



## Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής

### Πρόγραμμα Μεταπτυχιακών Σπουδών

«Ψηφιακός Πολιτισμός, Έξυπνες Πόλεις, IoT και Προηγμένες Ψηφιακές Τεχνολογίες»

#### Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΙΚΕΣ ΑΥΤΟΜΑΤΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ MODERN TECHNIQUES OF AUTOMATIC DOCUMENT CLASSIFICATION
Όνοματεπώνυμο Φοιτητή	ΑΝΤΩΝΙΟΣ ΠΟΝΤΙΚΑΚΟΣ
Πατρώνυμο	ΠΕΤΡΟΣ
Αριθμός Μητρώου	ΨΠΟΛ/19050
Επιβλέπων	Δρ Εμμανουήλ Σκόνδρας

Ημερομηνία Παράδοσης ΣΕΠΤΕΜΒΡΙΟΣ 2021

---

Τριμελής Εξεταστική Επιτροπή

Δρ. Εμμανουήλ Σκόνδρας  
Διδάσκων ΠΜΣ

Δημήτριος Βέργαδος  
Καθηγητής

Άγγελος Μιχάλας  
Καθηγητής

---

## ΠΕΡΙΛΗΨΗ

Στο πλαίσιο της εργασίας θα μελετηθούν οι σύγχρονες τεχνικές ψηφιακής ταξινόμησης εγγράφων μέσα από ένα πλήθος προδιαγεγραμμένων κλάσεων. Υπάρχουν εκατομμύρια σύγχρονα έγγραφα που ανήκουν και σε δημόσιους αλλά και σε ιδιωτικούς φορείς, τα οποία δεν έχουν ψηφιοποιηθεί και παραμένουν ανεκμετάλλευτα σε φυσική και μόνο μορφή. Η ταξινόμηση εικόνων που προέρχονται από τέτοια ψηφιοποιημένα έγγραφα αποτελεί ένα σημαντικό βήμα με στόχο την άντληση της πληροφορίας μέσω αυτών, καθώς και την ραγδαία επιτάχυνση διαδικασιών εισαγωγής δεδομένων, επιτρέποντας έτσι τη βέλτιστη διαχείριση ανθρώπινου δυναμικού αλλά και τη διατήρηση της πληροφορίας στο πέρασμα του χρόνου. Ενδεικτικές κατηγορίες σύγχρονων εγγράφων αποτελούν οι φόρμες, τα τιμολόγια/αποδείξεις, άρθρα από εφημερίδες, επιστολές και επιστημονικά άρθρα. Ένα τέτοιο ετερογενές δείγμα αποτελεί πρόκληση για ένα σύστημα αναγνώρισης εγγράφων, καθώς απαιτείται η αρχική ταξινόμηση τους στις παραπάνω κατηγορίες με στόχο την καλύτερη επεξεργασία τους (πχ. εξαγωγή περιοχών ενδιαφέροντος, δεικτοδότηση). Παράλληλα, στη βιβλιογραφία αναδεικνύεται το γεγονός πως καινοτόμες τεχνικές επεξεργασίας εικόνας με χρήση αλγορίθμων Τεχνητής Νοημοσύνης (πχ. Νευρωνικά Δίκτυα) έχουν δώσει σημαντική ώθηση στην αντιμετώπιση του παραπάνω προβλήματος. Στόχος της παρούσας εργασίας είναι να μελετηθούν τα παραπάνω συστήματα Τεχνητής Νοημοσύνης, καθώς και πως αυτά εκπαιδεύονται κατάλληλα και εντάσσονται σε ένα πλήρες σύστημα ταξινόμησης εγγράφων.

**ABSTRACT**

This paper studies recent advances in the Document Classification domain through a predefined set of classes. There are millions of modern documents which belong to public and private entities. Unfortunately, these documents stay unexploited, because they exist only in physical form. Image Classification of such documents is a crucial step towards the exploitation of useful information. Moreover, such processing facilitates the data entry procedure, saves information about these documents in a time-enduring digital form and allows the optimal use of human resources. Common classes of document images are forms, invoices/receipts, newspaper articles, letters and scientific reports. Such heterogeneous sample is a challenge for an Optical Character Recognition system, because an initial categorization in the respective document class is usually required, in order to further assist the processing procedure (eg. extraction of regions of interest, indexing). Recent technological advances in the field of Artificial Intelligence - AI (eg. Neural Networks) and innovative Image Processing techniques have been proven to be a priceless asset for the Document Classification problem. This paper aims to provide an introductory presentation of such AI systems and especially how do they apply on a Document Classification system.

<b>ΠΕΡΙΛΗΨΗ</b>	3
<b>ABSTRACT</b>	3
<b>ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ</b>	8
<b>ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ</b>	10
<b>ΔΟΜΗ ΕΡΓΑΣΙΑΣ</b>	11
<b>ΕΙΣΑΓΩΓΗ</b>	12
<b>ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΜΕΛΕΤΗ ΑΥΤΟΜΑΤΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ</b>	15
2.1. ΓΕΝΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΚΑΙ ΕΦΑΡΜΟΓΩΝ	15
2.2. ΣΥΣΤΗΜΑΤΑ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΓΙΑ ΤΑΞΙΝΟΜΗΣΗ ΕΓΓΡΑΦΩΝ	35
<b>ΕΠΙΣΚΟΠΗΣΗ ΒΑΣΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ ΚΑΙ ΤΕΧΝΙΚΩΝ ΕΚΠΑΙΔΕΥΣΗΣ ΣΥΣΤΗΜΑΤΩΝ ΨΗΦΙΑΚΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ</b>	44
<b>ΣΥΓΚΡΙΤΙΚΗ ΜΕΛΕΤΗ</b>	49
4.1. ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΨΗΦΙΑΚΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ	49
4.2. ΜΕΤΡΙΚΕΣ ΑΠΟΤΙΜΗΣΗΣ ΨΗΦΙΑΚΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ	51
4.3. ΣΥΓΚΡΙΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ	53
<b>ΟΛΟΚΛΗΡΩΜΕΝΟ ΣΥΣΤΗΜΑ ΑΥΤΟΜΑΤΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ</b>	61
5.1. ΕΠΙΣΚΟΠΗΣΗ ΣΥΣΤΗΜΑΤΩΝ ΑΥΤΟΜΑΤΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ	61
5.2. ΓΕΝΙΚΕΣ ΠΡΟΔΙΑΓΡΑΦΕΣ	63
5.3. ΑΝΑΛΥΣΗ ΑΠΑΙΤΗΣΕΩΝ ΛΕΙΤΟΥΡΓΙΚΟΤΗΤΑΣ	66
5.4. ΣΧΕΔΙΑΣΜΟΣ ΣΥΣΤΗΜΑΤΟΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ	70
<b>ΣΥΝΟΨΗ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ</b>	74
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ - ΑΝΑΦΟΡΕΣ</b>	76
<b>ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ</b>	78
<b>ΣΥΝΤΜΗΣΕΙΣ - ΑΡΚΤΙΚΟΛΕΞΑ - ΑΚΡΩΝΥΜΙΑ</b>	82



## ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

	Σελίδα
Σχήμα 1. α) Απλοποιημένο σχήμα βιολογικών νευρώνων. β) μηχανισμός μεταφοράς σημάτων μεταξύ νευρώνων	14
Σχήμα 2. Αναλογία βιολογικού και τεχνητού νευρώνα	15
Σχήμα 3. Πεδίο Εφαρμογών Τεχνητής Νοημοσύνης	16
Σχήμα 4. Αυτόματη Αναγνώριση Αντικειμένων/Οντοτήτων	18
Σχήμα 5. Αναγνώριση σκηνής από ένα σύστημα αυτόνομης οδήγησης	18
Σχήμα 6. Αυτόματη αναγνώριση περιοχών ενδιαφέροντος σε μαγνητική τομογραφία	19
Σχήμα 7. Αυτόματη περιγραφή και περίληψη Εικόνας	19
Σχήμα 8. Βασικές Αρχιτεκτονικές Νευρωνικών Δικτύων	26
Σχήμα 9. Προϋπολογισμός που αφορά την επενδύσεις σε θέματα Τεχνητής Νοημοσύνης στην Ευρώπη	27
Σχήμα 10. Βασικό διάγραμμα εκπαίδευσης Νευρωνικών Δικτύων	30
Σχήμα 11. Σύστημα Τεχνητού Νευρωνικού Δικτύου	31
Σχήμα 12. Δείγματα εγγράφων προς ταξινόμηση	32
Σχήμα 13. Διαχωρισμός εικόνας σε άνω, κάτω, αριστερό και δεξιό σώμα	33
Σχήμα 14. Υβριδική αρχιτεκτονική εξαγωγής “ρηχών” και “βαθέων” χαρακτηριστικών	34
Σχήμα 15. Συγκριτική μελέτη για διάφορες αρχιτεκτονικές Νευρωνικών Δικτύων και διάφορους τρόπους αρχικοποίησης παραμέτρων	34
Σχήμα 16. Δείγματα από τις βάσεις δεδομένων φυσικών εικόνων (άνω) και εικόνων εγγράφων (κάτω)	35
Σχήμα 17. Παράδειγμα μετασχηματισμού εικόνας για επαύξηση δεδομένων	35
Σχήμα 18. Αρχιτεκτονική Νευρωνικού Δικτύου συστήματος ταξινόμησης εγγράφων	37
Σχήμα 19. Αναπαράσταση υβριδικής αρχιτεκτονικής συστήματος ταξινόμησης εγγράφων	38
Σχήμα 20. Αρχιτεκτονική συστήματος ταξινόμησης εγγράφων που συνδυάζει οπτική και κειμενική πληροφορία	38
Σχήμα 21. Αρχιτεκτονική συστήματος ταξινόμησης εγγράφων που συνδυάζει	39

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

οπτική, κειμενική και δομική πληροφορία

Σχήμα 22. Σιγμοειδής συνάρτηση ενεργοποίησης νευρώνα	40
Σχήμα 23. “Απενεργοποίηση” Νευρώνων βάσει της τεχνικής Dropout	41
Σχήμα 24. Τεχνική επαύξησης δεδομένων μέσω τυχαίων τεχνητών παραμορφώσεων της εικόνας	42
Σχήμα 25. Παραδείγματα εικόνων από τη βάση δεδομένων Tobacco3482	44
Σχήμα 26. Παραδείγματα εικόνων από τη βάση δεδομένων RVL-CDIP	45
Σχήμα 27. Παράδειγμα Error Analysis για τη βάση δεδομένων Tobacco3482	53
Σχήμα 28. Στιγμιότυπο εφαρμογής Ταξινόμησης Εγγράφων της SAP	55
Σχήμα 29. Στιγμιότυπο εφαρμογής Ταξινόμησης Εικόνας της Google (Cloud Vision API)	57
Σχήμα 30. Σύγκριση εύρους (bandwidth) καρτών γραφικών (GPU) με επεξεργαστές (CPU) για την εκπαίδευση Νευρωνικών Δικτύων	59
Σχήμα 31. Σύγκριση εύρους (bandwidth) καρτών γραφικών (GPU) με επεξεργαστές (CPU) για την εκπαίδευση Νευρωνικών Δικτύων	59
Σχήμα 32. Διάγραμμα Ροής προτεινόμενης λειτουργικότητας εφαρμογής Αυτόματης Ταξινόμησης Εγγράφων.	60
Σχήμα 33. Παραδείγματα καμπυλών μάθησης, με χρήση της Βιβλιοθήκης TensorFlow	62
Σχήμα 34. Προτεινόμενη αρχιτεκτονική συστήματος Ταξινόμησης Εγγράφων	63



## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

	Σελίδα
ΠΙΝΑΚΑΣ 1: Χαρακτηριστικά Τεχνητών και Βιολογικών Νευρωνικών Δικτύων	15
ΠΙΝΑΚΑΣ 2: Βασικές αρχιτεκτονικές Νευρωνικών Δικτύων και ο σκοπός τους	20
ΠΙΝΑΚΑΣ 3: Σύγκριση Machine Learning και Deep learning	28
ΠΙΝΑΚΑΣ 4: Μήτρα σύγχυσης αποτελεσμάτων ενός συστήματος πρόβλεψης καρκινικών όγκων σε μαστογραφίες	47
ΠΙΝΑΚΑΣ 5: Μήτρα σύγχυσης αποτελεσμάτων για περισσότερες από δύο κλάσεις	47
ΠΙΝΑΚΑΣ 6: Αποτελέσματα ταξινόμησης εγγράφων (F-Measure) χρησιμοποιώντας διαφορετικά μέρη του εγγράφου	48
ΠΙΝΑΚΑΣ 7: Αποτελέσματα ταξινόμησης εγγράφων (F-Measure), κάνοντας χρήση διάφορων τεχνικών αρχικοποίησης	50
ΠΙΝΑΚΑΣ 8: Αποτελέσματα ταξινόμησης εγγράφων (F-Measure) χρησιμοποιώντας διαφορετικά είδη μετασχηματισμών για επαύξηση δεδομένων	50
ΠΙΝΑΚΑΣ 9: Αποτελέσματα ταξινόμησης εγγράφων (F-Measure) για διάφορες μεθοδολογίες που προτείνονται στη βιβλιογραφία	51

---

## ΔΟΜΗ ΕΡΓΑΣΙΑΣ

Στο Κεφάλαιο 1 γίνεται η εισαγωγή στο πρόβλημα, παρουσιάζεται η έννοια της Αυτόματης Ταξινόμησης Εγγράφων και πως αυτό αντιμετωπίζεται στην καθημερινότητα κατά την ιστορία, καθώς και η ανάγκη επίλυσής του με στόχο τη διευκόλυνση και επιτάχυνση καθημερινών διαδικασιών.

Στο Κεφάλαιο 2 δίνεται η επισκόπηση των Τεχνητών Νευρωνικών Δικτύων και Εφαρμογών, καθώς και η τρέχουσα τεχνολογική στάθμη στο χώρο της Ταξινόμησης Εγγράφων μέσα από την παρουσίαση των σύγχρονων τεχνικών που συνδυάζουν αλγόριθμους Μηχανικής Μάθησης (Machine Learning) για την εκπαίδευση Νευρωνικών Δικτύων.

Στο Κεφάλαιο 3 παρουσιάζονται διάφορες τεχνικές που επιτρέπουν την εκπαίδευση συστημάτων καθώς και τη βελτιστοποίηση της αποδοτικότητάς τους.

Στο Κεφάλαιο 4 παρουσιάζονται συγκριτικά αποτελέσματα των παραπάνω μεθοδολογιών μέσα από την εφαρμογή τους πάνω σε διάφορες βάσεις δεδομένων εικόνων εγγράφων.

Στο Κεφάλαιο 5, γίνεται μία παρουσίαση ενός Ολοκληρωμένου Συστήματος Αυτόματης Ταξινόμησης Εγγράφων. Δίνονται οι γενικές προδιαγραφές που πρέπει αυτό να πληροί, καθώς και μία σύντομη επισκόπηση παρόμοιων εφαρμογών που υπάρχουν ήδη. Στη συνέχεια, αναλύονται οι απαιτήσεις λειτουργικότητας που πρέπει να καλύπτονται ώστε γίνεται σωστή εκμετάλλευση της τεχνολογίας αυτόματης ψηφιακής ταξινόμησης εγγράφων και δίνεται η προτεινόμενη αρχιτεκτονική που πρέπει να ακολουθεί ένα τέτοιο σύστημα.

Τέλος, στο Κεφάλαιο 6 πραγματοποιείται μία συμπερασματική σύνοψη πάνω στο πρόβλημα της Ταξινόμησης Εγγράφων.

## 1. ΕΙΣΑΓΩΓΗ

Στη σημερινή εποχή υπάρχουν εκατομμύρια σύγχρονα έγγραφα που ανήκουν και σε δημόσιους αλλά και σε ιδιωτικούς φορείς, τα οποία δεν έχουν ψηφιοποιηθεί και παραμένουν ανεκμετάλλευτα σε φυσική και μόνο μορφή. Κατά τους προηγούμενους αιώνες ακόμα και από τα αρχαία χρόνια, όλες οι συναλλαγές των ανθρώπων συνοδεύοντουσαν σχεδόν πάντα από κάποιο έγγραφο που αποτελούσε την απόδειξη της εκάστοτε συναλλαγής. Αγοραπωλησίες ακίνητης περιουσίας και έγγραφα επιχειρηματικών συναλλαγών (τιμολόγια/αποδείξεις), έντυπα/φόρμες προς συμπλήρωση και υποβολή σε κρατικούς φορείς ή τράπεζες, επιστολές, γράμματα επίσημου ή ανεπίσημου χαρακτήρα αποτελούν ενδεικτικά παραδείγματα μέσα από ένα τεράστιο πλήθος παρόμοιων κατηγοριών εγγράφων. Η ενημέρωση των πολιτών λάμβανε χώρα κατά συντριπτική πλειοψηφία μέσω των εφημερίδων και διαφόρων περιοδικών ψυχαγωγικού/ενημερωτικού χαρακτήρα.

Τις τελευταίες δεκαετίες και παράλληλα με την είσοδο των υπολογιστών στην καθημερινότητα, ξεκίνησαν βήματα ώστε όλες οι παραπάνω διαδικασίες να “ψηφιοποιηθούν”. Διαπιστώθηκε, δηλαδή, ότι είναι προτιμότερο όλα τα υπάρχοντα έγγραφα να υπάρχουν σε έναν υπολογιστή, παρά να στοιβάζονται σε φυσική μορφή και να καταλαμβάνουν τεράστιους χώρους που απαιτείται να δεσμευτούν για τη φύλαξή τους. Επίσης, η αποθήκευση εγγράφων σε υπολογιστή αποτελεί μία ασφαλέστερη επιλογή καθώς μειστοποιούνται οι δυνατότητες ανάκτησης σε περίπτωση απώλειας, φυσικών καταστροφών κ.λπ. . Επιπλέον, ο χρόνος αναζήτησης πληροφορίας είναι συντριπτικά μικρότερος όπως για παράδειγμα σε σύγκριση με την αναζήτηση ενός ονόματος μέσα από ένα έγγραφο καταλόγου ονομάτων δεκάδων ή/και εκατοντάδων σελίδων.

Οι παραπάνω διαπιστώσεις έχουν οδηγήσει σήμερα στη δημιουργία πολύπλοκων εφαρμογών που διαχειρίζονται τέτοια έγγραφα. Βάσεις δεδομένων που περιέχουν εκατομμύρια εικόνες εγγράφων δίνουν τη δυνατότητα πρόσβασης μέσω Διαδικτύου σε όλους τους ενδιαφερόμενους και η σημερινές ανάγκες επιτάσσουν τον “ψηφιακό μετασχηματισμό” όλων των διαδικασιών συναλλαγής και διακίνησης κάθε είδους εγγράφων. Οι εφαρμογές εμπλουτίζονται διαρκώς με εργαλεία λογισμικού που προσπαθούν α) να βελτιώσουν την ποιότητα των εγγράφων που ψηφιοποιούνται μέσω της σάρωσης (scanning), ώστε τα ψηφιακά αντίγραφα που προκύπτουν να αποτυπώνουν σε ρεαλιστικό βαθμό το αυθεντικό και ακόμα και να βελτιώνουν σημεία που έχουν υποστεί φθορά και β) να μετατρέψουν το έγγραφο σε ένα πλήρως εμπλουτισμένο ηλεκτρονικό αντίγραφο, όπως θα ήταν αν είχε δημιουργηθεί εξ αρχής ηλεκτρονικά, καθώς κάτι τέτοιο δίνει αμέτρητες δυνατότητες όπως αυτές της αναζήτησης, επεξεργασίας και κατανόησης του.

Για την επίτευξη των παραπάνω στόχων λοιπόν, δηλαδή της μετατροπής ενός εγγράφου από απλή εικόνα αποθηκευμένη σε υπολογιστή σε ένα πλήρες ηλεκτρονικό αντίγραφο, έχουν αναπτυχθεί σημαντικά εργαλεία λογισμικού που πραγματοποιούν διάφορες λειτουργίες, όπως ενδεικτικά:

- Μετατροπή της εικόνας από έγχρωμη σε ασπρόμαυρη για συμπίεση μεγέθους και αύξηση επεξεργασίας (Διαδικτικοποίηση - Binarization)
- Βελτίωση Ποιότητας Εικόνας (Image Enhancement) και Αφαίρεση Θορύβου (Noise Removal)
- Κατάτμηση Εικόνας (Image Segmentation)
- Οπτική Αναγνώριση Χαρακτήρων (Optical Character Recognition - OCR)

Γίνεται εύκολα αντιληπτό πως η Οπτική Αναγνώριση Χαρακτήρων αποτελεί τον απώτερο στόχο της ψηφιοποίησης, διότι μέσω αυτής επιτυγχάνεται η κατανόηση και αναγνώριση του περιεχομένου μιας εικόνας εγγράφου, άρα και επίτευξη δημιουργίας ηλεκτρονικού αντιγράφου.

### Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

Παρά τη διαρκή εξέλιξη των παραπάνω λειτουργιών και την ενσωμάτωσή τους στην ψηφιοποίηση εγγράφων, δισεκατομμύρια έγγραφα σε όλο τον κόσμο παραμένουν ακόμα χωρίς ηλεκτρονικά αντίγραφα είτε αυτά αποτελούν αρχαικό υλικό ή νέα έγγραφα που απλά σαρώνονται και αποθηκεύονται στον υπολογιστή ως εικόνες. Η ποικιλομορφία τους καθιστά σχεδόν αδύνατη τη δημιουργία μιας εφαρμογής επεξεργασίας καθολικής χρήσης και απαιτείται συνήθως ιδιαίτερη επεξεργασία ανάλογα με το είδος στην κάθε περίπτωση.

Απαιτείται λοιπόν η γνώση ορισμένων πληροφοριών εκ των προτέρων (*a priori*) για τη φύση του εγγράφου, όπως για παράδειγμα η κατηγορία του ή αλλιώς κλάση που ανήκει. Ως μία φυσική αναλογία μπορεί να θεωρηθεί η ευκολία με την οποία ένας υπάλληλος νοσοκομείου αναζητά πληροφορίες μέσα από τεράστιο πλήθος εξετάσεων ασθενών που είναι της τάξης των χιλιάδων τουλάχιστον. Η αναζήτησή του είναι πολύ πιο αποτελεσματική όταν ο υπάλληλος γνωρίζει που βρίσκεται ο κάθε φάκελος, δηλαδή τα αρχεία του είναι **ταξινομημένα** (πχ. αλφαβητικά ή χρονολογικά ή ανά είδος εξέτασης). Αντίστοιχα, ένα λογισμικό OCR μπορεί να έχει πολύ καλύτερα αποτελέσματα αν γνωρίζει *a priori* την κατηγορία του εγγράφου. Ως συμπέρασμα των παραπάνω διαπιστώνει κανείς ότι απαιτείται η **Ψηφιακή Ταξινόμηση των Εγγράφων** προτού λάβει χώρα η διαδικασία αναγνώρισής τους (OCR). Μιλώντας για τεράστιο όγκο δεδομένων, η ταξινόμηση αυτή είναι απαραίτητο να γίνεται με αυτόματο τρόπο (μέσω ενός αλγορίθμου ταξινόμησης) και με τη βέλτιστη ακρίβεια.

Η παρούσα εργασία μελετά λοιπόν την τρέχουσα τεχνολογική στάθμη που αφορά σε τεχνικές και μεθόδους που αντιμετωπίζουν το πρόβλημα της αυτόματης **Ταξινόμησης Εγγράφων (Document Classification)**, μέσα από ένα πλήθος προδιαγεγραμμένων κατηγοριών. Η μελέτη αυτή ακολουθεί τρεις βασικούς άξονες:

- Παρουσιάζονται συστήματα που επεξεργάζονται κατά κύριο λόγο σύγχρονα έγγραφα που περιέχουν κυρίως τυπωμένο κείμενο αλλά και χειρόγραφο σε ορισμένες περιπτώσεις. Δεν αναλύεται η επεξεργασία ιστορικών εγγράφων όπως ιστορικά χειρόγραφα, καθώς τα τελευταία απαιτούν αρκετά διαφορετικό τρόπο προσέγγισης που εκτείνεται εκτός πλαισίων της παρούσας μελέτης.
- Η παρουσίαση των τεχνικών και των μεθόδων που συναντώνται στη βιβλιογραφία και παρουσιάζονται δεν αναλύονται σε τεχνικό επίπεδο ούτε αναλύεται το μαθηματικό τους υπόβαθρο. Η εστίαση πραγματοποιείται σε επίπεδο ερμηνείας και κατανόησης τρόπου λειτουργίας των συστημάτων και πως αυτός αποτυπώνεται συγκριτικά μέσα από την εφαρμογή σε δημοφιλείς βάσεις δεδομένων με εικόνες εγγράφων. Παράλληλα, στόχος είναι τα περιεχόμενα της μελέτης αυτής να είναι αντιληπτά ακόμα και από μη εξοικειωμένους με το χώρο αναγνώστες, καθώς και να είναι εφικτή η χρήση τους μέσα από φιλικές-προς-το-χρήστη πλατφόρμες.
- Επιπλέον, η μελέτη της τρέχουσας τεχνολογικής στάθμης επικεντρώνεται στα συστήματα που κάνουν χρήση αλγορίθμων Νευρωνικών Δικτύων (Neural Networks) και Βαθιάς Μάθησης (Deep Learning), με στόχο να δοθεί έμφαση στην είσοδο της Τεχνητής Νοημοσύνης (Artificial Intelligence - AI) στην καθημερινότητα, πράγμα το οποίο παρατηρείται έντονα τα τελευταία χρόνια.

Με βάση τα παραπάνω, η συνέχεια της εργασίας αυτής οργανώνεται ως ακολούθως:

- Στο Κεφάλαιο 2 δίνεται μία επισκόπηση των Νευρωνικών Δικτύων και των Εφαρμογών, καθώς και η τρέχουσα τεχνολογική στάθμη στο χώρο της Ταξινόμησης Εγγράφων μέσα από την παρουσίαση των σύγχρονων τεχνικών που συνδυάζουν αλγόριθμους Μηχανικής Μάθησης (Machine Learning) για την εκπαίδευση Νευρωνικών Δικτύων.
- Στο Κεφάλαιο 3 παρουσιάζονται διάφορες τεχνικές που επιτρέπουν την εκπαίδευση συστημάτων καθώς και τη βελτιστοποίηση της αποδοτικότητάς τους.

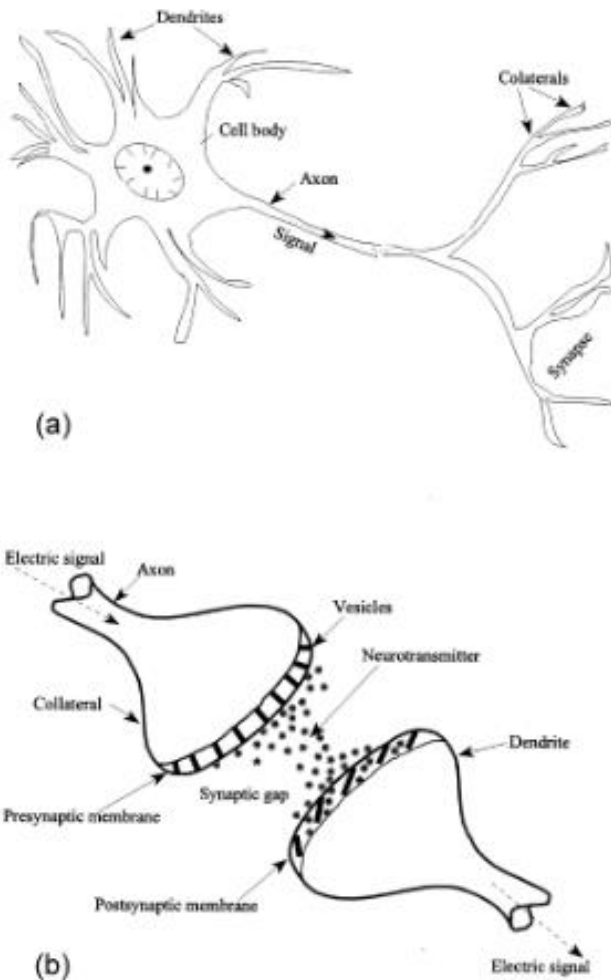
### Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

- Στο Κεφάλαιο 4 παρουσιάζονται συγκριτικά αποτελέσματα των παραπάνω μεθοδολογιών μέσα από την εφαρμογή τους πάνω σε διάφορες βάσεις δεδομένων εικόνων εγγράφων.
- Στο Κεφάλαιο 5, γίνεται μία παρουσίαση ενός Ολοκληρωμένου Συστήματος Αυτόματης Ταξινόμησης Εγγράφων. Δίνονται οι γενικές προδιαγραφές που πρέπει αυτό να πληροί, καθώς και μία σύντομη επισκόπηση παρόμοιων εφαρμογών που υπάρχουν ήδη. Στη συνέχεια, αναλύονται οι απαιτήσεις λειτουργικότητας που πρέπει να καλύπτονται ώστε γίνεται σωστή εκμετάλλευση της τεχνολογίας αυτόματης ψηφιακής ταξινόμησης εγγράφων και δίνεται η προτεινόμενη αρχιτεκτονική που πρέπει να ακολουθεί ένα τέτοιο σύστημα.
- Τέλος, στο Κεφάλαιο 6 πραγματοποιείται μία συμπερασματική σύνοψη πάνω στο πρόβλημα της Ταξινόμησης Εγγράφων.

## 2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΜΕΛΕΤΗ ΑΥΤΟΜΑΤΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ

### 2.1. ΓΕΝΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΚΑΙ ΕΦΑΡΜΟΓΩΝ

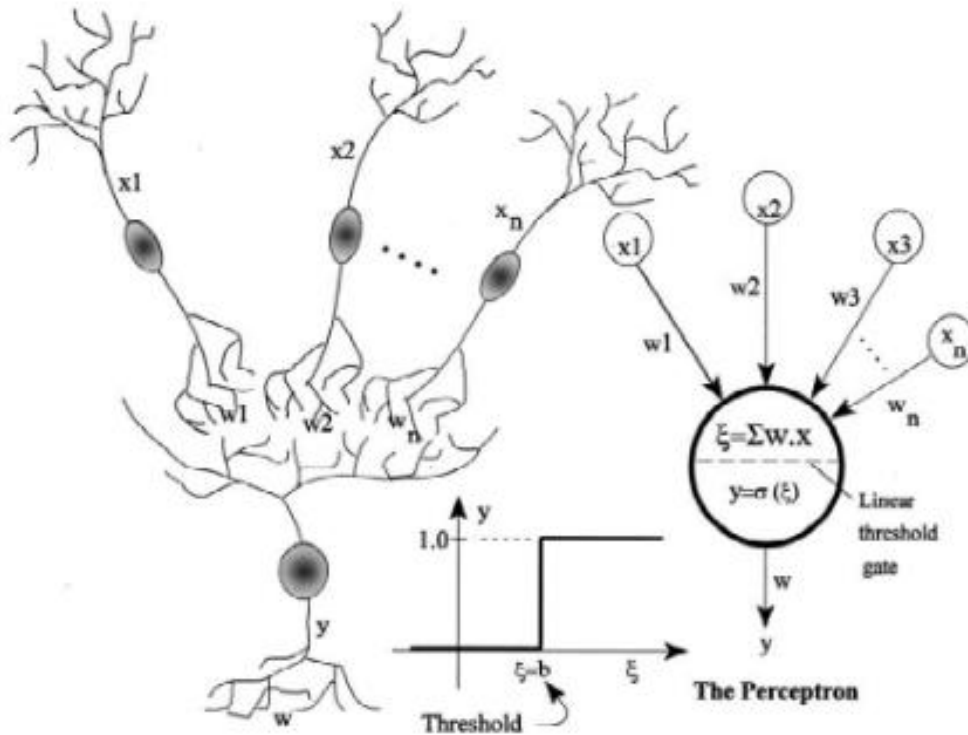
Η έννοια των Τεχνητών Νευρωνικών Δικτύων πηγάζει από αυτήν των βιολογικών νευρώνων του ανθρώπινου οργανισμού, οι οποίοι αποτελούν το βασικό δομικό συστατικό του νευρικού συστήματος. Συνεπώς, είναι απαραίτητη η κατανόηση της αναλογίας μεταξύ τους και του τρόπου λειτουργίας τους. Όπως αναφέρεται στο [1] (Σχήμα 1), το ανθρώπινο νευρικό σύστημα αποτελείται από δισεκατομμύρια νευρώνες διαφορετικού τύπου και μήκους, συσχετιζόμενοι ταυτόχρονα και με τη θέση τους στο ανθρώπινο σώμα. Ένας νευρώνας αποτελείται από 3 λειτουργικά μέρη. Τους δενδρίτες, το κυτταρικό σώμα και τους άξονες. Το κυτταρικό σώμα έχει έναν πυρήνα που περιέχει πληροφορίες σχετικά με τα χαρακτηριστικά κληρονομικότητας, καθώς και πλάσμα που κρατά τον μοριακό εξοπλισμό που παράγει το υλικό που απαιτείται από τον νευρώνα. Οι δενδρίτες λαμβάνουν σήματα από άλλους νευρώνες και τα προωθούν στο κυτταρικό σώμα. Ο άξονας με τις διακλαδώσεις του λαμβάνει σήματα από το κυτταρικό σώμα και τα μεταφέρει μέσω συνάψεων (μικροσκοπικά κενά) σε δενδρίτες γειτονικών νευρώνων.



Σχήμα 1. α) Απλοποιημένο σχήμα βιολογικών νευρώνων. β) μηχανισμός μεταφοράς σημάτων μεταξύ νευρώνων

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

Μία απλουστευμένη αναλογία μεταξύ βιολογικών και τεχνητών νευρώνων (Σχήμα 2) είναι ότι οι συνδέσεις μεταξύ των κόμβων του δικτύου είναι οι άξονες και οι δενδρίτες, οι συνάψεις αναπαριστώνται από βάρη (πολλαπλασιαστές ή αλλιώς τιμές βαρύτητας) και το κατώφλι προσεγγίζει τη δραστηριότητα του κυτταρικού σώματος. Πολλές φορές ο όρος νευρώνας μεταφράζεται και ως perceptron.



Σχήμα 2. Αναλογία βιολογικού και τεχνητού νευρώνα [1]

Στον παρακάτω πίνακα (Πίνακας 1) παρουσιάζονται σε συγκριτική μορφή διάφορα χαρακτηριστικά που αφορούν τα δύο αυτά είδη δικτύων (τεχνητά/βιολογικά):

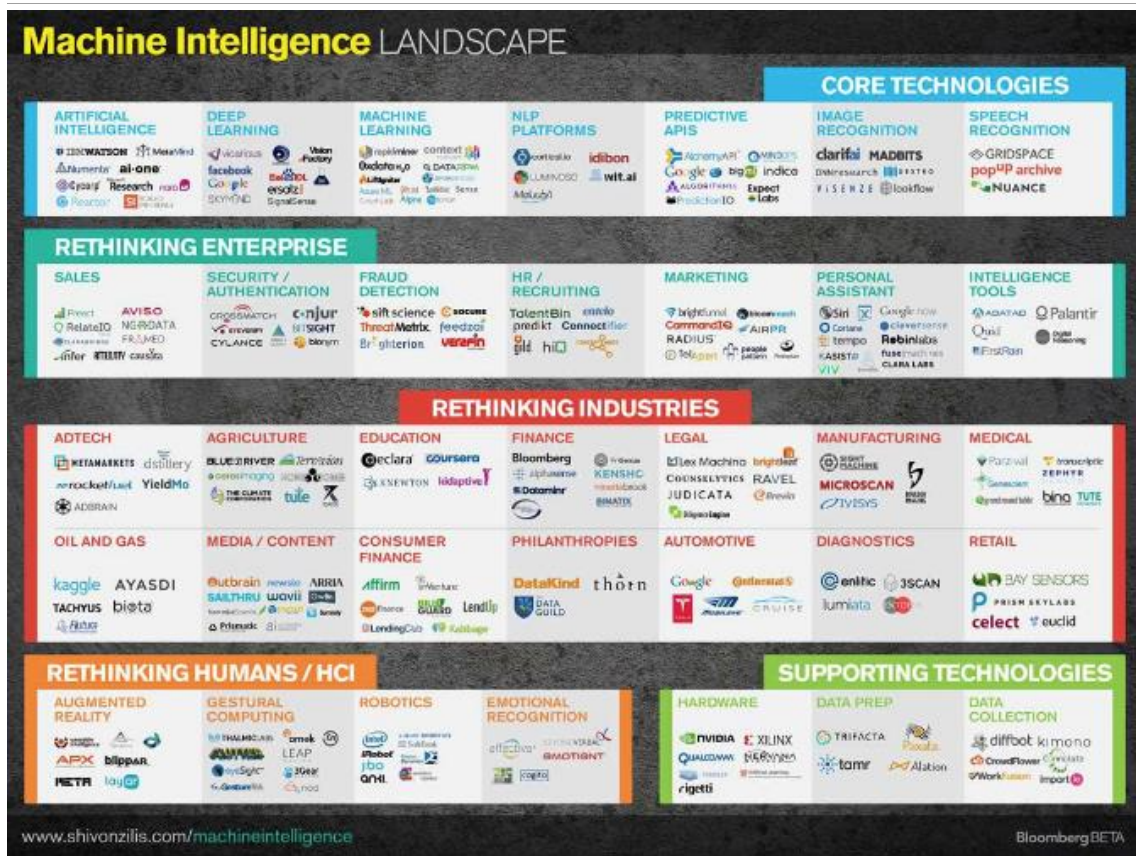
**ΠΙΝΑΚΑΣ 1: Χαρακτηριστικά Τεχνητών και Βιολογικών Νευρωνικών Δικτύων**

Χαρακτηριστικό	Τεχνητό Νευρωνικό Δίκτυο	Βιολογικό (Πραγματικό Νευρωνικό Δίκτυο)
Ταχύτητα	Γρηγορότερο στην	Αργότερο στην επεξεργασία

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

	επεξεργασία πληροφορίας με χρόνο απόκρισης σε nanoseconds.	πληροφορίας με χρόνο απόκρισης σε milliseconds.
<b>Επεξεργασία</b>	Σειριακή επεξεργασία	Μαζική παράλληλη επεξεργασία
<b>Μέγεθος &amp; Πολυπλοκότητα</b>	Μικρότερο μέγεθος και πολυπλοκότητα.	Τεράστιο μέγεθος και πολυπλοκότητα.
<b>Αποθήκευση</b>	Αποθήκευση νέας πληροφορίας μπορεί να γίνει αντικαθιστώντας παλαιότερη	Ένα πολύ περίπλοκο και πυκνό δίκτυο διασυνδεδεμένων νευρώνων που περιέχουν νευρώνες της τάξης των 10 <sup>11</sup> με 10 <sup>15</sup> διασυνδέσεων.
<b>Ανοχή Σφάλματος</b>	Μη ανοχή σε σφάλματα. Δεν είναι δυνατή η ανάκτηση κατεστραμμένων πληροφοριών σε περίπτωση βλάβης του συστήματος.	Η αποθήκευση πληροφοριών είναι προσαρμόσιμη σημαίνει ότι προστίθενται νέες πληροφορίες προσαρμόζοντας τα δυνατά σημεία διασύνδεσης χωρίς να καταστρέφονται οι παλιές πληροφορίες.
<b>Μηχανισμός Ελέγχου</b>	Υπάρχει συγκεκριμένη μονάδα ελέγχου που ελέγχει τις διαδικασίες υπολογισμών	Δεν υπάρχει συγκεκριμένη μονάδα ελέγχου.





Σχήμα 3. Πεδίο Εφαρμογών Τεχνητής Νοημοσύνης<sup>1</sup>

Οι εφαρμογές των Τεχνητών Νευρωνικών Δικτύων εκτείνονται σε ένα τεράστιο φάσμα. Στην ιδεατή περίπτωση μπορεί να υποθέσει κανείς ότι δύναται να εφαρμοστούν σε όλες τις περιπτώσεις λειτουργίας του ανθρώπινου εγκεφάλου, κάτι που προκύπτει και από την παρουσίαση της αναλογίας Τεχνητού/Βιολογικού Νευρώνα. Καταλαβαίνει λοιπόν κανείς ότι το πλήθος των δυνατών εφαρμογών τείνει προς το άπειρο. Ενδεικτικά παραδείγματα εφαρμογής δίνονται στη συνέχεια:

- **Υπολογιστική Όραση:** Νευρωνικά δίκτυα ανίχνευσης αντικειμένων, ταξινόμηση εικόνας, αποκατάσταση εικόνας και τμηματοποίηση εικόνας. Ενισχύουν ακόμη και την αναγνώριση των χειρόγραφων ψηφίων σε ένα σύστημα υπολογιστή.
- **Ανάλυση συναισθήματος:** Συστήματα ανάλυσης ειδήσεων χρησιμοποιούν τεχνολογίες νευρωνικών δικτύων βαθιάς μάθησης για να φιλτράρουν αρνητικά νέα και να δείχνουν μόνο τα θετικά πράγματα που συμβαίνουν.
- **Αυτόματες Μεταφράσεις:** Αυτόματη αναγνώριση και μετάφραση κειμένου σε οποιαδήποτε γλώσσα, ακόμα και διάλεκτο.
- **Εμπειρία πελατών:** Πολλές επιχειρήσεις χρησιμοποιούν ήδη συστήματα Νευρωνικών Δικτύων για να εργαστούν στην εμπειρία των πελατών. Στα παραδείγματα περιλαμβάνονται διαδικτυακές πλατφόρμες αυτόματης εξυπηρέτησης. Επιπλέον, πολλοί οργανισμοί εξαρτώνται πλέον από τη βαθιά μάθηση για να δημιουργήσουν αξιόπιστες ροές εργασίας. Οι περισσότεροι από εμάς είναι ήδη εξοικειωμένοι με τη χρήση των ρομπότ-επικοινωνίας.

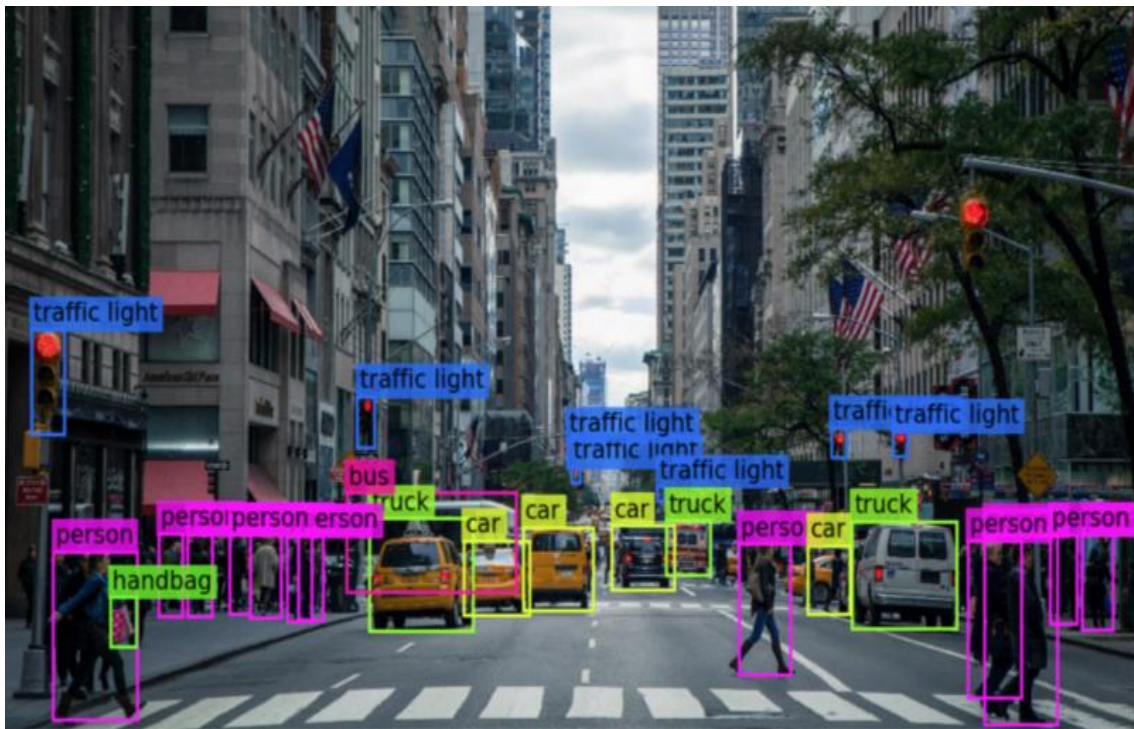
<sup>1</sup> <https://shivonzilis.com/machineintelligence>

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

- **Ανάλυση Εικόνας/Βίντεο:** Αυτόνομη Οδήγηση, αυτόματη περιγραφή γεγονότων, τρισδιάστατες απεικονίσεις και εφαρμογές επαυξημένης και εικονικής πραγματικότητας. Αναγνώριση Γραφικού Χαρακτήρα, Ταξινόμηση Εγγράφων
- **Αυτόματη Απομαγνητοφώνηση:** Ιατρικές Γνωματεύσεις, δικαστικές αποφάσεις, πρακτικά συνεδριάσεων.
- **Εφαρμογές Τραπεζικού τομέα:** Φθορά πιστωτικών καρτών, αξιολόγηση πιστωτικών και αιτήσεων δανείων, αξιολόγηση απάτης και κινδύνου, και καθυστερήσεις δανείων
- **Επιχειρηματική Ανάλυση Δεδομένων:** Μοντελοποίηση συμπεριφοράς πελατών, τμηματοποίηση πελατών, τάση απάτης, έρευνα αγοράς, συνδυασμός αγοράς, δομή αγοράς και μοντέλα τριβής, προεπιλογή, αγορά και ανανεώσεις
- **Εφαρμογές Άμυνας:** Αντιτρομοκρατία, αναγνώριση προσώπου, εξαγωγή χαρακτηριστικών, καταστολή θορύβου, διάκριση αντικειμένων, αισθητήρες, σόναρ, ραντάρ και επεξεργασία σήματος εικόνας, αναγνώριση σήματος / εικόνας, παρακολούθηση στόχου και καθοδήγηση όπλου
- **Εφαρμογές Εκπαίδευσης:** Προσαρμοστικό λογισμικό μάθησης, δυναμική πρόβλεψη, ανάλυση και πρόβλεψη εκπαιδευτικού συστήματος, μοντελοποίηση επιδόσεων μαθητών και προφίλ προσωπικοτήτων
- **Εφαρμογές Χρηματοοικονομικού τομέα:** Αξιολογήσεις εταιρικών ομολόγων, εταιρική χρηματοοικονομική ανάλυση, ανάλυση χρήσης πιστωτικών ορίων, πρόβλεψη τιμής νομίσματος, παροχή συμβουλών δανείων, έλεγχος στεγαστικών δανείων, αξιολόγηση ακινήτων και συναλλαγές χαρτοφυλακίου
- **Εφαρμογές Ιατρικής:** Ανάλυση καρκινικών κυττάρων, μείωση δαπανών και βελτίωση ποιότητας για νοσοκομειακά συστήματα, βελτιστοποίηση διαδικασίας μεταμόσχευσης και σχεδιασμός προσθετικών, επεξεργασία ακτινογραφιών, μαγνητικών και αξονικών τομογραφιών, αυτόματη διάγνωση.

Ένα γράφημα που αποτυπώνει το ευρύτερο πεδίο της Τεχνητής Νοημοσύνης και των εφαρμογών του, δόθηκε στο Σχήμα 3. Εστιάζοντας σε εφαρμογές Υπολογιστικής Όρασης οι οποίες συνάδουν και με το πρόβλημα της Ταξινόμησης Εγγράφων, παρατίθενται διάφορα παραδείγματα αναγνώρισης πληροφορίας (Σχήματα 4-7).

Καταλαβαίνει κανείς πως οι δυνατότητες που παρέχονται πλέον από τα συστήματα Νευρωνικών Δικτύων τείνουν να είναι απεριόριστες. Καλύπτουν και υποβοηθούν προβλήματα και ανάγκες υψηλής σημασίας για την κοινωνία (πχ. Ιατρικής), όσο και εφαρμογές που συνεισφέρουν στη βελτίωση της καθημερινότητας (πχ. Αυτόματη Ταξινόμηση Εγγράφων) ή διασκέδασης (πχ. Gaming).



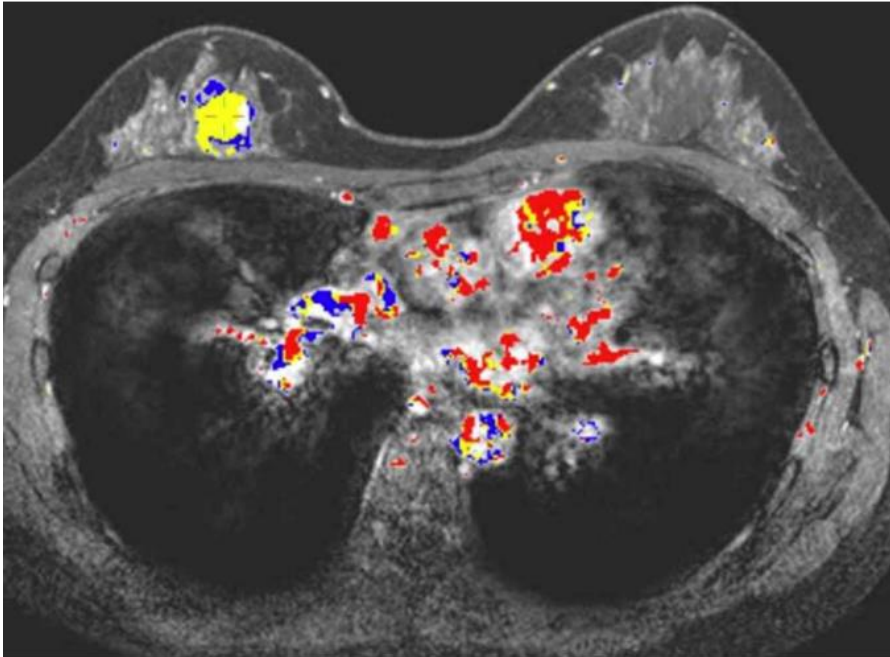
Σχήμα 4. Αυτόματη Αναγνώριση Αντικειμένων/Οντοτήτων<sup>2</sup>



Σχήμα 5. Αναγνώριση σκηνής από ένα σύστημα αυτόνομης οδήγησης<sup>3</sup>

<sup>2</sup> <https://towardsdatascience.com>

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων



Σχήμα 6. Αυτόματη αναγνώριση περιοχών ενδιαφέροντος σε μαγνητική τομογραφία<sup>4</sup>

<sup>3</sup> <https://heartbeat.fritz.ai>

<sup>4</sup> <https://towardsdatascience.com>



Labels: **Sombbrero**, **Woman**, **Clothing**

**No Priming:** A brown haired girl with a big straw hat.

**Priming:** **Woman** wearing a giant **sombbrero**-type sun hat.



Labels: **Red Panda**, **Tree**

**No Priming:** A brown rodent climbing up a **tree** in the woods.

**Priming:** A **red panda** is sitting in grass next to a **tree**.



Labels: **Gondola**, **Tree**, **Vehicle**

**No Priming:** A man and a woman being transported in a boat by a sailor through canals

**Priming:** Some people enjoying a nice ride on a **gondola** with a **tree** behind them.



Labels: **Woman**, **Man**, **Flower**, **Cake**

**No Priming:** A wedding **cake** with bouquet and lighted candles in the foreground.

**Priming:** A vase of **flowers** next to a wedding **cake** with a bride and groom on top.

#### Σχήμα 7. Αυτόματη περιγραφή και περίληψη Εικόνας<sup>5</sup>

Οι παραπάνω εφαρμογές Νευρωνικών Δικτύων περιλαμβάνουν μεγάλη ποικιλία συστημάτων είτε χαμηλής είτε υψηλής πολυπλοκότητας. Οι αρχιτεκτονικές πάνω στις οποίες βασίζονται τα συστήματα αυτά καθώς και το είδος τους, εξελίσσονται και βελτιώνονται διαρκώς και αποτελούν ένα ιδιαίτερα ενεργό πεδίο έρευνας.

Στα πλαίσια της εργασίας και στον παρακάτω πίνακα (Πίνακας 2) αυτής παρουσιάζονται συνοπτικά οι βασικές και δημοφιλέστερες αρχιτεκτονικές σε λογικό επίπεδο και αναφέρονται τα κύρια χαρακτηριστικά λειτουργίας και οι ιδιαιτερότητες τους όπως συναντώνται στη βιβλιογραφία και το διαδίκτυο<sup>6</sup>. Επιλέγεται συνειδητά οι ονομασίες να παραμείνουν στην αγγλική γλώσσα, καθώς διεθνώς χρησιμοποιούνται με τον τρόπο αυτό.

#### ΠΙΝΑΚΑΣ 2: Βασικές αρχιτεκτονικές Νευρωνικών Δικτύων και ο σκοπός τους

Αρχιτεκτονική/Αλγόριθμος	Σκοπός/Σύντομη Περιγραφή
--------------------------	--------------------------

<sup>5</sup> <https://medium.com>

<sup>6</sup> <https://www.smartsheet.com>

Autoencoder (AE)	Χρησιμοποιούνται συνήθως για επεξεργασία μοντέλων γενετικών δεδομένων
Bidirectional Recurrent Neural Network (BRNN)	Εφαρμόζονται για την αύξηση εισόδων πληροφοριών που είναι διαθέσιμες στο δίκτυο συνδέοντας δύο κρυμμένα, αντίθετα προς την κατεύθυνση στρώματα στην ίδια έξοδο. Χρησιμοποιώντας τα BRNN, το επίπεδο εξόδου μπορεί να λάβει πληροφορίες τόσο από προηγούμενες όσο και από μελλοντικές καταστάσεις.
Boltzmann Machine (BM)	Ένα επαναλαμβανόμενο νευρωνικό δίκτυο, που είναι ικανό να μάθει εσωτερικές αναπαραστάσεις και μπορεί να αναπαραστήσει και να λύσει προβλήματα υψηλής πολυπλοκότητας.
Convolutional Neural Network (CNN)	Συνήθως χρησιμοποιούνται για την ανάλυση οπτικών εικόνων, τα CNN είναι ένα νευρωνικά δίκτυα που έχουν σχεδιαστεί για να ελαχιστοποιούν την προεπεξεργασία.
Deconvolutional Neural Network (DNN)	Επιτρέπουν την επιβλεπόμενη κατασκευή ιεραρχικών αναπαραστάσεων εικόνων. Κάθε επίπεδο της ιεραρχίας ομαδοποιεί πληροφορίες από το προηγούμενο επίπεδο για να προσθέσει πιο περίπλοκα χαρακτηριστικά σε μια εικόνα.
Deep Belief Network (DBN)	Όταν εκπαιδεύεται με ένα μη εποπτευόμενο σύνολο παραδειγμάτων, ένα DBN μπορεί να μάθει να ανακατασκευάζει τις εισόδους του πιθανολογικά χρησιμοποιώντας επίπεδα ως ανιχνευτές χαρακτηριστικών. Μετά από αυτήν τη διαδικασία, μπορούν να εκπαιδευτούν για την εκτέλεση εποπτευόμενων ταξινομήσεων.

Deep Convolutional Inverse Graphics Network (DCIGN)	Στοχεύουν στην εκμάθηση μιας ερμηνεύσιμης αναπαράστασης εικόνων που διαχωρίζει το σύστημα σύμφωνα με τα στοιχεία της τρισδιάστατης δομής σκηνής, όπως παραλλαγές φωτισμού και περιστροφές βάθους. Ένα DCIGN χρησιμοποιεί πολλά επίπεδα όπως CNN και DNN.
Deep Residual Network (DRN)	Τα DRN εφαρμόζονται σε σύνθετα προβλήματα βαθιάς μάθησης. Έχοντας πολλά επίπεδα, ένα DRN αποτρέπει την υποβάθμιση των αποτελεσμάτων.
Denoising Autoencoder (DAE)	Χρησιμοποιούνται για την ανακατασκευή δεδομένων από κατεστραμμένες εισόδους δεδομένων. Ο αλγόριθμος αναγκάζει το κρυφό στρώμα να μάθει πιο σημαντικά χαρακτηριστικά. Ως αποτέλεσμα, η έξοδος αποδίδει μια πιο εκλεπτυσμένη έκδοση των δεδομένων εισαγωγής.
Echo State Network (ESN)	Ένα ESN λειτουργεί με ένα τυχαίο, μεγάλο, σταθερό επαναλαμβανόμενο νευρωνικό δίκτυο, όπου κάθε κόμβος λαμβάνει ένα μη γραμμικό σήμα απόκρισης. Ο αλγόριθμος ορίζει και εκχωρεί τυχαία βάρη και συνδεσιμότητα προκειμένου να επιτύχει μαθησιακή ευελιξία.
Extreme Learning Machine (ELM)	Αυτός ο αλγόριθμος μαθαίνει βάρη κρυφών κόμβων σε ένα βήμα, δημιουργώντας ένα γραμμικό μοντέλο. Τα ELM μπορούν να γενικευθούν καλά και να μάθουν πολλές φορές πιο γρήγορα από τα δίκτυα οπισθοδρομικής διάδοσης πληροφορίας.
Feed Forward Neural Network (FF or FFNN) and Perceptron (P)	Αυτοί είναι οι βασικοί αλγόριθμοι για τα νευρωνικά δίκτυα. Ένα νευρωνικό δίκτυο εμπρόσθιας διάδοσης είναι ένα τεχνητό νευρωνικό δίκτυο στο οποίο οι συνδέσεις κόμβων δεν σχηματίζουν κύκλο. Ο νευρώνας είναι μια δυαδική συνάρτηση με δύο μόνο αποτελέσματα (πάνω / κάτω, ναι / όχι, 0/1).

Gated Recurrent Unit (GRU)	Χρησιμοποιούν συνδέσεις μέσω ακολουθιών κόμβων για την εκτέλεση εργασιών μηχανικής εκμάθησης που σχετίζονται με ομαδοποίηση και μνήμη. Βελτιώνουν τις εξόδους μέσω του ελέγχου ροής πληροφοριών μοντέλου
Generative Adversarial Network (GAN)	Αυτό το σύστημα χωρίζει δύο νευρωνικά δίκτυα - Διάκρισης και Δημιουργίας - το ένα εναντίον του άλλου. Ο στόχος είναι να γίνει διάκριση μεταξύ πραγματικών και συνθετικών αποτελεσμάτων για την προσομοίωση εννοιολογικών εργασιών υψηλού επιπέδου.
Hopfield Network (HN)	Αυτή η μορφή επαναλαμβανόμενου τεχνητού νευρωνικού δικτύου είναι ένα συσχετισμένο σύστημα μνήμης με δυαδικούς κόμβους κατωφλίου. Σχεδιασμένο για σύγκλιση σε ένα ελάχιστο τοπικό, τα HN παρέχουν ένα μοντέλο για την κατανόηση της ανθρώπινης μνήμης.
Kohonen Network (KN)	Ένα KN οργανώνει έναν προβληματικό χώρο σε έναν διδιάστατο χάρτη. Η διαφορά μεταξύ των αυτο-οργανωμένων χαρτών (SOMs) και άλλων προσεγγίσεων επίλυσης προβλημάτων είναι ότι οι SOM χρησιμοποιούν ανταγωνιστική μάθηση αντί για μάθηση διόρθωσης σφαλμάτων.
Liquid State Machine (LSM)	Γνωστό ως μηχανική μάθηση τρίτης γενιάς, ένα LSM προσθέτει την έννοια του χρόνου ως στοιχείο. Τα LSM δημιουργούν ενεργοποίηση δικτύου χωροχρονικών νευρώνων καθώς διατηρούν τη μνήμη κατά την επεξεργασία. Η φυσική και η υπολογιστική νευροεπιστήμη χρησιμοποιούν LSM.



Long/Short-Term Memory (LSTM)	<p>Το LSTM είναι ικανό να μάθει ή να θυμάται την εξάρτηση από την τάξη σε προβλήματα πρόβλεψης σχετικά με την ακολουθία. Μια μονάδα LSTM κρατά ένα κελί, μια πύλη εισόδου, μια πύλη εξόδου και μια πύλη μνήμης. Τα κελιά διατηρούν τιμές σε αυθαίρετα χρονικά διαστήματα. Κάθε μονάδα ρυθμίζει ροές τιμών μέσω συνδέσεων LSTM. Αυτή η ικανότητα αλληλουχίας είναι απαραίτητη σε σύνθετους τομείς προβλημάτων, όπως η αναγνώριση ομιλίας και η αυτόματη μετάφραση.</p>
Markov Chain (MC)	<p>Είναι μια μαθηματική διαδικασία που περιγράφει μια ακολουθία πιθανών συμβάντων στα οποία η πιθανότητα κάθε συμβάντος εξαρτάται αποκλειστικά από την κατάσταση που επιτεύχθηκε στο προηγούμενο συμβάν. Τα παραδείγματα χρήσης περιλαμβάνουν προβλέψεις πληκτρολόγησης-λέξεων και Google PageRank.</p>
Neural Turing Machine (NTM)	<p>Με βάση το έργο του επιστήμονα δεδομένων Alan Turing στα μέσα του 20ου αιώνα, ένα NTM εκτελεί υπολογισμούς και επεκτείνει τις δυνατότητες των νευρωνικών δικτύων σε συνδυασμό με την εξωτερική μνήμη. Οι προγραμματιστές χρησιμοποιούν το NTM σε ρομπότ και το θεωρούν ως ένα από τα μέσα για την κατασκευή ενός τεχνητού ανθρώπινου εγκεφάλου.</p>
Radial Basis Function Networks (RBF nets)	<p>Οι προγραμματιστές χρησιμοποιούν δίκτυα RBF για να μοντελοποιήσουν δεδομένα που αντιπροσωπεύουν μια υποκείμενη τάση ή συνάρτηση. Τα δίκτυα RBF μαθαίνουν να προσεγγίζουν την υποκείμενη τάση χρησιμοποιώντας καμπύλες καμπάνας ή μη γραμμικούς ταξινομητές. Οι μη γραμμικοί ταξινομητές αναλύουν πιο βαθιά από ό, τι οι απλοί γραμμικοί ταξινομητές που λειτουργούν σε διανύσματα χαμηλότερης διάστασης. Χρησιμοποιείτε αυτά τα δίκτυα στον έλεγχο συστήματος και στις προβλέψεις χρονοσειρών.</p>

Recurrent Neural Network (RNN)	<p>Τα RNNs διαμορφώνουν διαδοχικές αλληλεπιδράσεις μέσω μνήμης. Σε κάθε βήμα, ένα RNN υπολογίζει μια νέα μνήμη ή μια κρυφή κατάσταση που εξαρτάται τόσο από την τρέχουσα είσοδο όσο και από την προηγούμενη κατάσταση μνήμης. Οι εφαρμογές περιλαμβάνουν σύνθεση μουσικής, έλεγχο ρομπότ και αναγνώριση ανθρώπινης δράσης.</p>
Restricted Boltzmann Machine (RBM)	<p>Ένα RBM είναι ένα πιθανοθεωρητικό γραφικό μοντέλο σε ένα μη εποπτευόμενο περιβάλλον. Ένα RBM αποτελείται από ορατά και κρυμμένα στρώματα, καθώς και τις συνδέσεις μεταξύ δυαδικών νευρώνων σε κάθε ένα από αυτά τα επίπεδα. Τα RBM είναι χρήσιμα για φιλτράρισμα, εκμάθηση χαρακτηριστικών και ταξινόμηση. Οι περιπτώσεις χρήσης περιλαμβάνουν ανίχνευση κινδύνων και επιχειρηματικές και οικονομικές αναλύσεις.</p>
Support Vector Machine (SVM)	<p>Με βάση εκπαιδευτικά σύνολα παραδειγμάτων που σχετίζονται με μία από τις δύο πιθανές κατηγορίες, ένας αλγόριθμος SVM δημιουργεί ένα μοντέλο που εκχωρεί νέα παραδείγματα σε μία από τις δύο κατηγορίες. Το μοντέλο στη συνέχεια αντιπροσωπεύει τα παραδείγματα ως χαρτογραφημένα σημεία στο διάστημα, διαιρώντας αυτά τα ξεχωριστά παραδείγματα κατηγοριών με το ευρύτερο δυνατό κενό. Στη συνέχεια, ο αλγόριθμος χαρτογραφεί νέα παραδείγματα στον ίδιο χώρο και προβλέπει σε ποια κατηγορία ανήκουν βασίζονται σε ποια πλευρά του κενού καταλαμβάνουν. Οι εφαρμογές περιλαμβάνουν ανίχνευση προσώπου και βιοπληροφορική.</p>

Variational Autoencoder (VAE)	<p>Το VAE είναι ένας συγκεκριμένος τύπος νευρωνικού δικτύου που βοηθά στη δημιουργία σύνθετων μοντέλων που βασίζονται σε σύνολα δεδομένων. Σε γενικές γραμμές, ένας αυτόματος κωδικοποιητής είναι ένα δίκτυο βαθιάς μάθησης που επιχειρεί να ανακατασκευάσει ένα μοντέλο ή να ταιριάξει τις εξόδους στόχου με τις παρεχόμενες εισόδους μέσω backpropagation. Ένα VAE αποδίδει επίσης υπερσύγχρονα αποτελέσματα μηχανικής εκμάθησης στους τομείς της δημιουργίας εικόνων και της μάθησης ενίσχυσης.</p>
-------------------------------	--

Όπως παρουσιάστηκε και στον Πίνακα 2, το εύρος των αρχιτεκτονικών νευρωνικών δικτύων είναι πολύ μεγάλο και πέρα από τους σκοπούς της εργασίας αυτής. Για την υποβοήθηση και την καλύτερη κατανόηση της λογικής τους από κάποιον αναγνώστη μη εξοικειωμένο με το μαθηματικό τους υπόβαθρο, δίνεται ένα συνοπτικό και άκρως απλουστευμένο σχήμα που περιλαμβάνει τις δημοφιλέστερες αρχιτεκτονικές (Σχήμα 8). Για περαιτέρω μελέτη και πληροφορίες, ο αναγνώστης θα πρέπει να ανατρέξει στη συσχετιζόμενη βιβλιογραφία.

Μία ενδιαφέρουσα στατιστική που πραγματοποιήθηκε πρόσφατα (Σχήμα 9)<sup>7</sup>, αφορά στην πρόβλεψη του προϋπολογισμού που επενδύεται στην Ευρώπη και αφορά την Τεχνητή Νοημοσύνη. Παρατηρείται ραγδαία αύξηση του προϋπολογισμού της τάξης του 75% για το 2023, σε σχέση με το 2021 και αφορά σε 21 δισεκατομμύρια δολάρια (US) έναντι 7 δισεκατομμυρίων δολαρίων. Ο αριθμός αυτός αποτελεί αύξηση 300% σε σχέση με το πρόσφατο παρελθόν (2019-7 δισεκατομμύρια). Τα παραπάνω καθιστούν τον τομέα της Τεχνητής Νοημοσύνης έναν από τους βασικούς οδηγούς στην εξέλιξη των ανθρώπινης δραστηριότητας.

Σύμφωνα με μελέτη του ITIF (<https://itif.org/>), οι ΗΠΑ συνεχίζουν να προηγούνται στον τομέα της Τεχνητής Νοημοσύνης έναντι της Ευρωπαϊκής Ένωσης και της Κίνας. Η τελευταία δείχνει να πλησιάζει την προπορευόμενη χώρα, ενώ η Ευρωπαϊκή Ένωση δείχνει να χάνει έδαφος. Σύμφωνα με τη μελέτη, η απουσία σημαντικών αλλαγών πολιτικής τόσο στην ΕΕ όσο και στις Ηνωμένες Πολιτείες - ιδίως στην ΕΕ που αλλάζει το ρυθμιστικό της σύστημα ώστε να είναι πιο φιλικό προς την καινοτομία και οι Ηνωμένες Πολιτείες αναπτύσσουν και χρηματοδοτούν μια πιο προληπτική εθνική στρατηγική για την τεχνητή νοημοσύνη - είναι πιθανό ότι η ΕΕ θα παραμείνει πίσω από τα δύο τις Ηνωμένες Πολιτείες και την Κίνα, και ότι η Κίνα θα κλείσει τελικά το χάσμα με τις Ηνωμένες Πολιτείες.

Το 2019, το Κέντρο Καινοτομίας Δεδομένων ανέλυσε τις δυνατότητες της Κίνας στο χώρο της Τεχνητής Νοημοσύνης, της Ευρωπαϊκής Ένωσης και των Ηνωμένων Πολιτειών χρησιμοποιώντας 30 μετρήσεις σε 6 κατηγορίες: ταλέντο, έρευνα, ανάπτυξη, υλικό, υιοθέτηση και δεδομένα. Διαπιστώσαμε ότι οι Ηνωμένες Πολιτείες ηγήθηκαν σε τέσσερις κατηγορίες (ταλέντο, έρευνα, ανάπτυξη και υλικό) και η Κίνα ηγήθηκε σε δύο (υιοθέτηση και δεδομένα). Από 100 συνολικά διαθέσιμους πόντους, οι Ηνωμένες Πολιτείες κατέκτησαν 44,2 πόντους, ακολουθούμενες από την Κίνα με 32,3 και την Ευρωπαϊκή Ένωση με 23,5. Αυτή η αναφορά μετρά την πρόοδο που έχει σημειώσει κάθε περιοχή Τεχνητής Νοημοσύνης χρησιμοποιώντας νέα δεδομένα για την ενημέρωση 15 των μετρήσεων και την προσθήκη 1 νέας μέτρησης. Διαπιστώνει ότι οι Ηνωμένες Πολιτείες εξακολουθούν να προηγούνται, με 44,6 πόντους,

<sup>7</sup> <https://www.statista.com>

ακολουθούμενες από την Κίνα με 32,0 και την Ευρωπαϊκή Ένωση με 23,3. Για να κατανοηθούν τα δυνατά σημεία της τεχνητής νοημοσύνης κάθε περιοχής σε σχέση με το μέγεθός τους, υπολογίζοντας επίσης βαθμολογίες για κάθε μέτρηση προσαρμόζοντας το μέγεθος των εργατικών τους δυνάμεων. Ελέγχοντας το μέγεθος, οι Ηνωμένες Πολιτείες (58,0 βαθμοί) οδηγούν την Ευρωπαϊκή Ένωση (24,2) και την Κίνα (17,8) - αν και η Κίνα έχει μειώσει την απόσταση μεταξύ της και των Ηνωμένων Πολιτειών από την τελευταία σχετική έκθεση.

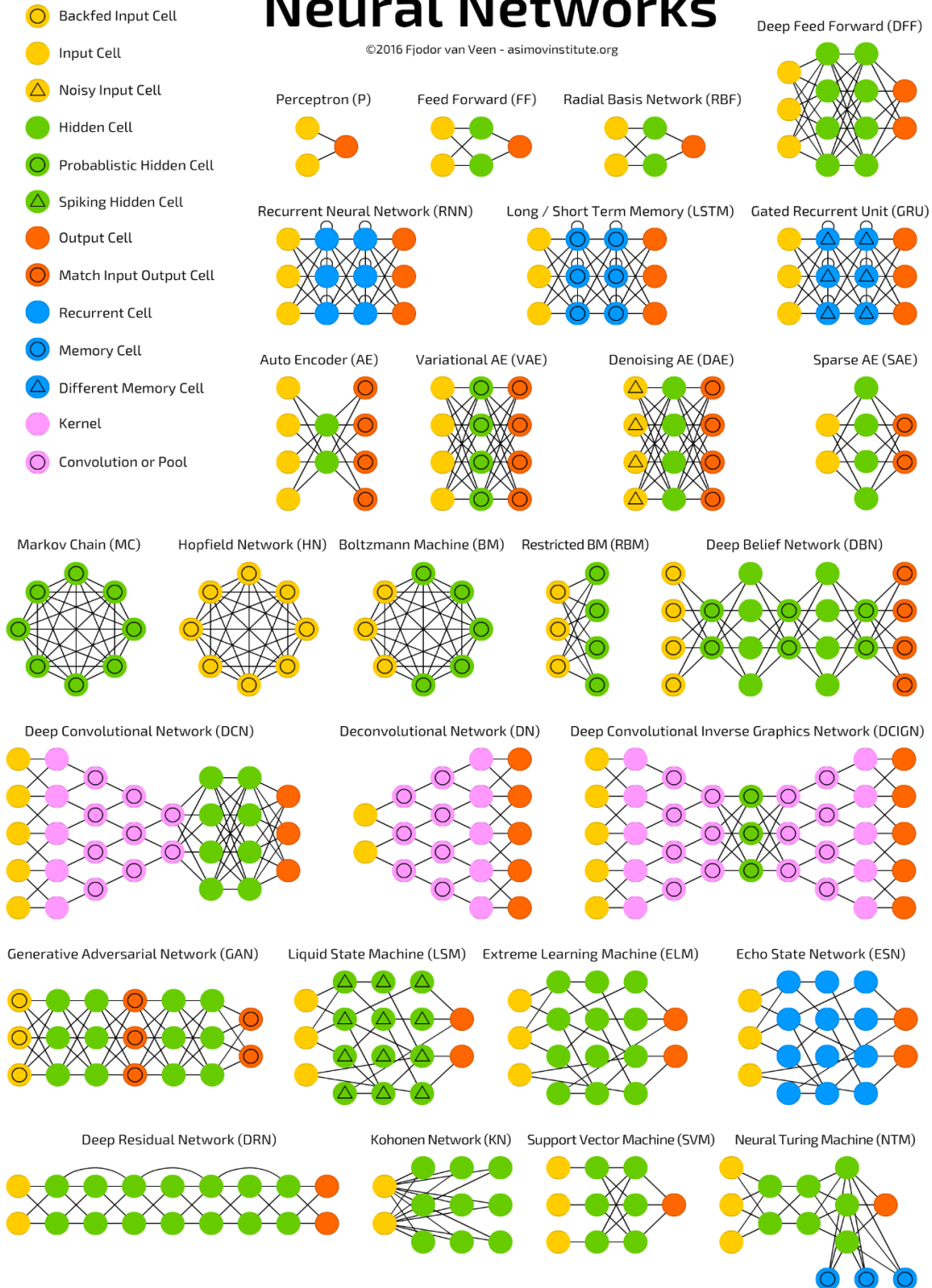
Βασικά, η Κίνα έχει σημειώσει σταδιακή πρόοδο - μειώνοντας το χάσμα ή επεκτείνοντας το προβάδισμά της στις Ηνωμένες Πολιτείες σε περισσότερες από τις μισές από τις ενημερωμένες μετρήσεις. Αντίθετα, η ΕΕ έχει σημειώσει πρόοδο σε σχέση με τις Ηνωμένες Πολιτείες σε λίγο περισσότερο από το ένα τέταρτο των ενημερωμένων μετρήσεων. Ως εκ τούτου, οι Ηνωμένες Πολιτείες έχουν διατηρήσει ή επεκτείνει το προβάδισμά τους έναντι της Ευρωπαϊκής Ένωσης σε σχεδόν 75% των ενημερωμένων μετρήσεων.

Παρά τη σταδιακή βελτίωση της Κίνας σε πολλούς δείκτες, οι Ηνωμένες Πολιτείες έχουν αυξήσει ελαφρώς τη συνολική τους πρωτοπορία στο σύστημα βαθμολογίας, επειδή έχουν αποδώσει εξαιρετικά καλά σε πολύ σταθμισμένους δείκτες, όπως το επιχειρηματικό κεφάλαιο και η χρηματοδότηση ιδιωτικών μετοχών. Για παράδειγμα, έχει έναν απaráμιλλο αριθμό νεοσύστατων επιχειρήσεων, οι οποίες έλαβαν 8 δισεκατομμύρια δολάρια περισσότερα σε κεφάλαια επιχειρηματικών συμμετοχών και χρηματοδότηση ιδιωτικών μετοχών από ό, τι η Κίνα το 2019. Οι Ηνωμένες Πολιτείες αποδίδουν επίσης καλά σε διάφορους δείκτες στους οποίους η Κίνα έχει μειώσει κάπως το κενό. Ένα παράδειγμα είναι οι δαπάνες έρευνας και ανάπτυξης (E&A) εταιρειών λογισμικού και υπηρεσιών πληροφορικής. Οι κινεζικές εταιρείες ξεπέρασαν ξεκάθαρα τις εταιρείες της ΕΕ στις δαπάνες E&A, αλλά οι εταιρείες λογισμικού και υπηρεσιών Η.Π.Α. δαπάνησαν ακόμη τρεις φορές περισσότερο στην E & A από ό, τι η Κίνα και η Ευρωπαϊκή Ένωση το 2019. Επιπλέον, η μέση ποιότητα έρευνας των ΗΠΑ είναι ακόμα υψηλότερη από εκείνη της Κίνας και η Ευρωπαϊκή Ένωση. Τέλος, παρά τις αυξανόμενες προσπάθειες της Κίνας να μειώσει την εξάρτησή της από τους ημιαγωγούς των ΗΠΑ, οι Ηνωμένες Πολιτείες εξακολουθούν να είναι ο παγκόσμιος ηγέτης στο σχεδιασμό τσιπ για συστήματα AI.

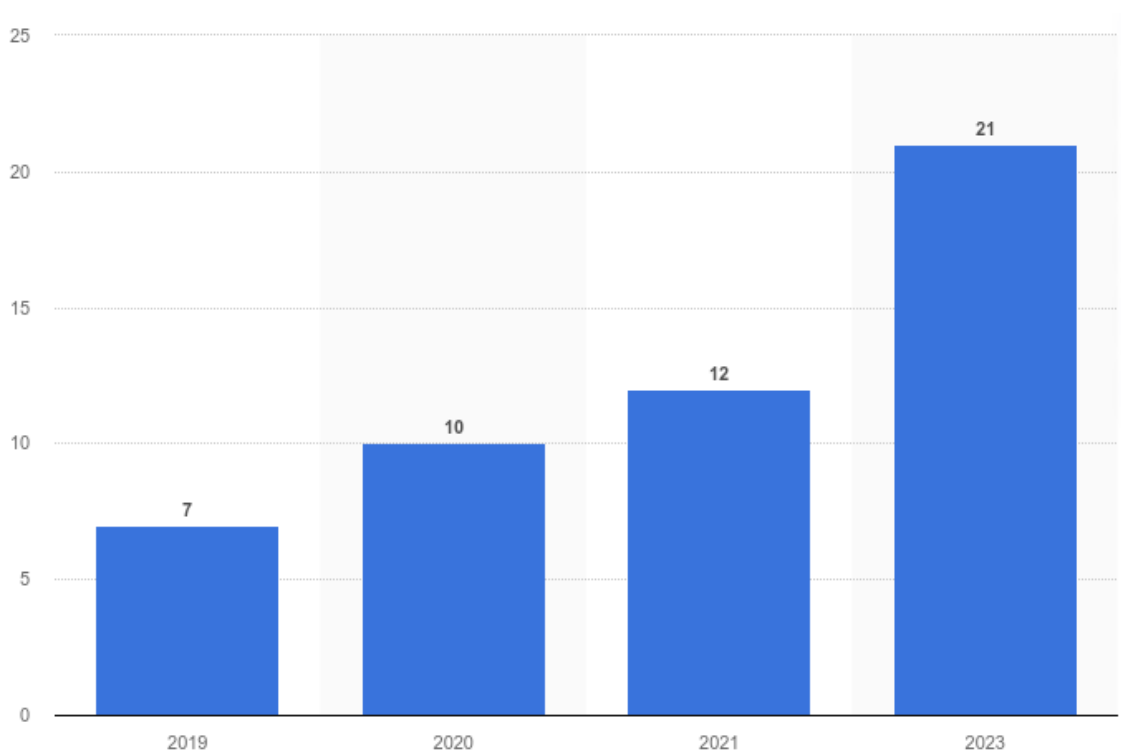
Η πρόοδος της Ευρωπαϊκής Ένωσης έναντι των Ηνωμένων Πολιτειών είναι ανάμεικτη. Για παράδειγμα, οι εταιρείες τεχνητής νοημοσύνης των ΗΠΑ συνεχίζουν να λαμβάνουν σημαντικά περισσότερες επενδύσεις από ό, τι οι ευρωπαϊκές. Ωστόσο, η χρηματοδότηση επιχειρηματικών κεφαλαίων και ιδιωτικών κεφαλαίων της ΕΕ ως ποσοστό της χρηματοδότησης των ΗΠΑ αυξήθηκε από 13% σε 22% μεταξύ 2016 και 2019. Επιπλέον, ο σταθμισμένος αντίκτυπος αναφοράς της ΕΕ (FWCI) για έγγραφα AI, ένα σχετικό μέτρο της ποιότητας, αυξήθηκε το 2018 ενώ το FWCI των Ηνωμένων Πολιτειών μειώθηκε. Ωστόσο, η Ευρωπαϊκή Ένωση έχει μείνει πιο πίσω από τις Ηνωμένες Πολιτείες όσον αφορά τον αριθμό των συμφωνιών χρηματοδότησης, τις εξαγορές εταιρειών AI και τις εταιρείες AI που έχουν συγκεντρώσει χρηματοδότηση τουλάχιστον 1 εκατομμυρίου δολαρίων από την τελευταία έκθεση. Επιπλέον, οι εταιρείες λογισμικού και υπηρεσιών πληροφορικής της ΕΕ απέτυχαν να καλύψουν το χάσμα μεταξύ τους και των αμερικανικών εταιρειών στις δαπάνες E&A. Η αποχώρηση του Ηνωμένου Βασιλείου από την Ευρωπαϊκή Ένωση θα μειώσει επίσης τις δυνατότητες AI της ΕΕ, τόσο σε απόλυτους όρους όσο και σε κατά κεφαλήν βάση.

# A mostly complete chart of Neural Networks

©2016 Fjodor van Veen - asimovinstitute.org



Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

Σχήμα 8. Βασικές Αρχιτεκτονικές Νευρωνικών Δικτύων<sup>8</sup>

Σχήμα 9. Προϋπολογισμός που αφορά την επενδύσεις σε θέματα Τεχνητής Νοημοσύνης στην Ευρώπη

Ένα άλλο σημείο που πρέπει να τονιστεί στη συνέχεια είναι το ακόλουθο: στην επιστημονική και επιχειρηματική κοινότητα, αναφέρονται συχνά δύο όροι: Μηχανική Μάθηση (Machine Learning) και Βαθιά Μάθηση (Deep Learning). Οι όροι αυτοί συχνά συγχέονται και θεωρούνται το ίδιο. Κάτι τέτοιο δεν είναι ακριβώς αληθές. Στην ουσία ο όρος Deep Learning αποτελεί υποσύνολο του Machine Learning. Με τη σειρά του, ο τελευταίος αποτελεί υποσύνολο της Τεχνητής Νοημοσύνης (Artificial Intelligence-AI). Πιο συγκεκριμένα<sup>9</sup>:

- **Machine Learning** είναι ένα υποσύνολο της Τεχνητής Νοημοσύνης που περιλαμβάνει τεχνικές και αλγόριθμους που επιτρέπουν στις μηχανές/υπολογιστές να βελτιωθούν στην αντιμετώπιση και επίλυση προβλημάτων, μέσω της απόκτησης εμπειρίας.
- **Deep Learning** είναι ένα υποσύνολο του Machine Learning που βασίζεται στα Νευρωνικά Δίκτυα και επιτρέπει στις μηχανές/υπολογιστές να εκπαιδευτούν με στόχο την επίλυση προβλημάτων

Η σύγκριση των παραπάνω δύο όρων μπορεί να γίνει και πιο λεπτομερής, όπως φαίνεται και στον επόμενο πίνακα (Πίνακας 3). Ως συμπέρασμα προκύπτει ότι οι τεχνικές Deep Learning αποτελούν ουσιαστικά επέκταση και εξέλιξη του Machine Learning, ώστε να είναι εφικτή η διαχείριση και εξαγωγή πληροφορίας μέσα από μεγάλο όγκο δεδομένων (Big Data). Παράλληλα, στόχος του Deep Learning είναι να μειωθεί/περιοριστεί ο παράγοντας της ανθρώπινης καθοδήγησης (πχ. κριτήρια αποφάσεων, διορθωτικές κινήσεις, κλπ) και το σύστημα να είναι σε θέση να “μάθει” από τα δεδομένα μόνο του, πραγματοποιώντας δικές του συγκρίσεις και θεωρώντας τα δικά του χαρακτηριστικά.

<sup>8</sup> <https://towardsdatascience.com>

<sup>9</sup> <https://docs.microsoft.com>

**ΠΙΝΑΚΑΣ 3: Σύγκριση Machine Learning και Deep learning**

Χαρακτηριστικό	Machine Learning & Deep Learning	Μόνο Deep Learning
Αριθμός Δεδομένων	Μπορούν να χρησιμοποιήσουν μικρό αριθμό δεδομένων.	Χρειάζονται τεράστιο αριθμό δεδομένων εκπαίδευσης για να είναι σε θέση να κάνουν προβλέψεις.
Ανάγκες Υλισμικού (hardware)	Μπορούν να δουλέψουν με συμβατικούς υπολογιστές και δεν απαιτούν ιδιαίτερη υπολογιστική ισχύ.	Απαιτούν υπολογιστές υψηλών προδιαγραφών. Πραγματοποιούν πολύ μεγάλο αριθμό πράξεων πολλαπλασιασμού πινάκων (χρονοβόρες), οι οποίες επιταχύνονται με χρήση καρτών γραφικών (GPU).
Διαδικασία Δημιουργίας Χαρακτηριστικών	Απαιτούν από τους χρήστες να αναγνωρίσουν και δημιουργήσουν με ακρίβεια διάφορα χαρακτηριστικά (πχ. κατώφλια απόφασης, δημιουργία σετ δεδομένων εκπαίδευσης)	Μαθαίνουν βασικά χαρακτηριστικά δεδομένων από τα ίδια τα δεδομένα και δημιουργούν μόνα τους νέα χαρακτηριστικά που αφορούν σε λεπτομέρειες.

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

Μεθοδολογία Εκμάθησης	Η μεθοδολογία εκμάθησης περιλαμβάνει τη διαίρεση της διαδικασίας σε μικρότερα βήματα. Έπειτα, συνδυάζονται τα βήματα αυτά σε μία έξοδο.	Η μεθοδολογία εκμάθησης για την επίλυση του προβλήματος βασίζεται σε διαδικασία από άκρη-σε-άκρη.
Χρόνος Εκτέλεσης	Μικρή διάρκεια εκπαίδευσης (από λίγα δευτερόλεπτα έως και λίγες ώρες)	Συνήθως απαιτούν αρκετό χρόνο εκπαίδευσης (πολλές ώρες ή/και μέρες) καθώς περιλαμβάνουν πολλά επίπεδα.
Έξοδος	Η έξοδος είναι συνήθως μία αριθμητική τιμή (πχ. πιθανότητα ή το νούμερο που αντιστοιχεί σε μία κατηγορία αναγνώρισης)	Η έξοδος μπορεί να έχει πολλές μορφές, όπως κείμενο, πιθανότητα, ήχος, εικόνα) Machine Learning & Deep Learning.

Η παραπάνω ανάλυση στοχεύει στη βασική κατανόηση της αξίας των σύγχρονων τεχνικών Νευρωνικών Δικτύων και Βαθιάς Μάθησης στην ανθρώπινη καθημερινότητα. Είναι εμφανές ότι ο τομέας της Τεχνητής Νοημοσύνης ηγείται ανάμεσα σε άλλους (πχ. Βιομηχανία Υγείας) των εξελίξεων και αποτελεί χώρο διαρκούς μελέτης και επιτευγμάτων.

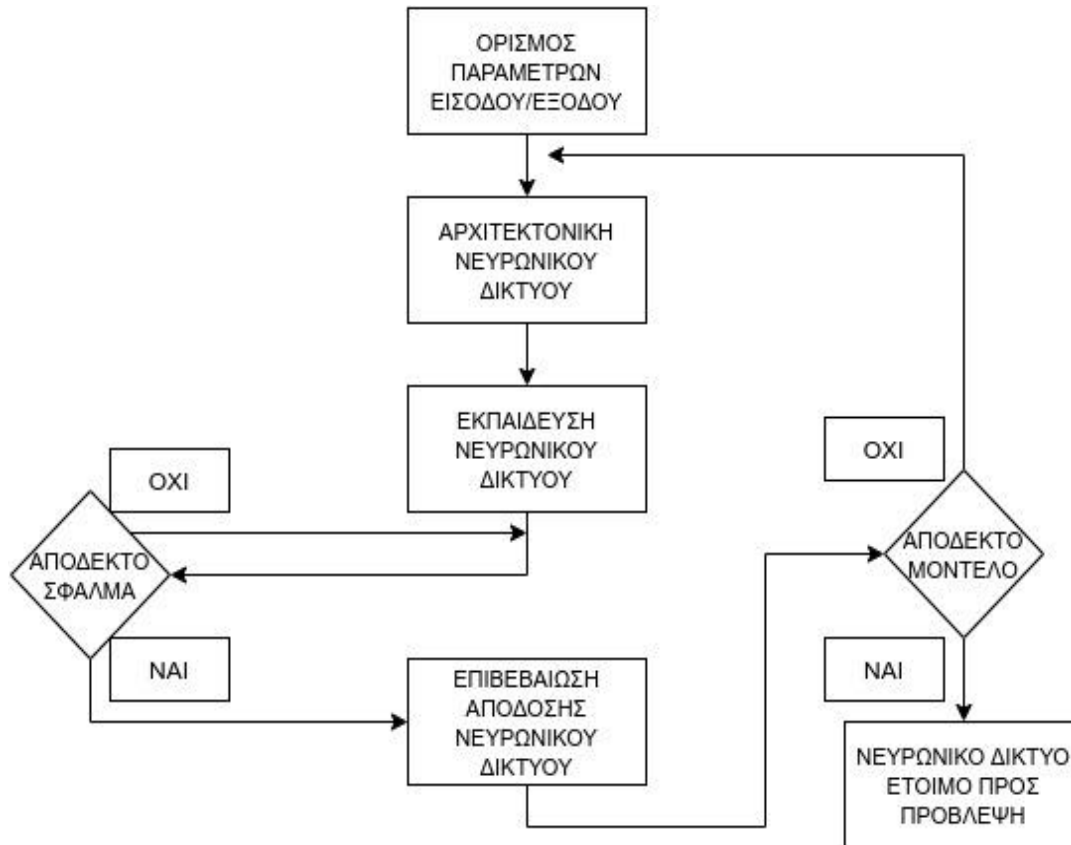
Στη συνέχεια του Κεφαλαίου αυτού πραγματοποιείται βιβλιογραφική ανασκόπηση που εστιάζει στο πρόβλημα της αυτόματης **Ταξινόμησης Εγγράφων**. Στο Κεφάλαιο 1 δόθηκε το πλαίσιο που κινείται η ανάλυση αυτή. Μέχρι την προηγούμενη δεκαετία και όπως παρουσιάζεται συνοπτικά στο [2], οι αλγόριθμοι που αναπτύχθηκαν με σκοπό την αντιμετώπιση του προβλήματος, βασίστηκαν κυρίως σε τεχνικές Επεξεργασίας Εικόνας (Image Processing) και σε αρκετές περιπτώσεις οι βασικές παράμετροι απόφασης των αλγορίθμων αυτών προέκυπταν σε αρκετές περιπτώσεις είτε μέσω πειραματικής αναζήτησης είτε μέσω της διαισθητικής ερμηνείας του προβλήματος. Σε άλλες περιπτώσεις έγινε χρήση τεχνικών Αναγνώρισης Προτύπων (Pattern Recognition) και αλγορίθμων ομαδοποίησης (clustering algorithms).

Η παρούσα μελέτη εστιάζει στην τελευταία δεκαετία, εξαιτίας της δημιουργίας βέλτιστων μέχρι τώρα συστημάτων που χρησιμοποιούν αλγόριθμους Τεχνητής Νοημοσύνης (AI). Ένας από τους λόγους επικράτησής τους στις αναπτυσσόμενες μεθόδους από την επιστημονική κοινότητα είναι, πέρα από τα ακριβέστερα αποτελέσματα, η δυνατότητα υποστήριξης εκπαίδευσης Νευρωνικών Δικτύων μέσω του υλικού (hardware). Η γνώση και η θεωρία γύρω από τα Νευρωνικά Δίκτυα πρωτοεμφανίστηκε το 1943 [3] αλλά η άνοδος του hardware και

### Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων



συγκεκριμένα η εκπαίδευση πάνω σε κάρτες γραφικών υπολογιστή (Graphical Processing Unit - GPU) επέτρεψαν την ταχεία εκπαίδευση και έρευνα πάνω στα Δίκτυα αυτά και συνεπώς την απότομη εξέλιξη του επιστημονικού αυτού κλάδου.



Σχήμα 10. Βασικό διάγραμμα εκπαίδευσης Νευρωνικών Δικτύων [4]

Με πολύ απλά λόγια, η βασική διαφορά των Τεχνητών Νευρωνικών Δικτύων (Artificial Neural Networks - ANN) συγκριτικά με τις παραδοσιακές τεχνικές, είναι ότι οι κατάλληλες τιμές για τις παραμέτρους/βάρη (ή αλλιώς τιμές των φίλτρων που εφαρμόζονται πάνω σε μία εικόνα) απόφασης “μαθαίνονται” από το ίδιο το σύστημα, μέσω της φάσης εκπαίδευσης (training phase), καθιστώντας τα έτσι ικανά για γενικότερη χρήση και εφαρμογή. Το πλήθος των παραμέτρων μπορεί να φτάνει και τα εκατομμύρια, συνεπώς δίνεται η δυνατότητα επίλυσης πολύπλοκων συστημάτων. Ένα γενικό διάγραμμα διαδικασίας εκπαίδευσης δίνεται στο Σχήμα 1.

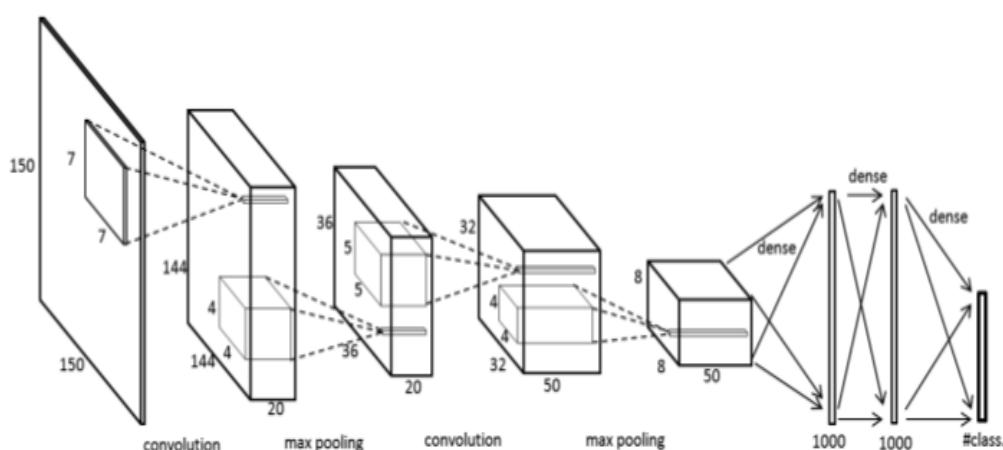
Κατά τη διάρκειά της και κρατώντας αναλογία με το πρόβλημα της **αυτόματης Ταξινόμησης Εγγράφων**, ο χρήστης παρέχει ως είσοδο στο σύστημα δεδομένα (εικόνες εγγράφων) για τα οποία γνωρίζει την κατηγορία/κλάση/ετικέτα στην οποία ανήκουν (class ή label) και την παρέχει και αυτή ως είσοδο στο σύστημα. Το Νευρωνικό Δίκτυο ξεκινά από τυχαίες τιμές των παραμέτρων του και προσπαθεί να προβλέψει την κατηγορία του εγγράφου, την οποία δίνει ως έξοδο. Στη συνέχεια, μετρά το σφάλμα της πρόβλεψης σε σχέση με την πραγματική τιμή της κατηγορίας, μέσω κάποιας μετρικής αποτίμησης (συνάρτηση κόστους - loss function) και αυτή είναι υψηλή (δηλ. υψηλό σφάλμα), μεταβάλλει τις τιμές των παραμέτρων, κάνοντας χρήση ενός αλγόριθμου εκμάθησης, ώστε την επόμενη φορά το σφάλμα να είναι χαμηλότερο. Η διαδικασία

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

επαναλαμβάνεται μέχρις ότου το σφάλμα να είναι αρκετά χαμηλό και ταυτόχρονα η ακρίβεια του συστήματος σε ένα άλλο παρόμοιο σύνολο εικόνων να μεγιστοποιείται (φάση επαλήθευσης - validation phase).

## 2.2. ΣΥΣΤΗΜΑΤΑ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΓΙΑ ΤΑΞΙΝΟΜΗΣΗ ΕΓΓΡΑΦΩΝ

Στην υπο-ενότητα αυτή παρουσιάζονται τα δημοφιλέστερα ANN συστήματα για Ταξινόμηση Εγγράφων. Πιο συγκεκριμένα, τα Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Network - CNN) είναι μία υποκατηγορία τέτοιων συστημάτων, απλούστερης μορφής και ταυτόχρονα σειριακής λογικής, που αντιμετωπίζουν αποτελεσματικά το πρόβλημα της ταξινόμησης (ή αλλιώς κατηγοριοποίησης).



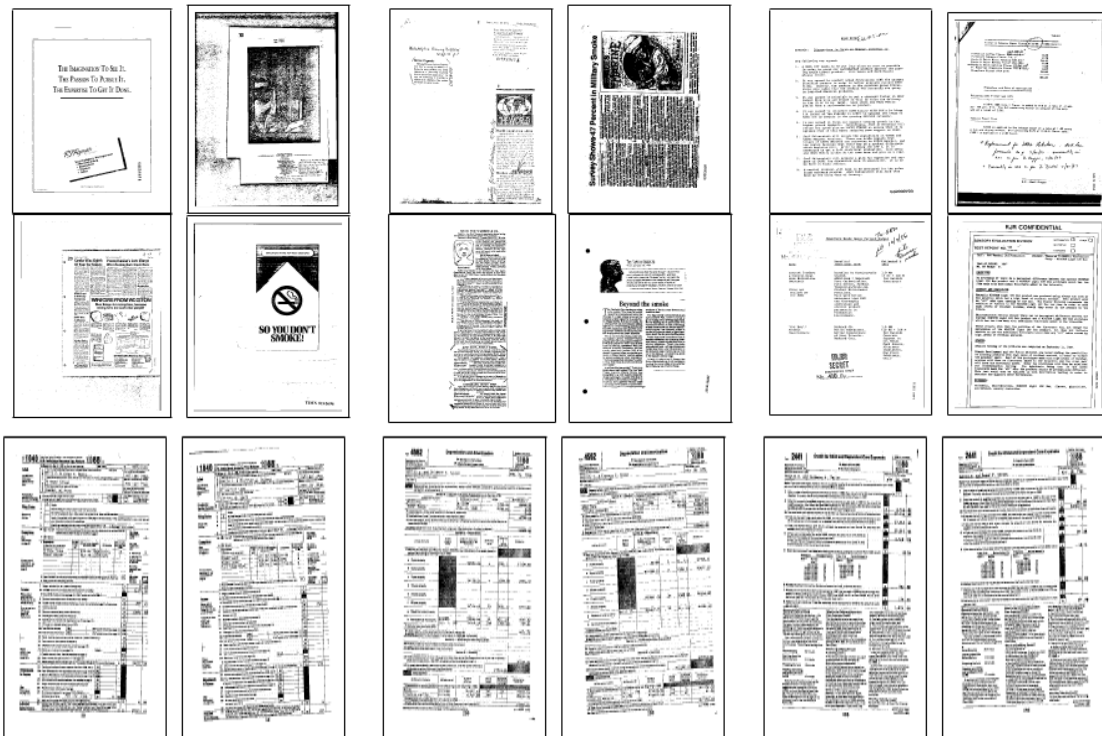
Σχήμα 11. Σύστημα Τεχνητού Νευρωνικού Δικτύου [5]

Στο [5] παρουσιάστηκε για πρώτη φορά η εφαρμογή ενός CNN για ταξινόμηση εγγράφων. Το σύστημα φαίνεται στο Σχήμα 11 και ο σκοπός του ήταν να ταξινομήσει εικόνες εγγράφων μέσα από διάφορες κλάσεις όπως άρθρα από περιοδικά, εικόνες διαφημίσεων, αναφορές, λογιστικές φόρμες κλπ, συνολικού αριθμού δέκα κλάσεων. Ενδεικτικές εικόνες φαίνονται στο Σχήμα 12. Βασικά σημεία του συστήματος είναι ότι η είσοδος του είναι μία εικόνα εγγράφου που έχει υποστεί αλλαγή στο μέγεθος (resizing) της (διαστάσεων Ύψος x Πλάτος: 150 x 150) έτσι ώστε το πλήθος των παραμέτρων (εξαρτάται από τις διαστάσεις) να είναι επαρκώς μικρό ώστε, ο χρόνος εκπαίδευσης να είναι σύντομος, το πλήθος των παραμέτρων να δύναται να αποθηκευτεί στη μνήμη της κάρτας γραφικών του υπολογιστή (GPU) και η ακρίβεια να είναι υψηλή.

Στο [6], αποδείχθηκε ότι ένα CNN μπορεί να “μάθει” τις κατάλληλες τιμές των παραμέτρων ακόμα και από τμήματα (κομμάτια) των εγγράφων. Προτάθηκε ο διαχωρισμός της εικόνας σε τέσσερις περιοχές (άνω, κάτω, αριστερό, δεξί σώμα, βλ. Σχήμα 13) και απέδειξε ότι ακόμα και αν χρησιμοποιηθεί ένα κομμάτι από αυτά, το αποτέλεσμα είναι υψηλό. Η φυσική ερμηνεία της παραπάνω υπόθεσης έγκειται στο γεγονός ότι πχ. το άνω μέρος (κεφαλίδα) ενός γράμματος διαφέρει σημαντικά από αυτό ενός λογιστικό εγγράφου, συνεπώς ένα CNN μπορεί να μάθει να τα διαχωρίζει μέσα από αυτό. Η εφαρμογή έγινε σε έγγραφα 16 κλάσεων συνολικά (οι βάσεις δεδομένων αναλύονται στο Κεφάλαιο 4). Στην πιο αποτελεσματική περίπτωση, έγινε χρήση πέντε Δικτύων (τα τέσσερα άνω και ένα που λαμβάνει ως είσοδο ολόκληρη την εικόνα) και ο

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

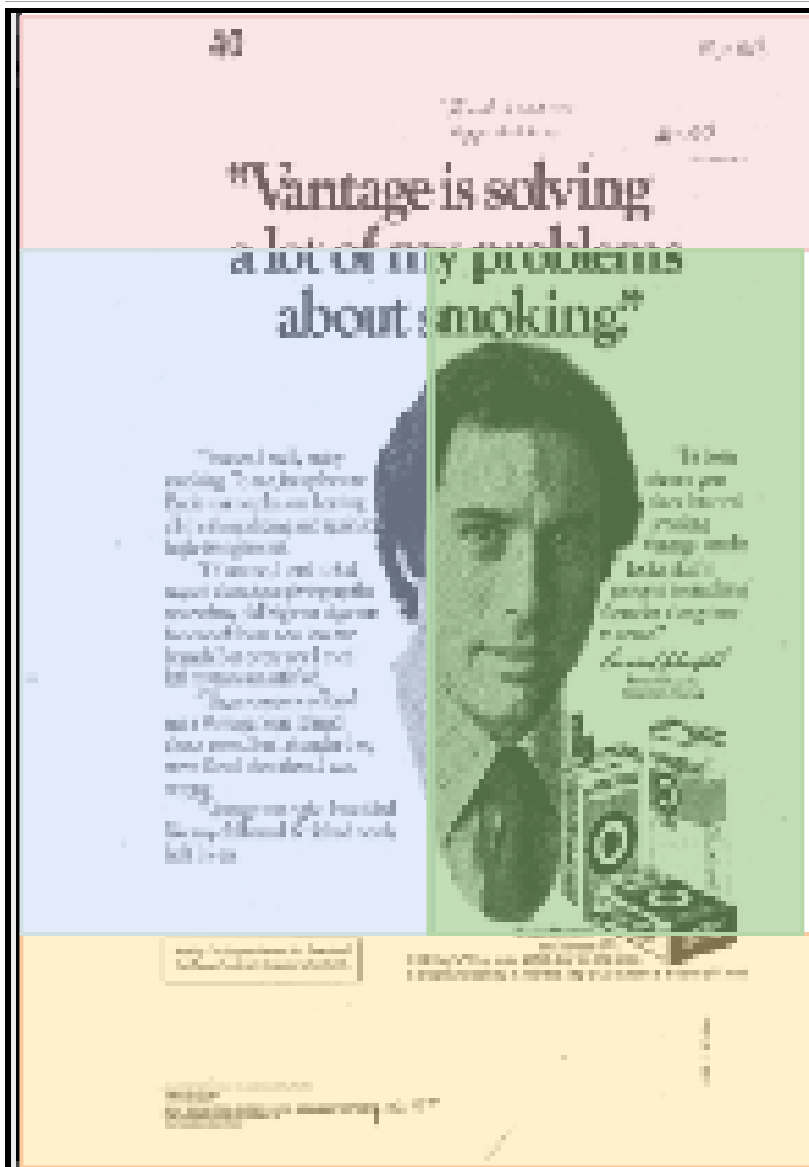
συνδυασμός τους έδωσε υψηλή αποτελεσματικότητα (~90%). Ταυτόχρονα, αποδείχθηκε ότι η επιλογή αρχικών τιμών είναι πολύ σημαντική και προκύπτει σημαντική βελτίωση αν χρησιμοποιηθούν αρχικές τιμές παραμέτρων από ένα άλλο ίδιο Δίκτυο, που έχει εκπαιδευτεί όμως σε άλλη βάση δεδομένων, πιο γενικής χρήσης. Η τεχνική αυτή ονομάζεται Transfer Learning (Μεταφορά Γνώσης) και η σημασία της παρουσιάζεται στο Κεφάλαιο 3. Επιπλέον, στην ίδια δημοσίευση εφαρμόστηκε μείωση των παραμέτρων της εξόδου του CNN μέσω της τεχνικής Ανάλυσης Πρωτευόντων Συστατικών (PCA - Principal Component Analysis). Τέλος, το ίδιο σύστημα χρησιμοποιήθηκε και για ένα παραπλήσιο πρόβλημα με αυτό της Ταξινόμησης Εγγράφων, αυτό της Ανάκτησης Εγγράφων (Document Retrieval), όπου για μία εικόνα εισόδου στόχος είναι να εντοπιστούν οι κορυφαίες κ-κοιτινότερες εικόνες που ανήκουν στην ίδια κατηγορία, μετρώντας την ευκλείδεια απόσταση των εξόδων του Νευρωνικού Δικτύου για το κάθε ζεύγος εικόνων.



Σχήμα 12. Δείγματα εγγράφων προς ταξινόμηση [5]

Η αποτελεσματικότητα των παραπάνω τεχνικών αποτιμάται σε διάφορα προβλήματα (όπως αυτά της Ταξινόμησης, Ομαδοποίησης και Ανάκτησης Εγγράφων). Επίσης, χρησιμοποιείται και εδώ η τεχνική του Transfer Learning. Το συμπέρασμα είναι πως η εξαγωγή χαρακτηριστικών που βασίζεται σε Νευρωνικά Δίκτυα με αρχικοποίηση παραμέτρων από Δίκτυο που έχει εκπαιδευτεί σε βάση δεδομένων παρόμοιου είδους εικόνων (δηλ. έγγραφα) είναι υψηλότερη σε σχέση με την εξαγωγή “ρηχών” χαρακτηριστικών. Αντίθετα, στην περίπτωση που η αρχικοποίηση γίνει από τιμές παραμέτρων που έχουν προκύψει από βάση δεδομένων διαφορετικής φύσης, τότε σε ορισμένες περιπτώσεις τα “ρηχά” χαρακτηριστικά δίνουν καλύτερα αποτελέσματα. Σημειώνεται επίσης πως στην μελέτη αυτή συγκρίνονται και διαφορετικές αρχιτεκτονικές Νευρωνικών Δικτύων.

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

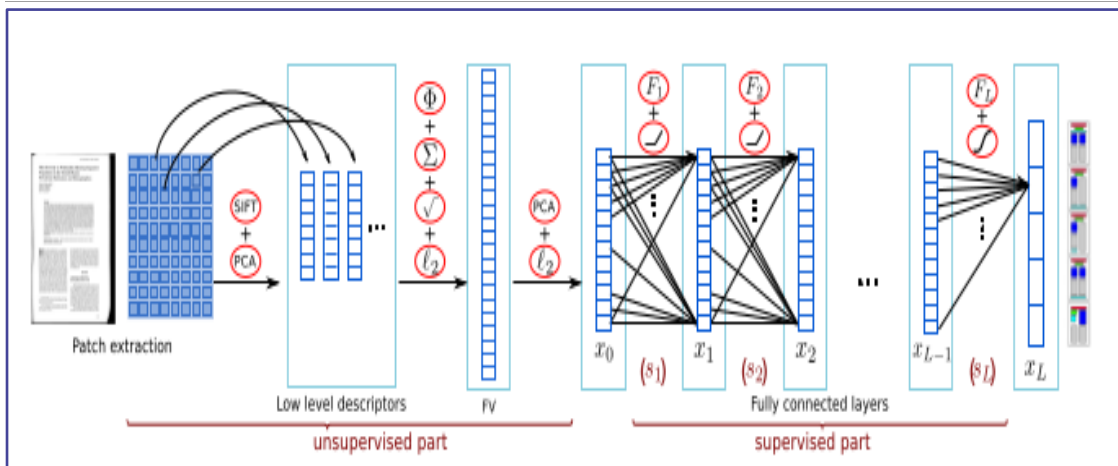


Σχήμα 13. Διαχωρισμός εικόνας σε άνω, κάτω, αριστερό και δεξιό σώμα

Στη δημοσίευση [7], πραγματοποιείται μία πειραματική ανάλυση των τρόπων που μπορούν να εξαχθούν οπτικά χαρακτηριστικά από εικόνες εγγράφων. Διακρίνονται τρεις τύποι εξαγωγής χαρακτηριστικών:

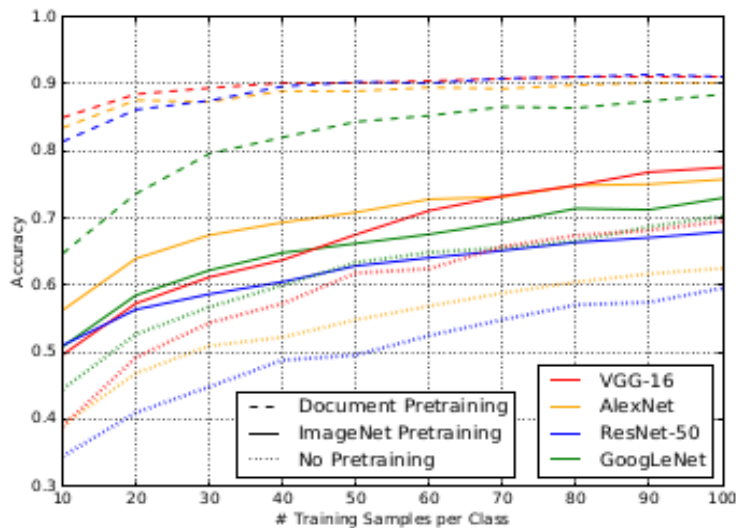
- “Παραδοσιακές” τεχνικές εξαγωγής “ρηχών” χαρακτηριστικών (shallow features) που δε βασίζονται σε αλγόριθμους Νευρωνικών Δικτύων.
- Τεχνικές που βασίζονται σε αλγόριθμους Νευρωνικών Δικτύων για εξαγωγή “βαθών” χαρακτηριστικών (deep features).
- Υβριδικές τεχνικές που συνδυάζουν τους δύο παραπάνω τρόπους (Σχήμα 14).

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων



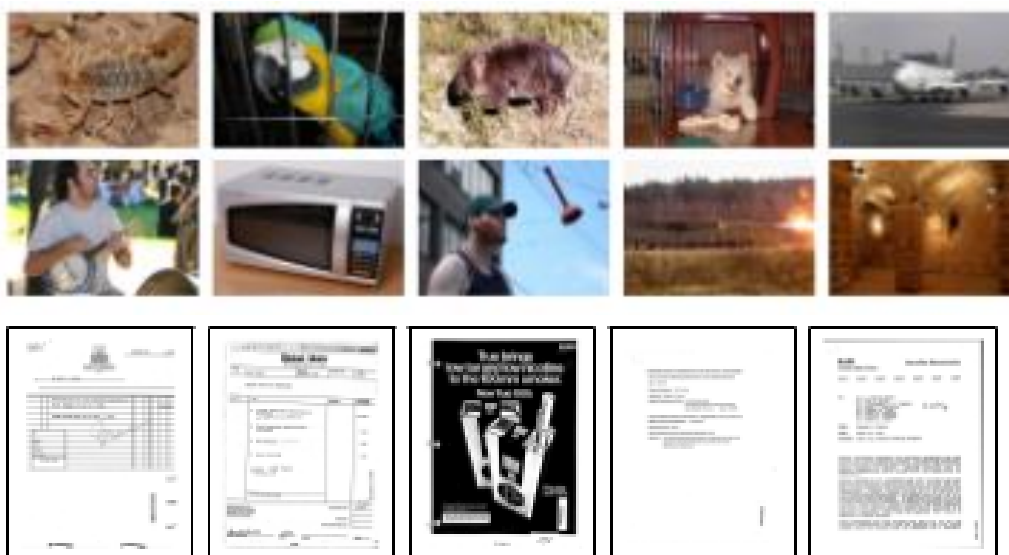
Σχήμα 14. Υβριδική αρχιτεκτονική εξαγωγής “ρηχών” και “βαθών” χαρακτηριστικών [7]

Στη συνέχεια [8] και ταυτόχρονα με την εξέλιξη των αρχιτεκτονικών Νευρωνικών Δικτύων (πιο αποτελεσματικές λύσεις, μείωση παραμέτρων κλπ) παρατηρείται ότι διάφορες τέτοιες αρχιτεκτονικές είναι αποτελεσματικές στην αντιμετώπιση του προβλήματος της Ταξινόμησης Εγγράφων, καθιστώντας τα Δίκτυα αυτά κατάλληλα και τη βέλτιστη λύση για το πρόβλημα. Παρουσιάζεται μία συγκριτική μελέτη (Σχήμα 15) απόδοσης και τονίζεται και εδώ η σημασία του Transfer Learning. Ειδικότερα, οι αρχικές τιμές των παραμέτρων του Νευρωνικού Δικτύου είναι επιλεγμένες από κάποιο άλλο Δίκτυο εκπαιδευμένο σε κάποια άλλη βάση εικόνων. Η άλλη βάση (Σχήμα 16) μπορεί να είναι βάση “φυσικών” εικόνων καθημερινότητας [9] ή ακόμα καλύτερα, κάποια βάση δεδομένων με εικόνες εγγράφων, όπως αυτή που παρουσιάστηκε στο [6].



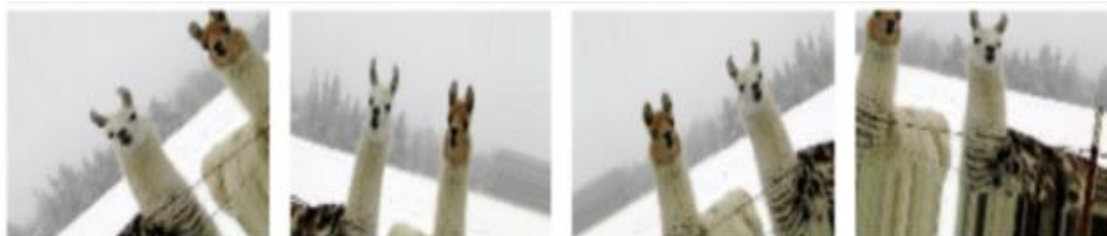
Σχήμα 15. Συγκριτική μελέτη για διάφορες αρχιτεκτονικές Νευρωνικών Δικτύων και διάφορους τρόπους αρχικοποίησης παραμέτρων [8]

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων



**Σχήμα 16.** Δείγματα από τις βάσεις δεδομένων φυσικών εικόνων (άνω) [9] και εικόνων εγγράφων (κάτω) [6]

Στο παραπάνω διάγραμμα (Σχήμα 15) δίνεται ξεκάθαρα η επιρροή του τρόπου αρχικοποίησης ή αλλιώς προ-εκπαίδευσης (Pretraining) ενός Νευρωνικού Δικτύου. Η χειρότερη απόδοση παρατηρείται αν δε χρησιμοποιηθεί κάποια συγκεκριμένη αρχικοποίηση και δοθούν τυχαίες αρχικές τιμές (No Pretraining). Η αρχικοποίηση με τιμές παραμέτρων Δικτύων που έχουν εκπαιδευτεί σε μεγάλες βάσεις δεδομένων ([9]) δίνει σημαντική βελτίωση (ImageNet Pretraining). Στην καλύτερη περίπτωση, αν χρησιμοποιηθεί Δίκτυο εκπαιδευμένο σε βάση δεδομένων παρόμοιας φύσης (δηλ. εγγράφων), τότε η απόδοση βελτιστοποιείται (Document Pretraining). Ταυτόχρονα, οι παραπάνω τρόποι αρχικοποίησης εφαρμόζονται σε διάφορες δημοφιλείς αρχιτεκτονικές Νευρωνικών Δικτύων (VGG-16, AlexNet, ResNet-50, GoogleNet)



**Σχήμα 17.** Παράδειγμα μετασχηματισμού εικόνας για επαύξηση δεδομένων

Στο [10], δίνεται έμφαση στον τρόπο προ-επεξεργασίας των εικόνων εγγράφων, προτού αυτές δοθούν ως είσοδοι σε ένα Νευρωνικό Δίκτυο. Αυτό επιτυγχάνεται με την τεχνική της Επαύξησης Δεδομένων (data augmentation, βλ. Κεφάλαιο 3), όπου πραγματοποιείται ο εμπλουτισμός της βάσης εικόνων με “παραμορφωμένα” αντίγραφα. Παράδειγμα τέτοιου είδους παραμόρφωσης, ή αλλιώς μετασχηματισμού εικόνας, δίνεται στο Σχήμα 17.

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

Οι μετασχηματισμοί αυτοί για την επαύξηση των δεδομένων είναι μεγάλου πλήθους και η επιρροή τους στην αποτελεσματικότητα ενός Νευρωνικού Δικτύου στην Ταξινόμηση Εγγράφων μελετάται στο [10]. Ενδεικτικά αναφέρονται ορισμένοι δημοφιλείς μετασχηματισμοί εικόνας:

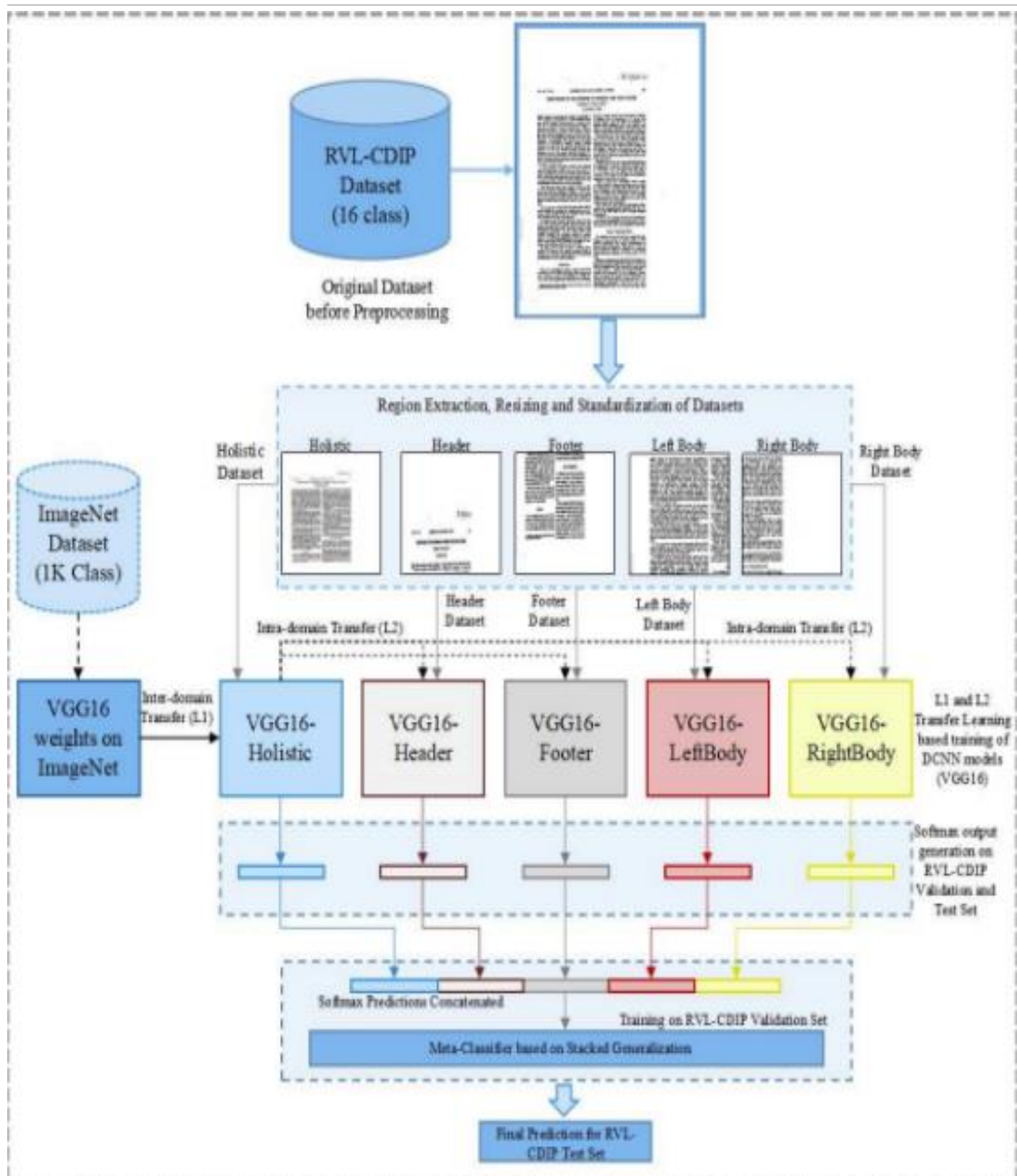
- Περιστροφή (rotation)
- Μετατόπιση (translation)
- Κλιμάκωση (scaling)
- Αντικατοπτρισμός (mirroring)
- Εισαγωγή θορύβου μορφής αλατοπίπερου (salt & pepper noise)
- Εισαγωγή θορύβου Gaussian κατανομής (gaussian noise)
- Στρέβλωση (shear)
- Παραμόρφωση χρώματος (color jitter)
- Θόλωση (blur)
- Περικοπή (crop)

Πλέον των παραπάνω μετασχηματισμό, στην ίδια δημοσίευση ([9]) μελετάται και η σημασία του τρόπου αναπαράστασης της εικόνας (έγχρωμη RGB, κλίμακας του γκρι-grayscale, ασπρόμαυρη-binary) καθώς και στη σημασία των διαστάσεων της εικόνας (αλλαγή μεγέθους, μεταβολή λόγου διαστάσεων πλευρών-aspect ratio).

Μετά την ανάλυση των πιο σειριακών/γραμμικών αρχιτεκτονικών Νευρωνικών Δικτύων (κατά κύριο λόγο των Συνελιγτικών Νευρωνικών Δικτύων), στη βιβλιογραφία παρατηρείται η πλέον η τάση χρήσης πιο σύνθετων αρχιτεκτονικών. Στο [11], γίνεται χρήση μίας πιο σύνθετης αρχιτεκτονικής (Σχήμα 18), συνδυάζοντας ταυτόχρονα επιμέρους παρόμοιες τεχνικές με τις προαναφερόμενες δημοσιεύσεις. Χρησιμοποιείται και εδώ πληροφορία από τμήματα της εικόνας για εκπαίδευση επιμέρους Δικτύων αλλά και από το σύνολό της, πραγματοποιείται αρχικοποίηση παραμέτρων από άλλη βάση δεδομένων και τέλος, οι προκύπτουσες πληροφορίες (πίνακες χαρακτηριστικών - feature vectors) συνενώνονται και τροφοδοτούν έναν τελικό ταξινομητή. Εξετάζονται διάφορες παραλλαγές του ταξινομητή αυτού και πραγματοποιείται η αντίστοιχη αποτίμησή της ακρίβειας στην κάθε περίπτωση.

Τα τελευταία χρόνια, πραγματοποιούνται προσπάθειες βελτίωσης του προβλήματος Ταξινόμησης Εγγράφων, χρησιμοποιώντας τεχνικές υβριδικής μορφής [12], [13], [14]. Στη μελέτη που παρουσιάστηκε πρόσφατα [12], συνδυάζονται τόσο τα οπτικά χαρακτηριστικά της εικόνας που εξάγονται μέσω εκπαίδευσης ενός Νευρωνικού Δικτύου, όσο και η πληροφορία που προκύπτει από το αποτέλεσμα Οπτικής Αναγνώρισης Χαρακτήρων (OCR). Η διαδικασία ανάλυσης της κειμενικής πληροφορίας που είναι αποτέλεσμα του OCR, έχει ως εξής: α) Εξάγονται χαρακτηριστικά (πχ. συχνά χρησιμοποιούμενες λέξεις) μέσα από τη διαδικασία Επιλογής Χαρακτηριστικών (Feature Selection) και αυτά αποτελούν τα κειμενικά χαρακτηριστικά της εικόνας (textual/text features) που αναπαριστώνται μέσω πινάκων χαρακτηριστικών. Για την επίτευξη αυτού, γίνεται χρήση λεξικού. Στη συνέχεια, β), εκπαιδεύεται ένα σύστημα (συνήθως μικρό Νευρωνικό Δίκτυο) όπου η έξοδος του “συνενώνεται” με αυτήν του Νευρωνικού Δικτύου υπεύθυνο για την εξαγωγή οπτικών χαρακτηριστικών και γ) μέσω ενός ταξινομητή αποφασίζεται η κατηγορία στην οποία ανήκει το έγγραφο. Η παραπάνω αρχιτεκτονική περιγράφεται στο Σχήμα 19.

### Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων



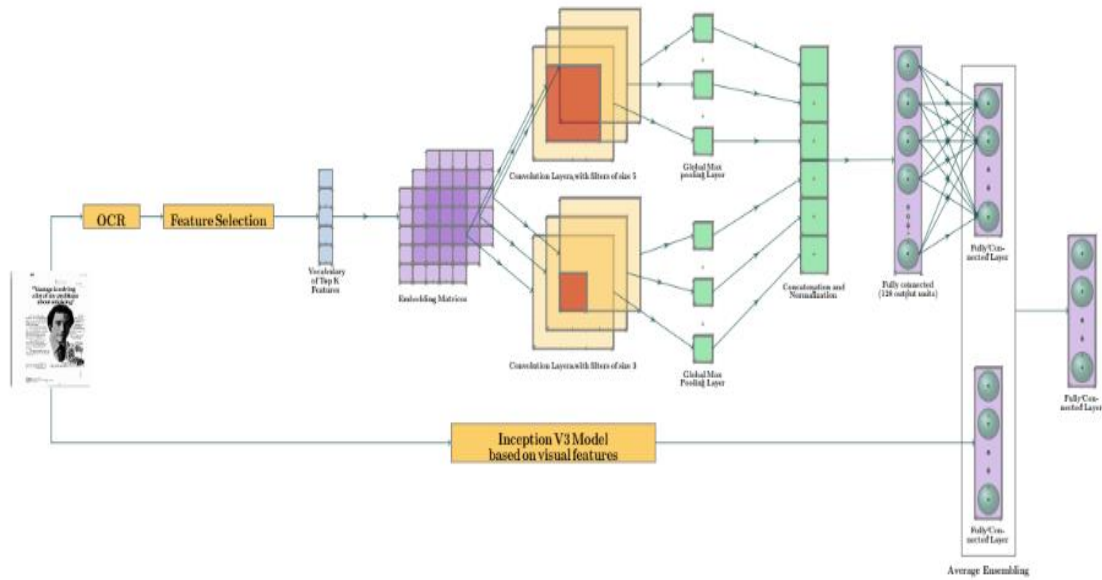
Σχήμα 18. Αρχιτεκτονική Νευρωνικού Δικτύου συστήματος ταξινόμησης εγγράφων [11]

Παρόμοια μεθοδολογία με το [12], παρουσιάστηκε και στη μελέτη που δίνεται στο [13]. Χρησιμοποιείται λογισμικό ανοιχτού κώδικα<sup>10</sup> για την Οπτική Αναγνώριση Χαρακτήρων. Έπειτα, παρουσιάζονται τεχνικές εξαγωγής των σημαντικών χαρακτηριστικών από την προκύπτουσα κειμενική πληροφορία, καθώς αυτή περιέχει θόρυβο (δηλ. λάθη αναγνώρισης). Η οπτική αναγνώριση της ίδιας λέξης μπορεί να συναντάται ως πολλά διαφορετικά αποτελέσματα, παρόμοια μεταξύ τους, που μπορεί να διαφέρουν σε έναν ή δύο χαρακτήρες, λόγω σφάλματος. Συνεπώς, οι δύο αυτές λέξεις θα πρέπει να “ομαδοποιηθούν” καθώς πρόκειται για την ίδια λέξη.

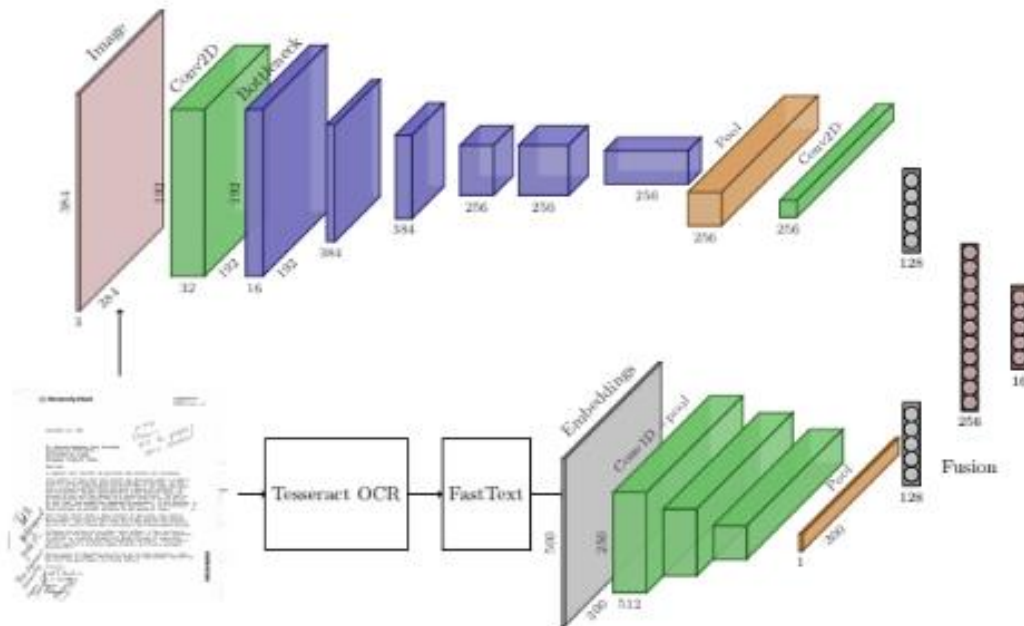
<sup>10</sup> <https://github.com/tesseract-ocr/>



Επιπλέον, τα συστήματα αναγνώρισης συνήθως δίνουν λάθος αποτελέσματα σε “θορυβώδεις” περιοχές, όπως για παράδειγμα πλεονάζοντα σημεία στίξης που στην πραγματικότητα είναι περιοχές με χαμηλή ποιότητα σάρωσης. Η εκπαίδευση του συστήματος ακολουθεί και εδώ την υβριδική αρχιτεκτονική όπου δύο παράλληλα συστήματα “συνενώνονται” και τροφοδοτούν έναν τελικό ταξινομητή κατηγοριοποίησης. Η προτεινόμενη αρχιτεκτονική δίνεται στο Σχήμα 20.

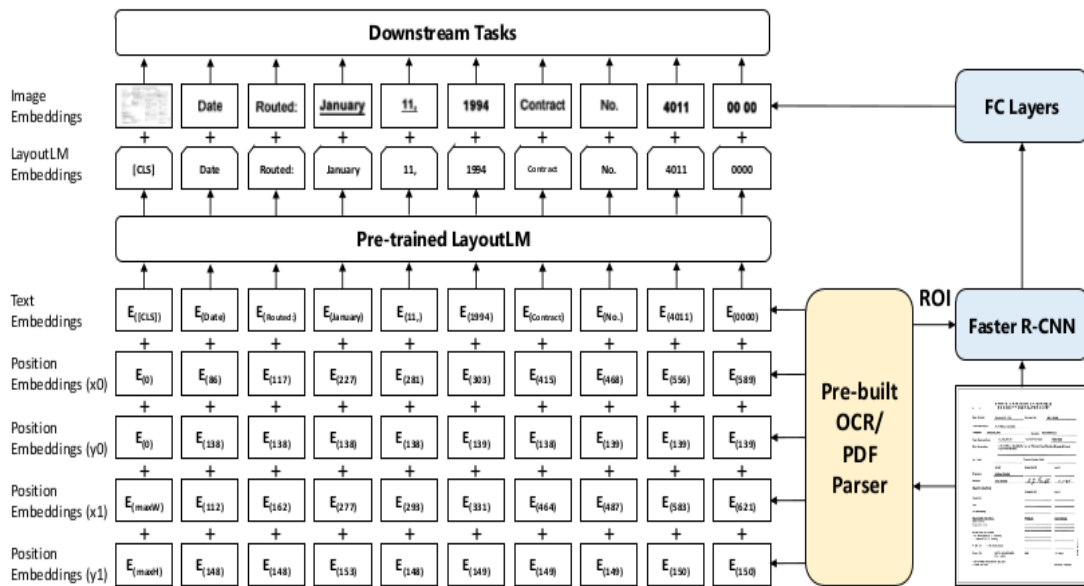


Σχήμα 19. Αναπαράσταση υβριδικής αρχιτεκτονικής συστήματος ταξινόμησης εγγράφων [12]



Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

**Σχήμα 20. Αρχιτεκτονικής συστήματος ταξινόμησης εγγράφων που συνδυάζει οπτική και κειμενική πληροφορία [13]**



**Σχήμα 21. Αρχιτεκτονική συστήματος ταξινόμησης εγγράφων που συνδυάζει οπτική, κειμενική και δομική πληροφορία [14]**

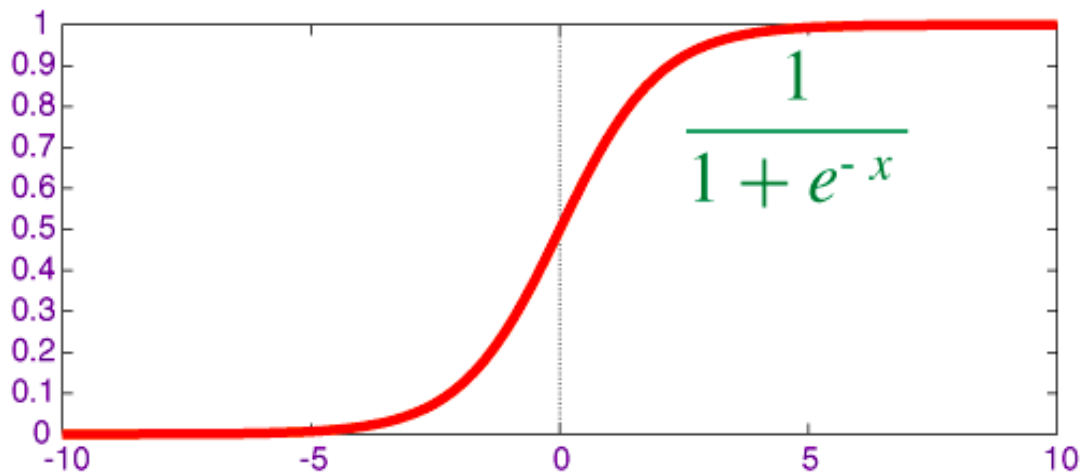
Σε μια άλλη πρόσφατη δημοσίευση [14], παρουσιάζεται μια νέα αρχιτεκτονική υβριδικής μορφής. Σύμφωνα με τους συγγραφείς, γίνεται για πρώτη φορά συνδυασμός πληροφορίας που προκύπτει μέσω της εικόνας (οπτικά χαρακτηριστικά - image/visual features), του κειμένου της (κειμενικά χαρακτηριστικά - textual/text features), καθώς και της δομής της εικόνας (δομικά χαρακτηριστικά - layout features). Η βασική δομή της προτεινόμενης αρχιτεκτονικής συστήματος που συνδυάζει τα παραπάνω δίνεται στο Σχήμα 21. Τα αποτελέσματα του παραπάνω συνδυασμού πληροφοριών έδωσαν βελτιωμένα αποτελέσματα (εκτενέστερη ανάλυση στο Κεφάλαιο 4) σε βάσεις δεδομένων εγγράφων (94.42%). Ταυτόχρονα, η δημοσίευση αυτή χρησιμοποιεί μία τεχνική ενσωμάτωσης κειμενικής πληροφορίας σε Νευρωνικά Δίκτυα, γνωστή και ως μοντέλο BERT [15].

### 3. ΕΠΙΣΚΟΠΗΣΗ ΒΑΣΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ ΚΑΙ ΤΕΧΝΙΚΩΝ ΕΚΠΑΙΔΕΥΣΗΣ ΣΥΣΤΗΜΑΤΩΝ ΨΗΦΙΑΚΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ

Στο Κεφάλαιο αυτό γίνεται μία επισκόπηση των βασικών παραμέτρων και τεχνικών εκπαίδευσης Νευρωνικών Δικτύων που αποτελούν συστήματα ή μέρη αυτών για την Ψηφιακή Ταξινόμηση Εγγράφων. Υπάρχει ένα σύνολο θεμάτων που πρέπει να ληφθούν υπόψη, ώστε να επιλεγθεί ο κατάλληλος συνδυασμός των ζητούμενων παραμέτρων ή αλλιώς υπερ-παραμέτρων (hyperparameters) για την επίτευξη μεγαλύτερης ακρίβειας.

Αρχικά [16], θα πρέπει να επιλεγθεί ο κατάλληλος αριθμός επιπέδων νευρώνων (layers) και ενδιάμεσων ή κρυφών επιπέδων (hidden layers) έτσι ώστε να καθοριστεί το πλήθος ή αλλιώς “βάθος” του Δικτύου. Η κατάλληλη επιλογή σχετίζεται τόσο με τον τρόπο που ορίζεται η ροή πληροφορίας, όσο και με την πολυπλοκότητα του προβλήματος. Η λανθασμένη επιλογή μπορεί να οδηγήσει σε φαινόμενα υπο-προσαρμογής (underfitting) ή υπερ-προσαρμογής (overfitting). Παράδειγμα της πρώτης περίπτωσης είναι αποτυχία μοντελοποίησης του προβλήματος λόγω μικρού αριθμού νευρώνων, ενώ στη δεύτερη περίπτωση η υπερβολική πολυπλοκότητα του Δικτύου (μεγάλος αριθμός νευρώνων) οδηγεί συχνά σε υπερβολική μοντελοποίηση των δεδομένων ή/και στο θόρυβο που υπάρχει πιθανόν σε αυτά (απομνημόνευση).

Στη συνέχεια, σημαντικό ρόλο παίζει και το είδος του αλγόριθμου εκπαίδευσης που θα επιλεγθεί. Δημοφιλέστερος είναι ο αλγόριθμος ανάστροφης μετάδοσης λάθους (back propagation algorithm), όπου η μεταβολή των βαρών γίνεται βάσει του υπολογισμού της συνεισφοράς κάθε βάρους στο συνολικό σφάλμα.



Σχήμα 22. Σιγμοειδής συνάρτηση ενεργοποίησης νευρώνα

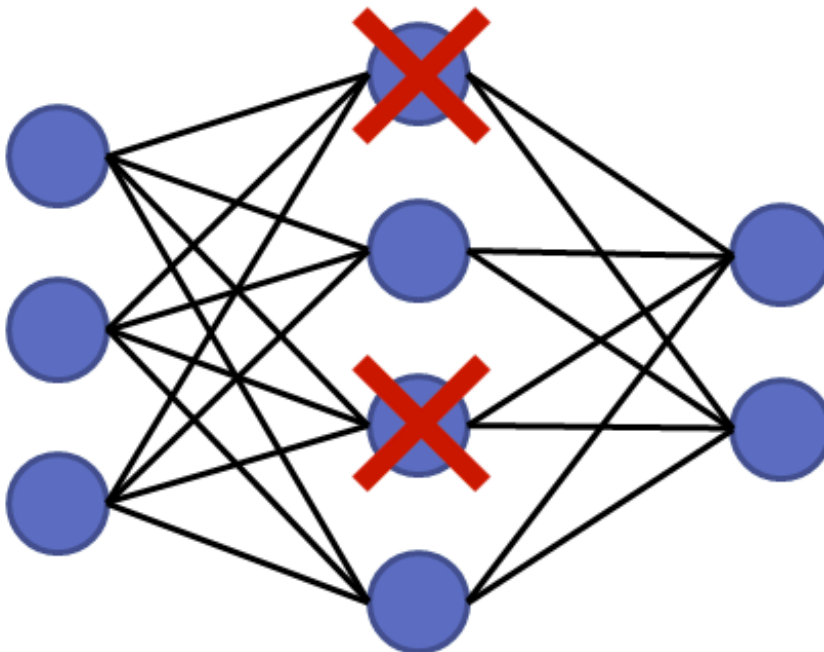
Τέλος, σημαντικό ρόλο στην εκπαίδευση των Νευρωνικών δικτύων έχουν οι παρακάτω υπερ-παραμέτροι:

- Ρυθμός μάθησης (learning rate): Ορίζει το “πόσο γρήγορα μαθαίνει ένα δίκτυο”, δηλαδή καθορίζει το ρυθμό μεταβολής των βαρών (weights)

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

- Κύκλοι εκπαίδευσης ή αλλιώς εποχές (epochs): Ορίζουν το πόσο πολύ θα εκπαιδευτεί ένα δίκτυο, δηλαδή το πόσες φορές (κύκλοι) θα χρησιμοποιηθούν όλα τα δεδομένα στη φάση εκπαίδευσης.
- Μέγεθος δέσμης δεδομένων (batch size): Ορίζει το πόσες εικόνες θα τροφοδοτούν το δίκτυο σε κάθε επανάληψη. Το σφάλμα στην εκάστοτε επανάληψη υπολογίζεται ως ο μέσος όρος των επιμέρους σφαλμάτων της δέσμης.
- Μέγεθος της εικόνας: Επιλογή κατάλληλου μεγέθους ώστε και να μη χάνεται πληροφορία λόγω της σμίκρυνσης και παράλληλα να μπορεί να υποστηριχθεί από τη μνήμη ενός υπολογιστή (ή της κάρτας γραφικών-GPU που χρησιμοποιείται στην πράξη)
- Συναρτήσεις ενεργοποίησης (activation functions) νευρώνων: Κάθε έξοδος ενός νευρώνα, μετασχηματίζεται μέσω μιας συνάρτησης ώστε οι τιμές να είναι σε κατάλληλη κλίμακα και να είναι εφικτή η εκπαίδευση και η μεταβολή βαρών μέσω του αλγόριθμου εκμάθησης. Παράδειγμα συνάρτησης σιγμοειδούς ενεργοποίησης δίνεται στο προηγούμενο σχήμα (Σχήμα 22).

Σημαντική τεχνική για την αποτελεσματικότερη εκπαίδευση νευρωνικών δικτύων είναι η τεχνική Dropout [17]. Σύμφωνα με την τεχνική αυτή, σε κάθε επανάληψη εκπαίδευσης, απενεργοποιούνται με τυχαίο τρόπο (Σχήμα 23) ένα ποσοστό νευρώνων του δικτύου και δε συνεισφέρουν στο τελικό σφάλμα της επανάληψης αυτής. Αποτέλεσμα αυτού είναι να μειώνεται η εξάρτηση απόφασης κάθε νευρώνα από τους υπόλοιπους. Έχει αποδειχθεί ότι η χρήση της συνεισφέρει σημαντικά στην αποφυγή του φαινομένου του over-fitting.



Σχήμα 23. “Απενεργοποίηση” Νευρώνων βάσει της τεχνικής Dropout

Μία ιδιαίτερα διαδεδομένη τεχνική εκπαίδευσης νευρωνικών δικτύων είναι αυτή της Επαύξησης Δεδομένων (Data Augmentation). Μία εκτενής μελέτη που αναλύει την τεχνική αυτή βρίσκεται στο [18]. Ο στόχος της επαύξησης δεδομένων είναι η αύξηση των δεδομένων μέσω παραγωγής τυχαίων και “παραμορφωμένων” αντιγράφων της αρχικής εικόνας. Με τον τρόπο αυτό εμπλουτίζεται η εκάστοτε βάση δεδομένων με εικόνες, κάτι το οποίο ενισχύει σημαντικά την εκπαίδευση δικτύων που εκπαιδεύονται σε βάσεις δεδομένων μικρού πλήθους εικόνων. Επιπλέον, η τεχνική αυτή επιτρέπει στη δημιουργία μοναδικών εικόνων αμέτρητου πλήθους με αποτέλεσμα το νευρωνικό δίκτυο να μη “βλέπει” ποτέ την ίδια εικόνα αλλά παρόμοια αντίγραφα τους και έτσι να είναι σε θέση να αποφασίζει ορθά παρά το θόρυβο που εισάγεται μέσα από τις τεχνητές παραμορφώσεις.

Στο επόμενο σχήμα δίνονται διάφορες περιπτώσεις επαύξησης δεδομένων μέσω τυχαίας παραμόρφωσης της εικόνας (Σχήμα 24). Σύμφωνα με αυτό, στόχος ενός νευρωνικού δικτύου είναι βρίσκεται σε θέση να δίνει τη σωστή απόφαση τόσο για την αρχική εικόνα (πχ. ένας καφέ σκύλος σε μία συγκεκριμένη στάση σώματος) αλλά και για οποιοδήποτε διαφορετική εικόνα που μπορεί να αναπαριστά κάτι παρόμοιο με την αρχική (πχ. ο ίδιος σκύλος μέσα από μία λήψη χαμηλότερης ευκρίνειας ή ένας παρόμοιος σκύλος με διαφορές ως προς το χρώμα, το μέγεθος, τη στάση κλπ).



**Σχήμα 24.** Τεχνική επαύξησης δεδομένων μέσω τυχαίων τεχνητών παραμορφώσεων της εικόνας [19]

Ορισμένες από τις πιο δημοφιλείς τεχνικές επαύξησης δεδομένων, είναι οι παρακάτω:

- Περιστροφή (rotation): Η εικόνα περιστρέφεται κατά μία γωνία επιλεγμένη με τυχαίο τρόπο.
- Μετατόπιση (translation): Η εικόνα μετατοπίζεται κατά ύψος και πλάτος κατά τυχαίο αριθμό pixels. Τα αρχικά pixels της εικόνας που μετατοπίζονται εκτός των αρχικών ορίων της, αγνοούνται.
- Κλιμάκωση (scaling): Αλλαγή μεγέθους εικόνας (σμίκρυνση ή μεγέθυνση)

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

- Αντικατοπτρισμός (mirroring): Αντικατοπτρισμός της εικόνας κατά ύψος ή/και πλάτος της.
- Περικοπή (crop): Περικοπή τμήματος της εικόνας έπειτα από τυχαίο ορισμό ορθογώνιου πλαισίου που εσωκλείει τη νέα εικόνα.
- Εφαρμογή φίλτρων στην εικόνα. Παραδείγματα: Εισαγωγή θορύβου μορφής αλατοπίπερου (salt & pepper noise), εισαγωγή θορύβου Gaussian κατανομής (gaussian noise), παραμόρφωση χρώματος (color jitter), θόλωση (blur)

Κατά την εφαρμογή της τεχνικής της επαύξησης δεδομένων, συνήθως εφαρμόζεται ταυτόχρονος συνδυασμός των παραπάνω μετασχηματισμών. Για παράδειγμα, ένα “παραμορφωμένο” αντίγραφο της εικόνας μπορεί να προκύψει από περικοπή της εικόνας, στη συνέχεια περιστροφή της κατά 30 μοίρες και τέλος εφαρμογής Θορύβου μορφής αλατοπίπερου.

Μία άλλη ιδιαίτερα σημαντική τεχνική εκπαίδευσης νευρωνικών δικτύων σχετίζεται με τον τρόπο αρχικοποίησης των τιμών για τα βάρη των νευρώνων (weight initialization) [20]. Πριν αρχίσει η εκπαίδευση των δικτύων, θα πρέπει να δοθούν αρχικές τιμές στα βάρη των νευρώνων, ώστε να υπολογιστεί η αρχική πρόβλεψη της πρώτης δέσμης εικόνων, στη συνέχεια να υπολογιστεί το σφάλμα της πρόβλεψης από την πραγματική έξοδο και τέλος οι τιμές αυτές να μεταβληθούν κατάλληλα ώστε στην επόμενη πρόβλεψη το σφάλμα πρόβλεψης να είναι μικρότερο.

Μία αρχική ιδέα είναι να αρχικοποιηθούν όλα τα βάρη με μηδενική τιμή. Η ιδέα αυτή βασίζεται στο γεγονός ότι η έξοδος των μισών περίπου νευρώνων είναι θετική και των υπόλοιπων μισών αρνητική. Κάτι τέτοιο όμως οδηγεί σε πλήρη συμμετρία του δικτύου, καθώς όλοι οι νευρώνες θα υπολογίζουν την ίδια έξοδο και η τιμή ενημέρωσής τους θα είναι ίδια για όλους. Συνεπώς όλοι οι νευρώνες θα υπολογίζουν την ίδια τιμή ενημέρωσης για τα βάρη τους.

Η εναλλακτική λύση είναι οι αρχικές τιμές των νευρώνων να οριστούν κρατώντας μεν μηδενική μέση τιμή, λαμβάνοντας δε την τιμή τους μέσα από κάποια κατανομή πιθανότητας (πχ. Gaussian κατανομή). Υπάρχουν και άλλες μελέτες που προτείνουν αποτελεσματικές τεχνικές αρχικοποίησης των βαρών των νευρώνων [21][22], ξεφεύγουν από τα πλαίσια της εργασίας αυτής.

Η τελευταία τεχνική που συνδέεται με την αρχικοποίηση των βαρών των νευρώνων και συνεισφέρει σημαντικά, τόσο στη μείωση του χρόνου εκπαίδευσης δικτύων, όσο και στην αποτελεσματικότητά τους (βλ. Σχήμα 15, Κεφάλαιο 2), είναι η τεχνική του της μεταφοράς γνώσης (transfer learning). Στην περίπτωση αυτή, η αρχικοποίηση δεν γίνεται μέσω τυχαίων τιμών ούτε μέσω κάποιας κατανομής, αλλά μέσω ενός ίδιου δικτύου που έχει εκπαιδευτεί για κάποιο άλλο σκοπό. Για παράδειγμα, συναντάται πολύ συχνά στη βιβλιογραφία, η αρχικοποίηση των τιμών να γίνεται μέσω βαρών από δίκτυα που έχουν εκπαιδευτεί στην ImageNet [9], η οποία είναι μία τεράστια βάση εικόνων που περιέχει αντικείμενα από πολλές κλάσεις. Ως μεταφορική αναλογία της μεταφοράς γνώσης μπορεί να σκεφτεί κανείς την επίλυση ενός γρίφου. Είναι πολύ πιο εύκολο να λυθεί ένας γρίφος αν διαθέτει ένα στοιχείο (clue), ακόμα και αν το στοιχείο αυτό δεν είναι ιδιαίτερα σημαντικό ή σχετικό με την επίλυση. Δίνει τη δυνατότητα όμως να ξεκινήσει κάπως η έρευνα προς την επίλυση. Από την άλλη, είναι πολύ πιο δύσκολο και χρονοβόρο να επιλυθεί ένας γρίφος χωρίς να υπάρχει κανένα αρχικό στοιχείο.

Επιστρέφοντας λοιπόν στην Ψηφιακή Ταξινόμηση Εγγράφων, η αρχικοποίηση βαρών μέσω της μεταφοράς γνώσης οδηγεί σε καλύτερα αποτελέσματα και μειώνει τον απαιτούμενο χρόνο εκπαίδευσης. Συνήθως επιτυγχάνεται είτε κάνοντας χρήση βαρών από δίκτυα που έχουν εκπαιδευτεί σε βάσεις δεδομένων εικόνων παρόμοιας φύσης (πχ. έγγραφα) (μεταφορά γνώσης εντός του τομέα - intra-domain transfer learning) είτε από δίκτυα που έχουν εκπαιδευτεί σε βάσεις εικόνων διαφορετικής φύσης (πχ. φυσικές εικόνες) (μεταφορά γνώσης μεταξύ τομέων - inter-domain transfer learning)

### Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων



## 4. ΣΥΓΚΡΙΤΙΚΗ ΜΕΛΕΤΗ

### 4.1. ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΨΗΦΙΑΚΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ

Για το πρόβλημα της Ψηφιακής Ταξινόμησης Εγγράφων, δύο είναι οι βασικές βάσεις δεδομένων εικόνων που χρησιμοποιούνται στη βιβλιογραφία για τη μελέτη των προτεινόμενων αλγορίθμων και της αποτελεσματικότητάς τους. Η πρώτη [23], θεωρείται μία βάση εικόνων μικρής κλίμακας (3.482 εικόνες) και ονομάζεται Tobacco3482. Περιλαμβάνει σαρωμένες εικόνες (ασπρόμαυρες) από 10 διαφορετικές κλάσεις. Παράδειγμα αυτών δίνεται στο επόμενο σχήμα (Σχήμα 25). Οι κλάσεις αυτές είναι:

- Memo
- E-mail
- Resume
- Letter
- Report
- Forms
- Advertisement
- Scientific
- Note
- Letter



Σχήμα 25. Παραδείγματα εικόνων από τη βάση δεδομένων Tobacco3482 [23]

Επίσης, η Tobacco3482 δε διαχωρίζεται σε σύνολο εκπαίδευσης, επικύρωσης και δοκιμής (train/validation/test sets). Στη βιβλιογραφία όμως, συνηθίζεται η βάση αυτή να διασπάται σε πέντε τυχαίους συνδυασμούς train-validation-test. Κάθε συνδυασμός περιέχει 100 train εικόνες

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων



από κάθε κλάση, εκ των οποίων πιθανόν οι 20 να αποτελούν το validation set. οι υπόλοιπες της κλάσης χρησιμοποιούνται ως test set. Για τις 10 κλάσεις, συνολικά χρησιμοποιούνται 1.000 εικόνες ως train/validation sets και 2.482 εικόνες ως test set.

Η δεύτερη βάση δεδομένων, ονομάζεται RVL-CDIP [6] και αποτελεί μία μεγάλης κλίμακας βάση δεδομένων, καθώς περιλαμβάνει 400.000 εικόνες (Σχήμα 26) που χωρίζονται σε 16 κλάσεις, οι οποίες είναι:

- letter
- memo
- email
- file folder
- form
- handwritten
- invoice
- advertisement
- budget
- news article
- presentation
- scientific publication
- questionnaire
- resume
- scientific report
- specification



Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

**Σχήμα 26. Παραδείγματα εικόνων από τη βάση δεδομένων RVL-CDIP [6]**

Η RVL-CDIP αποτελεί αυτή τη στιγμή τη σημαντικότερη βάση δεδομένων για την Ψηφιακή Ταξινόμηση Εγγράφων, καθώς όλες οι σχετικές τεχνικές που συναντώνται στη βιβλιογραφία εκπαιδεύονται και συγκρίνονται πάνω σε αυτήν. Υπάρχει επίσημος διαχωρισμός της σε train/validation/test sets και είναι 320.000 εικόνες ως train set, 40.000 ως validation, και οι υπόλοιπες 40.000 ως test set.

**4.2. ΜΕΤΡΙΚΕΣ ΑΠΟΤΙΜΗΣΗΣ ΨΗΦΙΑΚΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ**

Στο υπο-κεφάλαιο αυτό παρουσιάζονται οι μετρικές αποτίμησης Ψηφιακής Ταξινόμησης Εγγράφων, οι οποίες συμβαδίζουν γενικότερα με αυτές που χρησιμοποιούνται ευρύτερα στη βιβλιογραφία [24] σχετικά με το πρόβλημα της Ταξινόμησης/Κατηγοριοποίησης (image classification).

Αρχικά, ορίζονται οι παρακάτω έννοιες/ποσότητες, για κάθε μία κλάση, έστω A:

- **True Positives (TP):** Το σύνολο των εικόνων που ταξινομήθηκαν ότι ανήκουν στην κλάση A (P) και η συγκεκριμένη πρόβλεψη ήταν σωστή (T).
- **False Positives (FP):** Το σύνολο των εικόνων που ταξινομήθηκαν ότι ανήκουν στην κλάση A (P) και η συγκεκριμένη πρόβλεψη ήταν λανθασμένη (F)
- **True Negatives (TN):** Το σύνολο των εικόνων που ταξινομήθηκαν ότι δεν ανήκουν στην κλάση A (N), δηλαδή ανήκουν σε άλλες κλάσεις, και η συγκεκριμένη πρόβλεψη ήταν σωστή (T).
- **False Negatives (FN):** Το σύνολο των εικόνων που ταξινομήθηκαν ότι ανήκουν στην κλάση A (N), δηλαδή ανήκουν σε άλλες κλάσεις, και η συγκεκριμένη πρόβλεψη ήταν λανθασμένη (F)

Έχοντας υπολογίσει τις παραπάνω ποσότητες (συνήθως μετατρέπονται και σε ποσοστά) για κάθε κλάση, μπορούμε να ορίσουμε τις παρακάτω μετρικές [25], για την κλάση, έστω A:

- **Ακρίβεια (Precision):** Είναι το ποσοστό των θετικών προβλέψεων της κλάσης A που ήταν σωστό. Ισχύει ότι:
  - **Precision = TP/(TP+FP)**
- **Ανάκληση (Recall):** Είναι το ποσοστό των πραγματικών σωστών εικόνων της κλάσης A, που αναγνωρίστηκε σωστά. Ισχύει ότι:
  - **Recall = TP/(TP+FN)**
- **Τελική μετρική ακρίβειας (F-Measure):** Είναι η τελική μετρική ακρίβειας που συνυπολογίζει τόσο την ακρίβεια, όσο και την ανάκληση. Ισχύει ότι:
  - **F-Measure = (2 x Recall x Precision) / (Recall + Precision)**

Οι παραπάνω ποσότητες υπολογίζονται για κάθε κλάση/κατηγορία και η μέση τιμή λαμβάνεται ως η πραγματική ακρίβεια του συστήματος. Εύκολα συμπεραίνει κανείς ότι όσο πιο υψηλές είναι οι παραπάνω τιμές (που συνοψίζονται τελικά στην F-Measure), τόσο πιο αποτελεσματικό είναι

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

ένα σύστημα ταξινόμησης. Ένας άλλος τρόπος να αντιληφθεί κάποιος τις ποσότητες TP, FP, TN, FN είναι μέσω της μήτρας σύγχυσης (confusion matrix).

**ΠΙΝΑΚΑΣ 4: Μήτρα σύγχυσης αποτελεσμάτων ενός συστήματος πρόβλεψης καρκινικών όγκων σε μαστογραφίες**

	ΑΛΗΘΙΝΟΣ ΚΑΡΚΙΝΟΣ: ΝΑΙ	ΑΛΗΘΙΝΟΣ ΚΑΡΚΙΝΟΣ: ΟΧΙ
ΠΡΟΒΛΕΨΗ ΚΑΡΚΙΝΟΥ: ΝΑΙ	<b>True Positive (TP)</b>	<b>False Positive (FP)</b>
ΠΡΟΒΛΕΨΗ ΚΑΡΚΙΝΟΥ: ΟΧΙ	<b>False Negative (FN)</b>	<b>True Negative (TN)</b>

Ένα παράδειγμα δίνεται στο Πίνακα 4 και σχετίζεται με το δυαδικό πρόβλημα (η απόφαση είναι επιλογή από ναι ή όχι, συνεπώς 2 κλάσεις), είναι αυτό της πρόβλεψης ενός συστήματος για ύπαρξη ή όχι καρκινικού όγκου (πχ. μέσω μαστογραφίας). Στο σχήμα δίνεται μία οπτικοποίηση του ορισμού των ποσοτήτων αυτών.

Εν αντιθέσει με τα δυαδικά προβλήματα όπου οι πιθανές αποφάσεις ενός συστήματος είναι ναι ή όχι, στα περισσότερα προβλήματα όπως αυτό της ταξινόμησης εγγράφων πιθανές κλάσεις είναι παραπάνω από δύο. Για παράδειγμα (βλ. υπο-κεφάλαιο 4.1), στην Tobacco3482 οι κλάσεις είναι 10, ενώ στην RVL-CDIP οι κλάσεις είναι 16.

**ΠΙΝΑΚΑΣ 5: Μήτρας σύγχυσης αποτελεσμάτων για περισσότερες από δύο κλάσεις**

		ΠΡΑΓΜΑΤΙΚΗ ΚΑΤΗΓΟΡΙΑ		
		Γάτα	Ψάρι	Κότα
ΠΡΟΒΛΕΠΟΜΕΝΗ ΚΑΤΗΓΟΡΙΑ	Γάτα	4	6	3
	Ψάρι	1	2	0

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

	<b>Κότα</b>	<b>1</b>	<b>2</b>	<b>6</b>
--	-------------	----------	----------	----------

Οι αναγωγή της μήτρας σύγχυσης για περισσότερες κλάσεις είναι αντίστοιχη και ένα παράδειγμα δίνεται στον Πίνακα 5. Έστω η κατηγορία Ψάρι. Τότε, από το σχήμα παρατηρεί κανείς ότι:

- TP: Το δεύτερο στοιχείο της διαγωνίου, δηλαδή 2
- TN: Το άθροισμα όλων των στοιχείων της διαγωνίου εκτός του δευτέρου, δηλαδή  $4+6=10$
- FP: Όλα τα στοιχεία της δεύτερης γραμμής, εκτός αυτού της δεύτερης στήλης, δηλαδή  $1+0=1$
- FN: Όλα τα στοιχεία της δεύτερης στήλης, εκτός αυτού της δεύτερης γραμμής, δηλαδή  $6+2=8$

Από την παραπάνω ανάλυση, μπορούμε εύκολα με τους προαναφερόμενους τύπους να υπολογίσουμε Recall, Precision και F-Measure για την κατηγορία Ψάρι. Επαναλαμβάνουμε τα αντίστοιχα για τις άλλες δύο κατηγορίες (Γάτα, Κότα) και τελικά η ακρίβεια του συστήματος δίνεται ως ο μέσος όρος των υπολογισμένων F-Measure των τριών κατηγοριών.

#### 4.3. ΣΥΓΚΡΙΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Στο υπο-κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα των καλύτερων μεθοδολογιών και τεχνικών που συναντώνται στη βιβλιογραφία ([5], [6], [7], [8], [10], [11], [12], [13], [14]), σχετικά με το πρόβλημα της Ψηφιακής Ταξινόμησης Εγγράφων. Οι βάσεις δεδομένων που εξετάζονται περιγράφηκαν προηγουμένως και είναι η Tobacco3482 και η RVL-CDIP. Ταυτόχρονα, παρουσιάζεται η συνεισφορά εφαρμογής διαφόρων τεχνικών και επιλογής παραμέτρων, όπως αυτές παρουσιάστηκαν στα Κεφάλαια 2 και 3.

Αρχικά, μελετάμε την περίπτωση όπου ένα Νευρωνικό Δίκτυο μπορεί να μάθει πληροφορία από μέρος της εικόνας, σύμφωνα και με την τεχνική που παρουσιάστηκε στο [6]. Στη συγκεκριμένη δημοσίευση, εκπαιδεύεται ένα ξεχωριστό σύστημα για κάθε ένα από τα παρακάτω μέρη της εικόνας:

- Κεφαλίδα (άνω μέρος)
- Δεξιό σώμα
- Αριστερό σώμα
- Υποσέλιδο (κάτω μέρος)
- Ολόκληρη έγγραφο

Τα αποτελέσματα των παραπάνω συστημάτων, καθώς και του συνδυασμού όλων των παραπάνω, δίνεται στον επόμενο πίνακα (Πίνακας 6):

#### Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

**ΠΙΝΑΚΑΣ 6: Αποτελέσματα ταξινόμησης εγγράφων (F-Measure) χρησιμοποιώντας διαφορετικά μέρη του εγγράφου [6]**

	<b>Tobacco 3482</b>	<b>RVL-CDIP</b>
<b>Σύστημα Κεφαλίδας</b>	71,00 %	84,90 %
<b>Σύστημα Αριστερού Σώματος</b>	66,70 %	82,70 %
<b>Σύστημα Δεξιού Σώματος</b>	70, 80 %	79,50 %
<b>Σύστημα Υποσέλιδου</b>	62,20 %	79,40 %
<b>Σύστημα Ολόκληρου Εγγράφου</b>	75, 60 %	<b>89,80 %</b>
<b>Συνδυασμός των παραπάνω</b>	<b>79,90 %</b>	89,30 %

Σύμφωνα με τον παραπάνω πίνακα, στη βάση Tobacco3482 (μικρή βάση 3482 εικόνων) ο συνδυασμός των προβλέψεων των επιμέρους δικτύων έδωσε με διαφορά το καλύτερο αποτέλεσμα. Αυτό αποδεικνύει την ικανότητα των δικτύων να μάθουν καλύτερα μέσα από τμηματική πληροφορία. Παράλληλα, στην RVL-CDIP το σύστημα που λαμβάνει ως είσοδο ολόκληρο το έγγραφο έδωσε τα καλύτερα αποτελέσματα και σχεδόν τα ίδια έδωσε και στο σύστημα του συνδυασμού όλων των επιμέρους. Αυτό δείχνει ότι για μεγάλες βάσεις δεδομένων (η RVL-CDIP περιέχει 400.000 εικόνες) και άρα μεγάλο αριθμό δεδομένων εκπαίδευσης, ένα νευρωνικό δίκτυο μπορεί να μάθει/κατανοήσει καλύτερα τοπική πληροφορία που μπορεί να αποτελεί και κριτήριο απόφασης/διαχωρισμού.

Η επόμενη τεχνική που μελετάται είναι αυτή της αρχικοποίησης των τιμών για τα βάρη των νευρώνων (weight initialization) και της μεταφοράς γνώσης (transfer learning). Με αφορμή τη δημοσίευση [8] και όπως φαίνεται στον επόμενο πίνακα (Πίνακας 7), διαπιστώνεται η σημασία της μη-τυχαίας αρχικοποίησης των τιμών για τα βάρη των νευρώνων. Επιπλέον, τονίζεται πως η μεταφορά γνώσης από βάση εικόνων παρόμοιας φύσης (Document Pretraining) δίνει τα

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

βέλτιστα αποτελέσματα συγκριτικά με μεταφορά γνώσης από βάση δεδομένων εικόνων διαφορετικής φύσης (ImageNet Pretraining).

Διευκρινίζεται ότι όπως παρουσιάζεται στη δημοσίευση [8], αποτελέσματα για την περίπτωση του Document Training δίνονται μόνο στην περίπτωση της Tobacco3482, όπου έχει προηγηθεί προ-εκπαίδευση στην RVL-CDIP. Η αντίστροφη περίπτωση δεν παρέχεται.

Σύμφωνα με τον παρακάτω πίνακα, διαπιστώνεται εύκολα πως ανεξαρτήτως της αρχιτεκτονικής Νευρωνικού Δικτύου που χρησιμοποιείται, η αρχικοποίηση των τιμών για τα βάρη των νευρώνων δε θα πρέπει να είναι μέσα από τυχαίες τιμές. Στην ιδανική περίπτωση, θα πρέπει να διατίθεται ένα εκπαιδευμένο δίκτυο πάνω κάτω παρόμοιας φύσης βάση δεδομένων εικόνων (intra-domain transfer learning). Εναλλακτικά, βάσεις δεδομένων “γενικής χρήσης”, όπως η ImageNet, μπορούν να χρησιμοποιηθούν για την προ-εκπαίδευση του δικτύου (inter-domain transfer learning) και τον καθορισμό των αρχικών τιμών για τα βάρη των νευρώνων.

**ΠΙΝΑΚΑΣ 7: Αποτελέσματα ταξινόμησης εγγράφων (F-Measure), κάνοντας χρήση διαφόρων τεχνικών αρχικοποίησης [8]**

	Tobacco3482			RVL-CDIP	
	Document Pretraining	ImageNet Pretraining	No Pretraining	ImageNet Pretraining	No Pretraining
AlexNet	90,04 %	75,73 %	62,49 %	88,60 %	88,19 %
GoogleNet	88,40 %	72,98 %	70,28 %	89,02 %	88,60 %

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

<b>VGG-16</b>	<b>91,01 %</b>	77,52 %	69,50 %	<b><u>90,97 %</u></b>	89,41 %
<b>ResNet50</b>	<b><u>91,13 %</u></b>	67,93 %	59,55 %	<b>90,40 %</b>	89,24 %

Στον επόμενο πίνακα (Πίνακας 8) και σύμφωνα με τις τεχνικές που παρουσιάστηκαν στη δημοσίευση [10], μελετώνται οι διάφορες τεχνικές επαύξησης δεδομένων (data augmentation) (παρουσιάζονται οι κυριότερες εξ αυτών) και πως αυτές επηρεάζουν την ακρίβεια ταξινόμησης.

**ΠΙΝΑΚΑΣ 8: Αποτελέσματα ταξινόμησης εγγράφων (F-Measure) χρησιμοποιώντας διαφορετικά είδη μετασχηματισμών για επαύξηση δεδομένων [10]**

<b>Είδος Μετασχηματισμού</b>	<b>RVL-CDIP</b>
<b>Χωρίς Μετασχηματισμό</b>	88,30 %
<b>Παραμόρφωση Χρώματος (color jitter)</b>	88,36 %
<b>Περικοπή (crop)</b>	88,83 %
<b>Περιστροφή (rotation)</b>	88,74 %

<b>Στρέβλωση (Shear)</b>	89,33 %
<b>Κλιμάκωση (Scale) (384x384) + Shear</b>	<b>90,94 %</b>

Παρατηρείται λοιπόν πως τεχνικές επαύξησης δεδομένων μπορούν να συνεισφέρουν στη βελτίωση της ταξινόμησης των εγγράφων, καθώς το δίκτυο είναι σε θέση να μάθει μέσα από πιο ετερογενή πληροφορία, περισσότερα δεδομένα και χωρίς να “βλέπει” την ίδια εικόνα παραπάνω από μια φορά. Ειδικότερα, ο συνδυασμός πολλών μετασχηματισμών δίνει σε αρκετές περιπτώσεις καλύτερα αποτελέσματα. Επιπλέον, το μέγεθος της εικόνας εισόδου έχει σημαντικό ρόλο, καθώς η σμίκρυνση της εικόνας σε μικρό ύψος και πλάτος μπορεί να οδηγήσει σε απώλεια πληροφορίας.

Τέλος, στον επόμενο πίνακα (Πίνακας 9) δίνονται τα καλύτερα αποτελέσματα από τις τεχνικές που προτείνονται στη βιβλιογραφία για την ταξινόμηση εγγράφων ([5], [6], [7], [8], [10], [11], [12], [13], [14]). Ταυτόχρονα, όπως φαίνεται και στον πίνακα, δίνεται επιγραμματικά και η γενική περιγραφή της εκάστοτε μεθοδολογίας με τη μορφή σχολίων. Διευκρινίζεται ότι και ορισμένες περιπτώσεις όπου από την εκάστοτε δημοσίευση δεν παρουσιάστηκαν αποτελέσματα ταξινόμησης σε κάποια από τις δύο βάσεις εικόνων που εξετάζονται.

**ΠΙΝΑΚΑΣ 9: Αποτελέσματα ταξινόμησης εγγράφων (F-Measure) για διάφορες μεθοδολογίες που προτείνονται στη βιβλιογραφία**

	<b>Tobacco 3482</b>	<b>RVL-CDIP</b>	<b>Παρατηρήσεις</b>
<b>Μέθοδος 1 [5]</b>	65,35 %	-	AlexNet, είσοδος 150x150
<b>Μέθοδος 2 [6]</b>	79,90 %	89,80 %	AlexNet, εκπαίδευση μέσω τμημάτων της εικόνας, μείωση δεδομένων με PCA

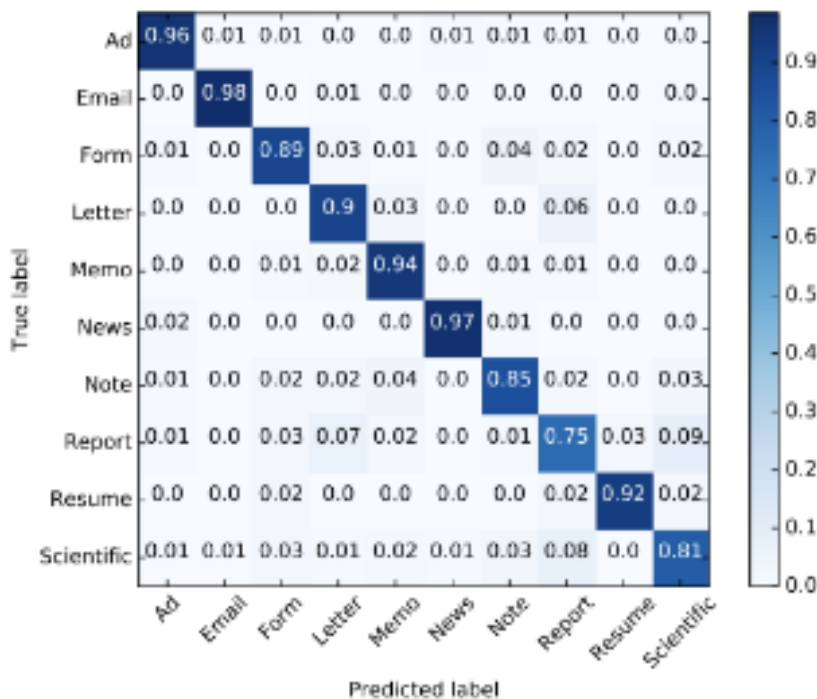


<b>Μέθοδος 3 [7]</b>	-	90,70 %	GoogleNet, αρχικοποίηση με ImageNet
<b>Μέθοδος 4 [8]</b>	<b>91,13 %</b>	90,97 %	Resnet-50, VGG-16, επαύξηση δεδομένων με τυχαία περικοπή, μεταφορά γνώσης από βάση παρόμοιας φύσης
<b>Μέθοδος 5 [10]</b>	-	90,94 %	AlexNet, επαύξηση δεδομένων με διάφορους μετασχηματισμούς, είσοδος 384x384
<b>Μέθοδος 6 [11]</b>	-	92,21 %	VGG-16, μεταφορά γνώσης από βάση παρόμοιας φύσης, εκπαίδευση σε τμήματα της εικόνας, μετα- ταξινομητής
<b>Μέθοδος 7 [12]</b>	<b>93,2%</b>	<b>96,4%</b>	Συνδυασμός image & textual features, αποτίμηση σε υποσύνολο κλάσεων
<b>Μέθοδος 8 [13]</b>	87,8%	90,6%	Συνδυασμός image & textual features

<b>Μέθοδος 9 [14]</b>	-	<b>94,42 %</b>	Συνδυασμός OCR, δομής της εικόνας (κατάτμηση) - μοντέλο BERT και χαρακτηριστικών εικόνας
-----------------------	---	----------------	--

Από τον παραπάνω πίνακα μπορούν εξαχθούν σημαντικά συμπεράσματα σχετικά με την Ταξινόμηση Εγγράφων. Αρχικά, ο συνδυασμός κειμενικής και χωρικής/δομικής πληροφορίας με χαρακτηριστικά του εγγράφου, δίνουν τα καλύτερα αποτελέσματα (94,42 % , Μέθοδος 9 [14]). Η συγκεκριμένη μέθοδος συνδυάζει επιμέρους τεχνικές όπως αυτή της επαύξησης δεδομένων, αρχικοποίησης βαρών των νευρώνων κλπ και ταυτόχρονα συνδυάζει πιο σύνθετες αρχιτεκτονικές νευρωνικών δικτύων που ξεφεύγουν από τα πλαίσια της εργασίας αυτής. Όμως, στις περισσότερες περιπτώσεις, η διάθεση κειμενικής και χωρικής πληροφορίας για ένα έγγραφο δεν είναι εφικτή ή ακριβής, καθώς τα συγκεκριμένα στάδια επεξεργασίας είναι αλληλένδετα με την κατηγορία του εγγράφου. Συνεπώς, η ταξινόμηση προ-απαιτείται του υπολογισμού και εντοπισμού της δομής, καθώς και της αναγνώρισης γραφής του εγγράφου. Στη Μέθοδο 7 δίνονται μεν τα κορυφαία αποτελέσματα, αλλά η αποτίμηση παρουσιάζεται σε υποσύνολο των κλάσεων και συνεπώς δεν μπορεί να θεωρηθεί ως η βέλτιστη.

Ως εναλλακτική λύση, παρατηρείται ότι τα Συνελικτικά Νευρωνικά Δίκτυα δίνουν ικανοποιητικά αποτελέσματα και όταν συνδυαστούν με διάφορες από τις προαναφερόμενες τεχνικές, βελτιώνουν ακόμα περισσότερο την αποδοτικότητά τους η οποία κυμαίνεται στο 89,80% - 92,21%.



Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

**Σχήμα 27. Παράδειγμα Error Analysis για τη βάση δεδομένων Tobacco3482 [8]**

Για την περαιτέρω βελτίωση της αποτελεσματικότητας, θα πρέπει κανείς να είναι σε θέση να κατανοήσει και να ερμηνεύσει τα σφάλματα (περιπτώσεις λανθασμένων κατηγοριοποιήσεων) που προκύπτουν από τα παραπάνω συστήματα. Απαιτείται, δηλαδή, μια Ανάλυση Σφαλμάτων (Error Analysis). Ένα τέτοιο ενδεικτικό παράδειγμα για τη βάση δεδομένων Tobacco3482 δόθηκε στο σύστημα που προτάθηκε στο [8] και παρουσιάζεται μέσω της μήτρας σύγχυσης αποτελεσμάτων και διακρίνεται στο Σχήμα 27.

Παρατηρούμε ότι οι χαμηλότερη αποτελεσματικότητα (F-Measure) εμφανίζεται σε είδη εγγράφων όπως:

- **Report:** Αναφορές, συνήθως δομημένα έγγραφα με εικόνες, σχεδιαγράμματα, μαθηματικούς τύπους, κ.λπ.
- **Scientific:** Επιστημονικές Δημοσιεύσεις, άρθρα από περιοδικά και συνέδρια, δομημένα και περιλαμβάνουν συνήθως εικόνες, σχήματα, σχεδιαγράμματα, μαθηματικούς τύπους, κ.λπ.
- **Note:** Σημειώσεις, συνήθως περιλαμβάνουν χειρόγραφο κείμενο, ίσως και τυπωμένο και είναι πιθανό να έχουν κάποια δομή.
- **Form:** Φόρμες, συνήθως πρόκειται για έγγραφα με προτυποποιημένη μορφή, που περιλαμβάνουν είτε τυπωμένο κείμενο είτε τυπωμένο κείμενο και χειρόγραφο, συμπληρωμένα πεδία, κλπ.

Αν προσπαθήσει κανείς να ερμηνεύσει τον παραπάνω πίνακα, θα διαπιστώσει πως υπάρχει φυσική ερμηνεία για τις κατηγορίες εγγράφων με τη χαμηλότερη μετρική αποτελεσματικότητας, διότι:

- Υπάρχει έντονη συσχέτιση μεταξύ των κλάσεων Report και Scientific, κάτι που αντικατοπτρίζεται και σε υψηλό αριθμό σφαλμάτων, δηλαδή είναι σχετικά σύνηθες η μία κλάση να αναγνωρίζεται ως η άλλη και αντίστροφα. Κάτι τέτοιο είναι αναμενόμενο καθώς πρόκειται για έγγραφα παρόμοιας φύσης και είναι πιθανόν ακόμη και κάποιος άνθρωπος να τα ταξινομήσει λανθασμένα λόγω αδυναμίας διαχωρισμού.
- Ομοίως, υπάρχει έντονη συσχέτιση μεταξύ των κατηγοριών Form και Report.
- Τέλος, παρατηρούμε ότι η κλάση Note (Σημειώσεις) συγγέεται συχνά με την κλάση Memo. Η τελευταία μπορεί να περιέχει μία υπενθύμιση (πχ χειρόγραφη ή ένα σύντομο μήνυμα) και άρα ερμηνεύεται ορθώς και το παραπάνω σφάλμα.

Από την παραπάνω ανάλυση προκύπτει η διαίσθηση ότι τα Νευρωνικά Δίκτυα αποτελούν ένα τεχνητό αντίγραφο της ανθρώπινης σκέψης. Μερικές αναλογίες διακρίνονται στη συνέχεια:

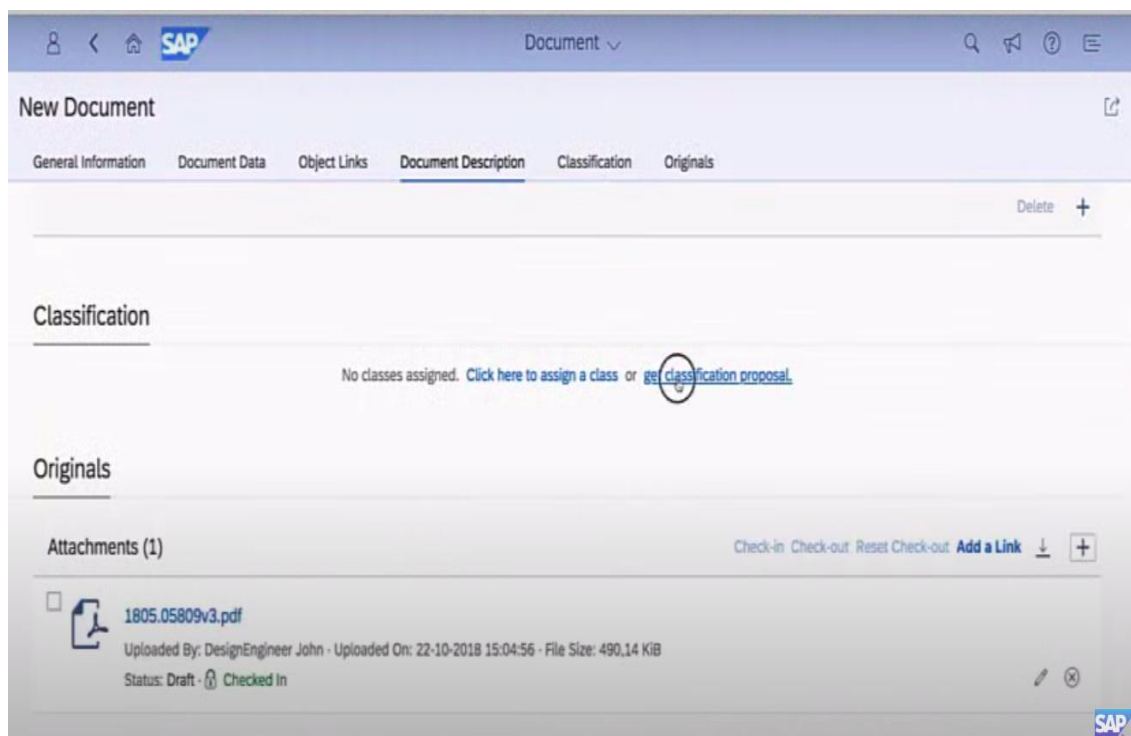
- Αναπτύσσονται και βελτιώνονται μέσω της διαδικασίας εκμάθησης.
- Η υπολογιστική ισχύς τους επιτρέπει να διαχειρίζονται (μεγαλύτερο από τον αυτόν του ανθρώπου) όγκο δεδομένων.
- Τα σφάλματα αναγνώρισης συχνά είναι παρόμοια με αυτά του ανθρώπινου εγκεφάλου.

## 5. ΟΛΟΚΛΗΡΩΜΕΝΟ ΣΥΣΤΗΜΑ ΑΥΤΟΜΑΤΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ

Στο Κεφάλαιο αυτό, παρουσιάζονται συνοπτικά οι υπάρχουσες εφαρμογές που υποστηρίζουν τεχνολογίες Αυτόματης Ταξινόμησης Εγγράφων. Έπειτα, αναλύονται οι γενικές τεχνολογικές προδιαγραφές που πρέπει να πληροί ένα τέτοιο σύστημα με στόχο να μπορούν να υποστηριχθούν μέσα από αυτό οι αλγόριθμοι Μηχανικής Μάθησης και εκπαίδευσης Νευρωνικών Δικτύων που σχετίζονται με το πρόβλημα και περιγράφηκαν στα προηγούμενα κεφάλαια. Τέλος, αναλύονται οι απαιτήσεις λειτουργικότητας που πρέπει να καλύπτει ένα τέτοιο σύστημα, ώστε να είναι δυνατή η χρήση του από διάφορα είδη χρηστών (πχ. απλοί, προχωρημένοι) με τρόπο άμεσο, εύκολο και κατανοητό.

### 5.1. ΕΠΙΣΚΟΠΗΣΗ ΣΥΣΤΗΜΑΤΩΝ ΑΥΤΟΜΑΤΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ

Αυτήν τη στιγμή δεν υπάρχει μεγάλο πλήθος εφαρμογών που να είναι δημοφιλείς στο κοινό και να υποστηρίζουν την ταξινόμηση εγγράφων με χρήση συστημάτων Νευρωνικών Δικτύων. Μία πρώτη περίπτωση μελετάται στο σύστημα Ταξινόμησης Εγγράφων της SAP [26], όπου παρέχεται ένα εργαλείο για την ταξινόμηση επιχειρηματικών εγγράφων σε κατηγορίες που καθορίζονται από τον χρήστη. Ταυτόχρονα υπάρχει υποβοήθηση ώστε να εφαρμοστεί μηχανική εκμάθηση για να αυτοματοποιηθεί η διαχείριση και η επεξεργασία μεγάλων ποσοτήτων επιχειρηματικών εγγράφων. Η ταξινόμηση εγγράφων αποτελεί μέρος του χαρτοφυλακίου SAP AI Business Services. Επιπλέον, αποτελεί συνδρομητική υπηρεσία υπολογιστικού νέφους (cloud-based) και το κόστος της σχετίζεται με το πλήθος εγγράφων που ο χρήστης επιθυμεί να επεξεργαστεί εντός διαφόρων χρονικών διαστημάτων (κλιμακωτή χρέωση). Τέλος, δίνεται η δυνατότητα προσαρμοζόμενης εκπαίδευσης συστήματος Νευρωνικού Δικτύου, για αναγνώρισης σε σύνολο εγγράφων που επιθυμεί ο χρήστης. Παράδειγμα του περιβάλλοντος της εφαρμογής δίνεται στο Σχήμα 28.



Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

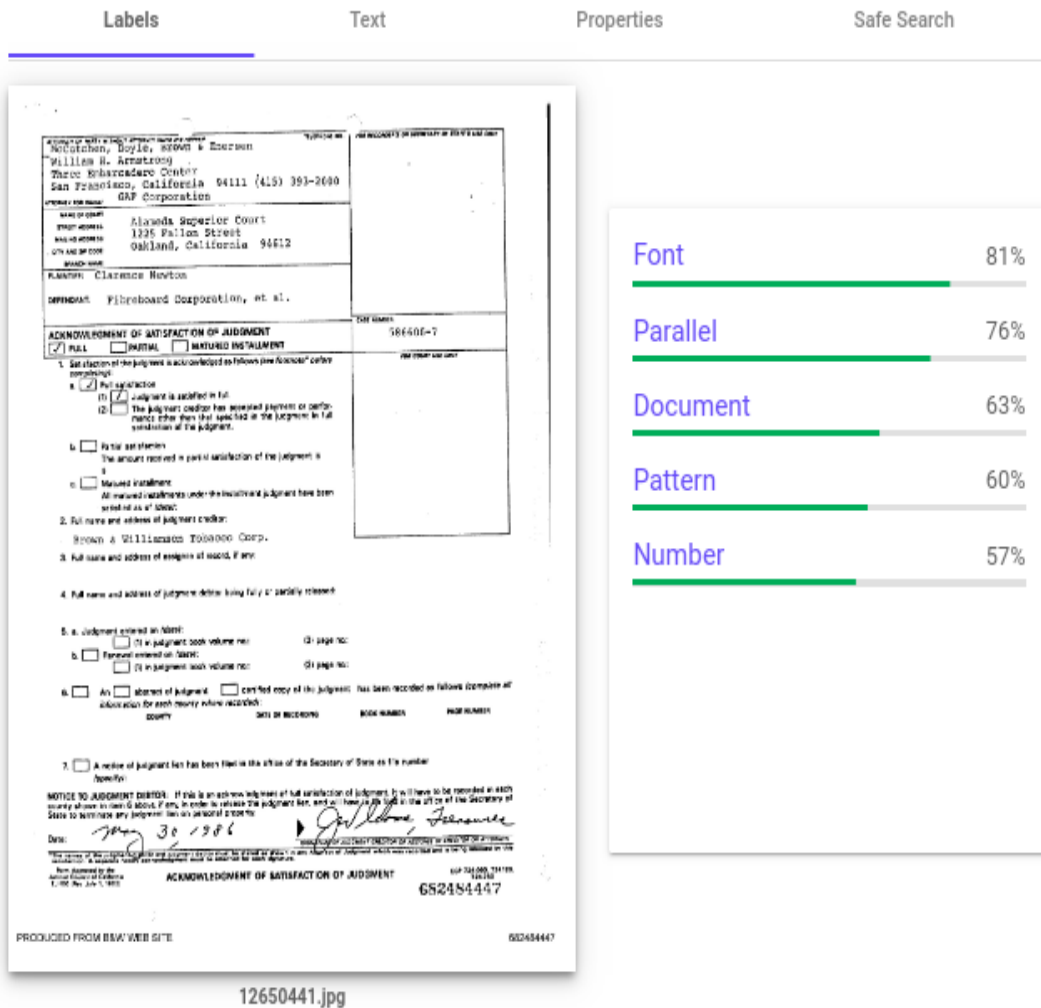
**Σχήμα 28. Στιγμιότυπο εφαρμογής Ταξινόμησης Εγγράφων της SAP [26]**

Μία άλλη περίπτωση συστήματος Ταξινόμησης Εγγράφων, είναι αυτό της ABBYY [27]. Πρόκειται και εδώ για εμπορικό προϊόν (ABBYY FineReader Engine), όπου ο χρήστης προμηθεύεται μία εφαρμογή που συνδέεται με τις υπηρεσίες νέφους που παρέχονται. Πρόκειται και εδώ για ένα σύστημα που κατηγοριοποιεί αυτόματα τα έγγραφα και τα ταξινομεί σε προκαθορισμένες κατηγορίες εγγράφων. Η προηγμένη ταξινόμηση εγγράφων αξιοποιεί σύγχρονες τεχνολογίες, όπως Μηχανική Μάθηση και Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP). Αυτές οι τεχνολογίες είναι σε θέση να ανιχνεύσουν ακόμη και λεπτές διαφορές μεταξύ μεμονωμένων κατηγοριών εγγράφων και επιτρέπουν τη δημιουργία ευέλικτων και επεκτάσιμων διαδικασιών ταξινόμησης που μπορούν να διακρίνουν λεπτομερώς μεταξύ πολλών κατηγοριών εγγράφων. Η διαδικασία ταξινόμησης αποτελείται από 3 βασικά στάδια:

1. Προετοιμασία συνόλων δεδομένων για εκπαίδευση ταξινόμησης. Σε αυτό το βήμα, καθορίζονται οι ζητούμενες κλάσεις εγγράφων. Για κάθε κατηγορία εγγράφων, επιλέγονται διάφορα παραδείγματα εγγράφων - με παρόμοια εμφάνιση ή / και περιεχόμενο -. Με τη βοήθεια αλγορίθμων Μηχανικής Μάθησης και Επεξεργασίας Φυσικής Γλώσσας, το σύστημα ταξινόμησης της ABBYY αναλύει τα έγγραφα εκπαίδευσης σε κάθε κατηγορία εγγράφων και καθορίζει παραμέτρους που πρέπει να χρησιμοποιηθούν για τον προσδιορισμό της αντίστοιχης κατηγορίας εγγράφων.
2. Εκπαίδευση του μοντέλου ταξινόμησης. Κατά τη διάρκεια αυτού του βήματος, πληροφορίες σχετικά με τις κατηγορίες εγγράφων και τις αντίστοιχες παραμέτρους εισάγονται στο μοντέλο ταξινόμησης και το μοντέλο ταξινόμησης εκπαιδεύεται. Το μοντέλο μπορεί να χρησιμοποιήσει Image Classifier, Text Classifier ή συνδυασμό και των δύο. Η απόδοση μπορεί να βελτιστοποιηθεί καθορίζοντας την ισορροπία μεταξύ υψηλής ανάκλησης (recall) και υψηλής ακρίβειας (precision). Ταυτόχρονα, υποστηρίζεται και αυτόματη επικύρωση δεδομένων για τη δοκιμή της ποιότητας και αποτελεσματικότητας του μοντέλου ταξινόμησης.
3. Εφαρμογή εκπαιδευμένου συστήματος ταξινόμησης. Κατά τη διαδικασία ταξινόμησης, το εκπαιδευμένο μοντέλο ταξινόμησης αναλύει κάθε εισερχόμενο έγγραφο. Για τον σωστό προσδιορισμό του τύπου εγγράφου, το Μοντέλο Ταξινόμησης υπολογίζει τις παραμέτρους που ζητήθηκαν για κάθε έγγραφο και τις συγκρίνει με τις πληροφορίες που έλαβε κατά τη διάρκεια του βήματος εκπαίδευσης. Οι προγραμματιστές μπορούν να δημιουργήσουν μια ρουτίνα, η οποία επιτρέπει στους χρήστες να ενημερώσουν με ευελιξία το σύνολο δεδομένων εκπαίδευσης και να εκπαιδεύσουν εκ νέου το μοντέλο ταξινόμησης.

Τέλος, παρουσιάζεται το σύστημα Ταξινόμησης της Google, Google Cloud Vision API [28]. Πρόκειται για ένα αρχικά προσανατολισμένο σύστημα ταξινόμησης για εικόνες και όχι για έγγραφα συγκεκριμένα. Παρόλα αυτά, παρέχεται η δυνατότητα προσαρμοζόμενης εκπαίδευσης και συνεπώς οι παρεχόμενες υπηρεσίες μπορούν να προσαρμοστούν με στόχο την εκπαίδευση συστημάτων ταξινόμησης μέσα από διαφορετικές κλάσεις εγγράφων. Αντίστοιχα με την προσφερόμενη λύση της SAP και εδώ πρόκειται για μία λύση υπηρεσιών υπολογιστικού νέφους (cloud-based services) όπου ο χρήστης δεσμεύει-αγοράζει υπολογιστικούς πόρους (αναλύονται στην επόμενη υπο-ενότητα) και ανάλογα με το πλήθος των δεδομένων που επιθυμεί να επεξεργαστεί, χρεώνεται αντίστοιχα βάσει κλίμακας. Στο επόμενο σχήμα, δίνεται ένα στιγμιότυπο όπου διακρίνονται διάφορες κατηγορίες όπου πιθανόν ανήκει το έγγραφο, μέσα από υπηρεσία που παρέχεται στους χρήστες δωρεάν και αποτελεί παρουσίαση της υπηρεσίας (Σχήμα 29). Τονίζεται, πως, όπως φαίνεται και στο σχήμα, ο προεπιλεγμένος ταξινομητής αφορά στην αναγνώριση της εικόνας μέσα από ένα πλήθος γενικών κατηγοριών και όχι μέσα από κατηγορίες που σχετίζονται αποκλειστικά με έγγραφα.

**Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων**



Σχήμα 29. Στιγμιότυπο εφαρμογής Ταξινόμησης Εικόνας της Google (Cloud Vision API) [28]

5.2. ΓΕΝΙΚΕΣ ΠΡΟΔΙΑΓΡΑΦΕΣ

Ένα σύστημα Αυτόματης Ταξινόμησης Εγγράφων πρέπει να πληροί διάφορες τεχνολογικές προδιαγραφές, ώστε να μπορεί να αποτελέσει ένα χρήσιμο εργαλείο στην καθημερινότητα. Η τεχνολογία εξελίσσεται διαρκώς, επομένως το προτεινόμενο σύστημα θα πρέπει να περιλαμβάνει: α) τεχνολογία που βρίσκεται στην αιχμή της επιστήμης, β) δυνατότητες πλήρους επεκτασιμότητας και αναβάθμισης και γ) δυνατότητα διαλειτουργικότητας με άλλα συστήματα.

Υπό αυτό το πρίσμα και λαμβάνοντας υπόψη και τη φιλοσοφία των υπαρχόντων συστημάτων που παρουσιάστηκαν συνοπτικά παραπάνω, ένα σύστημα Αυτόματης Ταξινόμησης Εγγράφων, θα πρέπει:

- Θα πρέπει να ακολουθεί τη λογική των υπηρεσιών υπολογιστικού νέφους (cloud based services), καθώς αυτό εκμηδενίζει τις απαιτήσεις εξοπλισμού από την πλευρά του

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

χρήστη. Με τον τρόπο αυτό, ο χρήστης θα μπορεί μέσα από ένα οικείο πρόγραμμα περιήγησης (browser) να χρησιμοποιήσει το σύστημα και τις δυνατότητές του.

- Η πλατφόρμα που θα υποστηρίζει το σύστημα αυτό, θα πρέπει να ακολουθεί σύγχρονες τεχνολογίες και να βασίζεται σε open-source λογισμικό και να υποστηρίζεται πολυνηματική επεξεργασία (multi-threaded).
- Η ανάπτυξη των εφαρμογών να γίνεται με χρήση αποθετηρίων λογισμικού ώστε να είναι εφικτή η συνεργατική επεξεργασία και παρακολούθηση. Παράδειγμα αποθετηρίων τέτοιου είδους είναι τα γνωστά αποθετήρια λογισμικού GitHub και το GitLab.
- Θα πρέπει να είναι mobile-friendly, η πρόσβαση και η ευρετηρίαση του site να είναι εύκολη.
- Η προβολή εικόνων εγγράφου υψηλής ανάλυσης και εκτέλεσης λειτουργιών υψηλής απαίτησης σε υπολογιστικούς πόρους (μνήμη-ταχύτητα επεξεργαστή) να είναι στα πλαίσια του ανεκτού και ρεαλιστικού χρόνου απόκρισης.
- Η εφαρμογή να συμμορφώνεται με τα κοινά αποδεκτά πρωτόκολλα ασφαλείας (πχ. HTTPS, SSL, λπ)
- Να παρέχεται sitemap στα πλαίσια του schema.org
- Να χρησιμοποιούνται τα κατάλληλα tags στις html λίστες και meta-descriptors για την εύκολη αναζήτηση
- Να γίνει η κατάλληλη έρευνα των λέξεων-κλειδιών (keywords) με εργαλεία όπως το Google Trends, το Google Keyword Planner και το SEMrush. Χρήση των keywords σε τίτλους, υπότιτλους, εναλλακτικές ονομασίες εικόνων, URLs και anchor texts.
- Να υπάρχει δυνατότητα επιλεγμένοι χρήστες να μπορούν να εισάγουν περιεχόμενο (πχ. σελίδα ή βιβλίο) στην εφαρμογή και να εκτελέσουν τις παρεχόμενες λειτουργίες εφόσον το επιθυμούν
- Να γίνεται χρήση σύγχρονων γλωσσών προγραμματισμού όπως Php, Laravel, MySQL, Javascript, HTML5, Python, Java, C#
- Σχεδιασμός Βάσης δεδομένων με κατάλληλες συσχετίσεις
- Σύγχρονα πρωτόκολλα επικοινωνίας (πχ. REST APIs)
- Εισαγωγή δεδομένων από διάφορες πηγές (πχ. json, xls, xml)

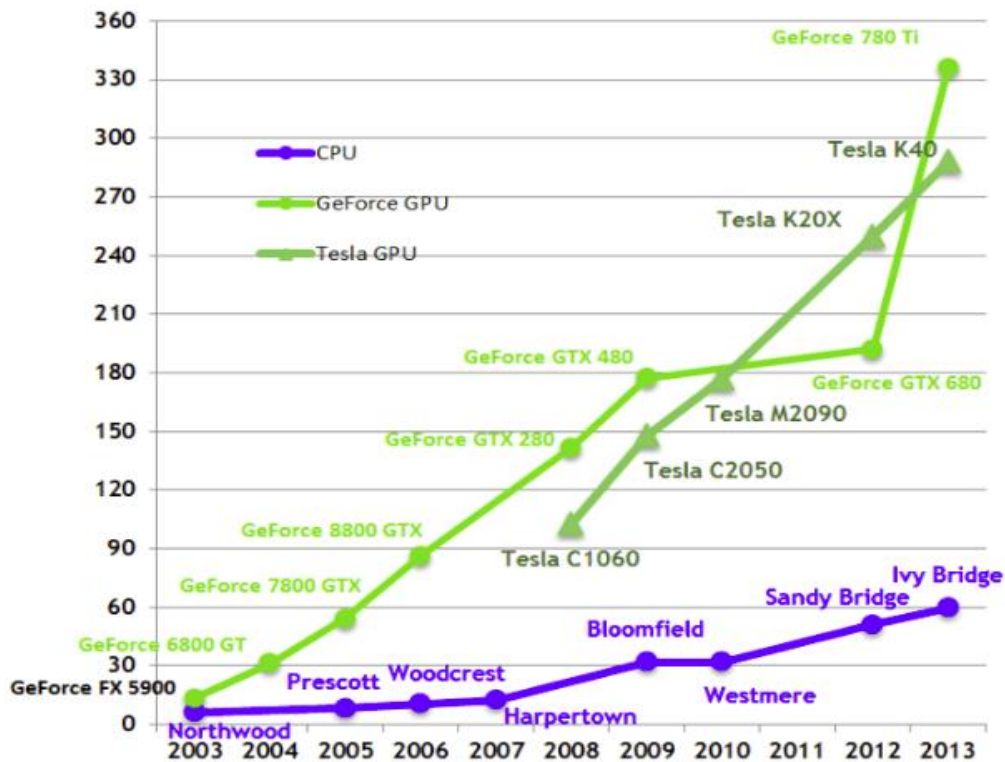
Όπως φαίνεται και από τα παραπάνω, η έννοια των υπολογιστικών πόρων είναι ιδιαίτερα σημαντική για την εύρυθμη λειτουργία μιας εφαρμογής που να υποστηρίζει ένα σύστημα αυτόματης ταξινόμησης. Ειδικότερα, ο εξυπηρετητής (server) θα πρέπει να πληροί:

- Να παρέχει τη δυνατότητα εκπαίδευσης σε κάρτα γραφικών (Graphical Process Unit - GPU). Η εκπαίδευση Νευρωνικών Δικτύων σε κάρτες γραφικών ελαχιστοποιεί τον απαιτούμενο χρόνο επεξεργασίας σε σχέση με την ταχύτητα των επεξεργαστών (Central Processing Unit - CPU), καθώς επιτρέπουν την παράλληλη επεξεργασία δεδομένων ιδιαίτερα υψηλής πολυπλοκότητας. Η διαφορά αυτή φαίνεται έντονα στο Σχήμα 30. Όπως φαίνεται και στο σχήμα, οι κάρτες γραφικών παρέχουν πολλαπλάσιο εύρος (bandwidth - GB/s) σε σχέση με τους επεξεργαστές. Ενδεικτικά αναφέρεται πως το 2013 η διαφορά ήταν εξαπλάσια του εύρους των αντίστοιχων επεξεργαστών και συνεχίζει να αυξάνεται σημαντικά μέχρι και σήμερα [30].
- Εκτός από τις κάρτες γραφικών, το σύστημα αυτό θα πρέπει να παρέχει επαρκής υπολογιστικούς πόρους (μνήμη RAM, επεξεργαστής, αποθηκευτικός χώρος) καθώς είναι πολλές οι λειτουργίες (όπως πχ η προεπεξεργασία εγγράφων που για λόγους χρονοπρογραμματισμού είναι προτιμότερο να μην εκτελούνται στην κάρτα γραφικών).
- Καθώς το σύστημα θα πρέπει να υποστηρίζει ταυτόχρονα πολλούς χρήστες, είναι απαραίτητος ο εικονικός διαχωρισμός του server σε επιμέρους servers, ώστε κάθε χρήστης να έχει αποκλειστική πρόσβαση στους υπολογιστικούς πόρους που έχει επιλέξει. Αυτό επιτυγχάνεται μέσα από τη χρήση Εικονικών Μηχανών (Virtual Machines) αι του αντίστοιχου λογισμικού (VMware).

### Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

- Παράλληλα, τα δεδομένα πρέπει να είναι οργανωμένα σε μία κατάλληλη Βάση Δεδομένων (ΒΔ) όπου θα περιγράφονται με ορθό τρόπο οι κατάλληλες σχέσεις και συσχετίσεις.
- Να χρησιμοποιεί τις κατάλληλες βιβλιοθήκες (Frameworks) που παρέχουν την Προγραμματιστική Διεπαφή Εφαρμογής (Application Programming Interface - API) ώστε να είναι εφικτή η επικοινωνία με τις κάρτες γραφικών και τους υπόλοιπους προγραμματιστικούς πόρους. Ταυτόχρονα, οι βιβλιοθήκες αυτές θα πρέπει να είναι πλήρως επεκτάσιμες και αναβαθμίσιμες, καθώς διαρκώς οι αλγόριθμοι που περιέχονται, επεκτείνονται και βελτιστοποιούνται τόσο από άποψη χρόνου όσο και απαιτήσεων σε υπολογιστικούς πόρους. Τέτοιες δημοφιλείς βιβλιοθήκες που συναντώνται στην επιστημονική κοινότητα, δίνονται στο Σχήμα 31.

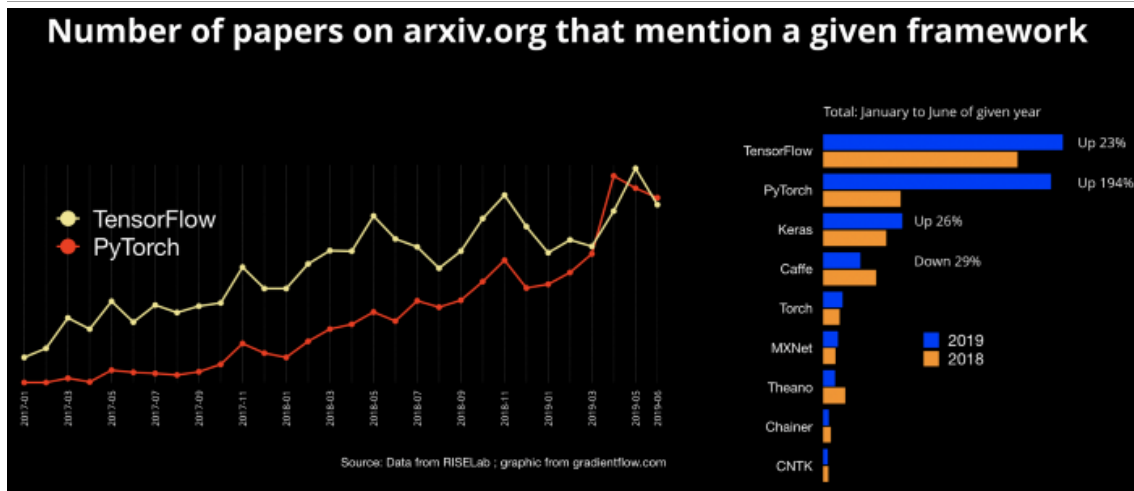
Theoretical GB/s



Σχήμα 30. Σύγκριση εύρους (bandwidth) καρτών γραφικών (GPU) με επεξεργαστές (CPU) για την εκπαίδευση Νευρωνικών Δικτύων [29]

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων





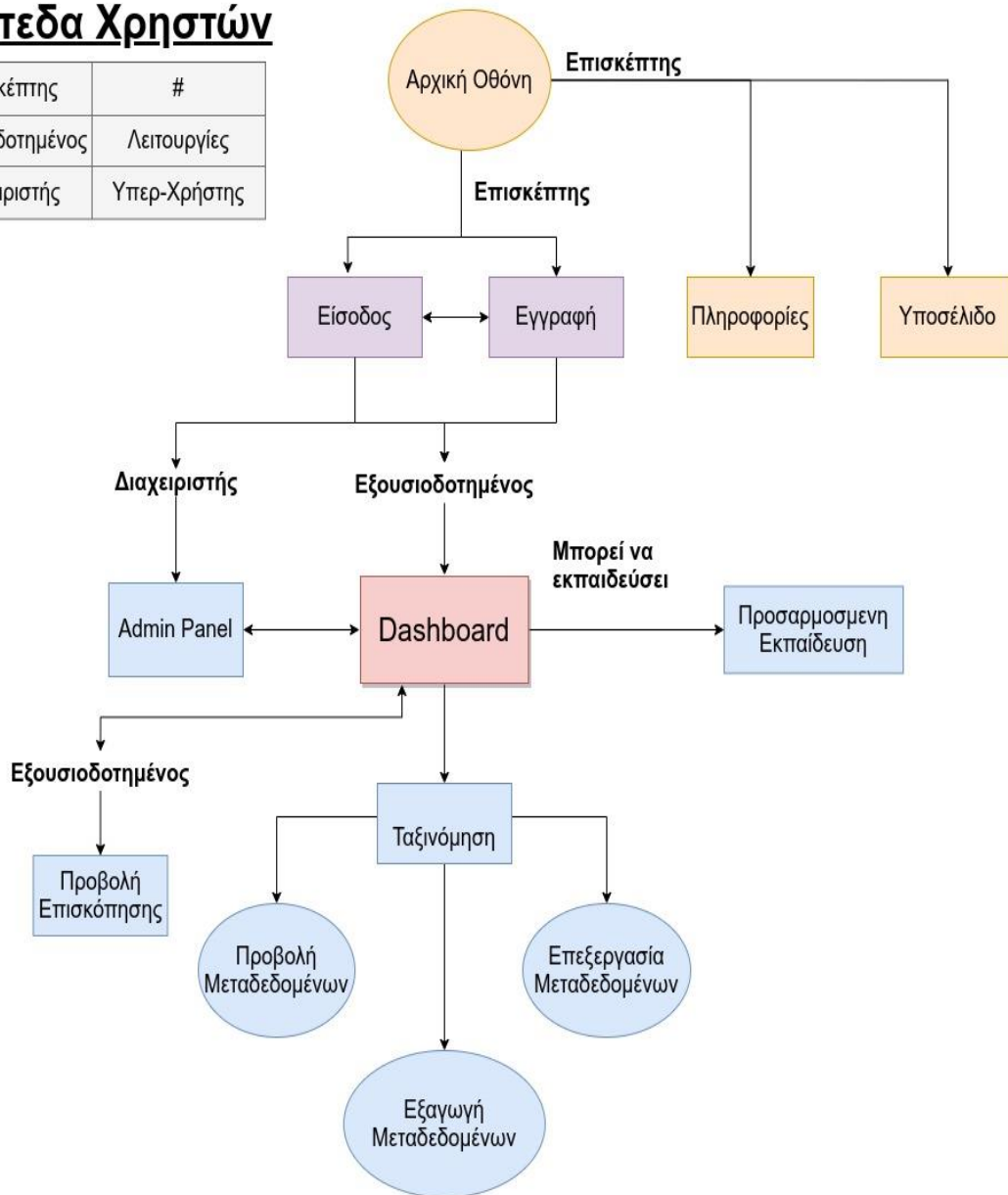
Σχήμα 31. Σύγκριση εύρους (bandwidth) καρτών γραφικών (GPU) με επεξεργαστές (CPU) για την εκπαίδευση Νευρωνικών Δικτύων [31]

### 5.3. ΑΝΑΛΥΣΗ ΑΠΑΙΤΗΣΕΩΝ ΛΕΙΤΟΥΡΓΙΚΟΤΗΤΑΣ

Μία web-based εφαρμογή που υποστηρίζει ένα σύστημα Αυτόματης Ταξινόμησης Εγγράφων, θα πρέπει να παρέχει στο χρήστη και κάποιες επιθυμητές λειτουργίες, πέραν των τεχνολογικών προδιαγραφών που πρέπει να πληρούνται και αναλύθηκαν προηγουμένως. Στόχος των λειτουργιών αυτών είναι να παρέχουν όλες τις διευκολύνσεις που απαιτεί ένας απλός χρήστης και ταυτόχρονα να ικανοποιούν και τις απαιτήσεις των έμπειρων χρηστών. Η λειτουργικότητα αυτή περιγράφεται στο Σχήμα 23 και αναλύεται στη συνέχεια.

## Επίπεδα Χρηστών

Επισκέπτης	#
Εξουσιοδοτημένος	Λειτουργίες
Διαχειριστής	Υπερ-Χρήστης



Σχήμα 32. Διάγραμμα Ροής προτεινόμενης λειτουργικότητας εφαρμογής Αυτόματης Ταξινόμησης Εγγράφων.

- Σύμφωνα με το παραπάνω διάγραμμα, στην εφαρμογή θα πρέπει να διακρίνονται τρία επίπεδα χρηστών, όπου κάθε ανώτερο επίπεδο θα έχει τις δυνατότητες των κατώτερων, μαζί με κάποιες επιπλέον. Τα προτεινόμενα επίπεδα και αναλύονται παρακάτω:
  - **Guest (επισκέπτης):** Είναι το κατώτερο επίπεδο χρηστών, όπου δε δίνονται ιδιαίτερα δικαιώματα, παρά μόνο πρόσβαση σε γενικές σελίδες περιεχομένου της πλατφόρμας (π.χ. πληροφορίες για την εφαρμογή (**Πληροφορίες**), περαιτέρω σύνδεσμοι (**Υποσέλιδο**)). Ταυτόχρονα, θα υπάρχει περιοχή όπου ο Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

χρήστης θα μπορεί να εισέλθει (**Είσοδος**) στο σύστημα με τον κωδικό του και το καταχωρημένο όνομά του. Τέλος, θα εξεταστεί στη συνέχεια του έργου αν θα δοθεί και ελεύθερη δυνατότητα εγγραφής χρηστών (**Εγγραφή**)

- **Authorized/Auth User (εξουσιοδοτημένος χρήστης):** Το επίπεδο αυτών των χρηστών θα διαθέτει επιπλέον δυνατότητες, μετά την είσοδο στην εφαρμογή. Σε γενικές γραμμές, το επίπεδο αυτό θα πρέπει να έχει δικαιώματα ανάγνωσης και όχι επεξεργασίας. Θα διαθέτει ένα πάνελ επισκόπησης, όπου θα μπορεί να προσπελάσει ή να αναζητήσει με χρήση βασικών φίλτρων αναζήτησης τα ψηφιοποιημένα έγγραφα ανά κατηγορία, μέσα από μενού δένδρικής μορφής (Dashboard - tree view) καθώς και να ανεβάσει στην πλατφόρμα τα έγγραφα που ο ίδιος επιθυμεί. Στη συνέχεια, θα μπορεί να προβάλει και να ξεφυλλίσει ολόκληρο το βιβλίο/έγγραφο (**Προβολή επισκόπησης**), οπτικοποιώντας παράλληλα το ξεφύλλισμα του εγγράφου με στόχο να δίνεται μία ρεαλιστική αίσθηση. Στη λειτουργία αυτή, θα πρέπει δοθεί και η δυνατότητα οπτικοποίησης μεταδεδομένων (αναλύονται παρακάτω) μετά από επιλογή του χρήστη. Στην περίπτωση αυτή, θα προβάλλονται πληροφορίες που αντιστοιχούν στο έγγραφο αυτό με τρόπο έντονα διαδραστικό. Εκτός από τη λειτουργία προβολής ολόκληρου του εγγράφου, ο χρήστης θα μπορεί να κάνει χρήση της λειτουργίας της **Ταξινόμησης** επιλέγοντας κάποια από τα παρεχόμενα εκπαιδευμένα μοντέλα. Θα μπορεί να περιηγηθεί μέσα από μενού με μικρογραφίες σελίδων (thumbnails) και να μία σελίδα από ένα επιλεγμένο έγγραφο (**Προβολή Μεταδεδομένων**), να δει τη σελίδα αυτή και με παρόμοιο τρόπο όπως αυτός της λειτουργίας Προβολής Προεπισκόπησης να προβάλει σχετικές πληροφορίες μέσα από τα μεταδεδομένα της σελίδας. Επιπλέον, μέσα από τη λειτουργία της **Επεξεργασίας Μεταδεδομένων** θα μπορεί να επεξεργαστεί το αποτέλεσμα αναγνώρισης. Τέλος, η κατηγορία χρηστών αυτών θα μπορεί να εξαγει (Εξαγωγή Μεταδεδομένων) τα μεταδεδομένα που επιθυμεί σε διάφορα επίπεδα, δηλαδή ανά κατηγορία εγγράφων (Category) (πχ. όλες τις επιστολές), ανά έγγραφο (Book) ή σελίδα (Image), σε διάφορες μορφές. Οι υποστηριζόμενες μορφές αυτές θα πρέπει να είναι JSON, XLS, CSV και ταυτόχρονα να δίνεται η δυνατότητα λήψης συμπιεσμένου αρχείου μορφής (ZIP ή RAR).
- **Administrator/Admin (διαχειριστής):** Πρόκειται για το ανώτατο επίπεδο χρήστη που υπάρχει στην εφαρμογή. Διαθέτει πλήρη πρόσβαση σε όλες τις λειτουργίες της πλατφόρμας και ταυτόχρονα διαθέτει πάνελ επισκόπησης (**Admin Panel**), παρακολούθησης και ελέγχου των λειτουργιών, διαδικασιών και δράσεων των χρηστών.

Όπως αναφέρεται και παραπάνω, σημαντικό ρόλο στην παραπάνω εφαρμογή αποτελούν τα μεταδεδομένα κάθε εικόνας. Η δυνατότητα πρόσβασης ανάλογα με το επίπεδο κάθε χρήστη με σκοπό την προβολή ή/και επεξεργασία ή/και επικύρωση, περιγράφηκε προηγουμένως. Από το διάγραμμα της προτεινόμενης αρχιτεκτονικής, διακρίνονται διάφορα είδη μεταδεδομένων για κάθε εικόνα εγγράφου (σελίδα). Αυτά είναι τα **Καθολικά Μεταδεδομένα (Global Metadata)** και αναλύονται στη συνέχεια. Αφορούν σε πληροφορία σχετική με τη σελίδα εγγράφου που σχετίζεται με ολόκληρη την εικόνα και συνήθως είναι πεδία με τιμές μέσα από ένα σύνολο δυνατών επιλογών. Αυτά είναι:

- Όνομα Αρχείου
- Θεματολογική / Ειδολογική Ταξινόμηση, δηλαδή η κατηγορία στην οποία ανήκει το έγγραφο
- Προσαρμοσμένα πεδία έπειτα από επιλογή του χρήστη. Παραδείγματα μπορεί να είναι:
  - Ημερομηνία
  - Υπογραφή
  - Σφραγίδα

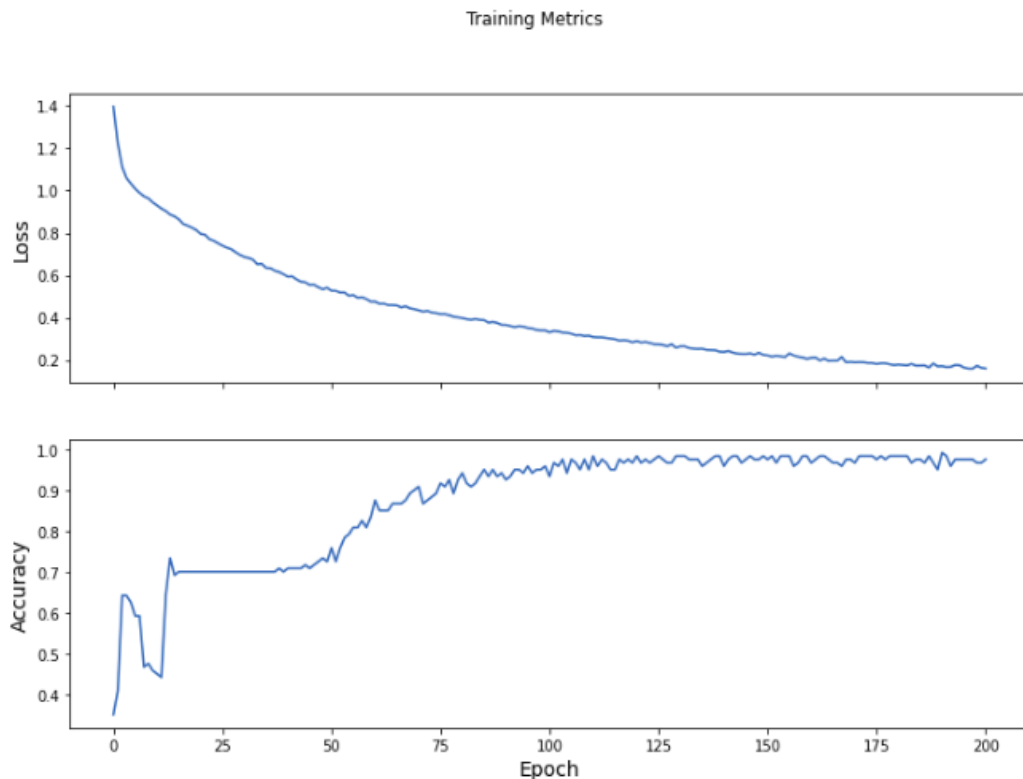
Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

- Αριθμός Σελίδας
- Φορέας που αναφέρεται
- Κωδικός Εγγράφου
- Αποστολέας
- Θέμα Εγγράφου

Στο παραπάνω διάγραμμα διακρίνεται και η σημαντικότερη λειτουργία που πρέπει να υπάρχει σε μία τέτοια εφαρμογή, αυτή της δυνατότητας **Προσαρμοσμένης Εκπαίδευσης** από το χρήστη. Σύμφωνα με αυτή, θα πρέπει να παρέχονται στο χρήστη όλα τα βήματα που απαιτούνται για την επίτευξη του επιθυμητού αποτελέσματος, δηλαδή την εκπαίδευση ενός Νευρωνικού Δικτύου για την Αυτόματη Ταξινόμηση Εγγράφων. Τα βήματα αυτά θα πρέπει κατά το ελάχιστο να περιλαμβάνουν:

- Ορισμό των κλάσεων που ενδιαφέρουν το χρήστη.
- Επιλογή μοντέλου Νευρωνικού Δικτύου που θα εκπαιδευτεί.
- Ορισμό ελάχιστου αριθμού εικόνων εγγράφων ανά κατηγορία.
- Ορισμός τιμών για τις υπερ-παραμέτρους του συστήματος.
- Δυνατότητα εισαγωγής αρχείων ανά κατηγορία στην εφαρμογή (upload).

Παράλληλα με το παραπάνω υποβοηθούμενο σύστημα εκπαίδευσης, ο χρήστης θα πρέπει να είναι σε θέση να παρακολουθεί την εξέλιξη της διαδικασίας εκπαίδευσης μέσα από κατάλληλα διαγράμματα (πχ. καμπύλη σφάλματος ή ακρίβειας) όπως και να γνωρίζει τον εκτιμώμενο χρόνο ολοκλήρωσης. Παραδείγματα τέτοιων καμπυλών μάθησης, με χρήση της Βιβλιοθήκης TensorFlow [32], φαίνονται στη συνέχεια (Σχήμα 33).



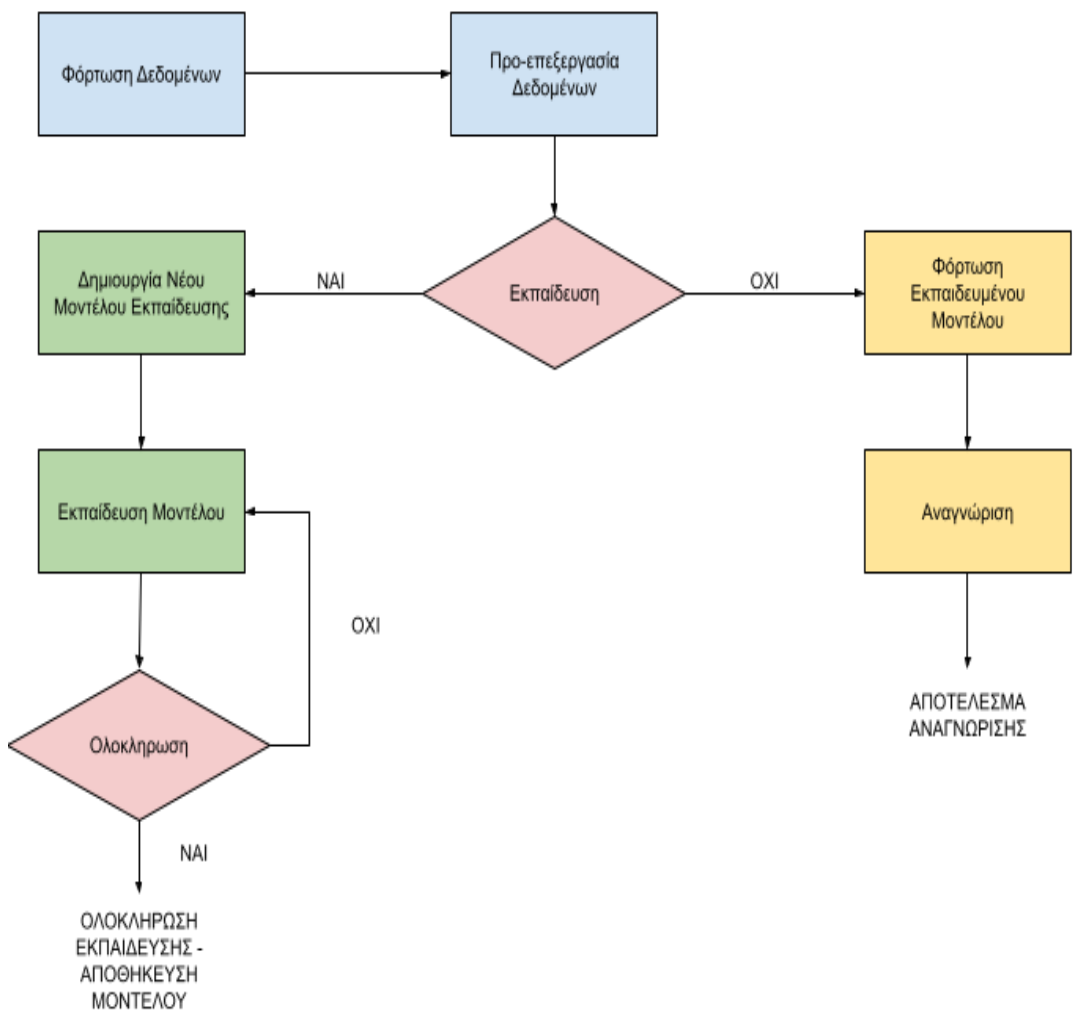
Σχήμα 33. Παραδείγματα καμπυλών μάθησης, με χρήση της Βιβλιοθήκης TensorFlow [32]

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

5.4. ΣΧΕΔΙΑΣΜΟΣ ΣΥΣΤΗΜΑΤΟΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΓΓΡΑΦΩΝ

Στα προηγούμενα Κεφάλαια παρουσιάστηκαν τόσο η τρέχουσα τεχνολογική στάθμη για την ψηφιακή ταξινόμηση εγγράφων με χρήση Νευρωνικών Δικτύων, όσο και οι σημαντικότερες τεχνικές που χρησιμοποιούνται για τη βελτιστοποίηση της απόδοσής τους. Παράλληλα, παρουσιάστηκαν οι απαιτήσεις των εφαρμογών που θα πρέπει να υποστηρίζουν τα συστήματα αυτά, με στόχο τη δημιουργία ολοκληρωμένων προϊόντων και υπηρεσιών που επιλύουν στο βέλτιστο βαθμό το συγκεκριμένο πρόβλημα.

Στο υπο-κεφάλαιο αυτό, παρουσιάζεται η δομή του συστήματος ταξινόμησης, λαμβάνοντας υπόψη όλα τα παραπάνω. Το γενικό διάγραμμα του συστήματος δίνεται στο Σχήμα 25 και αναλύεται στα επόμενα.



Σχήμα 34. Προτεινόμενη αρχιτεκτονική συστήματος Ταξινόμησης Εγγράφων

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

Σύμφωνα με το παραπάνω σχήμα, διακρίνονται τα εξής επιμέρους τμήματα, τα οποία σε μια γλώσσα προγραμματισμού (πχ Python, C, C++, Java) μπορούν να θεωρηθούν και ως ρουτίνες-συναρτήσεις:

**Φόρτωση Δεδομένων:** Η ρουτίνα αυτή θα πρέπει να είναι υπεύθυνη για τη φόρτωση των δεδομένων (πχ 1 ή περισσότερες εικόνες) από το σκληρό δίσκο του υπολογιστή στη μνήμη του (ή στην GPU).

- **Είσοδος:** Το όνομα και η διαδρομή αποθήκευσης μιας ή περισσότερων εικόνων, ή ενός φακέλου ή περισσότερων που περιέχει/περιέχουν εικόνες. Επιπλέον παρέχεται (μέσω αρχείου ή συνήθως προκύπτει από τους φακέλους αποθήκευσης) η κατηγορία (κλάση) για κάθε εικόνα που φορτώνεται.
- **Έξοδος:** Οι εικόνες αποθηκευμένες στη μνήμη του υπολογιστή, σε μορφή πινάκων (πχ βάσει προτύπου RGB) μαζί με την κατηγορία ανά εικόνα.

**Προ-επεξεργασία Δεδομένων:** Η ρουτίνα αυτή είναι υπεύθυνη για τη μετατροπή των φορτωμένων δεδομένων σε μορφή κατάλληλη ώστε να χρησιμοποιηθούν ως δεδομένα εισόδου σε ένα Νευρωνικό Δίκτυο. Περιλαμβάνει όλα τα στάδια τεχνητής επαύξησης δεδομένων, κλιμάκωσης σε μέγεθος κατάλληλο για το Νευρωνικό Δίκτυο και διαχωρισμού σε τμήματα (batches).

- **Είσοδος:** Φορτωμένες εικόνες σε μορφή Πίνακα (πχ βάσει προτύπου RGB)
- **Παράμετροι:** Επιλογή τρόπου επαύξησης δεδομένων, μεγέθους εικόνων, μεγέθους τμηματοποίησης δεδομένων.
- **Έξοδος:** Σύνολο εικόνων που περιλαμβάνει τόσο τις αυθεντικές εικόνες, όσο και τεχνητά-παραμορφωμένα αντίγραφα που έχουν προκύψει μέσω της τεχνητής επαύξησης.

**Δημιουργία Νέου Μοντέλου Εκπαίδευσης:** Στη ρουτίνα αυτή ορίζεται η αρχιτεκτονική του μοντέλου Νευρωνικού Δικτύου που θα χρησιμοποιηθεί. Λαμβάνεται υπόψη πιθανή αρχικοποίηση βαρών, και διάφορες υπερ-παραμέτροι που υπάρχουν, καθώς και το πλήθος των κατηγοριών που αντικατοπτρίζουν τις δυνατές εξόδους του συστήματος. Η ρουτίνα αυτή χρησιμοποιείται εφόσον ικανοποιείται η συνθήκη εκπαίδευσης.

- **Είσοδος:** Σύνολο εικόνων προς εκπαίδευση, χωρισμένες σε τμήματα (batches)
- **Παράμετροι:** Επιλογή αρχιτεκτονικής Νευρωνικού Δικτύου (βάθος δικτύου, πλήθος νευρώνων, συναρτήσεις ενεργοποίησης), επιλογή αρχικών τιμών για τα βάρη του δικτύου, επιλογή ποσοστού απενεργοποίησης νευρώνων (dropout), ορισμός δυνατών επιλογών εξόδου.
- **Έξοδος:** Πλήρως ορισμένο μοντέλο Νευρωνικού Δικτύου έτοιμου προς εκπαίδευση.

**Εκπαίδευση Μοντέλου:** Η ρουτίνα αυτή χρησιμοποιείται εφόσον ικανοποιείται η συνθήκη εκπαίδευσης. Πρόκειται για τη σημαντικότερη συνάρτηση της προτεινόμενης αρχιτεκτονικής. Σε αυτή πραγματοποιείται η διαδικασία της εκπαίδευσης. Το σύστημα τροφοδοτείται διαρκώς με εικόνες μέχρις ότου ικανοποιούνται οι επιλεγμένες συνθήκες τερματισμού. Επιπλέον, η συνάρτηση αυτή πρέπει να παρέχει/ενημερώνει και άλλες λειτουργίες, όπως αυτή της παρακολούθησης (monitoring) και της ικανοποίησης συνθήκης ολοκλήρωσης μέσω χρήσης δεδομένων επαλήθευσης. Τέλος, στη ρουτίνα αυτή καταγράφονται όλες οι μετρικές (πχ ακρίβεια, σφάλμα) που βοηθούν την διαδικασία διαρκούς ενημέρωσης των παραμέτρων (βαρών) του δικτύου, μέσω του αλγόριθμου ανάστροφης μετάδοσης λάθους (back propagation algorithm). Στο σημείο αυτό πρέπει να επιλεγθούν κατάλληλα διάφορες υπερ-παραμέτροι όπως ο ρυθμός εκπαίδευσης/μάθησης, οι συνολικές επαναλήψεις ή αλλιώς εποχές εκπαίδευσης, το μέγιστο αποδεκτό σφάλμα, κλπ.

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

- **Είσοδος:** Σύνολο εικόνων που περιλαμβάνει τόσο τις αυθεντικές εικόνες, όσο και τεχνητά-παραμορφωμένα αντίγραφα που έχουν προκύψει μέσω της τεχνικής επαύξησης, καθώς και ένα πλήρως ορισμένο μοντέλο Νευρωνικού Δικτύου έτοιμο προς εκπαίδευση.
- **Παράμετροι:** Ρυθμός μάθησης, εποχές εκπαίδευσης, μέγιστο αποδεκτό σφάλμα
- **Παράλληλες λειτουργίες:** Ενημέρωση καμπυλών σφάλματος, ακρίβειας, στατιστικά στοιχεία για βάρη νευρώνων. Επαλήθευση αποτελεσμάτων σε δεδομένα επαλήθευσης.
- **Έξοδος:** Εκπαιδευμένο μοντέλο που πληροί τη συνθήκη Ολοκλήρωσης ή παρήλθαν οι συνολικές επαναλήψεις ή εποχές εκπαίδευσης.

**Φόρτωση Εκπαιδευμένου Μοντέλου:** Η ρουτίνα αυτή εκτελείται στην περίπτωση που η συνθήκη εκπαίδευσης δεν ικανοποιείται, δηλαδή το σύστημα έχει εκπαιδευτεί ήδη και καλείται να χρησιμοποιηθεί ώστε να αναγνωρίσει ένα νέο σύνολο εικόνων.

- **Είσοδος:** Ένα αποθηκευμένο αρχείο στο σκληρό δίσκο (διαδρομή αποθήκευσης) που έχει αποθηκευμένη όλη την πληροφορία που προέκυψε κατά τη διαδικασία της εκπαίδευσης τελικά βάρη νευρώνων δικτύου και αρχιτεκτονική του μοντέλου.
- **Έξοδος:** Το αποθηκευμένο μοντέλο, φορτωμένο στη μνήμη του υπολογιστή (ή στη GPU) και έτοιμο προς χρήση.

**Αναγνώριση:** Η ρουτίνα αυτή εκτελείται στην περίπτωση που η συνθήκη εκπαίδευσης δεν ικανοποιείται, δηλαδή το σύστημα έχει εκπαιδευτεί ήδη και καλείται να χρησιμοποιηθεί ώστε να αναγνωρίσει ένα νέο σύνολο εικόνων. Το φορτωμένο μοντέλο δέχεται εικόνες που έχουν προεπεξεργαστεί, οι οποίες και τροφοδοτούν το δίκτυο. Για κάθε μία από τις εικόνες προκύπτει μία έξοδος που είναι και το αποτέλεσμα ταξινόμησης/κατηγοριοποίησης της εικόνας εισόδου. Σημειώνεται ότι κατά την προ-επεξεργασία δε χρησιμοποιείται η τεχνική της επαύξησης/παραμόρφωσης εικόνων, παρά μόνο η κλιμάκωση της εικόνας σε μέγεθος κατάλληλο για την είσοδο του Νευρωνικού Δικτύου.

- **Είσοδος:** Εικόνες Εγγράφων που έχουν τις κατάλληλες διαστάσεις (έπειτα από κλιμάκωση) για την είσοδο του Δικτύου.
- **Έξοδος:** Αποτέλεσμα ταξινόμησης/κατηγοριοποίησης για κάθε εικόνα εισόδου.

**Συνθήκη Ολοκλήρωσης:** Κατά τη συνθήκη αυτή και στη διάρκεια εκπαίδευσης, ελέγχεται έπειτα από κάθε κύκλο εκπαίδευσης η αποτελεσματικότητα του μοντέλου σε ένα σύνολο δεδομένων επαλήθευσης (φάση επαλήθευσης - validation phase). Κατά τη φάση αυτή, δεν ενημερώνονται τα βάρη του μοντέλου και πραγματοποιείται ένας “μίνι” έλεγχος, θεωρώντας το μοντέλο ήδη εκπαιδευμένο. Αν η μετρική ακρίβειας είναι η μέγιστη σε σχέση με τα προηγούμενα “στιγμιότυπα” του μοντέλου, τότε το μοντέλο της τρέχουσας φάσης θεωρείται προσωρινά βέλτιστο. Η συνθήκη ικανοποιείται όταν συμβαίνει ένα ή και τα δύο από τα παρακάτω: α) Το σφάλμα ή η ακρίβεια στα δεδομένα επαλήθευσης βρίσκονται εντός επιθυμητών ορίων και β) όταν ολοκληρώνεται η διαδικασία εκπαίδευσης. Μετά την ολοκλήρωση, αποθηκεύεται το μοντέλο που θεωρείται βέλτιστο, δηλαδή, το μοντέλο στο οποίο μεγιστοποιήθηκε η μετρική αποτελεσματικότητας (F-Measure).

- **Είσοδος:** Τρέχον μοντέλο που προκύπτει έπειτα από έναν κύκλο εκπαίδευσης
- **Έξοδος:** Εφόσον πληρούνται συνθήκες ολοκλήρωσης αποθηκεύεται το βέλτιστο μοντέλο. Αν όχι, ελέγχεται αν το μοντέλο είναι βέλτιστο και στην περίπτωση αυτή θεωρείται αυτό τρέχον βέλτιστο.

**Συνθήκη Εκπαίδευσης:** Απλή συνθήκη που ελέγχει αν η πρόκειται να ακολουθήσει εκπαίδευση ενός μοντέλου ή αν πρόκειται να χρησιμοποιηθεί ένα ήδη εκπαιδευμένο μοντέλο για αναγνώριση. Στη δεύτερη περίπτωση, δεν πραγματοποιείται επαύξηση δεδομένων, παρά μόνο αλλαγή μεγέθους της εικόνας ώστε αυτό να είναι κατάλληλου μεγέθους για είσοδο στο εκπαιδευμένο μοντέλο.

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

Τα παραπάνω αποτελούν τις βασικές συναρτήσεις/ρουτίνες που πρέπει να χρησιμοποιηθούν από κάποιον με στόχο την υλοποίηση ενός συστήματος ταξινόμησης εικόνων εγγράφων. Ακολουθώντας το παραπάνω σχεδιάγραμμα (Σχήμα 25), μπορεί εύκολα κανείς να δημιουργήσει ένα σύστημα που σε βάσεις δεδομένων που συναντώνται στη βιβλιογραφία, να πετύχει μετρικές αποτελεσματικότητας της τάξης του 90%. Σε περίπτωση που είναι επιθυμητή η χρήση άλλης βάσης δεδομένων με έγγραφα, η ακρίβεια μπορεί να μεταβληθεί θετικά ή αρνητικά και ανάλογα με την πολυπλοκότητα και τη δυσκολία επίλυσης του προβλήματος.

Ένα τόσο απλό σύστημα, δηλαδή, μπορεί και να επιλύσει σε ικανοποιητικό βαθμό ένα οποιοδήποτε πρόβλημα ταξινόμησης εγγράφων, εξοικονομώντας παράλληλα σημαντικούς πόρους (ανθρώπινο δυναμικό, κόστος, χρόνος) και να εντάξει μία χρήσιμη λειτουργία σε μία ευρύτερη εφαρμογή που αποτελείται από ένα σύνολο υπηρεσιών/εφαρμογών (πχ. σύνολο υπηρεσιών/εφαρμογών επεξεργασίας εγγράφων).



## 6. ΣΥΝΟΨΗ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην εργασία αυτή μελετήθηκε το πρόβλημα της Ψηφιακής Ταξινόμησης Εγγράφων, μέσα από ένα πλήθος προκαθορισμένων κλάσεων. Η λύση του θα βοηθήσει σημαντικά διάφορα ζητήματα της καθημερινότητας, όπως η γραφειοκρατία. Επιπλέον, η επίλυσή του θα προσφέρει πρόσβαση σε εκατομμύρια έγγραφα που παραμένουν ανεκμετάλλευτα είτε σε φυσική μορφή είτε αποθηκευμένα σε διάφορους ξεχασμένους φακέλους υπολογιστών. Η πρόσβαση αυτή θα δώσει την εναρκτήρια ώθηση ώστε να γίνει γνωστή η πλούσια πληροφορία που μπορεί να περιέχουν τέτοιου είδους έγγραφα.

Τις προηγούμενες δεκαετίες, η προσπάθεια για την Ψηφιακή Ταξινόμηση Εγγράφων γινόταν κάνοντας χρήση “παραδοσιακών” τεχνικών Επεξεργασίας Εικόνας. Τα αποτελέσματα απείχαν πολύ από το να θεωρηθούν ικανοποιητικά, καθώς η ποικιλομορφία και η έντονη πολυπλοκότητα αμέτρητων κατηγοριών εγγράφων δεν μπορούσε να οδηγήσει στην πρόταση μιας επιστημονικής μεθόδου καθολικής χρήσης.

Την τελευταία δεκαετία και με την είσοδο της Τεχνητής Νοημοσύνης και των Νευρωνικών Δικτύων, γίνονται σημαντικές προσπάθειες προς την επίλυση του παραπάνω προβλήματος. Πιο συγκεκριμένα, τα Συνελικτικά Νευρωνικά Δίκτυα που προτείνονται στην επιστημονική κοινότητα βρίσκουν κατάλληλη εφαρμογή και στην κατηγοριοποίηση εικόνων εγγράφων. Η συνεχόμενη εξέλιξή τους δείχνει να συνάδει και με την αύξηση της αποτελεσματικότητάς τους στην ταξινόμηση εικόνων εγγράφων μέσα από βάσεις δεδομένων μεγάλης κλίμακας με χιλιάδες εικόνες και δεκάδες κλάσεις.

Όπως παρουσιάστηκε και στα προηγούμενα Κεφάλαια, δεν αρκεί μόνο η επιλογή μιας αρχιτεκτονικής Νευρωνικού Δικτύου για την επίτευξη του στόχου. Για το λόγο αυτό, η εργασία αυτή επικεντρώθηκε στην απλή περιγραφή τους και ταυτόχρονα στον τρόπο χρήσης και συνδυασμού τους με άλλες τεχνικές, καθώς και στην ενσωμάτωσή τους σε προτεινόμενες μεθόδους που συναντώνται στη βιβλιογραφία και παρουσιάστηκαν στο Κεφάλαιο 2.

Παράλληλα λοιπόν με την επιλογή μιας αρχιτεκτονικής Νευρωνικού Δικτύου, είναι πολύ σημαντικό να επιλεγθούν κατάλληλες τιμές για διάφορες υπερ-παραμέτρους. Μερικές από τις σημαντικότερες εξ αυτών είναι ο ρυθμός μάθησης ενός δικτύου και οι κύκλοι εκπαίδευσης. Είναι πολύ σημαντικό το πόσο γρήγορα θα εκπαιδευτεί ένα δίκτυο και για πόσο πολύ. Μία φυσική αναλογία είναι ο τρόπος που ένα μικρό παιδί μαθαίνει να αναγνωρίζει αυτοκίνητα. Στην αρχή, δε γνωρίζει καν τι είναι αυτοκίνητο. Στη συνέχεια, σιγά-σιγά και με τη βοήθεια των γονιών, μαθαίνει τι είναι και ορισμένες μάρκες. Αν σταματήσει στο σημείο αυτό η εκπαίδευση, σε καμία περίπτωση δε θα είναι σε θέση το παιδί να αναγνωρίζει αρκετές μάρκες, καθώς απαιτείται και άλλος χρόνος εκπαίδευσης που θα το βοηθήσουν να διακρίνει παρόμοιες μάρκες και σήματα, διαφορετικά μοντέλα της ίδιας μάρκας, διαφορετικά μεγέθη, σχήματα, χρώματα αυτοκινήτων, κλπ και όχι απλά να “αποστηθίζει” έναν μικρό αριθμό από μάρκες. Συνεπώς, η επιλογή του ρυθμού μάθησης (το πόσο γρήγορα) καθώς και ο χρόνος της συνολικής εκπαίδευσης (για πόσο καιρό) παίζουν ιδιαίτερα σημαντικό ρόλο.

Άλλη σημαντική παράμετρος είναι το μέγεθος της εικόνας. Πολύ μικρές εικόνες μπορεί να οδηγήσουν σε απώλεια πληροφορίας ιδιαίτερα σημαντικής για τη διάκριση και ταξινόμηση. Από την άλλη, μεγάλο μέγεθος εικόνας δεν μπορεί να υποστηριχθεί αποτελεσματικά από τη διαθέσιμη μνήμη των υπολογιστών, καθώς και περιορίζει τις δυνατότητες παράλληλης επεξεργασίας εικόνων σε κάθε επανάληψη εκπαίδευσης. Εκτός από το μέγεθος της εικόνας εισόδου, ένα πάρα πολύ σημαντικό ρόλο παίζει και το πλήθος των εικόνων εκπαίδευσης. Καθώς οι βάσεις δεδομένων αποτελούνται συχνά από λίγες εικόνες (μερικές χιλιάδες), είναι πολύ σημαντικό αυτές να ενισχυθούν προσθέτοντας νέες εικόνες. Ένα τέχνασμα που εξυπηρετεί κάτι τέτοιο, είναι αυτό της επαύξησης δεδομένων. Στη βιβλιογραφία (Κεφάλαια 2 και 3) προτείνεται ένας μεγάλος αριθμός μετασχηματισμών της εικόνας (πχ. περιστροφή, μετατόπιση) ώστε να δημιουργηθούν παραμορφωμένα αντίγραφα. Έχει αποδειχθεί (Κεφάλαιο

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

4) ότι οι μετασχηματισμοί αυτοί εμπλουτίζουν την ποικιλομορφία μιας βάσης εικόνων και βοηθούν το ίδιο το δίκτυο να μάθει καλύτερα και πιο αποτελεσματικά μέσα από αυτές, καθώς δέχεται πάντα ως είσοδο μία διαφορετική εικόνα και ποτέ την ίδια.

Ο τρόπος έναρξης της διαδικασίας εκπαίδευσης είναι ίσως ο πιο σημαντικός παράγοντας ενίσχυσης της αποτελεσματικότητας ενός δικτύου, καθώς και μείωσης του χρόνου εκπαίδευσής του. Αποδεικνύεται στη βιβλιογραφία (βλ. Πίνακα 5 Κεφάλαιο 4) ότι η αρχικοποίηση των βαρών των νευρώνων ενός δικτύου δεν οδηγεί σε καμία περίπτωση σε ικανοποιητικά αποτελέσματα και ιδιαίτερα για βάσεις δεδομένων μικρής κλίμακας. Απαιτείται αρχικοποίηση με ένα προ-εκπαιδευμένο ίδιο δίκτυο σε κάποια άλλη βάση δεδομένων με εικόνες είτε διαφορετικής φύσης (πχ. ImageNet) από τα έγγραφα, είτε, στην ιδανική περίπτωση, με βάσεις εικόνων εγγράφων. Για παράδειγμα, στη βάση μικρής κλίμακας Tobacco3482, με προ-εκπαίδευση στην ImageNet η ακρίβεια ταξινόμησης αυξάνεται κατά μέσο όρο περίπου κατά 8% σε σχέση με την περίπτωση της τυχαίας αρχικοποίησης. Στην ιδανική περίπτωση, η προ-εκπαίδευση σε βάση δεδομένων εγγράφων, δίνει αντίστοιχη αύξηση κατά μέσο όρο άνω του 25% σε σχέση με την τυχαία αρχικοποίηση. Για βάσεις δεδομένων μεγάλης κλίμακας (RVL-CDIP), η προ-εκπαίδευση στην ImageNet δίνει αύξηση ακρίβειας κατά μέσο όρο 1%.

Μία άλλη τεχνική που εφαρμόζεται με επιτυχία πλέον των προηγούμενων είναι αυτή της τμηματικής μάθησης. Ένα Νευρωνικό Δίκτυο μπορεί να μάθει την πληροφορία που απαιτείται για να μπορεί να διακρίνει σωστά μέσα από τμήματα της εικόνας και όχι από ολόκληρη την εικόνα ή στην ιδανική περίπτωση, να συνδυάσει και τους δύο τρόπους. Οι κορυφαίες μεθοδολογίες συνδυάζουν είτε άμεσα (διαχωρισμός του εγγράφου σε άνω, κάτω, αριστερό και δεξί μέρος [6], [11]) είτε έμμεσα (ανάλυση δομής εγγράφου [14]) χωρική πληροφορία με χαρακτηριστικά της εικόνας για να πετύχουν βέλτιστα αποτελέσματα.

Συνοψίζοντας, βάσει και των μεθοδολογιών που αναλύθηκαν και στα προηγούμενα κεφάλαια, παρατηρείται ότι το πρόβλημα της Ταξινόμησης Εγγράφων έχει επιλυθεί ως ένα βαθμό, αλλά όχι ικανοποιητικό, καθώς έχουν προταθεί μεθοδολογίες που επιτυγχάνουν μετρικές ακρίβειας μεγαλύτερες από 92% και πλέον πλησιάζουν το 95%. Διαπιστώνεται ότι ένα Νευρωνικό Δίκτυο δεν είναι από μόνο του ο τρόπος επίλυσης, αλλά σίγουρα είναι η ραχοκοκαλιά ευρύτερων συστημάτων που συνδυάζουν διάφορες τεχνικές και μεθοδολογίες. Χρειάζεται να μελετηθούν ζητήματα που σχετίζονται τόσο με την είσοδο των δικτύων αυτών, καθώς και πως τα χαρακτηριστικά της εικόνας που μαθαίνονται μέσω της εκπαίδευσης ενός τέτοιου δικτύου, συνδυάζονται περαιτέρω με ετερογενή πληροφορία (πχ. χωρική, κειμενική) για την ακριβέστερη κατηγοριοποίηση εγγράφων. Τέλος, οι εφαρμογές που υποστηρίζουν τέτοια συστήματα θα πρέπει να συνδυάζουν τόσο την ευχρηστία και φιλικότητα προς το χρήστη, αλλά παράλληλα να παρέχουν ένα πλήρες σύνολο τεχνολογικών καινοτομιών ώστε να καλύπτουν ταυτόχρονα και τους απλούς αλλά και τους έμπειρους χρήστες.

## ΒΙΒΛΙΟΓΡΑΦΙΑ - ΑΝΑΦΟΡΕΣ

- [1] I.A Basheer, and M Hajmeer 2000. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43(1), p.3-31.
- [2] Chen, N., & Blostein, D. (2007). A survey of document image classification: problem statement, classifier architecture and performance evaluation. *International Journal of Document Analysis and Recognition (IJ DAR)*, 10(1), 1-16.
- [3] McCulloch, Warren S., and Walter Pitts. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, Dec. 1943, pp. 115–33. DOI.org (Crossref), doi:10.1007/BF02478259.
- [4] Ismail, Mohd Idris Shah, et al. "Neural Network Modeling for Prediction of Weld Bead Geometry in Laser Microwelding." *Advances in Optical Technologies*, vol. 2013, 2013, pp. 1–7. DOI.org (Crossref), doi:10.1155/2013/415837.
- [5] L. Kang, J. Kumar, P. Ye, Y. Li, & D. Doermann (2014). Convolutional Neural Networks for Document Image Classification. In 2014 22nd International Conference on Pattern Recognition (pp. 3168-3172).
- [6] Harley, A., Ufkes, A., & Derpanis, K. (2015). Evaluation of Deep Convolutional Nets for Document Image Classification and Retrieval. In *Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (pp. 991–995). IEEE Computer Society.
- [7] G. Csurka, Diane Larlus, A. Gordo, & J. Almazán (2016). What is the right way to represent document images? *ArXiv*, abs/1603.01076.
- [8] Muhammad Zeshan Afzal and Andreas Kölsch and Sheraz Ahmed and Marcus Liwicki (2017). Cutting the Error by Half: Investigation of Very Deep CNN and Advanced Training Strategies for Document Image Classification. In 14th IAPR International Conference on Document Analysis and Recognition, ICDAR 2017, Kyoto, Japan, November 9-15, 2017 (pp. 883–888).
- [9] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.
- [10] C. Tensmeyer, & T. Martinez (2018). Analysis of Convolutional Neural Networks for Document Image Classification. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (pp. 388-393).
- [11] A. Das, S. Roy, U. Bhattacharya, & S. K. Parui (2018). Document Image Classification with Intra-Domain Transfer Learning and Stacked Generalization of Deep Convolutional Neural Networks. In 2018 24th International Conference on Pattern Recognition (ICPR) (pp. 3180-3185).
- [12] M. N. Asim, M. U. G. Khan, M. I. Malik, K. Razzaque, A. Dengel, and S. Ahmed 2019. Two Stream Deep Network for Document Image Classification. In 2019 International Conference on Document Analysis and Recognition (ICDAR) (pp. 1410-1416).
- [13] Nicolas Audebert, Catherine Herold, Kuider Slimani, and Cédric Vidal. (2019). Multimodal deep networks for text and image-based document classification.
- [14] Xu, Y., Li, M., Cui, L., Huang, S., Wei, F., & Zhou, M. (2020). LayoutLM: Pre-Training of Text and Layout for Document Image Understanding. In *Proceedings of the 26th ACM SIGKDD*

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

International Conference on Knowledge Discovery & Data Mining (pp. 1192–1200). Association for Computing Machinery.

[15] Devlin, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171–4186). Association for Computational Linguistics.

[16]

<https://opencourses.auth.gr/modules/document/file.php/OCRS118/Σημειώσεις%20μαθήματος/E11%20-%20Τεχνητά%20Νευρωνικά%20Δίκτυα.pdf> , Τελευταία Προσπέλαση: 15/07/2021

[17] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.*, 15(1), 1929–1958.

[18] Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J Big Data* 6, 60 (2019). <https://doi.org/10.1186/s40537-019-0197-0>

[19] <https://medium.com/analytics-vidhya/data-augmentation-in-deep-learning-3d7a539f7a28> , Τελευταία Προσπέλαση: 15/07/2021

[20] <https://cs231n.github.io/neural-networks-2/> , Τελευταία Προσπέλαση: 15/07/2021

[21] K. He, X. Zhang, S. Ren and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1026-1034, doi: 10.1109/ICCV.2015.123.

[22] Xavier Glorot, & Yoshua Bengio (2010). Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (pp. 249–256). JMLR Workshop and Conference Proceedings.

[23] Jayant Kumar, Peng Ye and David Doermann. "Structural Similarity for Document Image Classification and Retrieval." *Pattern Recognition Letters*, November 2013.

[24] Ting K.M. (2011) Precision and Recall. In: Sammut C., Webb G.I. (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-30164-8\\_652](https://doi.org/10.1007/978-0-387-30164-8_652)

[25] <https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall> , Τελευταία Προσπέλαση: 15/07/2021

[26] <https://www.sapstore.com/solutions/53315/Document-Classification> , Τελευταία Προσπέλαση: 15/07/2021

[27] <https://www.abbyy.com/ocr-sdk/key-features/classification/> , Τελευταία Προσπέλαση: 15/07/2021

[28] <https://cloud.google.com/vision> , Τελευταία Προσπέλαση: 15/07/2021

[29] <https://medium.com/@shachishah.ce/do-we-really-need-gpu-for-deep-learning-47042c02efe2> , Τελευταία Προσπέλαση: 15/07/2021

[30] E. BUBER and B. DIRI, "Performance Analysis and CPU vs GPU Comparison for Deep Learning," 2018 6th International Conference on Control Engineering & Information Technology (CEIT), Istanbul, Turkey, 2018, pp. 1-6, doi: 10.1109/CEIT.2018.8751930.

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

[31] <https://rise.cs.berkeley.edu> , Τελευταία Προσπέλαση: 15/07/2021

[32] <https://www.tensorflow.org> , Τελευταία Προσπέλαση: 15/07/2021

### ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

ΑΓΓΛΙΚΟΣ ΟΡΟΣ	ΕΛΛΗΝΙΚΟΣ ΟΡΟΣ
Scanning	Σάρωση
A priori	Εκ των προτέρων
Activation Function	Συνάρτηση Ενεργοποίησης
Administrator	Διαχειριστής
Application Programming Interface	Προγραμματιστική Διεπαφή Εφαρμογής
Artificial Intelligence	Τεχνητή Νοημοσύνη
Artificial Neural Network	Τεχνητό Νευρωνικό Δίκτυο
Aspect Ratio	Λόγος Διαστάσεων πλευρών
Authorized User	Εξουσιοδοτημένος Χρήστης
Back Propagation Algorithm	Αλγόριθμος Ανάστροφης Μετάδοσης Λάθους
Bandwith	Εύρος
Batch Size	Μέγεθος Δέσμης Δεδομένων
Binarization	Διαδικοποίηση
Binary	Ασπρόμαυρη
Blur	Θόλωση
Browser	Πρόγραμμα Περιήγησης
Central Processing Unit	Κεντρική Μονάδα Επεξεργασίας
Class	Κλάση
Cloud-Based-Services	Υπηρεσίες Υπολογιστικού Νέφους
Clue	Στοιχείο
Clustering Algorithms	Αλγόριθμοι Ομαδοποίησης
Color Jitter	Παραμόρφωση Χρώματος

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

Confusion Matrix	Μήτρα Σύγχυσης
Convolutional Neural Networks	Συνελκτικά Νευρωνικά Δίκτυα
Crop	Περικοπή
Dashboard	Ταμπλό
Data Augmentation	Επαύξηση Δεδομένων
Deep Features	“Βαθιά” Χαρακτηριστικά
Deep Learning	Βαθιά Μάθηση
Document Classification	Ταξινόμηση Εγγράφων
Document Retrieval	Ανάκτηση Εγγράφων
Dropout Technique	Τεχνική Απενεργοποίησης Νευρώνων
Error Analysis	Ανάλυση Σφαλμάτων
F-Measure	Τελική μετρική ακρίβειας
False Negative	Ψευδώς Αρνητικό
False Positive	Ψευδώς Θετικό
Feature Selection	Επιλογή Χαρακτηριστικών
Feature Vector	Πίνακας Χαρακτηριστικών
Framework	Βιβλιοθήκη Λογισμικού
Global Metadata	Καθολικά Μεταδεδομένα
Graphical Processing Unit	Κάρτα Γραφικών Υπολογιστή
Grayscale	Κλίμακα του Γκρι
Guest User	Επισκέπτης-Χρήστης
Hardware	Υλισμικό
Hidden Layers	Κρυφά Επίπεδα Νευρώνων
Hyper-parameters	Υπερ-παράμετροι
Image Enhancement	Βελτίωση Ποιότητας Εικόνας
Image Processing	Επεξεργασία Εικόνας

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

Image Segmentation	Κατάτμηση Εικόνας
Inter-Domain	Μεταξύ Τομέων
Intra-Domain	Εντός του Τομέα
Label	Ετικέτα
Layers	Επίπεδα Νευρώνων
Layout Features	Δομικά Χαρακτηριστικά
Learning Rate	Ρυθμός Μάθησης
Loss Function	Συνάρτηση Κόστους
Machine Learning	Μηχανική Μάθηση
Mirroring	Αντικατοπτρισμός
Multi-threaded Processing	Πολυνηματική Επεξεργασία
Natural Language Processing	Επεξεργασία Φυσικής Γλώσσας
Neural Networks	Νευρωνικά Δίκτυα
Open-source Software	Λογισμικό Ανοικτού Κώδικα
Optical Character Recognition	Οπτική Αναγνώριση Χαρακτήρων
Overfitting	Υπερ-προσαρμογή
Pattern Recognition	Αναγνώριση Προτύπων
Pre-training	Προ-εκπαίδευση
Precision	Ακρίβεια
Principal Component Analysis	Ανάλυση Πρωτευόντων Συστατικών
Recall	Ανάκληση
Resize	Αλλαγή Μεγέθους
Rotation	Περιστροφή
Salt & Pepper Noise	Θόρυβος Μορφής αλατοπίπερου
Scaling	Κλιμάκωση
Server	Εξυπηρετητής
Shallow Features	“Ρηχά” Χαρακτηριστικά

Σύγχρονες τεχνικές αυτόματης ταξινόμησης εγγράφων

Shear	Στρέβλωση
Sigmoid Function	Σιγμοειδής Συνάρτηση
Textual Features	Κειμενικά Χαρακτηριστικά
Thumbnail	Μικρογραφία
Training Epochs	Κύκλοι Εκπαίδευσης
Training Phase	Φάση Εκπαίδευσης
Transfer Learning	Μεταφορά Γνώσης
Translation	Μετατόπιση
Tree View	Οπτικοποίηση Δενδρικής Μορφής
True Negative	Αληθώς Αρνητικό
True Positive	Αληθώς Θετικό
Underfitting	Υπο-προσαρμογή
Validation Phase	Φάση Επαλήθευσης
Virtual Machine	Εικονική Μηχανή
Visual Features	Οπτικά Χαρακτηριστικά
Weight Initialization	Αρχικοποίηση Τιμών για τα Βάρη των Νευρώνων
Weights	Βάρη



## ΣΥΝΤΜΗΣΕΙΣ - ΑΡΚΤΙΚΟΛΕΞΑ - ΑΚΡΩΝΥΜΙΑ

OCR	Optical Character Recognition
AI	Artificial Intelligence
GPU	Graphical Processing Unit
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
PCA	Principal Component Analysis
RGB	Red-Green-Blue
BERT	Bidirectional Encoder Representations from Transformers
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
NLP	Natural Language Processing
CPU	Central Processing Unit
GB/s	Gigabit per second
HTTPS	Hypertext Transfer Protocol Secure
SSL	Secure Sockets Layer
REST	Representational State Transfer
API	Application Programming Interface
RAM	Random Access Memory
VMware	Virtual Machine Software
ΒΔ	Βάση Δεδομένων