkdir(letter_directory)

    for font_name in font_name_array:
        try:
            font = ImageFont.truetype(font_name, font_size)

            # Setting up the picture name
            picture_name = letters[l] + '_' + str(counter)
                        + '.jpg'

            # Create a new RGB image for every character
            # with size x,y (PIL)
            image = Image.new('RGB', (img_width, img_height))

            # Enable drawing on the image
            draw = ImageDraw.Draw(image)

            # Draw the character on the image
            draw.text((10, int(img_height/(font_size/2))),
                    letters_in_image[l],
                    fill = (255, 255, 255), font = font)

            image = Augment_Image(image)

            # Moving the directory to the character folder
            os.chdir(letter_directory)

            # Using PIL to save the image
            image.save(picture_name)

            counter = counter + 1
        except:
            pass

TrainModel(dataset_directory_root, exec_directory_root,
```

```
img_width, img_height)
```

Πίνακας 3: Γεννήτρια εικόνων.

Αμέσως μετά την εισαγωγή των επιλεγμένων χαρακτήρων εντός των λιστών, δημιουργείται ένας φάκελος με όνομα Dataset στο ίδιο σημείο από το οποίο εκτελείται η εφαρμογή. Ο συγκεκριμένος φάκελος προορίζεται ως μέσο αποθήκευσης των εικόνων που δημιουργούνται από την γεννήτρια, και κατά συνέπεια ως ένας τρόπος ελέγχου της ορθότητας των εικόνων προτού εισαχθούν στο μοντέλο εκπαίδευσης. Εάν υπάρχει ήδη ο φάκελος Dataset, από κάποια προηγούμενη εκτέλεση του προγράμματος, τότε θα το ήδη υπάρχον περιεχόμενο θα διαγραφτεί, και στη συνέχεια θα αντικατασταθεί με τις νέες εικόνες.

Εντός του φακέλου Dataset δημιουργούνται υποφάκελοι με τα ονόματα των χαρακτήρων προς δημιουργία, όπου και αποθηκεύονται οι εικόνες που παράγονται από την γεννήτρια.

Χρησιμοποιώντας τις συναρτήσεις ImageFont, Image και ImageDraw από την βιβλιοθήκη Python Image Library (PIL ή αλλιώς Pillow), ορίζεται πως στην μεταβλητή font θα εισαχθεί η τρέχουσα επιλεγμένη γραμματοσειρά από την λίστα font_name_array, μαζί την τιμή της font_size για το μέγεθος των χαρακτήρων εντός της εικόνας.

Στην συνέχεια προετοιμάζεται το όνομα που θα λάβει η εικόνα κατά την ολοκλήρωση της δημιουργίας της όταν αποθηκευτεί στον δίσκο, το οποίο λαμβάνει την μορφή: χαρακτήρας_αριθμός, ο οποίος αριθμός αποτελείται από έναν αυξανόμενο μετρητή counter ως τρόπο διαφοροποίησης μεταξύ όλων των εικόνων. Έπειτα δημιουργείται μια εικόνα image στην μνήμη του υπολογιστικού συστήματος με χρωματική μορφή RGB και μέγεθος τις τιμές των μεταβλητών img_width και img_height. Όταν έχει δημιουργηθεί η εικόνα, καλείται η ImageDraw επάνω στην εικόνα για να επιτρέψει την μετέπειτα επεξεργασία της. Η draw πραγματοποιεί την αρχική τοποθέτηση κάθε χαρακτήρα επάνω στην εικόνα, με λευκό χρώμα και την τρέχουσα γραμματοσειρά.

4.2.2 Ενίσχυση εικόνων

Αμέσως μετά η εικόνα περνάει από την συνάρτηση `Augment_Image`, η οποία τροποποιεί ελάχιστα την εικόνα και στη συνέχεια την επιστρέφει πίσω στην συνάρτηση `generate_store_dataset`, όπου και στη συνέχεια θα αποθηκευτεί εντός του κατάλληλου φακέλου, ενώ τέλος θα ξεκινήσει η διαδικασία της εκπαίδευσης.

```
def Augment_Image(image):

    global img_height
    global img_width

    #Converting the image from PIL to CV2 format.
    cv2Image = np.array(image)
    cv2Image = cv2.cvtColor(cv2Image, cv2.COLOR_RGB2BGR)

    #Grayscaleing the image and getting its threshold.
    gray = cv2.cvtColor(cv2Image, cv2.COLOR_BGR2GRAY)
    thresh = cv2.threshold(gray, 0, 255, cv2.THRESH_BINARY
                           + cv2.THRESH_OTSU)[1]

    #Morphologically closing the image to connect
    #all the character parts.
    kernel = np.ones((13,13),np.uint8)
    closing = cv2.morphologyEx(thresh, cv2.MORPH_CLOSE,
                               kernel)

    #Detecting contours inside the images.
    cnts, hierarchy = cv2.findContours
                        (closing, cv2.RETR_EXTERNAL,
                         cv2.CHAIN_APPROX_SIMPLE)

    for c in cnts:

        #Bounding the contours into a rectangle,
        #then expanding the rectangle.
        (x, y, w, h) = cv2.boundingRect(c)
        x = x-3
        y = y-3
        w = w+6
        h = h+6

        if((x-3) < 0):
            x = 0
        if((y-3) < 0):
            y = 0

        #Creating a new image based on the expanded rectangle
        roi = thresh[y:y+h, x:x+w]
```

```

#Resizing the new image.
roi = cv2.resize(roi, (img_width, img_height))

cv2Image = cv2.cvtColor(roi, cv2.COLOR_BGR2RGB)
pilImage = Image.fromarray(roi)

return pilImage

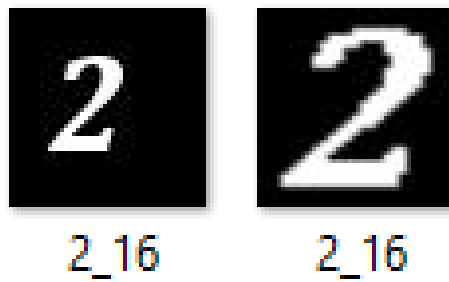
```

Πίνακας 4: Συνάρτηση `Augment_Image`.

Συγκεκριμένα, η `Augment_Image` μετατρέπει την εικόνα από PIL σε μορφή συμβατή με την βιβλιοθήκη OpenCV για την ευκολότερη επεξεργασία τους. Αρχικά η εικόνα λαμβάνει μια ασπρόμαυρη μορφή στα χρωματικά της κανάλια μέσω της `cv2.cvtColor`, κάτι που στην συνέχεια επιτρέπει την χρωματική οριοθέτηση εντός της εικόνας με την χρήση της `cv2.threshold`, και της δυαδικής μεθόδου του Otsu για την καλύτερη δυνατή οριοθέτηση μεταξύ των μαύρων pixel του φόντου και των λευκών pixel των χαρακτήρων. Το αποτέλεσμα αποθηκεύεται στην μεταβλητή `thresh`, και στην συνέχεια δέχεται έναν μορφολογικό μετασχηματισμό με σκοπό να ενωθούν τυχόν σημεία που μπορεί να περιέχουν «τρύπες» επάνω στον κάθε χαρακτήρα. Η συγκεκριμένη ενέργεια χρησιμεύει κυρίως στο επόμενο βήμα της συνάρτησης, όπου χρησιμοποιείται η συνάρτηση `cv2.findContours`, για τον εντοπισμό του περιγράμματος κάθε χαρακτήρα. Το περίγραμμα αποθηκεύεται στην `cnts` με μορφή συντεταγμένων που φανερώνουν σημεία επάνω στις ακμές και τις γωνίες ενός χαρακτήρα.

Στη συνέχεια, με την χρήση της `cv2.boundingRect`, το περίγραμμα του χαρακτήρα περιτριγυρίζεται από ένα ορθογώνιο πλαίσιο, και έπειτα δημιουργείται μια νέα εικόνα Region Of Interest – `roi`, η οποία περιλαμβάνει τον αρχικό χαρακτήρα στο κέντρο της εικόνας. Με το τρέχον βήμα γίνεται μια προσπάθεια ομοιομορφίας τόσο των χαρακτήρων προς εκπαίδευση, όσο και προς πρόβλεψη, μιας και η ίδια τεχνική χρησιμοποιείται και στο κομμάτι της αναγνώρισης.

Τέλος, η νέα πλέον εικόνα παίρνει το απαραίτητο μέγεθος 50x50 pixels από τις μεταβλητές `img_width` και `img_height` μέσω της συνάρτησης `cv2.resize`, και στη συνέχεια η εικόνα λαμβάνει ξανά την μορφή PIL από OpenCV προτού επιστρέψει στην γεννήτρια για αποθήκευση.



Εικόνα 12: Πριν και μετά την συνάρτηση `Augment_Image`

4.3 ΚΑΤΑΣΚΕΥΗ ΜΟΝΤΕΛΟΥ

Κατά την διάρκεια της αρχικής μου έρευνας, όταν είχα ξεκινήσει να εξετάζω το αντικείμενο της οπτικής αναγνώρισης χαρακτήρων και ταυτόχρονα ανακάλυπτα τις διαφορετικές διαθέσιμες τεχνολογίες υλοποίησης εφαρμογών OCR, βρέθηκα αντιμέτωπος με το Keras, και το Pytorch. Αρχικά ξεκίνησα να πραγματοποιώ μια μικρή υλοποίηση ενός demo με το Keras, στο οποίο εξέταζα την ευκολία υλοποίησης, επεξεργασίας, καθώς και τα αποτελέσματα της εκπαίδευσης σε ένα μικρό dataset.

Παρατήρησα πως υπήρχε μια ιδιαίτερη απλότητα στην κατασκευή μοντέλων, σε σημείο όπου έμοιαζε με παζλ με τον τρόπο που όλα τα κομμάτια αλληλοεπιδρούν μεταξύ τους και παράγουν διαφορετικά αποτελέσματα ανάλογα με τις ορισμένες παραμέτρους. Αργότερα, όταν επιχείρησα να πειραματιστώ με το Pytorch, δεν ένοιωθα το ίδιο άνετα με την συγκεκριμένη τεχνολογία, με αποτέλεσμα να επιλέξω την επιστροφή μου στην πρώτη μου επιλογή, λαμβάνοντας την τελική απόφαση να υλοποιήσω την εφαρμογή με το Keras.

Από τεχνικής άποψης, αυτές οι δύο τεχνολογίες χαρακτηρίζονται ως εξής: (Sayantini, 2020)

Keras:

- Τεχνολογία ανοιχτού κώδικα (Open Source), κατασκευασμένη σε γλώσσα Python.

- Χαρακτηρίζεται ως μια υψηλόβαθμη διεπαφή προγραμματισμού εφαρμογών (high-level API), το οποίο μπορεί να λειτουργήσει επάνω σε τεχνολογίες όπως το TensorFlow, το CNTK, και το Theano.
- Η ταχύτητα του Keras θεωρείται πιο αργή συγκριτικά με άλλες τεχνολογίες υλοποίησης και εκπαίδευσης νευρωνικών δικτύων, με αποτέλεσμα να χρησιμοποιείται κυρίως σε μικρότερα dataset.
- Η αρχιτεκτονική του χαρακτηρίζεται ως αρκετά απλή, ενώ παράλληλα θεωρείται ευανάγνωστη και περιεκτική, κάτι που επιτρέπει στην λιγότερο συχνή αποσφαλμάτωση του δικτύου.

PyTorch:

- Επίσης τεχνολογία ανοιχτού κώδικα (Open Source), κατασκευασμένη για την γλώσσα Python, βασισμένη στην τεχνολογία Torch.
- Αντίθετα με το Keras, το PyTorch χαρακτηρίζεται ως μια χαμηλόβαθμη διεπαφή προγραμματισμού εφαρμογών (lower-level API), επικεντρωμένο κυρίως στην απευθείας λειτουργία με array expressions.
- Θεωρείται ιδιαίτερα γρήγορη τεχνολογία, σχεδιασμένη με υψηλή αποδοτικότητα, με αποτέλεσμα να λειτουργεί καλύτερα με μεγάλου μεγέθους dataset.
- Η αρχιτεκτονική του είναι ιδιαίτερα περίπλοκη, σχετικά δυσανάγνωστη, αλλά παρέχει μεγαλύτερες δυνατότητες αποσφαλμάτωσης του δικτύου.

4.3.1 Το μοντέλο

Έχοντας πλέον επιλέξει την τεχνολογία βαθιάς μάθησης με την οποία θα υλοποιούσα την εφαρμογή οπτικής αναγνώρισης, ξεκίνησα να υλοποιώ το νευρωνικό μοντέλο που θα μπορούσε να υποστηρίξει το συγκεκριμένο εγχείρημα. Συγκεκριμένα κατασκευάστηκαν 5 μοντέλα συνολικά, με το τελευταίο να αποτελεί και το μοντέλο το οποίο χρησιμοποιείται κατά την εκτέλεση της εφαρμογής, ενώ όλα τα μοντέλα βρίσκονται στο αρχείο `predict.py` της εφαρμογής.

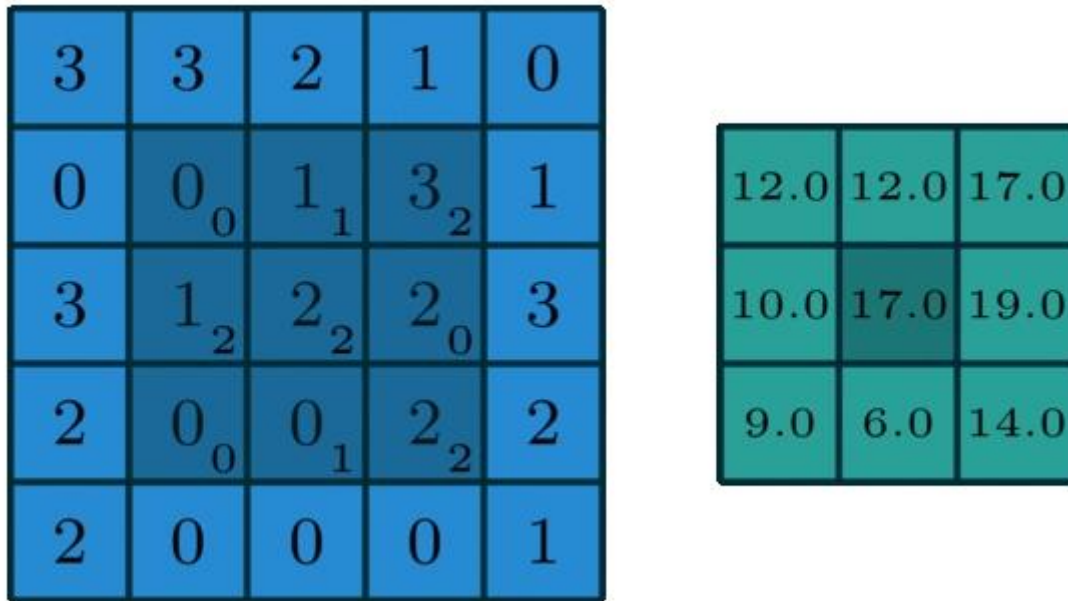
Ο πίνακας 5 παρουσιάζει το περιεχόμενο του συγκεκριμένου μοντέλου πριν την εκπαίδευση με αριθμητικούς χαρακτήρες, μέσω της συνάρτησης `summary` που είναι διαθέσιμη έπειτα από την δημιουργία ενός νευρωνικού μοντέλου μέσω της βιβλιοθήκης Keras.

Model: "sequential"		
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 50, 50, 32)	896
max_pooling2d (MaxPooling2D)	(None, 25, 25, 32)	0
conv2d_1 (Conv2D)	(None, 25, 25, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 12, 12, 64)	0
conv2d_2 (Conv2D)	(None, 12, 12, 64)	36928
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 64)	0
flatten (Flatten)	(None, 2304)	0
dropout (Dropout)	(None, 2304)	0
dense (Dense)	(None, 128)	295040
dense_1 (Dense)	(None, 10)	1290
Total params: 352,650		
Trainable params: 352,650		
Non-trainable params: 0		

Πίνακας 5: Το διαδοχικό μοντέλο βαθιάς μάθησης της εφαρμογής.

Το συγκεκριμένο μοντέλο είναι διαδοχικού τύπου (sequential), αυτό σημαίνει πως έχει την ιδιαιτερότητα να περιέχει έναν μόνο τανυστή (tensor) τόσο στην είσοδο, όσο και στην έξοδο του. Η επιλογή του συγκεκριμένου τύπου μοντέλου είναι λόγω του τρόπου λειτουργίας της συγκεκριμένης αναγνώρισης χαρακτήρων, καθώς θα εισάγεται μόνο μια εικόνα κάθε φορά ως είσοδος, και θα εξάγεται ένα μόνο αποτέλεσμα κατά την διαδικασία της πρόβλεψης.

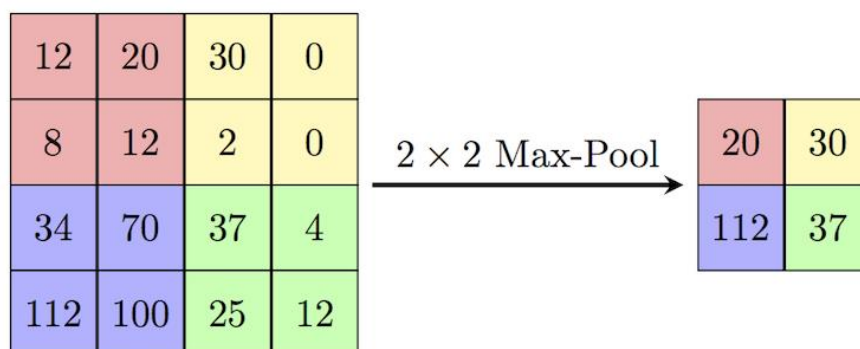
Χρησιμοποιούνται 3 στρώσεις Conv2D για την κατασκευή συνελκτικών πυρήνων διαστάσεων 3x3 (convolution kernels), οι οποίοι συνελίσσονται με τις προηγούμενες στρώσεις σαρώνοντας τα δεδομένα της εισόδου και πραγματοποιώντας πολλαπλασιασμό με τα δεδομένα που βρίσκονται την τρέχουσα στιγμή εντός του πυρήνα, και στη συνέχεια αθροίζοντας αυτά τα αποτελέσματα για να παράγουν τους τανυστές εξόδων. Στο μοντέλο της εφαρμογής χρησιμοποιείται η συνάρτηση ενεργοποίησης Relu σε κάθε Conv2D.



Εικόνα 13: Τρόπος λειτουργίας του Conv2D.

Η πρώτη στρώση Conv2D περιλαμβάνει το μέγεθος της εισόδου που περιμένει να λάβει το μοντέλο κατά την διαδικασία της εκπαίδευσης και της πρόβλεψης. Αυτή η είσοδος ορίζεται ως 50x50x3, δηλαδή 50 pixels ύψος, 50 pixels πλάτος, και 3 χρωματικά κανάλια.

Αμέσως μετά από κάθε στρώση Conv2D χρησιμοποιείται μια στρώση MaxPooling2D (3 στρώσεις συνολικά), μεγέθους 2x2, η οποία έχει ως στόχο την μείωση των δειγμάτων που αντιπροσωπεύει κάθε είσοδο. Αυτό πραγματοποιείται μειώνοντας τις διαστάσεις της εισόδου, επιτρέποντας έτσι στο μοντέλο βαθιάς μάθησης να δημιουργεί εικασίες σχετικά με χαρακτηριστικά που υπάρχουν εντός της περιοχής που αντιπροσωπεύει το MaxPooling2D.



Εικόνα 14: Γραφική αναπαράσταση της λειτουργίας MaxPooling2D.

Η συγκεκριμένη τεχνική βοηθάει κυρίως με το πρόβλημα της υπερπροσαρμογής του μοντέλου (overfitting), μέσω της παραχώρησης μιας συνοπτικής έκδοσης της εισόδου, ενώ ταυτόχρονα μειώνει το υπολογιστικό κόστος ελαχιστοποιώντας το πλήθος των παραμέτρων του δικτύου. Αυτό επιτυγχάνεται χωρίζοντας την αρχική εικόνα σε τμήματα ίσων διαστάσεων (πχ. 2x2), και εξάγοντας την μέγιστη τιμή κάθε τμήματος σε έναν ξεχωριστό πίνακα προς την έξοδο.

Στην συνέχεια χρησιμοποιούνται οι συναρτήσεις Flatten και Dropout. Η Flatten συμπίεζει τις διαστάσεις της εισόδου σε μια μόνο διάσταση, για παράδειγμα στον πίνακα 5, η είσοδος (None, 6, 6, 64) μετατρέπεται σε (None, 2304) πολλαπλασιάζοντας τις τιμές κάθε διάστασης μεταξύ τους. Αμέσως μετά, η Dropout έχει ως στόχο τον μηδενισμό, και κατά συνέπεια την εξαίρεση των νευρώνων από την διαδικασία εκπαίδευσης μέσω τυχαίας επιλογής, με πιθανότητα συνήθως 20% - 50%. Στο συγκεκριμένο μοντέλο, χρησιμοποιείται Dropout πιθανότητας 20%.

Όπως και η MaxPooling2D, έτσι και η Dropout προσπαθούν να περιορίσουν το φαινόμενο της υπερπροσαρμογής του δικτύου, με σκοπό την καλύτερη δυνατή γενίκευση στο συγκεκριμένο dataset.

Τέλος το μοντέλο αποτελείται από δύο βαθιές στρώσεις νευρώνων Dense, με την μια από αυτές να αποτελείται από μέγεθος 128 εξόδων, και την δεύτερη από εξόδους ίσες με το πλήθος των συνολικών διαθέσιμων χαρακτήρων, μιας και επιθυμούμε να λαμβάνουμε μια τιμή που θα αντιπροσωπεύει σε έναν μόνο χαρακτήρα. Κάθε νευρώνας μιας Dense στρώσης συνδέεται ολοκληρωτικά με της εξόδους της προηγούμενης στρώσης, ενώ η συγκεκριμένη στρώση θεωρείται από της πιο συχνές σε στρώσεις που χρησιμοποιούνται σε μοντέλα βαθιάς μάθησης.

4.3.2 Datagen, Ανακλητές, Μεταγλώττιση

Έχοντας πλέον ολοκληρώσει το μοντέλο που θα εκπαιδευτεί κατά την διαδικασία εκπαίδευσης, το επόμενο στάδιο αποτελεί τον τρόπο με τον οποίο οι παραγόμενες εικόνες της γεννήτριας θα εισαχθούν εντός του δικτύου από τον φάκελο στον οποίο βρίσκονται αποθηκευμένες.

Για να πραγματοποιηθεί το συγκεκριμένο εγχείρημα, θα πρέπει κάθε εικόνα να χωριστεί σε δύο σετ δεδομένων, συγκεκριμένα σε `train_set` και `validation_set`. Αυτός ο χωρισμός χρειάζεται για την ορθή εκπαίδευση του μοντέλου επάνω στο κύριο `train_set`, ενώ το `validation_set` λειτουργεί ως ένα άγνωστο σετ δεδομένων, το οποίο είναι άγνωστο από το μοντέλο. Αν τα τελικά αποτελέσματα εκπαίδευσης τόσο του `train_set` όσο και του `validation_set` είναι ικανοποιητικά, τότε μπορούμε να θεωρήσουμε πως το μοντέλο έχει εκπαιδευτεί σωστά.

```
datagen = ImageDataGenerator(  
    rescale = 1. / 255,  
    validation_split = 0.15,  
    rotation_range=8,  
    width_shift_range=0.1,  
    height_shift_range=0.1,  
    shear_range=0.1,  
    zoom_range=0.1,  
)  
  
ds_train = datagen.flow_from_directory(  
    dataset_directory_root,  
    target_size = (img_width, img_height),  
    batch_size = batch,  
    class_mode = 'categorical',  
    shuffle = True,  
    seed = 123,  
    subset = 'training',  
)  
  
ds_validate = datagen.flow_from_directory(  
    dataset_directory_root,  
    target_size = (img_width, img_height),  
    batch_size = batch,  
    class_mode = 'categorical',  
    shuffle = True,  
    seed = 123,  
    subset = 'validation',  
)
```

Πίνακας 6: `Datagen`, `train_set` και `validation_set`

Προτού χωρίσουμε τα δεδομένα μπορούμε να ορίσουμε μια μεταβλητή `datagen`, με την οποία έχουμε την δυνατότητα να δημιουργήσουμε περισσότερες εικόνες από το ήδη υπάρχον `dataset` μέσω της συνάρτησης `ImageDataGenerator` του Keras. Η συγκεκριμένη τεχνική χρησιμοποιείται για την ενίσχυση του `dataset` δημιουργώντας

αντίγραφα των εικόνων και στην συνέχεια αλλάζοντας τα ελάχιστα ως προς την περιστροφή της εικόνας, το ύψος, το πλάτος, το επίπεδο zoom, κλπ. Τα συγκεκριμένα αντίγραφα εισάγονται στην συνέχεια κανονικά μέσα στο μοντέλο κατά την εκπαίδευση. Επιπλέον, εντός του `datagen` ορίζεται και το `validation_split`, το οποίο είναι υπεύθυνο για τον χωρισμό του συνολικού dataset σε `training_set` και `validation_set`. Εντός του προγράμματος, ο χωρισμός γίνεται με τιμή 0,15, δηλαδή το 15% των εικόνων θα εισαχθούν στο `validation_set`.

Το `datagen` σε σχεδιασμό με το `flow_from_directory` μπορούν να παράγουν αυτά τα δύο απαραίτητα σετ μαζί με ορισμένες παραμέτρους που ορίζονται ανάλογα με τα χαρακτηριστικά που επιθυμούμε να έχουν. Στην εφαρμογή, δηλώνουμε πως οι εικόνες μας βρίσκονται στον φάκελο `/Dataset`, μεγέθους 50x50, και στην συνέχεια χωρίζουμε τις εικόνες σε μικρότερες ομάδες (batches) των 22 εικόνων. Τέλος ορίζουμε πως τα δύο dataset θα κατηγοριοποιούνται σύμφωνα με την κατηγορία, δηλαδή τον χαρακτήρα, όπου ανήκουν.

Το τελευταίο βήμα προτού ξεκινήσει η διαδικασία εκπαίδευσης είναι ο ορισμός των callbacks. Τα callbacks, ή αλλιώς ανακλητές, λειτουργούν σαν συναρτήσεις που καλούνται όταν πληρούνται ορισμένα κριτήρια κατά την εκπαίδευση του μοντέλου. Μπορούν να χρησιμοποιηθούν για την δημιουργία σημείων ελέγχου, τροποποίηση του ρυθμού εκμάθησης (learning rate), ακόμα και για πρόωρη ολοκλήρωση της εκπαίδευσης.

```
#Defining the model callbacks.
class Callback(tf.keras.callbacks.Callback):
    #Callbacks act as functions that get executed during
    #specific moments during training.
    #Here they work as a way of early stopping the training
    #once the model has reached accuracy and loss within
    #a specified threshold.
    def on_epoch_end(self, epoch, logs={}):
        if((logs.get('accuracy') > Accuracy_Threshold) &
            (logs.get('loss') < loss_threshold) &
            (logs.get('val_accuracy') > Accuracy_Threshold)
            & (logs.get('val_loss') < loss_threshold)):
            self.model.stop_training = True

#Initiating the model callbacks.
callbacks = Callback()
```

Πίνακας 7: Οι ανακλητές του προγράμματος.

Για την υλοποίηση της εφαρμογής επέλεξα να χρησιμοποιήσω πρόωρο τερματισμό της εκμάθησης όταν στο τέλος μιας εποχής οι τιμές των `accuracy` και `val_accuracy` είναι μεγαλύτερες από το `Accuracy_Threshold` που ορίζεται με τιμή 0,999, και οι τιμές των `loss` και `val_loss` είναι μικρότερες του `loss_threshold` με τιμή 0,02. Η επιλογή του συγκεκριμένου callback πάρθηκε από το γεγονός πως πιθανώς να μην είναι απαραίτητη η ολοκλήρωση των ορισμένων εποχών για τα καλύτερα αποτελέσματα του μοντέλου, καθώς υπάρχει μεγάλη πιθανότητα τα αποτελέσματα να αρχίσουν σταδιακά να χειροτερεύουν κατά την συμπλήρωση του ορίου εποχών, ενώ παράλληλα χρησιμοποιούνται πόροι του συστήματος για μεγαλύτερο χρονικό διάστημα από όσο χρειάζεται.

```
#The model will be compiled with the following parameters.
model.compile(
    loss='categorical_crossentropy',
    optimizer = keras.optimizers.Adam(learning_rate =
        0.0005), #rmsprop, adam, SGD
    metrics = ['accuracy']
)
```

Πίνακας 8: Μεταγλώττιση του μοντέλου.

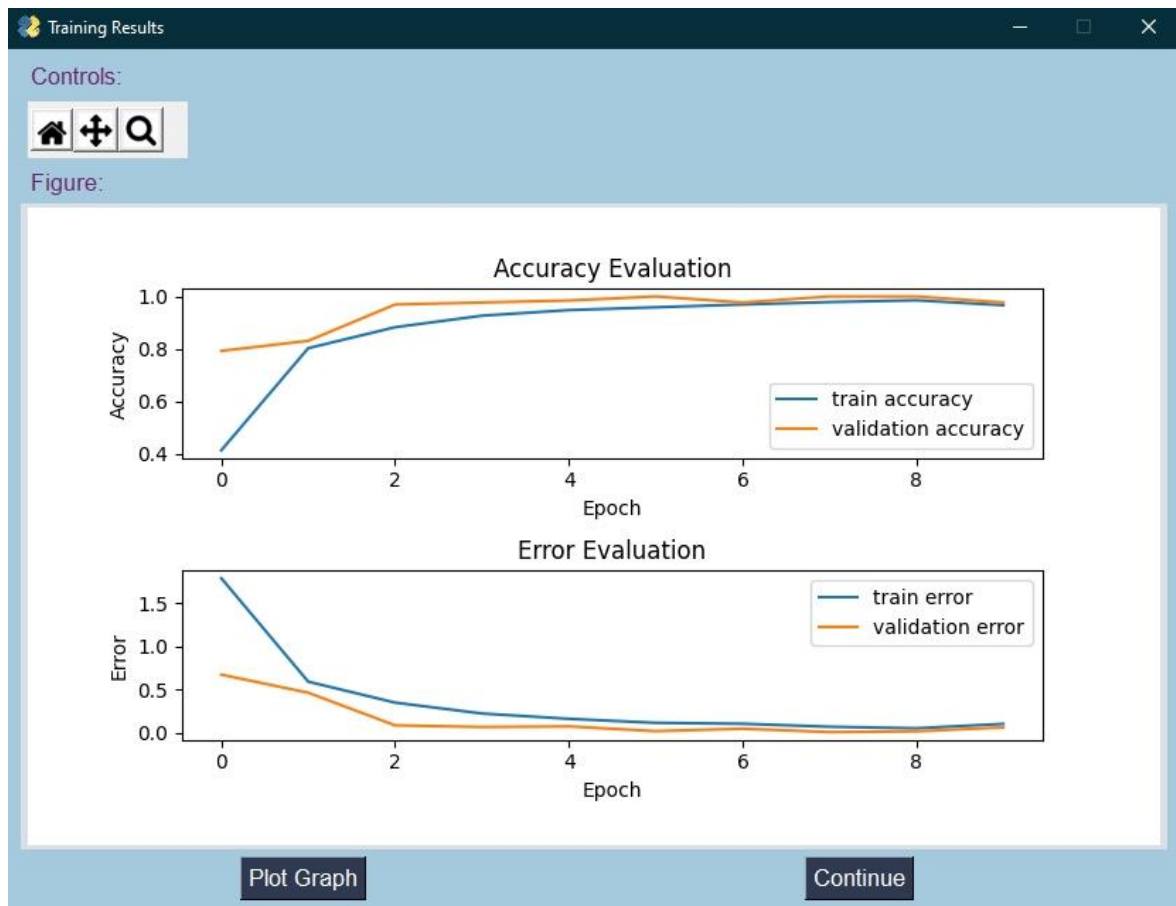
Τέλος πραγματοποιείται η μεταγλώττιση του νευρωνικού μοντέλου χρησιμοποιώντας `categorical_crossentropy` ως την απώλεια του μοντέλου, τον βελτιστοποιητή Adam με ρυθμό εκμάθησης 0,0005, και την μεταβλητή `metrics` με τιμή `accuracy` για τον υπολογισμό της ακρίβειας του μοντέλου.

4.3.3 Αξιολόγηση του μοντέλου

```
Epoch 1/10
35/35 - 5s - loss: 1.7895 - accuracy: 0.4133 - val_loss:
0.6705 - val_accuracy: 0.7923
Epoch 2/10
35/35 - 2s - loss: 0.5893 - accuracy: 0.8027 - val_loss:
0.4610 - val_accuracy: 0.8308
Epoch 3/10
35/35 - 2s - loss: 0.3460 - accuracy: 0.8827 - val_loss:
0.0824 - val_accuracy: 0.9692
Epoch 4/10
35/35 - 2s - loss: 0.2197 - accuracy: 0.9267 - val_loss:
0.0631 - val_accuracy: 0.9769
Epoch 5/10
35/35 - 2s - loss: 0.1579 - accuracy: 0.9480 - val_loss:
0.0703 - val_accuracy: 0.9846
Epoch 6/10
35/35 - 2s - loss: 0.1116 - accuracy: 0.9587 - val_loss:
0.0150 - val_accuracy: 1.0000
Epoch 7/10
35/35 - 2s - loss: 0.1008 - accuracy: 0.9693 - val_loss:
0.0435 - val_accuracy: 0.9769
Epoch 8/10
35/35 - 2s - loss: 0.0663 - accuracy: 0.9787 - val_loss:
0.0044 - val_accuracy: 1.0000
Epoch 9/10
35/35 - 2s - loss: 0.0480 - accuracy: 0.9853 - val_loss:
0.0129 - val_accuracy: 1.0000
Epoch 10/10
35/35 - 2s - loss: 0.0992 - accuracy: 0.9667 - val_loss:
0.0588 - val_accuracy: 0.9769

Training is Complete!
```

Πίνακας 9: Αποτελέσματα εκπαίδευσης αριθμητικών χαρακτήρων για 10 εποχές



Εικόνα 15: Γραφική αναπαράσταση αποτελεσμάτων Πίνακα 9.

- evaluating the model
- overfitting and underfitting

4.4 ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ

- Process
- Characters being too close to each other
- Unpredictable results

4.5 ΓΡΑΦΙΚΗ ΔΙΕΠΑΦΗ

- pysimplegui

- selecting areas inside images

4.6 ΑΠΟΤΕΛΕΣΜΑΤΑ

- Results process
- Code analysis?

4.7 ΒΕΛΤΙΩΣΕΙΣ

- More clever character detection system
- Model improvements
- Adding an interactive model generation system
- More dataset options.
- Specific font detection.

ΠΑΡΑΡΤΗΜΑ Α – ΠΗΓΑΙΟΣ ΚΩΔΙΚΑΣ ΕΦΑΡΜΟΓΗΣ

[Υλικό που είναι ογκώδες, παραδείγματος χάριν αποτελέσματα ερωτηματολογίων, σχήματα, πίνακες κ.ά. που εμποδίζουν τον αναγνώστη να κατανοήσει τη συνέχεια του

κειμένου, μπορεί να τοποθετηθεί σε παραρτήματα. Η αρίθμηση των παραρτημάτων γίνεται με κεφαλαία ελληνικά γράμματα *A, B, Γ, ...* ενώ σχήματα, σχέδια, πίνακες κ.λπ., που περιλαμβάνονται πρέπει να ονομάζονται *Σχήμα Α1, Πίνακας Β2, κ.λπ.*]

ΑΝΑΦΟΡΕΣ

Ask Any Difference, n.d. *Difference Between OMR and OCR (With Table)* / Ask Any Difference. [Ηλεκτρονικό]

Available at: <https://askanydifference.com/difference-between-omr-and-ocr/>

[Πρόσβαση 5 Ιουνίου 2021].

Brenner, M., 2018. *Artificial Neural Networks: What Every Marketer Should Know* / *Marketing Insider Group*. [Ηλεκτρονικό]

Available at: <https://marketinginsidergroup.com/content-marketing/artificial-neural-networks-every-marketer-know/>

[Πρόσβαση 28 Μαΐου 2021].

Bresnick, J., 2018. *What Is Deep Learning And How Will It Change Healthcare?* / *Health IT Analytics*. [Ηλεκτρονικό]

Available at: <https://healthitanalytics.com/features/what-is-deep-learning-and-how-will-it-change-healthcare>

[Πρόσβαση 28 Μαΐου 2021].

Bridgwater, A., 2018. *The 13 Types Of Data* / *Forbes*. [Ηλεκτρονικό]

Available at: <https://www.forbes.com/sites/adrianbridgwater/2018/07/05/the-13-types-of-data/?sh=10e795633624>

[Πρόσβαση 16 Απριλίου 2021].

Britton, S., 2019. *A Brief History Of OCR* / *CloudTrade*. [Ηλεκτρονικό]

Available at: <https://www.cloud-trade.com/blogs/2019/03/19/a-brief-history-of-ocr>

[Πρόσβαση 30 Μαΐου 2021].

Chen, J., 2020. *Neural Network* / *Investopedia*. [Ηλεκτρονικό]

Available at: <https://www.investopedia.com/terms/n/neuralnetwork.asp>

[Πρόσβαση 19 Μαΐου 2021].

Computer Science Wiki, 2018. *Max-pooling* / *Pooling* / *Computer Science Wiki*.

[Ηλεκτρονικό]

Available at: https://computersciencewiki.org/index.php/Max-pooling_/Pooling

[Πρόσβαση 28 Ιουνίου 2021].

'Data' in Lexico Dictionaries, n.d. *DATA / Definition of DATA by Oxford Dictionary / Lexico Dictionaries*. [Ηλεκτρονικό]

Available at: <https://www.lexico.com/definition/data>

[Πρόσβαση 15 Απριλίου 2021].

datareportal, 2021. *Digital Around The World / statcounter*. [Ηλεκτρονικό]

Available at: <https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet>

[Πρόσβαση 17 Μαΐου 2021].

DeepAI a, n.d. *What is a Hidden Layer? / DeepAI*. [Ηλεκτρονικό]

Available at: <https://deepai.org/machine-learning-glossary-and-terms/hidden-layer-machine-learning>

[Πρόσβαση 19 Μαΐου 2021].

DeepAI b, n.d. *What is Weight (Artificial Neural Network)? / DeepAI*. [Ηλεκτρονικό]

Available at: <https://deepai.org/machine-learning-glossary-and-terms/weight-artificial-neural-network>

[Πρόσβαση 24 Μαΐου 2021].

Diffen, n.d. *Data vs Information - Difference and Comparison / Diffen*. [Ηλεκτρονικό]

Available at: https://www.diffen.com/difference/Data_vs_Information

[Πρόσβαση 15 Απριλίου 2021].

Domo, 2020. *Data Never Sleeps 8.0 / Domo*. [Ηλεκτρονικό]

Available at: <https://www.domo.com/learn/data-never-sleeps-8>

[Πρόσβαση 10 Μαΐου 2021].

Dumoulin, V. & Visin, F., 2018. *A guide to convolution arithmetic for deep learning*.

[Ηλεκτρονικό]

Available at: <https://arxiv.org/abs/1603.07285>

[Πρόσβαση 28 Ιουνίου 2021].

FPT.AI, 2019. *Practical applications of OCR / FPT.AI*. [Ηλεκτρονικό]

Available at: <https://fpt.ai/practical-applications-ocr>

[Πρόσβαση 2 Ιουνίου 2021].

Great Learning Team, 2020. *Types of Neural Networks and Definition of Neural Network / Great Learning*. [Ηλεκτρονικό]

Available at: <https://www.mygreatlearning.com/blog/types-of-neural-networks/>
[Πρόσβαση 24 Μαΐου 2021].

Hardesty, L., 2017. *Explained: Neural Networks / MIT News Office*. [Ηλεκτρονικό]

Available at: <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>
[Πρόσβαση 19 Μαΐου 2021].

History Computer, n.d. *GISMO of David Shepard / History Computer*. [Ηλεκτρονικό]

Available at: <https://history-computer.com/gismo-of-david-shepard/>
[Πρόσβαση 30 Μαΐου 2021].

How We Read, n.d. *Reading Machines / How We Read*. [Ηλεκτρονικό]

Available at: <http://www.howweread.co.uk/gallery/reading-machines/>
[Πρόσβαση 1 Ιουνίου 2021].

IBM Cloud Education, 2020. *Deep neural network / IBM / Neural Networks*.

[Ηλεκτρονικό]

Available at: <https://www.ibm.com/cloud/learn/neural-networks>

[Πρόσβαση 19 Μαΐου 2021].

Internet Live Stats, n.d. *Google Search Statistics / Internet Live Stats*. [Ηλεκτρονικό]

Available at: <https://www.internetlivestats.com/google-search-statistics/>
[Πρόσβαση 10 Μαΐου 2021].

Karandish, F., n.d. *The Comprehensive Guide to Optical Character Recognition (OCR)*

/ Moov AI. [Ηλεκτρονικό]

Available at: <https://moov.ai/en/blog/optical-character-recognition-ocr/>

[Πρόσβαση 9 Ιουνίου 2021].

Khurana, S., Raghavan, R. G. & Kakde, P. M., 2018. *Applications of OCR You Haven't Thought Of / Medium*. [Ηλεκτρονικό]

Available at: <https://medium.com/swlh/applications-of-ocr-you-havent-thought-of->

69a6a559874b

[Πρόσβαση 2 Ιουνίου 2021].

Liu, D., 2017. *A Practical Guide to ReLU* / *Medium*. [Ηλεκτρονικό]

Available at: <https://medium.com/@danqing/a-practical-guide-to-relu-b83ca804f1f7>

[Πρόσβαση 20 Μαΐου 2021].

McLaughlin, E., 2020. *Data Collection* / *TechTarget*. [Ηλεκτρονικό]

Available at: <https://searchcio.techtarget.com/definition/data-collection>

[Πρόσβαση 11 Μαΐου 2021].

M., n.d. *Survey on Image Preprocessing Techniques to Improve OCR Accuracy* / *Medium*. [Ηλεκτρονικό]

Available at: <https://medium.com/technovators/survey-on-image-preprocessing-techniques-to-improve-ocr-accuracy-616ddb931b76>

[Πρόσβαση 9 Ιουνίου 2021].

Sayantini, D., 2020. *Keras vs TensorFlow vs PyTorch : Comparison of the Deep Learning Frameworks* / *Edureka*. [Ηλεκτρονικό]

Available at: <https://www.edureka.co/blog/keras-vs-tensorflow-vs-pytorch/>

[Πρόσβαση 27 Ιουνίου 2021].

Seetharaman, K., 2018. *Financial Applications of Neural Networks* / *Aspire Systems*. [Ηλεκτρονικό]

Available at: <https://blog.aspiresys.com/banking-and-finance/financial-applications-neural-networks>

[Πρόσβαση 28 Μαΐου 2021].

Sharma, A. V., 2017. *Understanding Activation Functions in Neural Networks* / *Medium*. [Ηλεκτρονικό]

Available at: <https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0>

[Πρόσβαση 20 Μαΐου 2021].

Shepard, D. H., 1953. *Apparatus For Reading*. US, Ευρεσιτεχνία Αρ. 2663758.

Steppan, J., 2017. *Sample images from MNIST test dataset*. [Τέχνη].

Talend, n.d. *What is Data Processing?* / *Talend*. [Ηλεκτρονικό]
Available at: <https://www.talend.com/resources/what-is-data-processing/>
[Πρόσβαση 11 Μαΐου 2021].

Tanz, O. & Cambron, C., 2017. *Why the future of deep learning depends on finding good data* / *TechCrunch*. [Ηλεκτρονικό]
Available at: <http://tcrn.ch/2vJhqp3>
[Πρόσβαση 23 Μαΐου 2021].

Taylor, K., n.d. *OCR vs. ICR: What Differentiates the two Character Recognition Software?* / *HiTechNectar*. [Ηλεκτρονικό]
Available at: <https://www.hitechnectar.com/blogs/ocr-vs-icr/>
[Πρόσβαση 7 Ιουνίου 2021].

Veen, F. v. & Leijnen, S., 2019. *The Neural Network Zoo* / *The Asimov Institute*. [Ηλεκτρονικό]
Available at: <https://www.asimovinstitute.org/neural-network-zoo/>
[Πρόσβαση 24 Μαΐου 2021].

Williams, L., 2019. *How Businesses Use Information* / *lumen*. [Ηλεκτρονικό]
Available at: <https://courses.lumenlearning.com/wmopen-introductiontobusiness/chapter/how-businesses-use-information/>
[Πρόσβαση 12 Μαΐου 2021].

Wood, T., n.d. *What is an Activation Function?* / *DeepAI*. [Ηλεκτρονικό]
Available at: <https://deepai.org/machine-learning-glossary-and-terms/activation-function>
[Πρόσβαση 20 Μαΐου 2021].