



**ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**  
UNIVERSITY OF WEST ATTICA

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**

**ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ**

**ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ  
ΥΠΟΛΟΓΙΣΤΩΝ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**« ΑΝΑΠΤΥΞΗ ΣΥΣΤΗΜΑΤΟΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ  
ΑΝΑΓΝΩΡΙΣΗΣ ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΝΟΗΜΑΤΙΚΗΣ ΓΛΩΣΣΑΣ »**

**Καψής Στυλιανός – Μιχαήλ**

**ΑΜ : 71347427**

**Εισηγητής Καθηγητής : Χρήστος Τρούσσας**

Διπλωματική εργασία υποβληθείσα προς μερική εκπλήρωση των απαραίτητων προϋποθέσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης.

**Αθήνα, Οκτώβριος, 2023**



**ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**  
UNIVERSITY OF WEST ATTICA

**UNIVERSITY OF WEST ATTICA**

**FACULTY OF ENGINEERING**

**DEPARTMENT OF INFORMATICS AND COMPUTER  
ENGINEERING**

**DIPLOMA THESIS**

**« DEVELOPMENT OF A GREEK SIGN LANGUAGE RECOGNITION  
SYSTEM WITH ARTIFICIAL INTELLIGENCE »**

**Kapsis Stylianos-Michail**

**RN : 71347427**

**Thesis Advisor and Supervisor : Christos Troussas**

Thesis submitted in partial fulfillment of the requirements for the attainment of the Master's Degree in  
Specialization.

**Athens, October, 2023**



**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**  
**« ΑΝΑΠΤΥΞΗ ΣΥΣΤΗΜΑΤΟΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ**  
**ΑΝΑΓΝΩΡΙΣΗΣ ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΝΟΗΜΑΤΙΚΗΣ ΓΛΩΣΣΑΣ »**

**Καψής Στυλιανός – Μιχαήλ**

**ΑΜ : 71347427**

**Εισηγητής :**

**Χρήστος Τρούσσας, Επιβλέπων καθηγητής**

**Εξεταστική Επιτροπή :**

**Χρήστος Τρούσσας, Επιβλέπων καθηγητής**

**Ακριβή Κρούσκα, Μέλος ΕΔΙΠ**

**Παναγιώτα Τσελέντη, Μέλος ΕΔΙΠ**

**Ημερομηνία εξέτασης :**

**Μάρτιος 2024**



## ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

«Βεβαιώνω ότι είμαι συγγραφέας της παρούσας διπλωματικής εργασίας και ότι έχω αναφέρει ή παραπέμπει σε αυτή, ρητά και συγκεκριμένα, όλες τις πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, προτάσεων ή λέξεων, είτε αυτές μεταφέρονται επακριβώς (στο πρωτότυπο ή μεταφρασμένες) είτε παραφρασμένες. Επίσης βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για την συγκεκριμένη διπλωματική εργασία»

Ο Δηλών

Όνοματεπώνυμο : Καυής Στυλιανός Μιχαήλ

(Υπογραφή φοιτητή)





## **ΕΥΧΑΡΙΣΤΙΕΣ**

Θα ήθελα να ευχαριστήσω ιδιαίτερα την μητέρα μου Ευτέρπη καθώς, τον αδερφό μου Δημήτρη, την κοπέλα μου Έλενα και τους φίλους μου, για την πολύτιμη βοήθεια τους στην ζωή μου. Επίσης θα ήθελα να ευχαριστήσω τους κοντινούς μου ανθρώπους για την συνεχή υποστήριξη τους. Τέλος, θα ήθελα να ευχαριστήσω τον καθηγητή μου για την ανάθεση του θέματος της παρούσας διπλωματικής εργασίας.





## ΠΕΡΙΛΗΨΗ

Το πεδίο της Όρασης των Υπολογιστών και της Μηχανικής Μάθησης έχει δει μεγάλες εξελίξεις τα τελευταία χρόνια, επιτρέποντας να δημιουργηθούν διάφορες εφαρμογές που βελτιώνουν την εμπειρία του χρήστη σε ποικίλους τομείς. Μία τέτοια εφαρμογή, η οποία αναπτύχθηκε στα πλαίσια της διπλωματικής εργασίας, αποτελεί η αναγνώριση χειρονομιών σε πραγματικό χρόνο. Μία εφαρμογή που διαθέτει το δυνατότητα να γεφυρώσει χάσματα επικοινωνίας και αν εμπλουτίσει την πραγματικότητα με ψηφιακές πληροφορίες. Σε αυτή τη διπλωματική παρουσιάζεται μία ολοκληρωμένη εξερεύνηση ενός συστήματος αναγνώρισης χειρονομιών πραγματικού χρόνου, σχεδιασμένο για την ερμηνεία του αλφαβήτου της Ελληνικής Νοηματικής Γλώσσας, χρησιμοποιώντας στατικές χειρονομίες, με δυνατότητα επέκτασης για περαιτέρω εκπαίδευση με επιπλέον χειρονομίες.

Το εισαγωγικό κεφάλαιο ορίζει τις βάσεις αναλύοντας τη σημασία της αναγνώρισης χειρονομιών, τονίζοντας τον στόχο της λύσης του προβλήματος της ερμηνείας της Ελληνικής Νοηματικής Γλώσσας και συζητώντας διάφορες προσεγγίσεις για την αντιμετώπιση αυτού του ζητήματος. Το δεύτερο κεφάλαιο αναφέρεται στις θεωρητικές βάσεις των εργαλείων τεχνολογίας που αξιοποιήθηκαν στην υλοποίηση της εφαρμογής αυτής. Ακόμα, εξετάζονται σχετικές θεωρητικές και πρακτικές μελέτες, που παρέχουν μια περιεκτική κατανόηση των βασικών αρχών και παραδείγματα ήδη υλοποιημένων εφαρμογών. Στο τρίτο κατά σειρά κεφάλαιο, παρουσιάζεται και αναλύεται η επιλεγμένη μεθοδολογία ανάπτυξης εφαρμογών, η γνωστή ως Rational Unified Process (RUP), καλύπτοντας και τις τέσσερις φάσεις της, υποστηριζόμενες από διαγράμματα περιπτώσεων χρήσης και διαγράμματα Gantt για να παρέχουν μία ολοκληρωμένη εικόνα της διαδικασίας ανάπτυξης. Η αρχιτεκτονική της εφαρμογής αναφέρεται στο τέταρτο κεφάλαιο. Ο πυρήνας της εφαρμογής αποτελείται από τέσσερα κύρια αρχεία, υλοποιημένα στη γλώσσα προγραμματισμού Python, αρχεία που αναλύονται διεξοδικά, εξηγώντας τις λειτουργίες και τον τρόπο αλληλεπίδρασής τους. Παρουσιάζεται στους αναγνώστες η εσωτερική λειτουργία του συστήματος, και τους δίνεται η ευκαιρία να την κατανοήσουν εις βάθος. Το πέμπτο κεφάλαιο αφιερώνεται στην εξήγηση και στην αξιολόγηση του πίνακα σύγχυσης, ο οποίος προκύπτει από την εκπαίδευση του μοντέλου μηχανικής μάθησης κατάλληλο για ταξινόμηση πολλών χειρονομιών. Διευκολύνει, με βάση μετρικών απόδοσης, στην καταμέτρηση του ύψους του βαθμού των ικανοτήτων του μοντέλου να αναγνωρίζει τις χειρονομίες της Ελληνικής Νοηματικής Γλώσσας. Το κλείσιμο της διπλωματικής, κεφάλαιο έξι, παρουσιάζει τα συμπεράσματα από την ολοκλήρωση και χρήση του συστήματος αναγνώρισης χειρονομιών, και περιγράφει πιθανούς τρόπους για την περαιτέρω μελλοντική ανάπτυξη και εξέλιξη του.

Ανακεφαλαιώνοντας, στα πλαίσια της διπλωματικής εργασίας αναπτύχθηκε ένα σύστημα αναγνώρισης χειρονομιών, που αντιστοιχούν σε γράμματα της Ελληνικής Νοηματικής Γλώσσας, αξιοποιώντας το API της MediaPipe-Hands για τον εντοπισμό των σημείων αναφοράς της παλάμης και μία διαδικασία συλλογής και προ-επεξεργασίας των δεδομένων. Έπειτα με την εκπαίδευση και τη χρήση ενός νευρωνικού δικτύου, καθίσταται, η αναγνώριση σε πραγματικό χρόνο των συγκεκριμένων χειρονομιών που αναφέρθηκαν παραπάνω, εφικτή. Πέρα από την ερμηνεία του αλφαβήτου της Ελληνικής Νοηματικής Γλώσσας, προσφέρει και δυνατότητα επέκτασης για την υποστήριξη παραπάνω χειρονομιών. Μέσω θεωρητικής μελέτης, μεθοδικής ανάπτυξης, δημιουργίας ευέλικτης αρχιτεκτονικής και ολοκληρωμένης αξιολόγησης, αυτό το έργο συμβάλλει στην ανάπτυξη εφαρμογών που εμπίπτουν στους τομείς της Όρασης Υπολογιστών και της Μηχανικής Μάθησης, θεμελιώνει σταθερές βάσεις για τη γεφύρωση της επικοινωνίας μεταξύ της κοινότητας των κωφών και βαρήκοων και των συνανθρώπων τους και ενισχύει τις αλληλεπιδράσεις των ανθρώπων με τους ηλεκτρονικούς υπολογιστές, αποφεύγοντας τη χρήση υπολογιστικού υλικού τελευταίας γενιάς.

**Λέξεις Κλειδιά:** Ελληνική νοηματική γλώσσα, Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Feedforward Neural Networks, Python, Αναγνώριση χειρονομιών, Σημεία ορόσημα παλάμης, Στατικές χειρονομίες, MediaPipe-Hands, Ταξινόμηση βάσει σημείων-κλειδιών

# ABSTRACT

The field of Computer Vision and Machine Learning has witnessed significant progress in recent years, enabling the creation of diverse applications that enhance user experience in various fields. Real-time gesture recognition is one of such applications, developed within the confines of this thesis. This application has the potential to bridge communication barriers and augment reality with digital information. This thesis presents a comprehensive investigation into a real-time gesture recognition system. The system is designed to interpret the Greek Sign Language alphabet using static gestures, with the potential for expansion to include additional gestures in future training. There are no subjective evaluations included, and technical term abbreviations are explained upon first use. The language is clear and concise, using high-level, standard language with consistent technical terms in a formal register. The structure is logical, with causal connections between statements and a balanced perspective. Precise subject-specific vocabulary is used for greater accuracy, and the text is free from grammatical errors, spelling mistakes, and punctuation errors. The formatting aligns with common academic sections, follows a consistent citation and footnote style, clearly marks quotes, and avoids filler words.

The first chapter sets out to analyze the significance of recognizing gestures. It highlights the objective of rectifying the issue of Greek Sign Language interpretation and deliberates several resolutions to this predicament. Chapter two is dedicated to the technical basis of the tools implicated in this system's execution. Furthermore, the text provides a thorough review of relevant theoretical and practical studies, allowing for a comprehensive understanding of basic principles and exemplified applications that have already been implemented. The third chapter details the chosen application development methodology, Rational Unified Process (RUP), presenting and analyzing all four phases. This analysis is further supported by use case diagrams and Gantt charts, offering an all-encompassing view of the development process. The fourth chapter presents an analysis of the application's architecture. The application's core is made up of four main files written in the Python programming language. These files are thoroughly examined, describing their functions and interactions with each other. The readers are provided with an in-depth explanation of the system's internal workings, allowing for comprehensive comprehension. The fifth chapter explains and evaluates the confusion matrix resulting from the training of a machine learning model for multi-gesture classification. Abbreviations are defined when first used. Clear structure and progression with causal relationships between statements are maintained. This matrix, based on performance metrics, measures the degree to which the model recognizes Greek Sign Language gestures. The language is formal and objective, avoiding biased or figurative language. The final chapter of this thesis provides an objective overview of the integration and use of the gesture recognition system, drawing conclusions from the

findings. Additionally, potential avenues for future development and improvement are discussed. The chapter follows a logical progression and maintains a formal, balanced tone throughout. Technical terms are explained where necessary, and the language is free from grammatical errors, colloquialisms, and filler words.

In summary, this thesis presents the development of a gesture recognition system for the Greek Sign Language. The API of Media Pipe-Hands was utilized to detect the reference points of the palm, and a data collection and pre-processing process was implemented. After training and applying a neural network, the system is able to perform real-time recognition of the specific gestures mentioned above. In addition to interpreting the alphabet of Greek Sign Language, this system also allows for further support of the mentioned gestures. This project contributes to the development of Computer Vision and Machine Learning applications through theoretical study, methodical development, creation of a flexible architecture, and integrated evaluation. It lays sturdy groundwork for enhanced communication between the deaf/hard of hearing community and others, and it elevates interactions with computers, all while avoiding the need for cutting-edge computing hardware.

**Keywords:** Greek sign language, Artificial Intelligence, Machine Learning, Feedforward Neural Networks, Python, Hand gesture recognition, Palm landmark points, Static gestures, Media Pipe-Hands, Key-Point based Classification

## ΠΕΡΙΕΧΟΜΕΝΑ

<b>Κατάλογος Εικόνων.....</b>	<b>σελίδα 17</b>
<b>Κατάλογος Εξισώσεων.....</b>	<b>σελίδα 22</b>
<b>Κεφάλαιο 1 , ΕΙΣΑΓΩΓΗ.....</b>	<b>σελίδα 23</b>
1.1 Αναφορά Προβλήματος.....	σελίδα 23
1.2 Στόχος Επίλυσης Προβλήματος .....	σελίδα 24
1.3 Τρόποι Επίλυσης Προβλήματος.....	σελίδα 25
1.4 Επιλεγμένος Τρόπος Επίλυσης Προβλήματος.....	σελίδα 26
<b>Κεφάλαιο 2 , ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ ΚΑΙ ΣΧΕΤΙΚΕΣ ΜΕΛΕΤΕΣ.....</b>	<b>σελίδα 28</b>
2.1 Θεωρητικό Υπόβαθρο.....	σελίδα 28
2.1.1 Τεχνητή Νοημοσύνη.....	σελίδα 28
2.1.2 Μηχανική Μάθηση.....	σελίδα 30
2.1.2.1 Επιβλεπόμενη Μάθηση.....	σελίδα 32
2.1.2.2 Μη-Επιβλεπόμενη Μάθηση.....	σελίδα 33
2.1.2.3 Ενισχυτική Μάθηση.....	σελίδα 34
2.1.2.4 Άλλοι τύποι Μηχανικής Μάθησης.....	σελίδα 35
2.1.3 Βαθιά Μάθηση.....	σελίδα 36
2.1.4 Νευρωνικά Δίκτυα.....	σελίδα 37
2.1.5 Συναρτήσεις Κόστους ή Απώλειας.....	σελίδα 39
2.1.6 Συναρτήσεις Ενεργοποίησης.....	σελίδα 40
2.1.7 Όραση Υπολογιστών.....	σελίδα 43
2.2 Ερευνητικά Άρθρα και Μελέτες.....	σελίδα 46

<b>Κεφάλαιο 3 , ΜΕΘΟΔΟΛΟΓΙΑ ΑΝΤΙΚΕΙΜΕΝΟΣΤΡΑΦΟΥΣ ΑΝΑΛΥΣΗΣ.....</b>	<b>σελίδα 74</b>
3.1 Φάση Σύλληψης.....	σελίδα 74
3.1.1 Περιγραφή Έργου.....	σελίδα 74
3.1.2 Στόχοι Έργου.....	σελίδα 74
3.1.3 Πεδίο Εφαρμογής Έργου.....	σελίδα 75
3.1.4 Τεχνολογίες που Χρησιμοποιήθηκαν.....	σελίδα 75
3.1.5 Δημιουργία Συνόλου Δεδομένων.....	σελίδα 76
3.1.6 Μοντέλο Μηχανικής Μάθησης.....	σελίδα 76
3.1.7 Δικαιούχοι .....	σελίδα 76
3.1.8 Περιορισμοί.....	σελίδα 77
3.1.9 Εκτίμηση Κινδύνων.....	σελίδα 77
3.1.10 Σχέδιο Έργου.....	σελίδα 78
3.1.11 Αρχικό Μοντέλο Περιπτώσεων Χρήσης.....	σελίδα 79
3.2 Φάση Επεξεργασίας.....	σελίδα 81
3.2.1 Επέκταση Μοντέλου Περιπτώσεων Χρήσης.....	σελίδα 81
3.2.2 Περιγραφή Αρχιτεκτονικής Λογισμικού.....	σελίδα 85
3.2.3 Επανεξέταση Κινδύνων.....	σελίδα 86
3.2.4 Τρόποι Αντιμετώπισης – Πιθανές Λύσεις.....	σελίδα 87
3.2.5 Σχέδιο Ανάπτυξης Ολικού Έργου.....	σελίδα 88
3.2.6 Τελικό Μοντέλο Περιπτώσεων Χρήσης.....	σελίδα 92
3.3 Φάση Κατασκευής.....	σελίδα 96
3.3.1 Προϊόν Λογισμικού.....	σελίδα 96
3.3.2 Εγχειρίδιο Χρήστη.....	σελίδα 98
3.4 Φάση Μετάβασης.....	σελίδα 101

<b>Κεφάλαιο 4 , ΑΡΧΙΤΕΚΤΟΝΙΚΗ και ΥΛΟΠΟΙΗΣΗ ΕΦΑΡΜΟΓΗΣ.....</b>	<b>σελίδα 104</b>
4.1 Αρχείο app.py.....	σελίδα 104
4.1.1 Ανάλυση Συναρτήσεων και Περιγραφή Λειτουργιών.....	σελίδα 104
4.2 Αρχείο keypoint_classification.py.....	σελίδα 111
4.2.1 Εκπαίδευση και Αξιολόγηση Μοντέλου Μηχανικής Μάθησης.....	σελίδα 111
4.3 Αρχείο KeyPointClassifier.py.....	σελίδα 119
4.3.1 Ανάλυση και Περιγραφή Αρχείου.....	σελίδα 119
4.4 Αρχείο cvfprscalr.py.....	σελίδα 121
4.4.1 Ανάλυση και Περιγραφή Αρχείου.....	σελίδα 121
<b>Κεφάλαιο 5 , ΑΞΙΟΛΟΓΗΣΗ ΠΙΝΑΚΑ ΣΥΓΧΥΣΗΣ.....</b>	<b>σελίδα 130</b>
5.1 Ανατομία Πίνακα Σύγχυσης.....	σελίδα 130
5.2 Ερμηνεία Αποτελεσμάτων Πίνακα Σύγχυσης.....	σελίδα 132
<b>Κεφάλαιο 6 , ΣΥΜΠΕΡΑΣΜΑΤΑ και ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ.....</b>	<b>σελίδα 139</b>
6.1 Συμπεράσματα.....	σελίδα 139
6.2 Μελλοντικές Επεκτάσεις.....	σελίδα 140
<b>Βιβλιογραφία.....</b>	<b>σελίδα 142</b>



## Κατάλογος Εικόνων

### Εικόνες Άρθρων

Εικόνα 2-1: Γενική Ανασκόπηση της προτεινόμενης Αρχιτεκτονικής.....σελίδα 47
Εικόνα 2-2: Δομή του προτεινόμενου κελιού GRU.....σελίδα 48
Εικόνα 2-3: Σύγκριση ποσοστών (%) Ακρίβειας Μοντέλων Μηχανικής και Βαθιάς Μάθησης.....σελίδα 48
Εικόνα 2-4: Σύγκριση Απώλειας Μοντέλων Μηχανικής και Βαθιάς Μάθησης .....σελίδα 50
Εικόνα 2-5: Σχηματική Αναπαράσταση του προτεινόμενου μοντέλου Νευρωνικών Δικτύων για την αναγνώριση χειρονομιών.....σελίδα 51
Εικόνα 2-6: Διάγραμμα Προσομοίωσης Ακρίβειας και Απώλειας για την εκπαίδευση του μοντέλου Νευρωνικών Δικτύων που προτείνεται .....σελίδα 52
Εικόνα 2-7: Διάγραμμα Ροής Ολοκληρωμένου Έργου.....σελίδα 53
Εικόνα 2-8: Διαμόρφωση Αρχιτεκτονικής προτεινόμενου Νευρωνικού Δικτύου Πρόσθιας Τροφοδότησης.....σελίδα 53
Εικόνα 2-9: Αναφορά Ταξινόμησης του προτεινόμενου Μοντέλου Νευρωνικού Δικτύου Πρόσθιας Τροφοδότησης.....σελίδα 54
Εικόνα 2-10: Πίνακας Συσχέτισης του προτεινόμενου Μοντέλου Νευρωνικού Δικτύου Πρόσθιας Τροφοδότησης.....σελίδα 54
Εικόνα 2-11: Συγκριτική Ανάλυση του προτεινόμενου Μοντέλου Νευρωνικών Δικτύων , με άλλα Μοντέλα δυοδιάστατης προσέγγισης .....σελίδα 55
Εικόνα 2-12: : Συγκριτική Ανάλυση του προτεινόμενου Μοντέλου Νευρωνικών Δικτύων , με άλλα Μοντέλα τρισδιάστατης προσέγγισης.....σελίδα 56
Εικόνα 2-13: Πίνακας Συσχέτισης για τα χαρακτηριστικά βασισμένα στην Απόσταση.....σελίδα 57
Εικόνα 2-14: : Πίνακας Συσχέτισης για τα χαρακτηριστικά βασισμένα στις γωνίες.....σελίδα 60

Εικόνα 2-15: Σύγκριση Ποσοστών (%) Απόδοσης διαφορετικών προσεγγίσεων Συνεχούς Αναγνώρισης της Νοηματικής Γλώσσας.....σελίδα 60	σελίδα 60
Εικόνα 2-16: Ένα μέρος του Συνόλου Δεδομένων , που χρησιμοποιήθηκε για την έρευνα.....σελίδα 61	σελίδα 61
Εικόνα 2-17: Δομή Αρχιτεκτονικής Δικτύου RPN.....σελίδα 61	σελίδα 61
Εικόνα 2-18: Δομή Αρχιτεκτονικής Δικτύου Faster-RCNN.....σελίδα 63	σελίδα 63
Εικόνα 2-19: Δομή Αρχιτεκτονικής Δικτύου Multiscale Faster-RCNN.....σελίδα 63	σελίδα 63
Εικόνα 2-20: Προτεινόμενη Αρχιτεκτονική για τον Εντοπισμό της Παλάμης και την Αναγνώριση χειρονομιών.....σελίδα 64	σελίδα 64
Εικόνα 2-21: Μέση Ακρίβεια που επιτυγχάνεται χρησιμοποιώντας αλγορίθμους Μηχανικής και Βαθιάς Μάθησης.....σελίδα 64	σελίδα 64
Εικόνα 2-22: Σύγκριση αποτελεσμάτων με παρόμοιες καινοτόμες μεθοδολογίες.....σελίδα 65	σελίδα 65
Εικόνα 2-23: Σύγκριση αποτελεσμάτων με παρόμοιες καινοτόμες μεθοδολογίες.....σελίδα 65	σελίδα 65
Εικόνα 2-24: Πίνακας Συσχέτισης: Αμερικανικής Νοηματικής Γλώσσας (αλφαβήτα) .....σελίδα 65	σελίδα 65
Εικόνα 2-25: Πίνακας Συσχέτισης: Αμερικανικής Νοηματικής Γλώσσας (αριθμοί) .....σελίδα 65	σελίδα 65
Εικόνα 2-26: Πίνακας Συσχέτισης: Ινδικής Νοηματικής Γλώσσας (αλφαβήτα) .....σελίδα 66	σελίδα 66
Εικόνα 2-27: Πίνακας Συσχέτισης: Τουρκικής Νοηματικής Γλώσσας (αριθμοί) .....σελίδα 67	σελίδα 67
Εικόνα 2-28: Πίνακας Συσχέτισης: Ιταλικής Νοηματικής Γλώσσας (αλφαβήτα) .....σελίδα 68	σελίδα 68
Εικόνα 2-29: Γραφική Αναπαράσταση της Αρχιτεκτονικής του Μοντέλου FCCNN.....σελίδα 69	σελίδα 69
Εικόνα 2-30: Σύγκριση του χρόνου εκπαίδευσης του 3DNN στο Σύνολο Δεδομένων MNIST ....σελίδα 69	σελίδα 69
Εικόνα 2-31: Χρόνος εκτέλεσης ανά εποχή(epoch) για την Εκπαίδευση των Μοντέλων AlexNet και VGG16 στο Σύνολο Δεδομένων CIFAR-10.....σελίδα 70	σελίδα 70
Εικόνα 2-32: Εκτιμώμενος χρόνος Εκπαίδευσης για το AlexNet (70% ποσοστό Ακριβείας) και VGG16 (75% ποσοστό Ακριβείας) στο Σύνολο Δεδομένων CIFAR-10.....σελίδα 70	σελίδα 70
Εικόνα 2-33: Ακρίβεια του Μοντέλου AlexNet , εκπαιδευμένο με τους αλγορίθμους Adam και SGD στο Σύνολο Δεδομένων CIFAR-10.....σελίδα 70	σελίδα 70

Εικόνα 2-34: Ακρίβεια του Μοντέλου VGG16 , εκπαιδευμένο με τους αλγορίθμους Adam και SGD στο Σύνολο Δεδομένων CIFAR-10.....σελίδα 71	σελίδα 71
Εικόνα 2-35: Λίστα των Συναρτήσεων Απώλειας.....σελίδα 71	σελίδα 71
Εικόνα 2-36: Καμπύλες Μάθησης για το Σύνολο Δεδομένων MNIST και CIFAR-10 , διαγράμματα ταχύτητας Μάθησης Μοντέλου.....σελίδα 72	σελίδα 72
Εικόνα 2-37: Καμπύλες Ακριβείας εκπαίδευσης και ελέγχου για το Σύνολο Δεδομένων MNIST.....σελίδα 72	σελίδα 72

## **Εικόνες Εφαρμογής**

Εικόνα 3-1: Αρχικό Μοντέλο Περιπτώσεων Χρήσης της εφαρμογής Αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας.....σελίδα 81	σελίδα 81
Εικόνα 3-2: Μοντέλο Περιπτώσεων Χρήσης για την εφαρμογή Αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας.....σελίδα 84	σελίδα 84
Εικόνα 3-3:Τελικό Μοντέλο Περιπτώσεων Χρήσης για την εφαρμογή Αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας.....σελίδα 95	σελίδα 95
Εικόνα 3-4: Πίνακας Gantt της διαδικασίας υλοποίησης της εφαρμογής Αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας.....σελίδα 96	σελίδα 96
Εικόνα 3-5: Διάγραμμα Gantt για τη διαδικασία υλοποίησης της εφαρμογής Αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας.....σελίδα 97	σελίδα 97
Εικόνα 4-1: Οπτικοποίηση Αρχιτεκτονικής του Ανεπτυγμένου Μοντέλου Βαθιάς Μάθησης ...σελίδα 113	σελίδα 113
Εικόνα 4-2: Εικόνα από το αρχείο ετικετών (label) keypoint_classifier_label.csv.....σελίδα 116	σελίδα 116
Εικόνα 4-3: Εικόνα από το αρχείο keypoint.csv , το Σύνολο Δεδομένων.....σελίδα 116	σελίδα 116
Εικόνα 4-4: Συλλογή Δεδομένων για το γράμμα "Α".....σελίδα 122	σελίδα 122
Εικόνα 4-5: Συλλογή Δεδομένων για το γράμμα "Β".....σελίδα 123	σελίδα 123
Εικόνα 4-6: Συλλογή Δεδομένων για το γράμμα "Δ".....σελίδα 123	σελίδα 123
Εικόνα 4-7: Συλλογή Δεδομένων για το γράμμα "Π".....σελίδα 124	σελίδα 124

Εικόνα 4-8: Συλλογή Δεδομένων για το γράμμα "X".....σελίδα	124
Εικόνα 4-9: Χειρονομία προς αναγνώριση του γράμματος "Γ".....σελίδα	125
Εικόνα 4-10: Χειρονομία προς αναγνώριση του γράμματος "Ζ".....σελίδα	126
Εικόνα 4-11: Χειρονομία προς αναγνώριση του γράμματος "Ι".....σελίδα	126
Εικόνα 4-12: Χειρονομία προς αναγνώριση του γράμματος "Κ".....σελίδα	126
Εικόνα 4-13: Χειρονομία προς αναγνώριση του γράμματος "Ν".....σελίδα	127
Εικόνα 4-14: Χειρονομία προς αναγνώριση του γράμματος "Ο".....σελίδα	127
Εικόνα 4-15: Χειρονομία προς αναγνώριση του γράμματος "Ρ".....σελίδα	127
Εικόνα 4-16: Χειρονομία προς αναγνώριση του γράμματος "Τ".....σελίδα	128
Εικόνα 4-17: Χειρονομία προς αναγνώριση του γράμματος "Υ".....σελίδα	128
Εικόνα 4-18: Χειρονομία προς αναγνώριση του γράμματος "Ψ".....σελίδα	128
Εικόνα 4-19: Χειρονομία προς αναγνώριση του γράμματος "Ω".....σελίδα	130
Εικόνα 5-1: Πίνακας Σύγκρισης ReLU.....σελίδα	133
Εικόνα 5-2: Πίνακας Σύγκρισης eLU.....σελίδα	133
Εικόνα 5-3: Πίνακας Σύγκρισης GeLU.....σελίδα	133
Εικόνα 5-4: Πίνακας Σύγκρισης Leaky ReLU.....σελίδα	133
Εικόνα 5-5: Πίνακας Σύγκρισης Sigmoid.....σελίδα	134
Εικόνα 5-6: Πίνακας Σύγκρισης Tanh.....σελίδα	134
Εικόνα 5-7: Πίνακας Σύγκρισης Swiss.....σελίδα	134
Εικόνα 5-8: Έλεγχος Κατηγοριοποίησης ReLU.....σελίδα	135
Εικόνα 5-9: Έλεγχος Κατηγοριοποίησης eLU.....σελίδα	135
Εικόνα 5-10: Έλεγχος Κατηγοριοποίησης GeLU.....σελίδα	136
Εικόνα 5-11: Έλεγχος Κατηγοριοποίησης Leaky ReLU.....σελίδα	136
Εικόνα 5-12: Έλεγχος Κατηγοριοποίησης Sigmoid.....σελίδα	136

Εικόνα 5-13: Έλεγχος Κατηγοριοποίησης Tanh.....σελίδα 136  
Εικόνα 5-14: Έλεγχος Κατηγοριοποίησης Swiss .....σελίδα 137

## Κατάλογος Εξισώσεων

Εξίσωση 2-1: Συνάρτηση Αθροίσματος.....σελίδα	39
Εξίσωση 2-2: Συνάρτηση Μετάβασης.....σελίδα	39
Εξίσωση 2-3: Συνάρτηση Μέσου Τετραγωνικού Σφάλματος.....σελίδα	40
Εξίσωση 2-4: Συνάρτηση Διωνυμικής Διασταυρωμένης Εντροπίας.....σελίδα	40
Εξίσωση 2-5: Συνάρτηση Κατηγορικής Διασταυρωμένης Εντροπίας.....σελίδα	41
Εξίσωση 2-6: Λογιστικής Συνάρτησης.....σελίδα	41
Εξίσωση 2-7: Συνάρτηση Υπερβολικής Εφαπτομένης.....σελίδα	42
Εξίσωση 2-8: Συνάρτηση Γραμμικής Μονάδας Ανόρθωσης.....σελίδα	42
Εξίσωση 2-9: Συνάρτηση Ενεργοποίησης Swish .....	σελίδα 43
Εξίσωση 2-10: Σιγμοειδής Συνάρτηση .....	σελίδα 43
Εξίσωση 2-11: Συνάρτηση Ενεργοποίησης GeLU.....σελίδα	43
Εξίσωση 2-12: Κανονικοποιημένη Εκθετική Συνάρτηση.....σελίδα	44
Εξίσωση 5-1: Συνάρτηση Ακριβείας.....σελίδα	132
Εξίσωση 5-2: Συνάρτηση Ανάκλησης .....	σελίδα 133
Εξίσωση 5-3: Συνάρτηση Αρμονικού Μέσου.....σελίδα	133
Εξίσωση 5-4: Συνάρτηση Ορθότητας.....σελίδα	133

# Κεφάλαιο 1<sup>ο</sup>

## ΕΙΣΑΓΩΓΗ

Σε έναν κόσμο που καθοδηγείται από την επικοινωνία, η αδυναμία μετάδοσης σκέψεων, συναισθημάτων και ιδεών λόγω των προβλημάτων ακοής αποτελεί μια βαθιά πρόκληση. Τα κωφά ή βαρήκοα άτομα αντιμετωπίζουν συχνά εμπόδια που εμποδίζουν την ικανότητά τους να αλληλοεπιδρούν άνετα, ενώ οι υπάρχουσες λύσεις επικοινωνίας μπορεί να είναι απαγορευτικά ακριβές. Ο πρωταρχικός στόχος της αντιμετώπισης αυτού του ζητήματος είναι η ενδυνάμωση των ατόμων με προβλήματα ακοής, δίνοντάς τους τη δυνατότητα να επικοινωνούν ελεύθερα και χωρίς αποκλεισμούς με το ευρύτερο κοινωνικό σύνολο. Το παρόν κεφάλαιο διεισδύει στην καρδιά του προβλήματος, αναγνωρίζοντας τις σημαντικές προκλήσεις που αντιμετωπίζουν τα κωφά ή βαρήκοα άτομα στην καθημερινή τους ζωή. Διερευνά την επιτακτική ανάγκη για καινοτόμες λύσεις που διευκολύνουν τη φυσική και αποτελεσματική επικοινωνία αυτής της κοινότητας. Για την αντιμετώπιση αυτής της πρόκλησης, διάφορες μεθοδολογίες και προσεγγίσεις έχουν προταθεί και διερευνηθεί κατά την πάροδο του χρόνου. Κάθε προσέγγιση επιφέρει τα μοναδικά της πλεονεκτήματα και τους περιορισμούς της. Η παρούσα διπλωματική εργασία, επικεντρώνεται σε μια νέα και τεχνολογικά προηγμένη λύση. Αξιοποιώντας το MediaPipe - Hands API, μια τεχνολογία αιχμής για τον εντοπισμό παλάμης, σε συνδυασμό με ένα Νευρωνικό Δίκτυο Πρόσθιας τροφοδότησης (Feedforward Neural Network), σχεδιασμένο για προσαρμοσμένη αναγνώριση χειρονομιών, το προτεινόμενο σύστημα αξιοποιεί τη δύναμη των οπτικών δεδομένων σε πραγματικό χρόνο. Μέσω της χρήσης μιας τυπικής ιστο-κάμερας, το σύστημα αυτό προσφέρει ένα δυναμικό και προσβάσιμο μέσο επικοινωνίας για άτομα με προβλήματα ακοής.

### 1.1 Αναφορά Προβλήματος

Τη σημερινή εποχή, η πλειοψηφία των ανθρώπων χρησιμοποιεί την ακοή και την ομιλία για την άμεση επικοινωνία. Δυστυχώς, έτσι δημιουργείται ένα φράγμα στην επικοινωνία μεταξύ της πλειοψηφίας των κατοίκων μίας χώρας, μίας πόλης ή ενός χωριού, και της κοινότητας των κωφών της εκάστοτε περιοχής. Το γεγονός αυτό, σε συνδυασμό με την ανεπαρκή προσπάθεια προώθησης εκμάθησης της νοηματικής γλώσσας της κάθε χώρας, στις περισσότερες περιοχές ανά τον κόσμο, έχει ως αποτέλεσμα η κοινότητα των κωφών να περιθωριοποιείται και να μην έχει πρόσβαση σε βασικές και απαραίτητες καθημερινές υπηρεσίες, όπως για παράδειγμα, όταν οι άνθρωποι αυτοί επιθυμούν να εξυπηρετηθούν σε κάποια δημόσια ή και ιδιωτική υπηρεσία, και είναι απαραίτητο να έχουν μαζί τους κάποιον διερμηνέα, διότι διαφορετικά δεν θα μπορέσουν να επικοινωνήσουν με τον υπάλληλο της υπηρεσίας [1].[3].

Τα μέλη της κοινότητας των κωφών, επικοινωνούν με τους ανθρώπους γύρω τους μέσω διάφορων χειρονομιών και κινήσεων, που απαρτίζουν τη νοηματική τους γλώσσα. Όπως η κάθε χώρα έχει τη δικιά της γλώσσα ή και γλώσσες, έτσι και σχεδόν κάθε γλώσσα συνοδεύεται από τη δικιά της νοηματική γλώσσα. Κάθε νοηματική γλώσσα είναι διαφορετική από κάθε άλλη, αν και συμβαίνει να χρησιμοποιούν ίδιες χειρονομίες ή κινήσεις, για διαφορετικές λέξεις, φράσεις ή γράμματα [4].

Δυστυχώς, η προώθηση εκμάθησης της νοηματικής γλώσσας δεν γίνεται αποτελεσματικά μέσα στις κοινωνίες. Το κενό στην επικοινωνία μεταξύ του πλήθους των ανθρώπων και της κοινότητας των κωφών, συνήθως καλύπτεται με κάποιον διερμηνέα. Το σύστημα αυτό όμως, είναι μία πολυτέλεια που δεν έχουν πολλοί άνθρωποι της κοινότητας αυτής, ειδικά αν σκεφτούμε πως πολλά από τα μέλη αυτής της κοινότητας δεν έχουν την δυνατότητα και την ευκαιρία να εργαστούν. Για αυτό το λόγο, και πρέπει να εξελιχθούν τα αυτόματα συστήματα αναγνώρισης της νοηματικής γλώσσας. Μέσω αυτών δίνεται η δυνατότητα να γεφυρωθεί το κενό της επικοινωνίας, μεταξύ των κωφών και των συνανθρώπων τους [12].[14].

## 1.2 Στόχος επίλυσης προβλήματος

Αναφέροντας το πρόβλημα, γίνεται κατανοητό πως η λύση είναι η γεφύρωση της επικοινωνίας μεταξύ της κοινότητας των κωφών και του πλήθους των ανθρώπων της ευρύτερης περιοχής. Για να συμβεί αυτό, πρέπει να εξετασθεί και να αναλυθεί η νοηματική γλώσσα που βρίσκεται στο επίκεντρο του ενδιαφέροντος. Οι νοηματικές γλώσσες αποτελούνται κυρίως από τις κινήσεις των χεριών και τις χειρονομίες, αλλά συνοδεύονται από τις εκφράσεις του προσώπου και της γλώσσας του σώματος, που δίνουν έμφαση στον τόνο, το ύφος και τις λέξεις του πομπού. Στην διπλωματική εργασία, θα αναλυθεί η ελληνική νοηματική γλώσσα, και πιο συγκεκριμένα το αλφάβητό της, το οποίο αποτελείται μόνο από στατικές χειρονομίες, δηλαδή από συμβολισμούς που γίνονται κυρίως με τα δάχτυλα των χεριών [13].

Οι άνθρωποι είναι σε θέση να αναγνωρίσουν τη γλώσσα του σώματος και τις νοηματικές κινήσεις πιο εύκολα από έναν υπολογιστή. Αυτό συμβαίνει λόγω του συνδυασμού της όρασης και των συν - απτικών αλληλεπιδράσεων που διαμορφώθηκαν κατά τη διάρκεια της εξέλιξης του εγκεφάλου. Προκειμένου να αναπαραχθεί αυτή η ικανότητα στους υπολογιστές είναι αναγκαίο να λυθούν ορισμένα προβλήματα: πως να μεταφερθούν τα αντικείμενα ενδιαφέροντος σε εικόνες, μαζί με τις απαραίτητες πληροφορίες τους, και ποιες τεχνολογίες καταγραφής εικόνας και τεχνικές κατηγοριοποίησης είναι οι καταλληλότερες να χρησιμοποιηθούν.



Συνεπώς, ο τελικός στόχος της εργασίας είναι η κατασκευή ενός αυτόματου συστήματος αναγνώρισης του αλφάβητου της ελληνικής νοηματικής γλώσσας. Ένα σύστημα το οποίο δεν θα χρειάζονται ιδιαίτερες προδιαγραφές υπολογιστικού υλικού για να έχουν οι χρήστες πρόσβαση, οικονομικό και τελευταίο και σημαντικότερο, οι αναγνωρίσεις του να έχουν μεγάλο ποσοστό ακριβείας [19].[20].

### 1.3 Τρόποι επίλυσης προβλήματος

Η αναγνώριση της νοηματικής γλώσσας έχει προσεγγιστεί με διάφορους τρόπους. Οι δύο πιο συνηθισμένες μέθοδοι προσέγγισης του προβλήματος χωρίζονται σε δύο κατηγορίες. Σε αυτές που είναι βασισμένες σε αισθητήρες ενσωματωμένους σε γάντια, και σε αυτές που βασίζονται στην όραση των υπολογιστών. Οι μέθοδοι βασισμένες σε γάντι λειτουργούν με δεδομένα που εκμαιεύονται από τις κινήσεις και τις θέσεις των χεριών του χρήστη τις οποίες και καταγράφουν, όταν αυτός φοράει και χρησιμοποιεί τον αντίστοιχο εξοπλισμό. Εν αντιθέσει, οι μέθοδοι που βασίζονται στην όραση των υπολογιστών, χρησιμοποιούν τεχνικές επεξεργασίας εικόνας και σήματος, αναλύοντας εικόνες ή βίντεο ως είσοδο για τον καθορισμό του σχετικού σήματος [4].

Οι προσεγγίσεις βασισμένες σε αισθητήρες εξάγουν τις μετρήσεις του χεριού, δηλαδή το προσανατολισμό των αρθρώσεων, τη θέση και την ταχύτητα των χεριών. Οι μετρήσεις αυτές πραγματοποιούνται με τη χρήση μικροελεγκτών και συγκεκριμένων αισθητήρων, όπως γάντια δεδομένων, ισχυρών γαντιών, ψηφιακών καμερών, επιταχυνσιομέτρων, καμερών βάθους, Kinect, ελεγκτών τύπου Leap Motion, επιδέξιων γαντιών μεταβλητών κ.α. Το σημείο στο οποίο υπερτερούν οι μέθοδοι βασισμένες σε αισθητήρια όργανα είναι ο υψηλότερος ρυθμός αναγνώρισης, λόγω των σκελετικών δεδομένων. Παράλληλα όμως, ενώ παρέχουν μεγαλύτερη ακρίβεια οι προσεγγίσεις βασισμένες σε αισθητήρες, επιτρέπουν περιορισμένες κινήσεις, απαιτούν εξειδικευμένες συσκευές και περιβάλλον, για την πλήρη χρήση των συστημάτων. Επιπρόσθετα, υπάρχει κίνδυνος μείωσης του ρυθμού αναγνώρισης των συστημάτων αυτών, κυρίως λόγω του θορύβου. Αισθητήρες όπως τα επιταχυνσιόμετρα, είναι ευαίσθητοι στο θόρυβο, με αποτέλεσμα, ακόμα και μία μικρή και ασήμαντη κίνηση να αναγνωριστεί ως κυματομορφή [1].[3].

Θέλοντας να καταπολεμήσουν τα εμπόδια αυτά, οι ερευνητές έχουν στραφεί σε μεθόδους βασισμένες στην όραση υπολογιστών, δηλαδή χρησιμοποιούν εισόδους από κάμερα, όπως είναι η διαδικτυακή (web) κάμερα του υπολογιστή. Τέτοιες προσεγγίσεις είναι πιο ελκυστικές, καθώς δεν απαιτούν κάποιο εξειδικευμένο εργαλείο ή συσκευή, που περιορίζει την κινησιολογία των χεριών, και μπορούν να πραγματοποιήσουν τις αναγνωρίσεις δίχως επαφή. Το κύριο πλεονέκτημά της, έναντι αυτής που βασίζεται σε αισθητήρια όργανα, είναι ότι αποτελεί μια πολύ πιο προσιτή οικονομικά προσέγγιση,

αλλά δεν προσφέρει τον ίδιο ρυθμό αναγνώρισης χειρονομιών και έχει υψηλό κόστος υπολογιστικής ισχύος [24].[34].[36].

Συνοψίζοντας, κάθε προσέγγιση παρουσιάζει τα δικά της πλεονεκτήματα και μειονεκτήματα, συμπεριλαμβανομένων των υπολογιστικών δαπανών, των προδιαγραφών ακριβείας και των απαιτήσεων εξοπλισμού. Υπάρχει πρόοδος και στις δύο μεθόδους. Οι μέθοδοι που βασίζονται σε αισθητήρες και γάντια για την απόκτηση των σκελετικών δεδομένων, είναι ιστορικά πιο ακριβείς, λόγω του εξειδικευμένου εξοπλισμού, αλλά οι μέθοδοι βασισμένες στην όραση των υπολογιστών, που ανιχνεύουν πρώτα το χέρι και έπειτα εξάγουν τα επιθυμητά χαρακτηριστικά του χεριού, γίνονται όλο και πιο διαδεδομένες λόγω του μικρού οικονομικού τους κόστους και της μεγάλης υπολογιστικής τους ακρίβειας, παρά όλα αυτά έχουν μικρότερο ρυθμό αναγνώρισης χειρονομιών [14].[30].[33].

## 1.4 Επιλεγμένος τρόπος επίλυσης προβλήματος

Σε αυτή τη μελέτη, η λύση στο πρόβλημα της αναγνώρισης και κατανόησης του ελληνικού αλφάβητου της νοηματικής, προσεγγίστηκε με τη μέθοδο της όρασης των υπολογιστών. Η υλοποίηση της εφαρμογής έγινε με τη χρήση της κάμερας του ηλεκτρονικού υπολογιστή, ιστο-κάμερα (webcam), ένα εργαλείο σχετικά εύκολο να αποκτηθεί με τιμή προσιτή. Ένα εμφανές πλεονέκτημα της μεθόδου αυτής, όπως ειπώθηκε προηγουμένως, είναι το ότι δεν απαιτείται ακριβός ή εξειδικευμένος εξοπλισμός, όπως ο αισθητήρας Leap Motion ή η κάμερα 3D [13].

Η εξαγωγή των χαρακτηριστικών έγινε με τη χρήση ενός πρόσφατα ανεπτυγμένου εκτιμητή συντεταγμένων από την Google, γνωστό ως «media-ripe», για την απόκτηση εικοσιένα εκτιμητών συντεταγμένων των χεριών από δισδιάστατες εικόνες. Μετά από πολυάριθμες δοκιμές εκτιμητών συντεταγμένων, επιλέχθηκε ο συγκεκριμένος, καθώς είναι λιγότερο επιρρεπής σε κατάρρευση, και δεν επηρεάζεται από εξωτερικούς συντελεστές, όπως το περιβάλλον – φόντο της δυσδιάστατης φωτογραφίας και το μέγεθος των χεριών. Από τους εκτιμητές συντεταγμένων ή αρθρώσεων, εξήχθησαν χαρακτηριστικά βασισμένα σε αποστάσεις και γωνίες. Λόγω της χρήσης των χαρακτηριστικών που αναφέρθηκαν παραπάνω, η διαδικασία εξαγωγής χαρακτηριστικών δεν είναι υπολογιστικά δαπανηρή, όπως τα συνελκτικά νευρωνικά δίκτυα, ή οι μηχανισμοί βασισμένοι στο χρώμα. Επίσης, με αυτή τη μέθοδο είναι εφικτό να εξαχθούν χαρακτηριστικά για τις αρθρώσεις σε τρεις διαστάσεις, από μία δισδιάστατη εικόνα, χωρίς εξειδικευμένο εξοπλισμό. Αυτό σημαίνει πως το προτεινόμενο σύστημα μας εξασφαλίζει τα πλεονεκτήματα των προσεγγίσεων βασισμένων σε αισθητήρια όργανα, και παράλληλα τα πλεονεκτήματα των προσεγγίσεων βασισμένων στην όραση των υπολογιστών, όπως είναι η ελευθερία κίνησης, η υψηλή ακρίβεια, η αναγνώριση δίχως αισθητήριου οργάνου και είναι και οικονομικά προσιτό [17].[35].[39].

Το σύστημα αναγνώρισης χειρονομιών που κατασκευάστηκε για την διπλωματική εργασία, ανιχνεύει την περιοχή του χεριού από την είσοδο της βίντεο ροής της ιστοκάμερας (webcam), χρησιμοποιώντας το API του MediaPipe και παρακολουθεί το χέρι από καρτέ σε καρτέ για να μειωθεί ο υπολογιστικός φόρτος. Για την ακρίβεια, παρέχεται ένα υπομοντέλο τοποθέτησης σημείων του χεριού, με μία εικόνα που έχει κοπεί κατάλληλα, για να επικεντρωθεί στην ακριβή πρόβλεψη των συντεταγμένων. Για να μειωθεί η αυξημένη επεξεργασία των δεδομένων, χρησιμοποιούνται τα προηγούμενα αναγνωρισμένα χέρια για τη δημιουργία του πλαισίου, μέσα στο οποίο φαίνεται το χέρι ή η χειρονομία που πρόκειται να αναγνωριστεί. Στην περίπτωση που το μοντέλο τοποθέτησης σημείων δεν μπορεί να αναγνωρίσει το χέρι, χρησιμοποιείται η ανίχνευση της παλάμης, ώστε να εντοπισθεί το χέρι ξανά. Η παρακολούθηση των σημείων του χεριού υλοποιήθηκε με τον εξειδικευμένο αλγόριθμο χεριών της MediaPipe, που χειρίζεται την απεικόνιση με των σημείων πάνω στο χέρι με ένα Hand – Renderer. Το μοντέλο ανίχνευσης παλάμης χρησιμοποιεί έναν αλγόριθμο τοποθέτησης σημείων χεριού και έναν αλγόριθμο ανίχνευσης παλάμης. Το αποτέλεσμα από το API του MediaPipe, που παρέχεται σε μορφή σημείων αναφοράς, αναλύεται και καταγράφεται σε ένα αρχείο CSV για εκπαίδευση μιας συγκεκριμένης χειρονομίας. Μόλις εκπαιδευτεί με αρκετές χειρονομίες, το μοντέλο εξόδου μπορεί να ανιχνεύσει τις εκπαιδευμένες χειρονομίες από οποιαδήποτε απλή βίντεο ροή σε πραγματικό χρόνο, μέσω της ιστοκάμερας [18].[19].[20].[33].

Εν συνεχεία, θα αναλυθούν λεπτομερώς το θεωρητικό υπόβαθρο του συστήματος, οι λειτουργίες και η αρχιτεκτονική της εφαρμογής, που πραγματοποιήθηκε σε επίπεδο κονσόλας, καθώς και άλλες παρόμοιες ερευνητικές εργασίες και άρθρα που σχετίζονται με το θέμα της διπλωματικής αυτής εργασίας και την τεχνολογία που έχει χρησιμοποιηθεί για την υλοποίηση της [12].[39].

## Κεφάλαιο 2<sup>ο</sup>

### ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ ΚΑΙ ΣΧΕΤΙΚΕΣ ΜΕΛΕΤΕΣ

Στο παρόν κεφάλαιο θα γίνει βιβλιογραφική ανασκόπηση της διπλωματικής εργασίας. Αρχικά θα αναλυθούν θεωρητικά όλες οι τεχνολογίες και τα εργαλεία που αξιοποιήθηκαν για την επιτυχή ολοκλήρωση της ανάπτυξης της εφαρμογής αναγνώρισης χειρονομιών που αντιστοιχούν σε γράμματα του αλφαβήτα της Ελληνικής Νοηματικής Γλώσσας. Επιγραμματικά μερικές από τις έννοιες που θα αναλυθούν παρακάτω είναι τεχνητή νοημοσύνη, μηχανική μάθηση, νευρωνικά δίκτυα και όραση υπολογιστών. Επιπλέον, θα εξεταστούν ερευνητικά άρθρα και μελέτες με παρεμφερές περιεχόμενο και θεωρητικό υπόβαθρο, όπως και παραδείγματα ήδη ανεπτυγμένων εφαρμογών που προσεγγίζουν λύσεις για το ίδιο ή για παρόμοια προβλήματα με διαφορετικούς τρόπους και μεθοδολογίες.

#### 2.1 Θεωρητικό Υπόβαθρο

Μπορούμε πλέον να υποθέσουμε πως η κοινότητα των κωφών εξαρτάται σε πολύ μεγάλο βαθμό από την τεχνολογία, μέσω της οποίας επιτυγχάνεται η ομαλή επικοινωνία με τα υπόλοιπα μέλη της κοινωνίας. Με την συνεχή εξέλιξη της όρασης των υπολογιστών και της μηχανικής μάθησης, εκπονούνται έρευνες που στοχεύουν στην ανάπτυξη αυτοματοποιημένων συστημάτων, ικανά να μετατρέπουν την νοηματική γλώσσα σε μορφή κειμένου ή σε ηχητική μορφή. Στο παρόν κεφάλαιο, θα γίνει η βιβλιογραφική ανασκόπηση της διπλωματικής εργασίας. Ως επί το πλείστον, θα αναλυθούν οι πιο πρόσφατες και καινοτόμες τεχνολογίες, μέθοδοι ερευνητικά άρθρα και μελέτες, που βασίζονται στη μηχανική μάθηση, χρησιμοποιούνται σε συστήματα Αναγνώρισης Νοηματικών Γλωσσών, ευρέως γνωστά ως SLR (Sign Language Recognition) συστήματα, ή και σε άλλα συστήματα που αξιοποιούν παρόμοια τεχνολογικά επιτεύγματα.

##### 2.1.1 Τεχνητή Νοημοσύνη

Ο όρος τεχνητή νοημοσύνη αναφέρεται σε ένα πεδίο τεχνολογίας που αποσκοπεί στην ανάπτυξη λογισμικών ικανά να μοντελοποιήσουν, να αντιγράψουν και να εφαρμόσουν την ανθρώπινη λογική και συμπεριφορά σε πολλά καθημερινά εμπόδια και προβληματισμούς. Αν και η Τεχνητή Νοημοσύνη θεωρείται μία επιστήμη που αναπτύχθηκε σχετικά πρόσφατα, οι ρίζες της χρονολογούνται από τη δεκαετία του 1950. Το 1956, πολλοί σημαντικοί επιστήμονες, όπως οι John McCarthy, Marvin Minsky, Claude Shannon κ.ά. συναντήθηκαν για να συζητήσουν για πρώτη φορά την ιδέα της μίμησης της ανθρώπινης νοημοσύνης, από τεχνητά μέσα. Τα λογισμικά συστήματα που χρησιμοποιούν αλγόριθμους τεχνητής

νοημοσύνης συνήθως στοχεύουν στην ανίχνευση μιας ή περισσότερων από τις ακόλουθες ανθρώπινες νοητικές ικανότητες:

- Μίμηση της ανθρώπινης σκέψης.
- Μίμηση της ανθρώπινης συμπεριφοράς.
- Μίμηση της ανθρώπινης λογικής και συλλογιστικής.
- Μίμηση των ανθρώπινων αντιδράσεων [42].[43].

Με αυτόν τον τρόπο, είναι εύκολο να παρουσιάσουμε μια πιο εξαντλητική περιγραφή της Τεχνητής Νοημοσύνης ως τον τομέα της Επιστήμης των Υπολογιστών, που αποσκοπεί στη σχεδίαση και στην υλοποίηση προγραμμάτων ικανών για να αναπαράγουν ανθρώπινες νοητικές ικανότητες, όπως, μεταξύ άλλων, η επίλυση προβλημάτων, η νοητική αντίληψη μέσω της όρασης, η μάθηση και η εκπαίδευση, η εξαγωγή συμπερασμάτων και η κατανόηση της φυσικής γλώσσας [42].[43].

Η τεχνητή νοημοσύνη εφαρμόζεται σε πολλούς και διαφορετικούς τομείς, οι οποίοι συνεχώς αυξάνονται. Μερικοί τομείς εφαρμογής που περιλαμβάνονται στην Τεχνητή Νοημοσύνη, φαίνονται επιγραμματικά παρακάτω:

- Μοντέλα Προβλέψεων Φαινομένων και Καταστάσεων (π.χ. καιρικών, οικονομικών και κοινωνικών)
- Επεξεργασία της Φυσικής Γλώσσας (π.χ. για τη δημιουργία αυτόματων τηλεφωνητών ειδικά για κέντρα εξυπηρέτησης πελατών, chat bots και συστημάτων κατανόησης και συλλογής δεδομένων κειμένων)
- Σχεδιασμός Ενεργειών και Χρονοπρογραμματισμός (π.χ. για τα προηγμένα συστήματα παραγωγής)
- Προσαρμοζόμενα και εξελισσόμενα ευφυή συστήματα (κατάλληλα για την ευφυή διαχείριση της πληροφορίας)
- Αυτόνομα Ρομποτικά Οχήματα (π.χ. self-driving cars και quadcopters-drones swarm)
- Έμπειρα Συστήματα και Συστήματα Γνώσης (που αποσκοπούν στη βέλτιστη διαχείριση γνώσης)
- Ευφυείς πράκτορες (agents)

- Τεχνητή Όραση (εκπαίδευση μοντέλων για την αναγνώριση και τη μοντελοποίηση φυσικών χώρων και αντικειμένων)
- Ευφυείς υπηρεσίες διαδικτύου και σημασιολογικό διαδίκτυο (π.χ. semantic web)
- Αυτόνομες Ρομποτικές Εφαρμογές (εφαρμογές σε ποικίλους τομείς της βιομηχανίας, εργαστηριακών δοκιμών κτλ)
- Απόδειξη και Θεμελίωση Θεωρημάτων (απαραίτητη για πολύπλοκα θεωρήματα υψηλής μεταβλητότητας δεδομένων)
- Επίλυση προβλημάτων γενικής Φύσεως (σύγχρονα προβλήματα υψηλής αβεβαιότητας)

[23].[42].[43].

Ο Alan Turing, ο γνωστός και ως ο πατέρας της Τεχνητής Νοημοσύνης, συνέλαβε το 1950 το Turing test, μια δοκιμασία για την αναγνώριση ευφύων μηχανών. Αυτή η δοκιμασία περιλαμβάνει την ταυτόχρονη αντιμετώπιση ερωτημάτων από έναν άνθρωπο και έναν υπολογιστή, ώστε μέσω σύγκρισης των αποτελεσμάτων, να αξιολογείται αν το υπολογιστικό σύστημα εμφανίζει ευφυΐα [42].[43].

Μία εφαρμογή της Τεχνητής Νοημοσύνης, η οποία θεωρείται και από τις πιο συνηθισμένες στον σύγχρονο υπολογισμό, είναι η προσομοίωση της ανθρώπινης συλλογιστικής ικανότητας, γνωστή και ως Υπολογιστική Νοημοσύνη. Κύριος σκοπός του τομέα, αποτελεί η έρευνα και η εφαρμογή ανθρώπινων νοητικών και γενετικών λειτουργιών σε υπολογιστικά συστήματα. Δύο από τους πιο μελετημένους και καινοτόμους τομείς είναι οι Γενετικοί Αλγόριθμοι και τα Νευρωνικά Δίκτυα, πανίσχυρα εργαλεία, κατασκευασμένα ειδικά για την υλοποίηση αρχών της Τεχνητής Νοημοσύνης σε υπολογιστικά συστήματα.

Συνοψίζοντας, η Τεχνητή Νοημοσύνη, ένας από τους πιο σημαντικούς και ραγδαία αναπτυσσόμενους τομείς της επιστήμης των υπολογιστών, ασχολείται με την ανάπτυξη λογισμικών ικανά να μιμηθούν ανθρώπινες νοητικές διαδικασίες και να επιλύσουν προβλήματα σε περιβάλλοντα με υψηλή αβεβαιότητα [23].[42].[43].

### **2.1.2 Μηχανική Μάθηση**

Όταν αναφερόμαστε στο πεδίο Μηχανικής Μάθησης (Machine Learning- ML) της πληροφορικής, ορίζουμε την ικανότητα των υπολογιστών να μπορούν να μαθαίνουν από διάφορα ήδη δεδομένων. Αποτελεί ένα υποσύνολο της τεχνητής νοημοσύνης (Artificial Intelligence - AI) που εστιάζει στην χρήση δεδομένων και αλγορίθμων, με σκοπό να προσομοιώσει τον τρόπο εκμάθησης των ανθρώπων,

βελτιώνοντας σταδιακά την ακρίβειά των αλγορίθμων και συνεπώς των μοντέλων Μηχανικής Μάθησης, σε επίπεδο προβλέψεων και αποφάσεων. Το έτος 1959 ο Arthur Samuel, ορίζει την Μηχανική Μάθηση ως “πεδίο σπουδών που δίνει στους υπολογιστές τη δυνατότητα μάθησης χωρίς να έχει προγραμματιστεί ρητά”. Ένας άλλος ορισμός που δόθηκε από Tom Mitchell το 1997 είναι πως “ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από την εμπειρία E σε σχέση με κάποια εργασία T και κάποιο μέτρο απόδοσης P, εάν η απόδοσή του στο T, όπως μετράτε με το P, βελτιώνεται με εμπειρία E”. Ένα μοντέλο μηχανικής μάθησης αποτελείται από αλγορίθμους, οι οποίοι χρησιμοποιώντας δεδομένα, γνωστά και ως δεδομένα εκπαίδευσης, κάνουν προβλέψεις και να λαμβάνουν αποφάσεις χωρίς να είναι προγραμματισμένοι συγκεκριμένα για αυτές τις διεργασίες. Πλέον, πολλές γνωστές εταιρείες κολοσσοί αξιοποιούν αλγόριθμους μηχανικής μάθησης. Μερικές από αυτές είναι η YouTube, η οποία είναι μια πλατφόρμα με βίντεο, κατηγοριοποιώντας τα βίντεο που παρακολουθεί ο κάθε χρήστης, εμφανίζοντας και προωθώντας προτεινόμενα βίντεο, σύμφωνα με την αρέσκεια των χρηστών. Μία ακόμα εταιρεία είναι η Netflix, η οποία με πανομοιότυπο τρόπο κατηγοριοποιεί τις ταινίες, και προτείνει στους χρήστες ταινίες με παρόμοιο περιεχόμενο, τίτλο, υπόθεση – πλοκή και είδος. Τέλος, οι αλγόριθμοι της μηχανικής μάθησης χρησιμοποιούνται και σε άλλους τομείς όπως είναι η υγεία, η εκπαίδευση, η αναγνώριση προσώπων, ομιλίας αντικειμένων και μοτίβων [44].[45].

Όπως αναφέρθηκε προηγουμένως η μηχανική μάθηση καθιστά τους υπολογιστές ικανούς να μαθαίνουν από τα δεδομένα, να προβλέπουν και να λαμβάνουν αποφάσεις. Συγκρίνοντας την με το παραδοσιακό μοντέλο, όπου οι εφαρμογές των υπολογιστών προγραμματίζονται και επαναπρογραμματίζονται από την αρχή, η μηχανική μάθηση δίνει την δυνατότητα στους υπολογιστές να μαθαίνουν αυτόματα από τα δεδομένα χωρίς να είναι απαραίτητος ο επαναπρογραμματισμός τους. Αυτός αποτελεί έναν από τους βασικούς λόγους, όπου η μηχανική μάθηση χρησιμοποιείται σχεδόν παντού στις μέρες μας. Ορισμένα πλεονεκτήματα της μηχανικής μάθησης είναι:

- Αναγνώριση μοτίβων
- Δεν είναι απαραίτητη η ανθρώπινη παρέμβαση
- Είναι ικανή να βελτιώνεται συνεχώς
- Προτείνει λύσεις σε περίπλοκα προβλήματα
- Λειτουργεί και σε περίπλοκα προβλήματα τα οποία χρησιμοποιούν τεράστιο αριθμό δεδομένων

Υπάρχουν τρεις τύποι μηχανικής μάθησης, όπου ο κάθε ένας από αυτούς είναι διαφορετικός. Οι τύποι αυτοί διακρίνονται στην επιβλεπόμενη μάθηση (supervised learning), στην μη-επιβλεπόμενη μάθηση (unsupervised learning) και τέλος στην ενισχυτική μάθηση (reinforcement learning) [23].[44].[45].

Κάθε τύπος είναι διαφορετικός και χρησιμοποιείται με διαφορετικό τρόπο σε κάθε πρόβλημα. Ο κάθε τύπος περιλαμβάνει και κάποιες υποκατηγορίες. Επιπρόσθετα, εκτός από τους τρεις τύπους μηχανικής μάθησης θα παρουσιαστούν και ορισμένοι άλλοι τύποι, οι οποίοι είτε χρησιμοποιούν ένα συνδυασμό από δύο διαφορετικούς τύπους μάθησης είτε ανήκουν σε κάποια κατηγορία από τους βασικούς τύπους [44],[45].

### 2.1.2.1 Επιβλεπόμενη Μάθηση

Η επιβλεπόμενη μάθηση (supervised learning) χρησιμοποιεί δεδομένα εισόδων (που ονομάζονται σύνολο εκπαίδευσης) σε γνωστές επιθυμητές εξόδους. Στόχος της είναι η εύρεση λύσης σε εισόδους που έχουν άγνωστη έξοδο. Στον τύπο αυτό περιλαμβάνονται δύο υποκατηγορίες, η κατηγοριοποίηση (classification) και η παλινδρόμηση (regression) [44].

- **Κατηγοριοποίηση:** Βασικός στόχος της Κατηγοριοποίησης είναι να κατατάσσει τα νέα δεδομένα που προκύπτουν. Ορισμένοι από τους πιο διαδεδομένους αλγόριθμους της Κατηγοριοποίησης είναι οι:
  - K-κοντινότεροι-γείτονες
  - Δέντρα απόφασης
  - Λογιστική παλινδρόμηση
  - Απλοϊκός Bayes
  - Νευρωνικά δίκτυα
  - Τυχαία δάση

Ένα πρόβλημα το οποίο μπορεί να αντιμετωπισθεί με τη διαδικασία της κατηγοριοποίησης είναι τα ανεπιθύμητα ηλεκτρονικά μηνύματα (spam emails). Στο ηλεκτρονικό ταχυδρομείο, κάθε νέο μήνυμα που λαμβάνεται, κατηγοριοποιείται σε ανεπιθύμητα και μη. Το συγκεκριμένο πρόβλημα ανήκει στο πεδίο της δυαδικής κατηγοριοποίησης (binary classification). Δηλαδή τα μηνύματα, μετά την κατηγοριοποίηση παίρνουν δύο τιμές: αληθές (true) και ψευδές (false). Παρά όλα αυτά, υπάρχουν και διακρίσεις όπως η πολλαπλή κατηγοριοποίηση (multiclass classification) όπου κατηγοριοποιεί τα δεδομένα σε περισσότερες από δύο κλάσεις (classes) [44].



- **Παλινδρόμηση:** Στόχος της Παλινδρόμησης αποτελεί η ικανότητα να μπορεί, ο υπολογιστής μέσω προγραμμάτων, να προβλέψει μια συνεχή μεταβλητή δεδομένου κάποιας εισόδου. Ευρέως γνωστοί αλγόριθμοι παλινδρόμησης είναι οι:

- Γραμμική παλινδρόμηση
- Πολυωνυμική γραμμική παλινδρόμηση
- K-κοντινότεροι-γείτονες
- Δέντρα Απόφασης
- Τυχαία δάση

Ένα παράδειγμα προβλήματος που μπορεί να λυθεί με τη διαδικασία της παλινδρόμησης είναι η πρόβλεψη της τιμής ενός σπιτιού. Χρησιμοποιώντας δεδομένα όπως το γεωγραφικό μήκος, γεωγραφικό πλάτος, την ηλικία του σπιτιού, τον αριθμό των δωματίων και τον αριθμό των κρεβατιών καθώς και πολλά άλλα, δίνεται η δυνατότητα να προβλεφθεί η μελλοντική τιμή του [44].

### 2.1.2.2 Μη Επιβλεπόμενη Μάθηση

Η μη επιβλεπόμενη μάθηση (unsupervised learning) σε σύγκριση με την επιβλεπόμενη μάθηση χρησιμοποιεί δεδομένες εισόδους (όπου ονομάζονται σύνολα εκπαίδευσης) σε άγνωστες εξόδους. Το σύνολο εκπαίδευσης δεν ανήκει σε κάποια συγκεκριμένη κλάση, με αποτέλεσμα το σύστημα ουσιαστικά να προσπαθεί να μάθει χωρίς κάποια βοήθεια. Επιπλέον, η μη επιβλεπόμενη μάθηση περιέχει δύο υποκατηγορίες. Οι υποκατηγορίες αυτές είναι γνωστές ως συσταδοποίηση (clustering) και μείωση διαστάσεων (dimensionality reduction) [45].

- **Συσταδοποίηση (Clustering):** Βασικός στόχος της συσταδοποίησης είναι ο κατάλληλος διαχωρισμός του συνόλου δεδομένων εκπαίδευσης, σε ένα σύνολο από υπο-ομάδες (clusters) χωρίς να υπάρχει κάποια προγενέστερη γνώση. Η κάθε υπο-ομάδα (cluster), η οποία προκύπτει κατά την διάρκεια της ανάλυσης, είναι μια ομάδα από αντικείμενα τα οποία έχουν ορισμένα κοινά χαρακτηριστικά. Ορισμένοι γνωστοί αλγόριθμοι συσταδοποίησης (clustering) φαίνονται παρακάτω:

- K-Means
- Πολυωνυμική Hierarchical Clustering

- DBSCAN
- BIRCH
- **Μείωση Διαστάσεων (Dimensionality Reduction):** Η μείωση διαστάσεων στην μηχανική μάθηση αποτελεί ένα μετασχηματισμό (transformation), όπου μετατρέπει τις υψηλές διαστάσεις (high dimensional) σε χαμηλές διαστάσεις (low dimensional). Η μείωση των διαστάσεων χρησιμοποιείται με σκοπό η αναπαράσταση να είναι ευκολότερη στην κατανόηση των ανθρώπων αλλά και των αλγόριθμων της μηχανικής μάθησης. Η εργασία των αλγόριθμων της μηχανικής μάθησης με δεδομένα που ανήκουν σε πίνακες υψηλών διαστάσεων είναι ανεπιθύμητη για πολλούς λόγους. Ένας από τους βασικούς και παράλληλα πολύ σημαντικούς λόγους είναι ότι απαιτείται περισσότερος χώρος αποθήκευσης και παρατηρείται μείωση της υπολογιστικής απόδοσης και ικανότητας των αλγόριθμων της μηχανικής μάθησης [45]. Οι πιο σημαντικοί από τους αλγόριθμους μείωσης διαστάσεων αναφέρονται επιγραμματικά παρακάτω:
  - Principal component analysis (PCA)
  - Linear discriminant analysis (LDA)
  - Non-negative matrix factorization (NMF)
  - Singular value decomposition (SVD)

### 2.1.2.3 Ενισχυτική Μάθηση

Η ενισχυτική μάθηση (reinforcement learning) σε σύγκριση με την επιβλεπόμενη μάθηση και τη μη επιβλεπόμενη μάθηση είναι ένας ξεχωριστός τύπος μηχανικής μάθησης. Το υπολογιστικό σύστημα μαθαίνει μια στρατηγική ενεργειών, καθώς αλληλοεπιδρά με το περιβάλλον. Στην ενισχυτική μάθηση το σύστημα περιέχει τον πράκτορα (agent), το περιβάλλον (environment) και την ενέργεια (action). Το σύστημα εκμάθησης αξιοποιεί τον πράκτορα στο περιβάλλον εκτελώντας ορισμένες ενέργειες. Από την κάθε ενέργεια ο πράκτορας λαμβάνει κάποια ανταμοιβή (reward) ή ποινή (penalty) με βάση τον στόχο του συστήματος. Ο πράκτορας χρειάζεται να μάθει από τον ίδιο του τον εαυτό ποια είναι η καταλληλότερη στρατηγική (strategy), η οποία ονομάζεται πολιτική (policy), με βάση τις ανταμοιβές ή τις τιμές τις οποίες λαμβάνει. Με βάση αυτήν την πολιτική, ο πράκτορας καταλαβαίνει ποια ενέργεια πρέπει να εκτελέσει σε κάθε διαφορετική κατάσταση [44].[45].

#### 2.1.2.4 Άλλοι τύποι μάθησης

Συνεχίζοντας, πέρα από την επιβλεπόμενη μάθηση, την μη επιβλεπόμενη μάθηση και την ενισχυτική μάθηση, δηλαδή τους τρεις τύπους Μηχανικής Μάθησης, που αναφέρθηκαν και αναλύθηκαν παραπάνω, υπάρχουν και ορισμένοι ακόμα τύποι Μηχανικής Μάθησης. Αυτοί οι τύποι Μηχανικής Μάθησης, είτε ανήκουν σε κάποια υποκατηγορία από τους τρεις βασικούς τύπους Μηχανικής Μάθησης, είτε χρησιμοποιούν ένα συνδυασμό από δύο διαφορετικούς τύπους μάθησης. Ένας από τους τύπους μάθησης που συνδυάζουν δύο διαφορετικούς τύπους μάθησης είναι η ημι-επιβλεπόμενη μάθηση. Επιπλέον, άλλοι τύποι Μηχανικής Μάθησης είναι η τεμπέλικη μάθηση (lazy learning) και η πρόθυμη μάθηση (eager learning) [23].[44].[45].

- **Ημι-επιβλεπόμενη Μάθηση (semi-supervised learning):** Η ημι-επιβλεπόμενη μάθηση είναι ένας τύπος Μηχανικής Μάθησης, που βρίσκεται ανάμεσα στην επιβλεπόμενη μάθηση και στη μη-επιβλεπόμενη μάθηση. Ουσιαστικά στην ημι-επιβλεπόμενη μάθηση στο σύνολο εκπαίδευσης ορισμένοι έξοδοι ανήκουν σε κάποια κλάση και κάποιοι έξοδοι έχουν άγνωστη έξοδο, δηλαδή δεν μπορούν να καταταχθούν σε κάποια συγκεκριμένη κλάση [44].[45].
- **Τεμπέλικη Μάθηση (Lazy learning):** Η τεμπέλικη μάθηση είναι μια μέθοδος στην οποία γενικεύονται τα δεδομένα εκπαίδευσης και παράλληλα καθυστερεί μέχρι να ερωτηθεί το σύστημα. Ονομάζονται τεμπέλικη (lazy), επειδή περιμένουν όσο το δυνατόν περισσότερο μέχρι να δημιουργηθεί κάποιο μοντέλο. Ένα από τα θετικά του συγκεκριμένου τύπου μάθησης είναι η ταχύτητα με την οποία μαθαίνει. Προϋποθέτει πως υπάρχει μεγάλος χώρος για την αποθήκευση όλων των δεδομένων εκπαίδευσης, ταξινομώντας τα όμως με αργό ρυθμό. Ένας αλγόριθμος μηχανικής μάθησης ο οποίος χρησιμοποιεί τεμπέλικη μάθηση είναι ο K-κοντινότεροι-γείτονες (KNearest Neighbors - KNN) [44].[45].
- **Πρόθυμη Μάθηση (Eager learning):** Σε αντίθεση με την τεμπέλικη μάθηση που προαναφέρθηκε, στην πρόθυμη μάθηση, το υπολογιστικό σύστημα κατά την διάρκεια της εκπαίδευσης του, προσπαθεί να δημιουργήσει έναν ανεξάρτητο στόχο εισόδου. Ένα κύριο πλεονέκτημα της μάθησης αυτής, είναι ότι απαιτεί λιγότερο χώρο για την αποθήκευση των δεδομένων εκπαίδευσης, σε σύγκριση με την τεμπέλικη μάθηση. Όμως, παρόλα αυτά η δημιουργία του μοντέλου είναι εξίσου αργή. Μερικοί αλγόριθμοι μηχανικής μάθησης οι οποίοι κατατάσσονται στην πρόθυμη μάθηση είναι ο Support Vector Machines και τα Neural Networks [23].[44].[45].

### 2.1.3 Βαθιά Μάθηση

Η Βαθιά Μάθηση (Deep Learning) αντιπροσωπεύει μια μέθοδο εκπαίδευσης υπολογιστών που αποσκοπεί στην αυτόνομη εκμάθηση, δηλαδή να μαθαίνουν οι υπολογιστές μέσω παραδειγμάτων, όπως ακριβώς και ο άνθρωπος. Λόγω της συνεχής ανάπτυξης της τεχνολογίας, αυτό πλέον, συμβαίνει χωρίς την ανάγκη προσδιορισμένων - συγκεκριμένων προτύπων γνώσης. Ουσιαστικά, μέσω της βαθιάς μάθησης, προσομοιάζεται η πολυ-επίπεδη εξελικτική διαδικασία του ανθρώπινου εγκεφάλου, επιτρέποντάς στις υπολογιστικές μηχανές, να αντιμετωπίζουν με επιτυχία προκλήσεις από τα δεδομένα που λαμβάνουν, τα οποία πολλές φορές είναι πολύπλοκα και αμυδρά. Στο πεδίο της Βαθιάς Μάθησης, μαθαίνουν τα υπολογιστικά συστήματα την διαδικασία ταξινόμησης – κατηγοριοποίησης, απευθείας από τα εισερχόμενα δεδομένα (όπως εικόνες, βίντεο, κείμενο, ήχος κλπ). Τα μοντέλα Βαθιάς Μάθησης είναι ικανά να πετύχουν ποσοστά ακρίβειας που συχνά υπερβαίνουν την ανθρώπινη επίδοση. Η Βαθιά Μάθηση επιτυγχάνεται με την εκπαίδευση μοντέλων, χρησιμοποιώντας ένα μεγάλο σύνολο τιτλοφορημένων δεδομένων, σε συνδυασμό με πολυ-επίπεδα Νευρωνικά Δίκτυα [3].[24].[24].

Αν και διατυπώθηκε και προτάθηκε τη δεκαετία του 1980, η τεχνολογία της Βαθιάς Μάθησης έγινε πραγματικότητα μόνο πρόσφατα, καθώς απαιτεί:

- Τη χρήση τεράστιου όγκου δεδομένων, που είναι πλέον διαθέσιμα χάρη στην ραγδαία αύξηση των αισθητήρων, τη διεύρυνση του αποθηκευτικού χώρου και την μεγάλη αύξηση της ταχύτητας μετάδοσης των δεδομένων [25].
- Σημαντική υπολογιστική ισχύ, για την παράλληλη επεξεργασία του μεγάλου όγκου των δεδομένων, με υψηλή ταχύτητα. Η ανάγκη παράλληλης επεξεργασίας δεδομένων, απαιτεί την αξιοποίηση συστημάτων παράλληλης επεξεργασίας, έναντι της σειριακής επεξεργασίας. Αυτό μπορεί να επιτευχθεί κυρίως με χρήση GPU αρχιτεκτονικής, ιδίως όταν η συγκεκριμένη αρχιτεκτονική συνδυάζεται με τεχνικές συσταδοποίησης (clusters) ή πρόκειται να ενσωματωθεί στο υπολογιστικό νέφος (cloud computing), ώστε να μειωθεί ο χρόνος εκπαίδευσης ενός δικτύου Βαθιάς Μάθησης [25].

Η τεχνική της Βαθιάς Μάθησης βασίζεται στη χρήση πολυ-επίπεδων Νευρωνικών Δικτύων, γνωστών και ως Βαθιά Νευρωνικά Δίκτυα. Ο όρος "Βαθιά" συνδέεται συνήθως με τον αριθμό των κρυφών επιπέδων στο Νευρωνικό Δίκτυο. Τα παραδοσιακά Νευρωνικά Δίκτυα έχουν συνήθως μόνο 2-3 κρυφά επίπεδα, ενώ τα Βαθιά δίκτυα μπορούν να έχουν ακόμη και 150. Τα μοντέλα Βαθιάς Μάθησης εκπαιδεύονται αξιοποιώντας μεγάλα σύνολα δεδομένων και πολυ-επίπεδα Νευρωνικά Δίκτυα που μαθαίνουν τις λειτουργίες απευθείας από τα δεδομένα, δίχως να χρειάζεται χειροκίνητη εξαγωγή των χαρακτηριστικών [28].[32].

Από τα πιο δημοφιλή και ευρέως γνωστά είδη βαθιών Νευρωνικών Δικτύων είναι τα Συνελικτικά Νευρωνικά Δίκτυα (CNN ή ConvNet).

Τα Συνελικτικά Νευρωνικά Δίκτυα περιγράφουν τις γνωστές λειτουργίες με δεδομένα εισόδου, χρησιμοποιώντας δυσδιάστατα στρώματα περιστροφής, καθιστώντας την αρχιτεκτονική αυτή κατάλληλη για την επεξεργασία δυσδιάστατων δεδομένων, όπως είναι τα δεδομένα που εξάγονται από εικόνες. Επιπλέον, εξαλείφουν την ανάγκη για χειροκίνητη εξαγωγή χαρακτηριστικών και αυτόματα αναγνωρίζουν χαρακτηριστικά και μοτίβα, από τα εισερχόμενα δεδομένα, που όπως προαναφέρθηκε έχουν ποικίλες μορφές, χωρίς να χρειάζεται να προ-καταχωρηθούν στο σύστημα, αφού το δίκτυο εκπαιδεύεται πάνω σε ένα σύνολο δεδομένων που λαμβάνει διαρκώς. Λόγω αυτής της ικανότητας τα μοντέλα Βαθιάς Μάθησης, θεωρούνται αξιόπιστα και κατάλληλα για λειτουργίες και εφαρμογές που περιλαμβάνουν τεχνικές Υπολογιστικής Όρασης, επεξεργασίας και ταξινόμησης δεδομένων [6].[26].

Τέλος, τα Συνελικτικά Νευρωνικά Δίκτυα μαθαίνουν να ανιχνεύουν και να αναγνωρίζουν διάφορα χαρακτηριστικά ενός αντικειμένου – μίας δομής, χρησιμοποιώντας δεκάδες ή εκατοντάδες κρυφά επίπεδα. Κάθε κρυφό επίπεδο αυξάνει την πολυπλοκότητα των χαρακτηριστικών της γνώσης της εικόνας. Δηλαδή, το πρώτο κρυφό επίπεδο ενδέχεται να μάθει να ανιχνεύει τις άκρες, ενώ το τελευταίο μπορεί να μάθει να ανιχνεύει πιο σύνθετα και αμυδρά σχήματα, ιδίως στις περιπτώσεις αναγνώρισης αντικειμένων [25].[28].[34].

#### 2.1.4 Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα αποτελούν ένα υπολογιστικό μοντέλο που χρησιμοποιείται για την επίλυση προβλημάτων τεχνητής νοημοσύνης. Αυτά τα δίκτυα εμπνέονται από τη δομή των νευρώνων του ανθρώπινου εγκεφάλου. Η πρώτη τους εφαρμογή, χρονολογείται το 1944 από τους ερευνητές Warren McCulloch και Walter Pitts. Χρησιμοποιούνται για προβλήματα κατηγοριοποίησης και παλινδρόμησης.

**Δομή των Νευρωνικών Δικτύων:** Τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks - ANN) αποτελούνται από τρία επίπεδα: το **επίπεδο εισόδου** (input layer), τα **κρυμμένα επίπεδα** (hidden layers) και το **επίπεδο εξόδου** (output layer). Το επίπεδο εισόδου (input layer) λαμβάνει την είσοδο ή τις εισόδους του τεχνητού νευρωνικού δικτύου και έπειτα επικοινωνεί με τα κρυμμένα επίπεδα [16]. Ο αριθμός των κρυφών επιπέδων (hidden layers) ποικίλλει. Τα κρυμμένα επίπεδα (hidden layers) συνδέονται με το επίπεδο εξόδου (output layer). Τέλος το επίπεδο εξόδου (output layer), με τη σειρά του, παράγει την έξοδο του δικτύου. Οι εισοδοί του τεχνητού νευρωνικού δικτύου,  $x_i$  μπορεί να είναι πολλές. Κάθε είσοδος  $x_i$ , συνδέεται με ένα βάρος (weight)  $w_i$  και τα αποτελέσματα υπολογίζονται με μια **συνάρτηση αθροίσματος** (summation function) [15].[16].[21].

**Συνάρτηση Αθροίσματος:** Η συνάρτηση αθροίσματος (summation function) υπολογίζει την είσοδο των ενδιάμεσων νευρώνων και περιλαμβάνει ένα κατώφλι (threshold value)  $\theta$  [16]. Για την παραγωγή της εξόδου του τεχνητού νευρωνικού δικτύου (Artificial Neural Network), η τιμή κατωφλίου (threshold value)  $\theta$  πρέπει να είναι μικρότερη από τη συνάρτηση μετάβασης (transfer function) [16], δηλαδή θα πρέπει να ισχύει  $I_j > \theta$ . Η συνάρτηση αθροίσματος υπολογίζεται από τον παρακάτω τύπο:

$$I_j = \sum_{i=1}^n x_{ij} w_{ij}$$

*Εξίσωση 2-1: Συνάρτηση Αθροίσματος*

**Τεχνητός Νευρώνας Perceptron:** Ο τεχνητός νευρώνας Perceptron δέχεται  $n$  εισόδους  $x$ , οι οποίες παράγονται μέσω της συνάρτησης μετάβασης  $g$ , και παράγει την έξοδο  $y$  [16]. Η συνάρτηση μετάβασης υπολογίζεται χρησιμοποιώντας τον παρακάτω τύπο:

$$y_j = g(I) = g\left(\sum_{i=1}^n x_{ij} w_{ij}\right)$$

*Εξίσωση 2-2: Συνάρτηση Μετάβασης*

Ορισμένες από τις διαθέσιμες συναρτήσεις μετάβασης είναι η βηματική συνάρτηση ή συναρτήσεις κατωφλίου (threshold function), με έξοδο 0 ή 1, η συνάρτηση πρόσημου (sign function), με έξοδο -1 ή +1 και η συνάρτηση αναρρίχησης (ramping function), όπου η έξοδος είναι 0 έως 1.

**Αρχιτεκτονικές Νευρωνικών Δικτύων:** Υπάρχουν δύο βασικές αρχιτεκτονικές νευρωνικών δικτύων, η **πρόσθια τροφοδότηση** (feed forward) και η **οπίσθια τροφοδότηση** (feed backward). Στα **νευρωνικά δίκτυα πρόσθιας τροφοδότησης**, η έξοδος μιας μονάδας αποτελεί την είσοδο της επόμενης μονάδας. Αυτό σημαίνει ότι οι μονάδες οργανώνονται σε διαφορετικά επίπεδα, έτσι ώστε οι μονάδες του πρώτου επιπέδου να τροφοδοτούν τις μονάδες του επόμενου επιπέδου, μέχρι να φτάσουν στις μονάδες του τελευταίου επιπέδου, του επιπέδου εξόδου [6].[16].[22].[29].

Τα **νευρωνικά δίκτυα οπίσθιας τροφοδότησης** (feed backward), διαχωρίζονται σε δύο ξεχωριστές κατηγορίες, τις αυτό-συσχετιζόμενες μνήμες (autoassociated memories) και τις ετερο-συσχετιζόμενες μνήμες (heteroassociated memories). Στις αυτό-συσχετιζόμενες μνήμες, η ανατροφοδότηση γίνεται μόνο στους κόμβους του ίδιου επιπέδου. Ενώ παράλληλα, στα νευρωνικά δίκτυα οπίσθιας τροφοδότησης με ετερο-συσχετιζόμενες μνήμες, η ανατροφοδότηση γίνεται σε όλα τα επίπεδα. Η παρακάτω εικόνα από το βιβλίο παρουσιάζει αυτά τα δύο νευρωνικά δίκτυα οπίσθιας τροφοδότησης [2].[30].[31].

Τα **νευρωνικά δίκτυα πολλών επιπέδων** ανήκουν στην κατηγορία της πρόσθιας τροφοδότησης. Τα νευρωνικά αυτά δίκτυα αποτελούνται από ένα επίπεδο εισόδου, το οποίο μπορεί να περιλαμβάνει έναν ή και περισσότερους κόμβους, έναν ή και περισσότερα κρυφά επίπεδα, και τέλος ένα μοναδικό επίπεδο εξόδου. Τα επίπεδα που βρίσκονται κοντά στο επίπεδο εισόδου ονομάζονται κατώτερα επίπεδα, ενώ αυτά που βρίσκονται κοντά στο επίπεδο εξόδου ονομάζονται ανώτερα επίπεδα [29].[30].[31].

Οι κόμβοι, σε κάθε επίπεδο, μπορεί να συνδέονται με δύο διαφορετικούς τρόπους: 1) **πλήρως συνδεδεμένους κόμβους** (fully connected), 2) **μερικώς συνδεδεμένους κόμβους** (partially connected). Στην πλήρως συνδεδεμένη διάταξη, κάθε κόμβος του ενός επιπέδου συνδέεται με όλους τους κόμβους του επόμενου επιπέδου, ενώ στη μερικώς συνδεδεμένη, κάθε κόμβος συνδέεται μόνο με έναν υποσύνολο του συνολικού αριθμού των κόμβων του επόμενου επιπέδου [16].[29].[30].

### 2.1.5 Συνάρτηση Κόστους ή Απώλειας (Loss Function)

Η συνάρτηση απώλειας, γνωστή επίσης και ως συνάρτηση κόστους, αποτελεί έναν δείκτη του μεγέθους της απόκλισης των αποτελεσμάτων αναγνώρισης ή πρόβλεψης από τις πραγματικές τιμές, κατά τη διάρκεια της εκπαίδευσης του μοντέλου. Η συνάρτηση αποτελεί ένα μέτρο αξιολόγησης του σφάλματος, απαραίτητη για την καλή προσαρμογή των παραμέτρων του μοντέλου με σκοπό την ελαχιστοποίηση, όσο το δυνατόν περισσότερο, του σφάλματος αυτού. Επομένως, αντιμετωπίζεται ένα πρόβλημα μαθηματικής βελτιστοποίησης, δηλαδή αναζητείται μία λύση για τη μείωση του κόστους [16].

Στον τομέα της Στατιστικής, αυτές οι συναρτήσεις κόστους χρησιμοποιούνται για την εκτίμηση των παραμέτρων των μοντέλων. Ένα συνηθισμένο παράδειγμα εκτίμησης παραμέτρων αποτελεί η γραμμική παλινδρόμηση, στην οποία χρησιμοποιείται το Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error ή MSE). Σε προβλήματα, όπου λαμβάνει χώρα η διαδικασία πρόβλεψης πραγματικών αριθμών, η μεταβλητή  $y$  αντιστοιχεί στις τιμές που παράγει το μοντέλο, ενώ με  $\hat{y}$  συμβολίζουμε τις πραγματικές τιμές των δεδομένων [16].

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

*Εξίσωση 2-3: Συνάρτηση Μέσου Τετραγωνικού Σφάλματος*

Ένα άλλο παράδειγμα είναι η Διωνυμική Διασταυρωμένη Εντροπία (Binary Cross Entropy ή BCE), η οποία εφαρμόζεται σε προβλήματα διωνυμικής κατηγοριοποίησης [16].

$$BCE = - \sum_{i=1}^2 y_i * \log \log (P(y_i))$$

*Εξίσωση 2-4: Συνάρτηση Διωνυμικής Διασταυρωμένης Εντροπίας*

Για τα προβλήματα κατηγοριοποίησης πολλαπλών κλάσεων, χρησιμοποιούμε την Κατηγορική Διασταυρωμένη Εντροπία (Categorical Cross Entropy ή CCE).[16].

$$CCE = - \sum_{i=1}^n y_i * \log \log (P(y_i))$$

*Εξίσωση 2-5: Συνάρτηση Κατηγορικής Διασταυρωμένης Εντροπίας*

Συνοψίζοντας, βελτιστοποιώντας τη συνάρτηση κόστους, υπολογίζονται οι βέλτιστες τιμές για τις παραμέτρους του μοντέλου, ώστε να επιτευχθεί η μέγιστη δυνατή ακρίβεια, και να λειτουργεί το εκπαιδευμένο μοντέλο όπως ακριβώς επιθυμεί ο κατασκευαστής του [16].

### 2.1.6 Συναρτήσεις Ενεργοποίησης (Activation Functions)

Οι συναρτήσεις ενεργοποίησης είναι κρίσιμες στα νευρωνικά δίκτυα και χρησιμοποιούνται για την υπολογιστική διαδικασία των νευρώνων. Αυτές οι συναρτήσεις, επιτρέπουν στους νευρώνες να υπολογίζουν την τελική τιμή εξόδου (output), συγκρίνοντας την τιμή αυτή με ένα κατώφλι (threshold). Με αυτόν το τρόπο, καθορίζεται εάν ο νευρώνας θα ενεργοποιηθεί και θα διαβιβάσει την έξοδό του στον επόμενο νευρώνα, ως είσοδο, ή αν θα παραμείνει ανενεργός. Υπάρχουν διάφορες συναρτήσεις ενεργοποίησης που χρησιμοποιούνται, και χωρίζονται σε δύο βασικές κατηγορίες, τις γραμμικές και τις μη-γραμμικές. Πρακτικά, οι μη-γραμμικές συναρτήσεις ενεργοποίησης είναι πιο χρήσιμες και μπορούν να αξιοποιηθούν σε περισσότερα προβλήματα [21].

Ακολουθούν συνοπτικά ορισμένες από αυτές:

#### Λογιστική Συνάρτηση (Logistic Regression):

- Η λογιστική συνάρτηση δίνεται από τον τύπο:

$$f(x) = \frac{1}{1 + e^{-kx}}$$

*Εξίσωση 2-6: Λογιστικής Συνάρτησης*

Για την προεπιλεγμένη τιμή του  $k = 1$ , υπολογίζεται η τυπική λογιστική συνάρτηση η οποία παράγει μία καμπύλη με σχήμα S-σχήμα, με πεδίο τιμών από 0 έως 1 και χρησιμοποιείται συχνά σε προβλήματα δυαδικής κατηγοριοποίησης και για τη μοντελοποίηση πιθανοτήτων. Αρχικά, χρησιμοποιήθηκε για τη μοντελοποίηση της αύξησης του πληθυσμού, όπου ο ρυθμός αναπαραγωγής είναι ανάλογος με τον τρέχοντα αριθμό του πληθυσμού και τον αριθμό των διαθέσιμων πόρων. Επιπλέον, χρησιμοποιείται ευρέως στον τομέα της Στατιστικής, όπου έχουμε την γνωστή Λογιστική Παλινδρόμηση, η οποία μοντελοποιεί την πιθανότητα ενός γεγονότος να συμβεί ανάλογα με τις τιμές των σχετικών μεταβλητών [21].[40].



### Συνάρτηση Υπερβολικής Εφαπτομένης (Hyperbolic Tangent - Tanh):

- Η συνάρτηση της υπερβολικής εφαπτομένης, γνωστή από την τριγωνομετρία, δίνεται από τον τύπο:

$$f(x) = \tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

*Εξίσωση 2-7: Συνάρτηση Υπερβολικής Εφαπτομένης*

Αυτή η συνάρτηση παράγει μια καμπύλη με S-σχήμα, με πεδίο τιμών από -1 έως 1 και χρησιμοποιείται συχνά σε νευρωνικά δίκτυα για προβλήματα διωνυμικής κατηγοριοποίησης [21].[40].

### Γραμμικές Μονάδες Ανόρθωσης (Rectified Linear Units - ReLU):

- Η συνάρτηση γραμμικής μονάδας ανόρθωσης, είναι απλή και δίνεται από τον τύπο:

$$f(x) = \max(0, x)$$

*Εξίσωση 2-8: Συνάρτηση Γραμμικής Μονάδας Ανόρθωσης*

Έχει πεδίο τιμών ίδιο με το πεδίο ορισμού, όταν η είσοδος είναι θετικός αριθμός, και μηδέν όταν είναι αρνητικός. Δηλαδή εάν η είσοδος είναι θετική, το αποτέλεσμα βρίσκεται εντός του πεδίου ορισμού της εισόδου, ενώ αν είναι αρνητική, το αποτέλεσμα είναι ο αριθμός μηδέν. Αυτή η συνάρτηση θεωρείται από τις πιο δημοφιλείς συναρτήσεις ενεργοποίησης και χρησιμοποιείται σε πολλά νευρωνικά δίκτυα για την επίλυση ποικίλων προβλημάτων. Ένα θέμα που προκύπτει με τη συνάρτηση γραμμικής μονάδας ανόρθωσης είναι όταν, καθώς απεικονίζονται όλες οι αρνητικές τιμές στο μηδέν, ίσως να προκληθεί αδράνεια στους νευρώνες καθώς δεν θα ενεργοποιούνται σχεδόν καθόλου. Για την αντιμετώπιση του προβλήματος αυτού, υπάρχουν και οι επεκτάσεις της συνάρτησης. Στις περιπτώσεις αυτές, αντί για μηδέν εισάγουμε μια πολύ μικρή τιμή (Leaky ReLU) ή μια σταθερά (Randomized Leaky ReLU) [21].[40].

### Συνάρτηση Ενεργοποίησης Swish:

- Η συνάρτηση ενεργοποίησης Swish είναι μια από τις πολλές συναρτήσεις ενεργοποίησης που χρησιμοποιούνται στη μηχανική μάθηση και τη βαθιά μάθηση. Παρουσιάστηκε αρχικά στο άρθρο "Searching for Activation Functions" από τους Ramachandran, Zoph, και Le το 2017. Είναι μια συνάρτηση ενεργοποίησης που χρησιμοποιείται κυρίως σε νευρωνικά δίκτυα, και έχει την ακόλουθη μορφή:

$$\text{Swish}(x) = x * \sigma(x)$$

*Εξίσωση 2-9: Συνάρτηση Ενεργοποίησης Swish*

Όπου  $\sigma(x)$ , η συνάρτηση σιγμοειδούς ( $\sigma$ ), που ορίζεται ως:

$$S(x) = \frac{1}{1 + e^{-x}}$$

*Εξίσωση 2-10: Σιγμοειδής Συνάρτηση*

Το κυρίως χαρακτηριστικό της συνάρτησης Swish είναι ότι είναι μια συνεχώς παραγωγίσιμη συνάρτηση, παρόμοια με την συνάρτηση ReLU (Rectified Linear Unit), αλλά πιο ομαλή. Αυτό την καθιστά κατάλληλη για χρήση σε νευρωνικά δίκτυα που χρησιμοποιούν τη μέθοδο οπισθοδιάδοσης (backpropagation) για την εκπαίδευσή τους. Επιπλέον, η συνάρτηση αυτή είναι γραμμική κοντά στο κέντρο της, δηλαδή πλησιάζει την ταυτοτική συνάρτηση στο μηδέν «0», και μη-γραμμική στις μεγαλύτερες απόλυτες τιμές της εισόδου. Η Swish έχει την ιδιότητα να επιτρέπει στο δίκτυο να μάθει νευρώνες που είναι πιο ευαίσθητοι σε περιοχές των δεδομένων με μεγάλη τιμή, ενώ παράλληλα διατηρεί μια μετρική ομαλότητα στον υπολογισμό των παραγώγων, κάτι που μπορεί να βοηθήσει στην εκπαίδευση [21].[40].

### **Συνάρτηση Ενεργοποίησης GeLU (Gaussian Error Linear Unit):**

- Η συνάρτηση ενεργοποίησης GeLU είναι μια συνάρτηση που χρησιμοποιείται συχνά στη μηχανική και τη βαθιά μάθηση ως μια εναλλακτική συνάρτηση ενεργοποίησης σε νευρωνικά δίκτυα. Η συνάρτηση GeLU σχεδιάστηκε για να βελτιώσει την απόδοση των νευρωνικών δικτύων σε σχέση με άλλες συναρτήσεις ενεργοποίησης. Παρακάτω θα αναλυθεί η συνάρτηση GeLU λεπτομερώς. Η μαθηματική μορφή της συνάρτησης είναι η ακόλουθη:

$$\text{GeLu} = \frac{x}{2} (1 + \text{tahn} (\sqrt{\frac{\pi}{2}}(x + 0,044715 * x^3)))$$

*Εξίσωση 2-11: Συνάρτηση Ενεργοποίησης GeLU*

Η GeLU είναι μια συνάρτηση που είναι και πεπερασμένη (bounded) και παραγωγίσιμη, κάτι που την καθιστά κατάλληλη για χρήση σε νευρωνικά δίκτυα. Η πρώτη παράγωγός της GeLU μπορεί να υπολογιστεί αναλυτικά, κάτι που είναι σημαντικό για την ανάδραση προς τα πίσω (backpropagation) κατά την εκπαίδευση του δικτύου. Η GeLU συμπεριφέρεται πολύ καλά σε σχέση με την αποφυγή του προβλήματος της "αποσβολωμένης" αντίδρασης (dying ReLU) που παρουσιάζουν ορισμένες άλλες

συναρτήσεις ενεργοποίησης. Αυτό σημαίνει ότι η GeLU διατηρεί μεγαλύτερη ποικιλία στις εξόδους των νευρώνων του δικτύου [21].[40].

Η συνάρτηση χρησιμοποιείται κυρίως σε πλήρως συνδεδεμένα (fully connected) επίπεδα και σε συνελκτικά (convolutional) επίπεδα νευρωνικών δικτύων.

Συνολικά, η συνάρτηση ενεργοποίησης GeLU είναι μια πολύ χρήσιμη επιλογή για τη μηχανική και τη βαθιά μάθηση, καθώς βοηθά στη βελτίωση της απόδοσης των νευρωνικών δικτύων και στην αποφυγή προβλημάτων που μπορεί να προκύψουν με άλλες συναρτήσεις ενεργοποίησης [21].[40].

### **Κανονικοποιημένη Εκθετική Συνάρτηση (SoftMax):**

Η κανονικοποιημένη εκθετική συνάρτηση, αποτελεί μία γενίκευση της λογιστικής συνάρτησης σε περισσότερες διαστάσεις και χρησιμοποιείται ως συνάρτηση ενεργοποίησης, συνήθως στο τελικό επίπεδο των νευρωνικών δικτύων, για προβλήματα κατηγοριοποίησης πολλαπλών κλάσεων. Δέχεται ως είσοδο πολυδιάστατα διανύσματα και τα κανονικοποιεί, μετατρέποντας τα σε διανύσματα ίδιας διάστασης, που αποτελούνται από πιθανότητες. Σύμφωνα με τις πιθανότητες αυτές γίνεται η κατηγοριοποίηση, και επιλέγεται η έξοδος με τη μεγαλύτερη πιθανότητα, εντάσσοντας τα δεδομένα σε μία τελική κατηγορία. Παρακάτω φαίνεται ο μαθηματικός τύπος:

$$f(z) = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}}$$

*Εξίσωση 2-12: Κανονικοποιημένη Εκθετική Συνάρτηση*

Όπου  $k$  είναι το πλήθος των διαστάσεων των δεδομένων, και  $z \in \mathbb{R}^k$ .

Κάθε συνάρτηση ενεργοποίησης έχει τις δικές της χρήσεις και πλεονεκτήματα. Η επιλογή εξαρτάται κυρίως από το πρόβλημα που έχει τεθεί προς επίλυση και την αρχιτεκτονική του νευρωνικού δικτύου έχει επιλεγεί [21].[40].

### **2.1.7 Όραση Υπολογιστών**

Ένας από τους δημοφιλείς τομείς της Τεχνητής Νοημοσύνης αφορά την Υπολογιστική Όραση, γνωστή και ως Όραση των Υπολογιστών (Computer Vision). Ο τομέας αυτός αναπτύχθηκε για την αποτελεσματική ανάλυση, και συνεπώς αναγνώριση, του περιεχομένου μιας εικόνας ή μίας ακολουθίας εικόνων (βίντεο) από ένα υπολογιστικό σύστημα. Με την τεχνολογία αυτή, οι ερευνητές προσπαθούν να προσομοιώσουν την αίσθηση της όρασης, όπως την αντιλαμβάνονται οι βιολογικοί οργανισμοί, στα υπολογιστικά συστήματα. Στη σύγχρονη εποχή, και με την ταχύτατη πρόοδο της τεχνολογίας, αυτός ο

τομέας βιώνει έντονη ανάπτυξη και συχνά συνδυάζεται με αλγόριθμους Μηχανικής Μάθησης. Η Όραση των Υπολογιστών έχει εφαρμογές σε πολλούς επιστημονικούς τομείς, συμπεριλαμβανομένων της Ιατρικής, της Ρομποτικής, της Νευροβιολογίας και της Γεωργίας [5].[19].

Οι αλγόριθμοι της Μηχανικής Μάθησης επιτρέπουν την γρήγορη και ουσιαστική αντιμετώπιση πολύπλοκων προβλημάτων. Εμφανίζονται καθημερινά καινούρια μοντέλα και τεχνικές, με την ακρίβεια των αλγορίθμων να αυξάνεται σταδιακά. Ένα υποσύνολο της Μηχανικής Μάθησης είναι η Βαθιά Μάθηση, που χρησιμοποιεί βαθιά Νευρωνικά Δίκτυα για την επίλυση προβλημάτων. Παραδείγματος χάριν, τα Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks ή αλλιώς CNN) αποτελούν μία από τις πιο δημοφιλείς αρχιτεκτονικές, που παρουσιάζει εξαιρετικά αποτελέσματα στην Υπολογιστική Όραση και θεωρείται η πιο κατάλληλη επιλογή για την ανάλυση εικόνων και βίντεο [5].[19].

Χρησιμοποιώντας τις παραπάνω αρχιτεκτονικές, μας δίνετε η δυνατότητα να δημιουργούμε μοντέλα με εξαιρετικές ικανότητες γενίκευσης της γνώσης. Τα εκπαιδευμένα αυτά μοντέλα μπορούν να επαναχρησιμοποιηθούν σε παρόμοια προβλήματα με πολύ μικρές προσαρμογές. Για παράδειγμα, στην αναγνώριση αντικειμένων από εικόνες, αυτά τα νευρωνικά δίκτυα μαθαίνουν γενικές διαδικασίες για τον εντοπισμό αντικειμένων και αναμένουν εκπαίδευση από το χρήστη για να προσαρμοστούν στα συγκεκριμένα προβλήματα [5].[19].[37].[38].

Όπως προαναφέραμε, η όραση υπολογιστών εφαρμόζεται σε ένα μεγάλο εύρος, που αποτελείται από πολλούς και διαφορετικούς τομείς, τους οποίους θα αναλύσουμε παρακάτω:

- **Ιατρική**

- Ένα αξιοσημείωτο παράδειγμα εφαρμογής της όρασης των υπολογιστών αφορά την ανάλυση ιατρικών εικόνων, όπως ακτινογραφίες και αποτελέσματα μαγνητικής τομογραφίας. Σε ιατρικές διαγνώσεις και παρακολούθηση πολλών παθήσεων, οι τεχνικές υπολογιστικής όρασης, έχουν αποδειχθεί εξίσου αποτελεσματικές με τις διαγνώσεις που πραγματοποιούνται από γιατρούς. Ιατροί, από διάφορες ειδικότητες, χρησιμοποιούν αυτά τα εργαλεία για την ανάλυση πολυδιάστατων εικόνων, εντοπίζοντας καλοήθεις και κακοήθεις όγκους, αλλά και άλλες δυνητικές σωματικές ανωμαλίες [37].[38].

- **Αυτό-οδηγούμενα οχήματα**

- Τα τελευταία χρόνια, οι βιομηχανίες αυτοκινήτων έχουν δώσει έμφαση στη δημιουργία αυτό-οδηγούμενων οχημάτων που μπορούν να λειτουργούν χωρίς την παρέμβαση ανθρώπινου παράγοντα. Τα οχήματα εφοδιάζονται με κάμερες και αισθητήρες και είναι

ικανά να αναγνωρίζουν αυτόματα οδικές λωρίδες, φωτεινούς σηματοδότες, πεζούς ακόμη και να προβλέπουν ατυχήματα, υπολογίζοντας τις πιθανές τροχιές των περιβαλλοντικών αντικειμένων. Με την ευρεία χρήση αυτών των έξυπνων οχημάτων αναμένεται να επηρεαστούν σημαντικά διάφοροι τομείς της κοινωνίας [37].[38].

- **Στρατιωτικές εφαρμογές**

- Η υπολογιστική όραση έχει ενσωματωθεί σε ποικίλες στρατιωτικές εφαρμογές, όπως είναι τα διάφορα οπτικά συστήματα. Σε αυτά περιλαμβάνονται η οπτική αναγνώριση αντίπαλων στρατευμάτων και οχημάτων, η καθοδήγηση πυραύλων και η αυτόματη ανάλυση περιοχών, που φέρουν κάποιο ενδιαφέρον. Τα μη-επανδρωμένα αεροσκάφη (drones) που χρησιμοποιούν πολλά στρατεύματα, είναι εξοπλισμένα με συστήματα αναγνώρισης προσώπων και αντικειμένων. Επιπλέον, είναι ικανά να συλλέγουν αυτόματα χρήσιμα δεδομένα, χωρίς ανθρώπινη επίβλεψη [37].[38].

- **Αυτοματισμοί σε γραμμές παραγωγής**

- Πολλές βιομηχανίες παραγωγής αξιοποιούν τεχνικές όρασης υπολογιστών για πληθώρα εργασιών. Οι γραμμές παραγωγής πλέον εφοδιάζονται με ρομποτικά συστήματα που εκτελούν αυτόματες διεργασίες, με ενσωματωμένες κάμερες. Καθιστώντας τα κατάλληλα μέλη ενός εργοστασίου ικανά να ελέγχουν αν οι εργαζόμενοι φορούν τον απαραίτητο εξοπλισμό, όπως μάσκες προστασίας, κράνη, γάντια, κ.λπ. Επιπρόσθετα, έχουν κατασκευαστεί εφαρμογές για τον έλεγχο της ποιότητας των προϊόντων, μειώνοντας έτσι το πλήθος του ανθρώπινου δυναμικού, καθώς παρόμοιες διεργασίες συχνά απαιτούν μεγάλο όγκο υπαλλήλων [37].[38].

- **Αλληλεπίδραση ανθρώπου-υπολογιστή**

- Σημαντική εφαρμογή της όρασης υπολογιστών αποτελεί η αλληλεπίδραση του ανθρώπου με υπολογιστικά συστήματα μέσω οπτικών μέσων. Οι υπολογιστές, μέσω των κατάλληλων αλγορίθμων, μπορούν πλέον να αντιληφθούν τις φυσικές κινήσεις του ανθρώπου και να εκτελέσουν προκαθορισμένες εντολές βάσει αυτών. Ρομποτικά συστήματα, όπως ρομποτικοί βραχίονες, είναι ικανά να μιμούνται τις ανθρώπινες κινήσεις. Αποτέλεσμα αυτού, είναι ο απομακρυσμένος έλεγχος τους, χωρίς την ανάγκη για επιπλέον εξοπλισμό, χρησιμοποιώντας μόνο μια κάμερα. Υπάρχουν διαθέσιμα πολλά μοντέλα που εστιάζουν στην ανάλυση της κίνησης και αναγνωρίζουν τα ανθρώπινα μέλη του σώματος (πρόσωπο,

χέρια, πόδια, κ.λπ.) με μεγάλη ακρίβεια, όπως και τις κινήσεις που πραγματοποιούν [37].[38].

- **Συστήματα παρακολούθησης**

- Στα συστήματα παρακολούθησης, είναι εφικτή η αυτόματη αναγνώριση συγκεκριμένων συμβάντων μέσω καμερών. Παραδείγματα περιλαμβάνουν την αναγνώριση της ανθρώπινης παρουσίας, την ταυτοποίηση ενός ατόμου βάσει των χαρακτηριστικών του προσώπου του (αναγνώριση προσώπου) και την αναγνώριση κακόβουλων ενεργειών, όπως η προσέγγιση σε απαγορευμένες περιοχές, την ανίχνευση οχημάτων και την αναγνώριση των αριθμών πινακίδων τους (αναγνώριση πινακίδας, αναγνώριση – «διάβασμα» αριθμών) [37].[38].

- **Γεωργία**

- Στη γεωργία, η υπολογιστική όραση χρησιμοποιείται κυρίως για τον υπολογισμό, δηλαδή πρόβλεψη, της αναμενόμενης παραγωγικότητας των φυτών (εκτίμηση αποδόσεων), την αναγνώριση των ασθενειών των φυτών, τη χρήση αυτόνομων οχημάτων για γεωργικές εργασίες, τον έλεγχο της ποιότητας των τροφίμων, και σε άλλες εφαρμογές [37].[38].

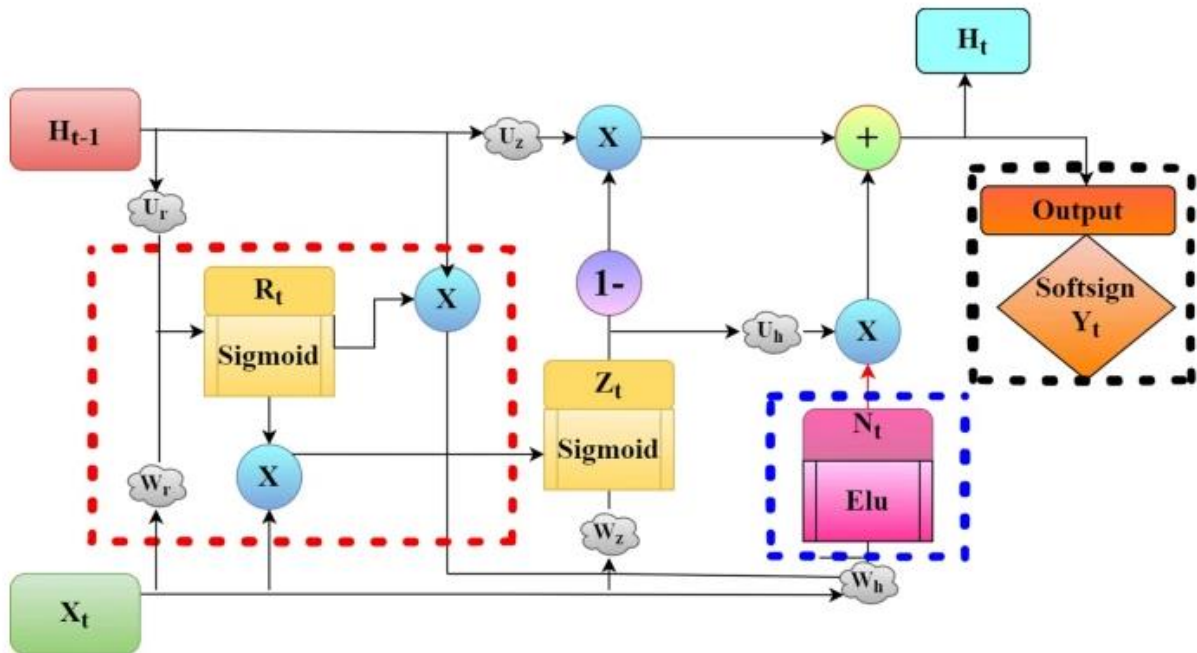
Συνοψίζοντας, αυτές είναι μερικές από τις εφαρμογές της όρασης υπολογιστών, που καλύπτουν πολλούς τομείς, από την ιατρική μέχρι τη γεωργία. Η τεχνολογία αυτή συνεχίζει να αναπτύσσεται και να εξελίσσεται, με δυνατότητες που διεκδικούνται διαρκώς για νέες καινοτόμες εφαρμογές και χρήσεις [5].[19].[37].[38].

## **2.2 Ερευνητικά Άρθρα και Μελέτες**

Το συγκεκριμένο άρθρο [1], ασχολείται με την αναγνώριση της νοηματικής γλώσσας. Αρχικά, οι συγγραφείς αναφέρονται στις δυσκολίες εκμάθησης της Ινδικής νοηματικής γλώσσας, στην πολύπλοκη γλωσσική δομή της, στη σημασία της για τους ανθρώπους με προβλήματα ακοής και την ανάγκη για αποτελεσματικά και εύχρηστα συστήματα αναγνώρισης χειρονομιών, ώστε να εξαλειφθούν σε σημαντικό βαθμό τα εμπόδια στην επικοινωνία μεταξύ των δύο κοινοτήτων, των κωφών και μη. Έπειτα, περιγράφονται οι προκλήσεις που αντιμετωπίζει ο τομέας της πληροφορικής που αναπτύσσεται για να ερμηνεύει νοηματικές γλώσσες, δηλαδή χειρονομίες, με χρήση τεχνικών όρασης υπολογιστών, όπως η ακριβής παρακολούθηση των χειρονομιών, η κάλυψη (occlusion) των χεριών και το υψηλό υπολογιστικό

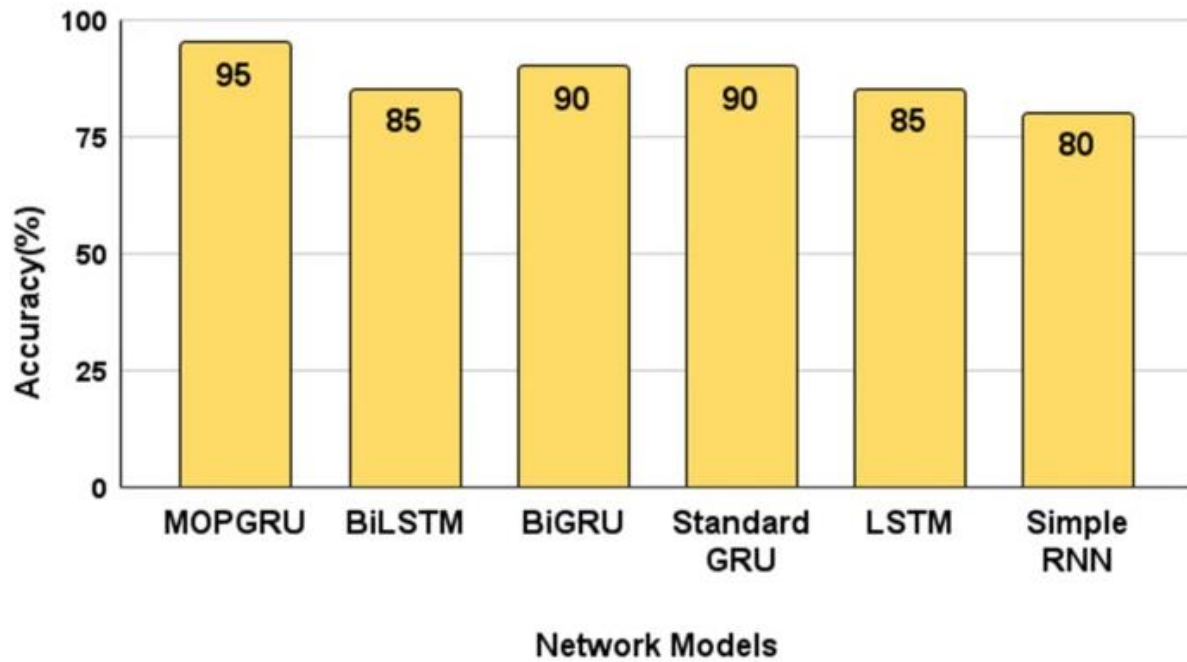
κόστος. Επιπλέον, οι ερευνητές και συγγραφείς του άρθρου προτείνουν ένα μοντέλο MOPGRU (MediaPipe-optimized gated recurrent unit) για την αναγνώριση της μίας διαλέκτου της Ινδικής νοηματικής γλώσσας.

Το μοντέλο αυτό βελτιώνει τη διαχείριση των πληροφοριών με τη χρήση ενισχυμένων πυλών ενημέρωσης και αλλάζοντας τις συναρτήσεις ενεργοποίησης, η αρχική συνάρτηση ενεργοποίησης candidate memory (tahn) μετατρέπεται στην Exponential Linear Units (ELU) συνάρτηση ενεργοποίησης και η αρχική συνάρτηση SoftMax αντικαθίσταται από την συνάρτηση ενεργοποίησης Softsign. Τα αποτελέσματα αποδεικνύουν ότι το ανεπτυγμένο MOPGRU μοντέλο επιτυγχάνει μεγαλύτερα ποσοστά ακρίβειας πρόβλεψης, υψηλή αποτελεσματικότητα μάθησης και γρηγορότερη σύγκλιση σε σύγκριση με άλλα ακολουθιακά-σειριακά μοντέλα μηχανικής μάθησης, που αξιοποιούν τις αρχικές συναρτήσεις, πριν αντικατασταθούν από τους ερευνητές.



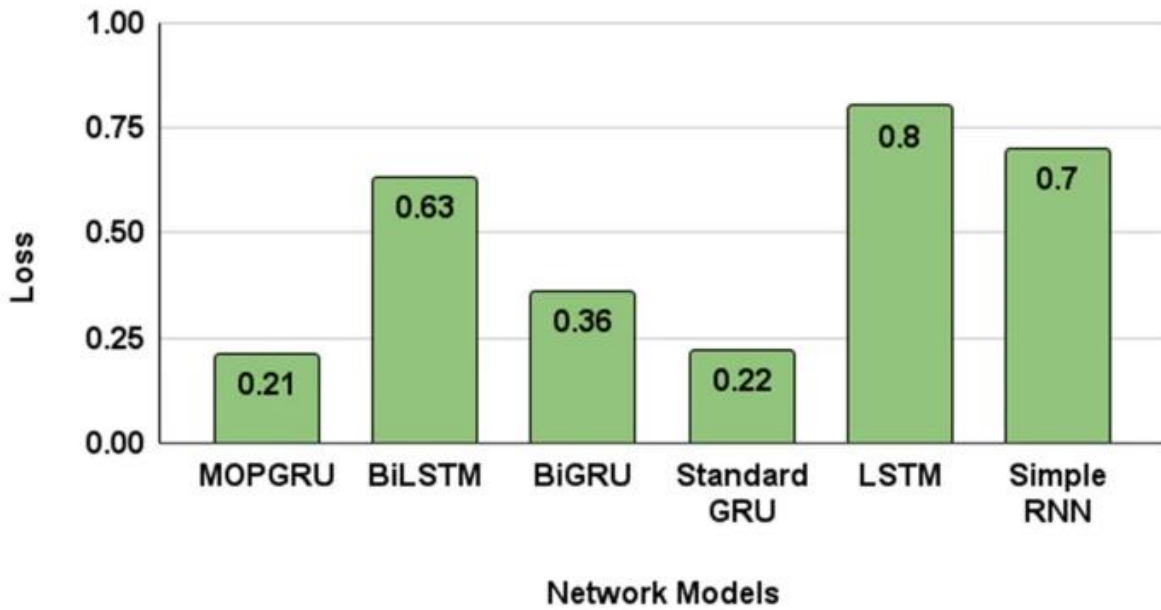
Εικόνα 2-1: Δομή του προτεινόμενου κελιού GRU

Πηγή: <https://www.nature.com/articles/s41598-022-15998-7>



Εικόνα 2-2: Σύγκριση ποσοστών (%) Ακριβείας Μοντέλων Μηχανικής και Βαθιάς Μάθησης

Πηγή: <https://www.nature.com/articles/s41598-022-15998-7>



Εικόνα 2-3: Σύγκριση Απώλειας Μοντέλων Μηχανικής και Βαθιάς Μάθησης

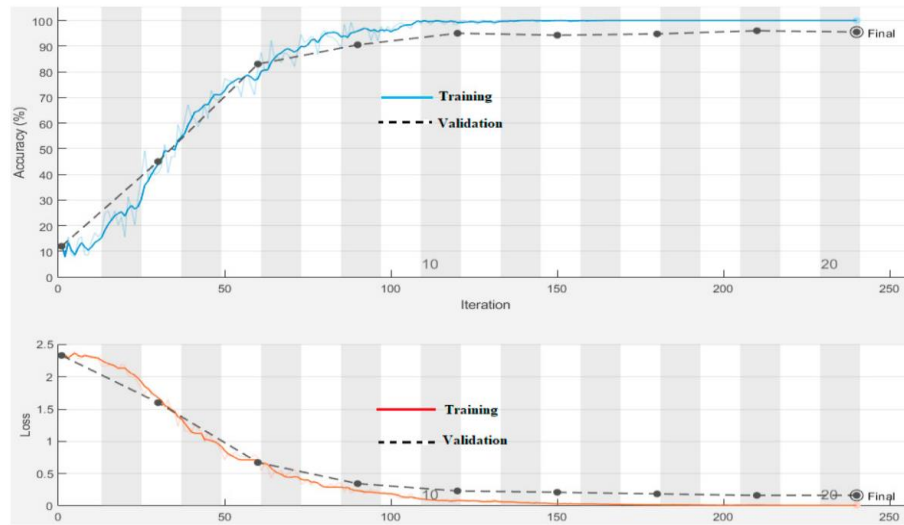
Πηγή: <https://www.nature.com/articles/s41598-022-15998-7>



Συνολικά, το άρθρο εξετάζει την ανάγκη για την ικανότητα αναγνώρισης της νοηματικής γλώσσας, προσφέροντας ένα νέο μοντέλο που βελτιώνει την απόδοση στον τομέα αυτό [1].

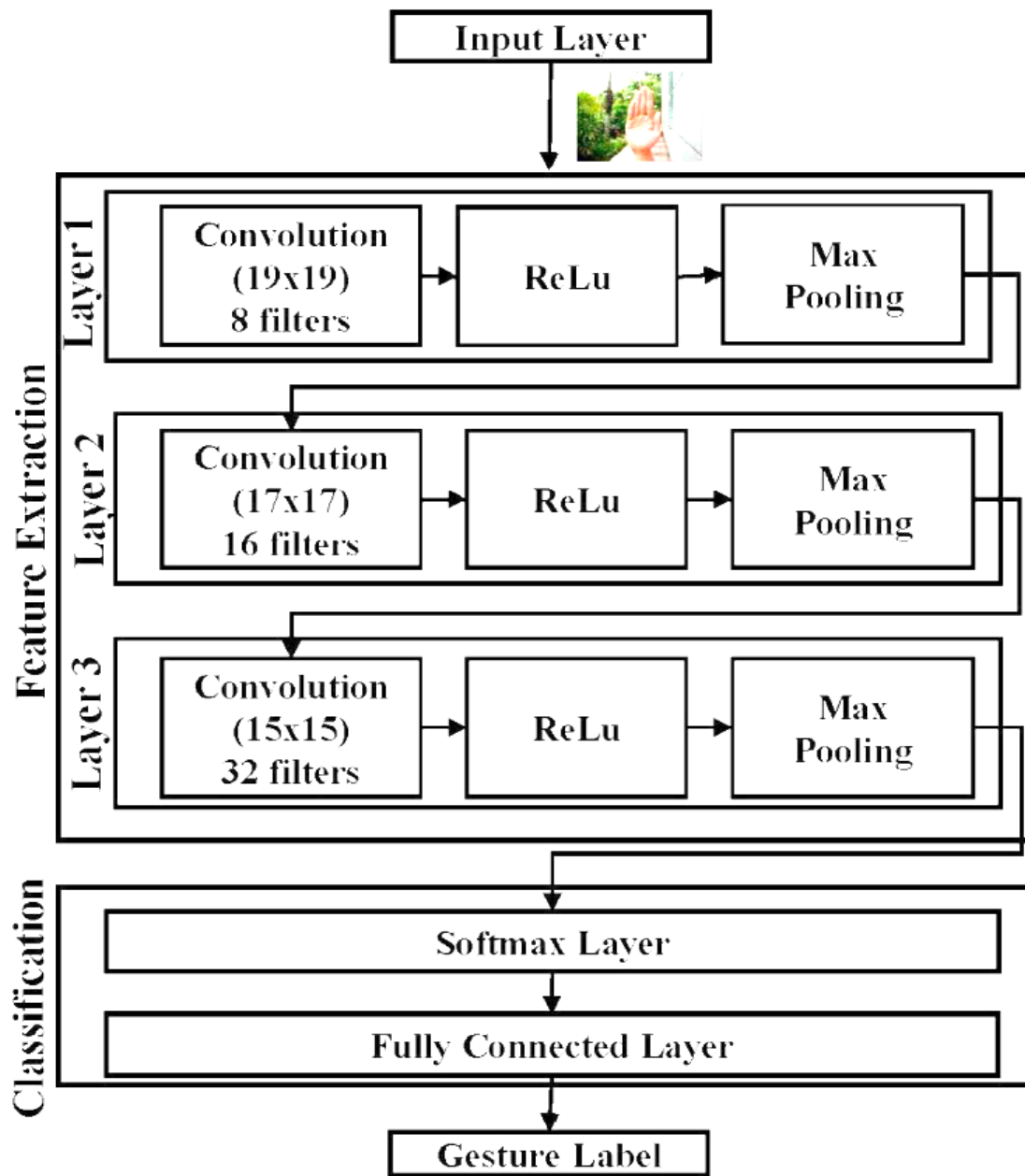
Το θέμα του ερευνητικού άρθρου [2], επικεντρώνεται στην αναγνώριση των νοηματικών χειρονομιών, που αποτελούν τμήμα της όλων των νοηματικών γλωσσών. Η ανάγκη για την αναγνώριση αυτών των χειρονομιών προέρχεται από το γεγονός ότι οι άνθρωποι με δυσκολία στην ομιλία και την ακοή (κωφοί) αντιμετωπίζουν προβλήματα στην καθημερινή επικοινωνία τους με την ακούσια πλειοψηφία. Αν και γίνονται τακτικά προσπάθειες για την ανάπτυξη συστημάτων αναγνώρισης της νοηματικής γλώσσας, αυτά τα συστήματα είναι περιορισμένα από πολλούς παράγοντες που καθιστούν την αναγνώριση σύνθετη. Ακόμα, στην έρευνα τονίζεται η σημασία της αναγνώρισης των χειρονομιών στο πλαίσιο της νοηματικής γλώσσας και η ανάγκη για αποτελεσματικές λύσεις, για έρθουν οι ερευνητές ένα βήμα πιο κοντά στην εξάλειψη του εμποδίου της επικοινωνίας μεταξύ των κωφών ατόμων και των ατόμων με κανονική ακοή.

Η πλειοψηφία των υφιστάμενων ερευνών στον τομέα ακολουθεί παραδοσιακές μεθόδους ανάλυσης μοτίβων για την αναγνώριση νοηματικών χειρονομιών. Ωστόσο, αυτές οι μέθοδοι είναι περιορισμένες στην αντιμετώπιση μεγάλων λεξικών νοηματικών χειρονομιών, σε συνδυασμό με πολύπλοκες συνθήκες φόντου. Στο άρθρο περιγράφεται μια μεθοδολογία για την αναγνώριση των στατικών χειρονομιών, βασιζόμενη σε τεχνικές βαθιάς μάθησης, και πιο συγκεκριμένα σε αρχιτεκτονική συνελκτικού νευρωνικού δικτύου(CNN). Η μέθοδος αυτή δοκιμάστηκε σε δύο δημόσια διαθέσιμα σύνολα δεδομένων και παρατηρήθηκαν καλύτερα αποτελέσματα στην αναγνώριση των χειρονομιών. Τέλος, σύμφωνα με τα αποτελέσματα που παρατηρήθηκαν, επισημαίνονται τα υψηλά ποσοστά επιτυχίας του ανεπτυγμένου μοντέλου βαθιάς μάθησης, και στη συγκεκριμένη περίπτωση ενός μοντέλου συνελκτικών νευρωνικών δικτύων, στην αναγνώριση χειρονομιών, αλλά και η αποφυγή περίπλοκων τεχνικών προ-επεξεργασίας δεδομένων εικόνων και χαρακτηριστικών των χειριών.



Εικόνα 2-4: Διάγραμμα Προσομοίωσης Ακρίβειας και Απώλειας για την εκπαίδευση του μοντέλου Νευρωνικών Δικτύων που προτείνεται

Πηγή: [https://www.sciencedirect.com/science/article/pii/S1877050920312473?ref=pdf\\_download&fr=RR-2&rr=7f37d006fd4e38d4](https://www.sciencedirect.com/science/article/pii/S1877050920312473?ref=pdf_download&fr=RR-2&rr=7f37d006fd4e38d4)



Εικόνα 2-5: Σχηματική Αναπαράσταση του προτεινόμενου μοντέλου Νευρωνικών Δικτύων για την αναγνώριση χειρονομιών

Πηγή: [https://www.sciencedirect.com/science/article/pii/S1877050920312473?ref=pdf\\_download&fr=RR-2&rr=7f37d006fd4e38d4](https://www.sciencedirect.com/science/article/pii/S1877050920312473?ref=pdf_download&fr=RR-2&rr=7f37d006fd4e38d4)

Συνολικά, το άρθρο επικεντρώνεται στην ανάγκη για αποτελεσματική αναγνώριση νοηματικών στατικών χειρονομιών και παρουσιάζει μια προσέγγιση βαθιάς μάθησης για την αντιμετώπιση αυτής της ανάγκης [2].

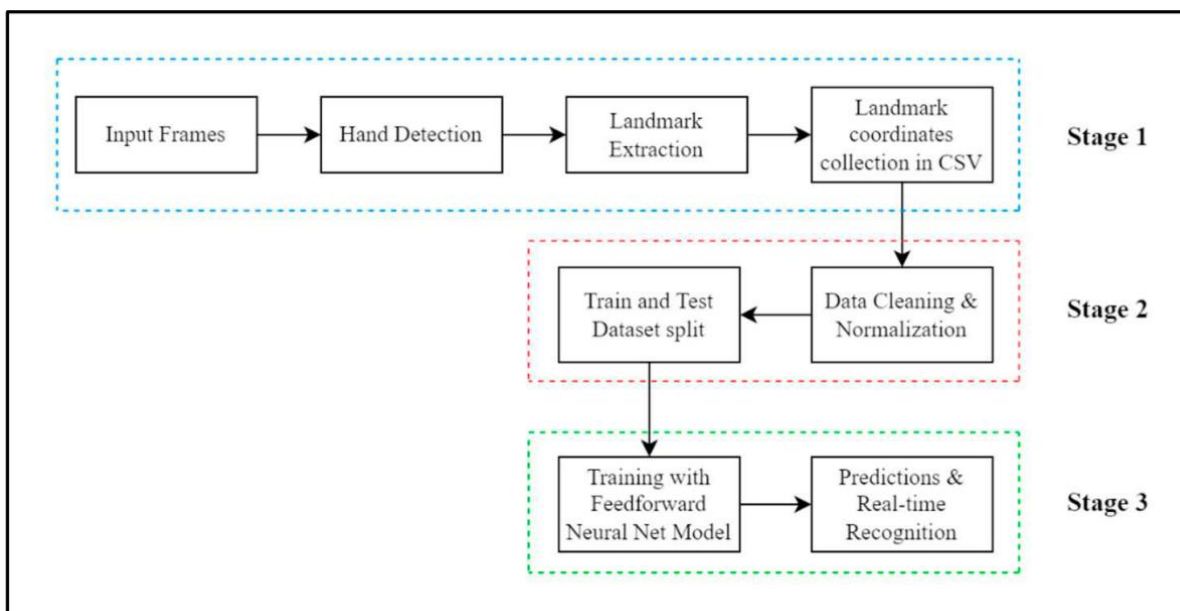
Επίκεντρο της ερευνητικής μελέτης [3] αποτελεί η αναγνώριση της νοηματικής γλώσσας, ιδίως της Ασαμέζικης νοηματικής γλώσσας, που χρησιμοποιείται από τα άτομα με δυσχέρεια ή αναπηρία στην ακοή και την ομιλία. Επιπρόσθετα, αναφέρεται στην αναγκαία ανάπτυξη τεχνολογικών εργαλείων για την ερμηνεία της νοηματικής γλώσσας. Παρά τις προσπάθειες για την αναγνώριση νοηματικών γλωσσών με τη χρήση τεχνολογίας, υπάρχει ακόμη χώρος για την ανάπτυξη εργαλείων για τις νοηματικές γλώσσες που χρησιμοποιούνται σε τοπικές διαλέκτους.

Το άρθρο περιγράφει την τεχνική προσέγγιση και την μεθοδολογία για την αναγνώριση της Ασαμέζικης νοηματικής γλώσσας χρησιμοποιώντας τεχνικές μηχανικής μάθησης. Χρησιμοποιήθηκαν δυσδιάστατες και τρισδιάστατες εικόνες των χειρονομιών της Ασαμέζικης νοηματικής γλώσσας για τη δημιουργία ενός προσαρμοσμένου συνόλου δεδομένων. Το MediaPipe framework χρησιμοποιήθηκε για την ανίχνευση σημείων αναφοράς πάνω στις εικόνες. Στο σύνολο δεδομένων της Ασαμέζικης νοηματικής γλώσσας, που δημιουργήθηκε από τον αρθρογράφο και ερευνητή, περιλαμβάνει εννέα στατικές χειρονομίες με φωνήεντα και σύμφωνα. Επιπλέον, η εφαρμογή του MediaPipe της Google, διασφαλίζει την ακριβή παρακολούθηση της κίνησης των χεριών, περιλαμβάνοντας διαφορετικά σημεία αναφοράς στις

Class label	Precision	Recall	F1 score	Support
1	1.00	1.00	1.00	57
2	1.00	1.00	1.00	55
3	0.96	1.00	0.98	26
4	0.99	1.00	0.99	72
5	1.00	0.99	0.99	70
6	1.00	1.00	1.00	50
7	0.98	0.98	0.98	48
8	1.00	0.99	0.99	68
9	1.00	1.00	1.00	53
accuracy			0.99	499
macro avg	0.99	0.99	0.99	499
weighted avg	0.99	0.99	0.99	499

Εικόνα 2-6: Έλεγχος Κατηγοριοποίησης του προτεινόμενου μοντέλου Νευρωνικών Δικτύων

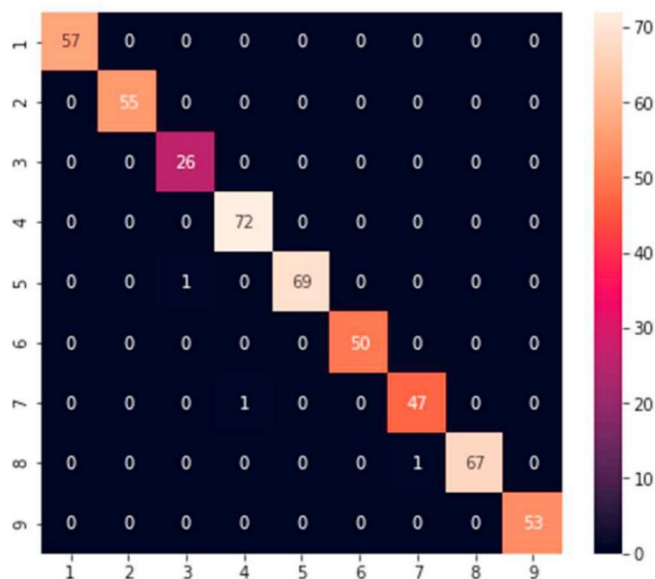
Πηγή: [https://www.sciencedirect.com/science/article/pii/S1877050923001175?ref=pdf\\_download&fr=RR-2&rr=7f5108830e29eeb4](https://www.sciencedirect.com/science/article/pii/S1877050923001175?ref=pdf_download&fr=RR-2&rr=7f5108830e29eeb4)



Εικόνα 2-7: Διαγραμμα Ροής Ολοκληρωμένου Έργου

Πηγή: [https://www.sciencedirect.com/science/article/pii/S1877050923001175?ref=pdf\\_download&fr=RR-2&rr=7f5108830e29eeb4](https://www.sciencedirect.com/science/article/pii/S1877050923001175?ref=pdf_download&fr=RR-2&rr=7f5108830e29eeb4)

φάλαγγες και στις αρθρώσεις των δακτύλων του χεριού του χρήστη. Το σύνολο δεδομένων αυτό αξιοποιήθηκε για την εκπαίδευση και την ανάπτυξη ενός νευρικού δικτύου, το οποίο και επέδειξε ποσοστό ακρίβειας 99%.



Εικόνα 2-8: Πίνακας Σύγκρισης του προτεινόμενου Μοντέλου Νευρωνικού Δικτύου Πρόσθιας Τροφοδότησης

Πηγή: [https://www.sciencedirect.com/science/article/pii/S1877050920312473?ref=pdf\\_download&fr=RR-2&rr=7f37d006fd4e38d4](https://www.sciencedirect.com/science/article/pii/S1877050920312473?ref=pdf_download&fr=RR-2&rr=7f37d006fd4e38d4)

Sign Language	Authors	Approach	Accuracy	Year of Development
American Sign Language	Das et al. [5]	CNN based Inception V3 model	90%	2018
Indian Sign language	Rekha et al. [8]	Skin color segmentation with SVM	86.3%	2011
Indian Sign language	Sahoo [7]	k-NN and Naïve Bayes classifier	98%	2021
American, Indian, Italian and Turkey Sign Language	Halder and Tayade [12]	MediaPipe with SVM	99%	2021
Assamese Sign Language (this work)	This approach	MediaPipe with Feedforward neural network	99%	2022

Εικόνα 2-9: Συγκριτική Ανάλυση του προτεινόμενου Μοντέλου Νευρωνικών Δικτύων, με άλλα Μοντέλα δυσδιάστατης προσέγγισης

Πηγή: [https://www.sciencedirect.com/science/article/pii/S1877050920312473?ref=pdf\\_download&fr=RR-2&rr=7f37d006fd4e38d4](https://www.sciencedirect.com/science/article/pii/S1877050920312473?ref=pdf_download&fr=RR-2&rr=7f37d006fd4e38d4)

Sign Language	Authors	Approach	Accuracy	Year of Development
Indian Sign language	Ansari and Harit [4]	k-means clustering	90.68%	2016
American Sign language	Keskin et al. [11]	Object recognition by parts	98%	2013
American Sign Language	Pugeault and Bowden [10]	MediaPipe with SVM	75%	2011
Indian Sign Language	Kumar et al. [13]	Spatio-temporal graph kernels	98.75%	2022
Assamese Sign Language (this work)	This approach	MediaPipe with Feedforward neural network	99%	2022

Εικόνα 2-10: : Συγκριτική Ανάλυση του προτεινόμενου Μοντέλου Νευρωνικών Δικτύων, με άλλα Μοντέλα τρισδιάστατης προσέγγισης

Πηγή: [https://www.sciencedirect.com/science/article/pii/S1877050920312473?ref=pdf\\_download&fr=RR-2&rr=7f37d006fd4e38d4](https://www.sciencedirect.com/science/article/pii/S1877050920312473?ref=pdf_download&fr=RR-2&rr=7f37d006fd4e38d4)

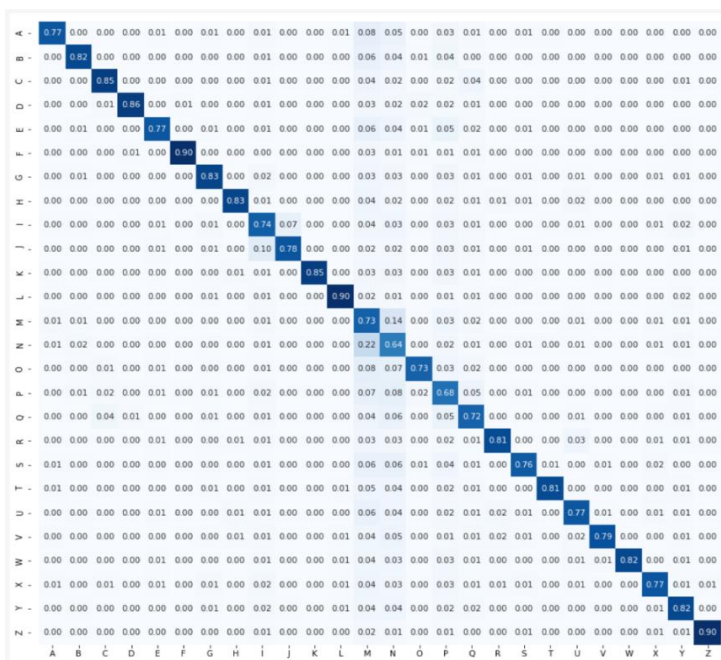
Συνοψίζοντας, τα αποτελέσματα υποδεικνύουν ότι η μέθοδος υλοποιημένη σε αυτό το έργο, είναι αποτελεσματικότερη, για την αναγνώριση στατικών χειρονομιών νοηματικής γλώσσας, που αντιστοιχούν σε γράμματα του αλφαβήτα της Ασαμέζικης γλώσσας. Τέλος, το άρθρο αναφέρει ότι η μέθοδος μπορεί να δοκιμαστεί και να εφαρμοστεί και σε άλλες τοπικές γλώσσες της Ινδίας για την αναγνώριση των τοπικών νοηματικών γλωσσών [3].

Το επόμενο άρθρο [4] που θα αναλυθεί, επικεντρώνεται στην αναγνώριση της αμερικανικής νοηματικής γλώσσας (ASL) χρησιμοποιώντας εικόνες χεριών που λαμβάνονται από την διαδικτυακή κάμερα (web camera) του υπολογιστή. Αρχικά, στο άρθρο αναφέρεται η σημασία της αναγνώρισης της νοηματικής γλώσσας για την ανεμπόδιστη επικοινωνία της κοινότητας των ατόμων που είναι κωφά ή πάσχουν από κάποια ακουστική αναπηρία, με τους υπόλοιπους συνανθρώπους τους. Το άρθρο υπογραμμίζει ότι παρά τις παλαιότερες προσπάθειες αναγνώρισης της νοηματικής γλώσσας, οι προηγούμενες μέθοδοι χρησιμοποιούσαν δαπανηρές και εξειδικευμένες συσκευές παρακολούθησης και

εξαγωγής των χαρακτηριστικών των χεριών, που δυστυχώς δεν έχουν πολλοί άνθρωποι πρόσβαση σε αυτές, αλλά και εμποδίζουν τις κινήσεις των χρηστών.

Η προσέγγιση αυτή χρησιμοποιεί τον αλγόριθμο MediaPipe Hands για τον εντοπισμό των συντεταγμένων των αρθρώσεων των χεριών, δηλαδή των σημείων αναφοράς των χεριών, από δυσδιάστατες εικόνες, οι οποίες λαμβάνονται από μία απλή κάμερα ενός υπολογιστή. Από τις συντεταγμένες των σημείων αναφοράς, παράγονται δύο τύποι χαρακτηριστικών: ο ένας βασίζεται στις αποστάσεις μεταξύ των σημείων των αρθρώσεων και ο άλλος στις γωνίες μεταξύ διανυσμάτων και των τριών διαστάσεων (X, Y, Z). Οι ερευνητές, χρησιμοποίησαν τους ταξινομητές Support Vector Machine (SVM) και Light Gradient Boosting Machine (GBM), για την κατηγοριοποίηση των χαρακτήρων.

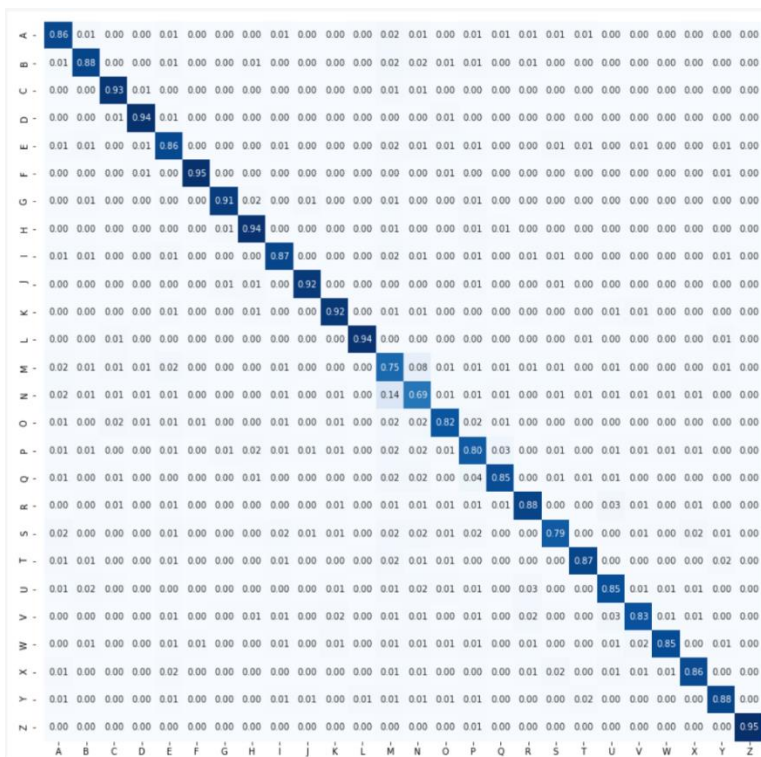
Χρησιμοποιήθηκαν τρία δημόσια σύνολα δεδομένων για την αναγνώριση και την εκμαίευση των αποτελεσμάτων : το ASL Alphabet dataset, το Massey dataset και το Finger Spelling A dataset. Με το πέρας της εκπαίδευσης, παρατηρήθηκε, πως αποτελέσματα κυμαίνονταν από 87,60% στο ASL Alphabet dataset έως 99,39% στο Massey dataset και 98,45% στο Finger Spelling A dataset. Η μέθοδος που προτείνεται για την ανάπτυξη ενός αυτοματοποιημένου συστήματος αναγνώρισης της αμερικανικής νοηματικής γλώσσας είναι οικονομικά αποδοτική, με μικρό υλικό υπολογιστικό φόρτο, δηλαδή δεν απαιτεί



Εικόνα 2-11: Πίνακας Σύγχυσης για τα χαρακτηριστικά βασισμένα στην Απόσταση

Πηγή: <https://www.mdpi.com/1424-8220/21/17/5856>

ειδικούς αισθητήρες ή συσκευές, και έχει ξεπεράσει τις προηγούμενες μελέτες, συγκριτικά με τα αποτελέσματά τους.



Εικόνα 2-12: Πίνακας Σύγχυσης για τα χαρακτηριστικά βασισμένα στις γωνίες

Πηγή: <https://www.mdpi.com/1424-8220/21/17/5856>

Ανακεφαλαιώνοντας, στο πλαίσιο της εργασίας αυτής, η αναγνώριση της Αμερικανικής νοηματικής γλώσσας βασίζεται στις εικόνες των χεριών που λαμβάνονται από μια κάμερα υπολογιστή. Με αυτή τη μέθοδο παρακάμπτεται η ανάγκη για ακριβό και εξειδικευμένο αισθητήριο εξοπλισμό, δίχως να επηρεάζονται τα υψηλά ποσοστά αναγνώρισης και πρόβλεψης του ανεπτυγμένου συστήματος. Τέλος, αποτελέσματα της έρευνας δείχνουν ότι η προτεινόμενη μέθοδος είναι αποδοτική και έχει υπερνικήσει προηγούμενες μελέτες, παρέχοντας μια προσιτή και αποδοτική λύση για την αναγνώριση της νοηματικής γλώσσας [4].

Οι ερευνητές του πέμπτου κατά σειρά άρθρου [13], ασχολούνται με τη συνεχή αναγνώριση της νοηματικής γλώσσας μέσω βίντεο, όπου διατίθενται ετικέτες για τις χειρονομίες της νοηματικής γλώσσας, αλλά όχι ακριβής χρονικούς περιορισμούς για την κάθε χειρονομία. Κατά αρχάς, οι συγγραφείς – ερευνητές του άρθρου αναφέρουν, ότι η συνεχής αναγνώριση της νοηματικής γλώσσας αποτελεί πρόκληση, για τους ίδιους και τους συναδέλφους τους, λόγω της έλλειψης ακριβούς επίβλεψης χρονικών ορίων στα δεδομένα βίντεο και της περιορισμένης ποσότητας δεδομένων εκπαίδευσης. Στην ερευνητική



εργασία, υπερνικούνται αυτές τις προκλήσεις και επιδιώκεται η αναγνώριση των αντίστοιχων σχέσεων μεταξύ των χρονικών ακολουθιών εικόνων και των ακολουθιών γλωσσών.

Η προσέγγιση αυτή αντιμετωπίζει το πρόβλημα της αναγνώρισης της νοηματικής γλώσσας, αντιστοιχίζοντας τα τμήματα του βίντεο με τις χειρονομίες της νοηματικής γλώσσας, χρησιμοποιώντας αναδρομικά συνελκτικά νευρωνικά δίκτυα για την εξαγωγή χωροχρονικών χαρακτηριστικών για τη δημιουργία του συνόλου δεδομένων και την εκμάθηση των ακολουθιών από το επιλεγμένο μοντέλο βαθιάς μάθησης νευρωνικών δικτύων. Επιπλέον, στο άρθρο παρουσιάζεται τρία επίπεδα βελτιστοποίησης για την αρχιτεκτονική του συνελκτικού νευρωνικού δικτύου, χρησιμοποιώντας συνάρτηση αντικειμενικού σκοπού της εκπαίδευσης της ακολουθίας, και υπογραμμίζεται η επίτευξη αποτελεσμάτων που είναι συγκρίσιμα με τα καλύτερα της τάξης στην κατηγορία αυτή.

Συνολικά, η προτεινόμενη μέθοδος βασίζεται σε αναδρομικά συνελκτικά νευρωνικά δίκτυα και εκμεταλλεύεται στο έπακρον τις ικανότητες εκπαίδευσης, αναγνώρισης και πρόβλεψης των βαθιών συνελκτικών νευρωνικών δικτύων. Επιπλέον, πραγματοποιεί μία τριαδική διαδικασία βελτιστοποίησης, χρησιμοποιώντας τις ακολουθίες γλωσσών για την εκπαίδευση του συστήματος και αποφεύγοντας την υπερπροσαρμογή (overfitting). Η αποτελεσματικότητα της προτεινόμενης μεθόδου επιδεικνύεται σε ένα δύσκολο πρόβλημα συνεχούς αναγνώρισης της νοηματικής γλώσσας, επιτυγχάνοντας αποτελέσματα που συγκρίνονται με τα καλύτερα της τάξης, όπως είναι τα παρακάτω μοντέλα μηχανικής μάθησης, ανεπτυγμένα για παρόμοιους σκοπούς: HOG-3D, CMLLR, 1-Mio-Hands[18], 1-Mio-Hands[18]+[16] και CNN-Hybrid [13].

Model setup	Extra supervision	Modality			Validation		Test	
		r-hand	traj	face	del / ins	WER	del / ins	WER
HOG-3D [16]		✓			25.8 / 4.2	60.9	23.2 / 4.1	58.1
[16] CMLLR		✓	✓	✓	21.8 / 3.9	55.0	20.3 / 4.5	53.0
1-Mio-Hands [18]	✓	✓			19.1 / 4.1	51.6	17.5 / 4.5	50.2
1-Mio-Hands [18]+[16]	✓	✓	✓	✓	16.3 / 4.6	47.1	15.2 / 4.6	45.1
CNN-Hybrid [19]	✓	✓			12.6 / 5.1	<b>38.3</b>	11.1 / 5.7	38.8
Our-end2end		✓			16.3 / 6.7	46.2	15.1 / 7.4	46.9
Ours		✓			13.7 / 7.3	39.4	12.2 / 7.5	<b>38.7</b>

Εικόνα 2-13: Σύγκριση Ποσοστών (%) Απόδοσης διαφορετικών προσεγγίσεων Συνεχούς Αναγνώρισης της Νοηματικής Γλώσσας, Όπου «r-hand» σημαίνει δεξί χέρι και όπου «traj» σημαίνει τροχιά.

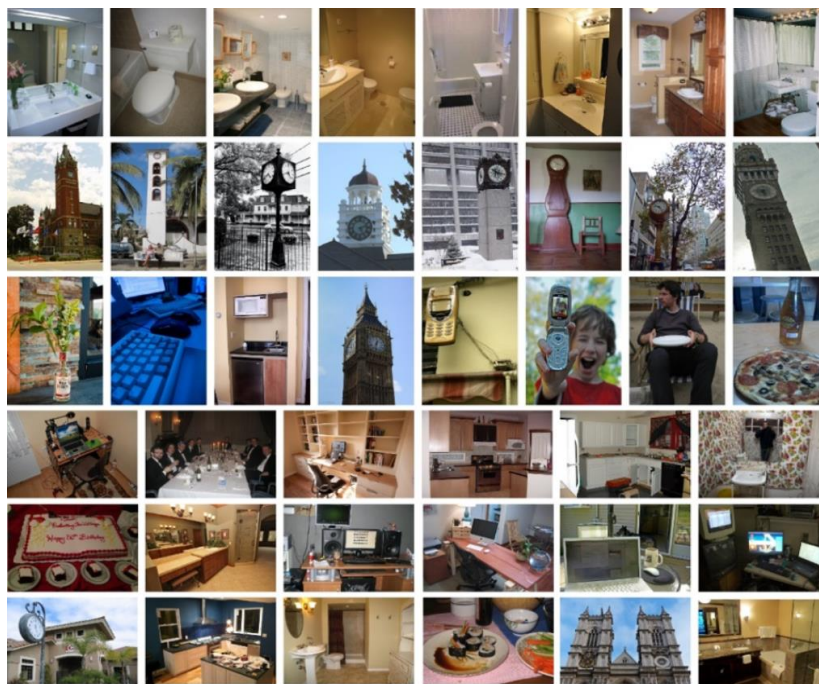
Πηγή: CVF, 'Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition by Staged Optimization', by Runpeng Cui, Hu Li, Changshui Zhang, Department of Automation, Tsinghua University, pp 7361-7368.

Στο επόμενο άρθρο [5] οι ερευνητές, αρχικά επισημαίνουν ότι η ανίχνευση αντικειμένων είναι άκρως απαραίτητη στην αυτόματη ανίχνευση εικόνων, με απαραίτητη προϋπόθεση την κατηγοριοποίηση

αντικειμένων σε εικόνες σε συνδυασμό με την πιο πιθανή τοποθεσία του αντικειμένου, με βάση στατιστικά, γεωγραφικά και γεωλογικά δεδομένα. Επιπλέον, αναφέρουν πως παρόμοια συστήματα εφαρμόζονται σε πολλούς τομείς, όπως η έξυπνη (αυτόματη) παρακολούθηση, η ανίχνευση αντικειμένων για στρατιωτικούς σκοπούς, η πλοήγηση μη-επανδρωμένων αεροσκαφών και η έξυπνη (αυτόματη) μεταφορά.

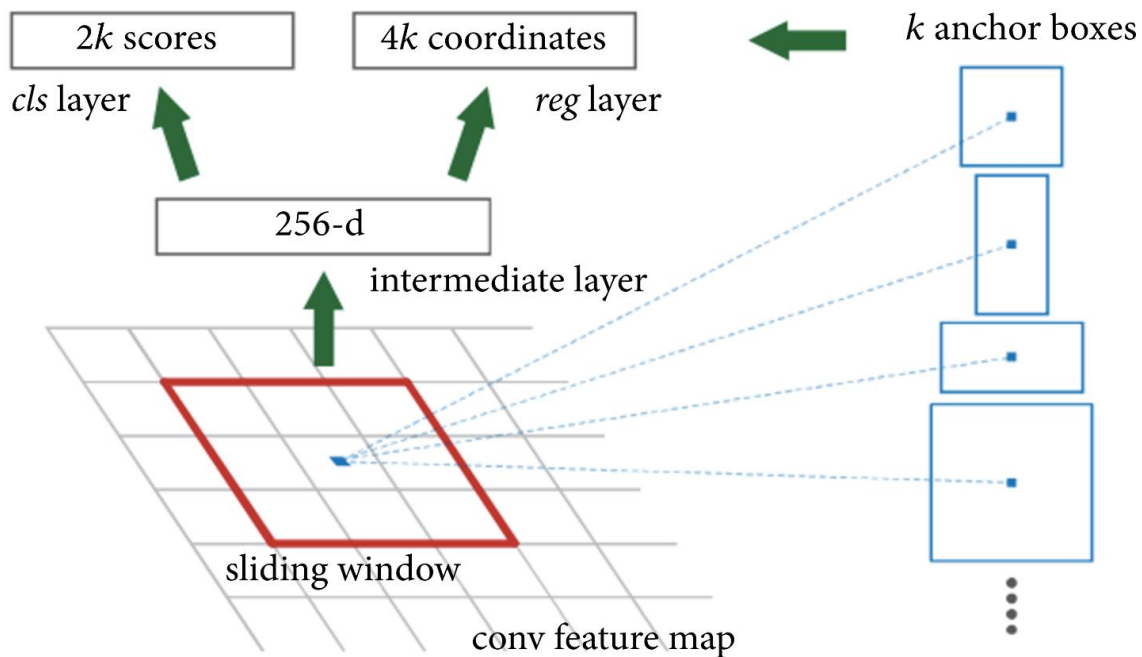
Συνεχίζοντας, κάνουν λόγο για τα εμπόδια και τις ανησυχίες τους, με τα οποία έρχονται αντιμέτωποι κατά τη διαδικασία της ανίχνευσης μικρών αντικειμένων σε εικόνες, αντιμετωπίζοντας ιδιαίτερη δυσκολία στα μικρά αντικείμενα των εικόνων που εμφανίζουν χαμηλή ανάλυση και επηρεάζονται από θόρυβο, αντικείμενα δηλαδή που είναι δυσδιάκριτα. Ακόμα, παραθέτουν, πως τα υπάρχοντα μοντέλα ανίχνευσης αντικειμένων, βασισμένα σε βαθιά συνελκτικά νευρωνικά δίκτυα, χρησιμοποιούν πολλαπλά στάδια συνέλιξης και συσσώρευσης για να εξάγουν τα απαραίτητα χαρακτηριστικά που φέρουν ενδιαφέρον, από ολόκληρη την εικόνα. Επιπλέον, σχολιάζουν πως τα συγκεκριμένα μοντέλα αποδίδουν καλά μόνο για μεγάλα αντικείμενα, και αποτυγχάνουν στην ανίχνευση μικρών αντικειμένων με χαμηλή ανάλυση. Αυτό συμβαίνει λόγω του ότι έπειτα από τα πολλαπλά στάδια συνέλιξης και συσσώρευσης, τα δυσδιάκριτα μικρά αντικείμενα αλλοιώνονται σε τέτοιο βαθμό, που τα βασικά τους χαρακτηριστικά ως αντικείμενα δεν τα αντιπροσωπεύουν πλέον.

Οι αρθρογράφοι και συνάμα ερευνητές της μελέτης, παρουσιάζουν ένα μοντέλο ανίχνευσης που επιτυγχάνει υψηλή ακρίβεια ανίχνευσης για τα μικρά αντικείμενα, εξάγοντας τα χαρακτηριστικά ενδιαφέροντος τους από διάφορα επίπεδα συνέλιξεων από τα οποία και φιλτράρεται η εικόνα και χρησιμοποιώντας τα για την ανίχνευση τους. Το σύστημα ανίχνευσης το οποίο περιγράφουν βασίζεται στο μοντέλο Faster-RCNN, ανεπτυγμένο συγκεκριμένα για την ανίχνευση μικρών και δυσδιάκριτων αντικειμένων, και την τοποθέτηση ειδικών πλαισίων, περιφερειακά των αντικειμένων. Το μοντέλο εκπαιδεύεται χρησιμοποιώντας ένα προσεγμένο και ειδικά δημιουργημένο για το σκοπό αυτό σύνολο δεδομένων για μικρά αντικείμενα. Τα αποτελέσματα δείχνουν ότι το προτεινόμενο μοντέλο επιτυγχάνει ακρίβεια ανίχνευσης 11% υψηλότερη από τα καλύτερα μοντέλα ανίχνευσης της αγοράς.



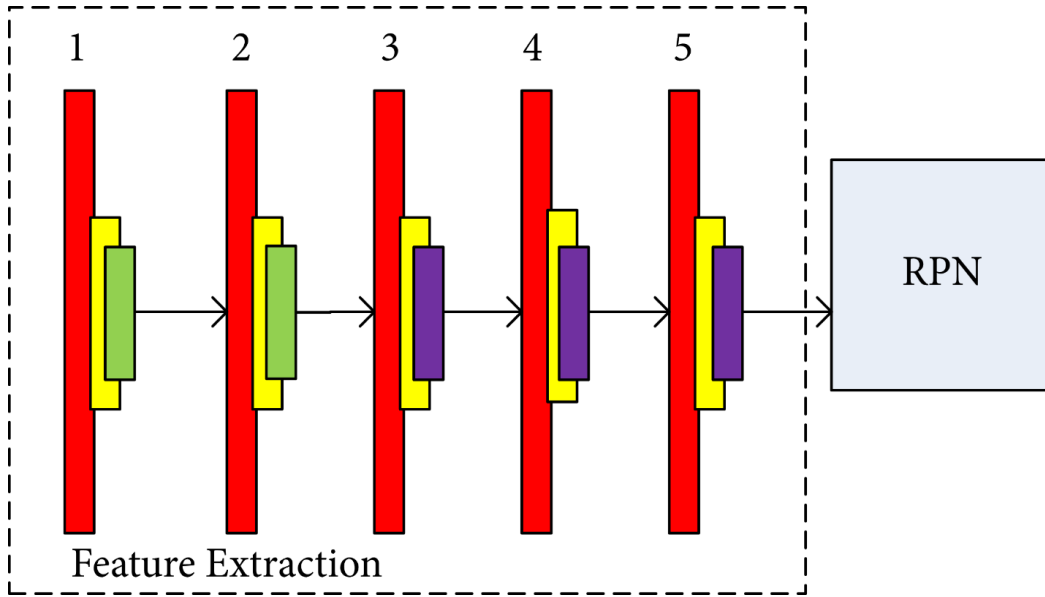
Εικόνα 2-14: Ένα μέρος του Συνόλου Δεδομένων, που χρησιμοποιήθηκε για την έρευνα

Πηγή: <https://www.hindawi.com/journals/ijdm/2018/4546896/>



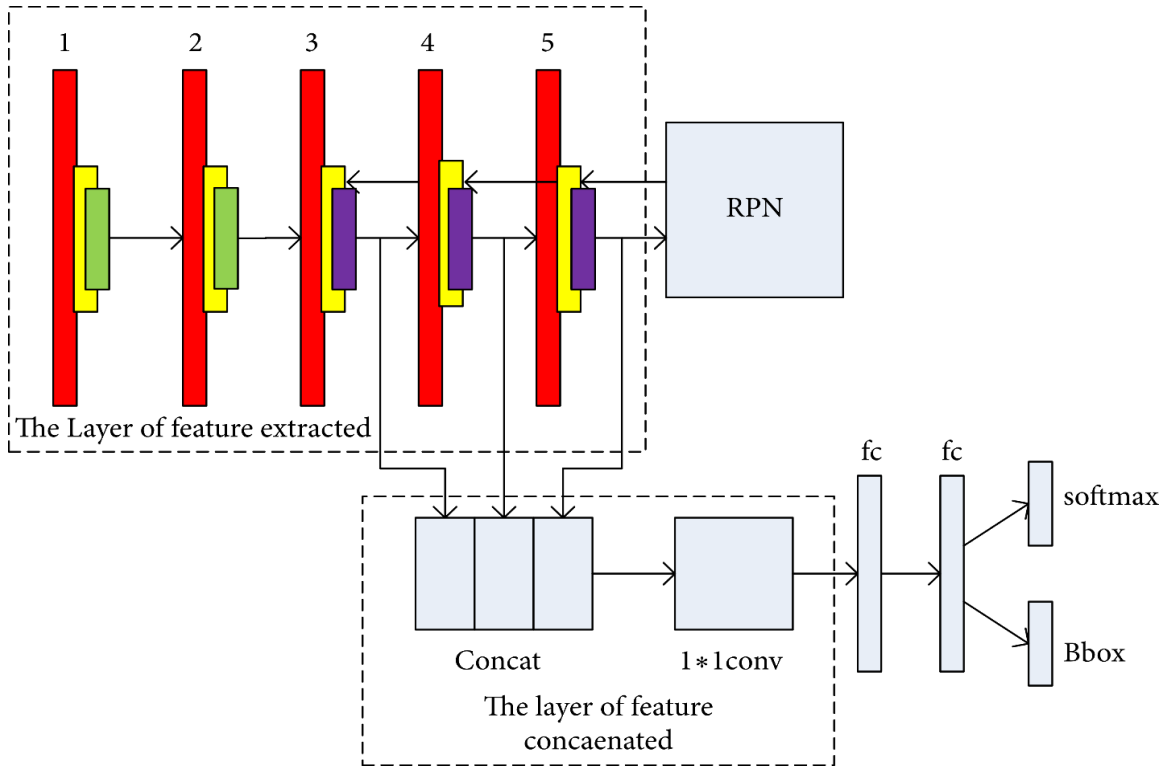
Εικόνα 2-15: Δομή Αρχιτεκτονικής Δικτύου RPN

Πηγή: <https://www.hindawi.com/journals/ijdm/2018/4546896/>



Εικόνα 2-16: Δομή Αρχιτεκτονικής Δικτύου Faster-RCNN

Πηγή: <https://www.hindawi.com/journals/ijdm/2018/4546896/>



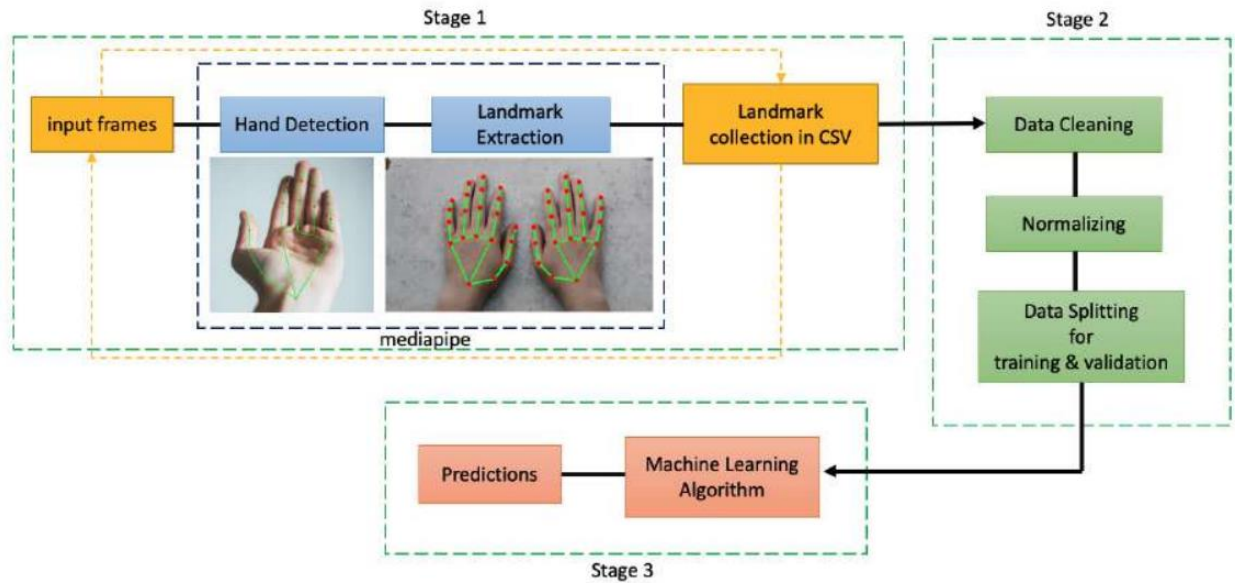
Εικόνα 2-17: Δομή Αρχιτεκτονικής Δικτύου Multiscale Faster-RCNN

Πηγή: <https://www.hindawi.com/journals/ijdm/2018/4546896/>

Συνοψίζοντας, το θέμα του παραπάνω άρθρου εστιάζει στην ανίχνευση μικρών αντικειμένων, που πολλές φορές είναι δυσδιάκριτα λόγω θορύβων και άλλων περιβαλλοντολογικών συνθηκών σε εικόνες και παρουσιάζει ένα νέο μοντέλο ανίχνευσης βασισμένο στην αρχιτεκτονική βαθιών συνελκτικών νευρωνικών δικτύων για την αντιμετώπιση αυτού του προβλήματος [5].

Στο έβδομο κατά σειρά άρθρο [14] που θα αναλυθεί, περιγράφεται μια μεθοδολογία για την αναγνώριση της νοηματικής γλώσσας μέσω χρήσης τεχνικών της τεχνητής νοημοσύνης και της όρασης υπολογιστών, σε συνδυασμό με το ανοικτό πλαίσιο του MediaPipe. Καταρχάς, επισημαίνεται η σημασία της αναγνώρισης της νοηματικής γλώσσας, όπως και η ανάγκη για την αυτοματοποίηση της διαδικασίας αυτής, γεγονός που θα βοηθούσε στη γεφύρωση του χάσματος της επικοινωνίας μεταξύ ατόμων με κανονική ακοή και ατόμων που ανήκουν στην κοινότητα των κωφών, η οποία αποτελεί και το 5% του παγκόσμιου πληθυσμού σύμφωνα με την παγκόσμια ομοσπονδία κωφών (World Federation of the Deaf (WFD)). Ο κύριος σκοπός του άρθρου είναι να απλοποιήσει την αναγνώριση της νοηματικής γλώσσας αξιοποιώντας αλγορίθμους μηχανικής μάθησης. Το προτεινόμενο μοντέλο είναι ελαφρύ και προσαρμόζεται εύκολα σε φορητές συσκευές, επιτρέποντας την αναγνώριση των χειρονομιών της νοηματικής γλώσσας σε πραγματικό χρόνο χωρίς τη χρήση αισθητήρων. Επιπλέον, υπογραμμίζεται ότι υπάρχει μεγάλη ποικιλομορφία στις νοηματικές γλώσσες ανά τις περιοχές, με διαφορετικά αλφάβητα, κινήσεις χεριών, δακτυλογραφίες και κινήσεις του σώματος. Επιπρόσθετα, γίνεται αναφορά στο γεγονός ότι υπάρχουν πολλές λέξεις που δεν έχουν ακόμη κανονικές αντιστοιχίες στα λεξικά της νοηματικής γλώσσας, και για αυτές χρησιμοποιείται η δακτυλογραφία. Στην συγκεκριμένη έρευνα χρησιμοποιούνται σύνολα δεδομένων από διάφορες νοηματικές γλώσσες, όπως η αμερικανική, η ινδική, η ιταλική και η τουρκική, για εκπαίδευση του μοντέλου που περιγράφεται.

Στο άρθρο εξηγείται ότι η χρήση αισθητήρων στην αναγνώριση νοηματικής γλώσσας ήταν περιοριστική και ακριβή, ενώ αντίθετα συστήματα στα οποία έχουν ενσωματωθεί τεχνικές βασισμένες στην όραση του υπολογιστή, είναι πιο οικονομικά και πολύ πιο εύκολα στη μετακίνηση. Οι πιο σύνηθες μέθοδοι προς ανίχνευση των χεριών περιλαμβάνουν την ανίχνευση χρώματος του δέρματος και την αφαίρεση του φόντου, ώστε να εξαχθεί το περίγραμμα του χεριού, και συνεπώς η χειρονομία που συμβολίζει. Το προτεινόμενο μοντέλο βασίζεται στο πλαίσιο MediaPipe και χρησιμοποιεί διάφορους αλγορίθμους μηχανικής μάθησης, όπως ο Support Vector Machine (SVM), ο k-Nearest-Neighbors (KNN), ο Random Forest, ο Decision Tree, ο Naive Bayes, το νευρωνικό δίκτυο Artificial Neural Network (ANN) και το νευρωνικό δίκτυο Multi-Layer Perception (MLP), για την αναγνώριση της νοηματικής γλώσσας.



Εικόνα 2-18:Προτεινόμενη Αρχιτεκτονική για τον Εντοπισμό της Παλάμης και την Αναγνώριση χειρονομιών

Πηγή: International Journal of Research Publication and Reviews, 'Real-time Vernacular Sign Language Recognition using

Στη συνέχεια, συγκρίνονται τα αποτελέσματα των μέσων όρων ποσοστών ακριβείας (Average Accuracy) όλων των αλγορίθμων, που εκπαιδεύτηκαν στα προαναφερόμενα σύνολα δεδομένων. Μεγαλύτερο ποσοστό ακριβείας φέρει ο αλγόριθμος Support Vector Machine (SVM) σε όλες τις νοηματικές γλώσσες, και παράλληλα λιγότερο αποτελεσματικός στη συγκεκριμένη έρευνα αποδεικνύεται ο Naive Bayes, ο οποίος φέρει και το μικρότερο μέσο όρο ποσοστού ακριβείας. Ακόμα, εμφανίζονται και τα ποσοστά επιτυχίας Ακρίβειας Εκπαίδευσης (Training Accuracy), Ακρίβεια Δοκιμής (Testing Accuracy), Ακρίβεια (Precision), Ανάκληση (Recall) και Βαθμολογία F1 (F1-Score), για τον αλγόριθμο Support Vector Machine (SVM).

Dataset	SVM	KNN	Random Forest	Decision Tree	Naive Bayes	ANN	MLP
ASL(alphabet)	99.15%	98.21%	98.57%	98.57%	53.74%	97.12%	94.69%
Indian(alphabet)	99.29%	98.87%	98.59%	98.59%	86.77%	94.79%	96.48%
Italian(alphabet)	98.19%	96.75%	97.83%	97.83%	77.19%	78.63%	72.14%
ASL(numbers)	99.18%	99.18%	97.56%	97.56%	96.74%	95.12%	97.56%
Turkey (numbers)	96.22%	93.08%	94.33%	94.33%	83.64%	93.71%	83.64%

Εικόνα 2-19: Μέση Ακρίβεια που επιτυγχάνεται χρησιμοποιώντας αλγορίθμους Μηχανικής και Βαθιάς Μάθησης

Πηγή: International Journal of Research Publication and Reviews, 'Real-time Vernacular Sign Language Recognition using

Sign Language	Reference	Type	Number of classes	Method	Accuracy
American	P.Das et al.,[14]	Alphabets	26	Deep CNN	94.3%
	M.Taskiran et al.,[15]	Alphabets and Numbers	36	CNN	98.05%
	N.Saquib and A.Rahman[16]	Alphabets	24	KNN	96.14%
				Random Forest	96.13%
				ANN	95.87%
				SVM	94.91%
<b>Ours</b>	Alphabets	26	SVM	<b>99.15%</b>	
	Numbers	10	SVM	<b>99.18%</b>	
Indian	K.K.Dutta et al.,[17]	Alphabets	24	KNN	94%-96%
	M.Sharma et al.,[18]	Numbers	10	KNN and Neural Network	97.10%
	J.L.Raheja et al.,[19]	Alphabets	24	SVM	97.5%
	<b>Ours</b>	Alphabets	26	SVM	<b>99.29%</b>
Italian	L.Pigou et al.,[20]	Alphabets	20	CNN	91.7%
	<b>Ours</b>	Alphabets	22	SVM	<b>98.19%</b>

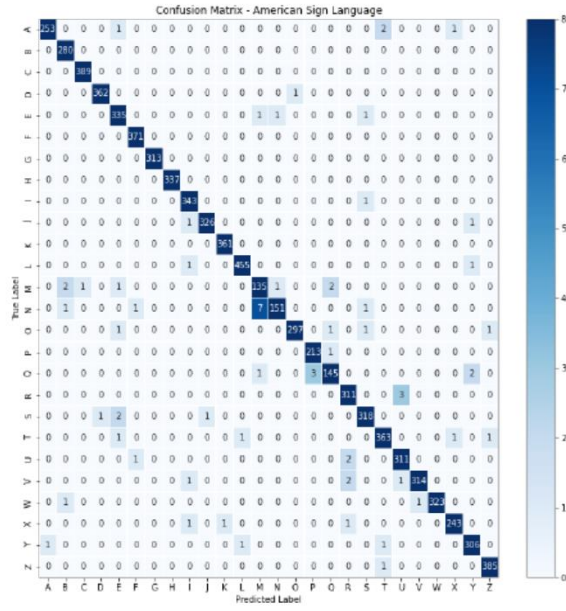
Εικόνα 2-20: Σύγκριση αποτελεσμάτων με παρόμοιες καινοτόμες μεθοδολογίες

Πηγή: International Journal of Research Publication and Reviews, 'Real-time Vernacular Sign Language Recognition using

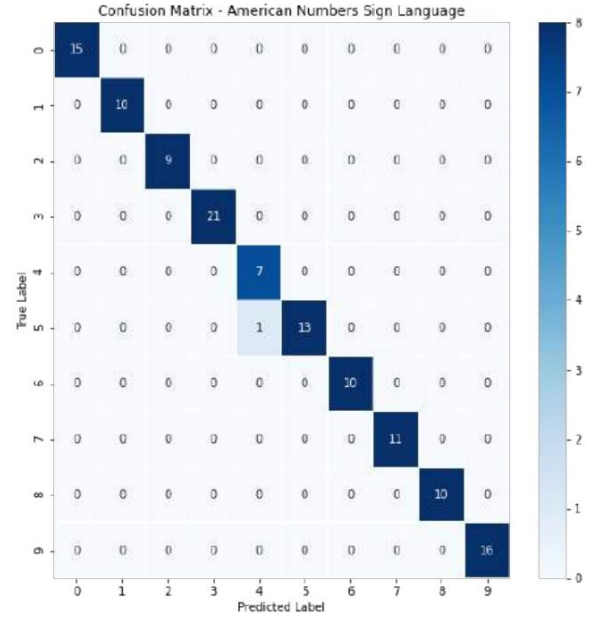
Dataset name	Training Accuracy	Testing Accuracy	Precision	Recall	F1-Score
ASL(alphabet)	99.50%	99.15%	99.15%	99.15%	99.15%
Indian(alphabet)	99.92%	99.29%	99.29%	99.29%	99.29%
Italian(alphabet)	99.72%	98.19%	98.19%	98.19%	98.19%
Turkey (numbers)	99.37%	96.22%	96.22%	96.22%	96.22%
American (numbers)	98.77%	99.18%	99.18%	99.18%	99.18%

Εικόνα 2-21: Ανάλυση Απόδοσης του αλγορίθμου SVM σε διάφορα Σύνολα Δεδομένων

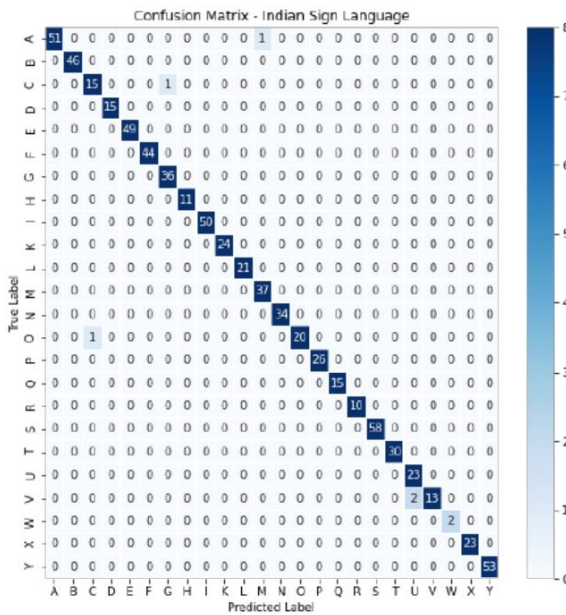
Πηγή: International Journal of Research Publication and Reviews, 'Real-time Vernacular Sign Language Recognition using



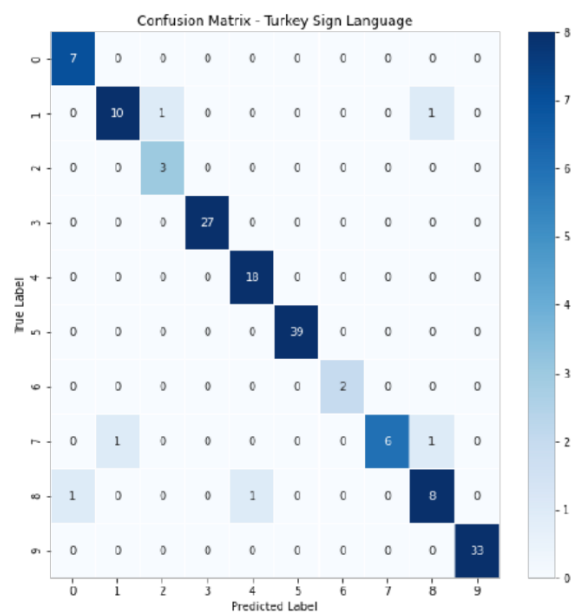
Εικόνα 2-22: Πίνακας Σύγκρισης: Αμερικανικής Νοηματικής



Εικόνα 2-23: Πίνακας Σύγκρισης: Αμερικανικής Νοηματικής

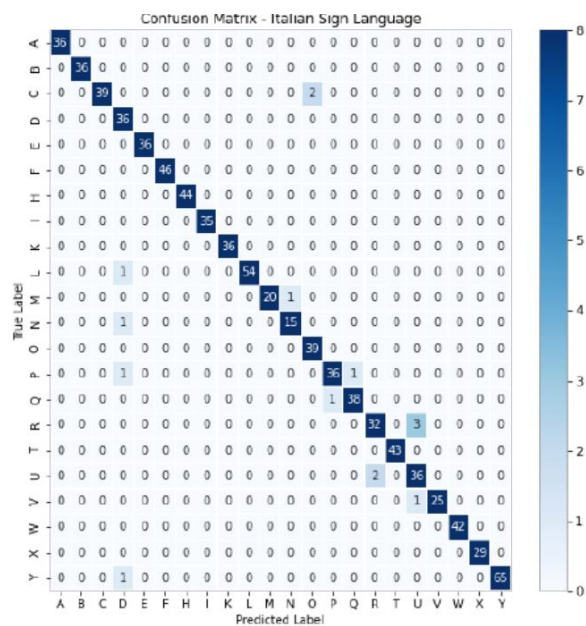


Εικόνα 2-24: Πίνακας Σύγκρισης: Ινδικής Νοηματικής



Εικόνα 2-25: Πίνακας Σύγκρισης: Τουρκικής Νοηματικής Γλώσσας (αριθμοί)





Εικόνα 2-26: Πίνακας Σύγχυσης: Ιταλικής Νοηματικής

Εικόνα 1 Πηγές Εικόνων 2-22. 2-23. 2-24. 2-25. 2-26: *International Journal of Research Publication and Reviews, 'Real-time Vernacular*

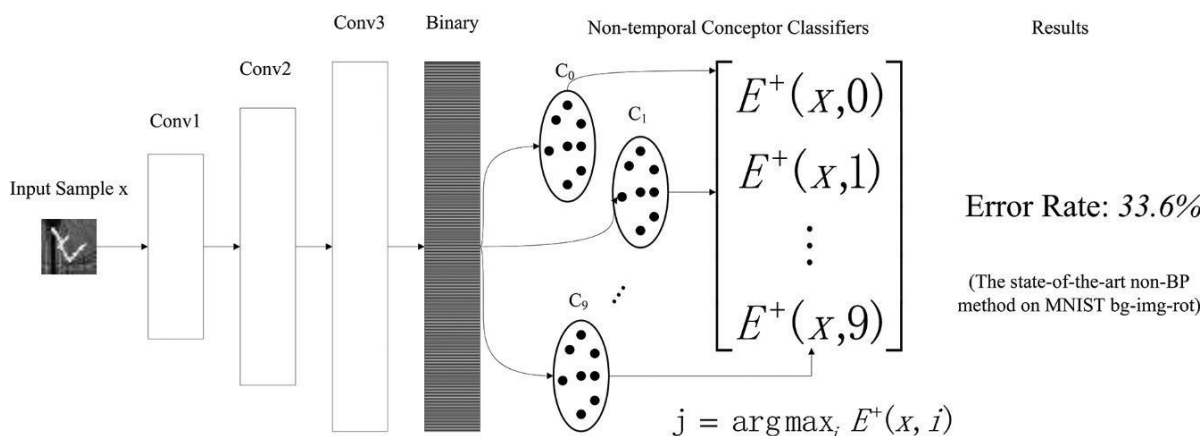
Συνοψίζοντας, το άρθρο παρουσιάζει έναν αποτελεσματικό και ακριβή τρόπο για την αναγνώριση της νοηματικής γλώσσας με χρήση τεχνικών μηχανικής και βαθιάς μάθησης, σε συνδυασμό με τεχνικές όρασης υπολογιστών και του πλαισίου του MediaPipe, με πολύ υψηλή μέση ακρίβεια, περίπου 99%, επιβεβαιωμένο σε διάφορα σύνολα δεδομένων, που αποτελούνται από χειρονομίες, που αντιστοιχίζονται σε σύμβολα νοηματικών γλωσσών. Τέλος, το ανεπτυγμένο μοντέλο αποδεικνύεται αποτελεσματικό, ακριβές και ανθεκτικό [14].

Στο επόμενο άρθρο [6], οι ερευνητές παρουσιάζουν μια μεθοδολογία εκπαίδευσης νευρικών δικτύων δίχως την ανάγκη για επαναληπτικές διαδικασίες (non-iterative training), με σκοπό την επιτάχυνση και απλοποίηση της διαδικασίας εκπαίδευσης των βαθιών νευρωνικών δικτύων. Η προτεινόμενη μέθοδος ονομάζεται Feedforward Convolutional Conceptor Neural Network (FCCNN) και χρησιμοποιείται στη συγκεκριμένη περίπτωση για κατηγοριοποίηση σε εικόνες.

Στο άρθρο γίνεται αναφορά για το γεγονός ότι η εκπαίδευση βαθιών νευρωνικών δικτύων, χρησιμοποιώντας τυπικές επαναληπτικές διαδικασίες (όπως το backpropagation), απαιτεί πολύ χρόνο εκπαίδευσης και υπολογιστικούς πόρους, πράγμα που καθιστά την εκπαίδευση, ακόμη και σε συστήματα υψηλής απόδοσης, δύσκολη. Σύμφωνα με τους συγγραφείς, παρατηρήσεις σε αυτό τον τομέα έχουν

οδηγήσει στην πρόταση μεθόδων εκπαίδευσης βαθιών δικτύων που δεν χρησιμοποιούν επαναληπτικές διαδικασίες.

Η μέθοδος FCCNN, που προαναφέρθηκε και αναλύεται εντός της επιστημονικής έρευνας, στηρίζεται σε ένα συνδυασμό τεχνικών που χρησιμοποιούν την Ανάλυση Κύριων Συνιστωσών (Principle Component Analysis (PCA)) για την ανάλυση χαρακτηριστικών των εικόνων, καθώς και σε μια νέα μέθοδο ταξινόμησης με την χρήση Συλληφτών (conceptors). Το αποτέλεσμα αυτής της διαδικασίας είναι ένα απλό και αποδοτικό νευρικό δίκτυο κατάλληλο ώστε να χρησιμοποιηθεί για την κατηγοριοποίηση εικόνων. Το FCCNN δοκιμάστηκε σε σύνολα δεδομένων βασισμένα στο MNIST και κατάφερε να επιτύχει ακρίβεια κατηγοριοποίησης παρόμοια με τις καλύτερες μεθόδους που χρησιμοποιούν επαναληπτικές διαδικασίες, ενώ παράλληλα απαιτεί σημαντικά λιγότερο χρόνο εκπαίδευσης.



Εικόνα 2-27: Γραφική Αναπαράσταση της Αρχιτεκτονικής του Μοντέλου FCCNN

Πηγή: <https://www.sciencedirect.com/science/article/abs/pii/S1568494617304982>

Ανακεφαλαιώνοντας, οι κύριες συνεισφορές αυτής της ερευνητικής αυτής εργασίας είναι η κατασκευή ενός ταξινομητή εικόνων, βασισμένο στην Ανάλυση Κύριων Συνιστωσών, ειδικός για μη-χρονικά δεδομένα, η αρχιτεκτονική του ανεπτυγμένου εμπρόσθιου νευρωνικού δικτύου, το FCCNN, ένα απλό μη-επαναληπτικό μοντέλο νευρικού δικτύου που χρησιμοποιεί τεχνικές όπως η ανάλυση των κυριότερων συνιστωσών και η χρήση δυαδικής κατώφλισης (binary thresholding) για την εξαγωγή χαρακτηριστικών. Τέλος, τα αποτελέσματα από των πειραμάτων που διεξάχθηκαν και αναλύθηκαν στο ερευνητικό άρθρο, δείχνουν ότι το FCCNN είναι ικανό να επιτύχει ακρίβεια κατηγοριοποίησης συγκρίσιμη με τις καλύτερες μεθόδους που χρησιμοποιούν επαναληπτικές διαδικασίες, αυξάνοντας σημαντικά την ταχύτητα εκπαίδευσης [6].

Οι συγγραφείς του άρθρου αυτού [15], επικεντρώνονται στην ανάγκη για ασφαλή και αποτελεσματικά συστήματα που διαφυλάσσουν το απόρρητο στην εφαρμογή της μηχανικής μάθησης (Machine Learning(ML)), ειδικά των Βαθιών Νευρωνικών Δικτύων (Deep Neural Networks(DNN)), σε ευαίσθητα δεδομένα όπως συστήματα συστάσεων, συστήματα ανίχνευσης απάτης στον τραπεζικό τομέα και συστήματα υγείας. Προτείνουν ένα πλαίσιο εργασίας που επιτρέπει την αποτελεσματική και ασφαλή αξιολόγηση σύγχρονων αλγορίθμων Μηχανικής Μάθησης χρησιμοποιώντας την τεχνική του ασφαλούς πολυμερικού υπολογισμού (Secure Multiparty Computation (MPC)).

Στο άρθρο, υπογραμμίζεται το γεγονός ότι πολλές παλαιότερες προσπάθειες για ασφαλή μηχανική μάθηση απαιτούν την αντικατάσταση προηγμένων αλγορίθμων του πεδίου, με προσεγγιστικές εκδοχές που είναι "MPC-friendly" πριν εφαρμοστούν τεχνικές ασφαλούς πολυμερικού υπολογισμού. Αυτή η προσέγγιση μπορεί να απαιτεί λεπτομερή ρύθμιση των αλγορίθμων και να οδηγήσει σε λιγότερο αποτελεσματικούς αλγορίθμους ή σε αλγορίθμους με χαμηλότερη ποιότητα μάθησης, η οποία διακρίνεται στα μειωμένα ποσοστά ακριβείας πρόβλεψης. Συνεχίζοντας σε αυτό το άρθρο, οι ερευνητές παρουσιάζουν έναν τρόπο που επιτρέπει την ασφαλή και αποτελεσματική υπολογιστική διεκπεραίωση των υπολογισμών που πρέπει να εκτελεστούν, για την εκπαίδευση των Βαθιών Νευρωνικών Δικτύων, ακόμη και όταν αυτοί οι υπολογισμοί θεωρούνται "MPC-unfriendly".

	Security/NW	Methods	Epochs	Time [s]	Accuracy [%]
FALCON	Passive/LAN	SGD	15	780	-
<b>Ours</b>		Adam	1	117	95.64
FALCON	Active/LAN	SGD	15	2,355	-
<b>Ours</b>		Adam	1	570	95.61
FALCON	Passive/WAN	SGD	15	16,110	-
<b>Ours</b>		Adam	1	4,537	95.64
FALCON	Active/WAN	SGD	15	37,185	-
<b>Ours</b>		Adam	1	11,516	95.61

Εικόνα 2-28: Σύγκριση του χρόνου εκπαίδευσης του 3DNN στο Σύνολο Δεδομένων MNIST

Πηγή: 'Adam in Private: Secure and Fast Training of Deep Neural Networks with Adaptive Moment Estimation', by Nuttapon Attrapadung, Koki Hamada, Dai Ikarashi, Ryo Kikuchi, Takahiro Matsuda, Ibuki Mishina, Hiraku Morita and Jacob C. N. Schuldt, (2021)

	Security	Setting	AlexNet [s]	VGG16 [s]
FALCON	Passive	LAN	10,892	523,127
<b>Ours</b>	Passive	LAN	3,139	43,150
FALCON	Active	LAN	41,537	2,051,751
<b>Ours</b>	Active	LAN	15,021	161,481
FALCON	Passive	WAN	23,489	575,699
<b>Ours</b>	Passive	WAN	49,833	347,928
FALCON	Active	WAN	75,838	2,240,515
<b>Ours</b>	Active	WAN	159,781	1,293,226

Εικόνα 2-29: Χρόνος εκτέλεσης ανά εποχή(epoch) για την Εκπαίδευση των Μοντέλων AlexNet και VGG16 στο Σύνολο Δεδομένων CIFAR-10

Πηγή: 'Adam in Private: Secure and Fast Training of Deep Neural Networks with Adaptive Moment Estimation', by Nuttapon Attrapadung, Koki Hamada, Dai Ikarashi, Ryo Kikuchi, Takahiro Matsuda, Ibuki Mishina, Hiraku Morita and Jacob C. N. Schuldt, (2021)

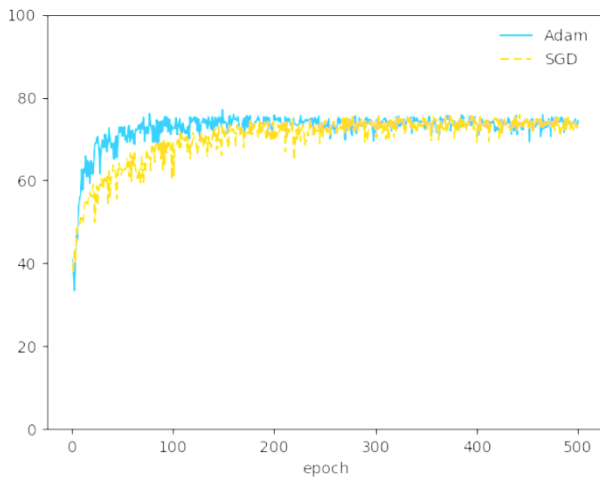
	Security	Setting	AlexNet [h]	VGG16 [h]
FALCON	Passive	LAN	324	3,342
<b>Ours</b>	Passive	LAN	22	72
FALCON	Active	LAN	1,235	13,108
<b>Ours</b>	Active	LAN	104	269
FALCON	Passive	WAN	698	3,678
<b>Ours</b>	Passive	WAN	346	580
FALCON	Active	WAN	2,254	14,314
<b>Ours</b>	Active	WAN	1,110	2,155

Εικόνα 2-30: Εκτιμώμενος χρόνος Εκπαίδευσης για το AlexNet (70% ποσοστό Ακριβείας) και VGG16 (75% ποσοστό Ακριβείας) στο Σύνολο Δεδομένων CIFAR-10

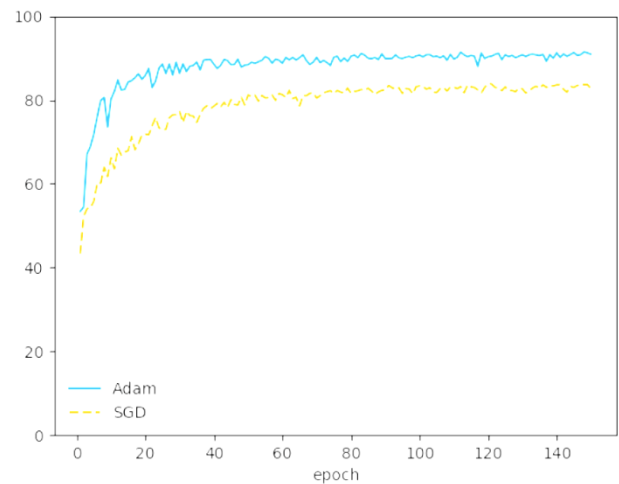
Πηγή: 'Adam in Private: Secure and Fast Training of Deep Neural Networks with Adaptive Moment Estimation', by Nuttapon Attrapadung, Koki Hamada, Dai Ikarashi, Ryo Kikuchi, Takahiro Matsuda, Ibuki Mishina, Hiraku Morita and Jacob C. N. Schuldt, (2021)

Συνοψίζοντας, το άρθρο περιγράφει πρωτόκολλα ασφαλών υπολογισμών για τους δύσκολους αυτούς υπολογισμούς, όπως η διαίρεση ακεραίων, η εκθετική λειτουργία, η αντιστροφή και η εξαγωγή της

τετραγωνικής ρίζας, υπολογισμοί που χαρακτηρίζονται ως "MPC-unfriendly". Τα πρωτόκολλα αυτά διατηρούν τόσο την ακρίβεια όσο και την αποτελεσματικότητα των υπολογισμών. Με αυτόν τον τρόπο, επιτυγχάνεται η δυνατότητα ασφαλούς εκπαίδευσης των Βαθιών Νευρωνικών Δικτύων, που αποδεικνύονται ισάξια, και σε μερικές περιπτώσεις καλύτερα από τα υπάρχοντα συστήματα τριών επιπέδων. Τέλος, κύριο πλεονέκτημα τους είναι η ταχύτητα σε σχέση με τα αντίστοιχα στάδια που προτείνει ένα πρόσφατο σύστημα FALCON, και ενσωματώνονται αβίαστα στα πολύπλοκα δίκτυα όπως το AlexNet και το VGG16 [15].



Εικόνα 2-31: Ακρίβεια του Μοντέλου AlexNet, εκπαιδευμένο με τους αλγορίθμους Adam και SGD στο Σύνολο Δεδομένων CIFAR-10



Εικόνα 2-32: Ακρίβεια του Μοντέλου VGG16, εκπαιδευμένο με τους αλγορίθμους Adam και SGD στο Σύνολο Δεδομένων CIFAR-10

Πηγή Εικόνων 2-31, 2-32: 'Adam in Private: Secure and Fast Training of Deep Neural Networks with Adaptive Moment Estimation', by Nuttapon Attrapadung, Koki Hamada, Dai Ikarashi, Ryo Kikuchi, Takahiro Matsuda, Ibuki Mishina, Hiraku Morita and Jacob C. N. Schuldt, (2021)

Το άρθρο αυτό [16], αναφέρεται στην επιρροή της εκπαίδευσης και της απόδοσης Βαθιών Νευρωνικών Δικτύων (Deep Neural Networks(DNN)), που φέρει η επιλογή συγκεκριμένων συναρτήσεων απώλειας. Τα Βαθιά Νευρωνικά Δίκτυα θεωρούνται από τα πιο διαδεδομένα μοντέλα βαθιάς μάθησης, είναι γνωστά για την ευελιξία τους, και συχνά χρησιμοποιούν τη συνάρτηση απώλειας log loss (cross-entropy) για τα προβλήματα ταξινόμησης.

Στο άρθρο, εξετάζονται δώδεκα διαφορετικές συναρτήσεις απώλειας, συμπεριλαμβανομένων των L1 loss, L2 loss, expectation loss, regularized expectation loss, Chebyshev loss, hinge loss, squared hinge loss, cubed hinge loss, log loss, squared log loss, Tanimoto loss Cauchy-Schwarz Divergence.

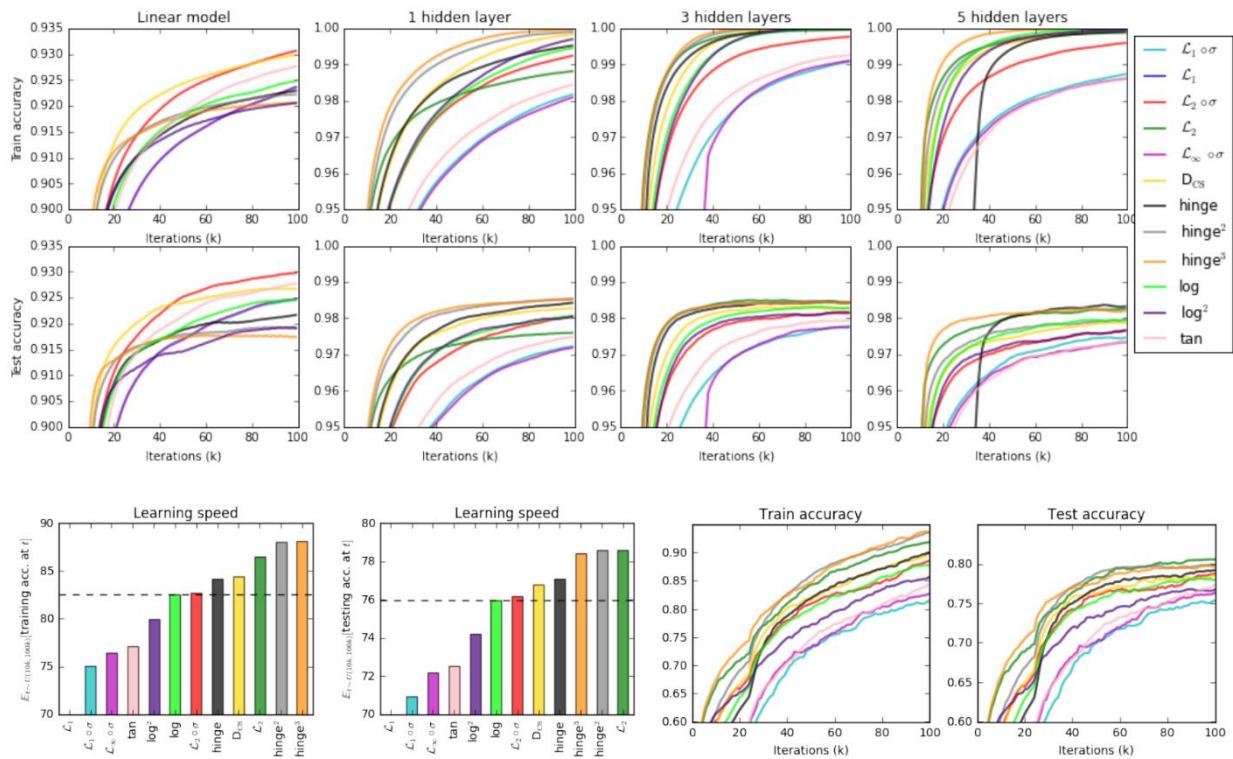
symbol	name	equation
$\mathcal{L}_1$	$L_1$ loss	$\ \mathbf{y} - \mathbf{o}\ _1$
$\mathcal{L}_2$	$L_2$ loss	$\ \mathbf{y} - \mathbf{o}\ _2^2$
$\mathcal{L}_1 \circ \sigma$	expectation loss	$\ \mathbf{y} - \sigma(\mathbf{o})\ _1$
$\mathcal{L}_2 \circ \sigma$	regularised expectation loss <sup>1</sup>	$\ \mathbf{y} - \sigma(\mathbf{o})\ _2^2$
$\mathcal{L}_\infty \circ \sigma$	Chebyshev loss	$\max_j  \sigma(\mathbf{o})^{(j)} - \mathbf{y}^{(j)} $
hinge	hinge [13] (margin) loss	$\sum_j \max(0, \frac{1}{2} - \hat{\mathbf{y}}^{(j)} \mathbf{o}^{(j)})$
hinge <sup>2</sup>	squared hinge (margin) loss	$\sum_j \max(0, \frac{1}{2} - \hat{\mathbf{y}}^{(j)} \mathbf{o}^{(j)})^2$
hinge <sup>3</sup>	cubed hinge (margin) loss	$\sum_j \max(0, \frac{1}{2} - \hat{\mathbf{y}}^{(j)} \mathbf{o}^{(j)})^3$
log	log (cross entropy) loss	$-\sum_j \mathbf{y}^{(j)} \log \sigma(\mathbf{o})^{(j)}$
log <sup>2</sup>	squared log loss	$-\sum_j [\mathbf{y}^{(j)} \log \sigma(\mathbf{o})^{(j)}]^2$
tan	Tanimoto loss	$\frac{-\sum_j \sigma(\mathbf{o})^{(j)} \mathbf{y}^{(j)}}{\ \sigma(\mathbf{o})\ _2^2 + \ \mathbf{y}\ _2^2 - \sum_j \sigma(\mathbf{o})^{(j)} \mathbf{y}^{(j)}}$
$D_{CS}$	Cauchy-Schwarz Divergence [3]	$-\log \frac{\sum_j \sigma(\mathbf{o})^{(j)} \mathbf{y}^{(j)}}{\ \sigma(\mathbf{o})\ _2 \ \mathbf{y}\ _2}$

Εικόνα 2-33: Λίστα των Συναρτήσεων Απώλειας, που αναλύονται στο άρθρο. Το  $y$  είναι η πραγματική ετικέτα ως κωδικοποίηση one-hot, το  $\hat{y}$  είναι η πραγματική ετικέτα ως κωδικοποίηση +1/-1, το  $\mathbf{o}$  είναι το αποτέλεσμα της τελευταίας στρώσης του δικτύου,  $\cdot^{(j)}$  υποδηλώνει την  $j$ -οστή διάσταση ενός δεδομένου διανύσματος, και η  $\sigma(\cdot)$  υποδηλώνει την πιθανότητα εκτίμησης.

Πηγή: ‘On Loss Functions for Deep Neural Networks in Classification’, by Katarzyna Janocha, Wojciech Marian Czarnecki, (2017)

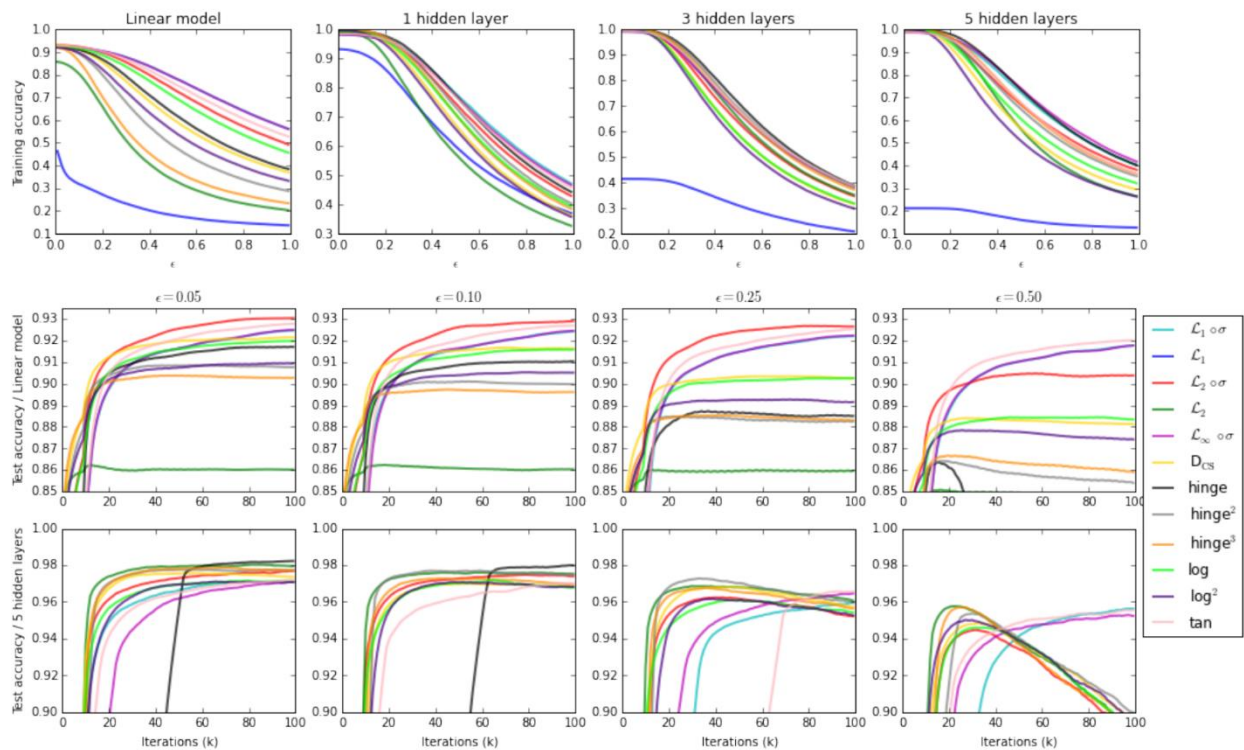
Πραγματοποιούνται θεωρητικές αναλύσεις για τις ιδιότητες των συναρτήσεων απώλειας και πειραματικές αξιολογήσεις για τα διάφορα μοντέλα που παράγονται σε συνδυασμό με αυτές τις συναρτήσεις απώλειας. Τα πειράματα αφορούν τόσο στατικά σύνολα δεδομένων όσο και την αντοχή του μοντέλου στην προσθήκη θορύβου στα δεδομένα εισόδου και στις ετικέτες των κατηγοριών.

Η σημαντικό συμπέρασμα από αυτή την εργασία είναι ότι η επιλογή της συνάρτησης απώλειας επηρεάζει την εκπαίδευση και την απόδοση των Βαθιών Νευρωνικών Δικτύων. Ενώ, η λογαριθμική απώλεια (log loss (cross-entropy)) είναι πολύ δημοφιλής, υπάρχουν εναλλακτικές συναρτήσεις απώλειας που μπορεί να είναι καταλληλότερες, ανάλογα με τον σκοπό της εφαρμογής ή του συστήματος. Στην περίπτωση που περισσότερο ενδιαφέρον, για τους μελλοντικούς ερευνητές ή προγραμματιστές, φέρει την ακρίβεια, η squared hinge loss αποδεικνύεται να είναι η καλύτερη επιλογή, καθώς συγκλίνει γρηγορότερα και παρέχει καλύτερη απόδοση. Ακόμα και στην περίπτωση όπου υπάρχει θόρυβος στα δεδομένα εκπαίδευσης, η συγκεκριμένη συνάρτηση απώλειας που φαίνεται να είναι η καλύτερη επιλογή.



Εικόνα 2-34: Οι δύο πρώτες γραμμές της εικόνας: Καμπύλες μάθησης για το Σύνολο Δεδομένων MNIST. Κάτω σειρά: (αριστερά) ταχύτητα μάθησης εκφρασμένη ως αναμενόμενη ακρίβεια εκπαίδευσης/δοκιμής όταν δειγματοληφτούνται οι επαναλήψεις ομοιόμορφα μεταξύ 10 χιλιάδων και 100 χιλιάδων. και (δεξιά) καμπύλες μάθησης για το σύνολο δεδομένων CIFAR-10.

Πηγή: 'On Loss Functions for Deep Neural Networks in Classification', by Katarzyna Janocha, Wojciech Marian Czarnecki, (2017)



Εικόνα 2-35:Επάνω σειρά: Καμπύλες ακρίβειας εκπαίδευσης για τα μοντέλα εκπαίδευσης MNIST, όταν παρουσιάζονται με παραδείγματα εκπαίδευσης με προσθήκη θορύβου από  $N(0, \epsilon)$ , σχεδιασμένες ως συνάρτηση του  $\epsilon$ . Μεσαία και κάτω σειρά: Καμπύλες ακρίβειας ελέγχου για το πείραμα MNSIT με την τιμή  $\epsilon$  των ετικετών εκπαίδευσης που άλλαξε, σχεδιασμένες ως συνάρτηση της επανάληψης της εκπαίδευσης. Αν το  $L_1$   $\circ$   $\sigma$  δεν είναι ορατό, σχεδόν ταυτίζεται απόλυτα με το  $L^\infty \circ \sigma$ .

Πηγή: ‘On Loss Functions for Deep Neural Networks in Classification’, by Katarzyna Janocha, Wojciech Marian Czarnecki, (2017)

Συνολικά, αυτή η ερευνητική εργασία επιχειρεί να ενισχύσει το εύρος των συναρτήσεων απώλειας που χρησιμοποιούνται στη βαθιά μάθηση, προσφέροντας πολλές εναλλακτικές που μπορούν να προσαρμοστούν στις ανάγκες της εκάστοτε εφαρμογής [16].

Στο τελευταίο άρθρο [40] που θα αναλυθεί, εξετάζεται η ανάπτυξη ενός συστήματος πρόβλεψης ενός στόχου-αντικειμένου χρησιμοποιώντας ένα Ρηχό Νευρωνικό Δίκτυο (Shallow Neural Network (SNN)) συνδυαστικά με τη συνάρτηση ενεργοποίησης Rectified Linear Unit (ReLU). Ο ερευνητής υποθέτει ότι τα εξωτερικά βάρη ακολουθούν μία πεπερασμένη  $L_1$ -νόρμα, ως προς το μέτρο Lebesgue στη σφαίρα. Επιπλέον, για τις μονοδιάστατες στοχοστραφείς συναρτήσεις, παρέχουν μια κλειστού τύπου φόρμουλα για όλες τις δυνατές αναπαράστασεις. Επιπρόσθετα, σε αυτή την περίπτωση, μπορεί να γίνει αναπαράσταση του Νευρωνικού Δικτύου με την ελάχιστη  $L_1$ -νόρμα για μια συγκεκριμένη συνάρτηση.



Η μέθοδος που συνιστούν οι ερευνητές στο άρθρο, για την διακριτοποίηση της αναπαράστασης του ολοκληρώματος της συνάρτησης που φέρει ενδιαφέρον, είναι η αναδιαμόρφωση της  $d$ -διάστατης συνάρτησης ως ένα ολοκλήρωμα της συγκεκριμένης συνάρτησης βάρους πάνω σε μια μονάδα σφαίρα διάστασης  $d + 1$ . Με βάση τα διαθέσιμα δεδομένα εκπαίδευσης, οι συγκεκριμένες ολοκληρώσεις προσεγγίζονται με μια διακριτή αθροιστική μέθοδο, παράγοντας την επιθυμητή αρχιτεκτονική του δικτύου.

Στη συνέχεια, εισάγεται ο χώρος  $W(\mathbb{R})$ , ο οποίος χαρακτηρίζει πλήρως την κατηγορία των συναρτήσεων που δέχονται την επιθυμητή αναπαράσταση μέσω ολοκληρώσεων. Αυτό επιτρέπει την επίλυση της αναπαράστασης του δικτύου με την ελάχιστη  $L1$ -νόρμα των εξωτερικών βαρών για μια συνάρτηση από τον χώρο  $W(\mathbb{R})$ .

Η προσέγγιση αυτή παρέχει μια θεωρητική κατανόηση για το πώς τα βάρη του δικτύου συμβάλλουν στην προσέγγιση των δεδομένων εκπαίδευσης. Η εργασία επικεντρώνεται στις αναπαραστάσεις ολοκληρωμάτων των Ρηχών Νευρωνικών Δικτύων, με τη συνάρτηση ενεργοποίησης ReLU και παρέχει πειραματική αξιολόγηση και τη δυνατότητα επίλυσης της αναπαράστασης με την ελάχιστη  $L1$ -νόρμα για ποικίλες συναρτήσεις [40].

## Κεφάλαιο 3<sup>ο</sup>

# ΜΕΘΟΔΟΛΟΓΙΑ ΑΝΤΙΚΕΙΜΕΝΟΣΤΡΑΦΟΥΣ ΑΝΑΛΥΣΗΣ – RATIONAL UNIFIED PROCESS (RUP)

Το συγκεκριμένο κεφάλαιο θα παρουσιάσει την μεθοδολογία ανάπτυξης που επιλέχθηκε για την εφαρμογής αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας, που υλοποιήθηκε στα πλαίσια της διπλωματικής εργασίας. Η μεθοδολογία αυτή είναι ευρέως γνωστή ως Rational Unified Process (RUP), και διευκόλυνε ιδιαίτερα στον τρόπο προσέγγισης επιλογής και ανάπτυξης της αρχιτεκτονικής και των λειτουργιών του συστήματος, όπως και στον τρόπο οργάνωσης της εξέλιξης των επιπέδων του, κατά την όλη πορεία από τη σύλληψη της ιδέας έως την ολοκλήρωσή της. Τα στάδια αυτά συνοδεύονται και από τα κατάλληλα διαγράμματα, όπου έχουν κριθεί απαραίτητα. Οι τέσσερις φάσεις στις οποίες γίνεται αναλυτική αναφορά παρακάτω είναι: η Φάση Σύλληψης, η Φάση Επεξεργασίας, η Φάση Κατασκευής και η Φάση Μετάβασης.

### 3.1 Φάση Σύλληψης (Inception Phase)

#### 3.1.1 Περιγραφή Έργου

Πρόκειται για μία εφαρμογή γραμμένη στη γλώσσα προγραμματισμού Python, η οποία αναπτύχθηκε με σκοπό την ενδυνάμωση της επικοινωνίας, της ελληνικής κοινότητας των κωφών με την υπόλοιπη κοινωνία. Η εφαρμογή διευκολύνει την αποτελεσματική επικοινωνία μεταξύ των κωφών και αφών ατόμων, καθώς επιτρέπει την αναγνώριση της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.), την αλφάβητό της, σε πραγματικό χρόνο μέσω της κάμερας του ηλεκτρονικού και φορητού υπολογιστή. Στο έργο αυτό αξιοποιήθηκαν το MediaPipe Hands API, για την αναγνώριση των χεριών, και πιο συγκεκριμένα για την αναγνώριση της παλάμης, τεχνικές όρασης υπολογιστών για την πρόσβαση στην κάμερα και για την δημιουργία εξατομικευμένου συνόλου δεδομένων.

#### 3.1.2 Στόχοι Έργου

- **Ενδυνάμωση της κοινότητας των κωφών** : Η ιδέα της εφαρμογής συνελήφθη με γνώμονα την αποτελεσματικότερη επικοινωνία των Ελλήνων κωφών με τους συνανθρώπους τους, παρέχοντας

τους ένα εργαλείο που τους επιτρέπει να επικοινωνούν ελεύθερα και πιο αποτελεσματικά στην καθημερινότητά τους.

- **Αναγνώριση της Νοηματικής Γλώσσας σε πραγματικό χρόνο** : Δευτερεύον στόχος που προσπαθεί η εφαρμογή αυτή να φέρει εις πέρας, είναι η αναγνώριση της Ελληνικής Νοηματικής Γλώσσας σε πραγματικό χρόνο και με όσο το δυνατόν μεγαλύτερη ακρίβεια, επιτρέποντας στους χρήστες να επικοινωνούν πιο φυσικά και ομαλά μέσω της νοηματικής γλώσσας, που συχνά αποτελεί και το κύριο μέσο έκφρασης τους.

### 3.1.3 Πεδίο εφαρμογής Έργου

- Επίκεντρο της εργασίας είναι η αναγνώριση του αλφαβήτου της Ελληνικής Νοηματικής Γλώσσας σε πραγματικό χρόνο, το οποίο αποτελείται από διαφορετικές μεμονωμένες χειρονομίες για κάθε γράμμα. Παρά όλο που σε αυτό το αρχικό επίπεδο, η εργασία εστιάζει στην αναγνώριση του ελληνικού αλφαβήτου, μελλοντικές επεκτάσεις μπορούν και να περιλαμβάνουν πιο περίπλοκες χειρονομίες ή σύνολα κινήσεων και εκφράσεις.

### 3.1.4 Τεχνολογίες που χρησιμοποιήθηκαν

- **Media Pipe Hands API** : Αξιοποιήθηκε το εργαλείο Media Pipe Hands API, ώστε να γίνεται η αναγνώριση, ο εντοπισμός και η παρακολούθηση της παλάμης σε πραγματικό χρόνο, δημιουργώντας έτσι τη βάση της αναγνώρισης της νοηματικής γλώσσας. Με τη χρήση και τη λειτουργία του εργαλείου εφαρμόζεται ένας «σκελετός» πάνω στην παλάμη, μέσω του οποίου μπορούμε να αντλήσουμε πολλές πληροφορίες για τη θέση και το σχήμα του χεριού [17].[18].[33].[34].[35].
- **Όραση Υπολογιστών** : Εφαρμόστηκαν τεχνικές της Όρασης Υπολογιστών (Computer Vision) για την πρόσβαση στην κάμερα του σταθερού ή φορητού ηλεκτρονικού υπολογιστή, την αποτύπωση και αποθήκευση στιγμιότυπων της κάμερας, που χρησιμοποιούνται για την δημιουργία του συνόλου δεδομένων, και για την επεξεργασία βίντεο ζωντανής μετάδοσης, για την αναγνώριση της ελληνικής νοηματικής γλώσσας και την εμφάνιση των γραμμάτων που αναγνωρίζονται, όπως και το ποσοστό ακρίβειας του εκπαιδευμένου μοντέλου [5].[37].

### 3.1.5 Δημιουργία Συνόλου Δεδομένων

- Ύστερα από μεγάλη και διεξοδική έρευνα, αποφασίστηκε πως δεν υπάρχει επαρκές διαδικτυακό υλικό – εικόνες, για τη δημιουργία του Συνόλου Δεδομένων ή κάποιο διαθέσιμο ολοκληρωμένο Σύνολο Δεδομένων. Λόγω της προαναφερθείσας απουσίας, στην εργασία περιλαμβάνεται η λήψη και η επιμέλεια στιγμιότυπων από την κάμερα του υπολογιστή, αφού πρώτα έχει εφαρμοστεί και είναι σε λειτουργία το εργαλείο Media Pipe Hands API, έτσι ώστε να έχουμε στην κατοχή μας πληροφορίες για την παλάμη (δεξιά ή αριστερή), για την θέση των δαχτύλων και των εικοσιένα οροσίων που βρίσκονται πάνω στο χέρι. Το σύνολο δεδομένων και η δημιουργία του αποτελούν ένα από τα πιο σημαντικά μέρη της εργασίας, καθώς είναι ζωτικής σημασίας για την εκπαίδευση και την αξιολόγηση του μοντέλου μηχανικής μάθησης που αναπτύσσουμε.

### 3.1.6 Μοντέλο Μηχανικής Μάθησης

- **Αρχιτεκτονική:** Το μοντέλο αυτό μηχανικής μάθησης είναι ένα νευρωνικό δίκτυο τροφοδοσίας (Feedforward Neural Network), με πλήρως συνδεδεμένα επίπεδα.
- **Συνάρτηση Ενεργοποίησης:** Επιλέχθηκαν δύο συναρτήσεις ενεργοποίησης για την κατασκευή της εφαρμογής. Η συνάρτηση ενεργοποίησης GeLU και η Swish. Τα δύο αυτά ανεπτυγμένα μοντέλα μπορούν να εντοπίσουν και να μάθουν πολύπλοκα χαρακτηριστικά και μοτίβα μεταξύ των δεδομένων. Η επιλογή των συναρτήσεων ενεργοποίησης έγινε μετά από σημαντική έρευνα, τα αποτελέσματα της οποίας θα αναλυθούν σε επόμενο κεφάλαιο [40].
- **Συνάρτηση Ενεργοποίησης Εξόδου:** Χρησιμοποιείται η συνάρτηση ενεργοποίησης εξόδου Softmax, και εφαρμόζεται στο τελευταίο επίπεδο του μοντέλου για να είναι δυνατή η κατηγοριοποίηση πολλαπλών κλάσεων [21].[40].
- **Συνάρτηση Απώλειας:** Εφαρμόζεται η κατάλληλη συνάρτηση απώλειας για την κατηγοριοποίηση πολλαπλών κλάσεων, η οποία κατευθύνει την εκπαίδευση του μοντέλου, προσπαθεί να ελαττώσει τις λανθασμένες προβλέψεις του προσαρμόζοντας συνεχώς τις παραμέτρους του και να μεγιστοποιήσει την επίδοση του [16].

### 3.1.7 Δικαιούχοι

- Το έργο αυτό προορίζεται κυρίως για τα μέλη της ελληνικής κοινότητας κωφών και βαρήκοων, δίνοντας τους τη δυνατότητα να επικοινωνούν και να εκφράζονται στην ελληνική νοηματική γλώσσα, πιο ελεύθερα και αποτελεσματικά. Επιπρόσθετα, από την εφαρμογή αυτή θα ωφεληθούν

εκπαιδευτικοί, διερμηνείς και οποιοσδήποτε άλλος ενδιαφέρεται να κατανοήσει και να μάθει την αλφαβήτα της ελληνικής νοηματικής γλώσσας.

### 3.1.8 Περιορισμοί

- **Διαθεσιμότητα Πόρων:** Λόγω της περιορισμένης διαθεσιμότητας πόρων του ηλεκτρονικού υπολογιστή, η εφαρμογή υλοποιήθηκε σε επίπεδο κονσόλας. Η έλλειψις υπολογιστική ικανότητα του υπολογιστή, επηρεάζει σημαντικά την απόδοση και την επίδοση της εφαρμογής.
- **Περιβαλλοντικοί Περιορισμοί:** Οι δυνατότητες του Jupyter Notebook από το Integrated Development Environment (IDE) του Anaconda, είναι περιορισμένες σε σχέση με την διαθεσιμότητα πληροφοριακών πόρων που χρειάζονται για την ανάπτυξη αυτόνομων εφαρμογών ή εφαρμογών με γραφική διεπαφή.
- **Περιορισμοί Εφαρμογής:** Το επιλεγμένο ανεπτυγμένο σύστημα λειτουργεί και δέχεται δεδομένα, δηλαδή τιμές των μεταβλητών που αντιστοιχούν στα εικοσιένα σημεία (landmarks), που εντοπίζονται πάνω στο χέρι, από τον αλγόριθμο Hands της MediaPipe και προέρχονται μόνο από στατικές δυσδιάστατες εικόνες. Δηλαδή, το συγκεκριμένο μοντέλο δεν είναι ικανό να αναγνωρίζει κινήσεις.

### 3.1.9 Εκτίμηση Κινδύνων

Κατά τη διάρκεια της ανάπτυξης σε επίπεδο κονσόλας της εφαρμογής αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.), στα πλαίσια της διπλωματικής εργασίας, έγινε μια αρχική εκτίμηση για την αναγνώριση και την αξιολόγηση ενδεχόμενων κινδύνων που ίσως εμφανιστούν κατά την εξέλιξη της εφαρμογής και γενικότερα κατά την πορεία του έργου.

- **Στόχος:** Ο πρωταρχικός στόχος της εκτίμησης των κινδύνων είναι η αναγνώριση, η καταγραφή, η κατανόηση, ο προσδιορισμός και η αξιολόγηση των πιθανών εμποδίων και των προσκομμάτων τους, όπως και των συνεπειών τους, που μπορεί να προκύψουν και να επηρεάσουν αρνητικά την σταθερή ανοδική πορεία της ολοκλήρωσης του έργου.
- **Μέθοδος:** Για την εκτίμηση των πιθανών κινδύνων, πραγματοποιήθηκε διεξοδικός έλεγχος του έργου. Έπειτα, από λεπτομερή ανάλυση και λαμβάνοντας υπόψη τους περιβαλλοντολογικούς περιορισμούς και γενικότερα τους περιορισμένους διαθέσιμους πόρους, αναγνωρίστηκαν πιθανά εμπόδια και προκλήσεις στη λειτουργικότητα της εφαρμογής.

- **Προσδιορισμός Ενδεχόμενων Κινδύνων:**

- **Δημιουργία εξειδικευμένου Συνόλου Δεδομένων:** Η αναζήτηση και η συλλογή κατάλληλων εικόνων και πληροφοριών, που θα αποτελέσουν το σύνολο δεδομένων για την Ελληνική Νοηματική Γλώσσα (Ε.Ν.Γ.), αποτελούν μια χρονοβόρα και εξαντλητική διαδικασία, καθώς πέρα από τη συλλογή εικόνων, έπεται η επεξεργασία και η τιτλοφόρηση τους.
- **Ανάπτυξη Αλγορίθμου Αναγνώρισης:** Η υλοποίηση του αλγορίθμου αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ) χρησιμοποιώντας το API του εργαλείου Media Pipe Hands, είναι πολύ πιθανό να προκαλέσει τεχνικά εμπόδια στην διαχείριση των δεδομένων που λαμβάνονται από την κάμερα σε πραγματικό χρόνο. Επιπρόσθετα, λόγω του περιβάλλοντος στο οποίο θα αναπτυχθεί η εφαρμογή, το οποίο έχει περιορισμένη δυνατότητα στην επεξεργασία μεγάλων σειρών δεδομένων, ενδέχεται να προκύψουν προκλήσεις στην εκπαίδευση και την βελτιστοποίηση του μοντέλου.
- **Ανάπτυξη Πηγαίου Κώδικα:** Καθ' όλη τη διάρκεια ανάπτυξης του συστήματος αυτού θα χρειαστούν και θα εκτελεστούν πολλές αλλαγές και μετατροπές στον πηγαίο κώδικα, μέχρι να φτάσει το έργο στην τελική του μορφή. Ενδέχεται να πραγματοποιηθούν αλλαγές, λόγω λογικών και τεχνικών σφαλμάτων, λανθασμένης δομής της αρχιτεκτονικής της εφαρμογής, επαναληπτικής εξέτασης, εκπαίδευσης και επιλογής του μοντέλου αναγνώρισης της νοηματικής γλώσσας, που θα ενσωματωθεί στην εφαρμογή, όπως και αλλαγές στο τυπικό περιβάλλον διεπαφής του χρήστη με την εφαρμογή.

### 3.1.10 Σχέδιο Έργου

- **Περιγραφή:** Το σχέδιο του Έργου αποτελεί τον σκελετό της διαδικασίας που θα ακολουθηθεί για την ανάπτυξη της εφαρμογής αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ), και επιδεικνύει τις καταστάσεις της πρώτης φάσης του RUP που θα πρέπει να επαναληφθούν, μέχρις ότου το αποτέλεσμα να είναι τουλάχιστον σε ικανοποιητικό επίπεδο.
- **Στόχος:** Σκοπός της επαναληπτικής αυτής διαδικασίας είναι η δόμηση μίας σαφής και σχολαστικά οργανωμένης πορείας αποτελώντας τον οδηγό, για τα διάφορα επίπεδα ανάπτυξης της εφαρμογής.

- **Μέθοδος:**

- **Επανάληψη 1<sup>η</sup>: Συλλογή Δεδομένων :** Η κύρια εργασία είναι η συλλογή και η δημιουργία ενός εξιδανικευμένου Συνόλου Δεδομένων από εικόνες που λήφθηκαν μέσω κάμερας, και απεικονίζουν τις αντίστοιχες χειρονομίες της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ) με εφαρμοσμένο τον αλγόριθμο Hands της media Pipe, ο οποίος είναι απαραίτητος για τη συλλογή των δεδομένων, αφού οι μετρήσεις βασίζονται στην εφαρμογή του.
- **Επανάληψη 2<sup>η</sup>: Προ-επεξεργασία Δεδομένων:** Στο στάδιο αυτό γίνεται ο έλεγχος, η επεξεργασία και η τιτλοφόρηση των δεδομένων που έχουν συλλεχθεί. Οι τίτλοι των δεδομένων αντιστοιχούν στα εικοσιτέσσερα γράμματα της ελληνικής αλφαβήτα, δηλαδή τα δεδομένα χωρίζονται σε 24 διαφορετικές κατηγορίες.
- **Επανάληψη 3<sup>η</sup>: Ανάπτυξη Μοντέλου:** Εδώ λαμβάνει χώρα η επιλογή του μοντέλου μηχανικής μάθησης που θα αξιοποιήσουμε για την αναγνώριση της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.), σε συνδυασμό με τον αλγόριθμο Hands της media Pipe και τεχνικές όρασης υπολογιστών.
- **Επανάληψη 4<sup>η</sup>: Αξιολόγηση Μοντέλου:** Πολύ σημαντικό στάδιο είναι η αξιολόγηση του μοντέλου μηχανικής μάθησης, που επιλέχθηκε να αναπτυχθεί και να εκπαιδευτεί, η οποία θα γίνει με δοκιμές ακριβείας και απόδοσης. Ουσιαστικά σε αυτό το στάδιο απορρίπτονται ή εγκρίνονται τα μοντέλα που ενδεχομένως να αναπτυχθούν και να χρησιμοποιηθούν.
- **Επανάληψη 5<sup>η</sup>: Συντακτική Αξιολόγηση:** Στο τελευταίο στάδιο, γίνεται η ανασκόπηση του κώδικα που έχει αναπτυχθεί μέχρι στιγμής για τη λειτουργία εφαρμογής, ακόμα, εντοπίζονται πιθανά σφάλματα και περιθώρια για βελτίωση.

### 3.1.11 Αρχικό Μοντέλο Περιπτώσεων Χρήσης (Use Case Diagram)

Παρακάτω φαίνεται ένα αρχικό μοντέλο περιπτώσεων χρήσης, στο οποίο φαίνονται επιγραμματικά όλες οι διαδικασίες ανάπτυξης ενός καλά εκπαιδευμένου και λειτουργικού μοντέλου αναγνώρισης μηχανικής μάθησης, που φέρει ως στόχο την αναγνώριση της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ) σε πραγματικό χρόνο. Το ανεπτυγμένο μοντέλο μηχανικής μάθησης, αναγνωρίζει επιτυχώς τα νοήματα και τις χειρονομίες που αντιστοιχούν σε γράμματα της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ), που απεικονίζουν οι χρήστες στην κάμερα τους.

### Διαχειριστής:

- Συλλογή εικόνων, που θα αποτελέσουν το **Σύνολο Δεδομένων**.
- Επεξεργασία εικόνων για την απόσπαση χαρακτηριστικών, και την δημιουργία του **Συνόλου Δεδομένων**.
- Επιλογή και εκπαίδευση **Μοντέλου Μηχανικής Μάθησης** που θεωρείται καταλληλότερο για την επίτευξη του σκοπού μας.
- Αξιολόγηση του **Μοντέλου Μηχανικής Μάθησης** που έχει επιλεγθεί, με βάση το ποσοστό ακριβείας του.
- Ανάπτυξη της **Εφαρμογής**.

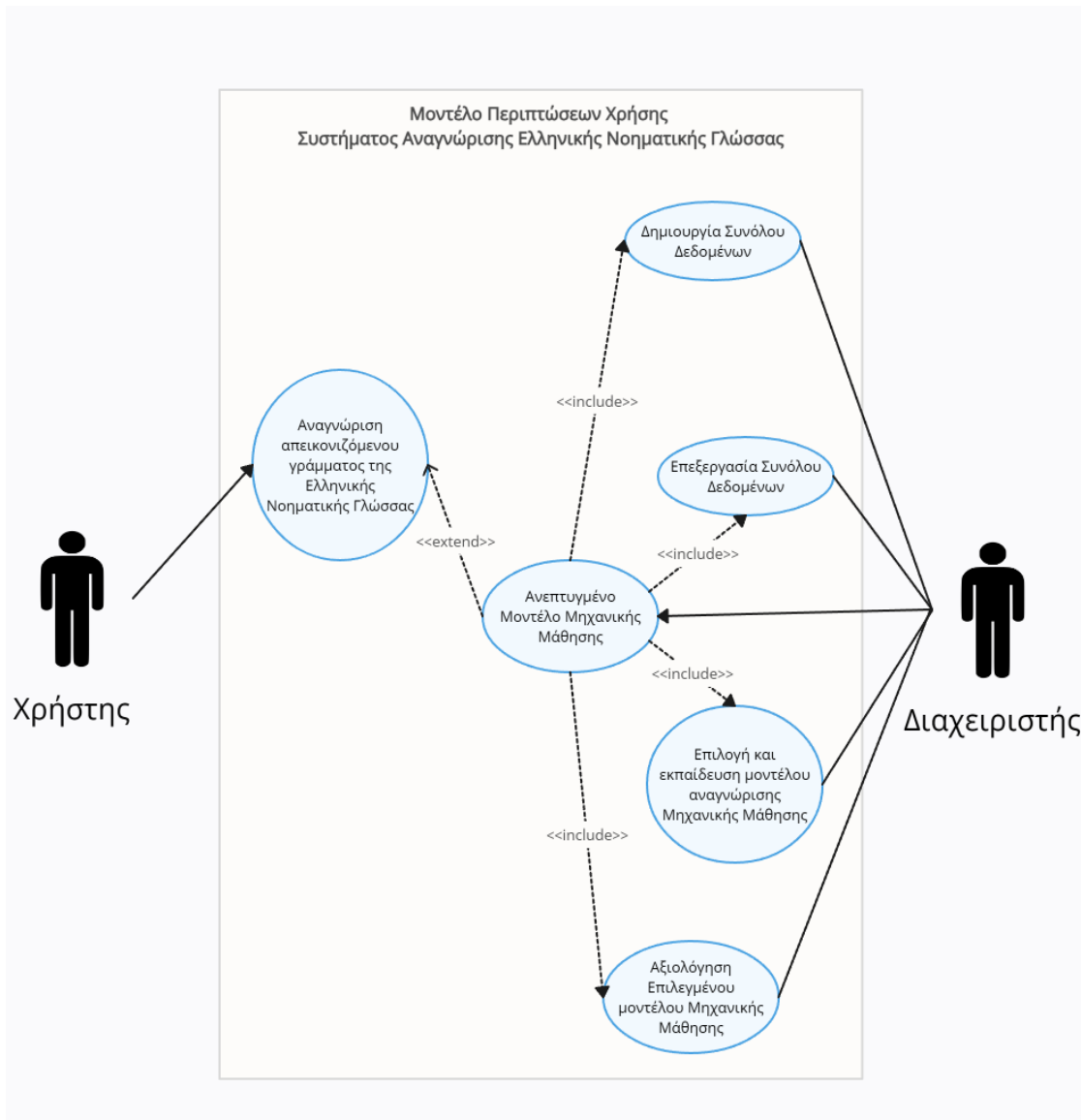
### Σύστημα:

- **Ανεπτυγμένο Μοντέλο** μηχανικής μάθησης, που **αναγνωρίζει** την Ελληνική Νοηματική Γλώσσα (Ε.Ν.Γ.), και πιο συγκεκριμένα τα σύμβολα – χειρονομίες που αποτελούν το ελληνικό αλφάβητα σε **πραγματικό χρόνο**.

### Χρήστης:

- Η **χειρονομία** ή αλλιώς το **σύμβολο** που ο χρήστης δείχνει στην κάμερα του υπολογιστή του, αναγνωρίζεται από το σύστημα.





Εικόνα 3-1: Αρχικό Μοντέλο Περιπτώσεων Χρήσης της εφαρμογής Αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας

## 3.2 Φάση Επεξεργασίας (Elaboration Phase)

### 3.2.1 Επέκταση Μοντέλου Περιπτώσεων Χρήσης (Use Case Diagram)

Παρακάτω φαίνεται ένα εξελιγμένο μοντέλο περιπτώσεων χρήσης, στο οποίο αναγράφονται επιγραμματικά όλες οι διαδικασίες ανάπτυξης που προστέθηκαν στο ανεπτυγμένο σύστημα, που φέρει ως

στόχο την αναγνώριση της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ) σε πραγματικό χρόνο, με τη χρήση της κάμερας του ηλεκτρονικού υπολογιστή.

#### **Διαχειριστής:**

- Η συλλογή εικόνων γίνεται μέσω της κάμερας του υπολογιστή και θα αποτελέσουν το **Σύνολο Δεδομένων**. Στις εικόνες αυτές έχει εφαρμοστεί ο κατάλληλος αλγόριθμος της Media Pipe, για την αναγνώριση και την παρακολούθηση των χειριών, ένας αλγόριθμος που βοηθάει στη συλλογή σημαντικών χαρακτηριστικών.
- Επεξεργασία εικόνων για την απόσπαση χαρακτηριστικών, που φέρουν κάποιο ενδιαφέρον, και την δημιουργία του **Συνόλου Δεδομένων**.
- Επιλογή και εκπαίδευση **Μοντέλου Μηχανικής Μάθησης** που θεωρείται καταλληλότερο για την επίτευξη του σκοπού μας.
- Αξιολόγηση του **Μοντέλου Μηχανικής Μάθησης** που έχει επιλεγεί, με βάση το ποσοστό ακριβείας του.
- Ανάπτυξη της **Εφαρμογής**.

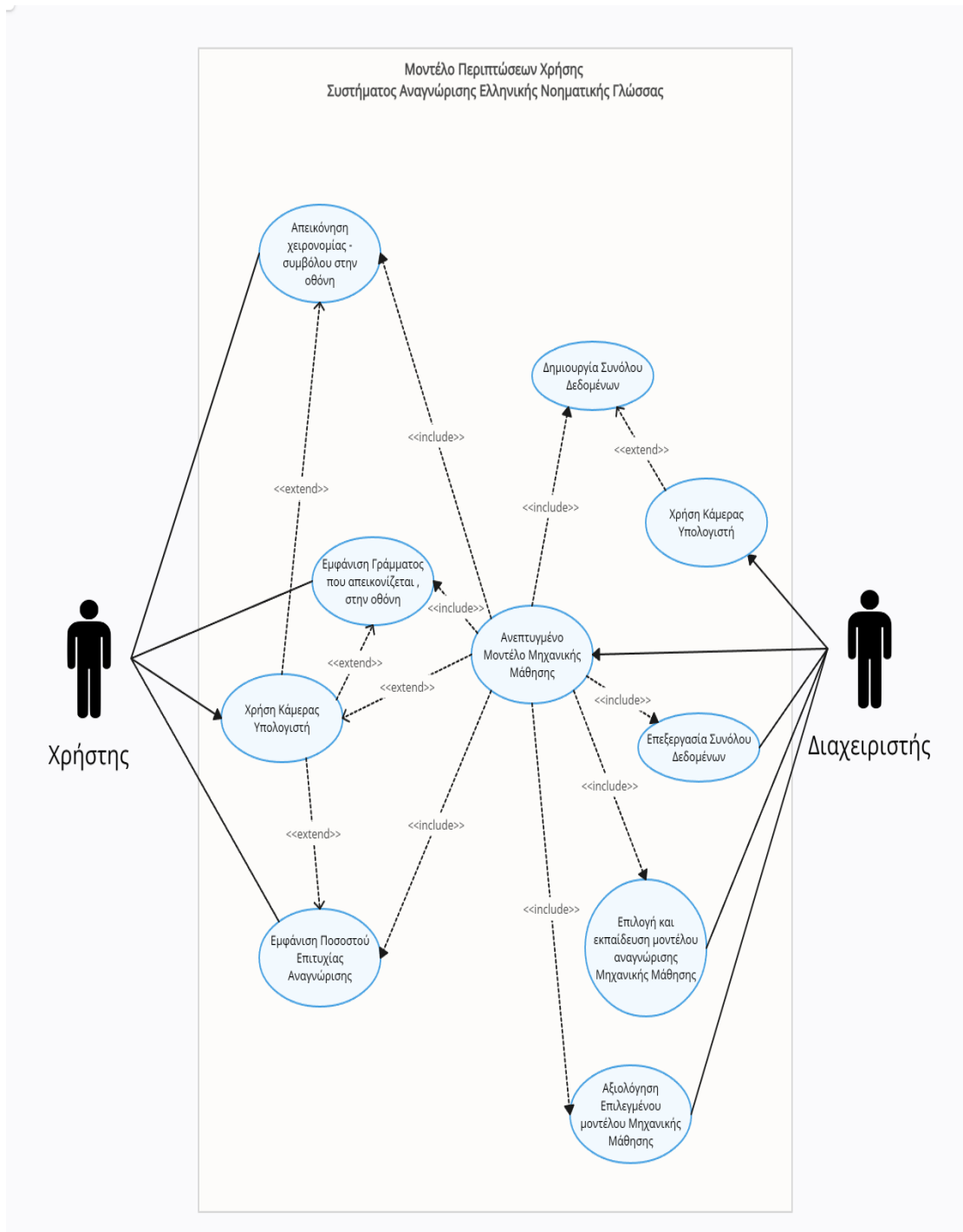
#### **Σύστημα:**

- **Ανεπτυγμένο Μοντέλο** μηχανικής μάθησης, που **αναγνωρίζει** την Ελληνική Νοηματική Γλώσσα (Ε.Ν.Γ.), και πιο συγκεκριμένα τα σύμβολα – χειρονομίες που αποτελούν το ελληνικό αλφαβήτα σε **πραγματικό χρόνο**.
- Το σύστημα είναι **απαραίτητο** να έχει **πρόσβαση** στην **κάμερα** υπολογιστή του **Χρήστη**. Θα την χρησιμοποιεί, ώστε να μπορεί να αναγνωρίζει τα σύμβολα και τις χειρονομίες που προβάλλει ο χρήστης.
- Το ανεπτυγμένο σύστημα είναι ικανό να **αναγνωρίζει** όλα τα **γράμματα** της **Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ)**.
- **Εμφανίζει** το **αναγνωρισμένο γράμμα** στην οθόνη ζωντανής μετάδοσης του Χρήστη.
- **Εμφανίζει** το **ποσοστό ακριβείας**, σύμφωνα με το οποίο αναγνωρίζει το γράμμα, στην οθόνη ζωντανής μετάδοσης του Χρήστη.

#### **Χρήστης:**

- Ο **Χρήστης** παρέχει στο σύστημα **πρόσβαση** στην **κάμερα** του ηλεκτρονικού του υπολογιστή.

- Η **χειρονομία** ή αλλιώς το **σύμβολο**, που ο χρήστης δείχνει στην κάμερα του υπολογιστή του, αναγνωρίζεται από το σύστημα.
- Η **χειρονομία** ή αλλιώς το **σύμβολο** που **αναγνωρίζεται** από το σύστημα **φαίνεται** στην οθόνη ζωντανής μετάδοσης του χρήστη.
- Η **χειρονομία** ή αλλιώς το **σύμβολο** που **αναγνωρίζεται**, φέρει και ένα **ποσοστό ακριβείας** το οποίο και αυτό **φαίνεται** στην οθόνη ζωντανής μετάδοσης του χρήστη.



Εικόνα 3-2: Μοντέλο Περιπτώσεων Χρήσης για την εφαρμογή Αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας

### 3.2.2 Περιγραφή Αρχιτεκτονικής Λογισμικού

Η Αρχιτεκτονική της εφαρμογής αποτελεί μια ολοκληρωμένη λύση για την αναγνώριση χειρονομιών σε πραγματικό χρόνο, χρησιμοποιώντας την ενσωματωμένη κάμερα μίας συσκευής. Αποτελείται από πέντε βασικά στοιχεία: την ανίχνευση των χεριών αξιοποιώντας τον αλγόριθμο Hands της MediaPipe, τη συλλογή των εικόνων με τα σημεία (landmarks) του αλγορίθμου Hands της MediaPipe για τη δημιουργία του συνόλου δεδομένων, το εκπαιδευμένο και αξιολογημένο μοντέλο μηχανικής μάθησης (M.M.), τον ταξινομητή, που σε συνδυασμό με το μοντέλο M.M., προσδιορίζει τις χειρονομίες που απεικονίζονται στο σύνολο δεδομένων και τη διεπαφή για την επικοινωνία με το χρήστη.

- **Ανίχνευση χεριών:** Η όλη διαδικασία της εφαρμογής ξεκινάει με την εκτέλεση του αλγορίθμου Hands της βιβλιοθήκης MediaPipe της Google, απαραίτητος για τον εντοπισμό των χεριών από καρέ εικόνας. Τα χέρια ανιχνεύονται σε πραγματικό χρόνο, ακόμα και όταν κινούνται. Επιπρόσθετα, ο αλγόριθμος αυτός εφαρμόζει και εικοσιένα σημεία (landmarks), πάνω στο χέρι, γεγονός που εξυπηρετεί στην εξαγωγή χαρακτηριστικών και τη δημιουργία του Συνόλου Δεδομένων.
- **Συλλογή εικόνων:** Συλλέγονται εικόνες και φωτογραφίες από την ενσωματωμένη κάμερα του υπολογιστή σε ζωντανή μετάδοση, πάνω στις οποίες έχει εφαρμοστεί ο αλγόριθμος Hands της MediaPipe. Αυτά που φέρουν ενδιαφέρον και αποθηκεύονται σε ένα αρχείο της μορφής .csv, είναι οι τιμές των χαρακτηριστικών των σημείων (landmarks) του αλγορίθμου. Καθώς αυτές οι τιμές «συγκρατούν» τις πληροφορίες για τη θέση του χεριού και των δαχτύλων.
- **Μοντέλο Μηχανικής Μάθησης:** Πάνω στο Σύνολο Δεδομένων που δημιουργήθηκε εκπαιδεύεται το μοντέλο μηχανικής μάθησης που έχει επιλεγεί. Έπειτα, το μοντέλο αυτό αξιολογείται από τα δεδομένα ελέγχου. Αν τα αποτελέσματα δεν είναι ικανοποιητικά, επιλέγεται άλλο μοντέλο μέχρις ότου παρατηρηθούν καλύτερα αποτελέσματα. Διαφορετικά, το μοντέλο αποθηκεύεται σε αρχείο HDF5, και μετατρέπεται σε αρχείο TensorFlow Lite, για χρήση σε φορητές συσκευές.
- **Ταξινομητής Χειρονομιών:** Δημιουργείται μία κλάση ταξινόμησης που χρησιμοποιεί το εκπαιδευμένο μοντέλο. Μία μέθοδος της κλάσης λαμβάνει μια λίστα από σημεία (landmarks) που ανιχνεύτηκαν. Το μοντέλο TensorFlow Lite που αναπτύχθηκε αναγνωρίζει τα σημεία. Στα αποτελέσματα της αναγνώρισης συμπεριλαμβάνονται ο τίτλος – ετικέτα της κάθε κατηγορίας – γράμματος και ο βαθμός εμπιστοσύνης της αναγνώρισης.
- **Διεπαφή Εφαρμογής:** Σε αυτό το σημείο λαμβάνει χώρα η διαχείριση της επικοινωνίας μεταξύ χρήστη και συστήματος. Τα αποτελέσματα που εμφανίζονται είναι το χέρι που ανιχνεύτηκε, δεξί

ή αριστερό, η χειρονομία που ανιχνεύτηκε, η βαθμολογία εμπιστοσύνης του μοντέλου και τα καρέ ανά δευτερόλεπτο (Frames Per Second). Άξιο αναφοράς, ο υπολογισμός των καρέ ανά δευτερόλεπτο, ο οποίος αποτελεί μία κλάση που κατασκευάζεται σε ξεχωριστό αρχείο .

Συνοψίζοντας, η αρχιτεκτονική του λογισμικού αποτελεί μία προηγμένη λύση για την αναγνώριση χειρονομιών σε πραγματικό χρόνο, και μπορεί να προσαρμοστεί σε ποικίλες εφαρμογές που απαιτούν αλληλεπίδραση με το χρήστη μέσω χειρονομιών. Είναι άκρως σημαντικό να σημειωθεί πως ο τρόπος που έχει δομηθεί το λογισμικό παρέχει ένα πολύ μεγάλο βαθμό ευελιξίας. Η αρχιτεκτονική του προσαρμόζεται πολύ εύκολα και είναι κατάλληλη για την ανάπτυξη εφαρμογών που πραγματεύονται την αναγνώριση κινήσεων. Το ανεπτυγμένο σύστημα είναι ικανό να εκπαιδεύεται συνεχώς σε καινούριες και πιο εξειδικευμένες χειρονομίες, και συνεπώς δεδομένα. Ναι μεν, το μοντέλο της μηχανικής μάθησης που αναπτύχθηκε στη συγκεκριμένη έκδοση της εφαρμογής, έχει εκπαιδευτεί σε εικοσιτέσσερις κατηγορίες - χειρονομίες, όσα και τα γράμματα της ελληνικής αλφαβήτα, αλλά αν ήταν επιθυμητό θα γινόταν να ταξινομήσει και συνεπώς να αναγνωρίσει παραπάνω κατηγορίες.

### 3.2.3 Επανεξέταση Κινδύνων

Καθ' όλη τη διαδικασία ανάπτυξης του λογισμικού και σε τακτά χρονικά διαστήματα, είναι απαραίτητο να διεξάγεται μια ανασκόπηση των κινδύνων και των εμποδίων που έχουν ήδη συζητηθεί. Παράλληλα με την επανεκτίμηση των κινδύνων, πραγματοποιείται μία διαδικασία οργάνωσης και προετοιμασίας στρατηγικών τεχνικών, για την αποφυγή και την αντιμετώπιση ενδεχόμενων εμποδίων. Παρακάτω φαίνονται οι δυνητικές δυσκολίες, αποτέλεσμα της εξέλιξης της πορείας της εφαρμογής και της επανεξέτασης των ήδη ορισμένων κινδύνων:

- **Δημιουργία εξειδικευμένου Συνόλου Δεδομένων:** Δεν υπάρχει διαθέσιμο σύνολο δεδομένων για την Ελληνική Νοηματική Γλώσσα (Ε.Ν.Γ), οπότε θα χρειαστεί να δημιουργηθεί ένα εκ νέου. Ενδεχομένως, η συγκεκριμένη διαδικασία να διαρκέσει περισσότερο χρόνο από ότι είχε εκτιμηθεί αρχικά. Βέβαια, για την μείωση του χρόνου, σε ένα βαθμό, αυτής της διαδικασίας, η τιτλοφόρηση και επεξεργασία των δεδομένων των εικόνων, έχουν ενσωματωθεί στις λειτουργίες του συστήματος.
- **Υποβάθμιση απόδοσης αλγορίθμου:** Η απόδοση του αλγορίθμου αναγνώρισης της νοηματικής γλώσσας, επηρεάζεται από ποικίλους παράγοντες, όπως ο φωτισμός του χώρου που βρίσκεται ο χειριστής την ώρα της σύνθεσης του συνόλου δεδομένων, ο φωτισμός του χώρου που βρίσκεται

ο χρήστης την ώρα της λειτουργίας της εφαρμογής και η ποιότητα της κάμερας που χρησιμοποιείται κατά τη διάρκεια της δημιουργίας του Συνόλου Δεδομένων, αλλά και για την αναγνώριση των χειρονομιών.

- **Νομοθεσία:** Οι λειτουργίες του συστήματος, είναι απαραίτητο να συμμορφώνονται με τις σχετικές νομοθετικές απαιτήσεις της εκάστοτε χώρας, περιοχής. Ιδιαίτερα πρέπει να ληφθούν υπόψη οι κανονισμοί περί προστασίας των προσωπικών δεδομένων και ιδιωτικότητας των χρηστών.
- **Ασφάλεια δεδομένων:** Η εφαρμογή ενδέχεται συλλέγει, να επεξεργάζεται και να αποθηκεύει εικόνες και να εκμαιεύει προσωπικά δεδομένα των χρηστών. Η ασφάλεια των δεδομένων να είναι υψηλή, είναι ζωτικής σημασίας και θα υπάρξουν σοβαρές συνέπειες με την κάθε πιθανή παραβίαση της.
- **Ανάπτυξη και συντήρηση λογισμικού:** Η διαδικασία ανάπτυξης ενός λογισμικού αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ), είναι μία συνεχής διαδικασία, που για να ακολουθήσουν τα επιθυμητά αποτελέσματα, αφού αναπτυχθεί και είναι σε λειτουργία το λογισμικό, πρέπει το σύστημα να ενημερώνεται και να βελτιώνεται με την πάροδο του χρόνου, ώστε να διασφαλιστούν υψηλά ποσοστά ακρίβειας και απόδοσης.
- **Ενσωμάτωση με άλλα λογισμικά:** Μελλοντικά, αν η εφαρμογή χρειαστεί να ενσωματωθεί με κάποιο άλλο λογισμικό ή σύστημα, πολύ πιθανό να προκύψουν προβλήματα συμβατότητας.

Επανεξετάζοντας τους κινδύνους και τα πιθανά εμπόδια, που ίσως επακολουθήσουν την πορεία ανάπτυξης της εφαρμογής αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.), ανακαλύφθηκαν καινούριοι κίνδυνοι και ρίσκα. Ο συνδυασμός όλων των πιθανών εμποδίων πρέπει να ληφθεί υπόψη στην ανάπτυξη της εφαρμογής, και πρέπει να διαχειριστούν με τέτοιο τρόπο, έτσι ώστε να ελαχιστοποιηθεί η επίδραση τους στην επιτυχή ολοκλήρωση του έργου.

### 3.2.4 Τρόποι Αντιμετώπισης – Πιθανές λύσεις

Παρακάτω παρουσιάζονται οι πιθανοί τρόποι επίλυσης των κινδύνων που έχουν εντοπιστεί και προαναφερθεί παραπάνω:

- **Περιορισμοί Υλικού και Απόδοσης Αλγορίθμου:** Λόγω του περιορισμού στις συσκευές υλικού, δηλαδή την έλλειψη κάμερας υψηλής ευκρίνειας, η εκπαίδευση του αλγορίθμου χρειάζεται να γίνει με εικόνες και δεδομένα σε ποικίλες συνθήκες φωτισμού, να βελτιστοποιηθεί ο αλγόριθμος για τις χαμηλές συνθήκες φωτισμού, και να χρησιμοποιηθούν τεχνικές επεξεργασίας εικόνας, όπως και

τεχνικές εξομάλυνσης και φιλτραρίσματος στην είσοδο της κάμερας που θα συμβάλλουν στη βελτίωση αναγνώρισης χειρονομιών σε ανακριβείς συνθήκες. Έτσι, θα επιτευχθεί η βελτίωση της απόδοσης του αλγορίθμου σε αυξομειωμένες συνθήκες φωτισμού.

- **Δημιουργία εξειδικευμένου Συνόλου Δεδομένων:** Διεξαγωγή συζήτησης με τις κοινότητες των κωφών και βαρήκοων του ελληνικού χώρου, ώστε να συνεισφέρουν στη συλλογή των δεδομένων.
- **Νομοθεσία:** Σχολαστική εξέταση και τήρηση όλων των νομοθετικών πλαισίων περί προστασίας των δεδομένων και ιδιωτικότητας. Περαιτέρω ανάπτυξη μηχανισμών ελέγχου και συνεχής ενημέρωση της εφαρμογής.
- **Ασφάλεια Δεδομένων:** Η αποθήκευση των δεδομένων των χρηστών είναι απαραίτητο να γίνεται μέσω ισχυρής κρυπτογράφησης και να συνοδεύεται με αυστηρούς κανόνες πρόσβασης και διαχείρισης δεδομένων.
- **Ανάπτυξη και Συντήρηση Λογισμικού:** Μεθοδολογίες ανάπτυξης λογισμικού, όπως είναι η Agile, πιθανώς να αποδειχτούν πολύ χρήσιμες στη διαχείριση των αλλαγών και των ενημερώσεων της εφαρμογής. Επίσης, με την αυτοματοποίηση των διαδικασιών ελέγχου και δοκιμών, θα γίνεται η ανίχνευση των σφαλμάτων με ταχύτερο ρυθμό.
- **Ανάπτυξη Πηγαίου Κώδικα:** Χρειάζεται να τεκμηριώνονται και να καταγράφονται όλες οι αλλαγές των λειτουργιών που ενδέχεται να συμβούν στον πηγαίο κώδικα, όπως και να υπάρχει πάντα η τελευταία λειτουργική και τεκμηριωμένη έκδοση της εφαρμογής πριν από την κάθε αλλαγή. Δηλαδή, να υπάρχει ένα σύστημα διαχείρισης των εκδόσεων του λογισμικού. Αποτέλεσμα αυτής της μεθοδολογίας θα είναι η διευκόλυνση της συντήρησης και της επέκτασης του ανεπτυγμένου συστήματος.

Η συνεχής παρακολούθηση της διαδικασίας ανάπτυξης του λογισμικού συστήματος είναι αναγκαία. Συχνά απαιτούνται αλλαγές ή προσθήκες, στις λειτουργίες οποιασδήποτε εφαρμογής που βρίσκεται υπό ανάπτυξη, σύμφωνα με τις ανάγκες και τις εξελίξεις που προκύπτουν. Η προσαρμογή σε καινούριες προκλήσεις είναι κρίσιμη και ζωτικής σημασίας για την επιτυχημένη ολοκλήρωση του έργου.

### 3.2.5 Σχέδιο Ανάπτυξης ολικού Έργου

Παρακάτω παρουσιάζεται επιγραμματικά ένα σχέδιο ανάπτυξης για το ολικό έργο που πραγματεύεται το θέμα της διπλωματικής εργασίας. Περιλαμβάνονται τα βήματα που ακολουθήθηκαν, όπως και τα αντίστοιχα χρονοδιαγράμματα που θεωρήθηκαν επαρκή, για την αντιμετώπιση των κινδύνων



που συναντήθηκαν κατά την πορεία του έργου. Οι κίνδυνοι προσεγγίστηκαν με τους τρόπους επίλυσης που έχουν προταθεί και καταγραφεί παραπάνω.

### **Προτεινόμενο Σχέδιο Ανάπτυξης:**

- **Φάση 1<sup>η</sup>: Προετοιμασία και Προσδιορισμός Απαιτήσεων (Διάρκεια: 3 Εβδομάδες)**

- **Καθορισμός Λειτουργιών και Περιορισμών Εφαρμογής:**

- **Λειτουργίες:** Εντοπισμός χεριού, παλάμης και αναγνώριση της Ελληνικής Νοηματικής Γλώσσας, μέσω της κάμερας του Ηλεκτρονικού Υπολογιστή, σε πραγματικό χρόνο, αξιοποιώντας ένα εκπαιδευμένο μοντέλο μηχανικής μάθησης. Εμφάνιση της αναγνωρισμένης χειρονομίας και της βαθμολογίας της ακριβείας του ανεπτυγμένου μοντέλου αναγνώρισης στην οθόνη του χρήστη.
- **Περιορισμοί:** Οι διαθέσιμοι πόροι είναι περιορισμένοι, για αυτό και σκοπός της εφαρμογής είναι να λειτουργεί και με κάμερες χαμηλής ευκρίνειας και συστήματα που δεν έχουν ισχυρή υπολογιστική δύναμη. Το σύστημα δεν θα εκπαιδευτεί στην αναγνώριση κινήσεων, αλλά μόνο σε στατικές χειρονομίες.

- **Συλλογή Δεδομένων:** Αποφασίστηκε, πως οι χειρονομίες πάνω στις οποίες θα εκπαιδευτεί το μοντέλο, θα είναι αυτές που αντιστοιχούν στα γράμματα της ελληνικής γλώσσας. Δηλαδή, θα κατασκευαστούν εικοσιτέσσερις διαφορετικές κατηγορίες χειρονομιών πάνω στις οποίες θα εκπαιδευτεί το μοντέλο. Στις εικόνες αυτές θα εφαρμοστεί ο αλγόριθμος Hands της MediaPipe, για τον εντοπισμό των σημείων (landmarks), που εφαρμόζει ο αλγόριθμος αυτός πάνω στο χέρι. Έτσι, θα εντοπίζονται οι διαφορετικές θέσεις της παλάμης και των δαχτύλων, και συνεπώς οι διαφορετικές χειρονομίες. Με την εξαγωγή των πληροφοριών των σημείων (landmarks), θα δημιουργηθεί το Σύνολο Δεδομένων. Το σύνολο Δεδομένων δεν θα αποτελείται από εικόνες, αλλά μόνο από σύνολα τιμών μεταβλητών αποθηκευμένα σε πίνακες.

- **Φάση 2<sup>η</sup>: Ανάπτυξη του Λογισμικού (Διάρκεια: 3 Μήνες)**

- **Αρχιτεκτονική Λογισμικού:** Αφού έχουν οριστεί οι λειτουργίες της εφαρμογής, έπεται ο σχεδιασμός της δομής του λογισμικού.

- **Διαχωρισμός Λειτουργιών:** Τα αρχεία πρέπει να είναι χωρισμένα και ξεκάθαρα ως προς τις λειτουργίες τους. **Υπάρχουν τρία κύρια αρχεία:** **1)** Το αρχείο που αποτελείται από τις κύριες λειτουργίες της εφαρμογής, περιλαμβάνει τη χρήση του εκπαιδευμένου μοντέλου μηχανικής μάθησης για την αναγνώριση της ελληνικής νοηματικής γλώσσας, την εμφάνιση των τιμών των χαρακτηριστικών ενδιαφέροντος, και τη συλλογή πληροφοριών από την επεξεργασία εικόνων, για το σύνολο δεδομένων. **2)** Το αρχείο κατασκευής, εκπαίδευσης και αξιολόγησης του μοντέλου μηχανικής μάθησης που προορίζεται για την ανάπτυξη της εφαρμογής. Σε αυτό το αρχείο αρχικά «διαβάζεται» το σύνολο δεδομένων, το επιλεγμένο μοντέλο εκπαιδεύεται σύμφωνα με αυτό, αξιολογείται, και τέλος εξάγεται ως αρχείο της μορφής TensorFlow Lite, για να μπορεί να αξιοποιηθεί. **3)** Το αρχείο αρχικοποίησης του ανεπτυγμένου μοντέλου, ως κλάση, που επιστρέφει την ετικέτα (label) που αντιστοιχεί σε κάποια από τις κατηγορίες των χειρονομιών και το ποσοστό ακριβείας του μοντέλου, για να μπορεί να αξιοποιηθεί στο κεντρικό αρχείο της εφαρμογής.

Όλα τα αρχεία είναι χωρισμένα και δομημένα σε κλάσεις.

- **Επαναχρησιμοποίηση Κώδικα:** Η δημιουργία μιας κλάσης σε ξεχωριστό αρχείο, που αρχικοποιεί το μοντέλο αναγνώρισης μηχανικής μάθησης αποτελεί μία πολύ καλή πρακτική για την αρχή της επαναχρησιμοποίησης του κώδικα.
  - **Διαχείριση Δεδομένων:** Η συλλογή και αποθήκευση των δεδομένων γίνεται σε αρχεία CSV. Πέρα από τις τιμές μεταβλητών που βρίσκονται αποθηκευμένες σε πίνακες, και υποδεικνύουν τη θέση της παλάμης και των δαχτύλων, δεν συγκεντρώνονται άλλα δεδομένα.
  - **Διεπαφή Χρήστη:** Επιλέχθηκε ως το επίπεδο ανάπτυξης της εφαρμογής η κονσόλα, καθώς μεγαλύτερο ενδιαφέρον είχαν οι λειτουργίες της εφαρμογής. Η απλοποίηση, σε κάποιο βαθμό, της διεπαφής του χρήστη διευκολύνει την ολοκλήρωση του έργου.
- **Επιλογή Γλώσσας Προγραμματισμού:** Επιλέχθηκε η γλώσσα προγραμματισμού Python, κατάλληλη για τη δημιουργία και την αξιοποίηση μοντέλων μηχανικής μάθησης, και γνωστή για το εύρος των τεχνικών ψηφιακής επεξεργασίας εικόνας και όρασης υπολογιστών.

- **Δοκιμές και Αποσφαλμάτωση:** Γίνονται συνεχής δοκιμές και αξιολογήσεις της εφαρμογής για την εξασφάλιση της ομαλής λειτουργίας της.
  
- **Δημιουργία Τεκμηρίωσης:** Δημιουργείται ένα αρχείο μέσα στο οποίο καταγράφονται όλες οι μεθοδολογίες και τεχνικές που χρησιμοποιήθηκαν, οι πιθανοί κίνδυνοι της εφαρμογής και οι τρόποι επίλυσής τους, οι εκδόσεις των αρχείων κώδικα που λειτουργούν.
  
- **Φάση 3<sup>η</sup>: Έλεγχος, Βελτιστοποίηση και Συντήρηση ( Διάρκεια: Συνεχής Εποπτεία)**
  - **Έλεγχος και Επαλήθευση:** Κατά τη διάρκεια ανάπτυξης της εφαρμογής προέκυψαν ποικίλα προβλήματα στην ομαλή εκτέλεση των αρχείων. Αρχικά, υπήρχε πρόβλημα με την εκτέλεση του αλγορίθμου Hands της MediaPipe, και κάθε φορά που εντοπιζόταν η παλάμη, τερμάτιζε η λειτουργία της εφαρμογής. Ακόμα, υπήρχε πρόβλημα με την ανάγνωση και την προβολή των ελληνικών γραμμάτων από το αρχείο κώδικα. Επίσης, αντιμετωπίστηκε μία δυσκολία στην εμφάνιση του ποσοστού ακριβείας του εκπαιδευμένου μοντέλου στην οθόνη του χρήστη. Τέλος, γίνεται μία επιβεβαίωση των λειτουργιών της εφαρμογής.
  
  - **Βελτιστοποίηση:** Πραγματοποιήθηκαν πολλές αλλαγές στην επιλογή του μοντέλου μηχανικής μάθησης. Πολλά διαφορετικά μοντέλα εκπαιδεύτηκαν και αξιολογήθηκαν, μέχρι που βρέθηκε αυτό με το μεγαλύτερο ποσοστό πρόβλεψης. Βέβαια, η αξιολόγηση του μοντέλου είναι μία συνεχής διαδικασία, που απαιτεί χρόνο, και σίγουρα ένα πολύ καλό Σύνολο Δεδομένων.
  
  - **Συντήρηση:** Η εφαρμογή, χρειάζεται να ενημερώνεται τακτικά, για να επιδιορθώνονται σφάλματα και να προστίθενται νέα χαρακτηριστικά και λειτουργίες, ανάλογα με τις ανάγκες των χρηστών. Το μοντέλο αναγνώρισης πρέπει να επανεκπαιδεύεται τακτικά, σύμφωνα με τις θεμιτές αλλαγές και προσθήκες που λαμβάνουν χώρα στο σύνολο δεδομένων.

- **Φάση 4<sup>η</sup> : Αξιολόγηση Εφαρμογής και Επανεξέταση Λειτουργιών ( Διάρκεια: Συνεχής Εποπτεία)**
  - **Συλλογή Σχολίων Χρηστών:** Η συμμετοχή των χρηστών ίσως κριθεί απαραίτητη για την ολοκλήρωση του έργου. Η ανατροφοδότηση και οι πληροφορίες που θα συλλεχτούν μετά από τη χρήση του συστήματος από τους ενδιαφερόμενους, ίσως κρίνουν την μελλοντική βελτιστοποίηση της εφαρμογής και τη διαδικασία επανεξέτασης των λειτουργιών. Επιπρόσθετα, είναι πολύ πιθανό οι χρήστες να αντιμετωπίσουν καινούρια προβλήματα που δεν προέκυψαν κατά την ανάπτυξη της εφαρμογής.
  - **Δοκιμές και Αποσφαλμάτωση:** Και σε αυτό το σημείο θα διεξαχθούν δοκιμές για τις λειτουργίες της εφαρμογής και για την ανίχνευση πιθανόν σφαλμάτων.
  - **Επανεξέταση Λειτουργιών:** Θα αξιολογηθούν οι λειτουργίες που έχουν ήδη ενσωματωθεί στο σύστημα, και σε συνδυασμό με τις παρατηρήσεις των χρηστών, θα αξιολογηθεί η ύπαρξη τρόπων βελτίωσης, και ποιες μπορεί να είναι οι μελλοντικές επεκτάσεις της εφαρμογής.
  - **Ενημέρωση Τεκμηρίωσης:** Το αρχείο της τεκμηρίωσης ενημερώνεται σε τακτά χρονικά διαστήματα, με τρόπο τέτοιο, που αντανακλά τις πιο πρόσφατες αλλαγές στον κώδικα, στην επιδιόρθωση των σφαλμάτων και τις καινούριες προσθήκες λειτουργιών.

### 3.2.6 Τελικό Μοντέλο Περιπτώσεων Χρήσης (Use Case Diagram)

Παρακάτω φαίνεται το τελικό μοντέλο περιπτώσεων χρήσης, στο οποίο αναγράφονται επιγραμματικά όλες οι λειτουργίες του ανεπτυγμένου συστήματος, κατάλληλο για την αναγνώριση της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ) σε πραγματικό χρόνο, με τη χρήση της κάμερας του ηλεκτρονικού υπολογιστή.

#### Διαχειριστής:

- Χρήση **κάμερας** ηλεκτρονικού υπολογιστή.
- Εφαρμογή του αλγορίθμου hands της MediaPipe, για τον **εντοπισμό** των χεριών και την **προβολή** των εικοσιένα σημείων(landmarks) ενδιαφέροντος στις παλάμες, σε πραγματικό χρόνο.
- Συλλογή **των δεδομένων των σημείων** σε μορφή πίνακα και αποθήκευση σε αρχείο της μορφής .CSV. Η αντιστοίχιση των δεδομένων με τις ετικέτες των χειρονομιών γίνεται αυτόματα, πατώντας

το κατάλληλο κουμπί. Οι φωτογραφίες από τις οποίες εξάγονται τα δεδομένα των σημείων δεν αποθηκεύονται.

- Επεξεργασία και έλεγχος των δεδομένων για τη δημιουργία του **Συνόλου Δεδομένων**.
- Επιλογή και εκπαίδευση **Μοντέλου Μηχανικής Μάθησης** που θεωρείται καταλληλότερο για την επίτευξη του σκοπού.
- Αξιολόγηση του **Μοντέλου Μηχανικής Μάθησης** που έχει επιλεγθεί, με βάση το ποσοστό ακριβείας του.
- Ανάπτυξη της **Εφαρμογής** γύρω από το εκπαιδευμένο μοντέλο.

#### **Σύστημα:**

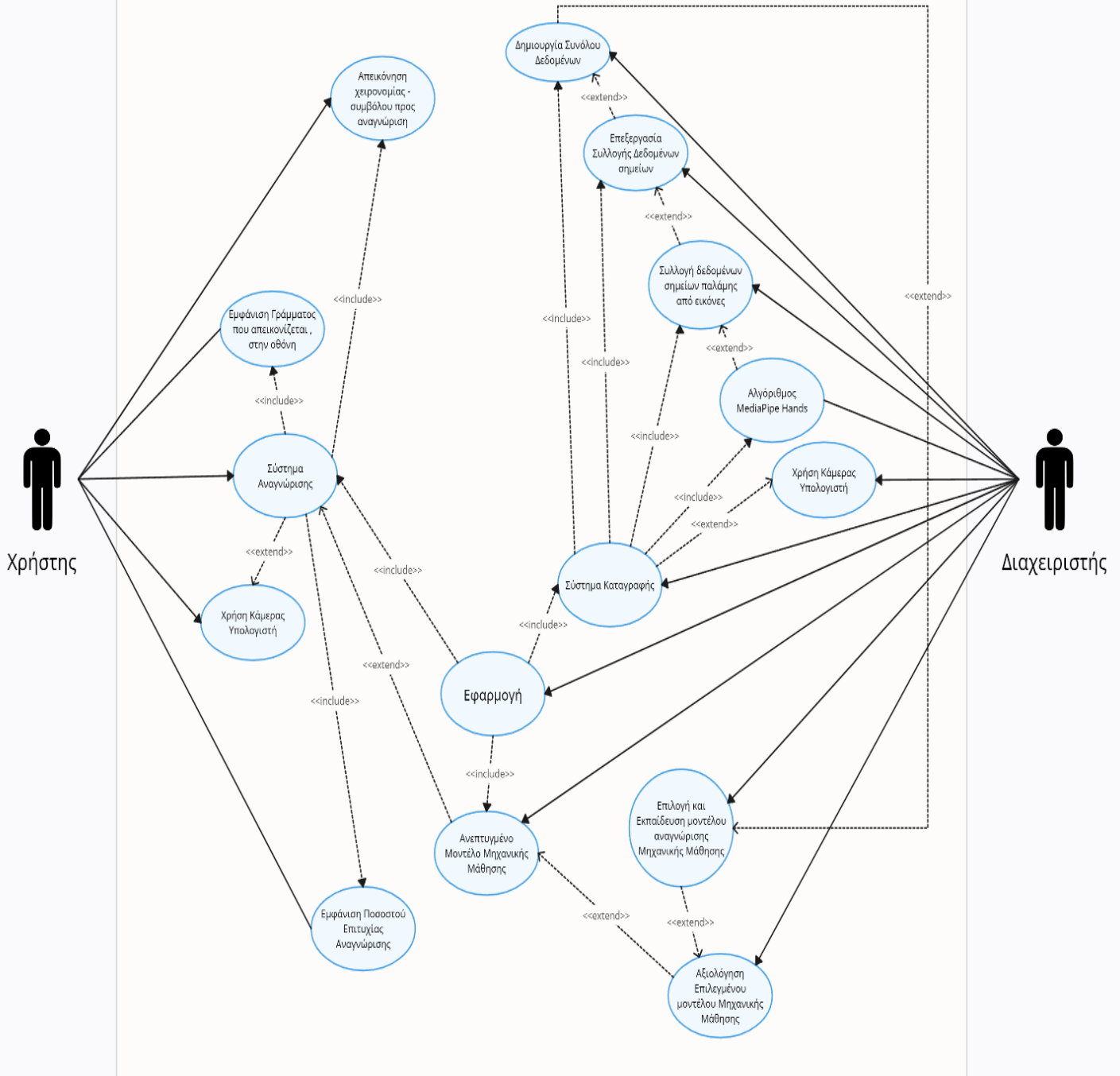
- **Ανεπτυγμένο Μοντέλο** μηχανικής μάθησης, που **αναγνωρίζει** την Ελληνική Νοηματική Γλώσσα (Ε.Ν.Γ.), και πιο συγκεκριμένα τα σύμβολα – χειρονομίες που αποτελούν το ελληνικό αλφαβήτα σε **πραγματικό χρόνο**.
- Το σύστημα είναι **απαραίτητο** να έχει **πρόσβαση** στην **κάμερα** υπολογιστή του **Χρήστη**. Θα την χρησιμοποιεί, ώστε να μπορεί να αναγνωρίζει τα σύμβολα και τις χειρονομίες που προβάλλει ο χρήστης.
- Το τελικό σύστημα είναι ικανό να **αναγνωρίζει** όλα τα **γράμματα της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ)**.
- **Εμφανίζει** το **αναγνωρισμένο γράμμα** στην οθόνη ζωντανής μετάδοσης του **Χρήστη**.
- **Εμφανίζει** το **ποσοστό ακριβείας** του μοντέλου στην οθόνη ζωντανής μετάδοσης του **Χρήστη**.

#### **Χρήστης:**

- Ο **Χρήστης** παρέχει στο σύστημα **πρόσβαση** στην **κάμερα** του ηλεκτρονικού του υπολογιστή.
- Η **χειρονομία** ή αλλιώς το **σύμβολο**, που ο χρήστης δείχνει στην κάμερα του υπολογιστή του, αναγνωρίζεται από το σύστημα.
- Η **χειρονομία** ή αλλιώς το **σύμβολο** που **αναγνωρίζεται** από το σύστημα **φαίνεται** στην οθόνη ζωντανής μετάδοσης του χρήστη.

- Η **χειρονομία** ή αλλιώς το **σύμβολο** που **αναγνωρίζεται**, φέρει και ένα **ποσοστό ακριβείας** το οποίο και αυτό **φαίνεται** στην οθόνη ζωντανής μετάδοσης του χρήστη.

Τελικό Μοντέλο Περιπτώσεων Χρήσης  
 Συστήματος Αναγνώρισης Ελληνικής Νοηματικής Γλώσσας



Εικόνα 3-3:Τελικό Μοντέλο Περιπτώσεων Χρήσης για την εφαρμογή Αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας

### 3.3 Φάση Κατασκευής (Construction Phase)

#### 3.3.1 Προϊόν Λογισμικού

Σε αυτό το σημείο, καθώς έχει γίνει όλη η προετοιμασία και προεργασία του σχεδίου ανάπτυξης του έργου, έχουν γίνει ξεκάθαρες οι λειτουργίες και οι ανάγκες που χρειάζεται να καλύψει η εφαρμογή, έχουν εντοπιστεί οι πιθανοί κίνδυνοι και έχουν βρεθεί ήδη λύσεις για την αντιμετώπιση τους, αρχίζει η ανάπτυξη της εφαρμογής σε επίπεδο κώδικα. Παρακάτω φαίνεται ένα διάγραμμα Gantt, που περιλαμβάνει όλα τα στάδια της ανάπτυξης του λογισμικού. Όλα τα βήματα και οι ενέργειες έχουν αναλυθεί σε προηγούμενες ενότητες.

#### Διάγραμμα Gantt

#### Εφαρμογής Αναγνώρισης Ελληνικής Νοηματικής Γλώσσας

Αρχικά, παρουσιάζεται ο πίνακας **Gantt**, στον οποίο φαίνονται επιγραμματικά ο τίτλος του ολικού έργου και κάποια δεδομένα των καταστάσεων της πορείας ανάπτυξης και ολοκλήρωσης του λογισμικού. Φαίνονται οι ημερομηνίες εκκίνησης και ολοκλήρωσης και η συνολική διάρκεια σε ημέρες της κάθε κατάστασης. Η κάθε κατάσταση αντιστοιχεί και σε ένα χρώμα, που βοηθάει στο να ξεχωρίζουν οι καταστάσεις μεταξύ τους, όταν μεταφράζονται σε οπτικά μέσα.

#### Πίνακας Gantt:

Name	Start Date	End Date	Duration	Progress %	Color
▼ Εφαρμογή Αναγνώρισης Ελληνικής Νοηματικής Γλώσσας	Mar 06, 2023	Sep 21, 2023	144 days	100	
Καθορισμός Λειτουργιών και Περιορισμών Εφαρμογής	Mar 06, 2023	Mar 14, 2023	7 days	100	Orange
Καθορισμός Αρχιτεκτονικής Εφαρμογής	Mar 15, 2023	Mar 28, 2023	10 days	100	Yellow
Ανάπτυξη Κώδικα για την Κατασκευή του Συνόλου Δεδομένων	Mar 29, 2023	Apr 10, 2023	9 days	100	Light Green
Κατασκευή Συνόλου Δεδομένων	Apr 11, 2023	Apr 24, 2023	10 days	100	Cyan
Ανάπτυξη Κώδικα για την Κατασκευή Μοντέλων Μηχανικής Μάθησης	Apr 25, 2023	May 15, 2023	15 days	100	Light Green
Εκπαίδευση και Αξιολόγηση Μοντέλων Μηχανικής Μάθησης	May 16, 2023	Jun 08, 2023	18 days	100	Blue
Επιλογή και Ανάπτυξη Μοντέλου Μηχανικής Μάθησης	Jun 09, 2023	Jun 15, 2023	5 days	100	Green
Ανάπτυξη Κώδικα για την ενσωμάτωση του ανεπτυγμένου Μοντέλου Μηχανικής Μάθησης στο σύστημα	Jun 16, 2023	Jul 10, 2023	17 days	100	Light Green
Ανάπτυξη Κώδικα για τη χρήση του ανεπτυγμένου μοντέλου M.M. στην Εφαρμογή	Jul 11, 2023	Jul 27, 2023	13 days	100	Light Green
Ανάπτυξη Κώδικα για την Διεπαφή του Χρήστη	Jul 28, 2023	Aug 14, 2023	12 days	100	Light Green
Έλεγχος και Αποσφαλμάτωση	Aug 15, 2023	Aug 23, 2023	7 days	100	Pink
Αξιολόγηση Εφαρμογής από μελλοντικούς Χρήστες	Aug 24, 2023	Sep 01, 2023	7 days	100	Magenta
Επανεξέταση Λειτουργιών Εφαρμογής και Βελτιστοποίηση	Sep 04, 2023	Sep 12, 2023	7 days	100	Purple
Κατασκευή Εγχειριδίου χρήσης , Ολοκλήρωση Έργου	Sep 13, 2023	Sep 21, 2023	7 days	100	Blue

Εικόνα 3-4: Πίνακας Gantt της διαδικασίας υλοποίησης της εφαρμογής Αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας



## Διάγραμμα Gantt:



Εικόνα 3-5: Διάγραμμα Gantt για τη διαδικασία υλοποίησης της εφαρμογής Αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας.

### 3.3.2 Εγχειρίδιο Χρήστη

Παρακάτω φαίνεται το εγχειρίδιο χρήστη για την εφαρμογή Αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ).

- **Εισαγωγή:**

- Πρόκειται για μία εφαρμογή κατάλληλη για την αναγνώριση της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.), με χρήση της κάμερας του υπολογιστή, σε πραγματικό χρόνο.
- Σκοπός του εγχειριδίου χρήστη είναι η σαφής περιγραφή των λειτουργιών της εφαρμογής και η εκτενής περιγραφή των βημάτων που χρειάζεται να ακολουθηθούν για την αποτελεσματική χρήση της εφαρμογής.

- **Εγκατάσταση και Εκκίνηση:**

- Η εφαρμογή υλοποιήθηκε στη γλώσσα προγραμματισμού python μέσω του Anaconda prompt στο περιβάλλον του Jupyter Notebook. Αποτελείται, από τρία βασικά αρχεία που, θα πρέπει να έχουν εγκατασταθεί όλα στο επιλεγμένο περιβάλλον ανάπτυξης της εφαρμογής, για την πρόσβαση σε όλες της λειτουργίες της.
- Απαραίτητο κομμάτι για την εκτέλεση των αρχείων είναι η εγκατάσταση των παρακάτω απαιτήσεων, πακέτων - βιβλιοθηκών της γλώσσας python, στις κατάλληλες εκδόσεις τους, που φαίνεται έπειτα του ονόματός τους :
  - **Mediapipe 0.8.1:** Η βιβλιοθήκη Mediapipe αποτελεί ένα πλαίσιο (API), που αναπτύχθηκε από την εταιρεία Google, χρήσιμο για τη δημιουργία εφαρμογών μηχανικής μάθησης που σχετίζονται με πολυμέσα (π.χ. ήχος και βίντεο). Η έκδοση 0.8.1 καθορίζει την ακριβή έκδοση της βιβλιοθήκης που απαιτείται για το σκοπό του έργου. Είναι σημαντικό να χρησιμοποιηθεί η συγκεκριμένη έκδοση για να εξασφαλιστεί η συμβατότητα και η ορθή λειτουργία.
  - **OpenCV 3.4.2 ή νεότερη έκδοση:** Το OpenCV (Open Source Computer Vision Library) είναι μια ευρέως γνωστή και χρησιμοποιούμενη, ανοικτού κώδικα, βιβλιοθήκη λογισμικού όρασης υπολογιστών και μηχανικής μάθησης. Απαραίτητη είναι η έκδοση της βιβλιοθήκης 3.4.2 ή μια νεότερη. Διάφορες

εκδόσεις μπορεί να έχουν προσθέσει χαρακτηριστικά, διορθώσεις σφαλμάτων ή αλλαγές, οπότε στη συγκεκριμένη περίπτωση, καθορίζεται η ελάχιστη έκδοση.

- **Tensorflow 2.3.0 ή νεότερη έκδοση:** Το TensorFlow είναι ένα πλαίσιο (API) μηχανικής μάθησης ανοικτού κώδικα. Η εταιρεία που το ανέπτυξε είναι η Google. Για να λειτουργεί η εφαρμογή σωστά αναγκαία είναι η εγκατάσταση του TensorFlow, στην έκδοση 2.3.0 ή μια νεότερη έκδοση. Όπως και στην περίπτωση του OpenCV, διάφορες εκδόσεις του TensorFlow μπορεί να εισάγουν νέα χαρακτηριστικά ή βελτιώσεις, επομένως συμβατότητα αποτελεί ένα πολύ κρίσιμο στοιχείο.
- **Scikit-learn 0.23.2 ή νεότερη έκδοση** (Απαραίτητο μόνο για την εμφάνιση του πίνακα σύγχυσης): Το scikit-learn είναι μια βιβλιοθήκη μηχανικής μάθησης για τη γλώσσα προγραμματισμού Python, συνήθως χρησιμοποιούμενη για εργασίες όπως η ταξινόμηση, η παλινδρόμηση και η συσταδοποίηση. Η βιβλιοθήκη αυτή δεν αποτελεί απαραίτητο στοιχείο των λειτουργιών του συστήματος, είναι απαραίτητη μόνο για την εμφάνιση του πίνακα σύγχυσης του μοντέλου, για την διευκόλυνση στην αξιολόγησή και την επιλογή του. Η καθορισμένη έκδοση είναι η 0.23.2 ή μεταγενέστερη, εξασφαλίζοντας την απαιτούμενη λειτουργικότητα για τη συγκεκριμένη εργασία.
- **Matplotlib 3.3.2 ή νεότερη έκδοση** (Απαραίτητο μόνο για την εμφάνιση του πίνακα σύγχυσης): Το matplotlib είναι μια δημοφιλής βιβλιοθήκη της γλώσσας προγραμματισμού Python κατάλληλη για τη δημιουργία στατικών, κινούμενων ή διαδραστικών οπτικοποιήσεων στατιστικών στοιχείων και δεδομένων. Όπως και η βιβλιοθήκη scikit-learn, έτσι και η συγκεκριμένη απαίτηση δεν είναι αναγκαία για την ορθή λειτουργία του συστήματος. Παρά όλα αυτά, αξιοποιείται και αυτή στην εμφάνιση του πίνακα σύγχυσης του εκπαιδευμένου μοντέλου αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας. Η καθορισμένη έκδοση είναι 3.3.2 ή μεταγενέστερη.

Συνοψίζοντας, αυτές οι απαιτήσεις εξασφαλίζουν ότι το λογισμικό έχει τις απαραίτητες απαιτήσεις-εξαρτήσεις και εκδόσεις για να λειτουργεί σωστά. Επιπλέον, καθορίζουν τις απαιτούμενες εκδόσεις για τις βιβλιοθήκες που αξιοποιούνται στην υλοποίηση του συστήματος, με βάση συγκεκριμένες

λειτουργίες του. Είναι σημαντικό να εγκατασταθούν οι παραπάνω απαιτήσεις, στις προαναφερόμενες ή και στις νεότερες εκδόσεις τους, στο ίδιο περιβάλλον ανάπτυξης τους συστήματος. Όπως έχει ήδη αναφερθεί, η εφαρμογή αναπτύχθηκε με τον μεταφραστή (compiler) του Jupyter Notebook, σε περιβάλλον Anaconda. Αφού δημιουργηθεί το περιβάλλον της εφαρμογής, μέσω του anaconda Prompt, τότε εγκαθίστανται οι όλες απαραίτητες βιβλιοθήκες, πάλι μέσω του anaconda Prompt, για να λειτουργήσει η εφαρμογή δίχως κάποιο σφάλμα ή και να αποφευχθούν δυσλειτουργίες στην κατασκευή του περιβάλλοντος ανάπτυξης.

- **Βασική Χρήση:**

- Μπαίνοντας στη διαδικασία χρήσης της εφαρμογής, ανοίγει η κάμερα του υπολογιστή του χρήστη. Το σύστημα μπαίνει κατευθείαν στη λειτουργία αναγνώρισης (mode = 0). Στην λειτουργία αυτή μπορείτε να επανέλθετε πατώντας το κουμπί 0 στο πληκτρολόγιο.
- Ο χρήστης απλά χρειάζεται να προβάλλει στην κάμερα του υπολογιστή του τις χειρονομίες που αντιστοιχούν στα Ελληνικά γράμματα της νοηματικής γλώσσας.
- Η χειρονομία αναγνωρίζεται. Το αναγνωριζόμενο γράμμα αναγράφεται στην οθόνη του υπολογιστή του χρήστη, σε συνδυασμό με το ποσοστό ακριβείας του εκπαιδευμένου μοντέλου.
- Στην οθόνη του χρήστη φαίνονται και τα καρτέ ανά δευτερόλεπτο (Frames per second).

- **Κρυφές Λειτουργίες (Διαχειριστής):**

- Είναι εφικτό, για κάποιον διαχειριστή να χρησιμοποιήσει το μοντέλο για την αναγνώριση πρόσθετων ή και διαφορετικών χειρονομιών εξ' ολοκλήρου, από αυτών που έχουν επιλεγεί.
- Κατά την εκτέλεση της εφαρμογής, είναι δυνατή η αλλαγή από τη λειτουργία αναγνώρισης, στη λειτουργία συλλογής δεδομένων, πατώντας το κουμπί 1 του πληκτρολογίου (mode = 1).
- Πατώντας το πλήκτρο 1, και αλλάζοντας λειτουργία, θα εμφανιστεί ο σκελετός και τα εικοσιένα σημεία (landmarks), που εφαρμόζονται πάνω στο εντοπισμένο χέρι, μέσω του κατάλληλου αλγορίθμου εντοπισμού.

- Έπειτα πατώντας τα πλήκτρα από «a» έως και «z», αποθηκεύονται τα δεδομένα των σημείων (landmarks) σε πίνακες, που κατέχουν τις πληροφορίες των χειρονομιών. Σε αυτό το σημείο γίνεται και η τιτλοφόρηση του κάθε δεδομένου αυτόματα, καθώς τα γράμματα από «a» έως και «z» αντιστοιχίζονται με τις κατάλληλες ετικέτες που βρίσκονται αποθηκευμένες σε ένα αρχείο της μορφής .CSV.
  - Εφικτές είναι και οι αλλαγές των ετικετών των χειρονομιών, στο κατάλληλο αρχείο.
  - Μετά τη δημιουργία του συνόλου δεδομένων, δίνεται η δυνατότητα στους διαχειριστές να εκπαιδεύσουν το μοντέλο αναγνώρισης στα νέα δεδομένα. Το μοντέλο εξάγεται αυτόματα σε μορφή TensorFlow Lite. Χωρίς καμία άλλη ενέργεια μπορούν οι διαχειριστές να χρησιμοποιήσουν το νέο τους μοντέλο στην εφαρμογή, και το σύστημα να αναγνωρίζει της χειρονομίες που επιθυμούν.
- **Έξοδος:**
    - Για να αποσυνδεθείτε από την εφαρμογή, κλείστε το αναδυόμενο παράθυρο που προκύπτει, πατώντας το κουμπί «escape» (esc).

### 3.4 Φάση Μετάβασης (Transition Phase)

Κατά τη φάση της μετάβασης στη μεθοδολογία RUP για την εφαρμογή αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας, πρέπει να προετοιμαστεί η εφαρμογή για παραγωγική χρήση, διανομή, και τεχνική στήριξη των χρηστών. Παρακάτω αναλύονται, λεπτομερής και διεξοδικά, πληροφορίες για κάθε στάδιο της φάσης μετάβασης:

- **Δοκιμές και Επαλήθευση**
  - **Δοκιμές Απόδοσης:** Θα διεξαχθούν δοκιμές απόδοσης για να εξασφάλιση της αξιόπιστης και αποδοτικής λειτουργίας της εφαρμογής κατά την επεξεργασία των εικόνων σε πραγματικό χρόνο. Θα πρέπει να υλοποιηθεί καταμέτρηση των χρόνων αναγνώρισης και της ακρίβειας.

- **Δοκιμές Ασφάλειας:** Θα επανεξεταστεί η ασφάλεια του συστήματος για να αποφευχθούν πιθανές επιθέσεις που μπορεί να απειλήσουν τα δεδομένα των χρηστών.
  - **Δοκιμές Χρήσης:** Η εφαρμογή χρειάζεται να αξιολογηθεί από διάφορους χρήστες για να διασφαλιστεί ότι η αναγνώριση, και συνεπώς η εκμάθηση, της Ελληνικής Νοηματικής Γλώσσας είναι εύκολη και κατανοητή.
  - **Δοκιμές Ενσωμάτωσης:** Είναι απαραίτητο να γίνει περαιτέρω προετοιμασία της εφαρμογής, για εγκατάσταση της σε διάφορα υπολογιστικά περιβάλλοντα, συμπεριλαμβανομένων διαφόρων λειτουργικών συστημάτων και υποστηριζόμενων κινητών συσκευών.
- **Βελτιστοποίηση και Απόδοση**
    - **Βελτιστοποίηση Κώδικα:** Θα λάβει χώρα εκτενέστερη εξέταση του κώδικα εφαρμογής για ευκαιρίες βελτιστοποίησης της απόδοσης, ιδίως σε ό,τι αφορά την επεξεργασία και τη συλλογή εικόνων.
    - **Βελτιστοποίηση Χρήσης Πόρων:** Θα αναθεωρηθούν και θα συζητηθούν τρόποι βελτιστοποίησης της χρήσης των πόρων, όπως της μνήμης και της CPU.
- **Εγκατάσταση και Διανομή**
    - **Εγκατάσταση στον Παραγωγικό Κατάλογο:** Η εφαρμογή θα πρέπει να περάσει από ένα ακόμα στάδιο αξιολόγησης, λειτουργώντας μέσα από το περιβάλλον ενός παραγωγικού καταλόγου. Είναι απαραίτητο η εφαρμογή να λειτουργήσει μέσω ενός τέτοιου περιβάλλοντος ώστε να έρθει αντιμέτωπη με πραγματικό φόρτο εργασίας, ταυτόχρονη εισαγωγή δεδομένων και διαχείριση χρηστών, ενώ παράλληλα να παρέχει σταθερά υψηλή απόδοση, για να ξεκινήσει η παραγωγή της.
    - **Διανομή σε Χρήστες:** Θα διανεμηθεί η εφαρμογή σε χρήστες, με προτεραιότητα ανθρώπους που πλήττουν από κάποια δυσχέρεια ή ανήκουν στην κοινότητα των κωφών και βαρήκοών, παρέχοντας λεπτομερείς οδηγίες εγκατάστασης και χρήσης.

- **Στήριξη και Συντήρηση:** Είναι απαραίτητο να δημιουργηθεί ένα σχέδιο στήριξης, προκειμένου να αντιμετωπιστούν πιθανοί προβληματισμοί, ερωτήσεις αλλά και προβλήματα κατά τη διάρκεια της χρήσης, των αρχικών καταναλωτών. Η εφαρμογή θα πρέπει να ενημερώνεται τακτικά και να επιλύονται τα επί μέρους εμπόδια.

- **Εκπαίδευση Διαχειριστών και Χρηστών**

- **Εκπαίδευση του Προσωπικού:** Εάν υπάρξει προσωπικό που χρησιμοποιεί την εφαρμογή σε θέση διαχειριστή, χρειάζεται να εκπαιδευτούν κατάλληλα σχετικά με τη χρήση της συλλογής δεδομένων και άλλων κρυφών λειτουργιών.
- **Εκπαίδευση των Χρηστών:** Με τη ολοκλήρωση και έπειτα με τη διανομή της εφαρμογής, θα παράσχετε ένας εύκολος και κατανοητός οδηγός χρήσης για τους τελικούς χρήστες. Ο οδηγός χρήσης θα περιλαμβάνει οδηγίες για την εκμάθηση και τη χρήση του αναγνώρισης Ελληνικής Νοηματικής Γλώσσας.
- **Στήριξη και Συμβουλές:** Στο σχέδιο στήριξης, που προαναφέρθηκε παραπάνω, θα πρέπει να συμπεριληφθούν η υποστήριξη και η συμβουλή των χρηστών για την αξιοποίηση της εφαρμογής σε ολόκληρο το ποσοστό της και την αποτελεσματική αναγνώριση της νοηματικής γλώσσας.

- **Συλλογή Και Ανάλυση Πληροφοριών**

- **Συλλογή Επιστροφών:** Θα ζητηθεί ανατροφοδότηση των χρηστών, σχετικά με την απόδοση και τη λειτουργία της εφαρμογής. Τα σχόλια αυτά θα αξιοποιηθούν περαιτέρω για τη βελτίωση της.
- **Παρακολούθηση Χρήσης:** Θα χρησιμοποιηθούν στατιστικά εργαλεία και εργαλεία ανάλυσης, για την καταγραφή του τρόπου χρήσης της εφαρμογής από τους χρήστες, και την βελτίωση της εμπειρίας τους.

Οι παραπάνω δράσεις που αναλύθηκαν διεξοδικά, ανταποκρίνονται στις απαιτήσεις του συστήματος αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας, κατά τη διάρκεια της φάσης της μετάβασης. Η εκπαίδευση του προσωπικού, η ασφάλεια των δεδομένων και η συνεχής επικοινωνία με τους χρήστες συμβάλουν στην επιτυχή μετάβαση και εισαγωγή της εφαρμογής στον παραγωγικό τομέα, την υποστήριξη της από το καταναλωτικό κοινό και την μελλοντική ανάπτυξη της ή ενσωμάτωσή της σε άλλα ήδη υπάρχοντα συστήματα.

## Κεφάλαιο 4<sup>ο</sup>

### ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΚΑΙ ΥΛΟΠΟΙΗΣΗ ΕΦΑΡΜΟΓΗΣ

Σε αυτό το κεφάλαιο θα αναλυθεί η εφαρμογή και οι λειτουργίες της σε επίπεδο κώδικα. Θα αναλυθούν τέσσερα αρχεία της γλώσσας προγραμματισμού python, που συντελούν στην κατασκευή της εφαρμογής αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ):

- Το αρχείο **app.py**, που αποτελεί το κύριο αρχείο. Οι περισσότερες λειτουργίες συμβαίνουν σε αυτό το αρχείο, και είναι το αρχείο που ουσιαστικά χρησιμοποιεί τα επόμενα.
- Το αρχείο **keypoint\_classification.py**. Το αρχείο αυτό είναι υπεύθυνο για την εκπαίδευση και την αξιολόγηση του μοντέλου μηχανικής μάθησης της εφαρμογής.
- Το αρχείο **KeyPointClassifier.py**. Σε αυτό το αρχείο δημιουργείται μία κλάση αρχικοποίησης του μοντέλου μηχανικής μάθησης, για την ευέλικτη χρήση του σε εφαρμογές.
- Το αρχείο **cvfpscale.py**. Σε αυτό το αρχείο ορίζεται μία κλάση υπεύθυνη για τον υπολογισμό των καρτέ ανά δευτερόλεπτο.

#### 4.1 Αρχείο app.py

##### 4.1.1 Ανάλυση Συναρτήσεων και Περιγραφή Λειτουργιών

Αρχικά θα αναλυθούν επιγραμματικά όλες οι συναρτήσεις του αρχείου, στο τέλος θα αναλυθεί και η κεντρική συνάρτηση (main) του αρχείου. Αφού γίνει η ανάλυση των συναρτήσεων θα ακολουθήσει μια περιγραφή των λειτουργιών του κώδικα του αρχείου.

- Συνάρτηση **get\_args()**:
  - Η συνάρτηση χρησιμοποιείται για την ανάλυση και την αλλαγή των παραμέτρων της γραμμής εντολών. Καθορίζει αρκετές διαφορετικές παραμέτρους, όπως device (συσκευή βιντεοσκόπησης), width (πλάτος παραθύρου προβολής ζωντανής μετάδοσης), height (ύψος παραθύρου προβολής ζωντανής μετάδοσης), use\_static\_image\_mode (χρήση στατικών εικόνων για τον εντοπισμό των χειρονομιών), min\_detection\_confidence και min\_tracking\_confidence (ελάχιστα όρια για τον εντοπισμό και την καταγραφή του χειριού). Η συνάρτηση επιστρέφει ένα αντικείμενο που περιέχει τις τιμές των παραμέτρων που δόθηκαν.



- Συνάρτηση **select\_mode(key, mode):**
  - Η συνάρτηση είναι υπεύθυνη για την επιλογή λειτουργίας της εφαρμογής, αναγνώριση των χειρονομιών με το εκπαιδευμένο μοντέλο (λειτουργία «0») ή συλλογή δεδομένων για τη δημιουργία του συνόλου δεδομένων (λειτουργία «1»). Παίρνει δύο ορίσματα : το key (πλήκτρο εισόδου που λαμβάνεται από τον χρήστη) και το mode (επιλεγμένη λειτουργία). Το πλήκτρο που θα πατήσει ο χρήστης (πλήκτρο «0» ή «1») λαμβάνεται από το σύστημα ως είσοδο, και γίνεται η αντίστοιχη επιλογή της λειτουργίας. Επιπρόσθετα, αν ο χρήστης έχει επιλέξει τη δεύτερη λειτουργία, πατώντας το πλήκτρο «1», δηλαδή τη λειτουργία συλλογής δεδομένων, και πατήσει ένα από τα πεζά γράμματα «a-z», τότε λαμβάνει χώρα μία ακόμα λειτουργία, στο επίπεδο συλλογής δεδομένων, που θα αναλυθεί παρακάτω. Τα γράμματα «a-z», που πληκτρολογούνται μετατρέπονται σε αριθμούς από «0-25» αντίστοιχα, και αποθηκεύονται προσωρινά στη μεταβλητή number. Με τη χρήση της συνάρτησης επιστρέφονται οι τιμές δύο μεταβλητών. Η τιμή της μεταβλητής number (0-25) και της μεταβλητής mode (0 ή 1).
- Συνάρτηση **clac\_bounding\_rect(image, landmarks):**
  - Με τη βοήθεια αυτής τη συνάρτησης, υπολογίζονται τα περιθώρια, πλευρές, του ορθογωνίου που περικλείει τα σημεία αναφοράς (landmarks) του χεριού. Δέχεται δύο ορίσματα : το image (πλαίσιο εικόνας) και τα landmarks (ανιχνευμένα σημεία αναφοράς του χεριού). Ανακτά τις συντεταγμένες (x, y), του κάθε σημείου αναφοράς και υπολογίζει τον ελάχιστο περιβάλλοντα χώρο του ορθογωνίου. Επιστρέφονται, σε μορφή λίστας, οι συντεταγμένες των σημείων: x, y, x+w, y+h. Οι συντεταγμένες των x και y, αναπαριστούν την πάνω αριστερά γωνία του ορθογωνίου, ενώ οι συντεταγμένες των x+w και y+h, αναπαριστούν την κάτω δεξιά γωνία του ορθογωνίου. Οι συντεταγμένες αυτές μπορούν να αξιοποιηθούν για την περαιτέρω επεξεργασία της περιοχής που περικλείεται το χέρι.
- Συνάρτηση **calc\_landmark\_list(image, landmarks):**
  - Η συνάρτηση αυτή ανακτά και επιστρέφει τις συντεταγμένες (x, y) όλων των εικοσιένα σημείων αναφοράς του χεριού. Λαμβάνει δύο ορίσματα: το image (πλαίσιο εικόνας) και τα landmarks (ανιχνευμένα σημεία αναφοράς του χεριού). Διαβάζει όλα τα σημεία

αναφοράς επαναληπτικά, μετατρέπει τις κανονικοποιημένες συντεταγμένες σε συντεταγμένες εικονοστοιχείων (pixel) και τις αποθηκεύει σε μία λίστα (`landmark_list`), την οποία και επιστρέφει.

- Συνάρτηση **`pre_process_landmark(landmark_list)`**:

- Εδώ διεξάγεται η προ επεξεργασία των δεδομένων των σημείων αναφοράς. Ως είσοδο στη συνάρτηση χρησιμοποιείται η λίστα των συντεταγμένων των αποθηκευμένων εικονοστοιχείων (`landmark_list`). Το σημείο αναφοράς των συντεταγμένων αλλάζει. Το νέο σημείο αναφοράς είναι πλέον ο καρπός του χεριού. Οι συντεταγμένες κανονικοποιούνται, σύμφωνα με το καινούριο σημείο αναφοράς, αποθηκεύονται σε μία μονοδιάστατη λίστα (`landmark_point`), και επιστρέφονται.

- Συνάρτηση **`logging_csv(number, mode, landmark_list)`**:

- Αυτή η συνάρτηση, σε συνδυασμό με τις συναρτήσεις `select_mode` και `calc_landmark_list`, που έχουν προαναφερθεί, υλοποιεί την καταγραφή των δεδομένων των σημείων αναφοράς του χεριού σε ένα αρχείο της μορφής CSV. Δέχεται τρία ορίσματα: το `number` (αντιστοιχίζεται από το πληκτρολογημένο γράμμα), το `mode` (λειτουργία αναγνώρισης χειρονομιών για `mode = 0`, λειτουργία συλλογής δεδομένων για `mode = 1`) και το `landmark_list` (λίστα συντεταγμένων σημείων αναφοράς). Αν κατά την εκτέλεση της εφαρμογής έχει επιλεγεί η λειτουργία «1» (`mode = 1`), για να συλλέξει δεδομένα, και παράλληλα ο χρήστης πληκτρολογήσει κάποιο από τα κουμπιά «a-z», δηλαδή βρίσκεται εντός του κανονικοποιημένου εύρους [0,25], τότε τα δεδομένα των σημείων αναφοράς, που φαίνονται εκείνη την ώρα στην οθόνη ζωντανής μετάδοσης, αποθηκεύονται στο αρχείο με όνομα «`keypoint.csv`».

- Συνάρτηση **`draw_landmarks(image, landmark_point)`**:

- Αυτή η συνάρτηση σχεδιάζει τα σημεία αναφοράς του χεριού στην εικόνα ζωντανής μετάδοσης. Παίρνει δύο ορίσματα: το `image` (πλαίσιο εικόνας) και το `landmark_point` (λίστα κανονικοποιημένων συντεταγμένων σημείων αναφοράς του χεριού). Οπτικοποιεί

τα κεντρικά σημεία του χεριού, όπως δηλαδή τα άκρα των δαχτύλων, ο καρπός και άλλοι σύνδεσμοι, σχεδιάζοντας κατάλληλα γραμμές και κύκλους. Επιστρέφεται η σχεδιασμένη εικόνα, στο παράθυρο ζωντανής μετάδοσης.

- **Συνάρτηση `draw_bounding_rect(use_brect, image, brect)`:**

- Αυτή η συνάρτηση κατασκευάστηκε για τη σχεδίαση ενός ορθογώνιου περιβλήματος γύρω από το χέρι του χρήστη, όταν αυτό έχει ανιχνευτεί. Τα ορίσματα της συνάρτησης είναι: το `use_brect` (μία λογική σημαία που παίρνει τιμές Αληθής - True ή Ψευδείς - False), το `image` (πλαίσιο εικόνας) και το `brect` (οι συντεταγμένες του ορθογώνιου περιγράμματος). Εάν, η τιμή του `use_brect` είναι Αληθής (True), δηλαδή όταν έχει ανιχνευτεί το χέρι, τότε προσαρμόζονται οι διαστάσεις του ορθογώνιου πλαισίου, και σχεδιάζεται το πλαίσιο στην οθόνη ζωντανής μετάδοσης.

- **Συνάρτηση `draw_info_text(image, brect, handedness, hand_sign_text)`:**

- Μέσω αυτής της συνάρτησης, στο ορθογώνιο πλαίσιο που περικλείει το χέρι του χρήστη, σχεδιάζεται το κείμενο των πληροφοριών, που θέλουμε να είναι διακριτό. Συμπεριλαμβάνονται πληροφορίες όπως οι ετικέτες - τίτλοι των χειρονομιών που αναγνωρίζονται από το εκπαιδευμένο μοντέλο μηχανικής μάθησης, η βαθμολογία εμπιστοσύνης του μοντέλου και το αν πρόκειται για το αριστερό ή το δεξί χέρι του χρήστη. Οι πληροφορίες φαίνονται σε ένα μαύρο πλαίσιο, τοποθετημένο πάνω από το ορθογώνιο πλαίσιο, του οποίου η θέση και οι διαστάσεις υπολογίζονται μέσω της συνάρτησης. Δέχεται τέσσερα ορίσματα: το `image` (πλαίσιο εικόνας), το `brect` (οι συντεταγμένες του ορθογώνιου περιγράμματος), `handedness` (πληροφορίες για το ποιο χέρι είναι) και `hand_sign_text` (κείμενο ετικέτας χειρονομίας).

- **Συνάρτηση `draw_info(image, fps, mode, number, confidence)`:**

- Η συνάρτηση αυτή δημιουργήθηκε, για το σχεδιασμό επιπρόσθετων πληροφοριών στην οθόνη ζωντανής μετάδοσης, όπως τα καρτέ ανά δευτερόλεπτο (Frames Per Second) και η τρέχουσα λειτουργία της εφαρμογής (`mode = 0` ή `mode = 1`). Χρειάζεται πέντε ορίσματα:

το `image` (πλαίσιο εικόνας), το `fps` (καρέ ανά δευτερόλεπτο), το `mode` (λειτουργία αναγνώρισης χειρονομιών για `mode = 0`, λειτουργία συλλογής δεδομένων για `mode = 1`), το `number` (αντιστοιχίζεται από το πληκτρολογημένο γράμμα) και το `confidence` (βαθμός εμπιστοσύνης εκπαιδευμένου μοντέλου αναγνώρισης). Εδώ πρέπει να σημειωθεί πως ο υπολογισμός των καρτέ ανά δευτερόλεπτο γίνεται σε άλλο εξωτερικό αρχείο, με τη χρήση κατάλληλης κλάσης, που θα αναλυθεί παρακάτω.

Οι παραπάνω συναρτήσεις είναι υπεύθυνες για τις εργασίες σχετικές με την ανίχνευση των χεριών, τον εντοπισμό των σημείων αναφοράς του χεριού, την οπτικοποίηση και την καταγραφή δεδομένων. Η συνάρτηση `main()` λειτουργεί ως το σημείο εισόδου και το κεντρικό ελεγκτικό στοιχείο της εφαρμογής. Επιπρόσθετα, συνδέει όλες αυτές τις συναρτήσεις, συντονίζει και διαχειρίζεται την εκτέλεση τους μέσα σε έναν ατέρμονο βρόγχο λήψης βίντεο σε πραγματικό χρόνο (εκτός αν πατηθεί το κατάλληλο πλήκτρο).

- **Συνάρτηση `main()`:**

- **Ρύθμιση και διαμόρφωση:**

- Η συνάρτηση ξεκινάει με την ανάλυση των παραμέτρων γραμμής εντολών χρησιμοποιώντας τη συνάρτηση `get_args()`. Οι παράμετροι αυτοί επιτρέπουν στο χρήστη να προσαρμόσει την καθορισμένη συσκευή κάμερας (`webcam`), τις διαστάσεις της εικόνας και τους ελάχιστους όρους εμπιστοσύνης του μοντέλου `Hands` της `MediaPipe`, για τον εντοπισμό και την καταγραφή του χεριού.

- **Ιστοκάμερα (`Webcam`):**

- Αρχικοποιείται και χρησιμοποιείται η ιστοκάμερα του ηλεκτρονικού υπολογιστή, για την λήψη βίντεο. Αξιοποιείται η βιβλιοθήκη της `OpenCV`, μέσω της οποίας ορίζονται οι διαστάσεις των καρτέ της κάμερας. Τα καταγεγραμμένα καρτέ αποθηκεύονται προσωρινά στην μεταβλητή `cap`.

- **Μοντέλο Εντοπισμού και Ανίχνευσης Χεριού `MediaPipe`:**

- Φορτώνεται ένα ήδη εκπαιδευμένο μοντέλο εντοπισμού και καταγραφής της παλάμης και των σημείων αναφοράς των χεριών σε πραγματικό χρόνο από την

τροφοδοσία της κάμερας του υπολογιστή, που παρέχεται από τη βιβλιοθήκη της MediaPipe. Χρησιμοποιείται σε ολόκληρη την εφαρμογή, τόσο στην κατάσταση λειτουργίας «0» όσο και στην κατάσταση λειτουργίας «1». Είναι απαραίτητο στην κατάσταση λειτουργίας «0», για τον εντοπισμό της παλάμης και την τοποθέτηση των σημείων αναφοράς στο χέρι, και στην κατάσταση λειτουργίας «1», για τη συλλογή δεδομένων της εντοπισμένης παλάμης σε πραγματικό χρόνο. Δεδομένα που προορίζονται για την περαιτέρω εκπαίδευση του προσαρμοσμένου αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.). Οι ρυθμίσεις του μοντέλου της MediaPipe, συμπεριλαμβανομένης της στατικής λειτουργίας εικόνας και των κατώτατων όρων εμπιστοσύνης, διαμορφώνονται βάσει των παραμέτρων της γραμμής εντολών.

➤ **Προσαρμοσμένο Μοντέλο Αναγνώρισης Ελληνικής Νοηματικής Γλώσσας:**

- Το προσαρμοσμένο μοντέλο μηχανικής μάθησης, για την αναγνώριση της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.) σε πραγματικό χρόνο, αξιοποιείται κυρίως στην κατάσταση λειτουργίας «0». Εντός αυτής της λειτουργίας, καταγράφονται τα καρτέ, εντοπίζονται τα σημεία αναφοράς της παλάμης, χρησιμοποιώντας το μοντέλο ανίχνευσης χεριών της MediaPipe, συλλέγονται και επεξεργάζονται τα δεδομένα των σημείων, σε πραγματικό χρόνο. Τα προεξεργασμένα δεδομένα τροφοδοτούνται στο ανεπτυγμένο, ειδικά για την αναγνώριση της Ε.Ν.Γ., μοντέλο μηχανικής μάθησης, το οποίο και εκτελεί την αναγνώριση των κατάλληλων στατικών χειρονομιών. Τα αποτελέσματα εμφανίζονται στο χρήστη επιτρέποντάς του να ερμηνεύει τις αναγνωρισμένες χειρονομίες.

➤ **Χειρισμός Πλήκτρων:**

- Οι χρήστες αλληλοεπιδρούν με την εφαρμογή πατώντας τα κατάλληλα πλήκτρα. Εντός ενός συνεχούς βρόχου, γίνεται έλεγχος για το πάτημα πλήκτρων χρησιμοποιώντας τη συνάρτηση `cv.waitKey()` της OpenCV. Η συνάρτηση `select_mode` ερμηνεύει το πάτημα του πλήκτρου και ενημερώνει αναλόγως την κατάσταση λειτουργίας της εφαρμογής. Επιλέγοντας τα κατάλληλα πλήκτρα «0»

ή «1», οι χρήστες μπορούν να μεταβούν σε δύο διαφορετικές καταστάσεις. Την κατάσταση λειτουργίας αναγνώρισης και την κατάσταση λειτουργίας συλλογής δεδομένων, «0» και «1» αντίστοιχα. Όταν βρίσκονται στην κατάσταση λειτουργίας «1», τους δίνεται η δυνατότητα, πληκτρολογώντας τα γράμματα «a-z», να αποθηκεύσουν τα δεδομένα της χειρονομίας που προβάλλουν στην κάμερα τους, στις αντίστοιχες θέσεις, με εύρος [0,25]., και με τις αντίστοιχες ετικέτες.

➤ **Λήψη και Επεξεργασία Εικόνας:**

- Εντός του βρόχου, όπως προαναφέρθηκε, λαμβάνονται καρέ από την κάμερα. Τα καρέ αναστρέφονται, με τέτοιο τρόπο ώστε η πλέον εμφάνιση τους στο παράθυρο της εφαρμογής να θυμίζει καθρέφτη. Η επεξεργασία των καρέ για την ανίχνευση των χεριών γίνεται αξιοποιώντας το μοντέλο Hands της MediaPipe. Τα εντοπισμένα σημεία αναφοράς επεξεργάζονται και καταγράφονται.

➤ **Οριοθέτηση Περιοχών και Σχεδίαση Σημείων Αναφοράς:**

- Ανάλογα με τη κατάσταση λειτουργίας, στην οποία βρίσκεται η εφαρμογή, οι οριοθετημένες περιοχές και τα σημεία των χεριών σχεδιάζονται στα καταγεγραμμένα καρέ για να οπτικοποιηθούν. Τις εργασίες σχεδίασης αυτές τις χειρίζονται οι συναρτήσεις `draw_bounding_rect` και `draw_landmarks`.

➤ **Εμφάνιση Πληροφοριών:**

- Τα καρέ, που εμφανίζονται στο παράθυρο εφαρμογής του χρήστη, φέρουν πληροφορίες που βοηθούν στην κατανόηση της κατάστασης της εφαρμογής. Τέτοιες πληροφορίες είναι: τα καρέ ανά δευτερόλεπτο, ο βαθμός εμπιστοσύνης του μοντέλου και η κατάσταση λειτουργίας. Οι πληροφορίες αυτές εμφανίζονται μέσω των συναρτήσεων `draw_info` και `draw_info_text`.

➤ **Εμφάνιση και Πραγματική Χρονική Απεικόνιση:**

- Τα επεξεργασμένα καρέ εμφανίζονται σε ένα παράθυρο με τίτλο «Hand Gesture Recognition», αξιοποιώντας την συνάρτηση `cv.imshow()` της OpenCV. Εδώ φαίνεται η πραγματική λειτουργία της εφαρμογής.

➤ **Συνθήκη Τερματισμού:**

- Ο ατέρμονος βρόχος συνεχίζεται μέχρι ο χρήστης να πατήσει το πλήκτρο `escape`, «Esc» που αντιστοιχεί στον αριθμό 27 στον ASCII κώδικα, δηλώνοντας την επιθυμία του χρήστη να τερματίσει την εφαρμογή. Αφού πατηθεί το κατάλληλο πλήκτρο και τερματιστεί ο βρόχος, απελευθερώνεται η λήψη βίντεο (`cap`) μέσω της κάμερας του υπολογιστή, και κλείνουν όλα τα παράθυρα της OpenCV, μέσω των συναρτήσεων `cv.release` και `cv.destroyAllWindows`.

Συνοψίζοντας, ο κώδικας του αρχείου `app.py`, ενσωματώνει διάφορα στοιχεία για να δημιουργήσει μια εφαρμογή αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ) σε πραγματικό χρόνο. Διαχειρίζεται τη ρύθμιση της ιστοκάμερας, τον εντοπισμό του χεριού, την εξαγωγή σημείων, τις αλληλεπιδράσεις του χρήστη, την οπτικοποίηση και καταγραφή δεδομένων, προσφέροντας μια αλληλεπιδραστική και ενημερωτική εμπειρία για το χρήστη.

## 4.2 Αρχείο `keypoint_classification.py`

### 4.2.1 Εκπαίδευση και Αξιολόγηση Μοντέλου Μηχανικής Μάθησης

Η παρακάτω ανάλυση του αρχείου `keypoint_classification.py`, γραμμένο στη γλώσσα προγραμματισμού Python, αναφέρεται στη εκπαίδευση και την αξιολόγηση του μοντέλου μηχανικής μάθησης που αναπτύχθηκε στα πλαίσια της εφαρμογής αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.) σε πραγματικό χρόνο. Το μοντέλο μετατρέπεται και εξάγεται σε μορφή TensorFlow Lite.

● **Φόρτωση και διαχωρισμός του Συνόλου Δεδομένων:**

- Αρχικά, διαβάζεται και φορτώνεται το επεξεργασμένο Σύνολο Δεδομένων, που έχει συλλεχθεί και κατασκευαστεί κατά την κατάσταση λειτουργίας «1», κατάσταση συλλογής δεδομένων, της εφαρμογής. Το Σύνολο Δεδομένων, βρίσκεται μέσα σε αρχείο της μορφής

.CSV. Μέσω της βιβλιοθήκης NumPy, καθορίζονται οι στήλες για τα χαρακτηριστικά εισόδου (`X_dataset`) και τις ετικέτες (`y_dataset`). Έπειτα το σύνολο δεδομένων διαιρείται σε σύνολα εκπαίδευσης και δοκιμής χρησιμοποιώντας το `train_test_split`.

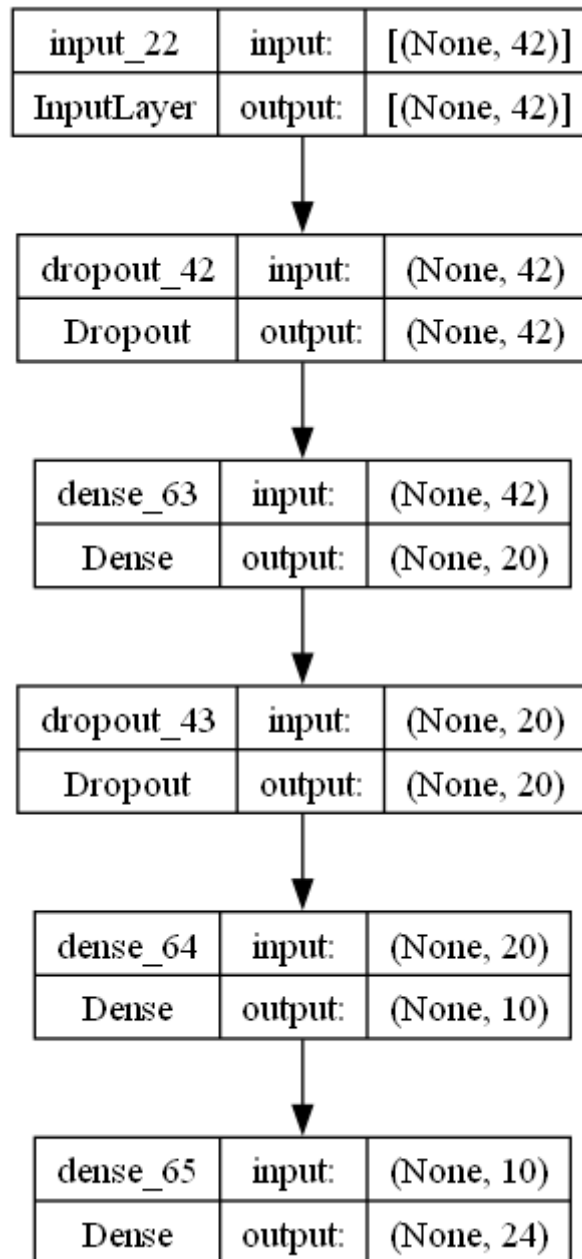
- **Κατασκευή Μοντέλου:**

- Κατασκευάζεται ένα νευρωνικό δίκτυο χρησιμοποιώντας το API Keras του TensorFlow. Η αρχιτεκτονική του μοντέλου αποτελείται από ένα επίπεδο εισόδου, επίπεδα απόρριψης (dropout layers), πυκνά κρυφά επίπεδα με συναρτήσεις ενεργοποίησης ReLU και ένα επίπεδο εξόδου με συνάρτηση ενεργοποίησης Softmax. Εμφανίζεται μία περιγραφή του μοντέλου και οπτικοποιείται η αρχιτεκτονική του.

- **Αρχιτεκτονική Μοντέλου:**

- Το περιγραφόμενο μοντέλο είναι ένα νευρωνικό δίκτυο πρόσω τροφοδότησης (Feedforward Neural Network), που χρησιμοποιεί πλήρως συνδεδεμένα επίπεδα. Αποτελείται από μία ακολουθία επιπέδων, όπου η έξοδος του κάθε επιπέδου λειτουργεί και ως είσοδος για το επόμενο επίπεδο. Το νευρωνικό δίκτυο αυτό είναι γνωστό και ως πυκνό νευρωνικό δίκτυο ή πολυστρωμική αντίληψη (Multilayer Perception). Το μοντέλο δέχεται ένα μονοδιάστατο πίνακα μήκους σαρανταδύο, σαν είσοδο, ενώ η έξοδος είναι μία κατανομή των πιθανοτήτων των κλάσεων εξόδου (κλάσεις που έχουν επιλεχτεί για την κατηγοριοποίηση των χειρονομιών). Η αρχιτεκτονική του μοντέλου είναι σειριακή, και ο πίνακας εισόδου του περιέχει προ-επεξεργασμένα δεδομένα.





Εικόνα 4-1: Οπτικοποίηση Αρχιτεκτονικής του Ανεπτυγμένου

➤ **Τα έξι Επίπεδα του Μοντέλου:**

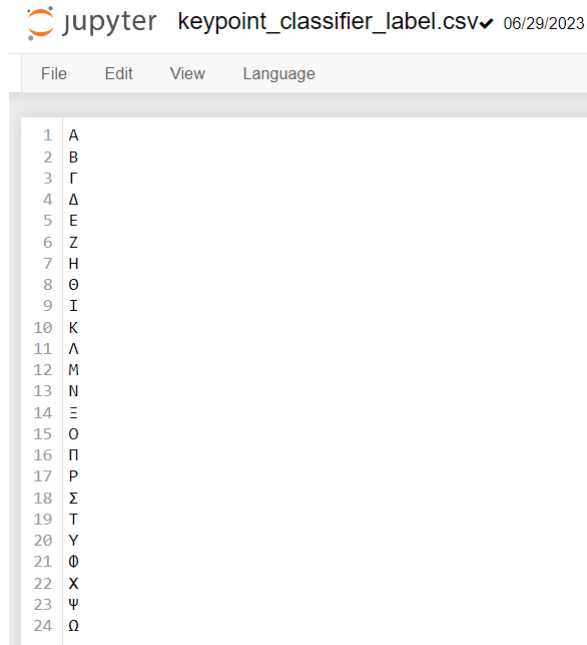
- Το πρώτο επίπεδο, το επίπεδο εισόδου, λαμβάνει δεδομένα της μορφής (21 \* 2), αντιπροσωπεύοντας τα εικοσιένα σημεία αναφοράς (landmarks) του χεριού, με συντεταγμένες (x, y).

- Το δεύτερο επίπεδο, είναι ένα επίπεδο πτώσης (Dropout layer) με ποσοστό 0.2, το οποίο ορίζει τυχαία το 20% των μονάδων εισόδου σε «0» κατά την εκπαίδευση για να αποφευχθεί το πρόβλημα της υπερπροσαρμογής (overfitting).
  - Το τρίτο επίπεδο, ένα πυκνό επίπεδο (Dense layer) με είκοσι μονάδες και μία συνάρτηση ενεργοποίησης με γραμμική και μη έξοδο (Swiss ή GeLU) που εφαρμόζεται στην έξοδο του προηγούμενου επιπέδου.
  - Το τέταρτο επίπεδο είναι ένα ακόμα επίπεδο πτώσης (Dropout layer) με ποσοστό 0.4, το οποίο ορίζει τυχαία, κατά τη διάρκεια κάθε παρτίδας εκπαίδευσης, το 40% των μονάδων εισόδου σε «0».
  - Το πέμπτο, όπως και το τρίτο επίπεδο, είναι ένα πυκνό επίπεδο (Dense layer) με δέκα μονάδες και μία συνάρτηση ενεργοποίησης με γραμμική και μη έξοδο (Swiss ή GeLU) που εφαρμόζεται στην έξοδο του προηγούμενου επιπέδου.
  - Το έκτο και τελευταίο επίπεδο, το επίπεδο εξόδου, είναι ένα πυκνό επίπεδο με N μονάδες και μία συνάρτηση ενεργοποίησης Softmax. Η συνάρτηση αυτή ενεργοποίησης στο επίπεδο εξόδου είναι ιδιαίτερα κατάλληλη για λύσεις σε προβλήματα πολυκατηγορικής ταξινόμησης. Μετατρέπει την έξοδο του προηγούμενου επιπέδου, δηλαδή ένα διάνυσμα αυθαίρετων τιμών, σε μία κατανομή πιθανοτήτων των κλάσεων ταξινόμησης. Η συνάρτηση Softmax, κανονικοποιεί τις τιμές εξόδου, με τέτοιο τρόπο ώστε όταν προστίθενται το αποτέλεσμα να είναι η μονάδα «1», εξασφαλίζοντας πως οι τιμές αυτές γίνεται να ερμηνευτούν ως πιθανότητες. Η υψηλότερη τιμή πιθανότητας αντιστοιχεί στην προβλεπόμενη κλάση για μία δεδομένη είσοδο ακολουθίας. Συνεπώς, η συνάρτηση ενεργοποίησης Softmax, είναι υπεύθυνη για την παραγωγή της κατανομής των πιθανοτήτων πάνω στις κλάσεις εισόδου και να κάνει την τελική πρόβλεψη.
- **Τα τέσσερα στάδια Προ-επεξεργασίας των Δεδομένων:** Τα στάδια που θα αναλυθούν περαιτέρω παρακάτω, έχουν ήδη συζητηθεί σε ένα βαθμό στο αρχείο «app.py», κατά τη διάρκεια συλλογής των δεδομένων, δηλαδή στην κατάσταση λειτουργίας «1». Η αρχιτεκτονική του μοντέλου σχεδιάστηκε με γνώμονα ένα πρόβλημα ταξινόμησης N κλάσεων εξόδου. Σε αυτό το μοντέλο, χρησιμοποιείται το API της MediaPipe, για τον εντοπισμό της παλάμης, και την εξαγωγή των συντεταγμένων των σημείων αναφοράς, που

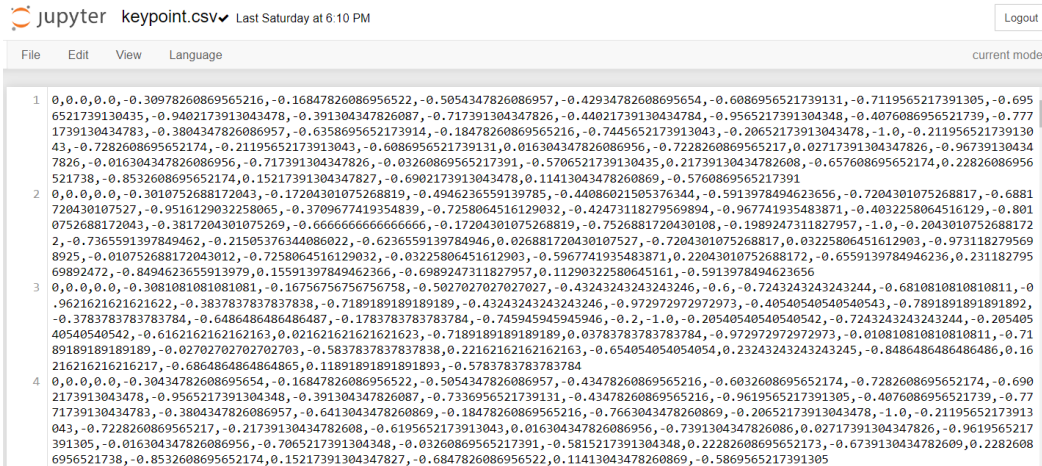
αντιστοιχούν σε σημεία - κλειδιά των χεριών, από μία ακολουθία βίντεο πραγματικού χρόνου. Στα σημεία – κλειδιά των χεριών περιλαμβάνονται τα άκρα των δαχτύλων, οι αρθρώσεις και ο καρπός. Εν συνεχεία, οι ακατέργαστες συντεταγμένες – κλειδιά διανύουν μία διαδικασία επεξεργασίας τεσσάρων σταδίων, για τη μετατροπή τους στην κατάλληλη μορφή για την εκπαίδευση του μοντέλου.

- Στο πρώτο στάδιο προ-επεξεργασίας, καταγράφονται και αποθηκεύονται οι συντεταγμένες των σημείων-κλειδιών (landmark), κάθε καρέ, σε έναν δυσδιάστατο πίνακα της μορφής (21,2). Αυτό συμβαίνει όταν ο χρήστης βρίσκεται στην κατάσταση λειτουργίας «1» στο αρχείο «app.py», τη λειτουργία συλλογής δεδομένων. Η κάθε γραμμή του πίνακα αντιπροσωπεύει τις συντεταγμένες (x, y) ενός μεμονωμένου σημείου αναφοράς (landmark). Αυτό το βήμα εξυπηρετεί στην αποτύπωση των χωρικών πληροφοριών των κινήσεων του χεριού και στην αναπαράσταση του δομημένου τρόπου τοποθέτησης του χεριού, ώστε να είναι εφικτή η χρήση του σαν είσοδος του μοντέλου.
- Στο δεύτερο στάδιο προ-επεξεργασίας των δεδομένων, οι συντεταγμένες κανονικοποιούνται ως προς το αναφερόμενο σημείο αναφοράς – κλειδί του καρπού, που φέρει το δείκτη 0. Με την μετατροπή αυτή επιτυγχάνεται η μείωση του αντίκτυπου της μετάφρασης του χεριού στο βίντεο, κανονικοποιείται η θέση του χεριού και το μοντέλο γίνεται πιο ανθεκτικό στην κίνηση του χεριού.
- Στο τρίτο στάδιο της προ-επεξεργασίας, οι σχετικές συντεταγμένες μετατρέπονται από δυσδιάστατο πίνακα της μορφής (21,2), σε μονοδιάστατο πίνακα μήκους  $21*2$ , με κάθε στοιχείο του πίνακα να αντιπροσωπεύει μία μόνο τιμή συντεταγμένης. Έτσι δημιουργείται μία σταθερού μήκους αναπαράσταση της εισόδου του μοντέλου. Με την μετατροπή αυτή, το μοντέλο μπορεί να χειρίζεται κάθε τιμή συντεταγμένης ως ξεχωριστό χαρακτηριστικό εισόδου, διευκολύνοντας την εκμάθηση πολύπλοκων σχέσεων μεταξύ της εισόδου και της εξόδου.
- Στο τέταρτο και τελευταίο στάδιο της προ-επεξεργασίας των δεδομένων των συντεταγμένων, οι τιμές στον μονοδιάστατο πλέον πίνακα κανονικοποιούνται στη μέγιστη απόλυτη τιμή. Η κανονικοποίηση των δεδομένων, είναι πολύ πιθανό να συνεισφέρει στην βελτίωση της απόδοσης του μοντέλου, μειώνοντας το αντίκτυπο της κλιμάκωσης στα δεδομένα εισόδου. Πιο συγκεκριμένα, κλιμακώνοντας τις τιμές, έτσι ώστε να κινούνται στο εύρος από «-1» έως «1»,  $([-1, 1].)$ , το

επιλεγμένο μοντέλο μπορεί να μάθει πιο αποτελεσματικά και να αποφύγει πιθανά προβλήματα όπως η εξαφάνιση των κλίσεων (gradients).



Εικόνα 4-2: Εικόνα από το αρχείο ετικετών (label) keypoint\_classifier\_label.csv



Εικόνα 4-3: Εικόνα από το αρχείο keypoint.csv, το Σύνολο Δεδομένων. Σε αυτό το αρχείο αναγράφονται όλα τα δεδομένα των σημείων αναφοράς της παλάμης της κάθε φωτογραφίας.

Συνολικά, τα βήματα της προ-επεξεργασίας των δεδομένων, εξυπηρετούν στη μετατροπή των ακατέργαστων συντεταγμένων των σημείων αναφοράς - κλειδιών (landmarks), σε μορφή κατάλληλη για να δημιουργηθεί το σύνολο δεδομένων, πάνω στο οποίο θα εκπαιδευτεί το μοντέλο. Η αναπαράσταση εισόδου που προκύπτει, αποτυπώνει τα βασικά χαρακτηριστικά των κινήσεων του χεριού και παρέχει την αρμόζουσα είσοδο για την εκπαίδευση του μοντέλου.

- **Εκπαίδευση Μοντέλου:**

- Το μοντέλο φέρει ως βελτιστοποιητή τον «adam», το «sparse\_categorical\_crossentropy» ως μέτρο απώλειας και το μετρητή ακριβείας. Η εκπαίδευση διεξάγεται με τα δεδομένα εκπαίδευσης ( $X_{train}$ ,  $y_{train}$ ). Στην όλη διαδικασία συμπεριλαμβάνονται επανακλήσεις (callbacks) για πρόωρη διακοπή της διαδικασίας εκπαίδευσης και αποθήκευση των παραμέτρων του μοντέλου. Η πρόωρη διακοπή της εκπαίδευσης συμβαίνει για την αποφυγή της υπερ-προσαρμογής (overfitting) του μοντέλου, δηλαδή εάν εντοπιστεί πως οι μετρήσεις επικύρωσης (validation metric) του μοντέλου σταματάνε να βελτιώνονται ή αρχίζουν να παρακαμάζουν, τότε σταματάει η εκπαίδευση του μοντέλου, ώστε αυτό να μην εκπαιδευτεί πάνω σε θόρυβο που βρίσκεται στο σύνολο δεδομένων. Ο μηχανισμός επανακλήσεων (callbacks), είναι και ένας μηχανισμός αποθήκευσης των παραμέτρων του μοντέλου, σε συγκεκριμένα σημεία κατά τη διάρκεια της εκπαίδευσης. Ο μηχανισμός αυτός βοηθάει στην αποτύπωση του καλύτερου μοντέλου που επιτεύχθηκε κατά τη διάρκεια της εκπαίδευσης. Εάν η απόδοση του μοντέλου βελτιωθεί, το κατάλληλο σημείο ελέγχου επανάκλησης ενημερώνει το αποθηκευμένο μοντέλο. Στο τέλος της εκπαίδευσης, είναι εφικτό να φορτωθεί το μοντέλο με την καλύτερη απόδοση, και όχι απαραίτητα το τελικό μοντέλο. Το μοντέλο εκπαιδεύεται συνολικά για χίλιες εποχές (epochs), δηλαδή το μοντέλο εκπαιδεύεται επαναληπτικά στο σύνολο δεδομένων χίλιες φορές. Σε κάθε εποχή (epoch), τροφοδοτούνται τα δεδομένα εκπαίδευσης στο μοντέλο, υπολογίζεται η απώλεια και ενημερώνονται οι παράμετροι του μοντέλου, για την ελαχιστοποίηση της απώλειας αυτής. Κατά τη διάρκεια της εκπαίδευσης, το σύνολο δεδομένων χωρίζεται σε μικρότερα υποσύνολα που ονομάζονται «παρτίδες». Κάθε παρτίδα περιλαμβάνει ένα σταθερό αριθμό δειγμάτων δεδομένων. Το μέγεθος της παρτίδας σε αυτό το αρχείο είναι εκατονεικοσιοχτώ, που σημαίνει πως κατά τη διάρκεια κάθε εποχής, το μοντέλο επεξεργάζεται εκατονεικοσιοχτώ δείγματα πριν ενημερώσει τις παραμέτρους του. Αυτός

ο τρόπος εκπαίδευσης συμβάλλει στη βελτίωση της αποτελεσματικότητας της εκπαίδευσης και στην αξιοποίηση των παράλληλων συστημάτων του σύγχρονου υλικού.

- **Αξιολόγηση Μοντέλου:**

- Η αξιολόγηση του μοντέλου λαμβάνει χώρα στα δεδομένα δοκιμής ( $X_{test}$ ,  $y_{test}$ ). Υπολογίζεται και εκτυπώνεται η απώλεια επικύρωσης και η ακρίβεια. Το αποθηκευμένο μοντέλο φορτώνεται για την πρόβλεψη.

- **Δοκιμή Πρόβλεψης:**

- Γίνεται μία δοκιμαστική πρόβλεψη στο φορτωμένο μοντέλο χρησιμοποιώντας ένα μέρος του Συνόλου Δεδομένων, δηλαδή χρησιμοποιώντας το σύνολο δοκιμής. Εκτυπώνονται οι προβλεπόμενες πιθανότητες της τάξης-κατηγορίας (χειρονομίας) και η προβλεπόμενη ετικέτα της.

- **Δημιουργία και Εκτύπωση του Πίνακα Σύγχυσης:**

- Στον υπολογισμό του πίνακα σύγχυσης χρησιμοποιούνται οι πραγματικές ετικέτες « $y_{true}$ » και οι προβλεπόμενες ετικέτες « $y_{pred}$ ». Αφού δημιουργηθεί ο πίνακας, εμφανίζεται ως θερμογραφικός χάρτης, μέσω τα βιβλιοθήκης «Seaborn», και εκτυπώνεται μία αναφορά ταξινόμησης (classification report). Τέλος, παρέχεται μία οπτική αναπαράσταση της απόδοσης του μοντέλου.

- **Μετατροπή του Μοντέλου σε Μορφή TensorFlow Lite:**

- Μετά τη δημιουργία του πίνακα σύγχυσης, το εκπαιδευμένο μοντέλο μηχανικής μάθησης μετατρέπεται σε μορφή TensorFlow Lite. Η νέα αυτή μορφή του μοντέλου αποθηκεύεται σε ένα συγκεκριμένο αρχείο, που καθορίζεται από το «μονοπάτι» «`tf_lite_save_path`».

- **Δοκιμή Πρόβλεψης με το Μοντέλο μορφής TensorFlow Lite:**

- Διεξάγεται μία δοκιμή πρόβλεψης με το μοντέλο μορφής TensorFlow Lite. Αποκτώνται τα δεδομένα εισόδου και εξόδου. Καθορίζεται ένα δείγμα δεδομένων εισόδου για την πρόβλεψη. Υπολογίζεται ο χρόνος πρόβλεψης και εκτυπώνονται τα αποτελέσματα, συμπεριλαμβανομένης της προβλεπόμενης ετικέτας τάξης (χειρονομίας).

- **Εξωτερικές Βιβλιοθήκες:**

- Το αρχείο εξαρτάται από εξωτερικές βιβλιοθήκες, όπως η NumPy, η TensorFlow, η scikit-learn, η pandas, η seaborn και η matplotlib. Η TensorFlow, χρησιμοποιείται για την εκπαίδευση και την αξιολόγηση του μοντέλου, η seaborn για την οπτικοποίηση του πίνακα σύγκυσης και το TensorFlow Lite για την μετατροπή του μοντέλου και τον έλεγχο της απόδοσης του σε πραγματικό χρόνο.

Σε αυτό το αρχείο γίνεται μία ολοκληρωμένη ροή εργασίας μηχανικής μάθησης, από τη φόρτωση του συνόλου δεδομένων και την εκπαίδευση του μοντέλου έως την αξιολόγηση του χρησιμοποιώντας έναν πίνακα σύγκυσης. Το μοντέλο αυτό έπειτα μετατρέπεται σε μορφή TensorFlow Lite. Ο πίνακας σύγκυσης εξυπηρετεί στην αξιολόγηση της απόδοσης της ταξινόμησης του ανεπτυγμένου μοντέλου, ενώ η μορφή TensorFlow Lite επιτρέπει την αποδοτική χρήση σε συσκευές ακμής, γνωστές ως «edge devices», για πρόβλεψη σε πραγματικό χρόνο. Οι «συσκευές ακμής» ή «συσκευές άκρων», είναι οποιοδήποτε κομμάτι υλικού που ελέγχει τη ροή δεδομένων στο όριο μεταξύ δύο δικτύων.

## 4.3 Αρχείο KeyPointClassifier.py

### 4.3.1 Ανάλυση και Περιγραφή αρχείου

Σε αυτό το αρχείο δημιουργείται μια κλάση της γλώσσας προγραμματισμού Python, με όνομα KeyPointClassifier.py, η οποία χρησιμοποιείται για την ταξινόμηση δεδομένων σημείων αναφοράς αξιοποιώντας το προσαρμοσμένο μοντέλο TensorFlow Lite, εκπαιδευμένο ειδικά για την αναγνώριση της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.). Η κλάση έχει σχεδιαστεί για να λειτουργεί με προ-εκπαιδευμένα μοντέλα TensorFlow Lite, και περιλαμβάνει διεπαφή για τη δημιουργία προβλέψεων.

- **Κλάση KeyPointClassifier :**

- Η κλάση ενθυλακώνει τη λειτουργικότητα που σχετίζεται με την ταξινόμηση των σημείων αναφοράς χρησιμοποιώντας ένα μοντέλο TensorFlow Lite. Μπορεί να αρχικοποιηθεί με προαιρετικές παραμέτρους: model\_path (καθορίζει το «μονοπάτι» προς το αρχείο μοντέλου TensorFlow Lite) και num\_threads (καθορίζει τον αριθμό των νημάτων). Ο κατασκευαστής (constructor) της κλάσης αρχικοποιεί τον διερμηνευτή TensorFlow Lite και εκχωρεί δέσμες για την είσοδο και την έξοδο.

- **Μέθοδος call(self, landmark\_list):**

- Αυτή η μέθοδος, επιτρέπει στα αντικείμενα της κλάσης KeyPointClassifier να καλούνται, σαν να ήταν συναρτήσεις. Τα δεδομένα εισέρχονται στην κλάση KeyPointClassifier μέσω της παραμέτρου «landmark\_list», μία λίστα που περιέχει δεδομένα σημείων αναφοράς προς ταξινόμηση. Η μέθοδος εκτελεί τα ακόλουθα βήματα : Ορίζει ως είσοδο του μοντέλου TensorFlow Lite τη λίστα «landmark\_list». Καλεί το διεργμηνέα TensorFlow Lite για να κάνει μία πρόβλεψη. Ανακτά το αποτέλεσμα της πρόβλεψης και το βαθμό εμπιστοσύνης. Επιστρέφει τον προβλεπόμενο τίτλο – ετικέτα της κλάσης και τον βαθμό εμπιστοσύνης, συνδυαστικά ως μία πλειάδα (tuple).

- **Εξωτερικές Διεπαφές:**

- Η κλάση αλληλοεπιδρά με το αρχείο μοντέλου TensorFlow Lite, που ορίζεται από την παράμετρο model\_path. Μέσω του διεργμηνευτεί TensorFlow Lite, γίνεται συμβατό με ποικίλες πλατφόρμες και συσκευές.

- **Έξοδος και Αποτελέσματα:**

- Η κύρια έξοδος της κλάσης KeyPointClassifier είναι ο προβλεπόμενος δείκτης της κλάσης και ο βαθμός εμπιστοσύνης. Οι χρήστες μπορούν να αναλύσουν το βαθμό εμπιστοσύνης για να καθορίσουν την αξιοπιστία του αποτελέσματος ταξινόμησης.

Συγκεφαλαιώνοντας, στο αρχείο αυτό δημιουργείται μία κλάση που ενθυλακώνει την ταξινόμηση σημείων αναφοράς, της παλάμης, χρησιμοποιώντας το προσαρμοσμένο μοντέλο TensorFlow Lite, για την αναγνώριση της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.). Παρέχει μια επαναχρησιμοποιήσιμη διεπαφή για τη δημιουργία προβλέψεων βασισμένων στην είσοδο των δεδομένων των σημείων ενδιαφέροντος. Η κλάση KeyPointClassifier, είναι ένα ευέλικτο στοιχείο που μπορεί εύκολα να ενσωματωθεί σε εφαρμογές που απαιτούν ταξινόμηση σημείων αναφοράς της παλάμης, δηλαδή να ενταχθεί σε εφαρμογές που απαιτούν αναγνώριση στατικών χειρονομιών.



## 4.4 Αρχείο cvfprscal.c.py

### 4.4.1 Ανάλυση και Περιγραφή του αρχείου

Σε αυτό το αρχείο δημιουργείται μία κλάση με όνομα «CvFpsCalc», υπεύθυνη για τον υπολογισμό των καρτέ ανά δευτερόλεπτο (Frames Per Second), ροής βίντεο ζωντανής μετάδοσης σε εφαρμογές Όρασης Υπολογιστών. Η κλάση σχεδιάστηκε για να μετρά την απόδοση της επεξεργασίας βίντεο, υπολογίζοντας τον μέσο όρο των καρτέ ανά δευτερόλεπτο.

- **Κλάση CvFpsCalc :**

- Η κλάση κατασκευάστηκε για να ενσωματώσει τη λειτουργία που σχετίζεται με τον υπολογισμό των καρτέ ανά δευτερόλεπτο στο κύριο αρχείο. Αρχικοποιείται με μία αρχική παράμετρο, τη «buffer\_len», η οποία καθορίζει το μήκος του προσωρινού ρυθμιστή (buffer) χρόνου καρτέ, που χρησιμοποιείται για τον υπολογισμό του μέσου όρου των καρτέ ανά δευτερόλεπτο. Η προεπιλεγμένη του τιμή είναι «1». Ο κατασκευαστής της κλάσης αρχικοποιεί εσωτερικές μεταβλητές, όπως χρονικές σημάνσεις έναρξης, συχνότητας και μίας ουράς με δύο άκρα (deque), για την αποθήκευση των διαφορών χρόνου καρτέ.

- **Μέθοδος get(self):**

- Η μέθοδος υπολογίζει και ανακτά τον τρέχοντα αριθμό καρτέ ανά δευτερόλεπτο (Frames Per Second). Αρχικά, εξάγει την τρέχουσα σήμανση χρόνου χρησιμοποιώντας τη συνάρτηση cv.getTickCount() της OpenCv. Υπολογίζει τη διαφορά χρόνου μεταξύ της τρέχουσας σήμανσης και της αρχικής σήμανσης. Το αποτέλεσμα της διαφοράς μετατρέπεται σε χιλιοστά του δευτερολέπτου αξιοποιώντας τη συχνότητα των σημάνσεων. Ενημερώνει την αρχική σήμανση στην τρέχουσα σήμανση. Προσθέτει την υπολογισμένη διαφορά χρόνου στην ουρά (deque), διατηρώντας ένα προκαθορισμένο μήκος του προσωρινού ρυθμιστή (buffer). Υπολογίζει τα καρτέ ανά δευτερόλεπτο ως το αντίστροφο της μέσης διαφοράς χρόνου ανάμεσα στα δεδομένα της ουράς (deque). Στρογγυλοποιεί το αποτέλεσμα των καρτέ ανά δευτερόλεπτο σε δύο δεκαδικά ψηφία και το επιστρέφει.

- **Εξωτερικές Διεπαφές:**

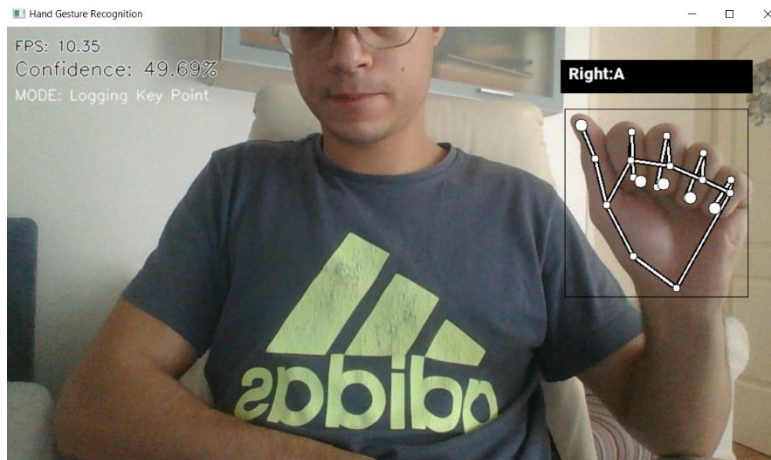
- Η κλάση χρησιμοποιεί τη βιβλιοθήκη OpenCV (cv2) για την πρόσβαση σε συναρτήσεις όπως η «cv.getTickCount()» και η «cv.getTickFrequency()» για τη μέτρηση του χρόνου.

- **Έξοδος και Αποτελέσματα:**

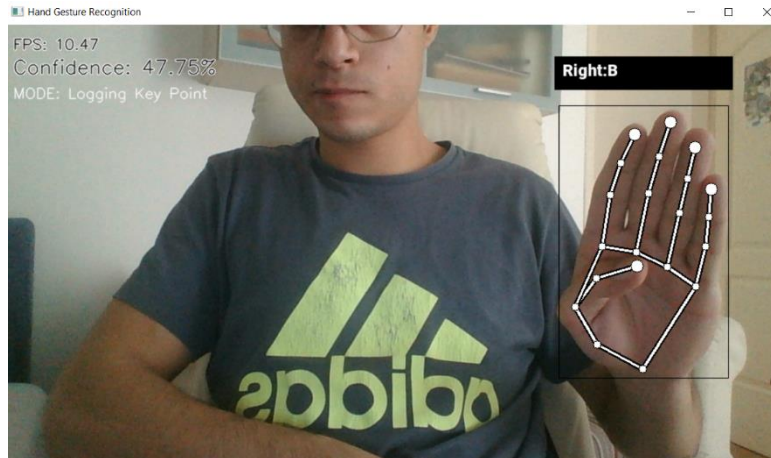
- Η κύρια έξοδος της κλάσης «CvFpsCalc» είναι η υπολογισμένη τιμή των καρτέ ανά δευτερόλεπτο (Frames Per Second). Ο υπολογισμός αυτός μπορεί να αξιοποιηθεί για την παρακολούθηση της απόδοσης και για βελτιστοποίηση σε εφαρμογές Όρασης Υπολογιστών.

Ανακεφαλαιώνοντας, αυτό το αρχείο περιλαμβάνει ένα επαναχρησιμοποιήσιμο και απλό μηχανισμό για τον υπολογισμό και την παρακολούθηση των καρτέ ανά δευτερόλεπτο σε εργασίες υπολογιστικής όρασης. Αποτελεί ένα χρήσιμο εργαλείο για προγραμματιστές που εργάζονται σε έργα, που ενσωματώνουν τεχνικές όρασης υπολογιστών, και χρειάζεται να παρακολουθούν και να βελτιστοποιήσουν την απόδοση των αλγορίθμων, υπεύθυνοι για την επεξεργασία βίντεο. Μέσω αυτού του υπολογισμού, παρέχεται ένας απλός τρόπος αξιολόγησης της αποτελεσματικότητας της ανάλυσης βίντεο σε πραγματικό χρόνο.

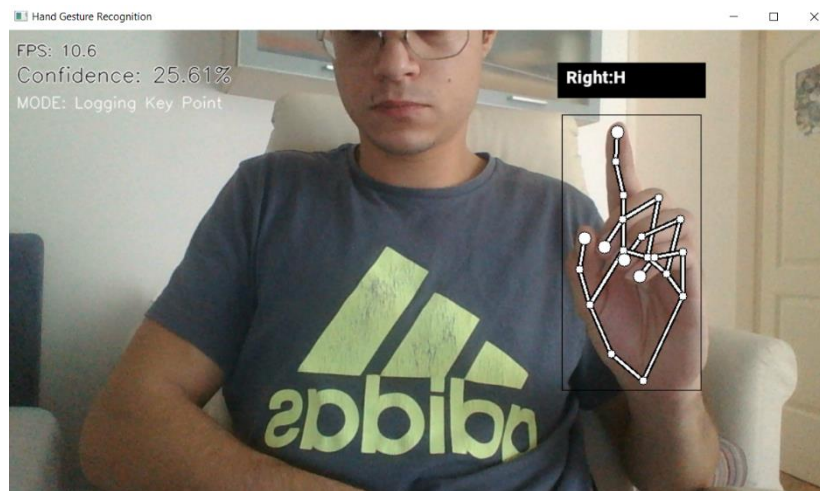
Παρακάτω φαίνονται κάποιες εικόνες από την υλοποιημένη εφαρμογή, που επιδεικνύουν τον τρόπο συλλογής δεδομένων από το δεξί και το αριστερό χέρι:



Εικόνα 4-4: Συλλογή Δεδομένων για το γράμμα "A"



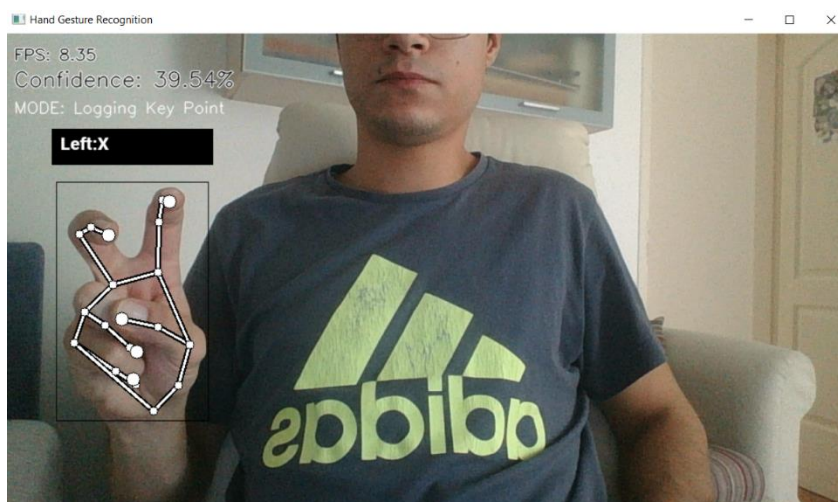
Εικόνα 4-5: Συλλογή Δεδομένων για το γράμμα "B"



Εικόνα 4-6: Συλλογή Δεδομένων για το γράμμα "Δ"



Εικόνα 4-7: Συλλογή Δεδομένων για το γράμμα "Π"

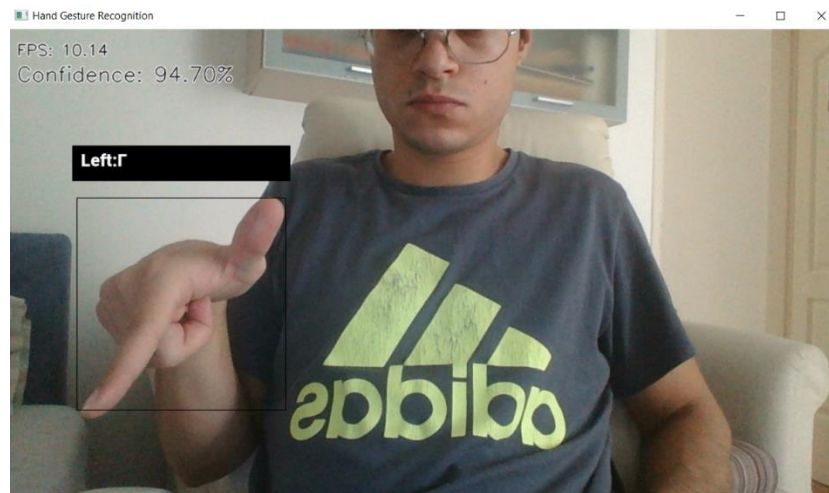


Εικόνα 4-8: Συλλογή Δεδομένων για το γράμμα "Χ"

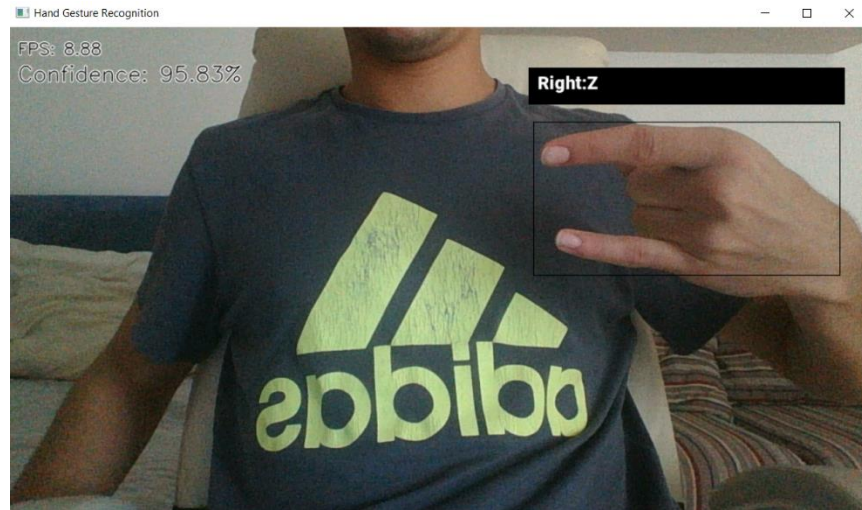
Σε αυτό το σημείο, που το μοντέλο δεν είναι εκπαιδευμένο καθόλου ή είναι μερικώς εκπαιδευμένο, πρέπει να σημειωθεί πως τα γράμματα που φαίνεται πως αναγνωρίζονται από το μοντέλο δεν ισχύουν. Για αυτό το λόγο, και σε κάποιες από τις εικόνες της εφαρμογής αναγράφονται λανθασμένα γράμματα, σε σχέση με τα γράμματα για τα οποία και θέλουμε να συλλέξουμε δεδομένα. Επίσης, το ποσοστό εμπιστοσύνης – ακριβείας του μοντέλου δεν πρέπει, σε αυτό το σημείο, να το λαμβάνουμε υπόψιν, καθώς τα παραπάνω είναι παραδείγματα συλλογής εικόνων για τη δημιουργία του Συνόλου Δεδομένων, και συνεπώς το μοντέλο δεν έχει εκπαιδευτεί ολοκλήρως. Τα χαρακτηριστικά που φαίνονται στα παραδείγματα των εικόνων της εφαρμογής και χρίζουν ενδιαφέρον είναι τα καρτέ ανά δευτερόλεπτο (Frames Per Second (FPS)), το πλαίσιο περιφερειακά του χεριού, που εμφανίζεται με τον εντοπισμό του και ο «σκελετός»

εφαρμοσμένος πάνω στην παλάμη και στα δάχτυλα, που έχουν εντοπιστεί, μέσω του προ-εκπαιδευμένου μοντέλου της MediaPipe.

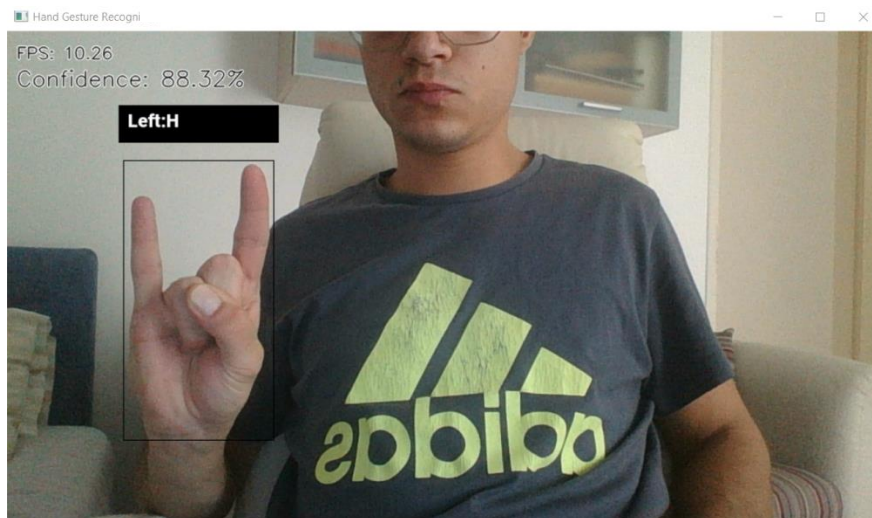
Στη συνέχεια, παρουσιάζονται παραδείγματα ως εικόνες, από την εφαρμογή που επιδεικνύουν τη λειτουργία αναγνώρισης των χειρονομιών, του δεξιού και του αριστερού χεριού, της Ελληνικής Νοηματικής Γλώσσας, που αντιστοιχούν σε γράμματα της ελληνικής αλφαβήτα, μετά την εκπαίδευση του μοντέλου που αναπτύχθηκε. Στις εικόνες αυτές αναγράφονται όλα τα στοιχεία που θα βλέπει ο χρήστης κατά τη διάρκεια χρήσης της εφαρμογής, τα οποία είναι τα καρτέ ανά δευτερόλεπτο (Frames Per Second (FPS)), ο βαθμός εμπιστοσύνης του μοντέλου (Confidence) και το γράμμα που αναγνωρίζεται σύμφωνα με τη χειρονομία του χρήστη («Αριστερό/Δεξί χέρι (Left/Right)» : «Αναγνωρισμένο Γράμμα»):



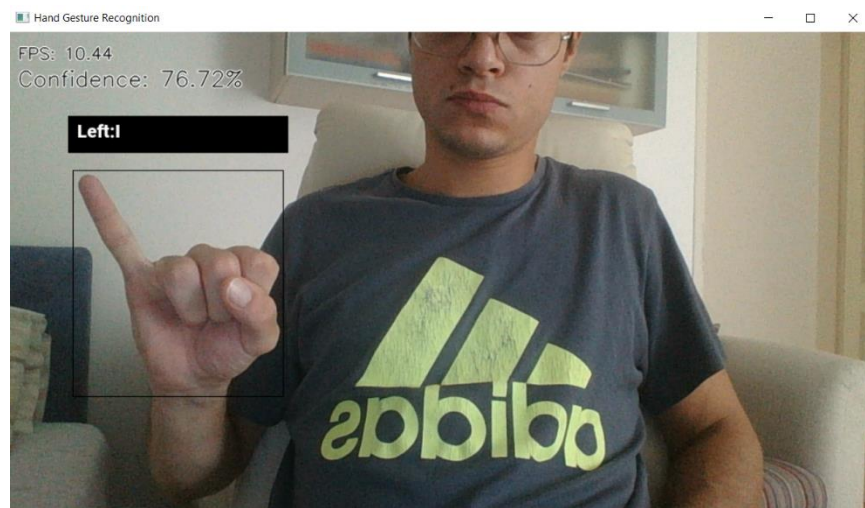
Εικόνα 4-9: Χειρονομία προς αναγνώριση του γράμματος "Γ"



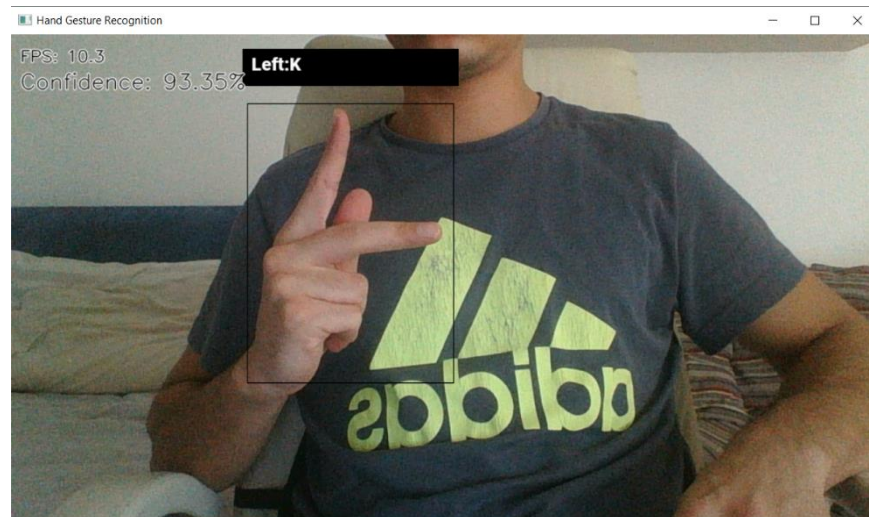
Εικόνα 4-10: Χειρονομία προς αναγνώριση του γράμματος "Ζ"



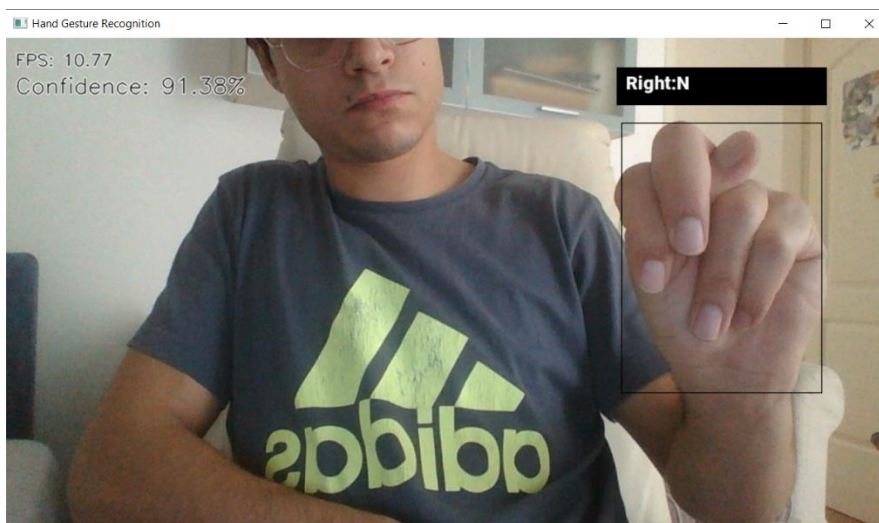
Εικόνα 4-11: Χειρονομία προς αναγνώριση του γράμματος "Η"



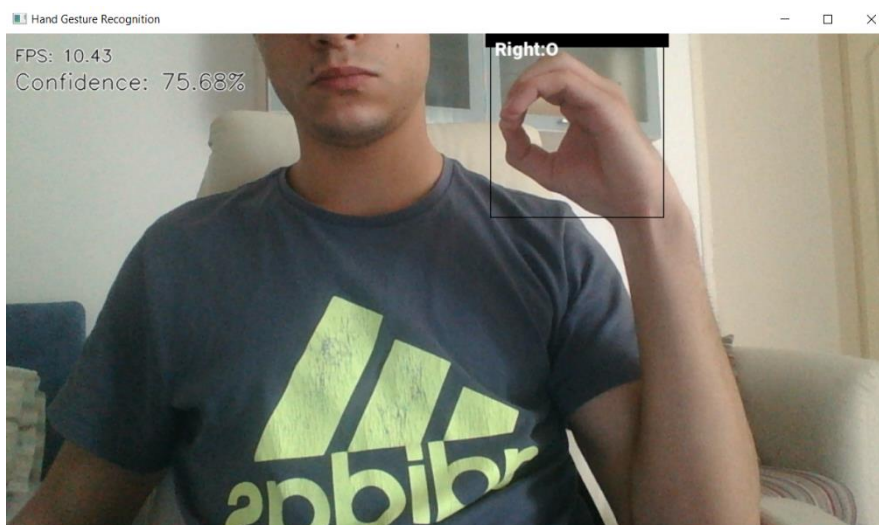
Εικόνα 4-12: Χειρονομία προς αναγνώριση του γράμματος "Ι"



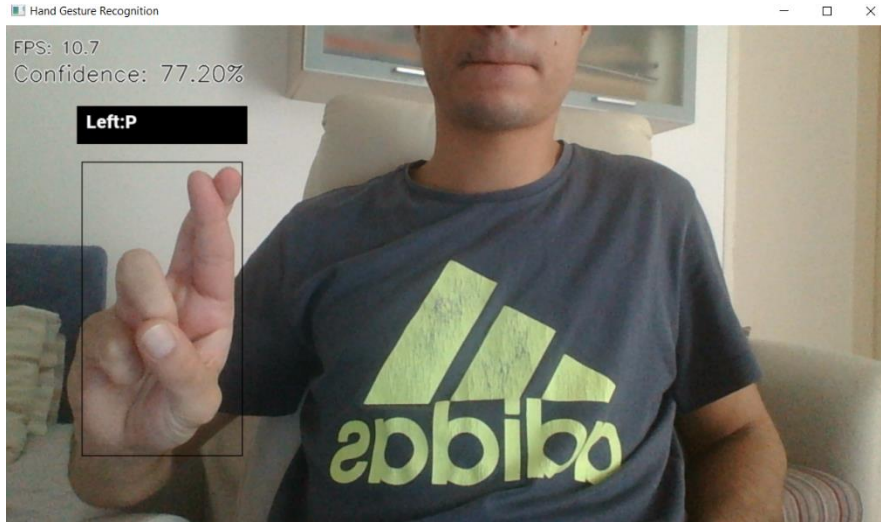
Εικόνα 4-13: Χειρονομία προς αναγνώριση του γράμματος "Κ"



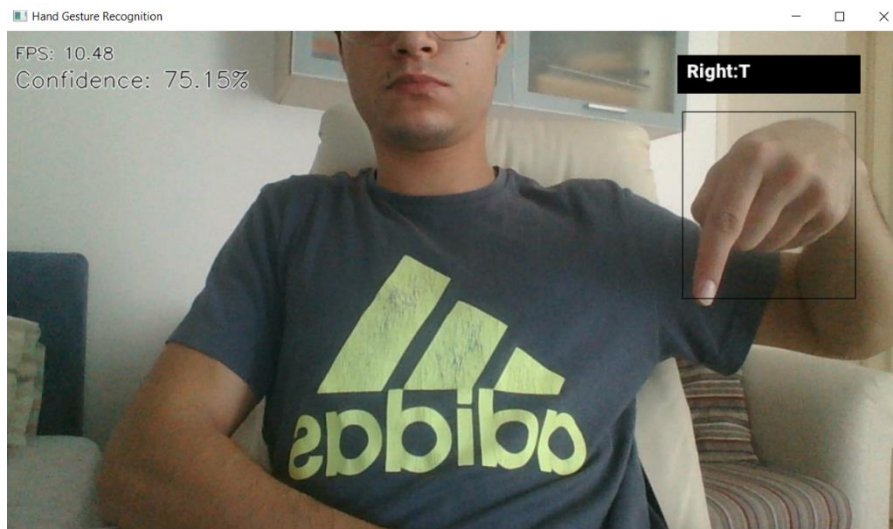
Εικόνα 4-14: Χειρονομία προς αναγνώριση του γράμματος "Ν"



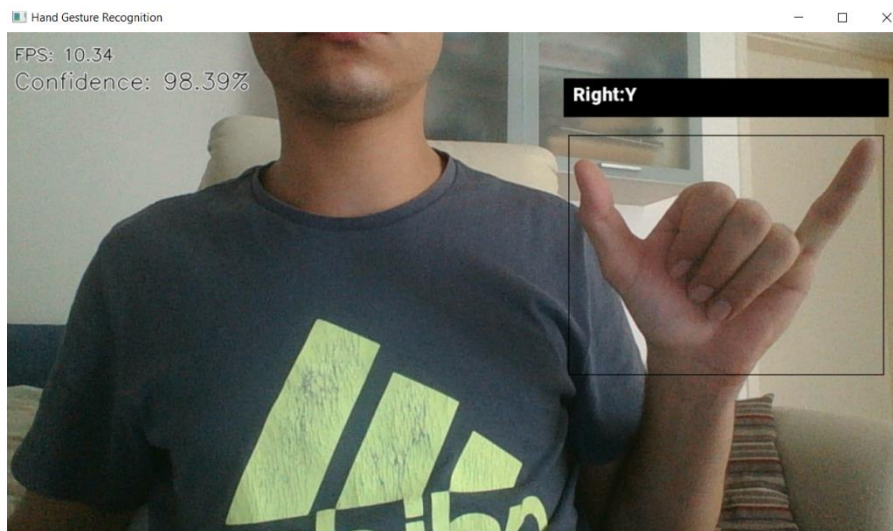
Εικόνα 4-15: Χειρονομία προς αναγνώριση του γράμματος "Ο"



*Εικόνα 4-16: Χειρονομία προς αναγνώριση του γράμματος "P"*



*Εικόνα 4-17: Χειρονομία προς αναγνώριση του γράμματος "T"*

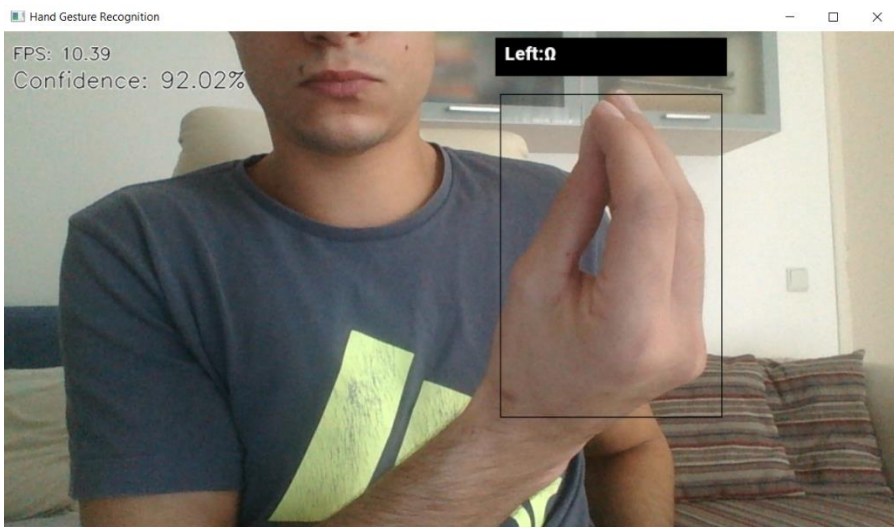


*Εικόνα 4-18: Χειρονομία προς αναγνώριση του γράμματος "Y"*





*Εικόνα 4-19: Χειρονομία προς αναγνώριση του γράμματος "Ψ"*



*Εικόνα 4-20: Χειρονομία προς αναγνώριση του γράμματος "Ω"*

Παρατηρώντας τις παραπάνω εικόνες, που αναρτήθηκαν ως παραδείγματα για την επίδειξη κάποιων κύριων λειτουργιών της εφαρμογής, συλλογή δεδομένων και αναγνώριση χειρονομιών με τη χρήση του εκπαιδευμένου πλέον μοντέλου, φαίνεται πως παρά τις ποικίλλεις και πολύπλοκες χειρονομίες, απαραίτητες για την αποκρυπτογράφηση και την αναγνώριση της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.) από ηλεκτρονικό υπολογιστή, τα προ-εκπαιδευμένα μοντέλα για τον εντοπισμό της παλάμης λειτουργούν καθωσπρέπει, όπως και τα μοντέλα βαθιάς μάθησης, που αναπτύχθηκαν και εκπαιδεύτηκαν πολλαπλές φορές, ώστε να αναγνωρίζουν τις προκαθορισμένες χειρονομίες της αλφαβήτα της Ε.Ν.Γ. με σχετικά αξιοσέβαστα ποσοστά (%) εμπιστοσύνης.

Στο επόμενο κεφάλαιο, θα αναλυθεί λεπτομερώς η διαδικασία επιλογής και αξιολόγησης των εκπαιδευμένων μοντέλων βαθιάς μάθησης, που αναπτύχθηκαν και εξετάστηκαν για το σκοπό της εφαρμογής.

## Κεφάλαιο 5<sup>ο</sup>

### ΑΞΙΟΛΟΓΗΣΗ ΠΙΝΑΚΑ ΣΥΓΧΥΣΗΣ (CONFUSION MATRIX)

Στον τομέα της μηχανικής μάθησης και των δραστηριοτήτων ανάπτυξης μοντέλων ταξινόμησης, η αξιολόγηση της απόδοσης του μοντέλου είναι κρίσιμη, και απαραίτητη για την ολοκλήρωση μία εφαρμογής που ενσωματώνει ένα τέτοιο μοντέλο. Στην προσπάθεια ανάπτυξης ενός μοντέλου μηχανικής μάθησης, εκπαιδευμένο στην αναγνώριση της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ), αξιοποιώντας τεχνικές ταξινόμησης, είναι αναγκαία η κατανόηση και η χρήση ενός πίνακα σύγχυσης (Confusion Matrix). Ο πίνακας σύγχυσης αποτελεί ένα ισχυρό εργαλείο που εξυπηρετεί στην αξιολόγηση της απόδοσης του μοντέλου που έχει επιλεγεί για να χρησιμοποιηθεί στην εφαρμογή που υλοποιείται. Ο πίνακας σύγχυσης, παρέχει μία λεπτομερή ανάλυση των ικανοτήτων πρόβλεψης του μοντέλου, ιδιαίτερα χρήσιμος σε περιπτώσεις ταξινόμησης πολλών κατηγοριών. Στο κεφάλαιο αυτό, λαμβάνει χώρα μία εκτενής διαδικασία εξερεύνησης του πίνακα σύγχυσης του μοντέλου αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.), αναλύοντας τα στοιχεία του και μελετώντας τις προκύπτουσες μετρήσεις [14].

#### 5.1 Ανατομία ενός Πίνακα Σύγχυσης

Ο πίνακας σύγχυσης ή αλλιώς γνωστός και ως πίνακας σφάλματος, είναι ένας δομημένος πίνακας που λειτουργεί ως βάση για την αξιολόγηση της απόδοσης ενός μοντέλου ταξινόμησης. Συγκρίνει, τις προβλέψεις του εκπαιδευμένου μοντέλου με τις πραγματικές ετικέτες των κατηγοριών, εντός ενός συνόλου δοκιμής. Στη συγκεκριμένη περίπτωση η κάθε κατηγορία ταυτίζεται με μία χειρονομία. Για κάθε γράμμα του ελληνικού αλφαβήτα υπάρχει μία ξεχωριστή χειρονομία, άρα η ανεπτυγμένη εφαρμογή στοχεύει στην κατηγοριοποίηση εικοσιτεσσάρων χειρονομιών. Σύμφωνα με τα παραπάνω, εδραιώνεται πως το πρόβλημα της αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.), που στοχεύει η εφαρμογή να επιλύσει, είναι ταξινόμηση πολυ-κατηγορικής φύσεως. Στην αξιολόγηση μοντέλων παρόμοιας φύσης, δηλαδή μοντέλα ανεπτυγμένα για την ταξινόμηση πολλών κατηγοριών, ο πίνακας σύγχυσης φέρει ανεκτίμητη αξία [14].

Ο πίνακας σύγχυσης αποτελείται από τέσσερα βασικά στοιχεία:

- **Αληθείς Θετικοί (ΑΘ) (True Positive)** : Οι Αληθείς Θετικοί υποδεικνύουν περιπτώσεις που το μοντέλο έχει ταξινομήσει σωστά, τα δείγματα – δεδομένα, στην κατάλληλη κατηγορία. Πιο

συγκεκριμένα, στο πλαίσιο της εφαρμογής που έχει αναπτυχθεί, εδώ περιλαμβάνονται οι περιπτώσεις όπου το μοντέλο αναγνωρίζει με επιτυχία μια χειρονομία σαν τη σωστή [14].

- **Αληθείς Αρνητικοί (ΑΑ) (True Negative)** : Οι Αληθείς Αρνητικοί αντιπροσωπεύουν περιπτώσεις που οι κατηγορίες έχουν ταξινομηθεί ορθά, ως μη – ανήκουσες σε κάποια κατηγορία. Στο ζήτημα της εφαρμογής που υλοποιήθηκε, και αναπτύχθηκε στα πλαίσια της διπλωματικής εργασίας, αυτές οι περιπτώσεις αφορούν τυχαίες χειρονομίες και μη – χειρονομίες, που ορθά κατηγοριοποιούνται πως δεν ανήκουν σε κάποια από τις τάξεις των χειρονομιών της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.) [14].
- **Ψευδείς Θετικοί (ΨΘ) (False Positive)** : Οι Ψευδείς Θετικοί περιλαμβάνουν περιπτώσεις που έχει γίνει λανθασμένη ταξινόμηση κατηγοριών. Ειδικότερα, στη συγκεκριμένη εφαρμογή κονσόλας, αυτό συμβαίνει όταν το μοντέλο αποτύχει στη σωστή αναγνώριση κάποιας χειρονομίας, δηλαδή όταν αναγνωριστεί η χειρονομία αλλά ενταχθεί σε λάθος κατηγορία [14].
- **Ψευδείς Αρνητικοί (ΨΑ) (False Negative)** : Οι Ψευδείς Αρνητικοί συμπεριλαμβάνουν περιπτώσεις που το μοντέλο έχει ταξινομήσει λανθασμένα τα δείγματα του συνόλου δοκιμής, ως μη – ανήκουσα σε μία συγκεκριμένη κατηγορία, ενώ αυτά ανήκουν. Για το μοντέλο μηχανικής μάθησης, αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ) που αναπτύχθηκε στα πλαίσια της διπλωματικής εργασίας, αυτό σημαίνει πως αποτυγχάνει να αναγνωρίσει σωστά μία χειρονομία [14].

Πέρα από τα θεμελιώδη στοιχεία, που αναφέρθηκαν παραπάνω, ο πίνακας σύγκρισης δημιουργεί μία σειρά από μετρικές που παρέχουν μια πιο βαθιά κατανόηση της απόδοσης του μοντέλου.

Αυτές οι μετρικές είναι:

- **Ακρίβεια (Precision)**: Η ακρίβεια ορίζεται ως τον λόγο των Αληθών Θετικών προς το συνολικό αριθμό των περιπτώσεων, που προβλέπονται ως θετικές από το μοντέλο. Αντιπροσωπεύει την ακρίβεια του μοντέλου στην ορθή θετική πρόβλεψη. [14].

$$\text{Τύπος: Ακρίβεια} = \frac{ΑΘ}{(ΑΘ + ΨΘ)}$$

*Εξίσωση 5-1: Συνάρτηση Ακριβείας*

- **Ανάκληση (Recall)**: Η ανάκληση, αντιπροσωπεύει τον λόγο των Αληθών Θετικών περιπτώσεων προς το συνολικό αριθμό των πραγματικών θετικών περιπτώσεων, εντός του συνόλου δοκιμής.

Διευκρινίζει αν το μοντέλο αιχμαλωτίζει όλες τις περιπτώσεις της θετικής κατηγορίας επαρκώς [14].

$$\text{Τύπος: Ανάκληση} = A\Theta / (A\Theta + \Psi\Theta)$$

*Εξίσωση 5-2: Συνάρτηση Ανάκλησης*

- **Αρμονικός Μέσος (F1-Score):** Η μετρική αυτή προκύπτει ως η αρμονική μέση της ακρίβειας και της ανάκλησης. Προσφέρει ένα ισορροπημένο μέτρο της απόδοσης του μοντέλου, ιδιαίτερα χρήσιμο για σύνολα δεδομένων διάφορων μεγεθών κατηγοριών ή ανισορροπιών [14].

$$\text{Τύπος: Αρμονικός Μέσος} = 2 * (\text{Ακρίβεια} * \text{Ανάκληση}) / (\text{Ακρίβεια} + \text{Ανάκληση})$$

*Εξίσωση 5-3: Συνάρτηση Αρμονικού Μέσου*

- **Ορθότητα (Accuracy):** Αυτή η μετρική αντιπροσωπεύει τη συνολική ορθότητα των προβλέψεων του μοντέλου. Ορίζεται από τον λόγο των ορθώς προβλεπόμενων περιπτώσεων προς το συνολικό αριθμό των περιπτώσεων στο σύνολο δοκιμής [14].

$$\text{Τύπος: Ορθότητα} = (A\Theta + AA) / (A\Theta + AA + \Psi\Theta + \Psi A)$$

*Εξίσωση 5-4: Συνάρτηση Ορθότητας*

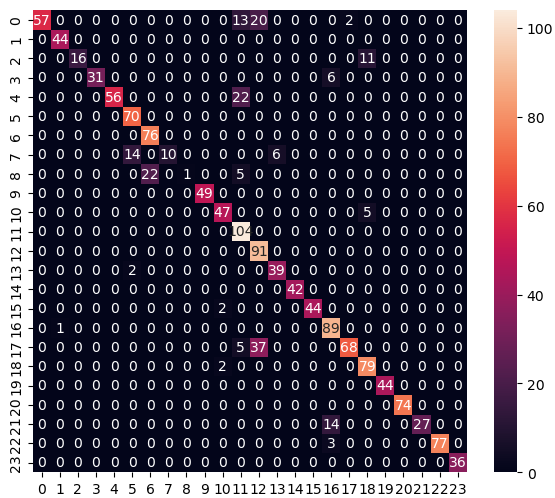
## 5.2 Ερμηνεία Αποτελεσμάτων Πίνακα Σύγκρισης

Στην περίπτωση της εφαρμογής αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.), που υλοποιήθηκε στα πλαίσια της διπλωματικής εργασίας, παρατηρείται μία τέλεια βαθμολογία για την Ακρίβεια, την Ανάκληση, τον Αρμονικό μέσο και την Ορθότητα σε όλες τις κατηγορίες. Υποδεικνύοντας ένα μοντέλο που επιδεικνύει εξαιρετική απόδοση στο σύνολο δοκιμής, αναγνωρίζοντας με επιτυχία όλες τις χειρονομίες, χωρίς κανένα ψευδώς θετικό ή ψευδώς αρνητικό αποτέλεσμα. Το ανεπτυγμένο μοντέλο της μηχανικής μάθησης, υποδηλώνει μία εντυπωσιακή ικανότητα να διακρίνει μεταξύ διάφορων χειρονομιών και μη – χειρονομιών, που σχετίζονται με την αλφαβήτα της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.).

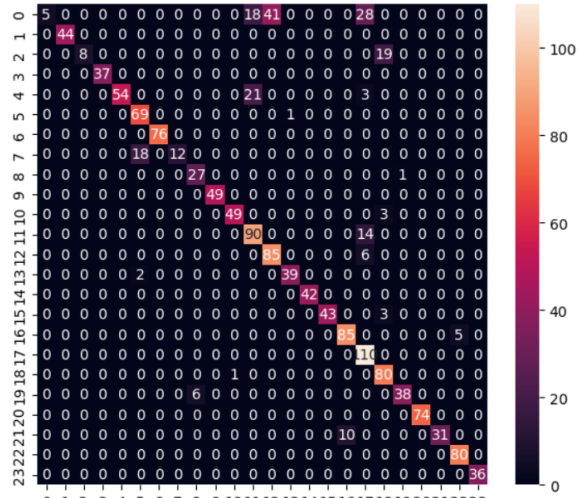
Παρακάτω φαίνονται οι απαραίτητες εικόνες για την σύγκριση των αποτελεσμάτων των ανεπτυγμένων μοντέλων, μοντέλα που φέρουν την ίδια αρχιτεκτονική, αλλά έχουν διαφορετική συνάρτηση ενεργοποίησης. Οι συναρτήσεις ενεργοποίησης που εξετάστηκαν είναι οι εξής: ReLU, eLU, GeLU, Leaky

ReLU, Tanh, Sigmoid και τέλος η Swiss. Παρακάτω παρέχονται οι πίνακες Σύγκρισης και οι έλεγχοι κατηγοριοποίησης, που δείχνουν τα αποτελέσματα των μετρικών όλων των μοντέλων, που έχουν ως συναρτήσεις ενεργοποίησης τις προαναφερόμενες:

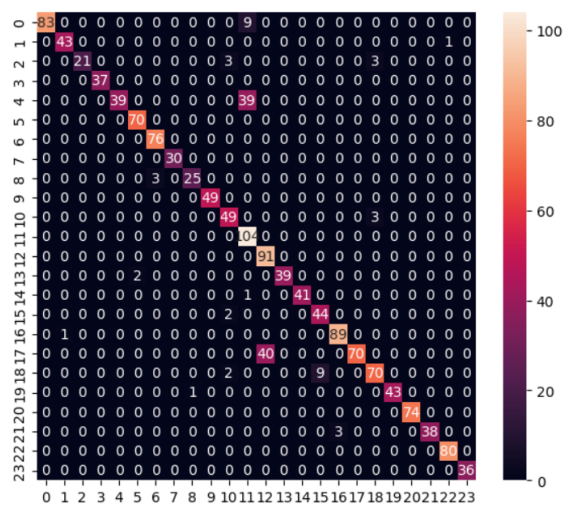
### Πίνακες Σύγκρισης:



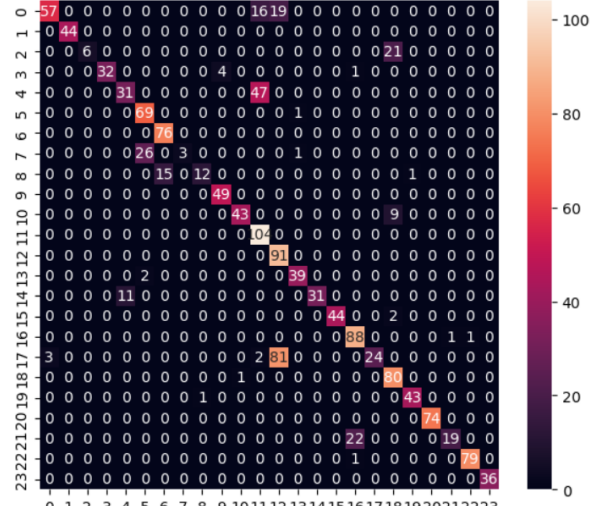
Εικόνα 5-1: Πίνακας Σύγκρισης ReLU

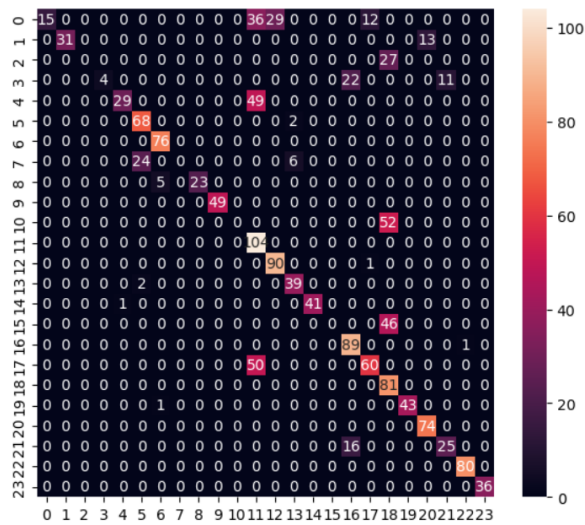


Εικόνα 5-2: Πίνακας Σύγκρισης eLU

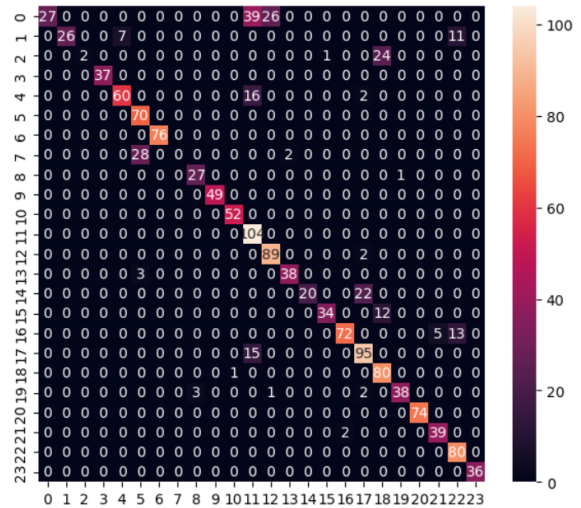


Εικόνα 5-3: Πίνακας Σύγκρισης GeLU

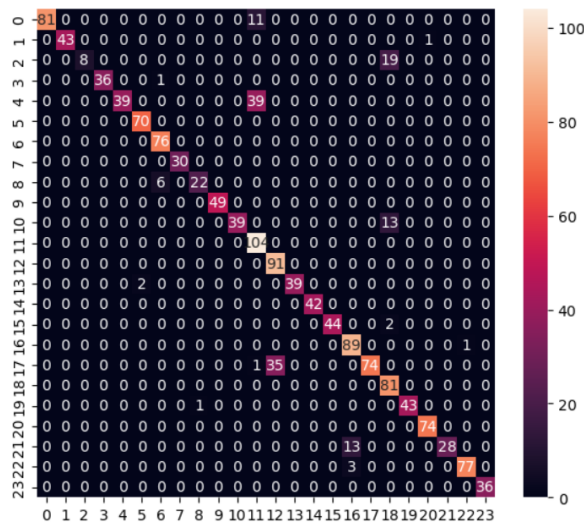




Εικόνα 5-5: Πίνακας Σύγκρισης Sigmoid



Εικόνα 5-6: Πίνακας Σύγκρισης Tanh



Εικόνα 5-7: Πίνακας Σύγκρισης Swiss

Οι παραπάνω πίνακες σύγκρισης έχουν οπτικοποιηθεί ως πίνακες θερμοκρασίας (heat maps). Τα χρώματα που χρησιμοποιούνται αντιπροσωπεύουν τις τιμές του εκάστοτε πίνακα σύγκρισης. Ένα βαθύ μπλε χρώμα ή μαύρο αντιστοιχεί σε μια τιμή 0, ενώ το λευκό αντιπροσωπεύει μια τιμή 100, ορίζοντας έτσι το σύνολο τιμών του πίνακα. Οι διαγώνιες γραμμές του πίνακα (ΑΘ, ΑΑ) σε έναν πίνακα θερμοκρασίας θα είναι πιο ανοιχτές αποχρώσεις, καθώς αντιπροσωπεύουν τις σωστές προβλέψεις. Οι αριθμοί επάνω σε αυτές τις γραμμές αποτελούν τα πραγματικά πλήθη (ή ποσοστά) των δειγμάτων σε αυτές τις κατηγορίες. Η παρουσία των τιμών 100 και 0 σε αυτές τις θέσεις στον πίνακα σύγκρισης είναι ένα θετικό σημάδι για

την καλή απόδοση του μοντέλου. Ωστόσο, στην πράξη, δεν είναι πάντα δυνατό να έχουμε τέλεια απόδοση, και τον πίνακα σύγκυσης συνήθως γεμίζουν και άλλα στοιχεία – χρώματα, που αντιπροσωπεύουν χαμηλότερα ποσοστά ακριβείας του μοντέλου, γνωστά και ως Ψευδείς Θετικοί (ΨΘ) και Ψευδείς Αρνητικοί (ΨΑ). Το μέρος του πίνακα που δεν διαχέεται από τη χρωματισμένη διαγώνια «γραμμή» (ΑΘ, ΑΑ), δηλαδή τα δύο τρίγωνα που δημιουργούνται από τη διχοτόμηση του πίνακα, έχει πιο σκούρο χρώμα, αντιπροσωπεύοντας τις λανθασμένες προβλέψεις (ΨΘ, ΨΑ). Αναλόγως της έντασης του χρώματος σε κάθε κελί, μπορεί να εκτιμηθεί η απόδοση του μοντέλου σε κάθε κατηγορία. Ένας πίνακας θερμοκρασίας παρέχει μια ευρεία οπτική αναπαράσταση των επιδόσεων, καθιστώντας ευκολότερη την ερμηνεία των αποτελεσμάτων.

Με απλά και πιο κατανοητά λόγια, θα επιλεγεί το μοντέλο, του οποίου ο πίνακας σύγκυσης θα φέρει τα πιο «ανοιχτά-έντονα χρώματα», στη νοητή διαγώνιο που προαναφέρθηκε. Αυτό συμβαίνει καθώς σε αυτή τη νοητή διαγώνιο παρατηρούνται τα Αληθείς αναγνωρισμένα δεδομένα (Αληθείς Θετικοί και Αληθείς Αρνητικοί), και πρέπει αυτά να έχουν μεγάλο ποσοστό ακριβείας, ώστε το μοντέλο να θεωρείται αξιόλογο.

Παρά όλα αυτά, τα ανεπτυγμένα μοντέλα πριν την τελική επιλογή θα ελεγχθούν και με τον παρακάτω τρόπο, με Έλεγχο Κατηγοριοποίησης (Classification Report), για να υπάρχει μία πιο ολοκληρωμένη εικόνα για τις ικανότητες του ανεπτυγμένου μοντέλου.

### Έλεγχοι Κατηγοριοποίησης:

Classification Report				
	precision	recall	f1-score	support
0	1.00	0.62	0.77	92
1	0.98	1.00	0.99	44
2	1.00	0.59	0.74	27
3	1.00	0.84	0.91	37
4	1.00	0.72	0.84	78
5	0.81	1.00	0.90	70
6	0.78	1.00	0.87	76
7	1.00	0.33	0.50	30
8	1.00	0.04	0.07	28
9	1.00	1.00	1.00	49
10	0.92	0.90	0.91	52
11	0.70	1.00	0.82	104
12	0.61	1.00	0.76	91
13	0.87	0.95	0.91	41
14	1.00	1.00	1.00	42
15	1.00	0.96	0.98	46
16	0.79	0.99	0.88	90
17	0.97	0.62	0.76	110
18	0.83	0.98	0.90	81
19	1.00	1.00	1.00	44
20	1.00	1.00	1.00	74
21	1.00	0.66	0.79	41
22	1.00	0.96	0.98	80
23	1.00	1.00	1.00	36
accuracy			0.87	1463
macro avg	0.93	0.84	0.84	1463
weighted avg	0.90	0.87	0.86	1463

Εικόνα 5-8: Έλεγχος Κατηγοριοποίησης ReLU

Classification Report				
	precision	recall	f1-score	support
0	1.00	0.05	0.10	92
1	1.00	1.00	1.00	44
2	1.00	0.30	0.46	27
3	1.00	1.00	1.00	37
4	1.00	0.69	0.82	78
5	0.78	0.99	0.87	70
6	1.00	1.00	1.00	76
7	1.00	0.40	0.57	30
8	0.82	0.96	0.89	28
9	1.00	1.00	1.00	49
10	0.98	0.94	0.96	52
11	0.70	0.87	0.77	104
12	0.67	0.93	0.78	91
13	0.97	0.95	0.96	41
14	1.00	1.00	1.00	42
15	1.00	0.93	0.97	46
16	0.89	0.94	0.92	90
17	0.68	1.00	0.81	110
18	0.76	0.99	0.86	81
19	0.97	0.86	0.92	44
20	1.00	1.00	1.00	74
21	1.00	0.76	0.86	41
22	0.94	1.00	0.97	80
23	1.00	1.00	1.00	36
accuracy			0.86	1463
macro avg	0.92	0.86	0.85	1463
weighted avg	0.90	0.86	0.84	1463

Εικόνα 5-9: Έλεγχος Κατηγοριοποίησης eLU

Classification Report				
	precision	recall	f1-score	support
0	1.00	0.90	0.95	92
1	0.98	0.98	0.98	44
2	1.00	0.78	0.88	27
3	1.00	1.00	1.00	37
4	1.00	0.50	0.67	78
5	0.97	1.00	0.99	70
6	0.96	1.00	0.98	76
7	1.00	1.00	1.00	30
8	0.96	0.89	0.93	28
9	1.00	1.00	1.00	49
10	0.88	0.94	0.91	52
11	0.68	1.00	0.81	104
12	0.69	1.00	0.82	91
13	1.00	0.95	0.97	41
14	1.00	0.98	0.99	42
15	0.83	0.96	0.89	46
16	0.97	0.99	0.98	90
17	1.00	0.64	0.78	110
18	0.92	0.86	0.89	81
19	1.00	0.98	0.99	44
20	1.00	1.00	1.00	74
21	1.00	0.93	0.96	41
22	0.99	1.00	0.99	80
23	1.00	1.00	1.00	36
accuracy			0.92	1463
macro avg	0.95	0.93	0.93	1463
weighted avg	0.94	0.92	0.91	1463

Εικόνα 5-10: Έλεγχος Κατηγοριοποίησης GeLU

Classification Report				
	precision	recall	f1-score	support
0	0.95	0.62	0.75	92
1	1.00	1.00	1.00	44
2	1.00	0.22	0.36	27
3	1.00	0.86	0.93	37
4	0.74	0.40	0.52	78
5	0.71	0.99	0.83	70
6	0.84	1.00	0.91	76
7	1.00	0.10	0.18	30
8	0.92	0.43	0.59	28
9	0.92	1.00	0.96	49
10	0.98	0.83	0.90	52
11	0.62	1.00	0.76	104
12	0.48	1.00	0.65	91
13	0.95	0.95	0.95	41
14	1.00	0.74	0.85	42
15	1.00	0.96	0.98	46
16	0.79	0.98	0.87	90
17	1.00	0.22	0.36	110
18	0.71	0.99	0.83	81
19	0.98	0.98	0.98	44
20	1.00	1.00	1.00	74
21	0.95	0.46	0.62	41
22	0.99	0.99	0.99	80
23	1.00	1.00	1.00	36
accuracy			0.80	1463
macro avg	0.90	0.78	0.78	1463
weighted avg	0.86	0.80	0.78	1463

Εικόνα 5-11: Έλεγχος Κατηγοριοποίησης Leaky ReLU

Classification Report				
	precision	recall	f1-score	support
0	1.00	0.16	0.28	92
1	1.00	0.70	0.83	44
2	0.00	0.00	0.00	27
3	1.00	0.11	0.20	37
4	0.97	0.37	0.54	78
5	0.72	0.97	0.83	70
6	0.93	1.00	0.96	76
7	0.00	0.00	0.00	30
8	1.00	0.82	0.90	28
9	1.00	1.00	1.00	49
10	0.00	0.00	0.00	52
11	0.44	1.00	0.61	104
12	0.76	0.99	0.86	91
13	0.83	0.95	0.89	41
14	1.00	0.98	0.99	42
15	0.00	0.00	0.00	46
16	0.70	0.99	0.82	90
17	0.82	0.55	0.66	110
18	0.39	1.00	0.56	81
19	1.00	0.98	0.99	44
20	0.85	1.00	0.92	74
21	0.69	0.61	0.65	41
22	0.99	1.00	0.99	80
23	1.00	1.00	1.00	36
accuracy			0.72	1463
macro avg	0.71	0.67	0.64	1463
weighted avg	0.73	0.72	0.67	1463

Εικόνα 5-12: Έλεγχος Κατηγοριοποίησης Sigmoid

Classification Report				
	precision	recall	f1-score	support
0	1.00	0.29	0.45	92
1	1.00	0.59	0.74	44
2	1.00	0.07	0.14	27
3	1.00	1.00	1.00	37
4	0.90	0.77	0.83	78
5	0.69	1.00	0.82	70
6	1.00	1.00	1.00	76
7	0.00	0.00	0.00	30
8	0.90	0.96	0.93	28
9	1.00	1.00	1.00	49
10	0.98	1.00	0.99	52
11	0.60	1.00	0.75	104
12	0.77	0.98	0.86	91
13	0.95	0.93	0.94	41
14	1.00	0.48	0.65	42
15	0.97	0.74	0.84	46
16	0.97	0.80	0.88	90
17	0.77	0.86	0.82	110
18	0.69	0.99	0.81	81
19	0.97	0.86	0.92	44
20	1.00	1.00	1.00	74
21	0.89	0.95	0.92	41
22	0.77	1.00	0.87	80
23	1.00	1.00	1.00	36
accuracy			0.84	1463
macro avg	0.87	0.80	0.80	1463
weighted avg	0.86	0.84	0.81	1463

Εικόνα 5-13: Έλεγχος Κατηγοριοποίησης Tanh



Classification Report				
	precision	recall	f1-score	support
0	1.00	0.88	0.94	92
1	1.00	0.98	0.99	44
2	1.00	0.30	0.46	27
3	1.00	0.97	0.99	37
4	1.00	0.50	0.67	78
5	0.97	1.00	0.99	70
6	0.92	1.00	0.96	76
7	1.00	1.00	1.00	30
8	0.96	0.79	0.86	28
9	1.00	1.00	1.00	49
10	1.00	0.75	0.86	52
11	0.67	1.00	0.80	104
12	0.72	1.00	0.84	91
13	1.00	0.95	0.97	41
14	1.00	1.00	1.00	42
15	1.00	0.96	0.98	46
16	0.85	0.99	0.91	90
17	1.00	0.67	0.80	110
18	0.70	1.00	0.83	81
19	1.00	0.98	0.99	44
20	0.99	1.00	0.99	74
21	1.00	0.68	0.81	41
22	0.99	0.96	0.97	80
23	1.00	1.00	1.00	36
accuracy			0.90	1463
macro avg	0.95	0.89	0.90	1463
weighted avg	0.93	0.90	0.89	1463

Εικόνα 5-14: Έλεγχος Κατηγοριοποίησης Swiss

Καταλαβαίνοντας αν ένα μοντέλο είναι καλύτερο από ένα άλλο μέσω ενός ελέγχου κατηγοριοποίησης, γίνεται με βάση τις μετρικές που περιλαμβάνονται στον έλεγχο κατηγοριοποίησης (classification report). Αναλόγως των αναγκών της εφαρμογής, επιλέγονται ποιες μετρικές είναι πιο σημαντικές για την αξιολόγησή του μοντέλου. Ορισμένα κοινά κριτήρια για τον προσδιορισμό του καλύτερου μοντέλου περιλαμβάνουν:

- **Ακρίβεια (Precision):** Η υψηλή ακρίβεια είναι σημαντική όταν θέλουμε να μειώσουμε τον αριθμό των false positives. Εάν η ακρίβεια είναι ένα κρίσιμο κριτήριο, επιλέγεται το μοντέλο που έχει υψηλότερη ακρίβεια.
- **Ανάκληση (Recall):** Η υψηλή ανάκληση είναι σημαντική όταν επιθυμούμε να μειώσουμε τον αριθμό των false negatives. Εάν η ανάκληση είναι ένα κρίσιμο κριτήριο, επιλέγεται το μοντέλο που έχει υψηλότερη ανάκληση.
- **F1-Score:** Το F1-Score είναι μια ισορροπημένη μετρική που λαμβάνει υπόψη τόσο την ακρίβεια όσο και την ανάκληση. Εάν έχουμε σκοπό να επιτύχουμε μια ισορροπημένη απόδοση, επιλέγουμε το μοντέλο που έχει υψηλότερο F1-Score.

- **Ακρίβεια (Accuracy):** Η ακρίβεια μετράει το συνολικό ποσοστό των σωστών προβλέψεων. Είναι σημαντική όταν οι κλάσεις είναι ισορροπημένες. Εάν η ακρίβεια είναι το πιο σημαντικό κριτήριο για την εφαρμογή που αναπτύσσεται, επιλέγουμε το μοντέλο με την υψηλότερη ακρίβεια.
- **Μέση Τιμή (Macro Avg) ή Ζυγισμένη Μέση Τιμή (Weighted Avg):** Ανάλογα με το αν προτιμάμε να δώσουμε ισορροπημένη σημασία σε όλες τις κλάσεις ή να λάβουμε υπόψη τον αριθμό των δειγμάτων, επιλέγουμε το μοντέλο με το υψηλότερο Macro Avg ή Weighted Avg.

Παρατηρώντας τα αποτελέσματα, των πινάκων Σύγκυσης και των Ελέγχων Κατηγοριοποίησης, δηλαδή συγκρίνοντας σχολαστικά τα αποτελέσματα των μετρικών των μοντέλων, και πιο συγκεκριμένα τις τιμές της ακρίβειας, της ανάκλησης, του αρμονικού μέσου και της ορθότητας, με τη μόνη διαφορά των μοντέλων να είναι η εκάστοτε συνάρτηση ενεργοποίησης, συμπεραίνουμε πως μεγαλύτερη επιτυχία φέρουν τα μοντέλα με συναρτήσεις ενεργοποίησης την GeLU και Swiss αντίστοιχα.

Παρά τα παραπάνω εντυπωσιακά αποτελέσματα, στον πραγματικό κόσμο, τέτοια τελειότητα συχνά είναι πολύ δύσκολο, έως και αδύνατο, να επιτευχθεί λόγω θορύβου στα δεδομένα, ποικιλία στις χειρονομίες ή και εξωγενών παραγόντων, όπως είναι οι συνθήκες φωτισμού. Επομένως, είναι απαραίτητο να γίνει κατανοητό πως οι πρακτικές εφαρμογές μπορεί να προκαλέσουν περαιτέρω προκλήσεις, απαιτώντας επιπλέον έρευνα και δυναμικές βελτιώσεις [14].

Επιπρόσθετα, οι τιμές του αρμονικού μέσου με ή δίχως συντελεστές, παρέχουν εισαγωγικά στοιχεία για την απόδοση του μοντέλου σε όλες τις κατηγορίες. Η τιμή που δεν επηρεάζεται από συντελεστές και αντιμετωπίζει όλες τις κατηγορίες ισότιμα ονομάζεται **macro**, ενώ η τιμή που επηρεάζεται από συντελεστές και λαμβάνει υπόψη την ανισορροπία των κατηγοριών, δίνοντας μεγαλύτερο βάρος στις κατηγορίες με μεγαλύτερο αριθμό περιπτώσεων, ονομάζεται **weighted**. Η υψηλή τιμή του Αρμονικού Μέσου σε όλα τα επίπεδα των μοντέλων τονίζει την ανθεκτική τους απόδοση [14].

Συμπερασματικά, ο πίνακας σύγκυσης αποτελεί μία βάση για την αξιολόγηση της απόδοσης των μοντέλων ταξινόμησης. Η εξερεύνηση των λεπτομερειών του, σε συνδυασμό με τις βαθμολογίες των μετρικών που παράγονται, διευκολύνει και κατευθύνει την απόφαση του προγραμματιστή για την κατάλληλη αξιολόγηση, συνεπώς και επιλογή, του μοντέλου μηχανικής μάθησης. Με αυτόν τον τρόπο, ενισχύεται σε πολύ μεγάλο βαθμό και η αποτελεσματικότητα της εφαρμογής. Ωστόσο, ανεξαρτήτως των εντυπωσιακών αποτελεσμάτων των μετρικών, χρειάζεται να διατηρείται μία ρεαλιστική προοπτική, λαμβάνοντας υπόψη, σε κάθε βήμα μεταγενέστερο της εκπαίδευσης, αξιολόγησης και τελικής επιλογής του μοντέλου ταξινόμησης μηχανικής μάθησης, τις προκλήσεις που έπεται να προκύψουν σε πρακτικές εφαρμογές, και ως εκ τούτου, να ανοίξει ο δρόμος για μελλοντικές προσπάθειες βελτίωσης [14].

## Κεφάλαιο 6<sup>ο</sup>

### ΣΥΜΠΕΡΑΣΜΑΤΑ και ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ

Στο παρόν κεφάλαιο θα παρουσιαστούν τα συμπεράσματα, από την υλοποίηση της εφαρμογής αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.), στα πλαίσια της παρούσας διπλωματικής εργασίας, όπως και προτάσεις για μελλοντικές κατευθύνσεις για περαιτέρω ανάπτυξη της εφαρμογής ή ενσωμάτωση της σε άλλα συστήματα. Ακολουθεί μία σύνοψη της διπλωματικής εργασίας, και έπειτα θα παρουσιαστούν τα αποτελέσματα. Τέλος, θα γίνει αναφορά σε ορισμένες προτάσεις, που αφορούν την κατεύθυνση που μπορεί να ακολουθήσει κανείς, για την επιπρόσθετη ανάπτυξη της ήδη υπάρχουσας εφαρμογής.

#### 6.1 Συμπεράσματα

Η μελέτη αυτή ολοκληρώθηκε με τη δημιουργία μιας καινοτόμου εφαρμογής αναγνώρισης χειρονομιών σε πραγματικό χρόνο, ειδικά σχεδιασμένη για την αναγνώριση των στατικών χειρονομιών, που αντιστοιχούν στα γράμματα της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.). Χρησιμοποιώντας το API της MediaPipe - Hands, σε συνδυασμό με ένα μοντέλο νευρωνικών δικτύων βαθιάς μάθησης, υλοποιημένα ως μοντέλα της μορφής TensorFlow Lite, αυτό το σύστημα φέρνει στο προσκήνιο αρκετά ενδιαφέροντα συμπεράσματα.

Κύριο επίτευγμα της έρευνας αποτελεί η ανάπτυξη ενός αποδοτικού και προσιτού εργαλείου αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας, που λειτουργεί άψογα τόσο σε ηλεκτρονικούς υπολογιστές, όσο και σε κινητές συσκευές. Πρέπει να σημειωθεί πως ο τρόπος προσέγγισης της υλοποίησης της εφαρμογής, εξαλείφει την ανάγκη για εξειδικευμένο υλικό, διευκολύνοντας στην ενσωμάτωσή της σε άλλα συστήματα ποικίλων μορφών. Η μετατροπή του μοντέλου σε μορφή TensorFlow Lite, και η χρήση του, βελτιστοποιεί επιπλέον τη διαδικασία υλοποίησης σε κινητά υλικά, διασφαλίζοντας ευρεία προσβασιμότητα.

Το λειτουργικό πλαίσιο του συστήματος περιλαμβάνει ένα διαδικαστικό σχήμα πολλών σταδίων. Αρχικά, το σύστημα αναγνωρίζει την περιοχή ενδιαφέροντος, δηλαδή το χέρι, μέσα στο εισερχόμενο βίντεο. Στη συνέχεια, παρακολουθεί τις κινήσεις του χεριού σε διαδοχικά καρέ, εξασφαλίζοντας αξιόπιστη και συνεχή αναγνώριση. Κρίσιμο στοιχείο της συγκεκριμένης προσέγγισης, είναι η ενσωμάτωση ενός μοντέλου σημείων αναφοράς χεριού, ενισχύοντας σημαντικά την ακρίβεια της αναγνώρισης. Επιπρόσθετα,

το πρόσθετο μέτρο ανίχνευσης της παλάμης διαδραματίζει κεντρικό ρόλο στον εντοπισμό του χεριού, όταν το μοντέλο αναγνώρισης σημείων αντιμετωπίζει προκλήσεις στην αναγνώριση. Αυτά τα συνεργατικά στοιχεία συμβάλλουν συλλογικά στη δυνατότητα του συστήματος να ερμηνεύει με αξιοπιστία τις χειρονομίες της Ελληνικής Νοηματικής Γλώσσας.

Η καρδιά του συστήματος αναγνώρισης, εξαρτάται από την ικανότητα του API της MediaPipe να παρέχει ακριβείς δεδομένα αναφοράς. Τα δεδομένα αναφοράς χρησιμοποιούνται για τη δημιουργία του συνόλου δεδομένων, απαραίτητο για την εκπαίδευση του μοντέλου αναγνώρισης μηχανικής μάθησης. Το εκπαιδευμένο μοντέλο είναι ικανό να αναγνωρίζει όλες τις χειρονομίες που αντιστοιχούν στα γράμματα της Ελληνικής Νοηματικής Γλώσσας, μέσα από ένα βίντεο. Η τεχνολογική πρόοδος αυτή, φέρει σημαντικές αποφάσεις σε ποικίλους τομείς, συμπεριλαμβανομένης, αλλά όχι περιοριστικά, της ερμηνείας της νοηματικής γλώσσας, των εφαρμογών εικονικής και επαυξημένης πραγματικότητας, των ηλεκτρονικών παιχνιδιών, και της τεχνολογίας βοήθειας για άτομα με ειδικές ανάγκες.

## 6.2 Μελλοντικές Επεκτάσεις

Οι μελλοντικές προοπτικές είναι ελπιδοφόρες και περιλαμβάνουν αρκετές κατευθύνσεις για περαιτέρω έρευνα ή ανάπτυξη παρόμοιων εφαρμογών με παραπάνω λειτουργίες.

- **Ενσωμάτωση Μοντέλου για Αναγνώριση Κινήσεων:** Το ανεπτυγμένο μοντέλο που χρησιμοποιείται στην εφαρμογή, δεν είναι ικανό να αναγνωρίζει χειρονομίες που αποτελούνται από κινήσεις ή από κάποιο σύνολο κινήσεων. Αυτό έχει ως αποτέλεσμα τον περιορισμό της αποκρυπτογράφησης νοηματικών γλωσσών, οι οποίες περιλαμβάνουν κινήσεις των χεριών και πολλές φορές και του σώματος. Για αυτό το λόγο περιορίστηκε το μοντέλο μόνο στην αναγνώριση των χειρονομιών που αντιστοιχούν στα γράμματα της αλφαβήτα της Ελληνικής Νοηματικής Γλώσσας.
- **Βελτίωση της Ταχύτητας και της Ακρίβειας:** Ενώ το παρόν σύστημα λειτουργεί σε πραγματικό χρόνο, υπάρχει διαρκής χώρος για βελτίωση τόσο στην ταχύτητα, όσο και στην ακρίβεια. Πιθανές βελτιώσεις περιλαμβάνουν τη βελτιστοποίηση της αρχιτεκτονικής του μοντέλου, την ρύθμιση των παραμέτρων του μοντέλου και τη χρήση πιο ισχυρού υπολογιστικού υλικού. Οι βελτιώσεις αυτές θα αυξήσουν τη συνολική απόδοση του συστήματος, καθιστώντας το ακόμα πιο αξιόπιστο και αποδοτικό.

- **Υποστήριξη Πολλαπλών Χρηστών:** Το τρέχον σύστημα σχεδιάστηκε για την αλληλεπίδραση ενός μόνο χρήστη. Χρειάζεται, να επεκταθεί παραπάνω η εφαρμογή, ώστε να συμπεριλαμβάνει πολλαπλούς χρήστες ταυτόχρονα. Η επέκταση αυτή είναι κρίσιμη για τις ομαδικές αλληλεπιδράσεις και το εκπαιδευτικό περιβάλλον που μπορεί να κατασκευαστεί γύρω από την εφαρμογή.
- **Ενσωμάτωση Επιπρόσθετων Εργασιών Όρασης Υπολογιστών:** Η αναγνώριση χειρονομιών μπορεί να ενσωματωθεί άριστα με άλλες εργασίες όρασης υπολογιστών, όπως είναι η ανίχνευση αντικειμένων και η αναγνώριση προσώπων. Η συνδυαστική αυτή χρήση θα δημιουργήσει πιο προηγμένες εφαρμογές σε διάφορους τομείς. Για την επίτευξη αυτής της ενσωμάτωσης, όμως, απαιτείται η ανάπτυξη πολύπλοκων αλγορίθμων και η συντονισμένη λειτουργία πολλών μοντέλων αναγνώρισης.
- **Ενσωμάτωση Μοντέλου Επεξεργασίας Φυσικής Γλώσσας:** Η ενσωμάτωση ενός μοντέλου φυσικής γλώσσας στην εφαρμογή, ενδέχεται να διευκολύνει περαιτέρω την κοινότητα των κωφών και βαρήκοων. Εξάγοντας και αποθηκεύοντας προσωρινά το γράμμα που προκύπτει από την αναγνώριση των χειρονομιών, είναι εφικτό οι χρήστες, σε συνδυασμό με ένα μοντέλο Επεξεργασίας Φυσικής Γλώσσας, να παράγουν ολόκληρες προτάσεις σε ελάχιστο χρονικό διάστημα και να τις στέλνουν ως μήνυμα ή να εμφανίζονται σαν κείμενο στην οθόνη του δέκτη. Τέλος, το μήνυμα - κείμενο αυτό μπορεί να μετατραπεί και σε αρχείο ήχου, ώστε οι πομποί να μην χρειάζεται να διαβάζουν κάποιο κείμενο και να γίνεται η διαδικασία της επικοινωνίας μεταξύ των ανθρώπων της κοινότητας των κωφών και βαρήκοων, και των συνανθρώπων τους ακόμα πιο άμεση, αποτελεσματική και φυσική.

Συνοψίζοντας, στα πλαίσια της διπλωματικής εργασίας παρουσιάζεται ένα καινοτόμο σύστημα αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (Ε.Ν.Γ.), που σκοπό έχει τη γεφύρωση του χάσματος μεταξύ ανθρώπων με ομιλητικά και ακουστικά εμπόδια και των συνανθρώπων τους. Η προσαρμοστικότητα του, η λειτουργία σε πραγματικό χρόνο και οι προοπτικές για επέκταση υπογραμμίζουν τη σημασία του στον τομέα της όρασης υπολογιστών και της αλληλεπίδρασης ανθρώπου υπολογιστή. Καθώς οι μελλοντικές προσπάθειες αρχίζουν, οι ορίζοντες για την αναγνώριση της Ελληνικής Νοηματικής Γλώσσας και την εφαρμογή της συνεχίζουν να διευρύνονται, υποσχόμενοι νέες διαστάσεις προσβασιμότητας για όλους, καινοτόμες τεχνολογίες και λιγότερους επικοινωνιακά περιθωριοποιημένους ανθρώπους.

## BIBΛΙΟΓΡΑΦΙΑ

1. Scientific Reports, ‘*An integrated mediapipe-optimized GRU model for Indian sign language recognition*’, [Online]. Available at: <https://www.nature.com/articles/s41598-022-15998-7>
2. ScienceDirect, ‘*A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition*’, [Online]. Available at: [https://www.sciencedirect.com/science/article/pii/S1877050920312473?ref=pdf\\_download&fr=R-R-2&rr=7f37d006fd4e38d4](https://www.sciencedirect.com/science/article/pii/S1877050920312473?ref=pdf_download&fr=R-R-2&rr=7f37d006fd4e38d4)
3. ScienceDirect, ‘*Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning*’, [Online]. Available at: [https://www.sciencedirect.com/science/article/pii/S1877050923001175?ref=pdf\\_download&fr=R-R-2&rr=7f5108830e29eeb4](https://www.sciencedirect.com/science/article/pii/S1877050923001175?ref=pdf_download&fr=R-R-2&rr=7f5108830e29eeb4)
4. MDPI, ‘*American Sign Language Alphabet Recognition by Extracting Feature from Hand Pose Estimation*’, [Online]. Available at: <https://www.mdpi.com/1424-8220/21/17/5856>
5. Hindaw, ‘*Emerging Technologies, Infrastructures, and Applications for Video Streaming over Future Communication Networks*’, [Online]. Available at: <https://www.hindawi.com/journals/ijdmb/2018/4546896/>
6. ScienceDirect, ‘*A simple feedforward convolutional conceptor neural network for classification*’ [Online]. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S1568494617304982>
7. RSP Science Hub, ‘*Real-Time Computer Vision Based Hand Gesture Recognition*’, (2023) [Online]. Available at: [https://rspsciencehub.com/article\\_23851\\_c2fe4bf164d70bdaeb7b2ce10af39268.pdf](https://rspsciencehub.com/article_23851_c2fe4bf164d70bdaeb7b2ce10af39268.pdf)
8. SpringerLink, ‘*Neural Networks – State of Art, Brief History, Basic Models and Architecture*’, (2016), [Online]. Available at: [https://link.springer.com/chapter/10.1007/978-3-319-45378-1\\_1](https://link.springer.com/chapter/10.1007/978-3-319-45378-1_1)
9. ScienceDirect, ‘*Robust sign language recognition by combining manual and non-manual features based on conditional random field and support vector machine*’, [Online]. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0167865513002559>
10. MediaPipe Github: <https://google.github.io/mediapipe/solutions/hands>. Access 2021.
11. ResearchGate, ‘*Computer vision based Hand gesture recognition system*’, [Online]. Available at: [https://www.researchgate.net/publication/368475746\\_Computer\\_vision\\_based\\_Hand\\_gesture\\_recognition\\_system](https://www.researchgate.net/publication/368475746_Computer_vision_based_Hand_gesture_recognition_system)
12. Scientific Research, ‘*Hand Gesture Recognition Approach for ASL Language Using Hand Extraction Algorithm*’, [Online]. Available at: <https://www.scirp.org/journal/paperinformation.aspx?paperid=59230>
13. CVF, ‘*Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition by Staged Optimization*’, by Runpeng Cui, Hu Li, Changshui Zhang, Department of Automation, Tsinghua University, pp 7361-7368.

14. International Journal of Research Publication and Reviews, '*Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning*', by Arpita Haldera, Akshit Tayadeb, (2021), pp 9-17.
15. '*Adam in Private: Secure and Fast Training of Deep Neural Networks with Adaptive Moment Estimation*', by Nuttapon Attrapadung, Koki Hamada, Dai Ikarashi, Ryo Kikuchi, Takahiro Matsuda, Ibuki Mishina, Hiraku Morita and Jacob C. N. Schuldt, (2021), pp 1-24.
16. '*On Loss Functions for Deep Neural Networks in Classification*', by Katarzyna Janocha, Wojciech Marian Czarnecki, (2017), pp 1-10.
17. Mediapipe and CNNs for Real-Time ASL Gesture Recognition, by Rupesh Kumar, Ashutosh Bajpai, Ayush Sinha, (2023).
18. '*MediaPipe Hands: On-device Real-time Hand Tracking*', by Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, (2020), pp 1-5.
19. ScienceDirect, '*Real time hand detection in a complex background*', by Ekaterini Stergiopoulou, Kyriakos Sgouropoulos, Nikos Nikolaou, Nikos Papamarkos, Nikos Mitianoudis, (2014), pp 54-70.
20. International Research Journal of Modernization in Engineering Technology and Science, '*RECOGNIZATION OF HAND GESTURES USING MEDIAPIPE HANDS*', by Kavana KM, Suma NR, (2022), pp 4149-4155.
21. '*SIGNRELU NEURAL NETWORK AND ITS APPROXIMATION ABILITY*', by Jianfei Li, Han Feng, Ding-Xuan Zhou, (2023) pp 1-28.
22. '*A Simple Way to Prevent Neural Networks from Overfitting*', by Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov, (2014), pp 1929–1958.
23. Journal of Machine Learning Research, '*Adaptive Subgradient Methods for Online Learning and Stochastic Optimization*', by Duchi, J., Hazan, E., & Singer, Y., (2011), pp 2121–2159.
24. World Academy of Science, Engineering and Technology, '*Vision Based Hand Gesture Recognition*', by Pragati Garg, Naveen Aggarwal and Sanjeev Sofat, (2009), pp 972-975.
25. ScienceDirect, '*Background subtraction in real applications: Challenges, current models and future directions*', by Belmar Garcia-Garcia, Thierry Bouwmans, (2019), pp 1-33.
26. International Journal of Machine Learning and Computing, '*Static Sign Language Recognition Using Deep Learning*', by Lean Karlo S. Tolentino, Ronnie O. Serfa Juan, August C. Thio-ac, Maria Abigail B. Pamahoy, Joni Rose R. Forteza, and Xavier Jet O. Garcia, (2019), pp 821-826.
27. E3S Web of Conferences 351, '*Sign Language Recognition : High Performance Deep Learning Approach Applied To Multiple Sign Languages*', by Abdellah El zaar, Nabil Benaya and Abderrahim El allati, (2022), pp 2-7.
28. '*Sign Language Recognition using Deep Learning*', by Mohamed Mahyoub, Friska Natalia, Sud Sudirman, pp 1-6.
29. International Conference on Learning Representations, '*Very Deep Convolutional Networks for Large-Scale Image Recognition*', by K. Simonyan and A. Zisserman, (2015).

30. D. Li, C. R. Opazo, X. Yu, and H. Li, '*Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison*', (2020), pp. 1448–1458.
31. D. Wu, L. Pigou, P.-J. Kindermans, N. Le, L. Shao, J. Dambre, and J.-M. Odobez, '*Deep dynamic neural networks for multimodal gesture segmentation and recognition*', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2016), pp 1583–1597.
32. CVF, '*Rich feature hierarchies for accurate object detection and semantic segmentation*', by Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, (2014), pp 1-8.
33. Atlantis Press International B.V., '*Applying Hand Gesture Recognition for User Guide Application Using MediaPipe*', by Indriani, Moh.Harris, Ali Suryaperdana Agoes1, (2021), pp 101-107.
34. Marco Maisto, '*An Accurate Algorithm for Identification of Fingertips Using an RGB-D Camera*', *IEEE Journal on Emerging and Selected Topics in Circuits and System*, (2013), pp. 272-283.
35. WSEAS, '*MediaPipe to Recognise the Hand Gestures*', by Lavanya Vaishnavi D. A., Anil Kumar C., Harish S., Divya M. L., (2021), pp 134-138.
36. ICPR, '*A framework for hand gesture recognition and spotting using sub-gesture modeling*', by M. R. Malgireddy, J. J. Corso, S. Setlur, V. Govindaraju, and D. Mandalapu', (2010), pp. 3780–3783.
37. JCSSE, '*Workflow translation and dynamic invocation for Image Processing based on OpenCV*', (2015), by A. Traisuwan, P. Tandayya and T. Limna, pp. 319-324.
38. D DIAS R MADEO T. ROCHA H. BISCARO S. PERES, '*Hand movement recognition for American sign language: a study using distance-based OpenCV*',(2009).
39. Scientific Research, '*Real-Time Static Hand Gesture Recognition for American Sign Language (ASL) in Complex Background*', by Jayashree R. Pansare, Shravan H. Gawande, Maya Ingle, (2012), pp 364-367.
40. Proceedings of Machine Learning Research, '*Neural network integral representations with the ReLU activation function*', (2020), by Armenak Petrosyan, Anton Dereventsov and Clayton G. Webster, pp vol 107:128–143.
41. Python 3.12.0 Documentation, '*History and license*', (2023), [Online]. Available at: <https://docs.python.org/3/license.html>
42. Science Direct, '*Artificial Intelligence in Medicine*', (2017), by Pavel\_Hamet and Johanne Tremblay, pp S36-S40, [Online]. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S002604951730015X>
43. Academic Press Series in Cognition and Perception, '*Artificial Intelligence*', (1975), by Earl B. Hunt
44. Machine Learning in Radiation Oncology, '*What is Machine Learning ?*', (2015), by El Naqa, I., Murphy, M.J., pp 3-11
45. American Journal of Epidemiology, '*What is Machine Learning? A Primer for the Epidemiologist*', (2019), by Qifang Bi, Katherine E Goodman, Joshua Kaminsky, Justin Lessler, Volume 188, Issue 12, December 2019, Pages 2222–2239