Ατομική Διπλωματική Εργασία

ΑΝΑΓΝΩΡΙΣΗ ΠΟΝΟΥ ΜΕΣΩ ΤΕΧΝΙΚΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ: ΑΝΑΛΥΣΗ ΒΙΝΤΕΟΓΡΑΦΗΜΕΝΩΝ ΕΚΦΡΑΣΕΩΝ ΠΡΟΣΩΠΟΥ ΣΤΗΝ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ BIOVID HEAT PAIN

Ανδρέου Κωνσταντίνος

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Σεπτέμβριος 2024

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΑΝΑΓΝΩΡΙΣΗ ΠΟΝΟΥ ΜΕΣΩ ΤΕΧΝΙΚΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ: ΑΝΑΛΥΣΗ ΒΙΝΤΕΟΓΡΑΦΗΜΕΝΩΝ ΕΚΦΡΑΣΕΩΝ ΠΡΟΣΩΠΟΥ ΣΤΗΝ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ BIOVID HEAT PAIN

Ανδρέου Κωνσταντίνος

Επιβλέπων Καθηγητής
Παττίχης Κωνσταντίνος

Η Ατομική Διπλωματική Εργασία υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων απόκτησης του πτυχίου Πληροφορικής του Τμήματος Πληροφορικής του Πανεπιστημίου Κύπρου

Σεπτέμβριος 2024

Ευχαριστίες

Με μεγάλη χαρά και βαθιά ευγνωμοσύνη θα ήθελα να εκφράσω τις ευχαριστίες μου προς τον Δρ. Κωνσταντίνο Παττίχη και την κα. Μέλπο Πιτταρά, οι οποίοι με καθοδήγησαν αδιάκοπα καθ' όλη τη διάρκεια της διπλωματικής μου εργασίας. Η αφοσίωση, η εμπιστοσύνη και οι πολύτιμες γνώσεις που μοιράστηκαν μαζί μου ήταν καθοριστικές για την πρόοδο και την επιτυχία της έρευνάς μου. Η εμπειρία και η βαθιά γνώση που διακατέχουν, με υποστήριξαν έμπρακτα στη διαδικασία της μάθησης και της προσωπικής ανάπτυξης, ενθαρρύνοντάς με να αντιμετωπίσω κάθε πρόκληση και να καταλήξω σε σημαντικά επιστημονικά επιτεύγματα. Τους είμαι βαθύτατα ευγνώμων για την ενθάρρυνση και την υποστήριξη που μου παρείχαν σε κάθε βήμα.

Επίσης, θα ήθελα να εκφράσω την εκτίμησή μου προς το ερευνητικό κέντρο CYENS, το οποίο μου παρείχε τους απαραίτητους πόρους και την τεχνογνωσία για να εξελιχθώ επαγγελματικά και επιστημονικά. Η υποδομή και οι ευκαιρίες για έρευνα που μου προσφέρθηκαν ήταν θεμελιώδεις για την επίτευξη των στόχων μου.

Η συνεισφορά όλων των παραπάνω ατόμων και φορέων ήταν αποφασιστική για την ολοκλήρωση αυτής της διπλωματικής εργασίας και τους ευχαριστώ από καρδιάς.

Περίληψη

Η αξιολόγηση του πόνου αποτελεί κρίσιμο μέρος της ιατρικής πρακτικής, καθώς επηρεάζει άμεσα τόσο τη διάγνωση όσο και την επιλογή θεραπείας, συμβάλλοντας καθοριστικά στην ευημερία των ασθενών. Οι παραδοσιακές μέθοδοι εκτίμησης βασίζονται κυρίως σε αυτό-αναφερόμενες κλίμακες και σε εκτιμήσεις των ιατρών, οι οποίες μπορεί να περιέχουν στοιχεία υποκειμενικότητας και να επηρεάζονται από τον ασθενή ή τις συνθήκες. Η υποκειμενικότητα αυτή ενδέχεται να οδηγήσει σε ανακριβείς εκτιμήσεις, με αρνητικές συνέπειες στη θεραπεία και τη διαχείριση του πόνου. Συνεπώς, είναι επιτακτική η ανάγκη για πιο αντικειμενικά και ακριβή συστήματα αξιολόγησης.

Η παρούσα έρευνα διερευνά τη χρήση Συνελικτικών Νευρωνικών Δικτύων (CNNs) και Μετασχηματιστών Όρασης (Vision Transformers, ViTs) για την ενίσχυση της αξιολόγησης του πόνου μέσω ανάλυσης εκφράσεων προσώπου, οι οποίες καταγράφηκαν σε βίντεο και αντιπροσωπεύουν μέγιστη και μηδενική ένταση πόνου. Τα προτεινόμενα μοντέλα εκπαιδεύτηκαν και αξιολογήθηκαν με τη βάση δεδομένων Biovid Heat Pain, εφαρμόζοντας διαφορετικούς συμμετέχοντες για τις φάσεις εκπαίδευσης και αξιολόγησης. Τα αποτελέσματα ήταν εντυπωσιακά, με το MobileVit να επιτυγχάνει ακρίβεια 60% στην ανάλυση εικόνων, ενώ το VideoMAE παρουσίασε μέση ακρίβεια 71% στην ανάλυση βίντεο.

Αξιοσημείωτο είναι ότι τα αποτελέσματα αυτά επιτεύχθηκαν υπό συνθήκες αξιολόγησης σε δεδομένα από άγνωστους συμμετέχοντες. Η παρούσα διπλωματική εργασία παρουσιάζει μια εκτενή συγκριτική ανάλυση διαφορετικών μοντέλων για την επεξεργασία εικόνων και βίντεο, περιλαμβάνοντας λεπτομερή αξιολόγηση της απόδοσής τους, τις προκλήσεις που αντιμετωπίστηκαν, τις τεχνικές βελτιστοποίησης που εφαρμόστηκαν και τα συμπεράσματα από κάθε προσέγγιση.

Λέξεις κλειδιά: εκτίμηση πόνου, μηχανική μάθηση, επεξεργασία εικόνας, βαθιά μάθηση, μετασχηματιστές, ανάλυση βίντεο

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

1.1 Ορισμός του Πόνου και Μέθοδοι αξιολόγησης

Ο πόνος αποτελεί μια πολυδιάστατη εμπειρία που περιλαμβάνει σωματικές και συναισθηματικές πτυχές και μπορεί να προέρχεται από διάφορες αιτίες, όπως τραύματα, ιατρικές καταστάσεις ή γενετικές διαταραχές. Το ανθρώπινο σώμα, αντιλαμβανόμενο τον πόνο, ενεργοποιεί προστατευτικούς μηχανισμούς και προειδοποιητικά σήματα για να ανταποκριθεί στις εσωτερικές ή εξωτερικές απειλές [22]. Η διαχείριση του πόνου είναι εξαιρετικά σημαντική, καθώς η έλλειψη κατάλληλης αντιμετώπισης μπορεί να έχει χ`αρνητικές επιπτώσεις στα όργανα και να παρεμποδίσει τις φυσικές διαδικασίες επούλωσης [28]. Επομένως, η ακριβής και αντικειμενική αξιολόγηση του πόνου είναι ζωτικής σημασίας για την ανάπτυξη αποτελεσματικών πρωτοκόλλων διαχείρισης, που στοχεύουν στην ανακούφιση της ταλαιπωρίας και την αποτροπή της μείωσης της λειτουργικότητας των ασθενών [7].

Η μέτρηση της έντασης του πόνου αποτελεί πρόκληση λόγω της εγγενούς μεταβλητότητας στην εμπειρία του. Μία από τις πιο κοινές μέθοδούς αξιολόγησης, είναι η Αριθμητική Κλίμακα Αξιολόγησης (NRS). Η αξιολόγηση του πόνου ως «πέμπτο ζωτικό σημείο» εισήχθη τη δεκαετία του 1990 για τη βελτίωση της διαχείρισης του πόνου, με την αριθμητική κλίμακα NRS (0–10) να χρησιμοποιείται ευρέως σε κλινικές πρακτικές για την παρακολούθηση της έντασης του πόνου μέσω ηλεκτρονικών ιατρικών

αρχείων [44][9]. Κάποιες άλλες επίσης γνωστές μεθόδους αξιολόγησης αποτελούν, η Οπτική Αναλογική Κλίμακα (VAS) και η Λεκτική Κλίμακα Αξιολόγησης (VRS) οι οποίες απαιτούν άμεση επικοινωνία με τον ασθενή [9].

Ωστόσο, αυτές οι μέθοδοι παρουσιάζουν περιορισμούς, ειδικά για ασθενείς που δεν μπορούν να επικοινωνήσουν, όπως όσοι βρίσκονται σε αναισθησία, νεογέννητοι ή άτομα με γνωστικές διαταραχές. Επιπλέον, οι αυτό-αναφορές δεν αντανακλούν πάντα με ακρίβεια την υποκειμενική εμπειρία του πόνου, καθώς μπορούν να επηρεαστούν από προκαταλήψεις, μνημονικά κενά και την ικανότητα έκφρασης [27], [43].

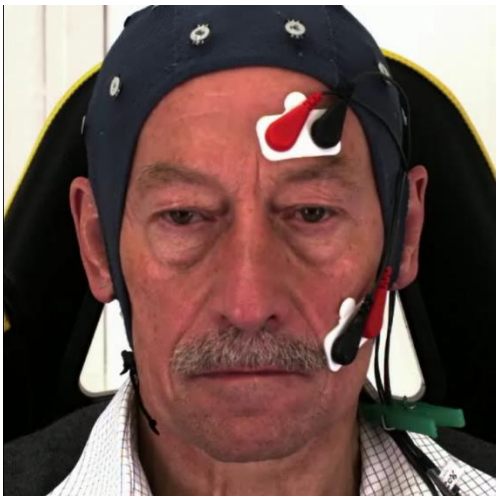
Η ανάγκη για αντικειμενικές μεθόδους αξιολόγησης του πόνου είναι συνεπώς προφανής, ειδικά σε περιπτώσεις όπου οι παραδοσιακές μέθοδοι αξιολόγησης δεν είναι εφαρμόσιμες ή αξιόπιστες.

1.2 Ανασκόπηση Προηγούμενων Μελετών

Λαμβάνοντας υπόψη τους περιορισμούς των υποκειμενικών μεθόδων αξιολόγησης του πόνου, η ερευνητική κοινότητα έχει στραφεί προς την αναζήτηση πιο αντικειμενικών προσεγγίσεων. Η ανάλυση βιοσημάτων, όπως το ηλεκτροεγκεφαλογράφημα (EEG) (*Εικόνα 1.2.1*), το ηλεκτροκαρδιογράφημα (ΗΚΓ) και η ηλεκτροδερμική δραστηριότητα (ΗΔΑ), σε συνδυασμό με την παρατήρηση των εκφράσεων του προσώπου και της κίνησης του σώματος, αναγνωρίζεται πλέον ως μια πιο αξιόπιστη μέθοδος για την αξιολόγηση του πόνου.

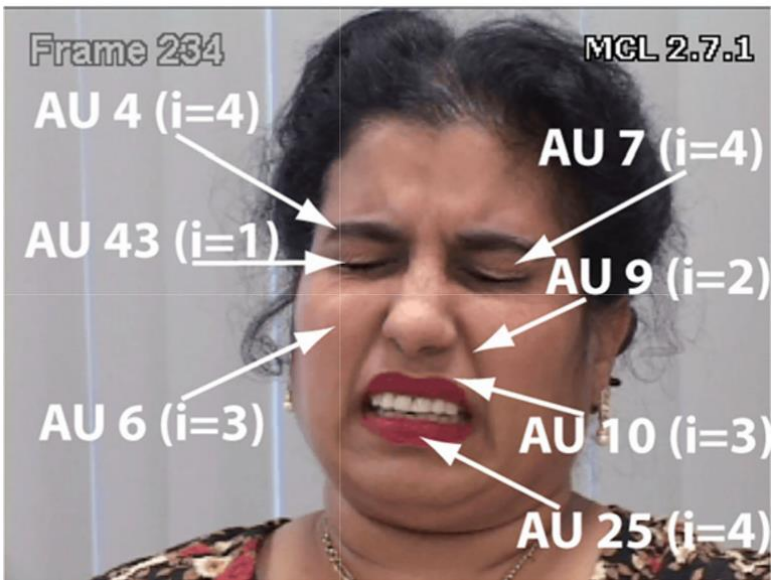
Τα βιοσήματα προσφέρουν μια αντανακλαστική και αυτοματοποιημένη απάντηση που δεν επηρεάζεται από τη συνειδητή διάθεση του ατόμου [24], [36]. Συγκεκριμένα, οι μέθοδοι ηλεκτροεγκεφαλογραφήματος (EEG) αξιολογούν τον πόνο αναλύοντας βιοσήματα που σχετίζονται με τη δραστηριότητα του εγκεφάλου. Ο πόνος προκαλεί χαρακτηριστικές αλλαγές στα πρότυπα EEG σε πολλές περιοχές του εγκεφάλου και σε διαφορετικές συχνότητες. Η διαδικασία περιλαμβάνει την εξαγωγή ενός «Δείκτη Πόνου» (Pain Index - Pi) από τα δεδομένα EEG, χρησιμοποιώντας αλγόριθμους κυματιδιακής ανάλυσης (wavelet), που μετασχηματίζουν τα στοιχεία συχνότητας των εγκεφαλικών κυμάτων ώστε να αντικατοπτρίζουν τα χαρακτηριστικά που σχετίζονται με τον πόνο. Ο Pi, ο οποίος κυμαίνεται από 0 έως 100, συσχετίζεται με υποκειμενικές κλίμακες πόνου

όπως η Οπτική Αναλογική Κλίμακα (VAS) και η Αριθμητική Κλίμακα Εκτίμησης (NRS). Αποτελεί ένα ποσοτικό μέτρο της έντασης του πόνου, ανεξάρτητο από εξωτερικές επιρροές, παρέχοντας έτσι ένα αντικειμενικό και αξιόπιστο εργαλείο για την αξιολόγηση του πόνου [45].



Εικόνα 1.2.1: Καταγραφή EEG σημάτων από τη βάση BioVid Heat Pain, PART B - Επίπεδο πόνου BL1

Σημαντικές προσπάθειες έχουν γίνει και στον τομέα της ανάλυσης των εκφράσεων του προσώπου. Ο Facial Action Coding System (FACS) αποτελεί έναν από τους πλέον διαδεδομένους και αξιόπιστους τρόπους κωδικοποίησης εκφράσεων του προσώπου στην έρευνα. Ο FACS, που προτάθηκε το 1978 από τους Ekman και Friesen, έχει σχεδιαστεί για την παροχή αντικειμενικών περιγραφών της δραστηριότητας του προσώπου που σχετίζεται με έξι βασικά συναισθήματα: χαρά, αποστροφή, θυμός, φόβο, έκπληξη και λύπη. Η κωδικοποίηση βασίζεται σε 44 "μονάδες δράσης" (Action Units - AUs), οι οποίες αναγνωρίζονται μέσω ανάλυσης βίντεο σε αργή κίνηση (*Εικόνα 1.2.2*). Επιπλέον, τόσο η συχνότητα όσο και η ένταση κάθε AU αξιολογούνται σε μια κλίμακα πέντε βαθμίδων. Ο FACS αποτελεί ένα αντικειμενικό εργαλείο, καθώς η ανάλυση γίνεται από εκπαιδευμένους κωδικοποιητές που έχουν πιστοποιηθεί με βάση συγκεκριμένα πρότυπα ακρίβειας [46].



Εικόνα 1.2.2: Παράδειγμα εκφράσεων προσώπου που εκφράζουν πόνο από το αρχείο UNBC-McMaster Shoulder Pain Expression Archive, με τις αντίστοιχες Μονάδες Δράσης (Action Units - AUs) και τις εντάσεις τους (i = ένταση κάθε AU) [47]

Εκτός από τον FACS, αξιοσημείωτη είναι και η συμβολή του PSPI (Prkachin and Solomon Pain Intensity). Πρόκειται για έναν καινοτόμο αλγόριθμο μέτρησης του πόνου, που βασίζεται στην ανάλυση συγκεκριμένων εκφράσεων του προσώπου (*Εικόνα 1.2.3*). Ο PSPI αξιολογεί την ένταση του πόνου μέσω της παρατήρησης τεσσάρων βασικών μονάδων δράσης: του χαμηλώματος των φρυδιών, του σφιξίματος γύρω από τα μάτια, της ανύψωσης του άνω χείλους και του κλεισίματος των ματιών. Η συνολική βαθμολογία υπολογίζεται ως το άθροισμα των εντάσεων αυτών των μονάδων, σε μια κλίμακα από 0 (καθόλου πόνος) έως 16 (μέγιστος πόνος). Η μετρική αυτή επιτρέπει μια ακριβή και διαβαθμισμένη αξιολόγηση του πόνου, καθώς οι τιμές ομαδοποιούνται περαιτέρω σε διακριτά επίπεδα: καθόλου πόνος, ίχνος πόνου, ασθενής πόνος και ισχυρός πόνος [47].

Εικόνα 1.2.3 : Παραδείγματα από τη βάση δεδομένων UNBC-McMaster Shoulder Pain Expression Archive, με την αντίστοιχη ένταση πόνου υπολογισμένη μέσω της μετρικής PSPI [44].

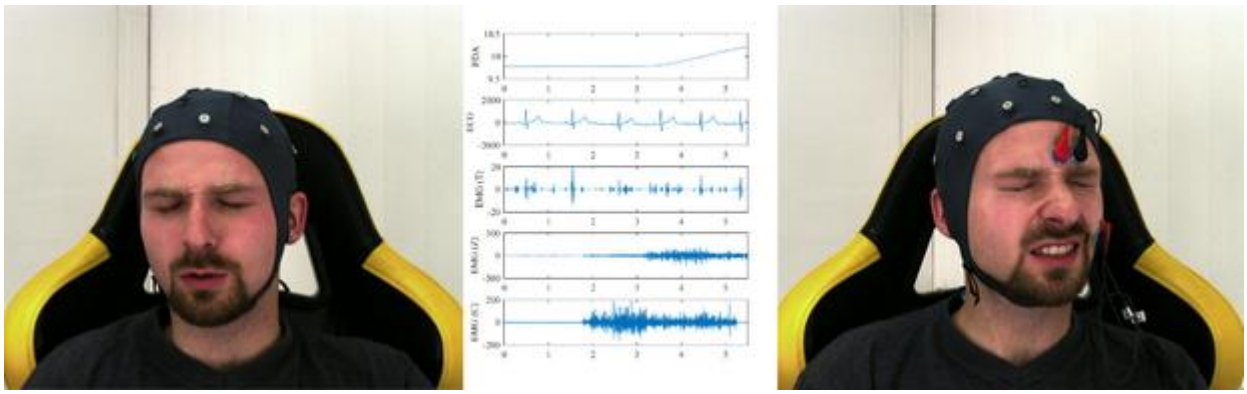
Επιπλέον, η δυνατότητα αυτόματης ανάλυσης αυτών των δεδομένων μέσω τεχνικών μηχανικής μάθησης και υπολογιστικής όρασης έχει προσελκύσει το ενδιαφέρον της ερευνητικής κοινότητας, με πολλές πρωτοβουλίες για την ανάπτυξη συστημάτων αναγνώρισης πόνου [31], [6], [3], [26]. Αυτές οι προσεγγίσεις επιτρέπουν την αντικειμενική και συνεχή παρακολούθηση του πόνου, ιδιαίτερα σε περιπτώσεις όπου η λεκτική επικοινωνία είναι αδύνατη ή αναξιόπιστη.

1.3 Βάση Δεδομένων BioVid Heat: Δομή και Χαρακτηριστικά

Η βάση δεδομένων BioVid Heat Pain αποτελεί μία από τις ελάχιστες αξιοσημείωτες βάσεις δεδομένων για τη μελέτη του πόνου. Η βάση δεδομένων BioVid Heat Pain διαρθρώνεται σε πέντε διακριτά μέρη (Part A έως Part E), καθένα από τα οποία εξυπηρετεί διαφορετικούς ερευνητικούς σκοπούς [48]. Το Part A περιλαμβάνει 8.700 βίντεο RGB χωρίς EMG προσώπου που καταγράφουν τις εκφράσεις προσώπου και

βιοϊατρικά σήματα (GSR, ECG, και EMG στον τραπεζοειδή μυ), ενώ το Part B περιλαμβάνει 8.600 δείγματα με μερικώς καλυμμένο πρόσωπο και επιπλέον EMG αισθητήρες στους μύες corrugator και zygomaticus. Το Part C περιέχει μεγαλύτερης διάρκειας βίντεο, το Part D περιλαμβάνει προσποιητό πόνο και βασικά συναισθήματα, και το Part E (γνωστό και ως BioVid Emo DB) επικεντρώνεται στην πρόκληση συναισθημάτων μέσω βίντεο κλιπ [48].

Το πρωτόκολλο συλλογής δεδομένων περιλαμβάνει πέντε διακριτά επίπεδα έντασης θερμικού ερεθίσματος: ένα επίπεδο βάσης (baseline/BL1) και τέσσερα προοδευτικά αυξανόμενα επίπεδα πόνου (PA1-PA4). Η μεθοδολογία εφαρμόστηκε σε ένα δείγμα 90 υγιών συμμετεχόντων, ηλικιακού εύρους 18-65 ετών, διασφαλίζοντας την αντιπροσωπευτικότητα διαφορετικών ηλικιακών ομάδων [35]. Για την αντιστάθμιση των διαφορετικών ευαισθησιών στον πόνο, οι θερμοκρασίες διέγερσης προσαρμόστηκαν με βάση το ατομικό κατώφλι πόνου και την ανοχή στον πόνο κάθε συμμετέχοντα [48].



Εικόνα 1.3.1 1 Παράδειγμα διαμόρφωσης μέτρησης με επαγόμενο ερέθισμα πόνου, [45].

Ιδιαίτερη σημασία έχει η μεθοδολογία πρόκλησης του πόνου, η οποία, αν και ελεγχόμενη εργαστηριακά, κατάφερε να προκαλέσει αυθεντικές εκφράσεις πόνου, εξασφαλίζοντας την οικολογική εγκυρότητα των δεδομένων [5], [31]. Κάθε επίπεδο πόνου διεγέρθηκε 20 φορές σε τυχαία σειρά, με τη μέγιστη θερμοκρασία να διατηρείται για 4 δευτερόλεπτα και τα διαλείμματα μεταξύ των ερεθισμάτων να κυμαίνονται τυχαία μεταξύ 8-12 δευτερολέπτων [48]. Η βάση δεδομένων ακολουθεί αυστηρά πρωτόκολλα τυποποίησης και διαθεσιμότητας, καθιστώντας την προσβάσιμη στην ερευνητική κοινότητα μέσω ανοικτού κώδικα [28], [30].

Η δομή των δεδομένων επιτρέπει την εξαγωγή χρονικά συγχρονισμένων χαρακτηριστικών τόσο από τις οπτικές καταγραφές όσο και από τα βιοσήματα, διευκολύνοντας την πολυτροπική ανάλυση και την ανάπτυξη ολοκληρωμένων συστημάτων αναγνώρισης πόνου [29]. Στην παρούσα έρευνα, για τον προσδιορισμό της έντασης του πόνου αποκλειστικά μέσω των εκφράσεων του προσώπου στην κορύφωση του πόνου και τη στιγμή χωρίς πόνο, χρησιμοποιήθηκαν τα δεδομένα από τα Μέρη Α και Β της βάσης δεδομένων [5].

1.4 Στόχος και Συμβολή της Παρούσας Μελέτης

Η παρούσα έρευνα επικεντρώνεται στην Αναγνώριση των διαφορετικών Επιπέδων Πόνου μέσω προηγμένων «Μοντέλων Μηχανικής Μάθησης», με στόχο την ανάπτυξη ενός καινοτόμου συστήματος αντικειμενικής αξιολόγησης του πόνου. Το βασικό όραμα της μελέτης είναι η δημιουργία ενός εξελιγμένου εργαλείου, το οποίο θα μπορεί να παρέχει αξιόπιστες μετρήσεις διαφορετικών επιπέδων πόνου, αξιοποιώντας τεχνικές επεξεργασίας και μηχανικής εκμάθησης εικόνων και βίντεο.

Η συνεισφορά της έρευνας είναι πολύ-επίπεδη, εστιάζοντας σε καίρια ζητήματα της σύγχρονης ιατρικής. Κύριος στόχος είναι η βελτίωση της ποιότητας ζωής των ασθενών μέσω της παροχής ενός αξιόπιστου εργαλείου που υποστηρίζει τους επαγγελματίες υγείας στην αντικειμενική εκτίμηση του πόνου. Το προτεινόμενο σύστημα απευθύνεται σε ένα ευρύ φάσμα ιατρικών εφαρμογών, όπως νοσοκομεία, κέντρα αποκατάστασης, γηροκομεία, παιδιατρικές μονάδες και δομές φροντίδας για ασθενείς με περιορισμένη λεκτική επικοινωνία.

Μία από τις βασικές καινοτομίες της μελέτης είναι η διερεύνηση και εφαρμογή προηγμένων μοντέλων μηχανικής μάθησης για την εκτίμηση του πόνου, με ιδιαίτερη έμφαση στην ανάλυση εκφράσεων προσώπου. Χρησιμοποιώντας τη βάση δεδομένων **BioVid Heat**, στοχεύουμε στη δημιουργία ενός μοντέλου που ξεπερνά τα όρια της υπάρχουσας έρευνας, προσφέροντας καινοτόμες προσεγγίσεις για την αντικειμενική μέτρηση του πόνου.

Μελλοντικές Προοπτικές

Η έρευνα ανοίγει τον δρόμο για τις παρακάτω κατευθύνσεις:

- **Ενσωμάτωση πολυτροπικών δεδομένων**, όπως βιολογικά σήματα (π.χ. καρδιακός ρυθμός, ηλεκτρομυογράφημα).
- **Ανάπτυξη συστημάτων πραγματικού χρόνου**, προσαρμοσμένων για χρήση σε κλινικά περιβάλλοντα.
- **Βελτίωση της γενικευσιμότητας των μοντέλων**, μέσα από τεχνικές προσαρμογής σε επιχώρια δεδομένα.
- **Χρήση τεχνικών Επεξηγήσιμης Τεχνητής Νοημοσύνης (Explainable AI[1])** για ενίσχυση της διαφάνειας στις αποφάσεις.

Απώτερος στόχος της έρευνας είναι η δημιουργία ενός ολοκληρωμένου και αξιόπιστου συστήματος που θα επιτρέπει στους επαγγελματίες υγείας να αποκτούν ακριβέστερα και πληρέστερα δεδομένα. Με αυτόν τον τρόπο, θα ενισχύεται η εξατομίκευση των θεραπευτικών προσεγγίσεων, συμβάλλοντας στη συνολική βελτίωση της φροντίδας των ασθενών.

[1] είναι η τεχνητή νοημοσύνη που επιτρέπει την κατανόηση και ερμηνεία των αποφάσεών της από τους ανθρώπους, αυξάνοντας τη διαφάνεια και την εμπιστοσύνη.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Θεωρητικό Υπόβαθρο στην Ανάλυση Εικόνας

Διάφορες προσεγγίσεις έχουν προταθεί για την αυτόματη αναγνώριση του πόνου και της έντασής του, χρησιμοποιώντας μετρήσιμα φυσιολογικά και οπτικοακουστικά δεδομένα από τη βάση δεδομένων Biovid Heat Pain. Επεκτείνοντας αυτό το σημείο, οι επόμενες παράγραφοι αναδεικνύουν τις πιο σημαντικές αρχιτεκτονικές και μεθοδολογίες στην αξιολόγηση του πόνου μόνο για το Μέρος Α της βάσης δεδομένων Biovid Heat Pain.

Στη βιβλιογραφία, οι περισσότερες μελέτες (7/12) αντιμετωπίζουν τον πόνο μέσω της αρχιτεκτονικής του Convolutional Neural Network (CNN) [31], [6], [3], [26], [43], [32], [40], η οποία αποτελεί μία από τις κυρίαρχες μεθοδολογίες για τη βάση δεδομένων Biovid Heat Pain. Τα CNN παρουσιάζουν υψηλή ακρίβεια στην ταξινόμηση, μειώνοντας την ανάγκη για σημαντική ανθρώπινη παρέμβαση, καθώς κατανοούν τα χαρακτηριστικά ανεξάρτητα [1]. Στην βιβλιογραφία του Devi et al., ανέπτυξαν ένα σύστημα αναγνώρισης πόνου βασισμένο στις εκφράσεις του προσώπου, χρησιμοποιώντας ρηχό CNN. Αρχικά, μετέτρεψαν τα καρέ από τα δεδομένα βίντεο σε εικόνες κλίμακας του γκρι και μείωσαν το μέγεθός τους σε 256x256 για να ελαχιστοποιήσουν τη διαδικασία υπολογισμού. Για την επίτευξη της μη γραμμικής δραστηριότητας, χρησιμοποίησαν τη συνάρτηση ενεργοποίησης ReLU. Επιπλέον, χρησιμοποιήθηκε ένα dropout layer για την αντιμετώπιση ζητημάτων υπερεκπαίδευσης. Η παρούσα μελέτη εκπαίδευσε 2371 δεδομένα και δοκίμασε 264 δεδομένα από 9 συμμετέχοντες, με ακρίβεια 94% [31]. Από την άλλη πλευρά, σε έρευνα του Dragomir et al., χρησιμοποίησαν την αρχιτεκτονική ResNet 18 CNN για την αξιολόγηση της έντασης του πόνου. Η ResNet 18 είναι μια

αρχιτεκτονική που αποτελείται από αρκετά residual blocks, αντί να μαθαίνει μη αναφερόμενες συναρτήσεις. Η residual connection δημιουργεί μια συντόμευση μέσα στο διαδοχικό δίκτυο. Αυτό επιτρέπει την προσθήκη της εισόδου από ένα προηγούμενο επίπεδο σε μια επόμενη ενεργοποίηση στο δίκτυο, διευκολύνοντας τη ροή των κλίσεων από το τελικό επίπεδο στο αρχικό επίπεδο [6]. Η υπάρχουσα μελέτη χρησιμοποίησε 9615 εικόνες από 60 συμμετέχοντες για εκπαίδευση και 2885 εικόνες από 27 συμμετέχοντες για δοκιμή. Για τη βελτίωση της ακρίβειας των αποτελεσμάτων, πραγματοποιήθηκε ενίσχυση δεδομένων (data augmentation) στις εικόνες εκπαίδευσης με διάφορους τυχαίους μετασχηματισμούς που παρέχονται στο Keras [30]. Εναλλακτικά, Lledo et al., έτρεξαν αρχιτεκτονική CNN για την αξιολόγηση των επιπέδων πόνου των συμμετεχόντων. Η παρούσα μελέτη συνέλεξε ένα υποσύνολο δεδομένων βίντεο συμμετεχόντων, που ήταν συνεπές με την ίδια κατηγορία, φύλο και ηλικία. Ενώ η εργασία τους έδειξε μεγάλη δυναμική στην αξιοποίηση των εκφράσεων του προσώπου για την ανίχνευση πόνου χρησιμοποιώντας Vision Transformer, η ακρίβεια των δεδομένων ήταν κάτω από 50% [3]. Η μελέτη του Zheng παρουσίασε ένα μοντέλο βασισμένο σε CNN 5 επιπέδων. Κατηγοριοποίησαν τα βίντεο σύμφωνα με το επίπεδο πόνου που αντιμετώπισαν οι συμμετέχοντες. Στη συνέχεια, διέκριναν τα καρέ που έδειχναν αντιδράσεις προσώπου και απέκοψαν τα μέρη που δεν είχαν αντιδράσεις. Έπειτα, επέλεξαν 1 καρέ από κάθε 10 στη σειρά ταινίας για να εξυπηρετήσει το σύνολο εκπαίδευσης και 1 από κάθε 22 καρέ για το σύνολο δοκιμής. Όλες οι εικόνες ήταν σε κλίμακα του γκρι, αν και εικόνες που δεν είχαν επαρκή χαρακτηριστικά από τον αλγόριθμο αναγνώρισης προσώπου παρακάμφθηκαν αυτόματα. Εφαρμόστηκε ενίσχυση δεδομένων για τη βελτίωση των δεδομένων. Αυτή η μελέτη έδειξε ακρίβεια 52% για τα δεδομένα πόνου και χωρίς πόνο [43].

Σε διαφορετική βιβλιογραφία, η Patania et al., χρησιμοποίησαν την αρχιτεκτονική Graph Neural Networks (GNN) για την αναγνώριση βίντεο με έντονο πόνο έναντι αυτών χωρίς πόνο. Επέλεξαν 20 βίντεο πόνου και 20 ουδέτερα βίντεο από συνολικά 87 συμμετέχοντες, με συνολικό αριθμό 3480 βίντεο. Κάθε ακολουθία συλλέχθηκε μέσω του MediaPipe Python [15]. Το μοντέλο GNN εκπαιδεύτηκε χρησιμοποιώντας τον αλγόριθμο βελτιστοποίησης Adam για τη μείωση της απώλειας διασταυρούμενης εντροπίας και αναλύθηκε με 5-fold cross-validation. Η μελέτη του Patania τροποποίησε τη δομή του δικτύου προσθέτοντας graph convolutional layers στο αρχικό δίκτυο και διπλασιάζοντας

τον αριθμό των πυρήνων. Πλήρως, μια συνάρτηση ενεργοποίησης ReLU αντικατέστησε τη συνάρτηση ενεργοποίησης υπερβολικού εφαπτόμενου στη δομή του δικτύου. Η προτεινόμενη μέθοδος ήταν αποδοτική, παρέχοντας ακρίβεια 73,2% [23]. Εναλλακτικά, οι Tavakolian et al. [32], χρησιμοποίησαν τα πρόσωπα από τα βίντεο σε Spatiotemporal Convolutional Network (SCN), πραγματοποιώντας διάφορα επίπεδα συνελικτικών λειτουργιών με διαφορετικά βάθη στον χρόνο. Η ακρίβεια της αρχιτεκτονικής SCN ήταν 86,02%. Τέλος, οι Xin et al. [40], ανέπτυξαν ένα πολυ-εργασιακό πλαίσιο με CNN. Χρησιμοποιώντας τη στοχαστική καθοδική κλίση (SGD), το CNN ενσωμάτωσε ένα autoencoder attention module. Αυτή η έρευνα πέτυχε ακρίβεια 85,65% για δεδομένα πόνου και χωρίς πόνο, και 40,4% για όλα τα επίπεδα.

Αντίθετα, τρεις μελέτες (3/12) διερεύνησαν τη δυνατότητα των μοντέλων αναγνώρισης πόνου τους μέσω της αρχιτεκτονικής Random Forest (RF) [26], [38], [36]. Το RF είναι μια μέθοδος ensemble που περιλαμβάνει πολλαπλά δέντρα αποφάσεων. Όταν παρουσιάζεται ένα μοτίβο δοκιμής, κάθε δέντρο το αξιολογεί και η τελική πρόβλεψη βασίζεται στην πλειοψηφική επιλογή. Για τη βελτίωση του αποτελέσματος, κάθε κόμβος του δέντρου εκπαιδεύεται σε ένα τυχαία επιλεγμένο υποσύνολο των δεδομένων εκπαίδευσης. Αυτή η προσέγγιση έχει αποδειχθεί επιτυχής, προσφέροντας πλεονεκτήματα όπως υψηλή προβλεπτική απόδοση, ταχεία εκπαίδευση και γρήγορη πρόβλεψη [33].

Το 2014, Werner et al., δημιούργησαν ένα αυτόματο σύστημα αναγνώρισης πόνου, συνδυάζοντας την ανάλυση εκφράσεων προσώπου σχετικών με χαρακτηριστικά πόνου μέσω του IntraFace [41], ενός προηγμένου ανιχνευτή σημείων χαρακτηριστικών προσώπου. Επιπλέον, ο αλγόριθμος iterative closest point (ICP) [24] χρησιμοποιήθηκε για την εκτίμηση της κλίσης του κεφαλιού και άλλων σημείων χαρακτηριστικών προσώπου. Εφαρμόστηκε διαστρωματωμένη 5-fold cross-validation στο σετ εκπαίδευσης για την επιλογή των βέλτιστων τιμών για τις παραμέτρους του RF. Ο αριθμός των δέντρων καθορίστηκε στα 100, με μέγιστο βάθος 6 και ελάχιστο αριθμό δειγμάτων για τη διάσπαση κόμβου 5. Η αρχιτεκτονική RF με τις παραπάνω παραμέτρους εκπαιδεύτηκε σε ένα σύνολο 8700 βίντεο. Η ακρίβεια του συστήματος ήταν 70% [38]. Το 2017, Werner et al., πρότειναν ένα νέο σύνολο χαρακτηριστικών για την περιγραφή της δραστηριότητας του προσώπου, που ονομάζεται facial activity descriptor

(FAD). Το χρησιμοποίησαν για την αναγνώριση πόνου και την εκτίμηση της έντασης του πόνου από σύνολα δεδομένων βίντεο. Η προτεινόμενη προσέγγιση ξεπέρασε προηγούμενες μεθόδους αιχμής στην ταξινόμηση πόνου στη βάση δεδομένων BioVid Heat Pain. Επιπλέον, παρατήρησαν ότι η προβλεπτική απόδοση της προηγούμενης μελέτης τους με RF επωφελείται περισσότερο από την αύξηση του αριθμού των δέντρων. Ως αποτέλεσμα, δημιουργήθηκαν 1000 δέντρα με μέγιστο βάθος 10 κόμβους. Όπως και προηγουμένως, αυτές οι παράμετροι εκπαιδεύτηκαν σε ολόκληρο το σετ εκπαίδευσης των 8700 βίντεο. Η ακρίβεια του παρόντος συστήματος αυξήθηκε στο 72,4% [36]. Τέλος, οι Othman et al., χρησιμοποίησαν το OpenFace και το FAD για την εξαγωγή χαρακτηριστικών προσώπου από κάθε καρέ για κάθε συμμετέχοντα. Το OpenFace μπορεί να ανιχνεύσει το πρόσωπο, τα σημεία αναφοράς του προσώπου, να εξαγάγει action units και να εκτιμήσει την κλίση του κεφαλιού [2]. Οι εικόνες μειώθηκαν σε μέγεθος 96x96 και μετατράπηκαν σε κλίμακα του γκρι. Επιπλέον, ολόκληρο το σύνολο δεδομένων εκπαίδευσης ενισχύθηκε τεχνητά. Η αρχιτεκτονική RF με 100 δέντρα χρησιμοποιήθηκε για την αξιολόγηση του πόνου. Τα αποτελέσματα έδειξαν ακρίβεια 65,8%.

Τέλος, τρεις μελέτες (3/12) χρησιμοποίησαν μοντέλα βασισμένα σε transformers για την πρόβλεψη της έντασης του πόνου χρησιμοποιώντας τη βάση δεδομένων Biovid Heat Pain. Οι Gigkas et al., εφάρμοσαν ένα πλαίσιο βασισμένο σε transformers που ενσωματώνει ένα χωρικό υποσύστημα εξαγωγής χαρακτηριστικών μέσω του μοντέλου Transformer in Transformer (TNT) και ένα χρονικό υποσύστημα εξαγωγής χαρακτηριστικών με μπλοκ cross και self-attention. Αυτό το πλαίσιο, που περιλαμβάνει 24 εκατομμύρια παραμέτρους, πέτυχε ακρίβεια 73,28% για δυαδική ανίχνευση πόνου (πόνος έναντι μη πόνου) και 31,52% για ταξινομήσεις πολλαπλών επιπέδων πόνου [8]. Το 2024, οι ίδιοι ερευνητές ανέπτυξαν ένα βελτιωμένο πλαίσιο, το οποίο περιλάμβανε ένα χωρικό υποσύστημα, έναν κωδικοποιητή καρδιακού ρυθμού, το AugmNet και ένα χρονικό υποσύστημα. Το χωρικό υποσύστημα, βασισμένο στην αρχιτεκτονική TNT, επεξεργάστηκε καρέ βίντεο με ανάλυση 448x448 pixels, με τις περιοχές του προσώπου να απομονώνονται μέσω του ανιχνευτή προσώπου MTCNN. Αυτό το βελτιωμένο πλαίσιο πέτυχε ακρίβεια 77,10% στη δυαδική ταξινόμηση και 35,39% στην ταξινόμηση πολλαπλών επιπέδων πόνου. Αξιοσημείωτα, όταν συνδυάστηκαν δεδομένα βίντεο και καρδιακού ρυθμού, η ακρίβεια αυξήθηκε στο 82,74% για τη δυαδική ταξινόμηση και στο

39,77% για την ταξινόμηση πολλαπλών επιπέδων πόνου [7]. Επιπλέον, οι Lledo et al., χρησιμοποίησαν το μοντέλο ViViT για την ταξινόμηση βίντεο, προτείνοντας διάφορες αρχιτεκτονικές βασισμένες σε transformers, όπως τον Transformer Encoder, που επεκτείνει το ViT στη χρονική διάσταση, τον Factorized Encoder, που διαχωρίζει την επεξεργασία της χωρικής και χρονικής προσοχής, και τον Factorized Dot Product, που τροποποιεί την πρώτη προσέγγιση σε χαμηλότερο επίπεδο. Από αυτά, ο Factorized Encoder παρήγαγε τα πιο ελπιδοφόρα αποτελέσματα, επιτυγχάνοντας ακρίβεια 30,07% στην ταξινόμηση πολλαπλών επιπέδων πόνου [3].

Πίνακας 2.1.1 : Ο παρακάτω πίνακας δείχνει τις αρχιτεκτονικές που χρησιμοποιούνται για την ανίχνευση πόνου από εικόνες, τις λεπτομέρειες προεπεξεργασίας στην είσοδο, το μέτρο απόδοσης, τις περιπτώσεις, τον αριθμό των εικονοστοιχείων στην είσοδο, τον αριθμό των εικόνων για εκπαίδευση, τον αριθμό των εικόνων για αξιολόγηση και την ακρίβεια των δεδομένων εκπαίδευσης και δοκιμής.

No	Study	Architecture	Preprocessing details in input	Performance measure	Cases (individuals)	No of pixels in input	Images for training	Images for evaluation	Accuracy training data	Accuracy test data
1	Devi et al., 2019 **Error! Reference source not found.**	CNN	ReLU layer, max pooling, dropout and softmax layers, Xaviers method, grayscale images, batch size = 32	Train- Test Split, Cross entropy, Accuracy, Classification Error, MSE, MAE	9 participants	256x256	2371	264	93.24% (all levels)	94.15% (all levels)
2	Dragomir et al., 2020 **Error! Reference source not found.**	CNN- ResNet 18 network	pre-trained facial landmark detector, cropped pain related frames around the face, ReLU and Batch Normalization, Dense Layer, data augmentation	Train- Test Split, Accuracy, loss metrics	60 participatns (for train), 27 participants (for test)	300x200	9651	2885	-	36.6% (all levels)
3	Zheng et al., 2022 **Error! Reference source not found.**	CNN	Gray rectangle, face recognition, data augmentation, ReLU, batch normalization	Train- Test Split, loss metrics, repeat classification task 4 times	87 participants	64x64	6980 (pictures)	1740 (pictures)	-	52% (pain – no pain), 51.1% (all levels)
4	Lledo et al., 2023 **Error! Reference source not found.**	CNN	Subset was selected considering classes, genres and ages	Custom split, Accuracy, loss metrics	34 participants	Not mentioned	1150	390	-	20% (all levels) <50% (pain-no pain)
5	Othman et al., 2019 **Error! Reference source not found.**	RFc with FAD	OpenFace, Time series Statistic Descriptor, grey scale	5 fold cross validation, Accuracy	87 participants	96x96	2760 (images)	720 (images)	-	65.8% (pain- no pain)
		MobilNetV2	OpenFace, Time series Statistic Descriptor, grey scale	5 fold cross validation, Accuracy	87 participants	96x96	2760 (images)	720 (images)	-	65.5% (pain- no pain)

CNN: Convolutional Networks, RNN: Recurrent Networks, ViViT: Vision Transformer, RF: Random Forest, RFc: Random Forest classifier, FAD: Facial Activity Descriptor, GNN: Graph Neural Network, ICP: Iterative Closest Point

2.2 Θεωρητικό Υπόβαθρο στην Ανάλυση Βίντεο

Η αυτοματοποίηση της ανίχνευσης πόνου μέσω εκφράσεων του προσώπου έχει προσελκύσει σημαντική προσοχή τα τελευταία χρόνια. Ο Lledo et al. [4] χρησιμοποίησε την αρχιτεκτονική Video Transformers, συγκεκριμένα το ViViT, για την αναγνώριση εκφράσεων πόνου. Το μοντέλο τους στόχευε να καταγράψει το επίπεδο πόνου σε διάφορες κατηγορίες, είδη και ηλικιακές ομάδες, ενσωματώνοντας ένα μεγάλο σύνολο δεδομένων βίντεο για εκπαίδευση και αξιολόγηση. Παρ' όλα αυτά, η μελέτη ανέφερε σχετικά χαμηλή ακρίβεια δοκιμής 30,07%, υποδεικνύοντας τις προκλήσεις που εξακολουθούν να υπάρχουν κατά τη χρήση αρχιτεκτονικών βασισμένων σε transformers για την ανίχνευση πόνου από εκφράσεις προσώπου. Μια πιο παραδοσιακή προσέγγιση χρησιμοποιήθηκε από τον Werner et al. [5], οι οποίοι χρησιμοποίησαν Random Forest (RF) classifiers σε συνδυασμό με εργαλεία ανίχνευσης προσώπου, όπως οι αλγόριθμοι IntraFace και Iterative Closest Point (ICP). Αυτή η μελέτη, που περιλάμβανε 87 συμμετέχοντες και 8.700 εκπαιδευτικά βίντεο, ανέφερε ακρίβεια 70% για τη διάκριση μεταξύ διαφορετικών επιπέδων πόνου, παρέχοντας ένα ισχυρό σημείο αναφοράς για τα πρώιμα αυτοματοποιημένα συστήματα ανίχνευσης πόνου. Επεκτείνοντας αυτήν την προσέγγιση, οι Werner et al. [6] ενσωμάτωσαν ανιχνευτές χαρακτηριστικών τύπου Haar-like, επιτυγχάνοντας βελτιωμένη ακρίβεια 72,4% σε όλα τα επίπεδα πόνου. Αυτές οι μελέτες καταδεικνύουν την αποτελεσματικότητα των RF classifiers όταν συνδυάζονται με ισχυρές μεθόδους εξαγωγής χαρακτηριστικών προσώπου.

Σε πιο πρόσφατες μελέτες, οι Gigkas et al. [3] και Gigkas et al. [7] διερεύνησαν τη χρήση Transformer Neural Networks (TNT) για την ανίχνευση πόνου. Τα μοντέλα τους χρησιμοποίησαν μεθόδους ανίχνευσης σημείων αναφοράς προσώπου, όπως τα MTCNN και FAN, για την εξαγωγή σχετικών χαρακτηριστικών. Οι Gigkas et al. [3] χρησιμοποίησαν Leave-One-Subject-Out (LOSO) cross-validation και ανέφεραν ακρίβεια 73,28% για την ανίχνευση πόνου έναντι μη πόνου, αλλά μόνο 31,52% για όλα τα επίπεδα πόνου. Στη συνέχεια της μελέτης τους [7], οι συγγραφείς ενσωμάτωσαν ένα ευρύτερο φάσμα συνόλων δεδομένων, όπως τα VGGFace2, AffectNet και RAF-DB, οδηγώντας σε αξιοσημείωτη βελτίωση της απόδοσης, με ακρίβεια 77,10% για την ταξινόμηση πόνου έναντι μη πόνου και 35,39% για την αναγνώριση διαφόρων επιπέδων πόνου. Οι Tavakolian et al. [8] πρότειναν μια νέα προσέγγιση χρησιμοποιώντας

Spatiotemporal Convolutional Networks (SCN), συνδυάζοντας 2D και 3D CNNs για να καταγράψουν τη χρονική δυναμική των εκφράσεων του προσώπου. Το μοντέλο τους, που εκπαιδεύτηκε στο σύνολο δεδομένων CASIA WebFace, πέτυχε εντυπωσιακή ακρίβεια 86,02% για τη διάκριση πόνου από μη πόνο, υποδεικνύοντας την αποτελεσματικότητα των χωροχρονικών χαρακτηριστικών στην αναγνώριση εκφράσεων πόνου με την πάροδο του χρόνου. Οι Xin et al. [9] χρησιμοποίησαν ένα Attention-based Autoencoder (AE-ATT), αξιοποιώντας τον μηχανισμό Landmark-based Attention Mechanism (LAM) και τον μηχανισμό Identity Attention Mechanism (IAM). Αυτή η προσέγγιση οδήγησε σε ακρίβεια 85,65% για την ταξινόμηση πόνου έναντι μη πόνου, αλλά μόνο 40,4% για όλα τα επίπεδα πόνου. Η μελέτη υπογραμμίζει την πρόκληση της μοντελοποίησης διαφορετικών εντάσεων πόνου, καθώς μικρές αλλαγές στις εκφράσεις του προσώπου μπορούν να επηρεάσουν σημαντικά την ικανότητα του μοντέλου να ανιχνεύσει τη σοβαρότητα του πόνου.

Οι Patania et al. [10] συνέβαλαν στον τομέα εφαρμόζοντας Graph Convolutional Neural Networks (GCNN) και Dynamic Graph Convolutional Neural Networks (DGCNN), ενσωματώνοντας σημεία αναφοράς προσώπου και SortPooling layers για τη βελτίωση της εξαγωγής χαρακτηριστικών. Τα μοντέλα τους πέτυχαν ακρίβεια 73,2% χρησιμοποιώντας GCNN και 83,4% με DGCNN για τη διάκριση εκφράσεων πόνου από μη πόνο. Αυτή η έρευνα αναδεικνύει το δυναμικό των νευρωνικών δικτύων βασισμένων σε γραφήματα στη μοντελοποίηση των σύνθετων χωρικών σχέσεων μεταξύ των σημείων αναφοράς του προσώπου για πιο ακριβή ανίχνευση πόνου. Παρόλο που αυτές οι μελέτες δείχνουν σημαντική πρόοδο στην αυτοματοποιημένη ανίχνευση πόνου, τα αποτελέσματα παρουσιάζουν σημαντική μεταβλητότητα στην απόδοση μεταξύ διαφορετικών συνόλων δεδομένων, τεχνικών προεπεξεργασίας και αρχιτεκτονικών μοντέλων που χρησιμοποιούνται.

Πίνακας 2.1.2 : 1

No	Study	No of participants and videos Train/Test	Preprocessing details in input	Performance measure	No of pixels in input	Accuracy training set	Accuracy test set
1	Lledo et al., 2023	Not mentioned 4600/1600	subset was selected considering different classes, genres and ages	Accuracy, Loss, Training Loss	Not mentioned	Not mentioned	30.07%
2	Werner et al., 2014	87 8700/ random sample	IntraFace, ICP	10-fold validation, LOOCV validation	180x180	70% (all levels of pain)	
3	Werner et al., 2017	87 174000/ random sample	IntraFace, Haar-like feature detector cascade, ICP	10-fold validation	180x180	72.4% (all levels of pain)	
4	Gigkas et al., 2023	87 8700/ not mentioned	MTCNN, FAN	LOSO validation, F1 score, Accuracy, Precision, Recall	224x224	73.28% (pain – no pain), 31.52% (all levels of pain)	
5	Gigkas et al., 2024	87 8700/ not mentioned	MTCNN, AugmNet, VGGFace2, AffectNet, RAF-DB basic, RAF-DB compound	LOSO validation, F1 score, Accuracy, Precision, Recall	224x224	77.10% (pain – no pain), 35.39% (all levels of pain)	
6	Tavakolian et al., 2019	87 8700/ random sample	CASIA WebFace	LOSO validation, accuracy	112x112	86.02% (pain – no pain)	
7	Xin et al., 2020	87 Not mentioned/ not mentioned	LAM, IAM	LOSO validation, PCC, ICC, MSE, MAE	112x112	85.65(pain- no pain) 40.4(all levels of pain)	
8	Patania et al., 2022	87 3480/3480	Face landmarks, SortPooling layer, Adam optimization algorithm	5-fold validation	64x64	73.2% (pain – no pain)	-
		178 videos Not mentioned/ not mentioned	Face landmarks, Adam optimization algorithm	5-fold validation	64x64	83.4% (pain-no pain)	-

CNN: Convolutional Networks, RNN: Recurrent Networks, ViViT: Vision Transformer, RF: Random Forest, RFc: Random Forest classifier, FAD: Facial Activity Descriptor, GNN: Graph Neural Network, ICP: Iterative Closest Point, SCN: Spatiotemporal Convolutional Network, PCC: Pearson correlation coefficient, ICC: Intraclass Correlation coefficient, MSE: Mean Square Error, MAE: Mean Absolute error, LAM: Locality Aware Module, IAM: Identity aware module, TNT: Transformer in Transformer, MTCNN: face detector, FAN: Face Alignment Network

Κεφάλαιο 3

Μεθοδολογία και Υλικά

3.1 Βάση Δεδομένων BioVid Heat Pain

Η βάση δεδομένων BioVid Heat Pain [48] αποτελεί έναν σημαντικό πόρο για την αυτοματοποιημένη αξιολόγηση του πόνου, προσφέροντας μια πλούσια συλλογή δεδομένων για τη μελέτη των αντιδράσεων σε θερμικά ερεθίσματα. Η βάση δεδομένων δημιουργήθηκε από το Πανεπιστήμιο Magdeburg και το Πανεπιστήμιο του Ulm στη Γερμανία, και είναι ευρέως αποδεκτή για τη χρήση της στην κοινότητα της μηχανικής μάθησης και της ιατρικής εικόνας.

Συλλογή Δεδομένων

Το σετ δεδομένων περιλαμβάνει βιντεοσκοπήσεις από 90 υγιείς ενήλικες, ηλικίας 20 έως 65 ετών. Οι συμμετέχοντες υποβλήθηκαν σε δοκιμές με τέσσερα διαφορετικά

επίπεδα έντασης πόνου, εφαρμόζοντας τα θερμικά ερεθίσματα είκοσι φορές σε τυχαία σειρά για κάθε ένταση.

Δομή Δεδομένων

Το Μέρος Α της βάσης, το οποίο χρησιμοποιήθηκε για αυτή την έρευνα, περιέχει συγκεκριμένα δεδομένα από 87 συμμετέχοντες, με κάθε έναν να παρέχει 20 βίντεο χωρίς πόνο (BL1) και τέσσερα σετ των 20 βίντεο υψηλής έντασης πόνου (PA1-PA4). Κάθε βίντεο διαρκεί 5.5 δευτερόλεπτα και παρουσιάζει τις αντιδράσεις των συμμετεχόντων πριν, κατά τη διάρκεια και μετά την εφαρμογή των θερμικών ερεθισμάτων.

Τα βίντεο αυτά είναι συγκεκριμένα σχεδιασμένα για την ανάλυση της προσωπικής αντίδρασης στον πόνο, παρέχοντας υψηλής ανάλυσης εικόνες που είναι κρίσιμες για την ανάπτυξη αλγορίθμων που ανιχνεύουν και αξιολογούν την ένταση του πόνου με βάση τις προσωπικές εκφράσεις.

Η χρήση της BioVid Heat Pain βάσης δεδομένων προσφέρει έναν πλούσιο πόρο για την καλύτερη κατανόηση των αντιδράσεων στον πόνο και για την ανάπτυξη πιο εξελιγμένων και ακριβών μεθόδων αξιολόγησης.

3.2 Ταξινομητές

3.2.1 Ταξινομητές Εικόνας

Η ανάλυση εικόνων αποτελεί κρίσιμο στοιχείο στην αυτόματη αξιολόγηση του πόνου, καθώς η ακριβής ερμηνεία των εκφράσεων του προσώπου μπορεί να προσφέρει άμεσες ενδείξεις για την ένταση του πόνου. Για την επεξεργασία εικόνων στην παρούσα διπλωματική εργασία, χρησιμοποιήθηκαν δύο κύρια μοντέλα: το VGG16 και το MobileViT.

VGG16 [11]: Πρόκειται για ένα μοντέλο Συνελικτικού Νευρωνικού Δικτύου (CNN), το οποίο είναι ευρέως γνωστό για την υψηλή του απόδοση στην ταξινόμηση εικόνων. Χρησιμοποιείται για την ανίχνευση χαρακτηριστικών από τις εικόνες προσώπου και

έχει αποδειχθεί ιδιαίτερα αποτελεσματικό στην εξαγωγή βαθιάς γνώσης από τα δεδομένα.

MobileViT [12]: Αυτό το μοντέλο ανήκει στην κατηγορία των οπτικών μετασχηματιστών (vision transformers), συνδυάζοντας τις τεχνολογίες CNN και Transformer για να αξιοποιήσει τα πλεονεκτήματα και των δύο. Το MobileViT είναι ιδανικό για εφαρμογές όπου απαιτείται μειωμένη υπολογιστική ισχύς σε σύγκριση με τα παραδοσιακά μοντέλα Transformer, ενώ διατηρεί εξαιρετική ακρίβεια στην ταξινόμηση εικόνων.

Η επιλογή αυτών των μοντέλων έγινε με στόχο την εκμετάλλευση των δυνατοτήτων τους στην ανίχνευση και ταξινόμηση των εκφράσεων πόνου από τις εικόνες που λαμβάνονται από τη βάση δεδομένων BioVid Heat Pain. Τα αποτελέσματα της εφαρμογής τους αναλύονται περαιτέρω στα επόμενα υποκεφάλαια και στη συζήτηση, υπογραμμίζοντας τη συμβολή τους στην ακριβέστερη αξιολόγηση του πόνου.

3.2.2 Ταξινομητές Βίντεο

Η ανάλυση βίντεο για την αυτόματη αξιολόγηση της έντασης του πόνου βασίζεται στην επεξεργασία και κατανόηση της χρονικής συνέχειας των εκφράσεων πόνου. Στο παρόν υποκεφάλαιο, εξετάζονται τρία προηγμένα μοντέλα για την ταξινόμηση των βιντεογραφημένων εκφράσεων πόνου, καθένα από τα οποία εξειδικεύεται στην αναγνώριση και ερμηνεία της χρονικής διάρκειας των εκφράσεων. Τα μοντέλα που εφαρμόζονται είναι το Temporal Segment Network (TSN), το Convolutional Long Short-Term Memory (ConvLSTM), και το Video Masked Autoencoder (VideoMAE). Αυτά τα μοντέλα επιτρέπουν την ακριβή καταγραφή και ανάλυση των χρονικών δυναμικών του πόνου, ενισχύοντας την αξιοπιστία της αξιολόγησης πόνου μέσα από βίντεο.

Temporal Segment Network (TSN) [17]: Το TSN επεξεργάζεται τα βίντεο χωρίζοντάς τα σε κομμάτια και αναλύοντας την πληροφορία κάθε τμήματος για να κατανοήσει τη χρονική δυναμική της έκφρασης πόνου. Χρησιμοποιεί την αρχιτεκτονική ResNet ως βάση για την εξαγωγή χαρακτηριστικών από τα βίντεο.

Convolutional Long Short-Term Memory (ConvLSTM) [16]: Το ConvLSTM ενσωματώνει τη δυνατότητα της μνήμης των LSTM μονάδων σε συνελικτικά δίκτυα για να αναλύσει τα βίντεο αναγνωρίζοντας χωρικά και χρονικά μοτίβα.

Video Masked Autoencoder (VideoMAE) [18]: Το VideoMAE χρησιμοποιεί μια τεχνική autoencoding με μάσκες για να εκμαθήσει αναπαραστάσεις βίντεο. Τα βίντεο είναι μερικώς μασκαρεμένα κατά τη διάρκεια της εκπαίδευσης, επιτρέποντας στο μοντέλο να μαθαίνει αποτελεσματικά αναπαραστάσεις από τις οπτικές πληροφορίες που λείπουν.

Αυτά τα μοντέλα έχουν αποδειχθεί αποτελεσματικά για την αναγνώριση και ταξινόμηση βίντεο με βάση την έκφραση πόνου, παρέχοντας σημαντικές πληροφορίες για τη βελτίωση των κλινικών διαδικασιών αξιολόγησης του πόνου.

3.3 Εργαλεία και Βιβλιοθήκες Επεξεργασίας

Η ανάπτυξη και η αξιολόγηση μοντέλων μηχανικής μάθησης απαιτεί τη χρήση ποικίλων εργαλείων και βιβλιοθηκών προγραμματισμού, τα οποία επιτρέπουν την αποδοτική επεξεργασία, ανάλυση και παρουσίαση δεδομένων. Σε αυτό το κεφάλαιο, θα παρουσιαστούν τα κύρια εργαλεία και βιβλιοθήκες που χρησιμοποιήθηκαν για την υλοποίηση και την αξιολόγηση των αλγορίθμων που αναπτύχθηκαν κατά τη διάρκεια αυτής της έρευνας.

Τα Εργαλεία και Τεχνολογίες Προγραμματισμού που χρησιμοποιήθηκαν ειναι:

Python (Έκδοση 3.11.7): Η Python είναι μία υψηλού επιπέδου, διασυνδεδεμένη και ερμηνευόμενη γλώσσα προγραμματισμού, ιδιαίτερα δημοφιλής για την ανάπτυξη αλγορίθμων μηχανικής μάθησης λόγω της εκτεταμένης υποστήριξης βιβλιοθηκών και του ευκολοχώρητου συντακτικού της.

TensorFlow: Αυτή η βιβλιοθήκη της Google παρέχει έναν εκτενή πυρήνα για τη δημιουργία και εκπαίδευση νευρωνικών δικτύων, υποστηρίζοντας τόσο εκπαίδευση σε CPU όσο και σε GPU. Περιλαμβάνει εργαλεία για τον ορισμό κρίσιμων επιπέδων δικτύων, όπως συνελικτικά επίπεδα, επίπεδα συγκέντρωσης και πλήρως συνδεδεμένα επίπεδα.

OpenCV (cv2): Προσφέρει ισχυρές λειτουργίες για την επεξεργασία εικόνων και βίντεο, όπως φόρτωση, μετατροπή, φιλτράρισμα και ανίχνευση αντικειμένων. Χρησιμοποιείται ευρέως για την ανίχνευση και την περικοπή προσώπων σε εικόνες.

NumPy: Είναι θεμελιώδης για την αναπαράσταση και την επεξεργασία εικόνων σε μορφή πινάκων. Η NumPy βοηθά στην αποτελεσματική διαχείριση και μετατροπή δεδομένων σε επίπεδο πίνακα, κρίσιμο στοιχείο για την προετοιμασία και την ανάλυση δεδομένων.

Matplotlib και Seaborn: Αυτές οι βιβλιοθήκες χρησιμοποιούνται για την οπτικοποίηση δεδομένων και μετρικών απόδοσης του μοντέλου. Παρέχουν εκτενείς δυνατότητες για την απεικόνιση της συμπεριφοράς και των χαρακτηριστικών των μοντέλων, βοηθώντας στην αναλυτική κατανόηση των αποτελεσμάτων.

Scikit-learn (sklearn): Παρέχει εργαλεία για την αξιολόγηση του μοντέλου, όπως τη μέτρηση F1 score, την καμπύλη ROC και τον πίνακα σύγχυσης. Είναι κρίσιμη για την αξιολόγηση της απόδοσης και την κατανόηση των αδυναμιών του συστήματος.

Η επιλογή και η χρήση αυτών των εργαλείων εξασφαλίζει την ορθή υλοποίηση των αλγορίθμων σε ένα περιβάλλον που υποστηρίζει GPU, διασφαλίζοντας την αποδοτικότητα και την ακρίβεια στην επεξεργασία και ανάλυση των δεδομένων.

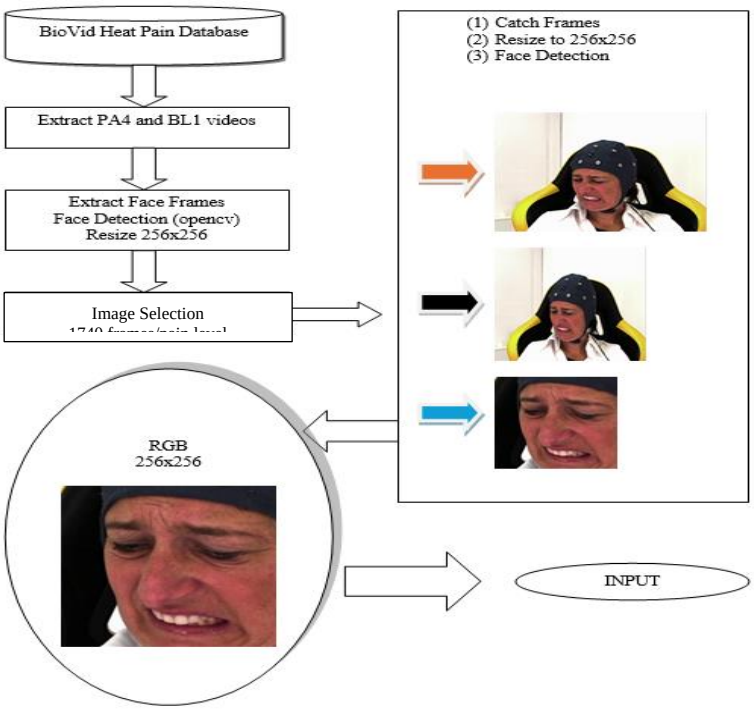
3.3.1 Προετοιμασία Δεδομένων

3.3.1.1 Προετοιμασία Δεδομένων για Εικόνες

Η προετοιμασία των εικόνων ήταν μια πολυεπίπεδη διαδικασία που ξεκίνησε με την εξαγωγή συνολικά 3,480 καρέ από τα βίντεο των 87 συμμετεχόντων. Κάθε συμμετέχων παρείχε 40 βιντεογραφήσεις, με 20 βίντεο για κάθε επίπεδο πόνου, με την κάθε καταγραφή να αντιστοιχεί στο επίπεδο πόνου BL1 (επίπεδο 0 πόνου) και PA4 (υψηλό επίπεδο πόνου). Από κάθε βίντεο επιλέγονταν από ένα ενδεικτικό καρέ, συνολικά 20 καρέ ανά επίπεδο πόνου για κάθε συμμετέχοντα, με το τελικό dataset να αποτελείται από 1,740 καρέ ανά επίπεδο πόνου.

Εξαγωγή και Περικοπή Προσώπων:

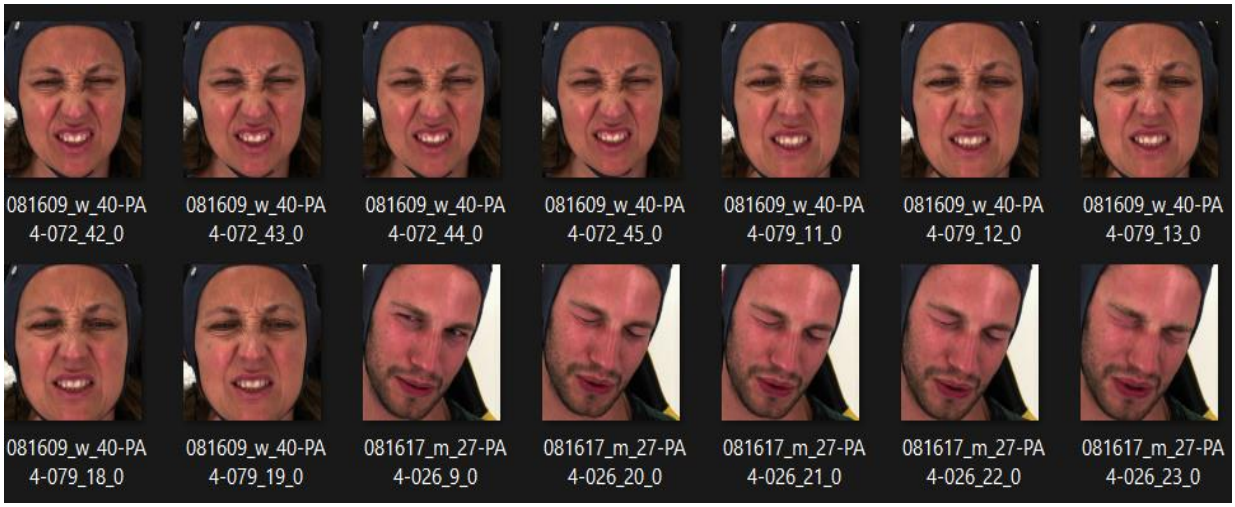
Για την ανίχνευση των προσώπων χρησιμοποιήθηκε ο αλγόριθμος Haarcascade Face Detection της OpenCV. Τα καρέ των βίντεο μετατράπηκαν πρώτα σε ασπρόμαυρες εικόνες για να ενισχύσουν την απόδοση του αλγορίθμου. Η μέθοδος detectMultiScale εφαρμόστηκε με παραμέτρους όπως scaleFactor=1.1, minNeighbors=5 και minSize=(30, 30) για την αποφυγή ψευδών ανιχνεύσεων. Τα ανιχνευμένα πρόσωπα περικόπηκαν και ενοποιήθηκαν σε διαστάσεις 256x256 pixels. Για την εξασφάλιση συμβατότητας με τα μοντέλα που απαιτούν είσοδο σε χρώμα RGB, τα ασπρόμαυρα καρέ μετατράπηκαν πίσω σε χώρο χρωμάτων RGB. Τα επεξεργασμένα καρέ αποθηκεύτηκαν ως αρχεία .png, χρησιμοποιώντας μια δομημένη ονοματολογία για να διατηρηθεί η οργάνωση ανά συμμετέχοντα και βίντεο.



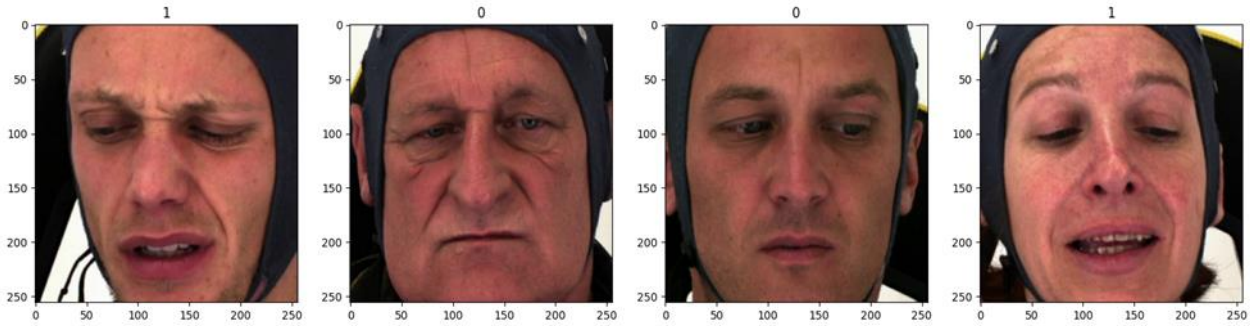
Εικόνα 3.3.1.1.1: Διάγραμμα ροής επεξεργασίας δεδομένων εικόνας.

Διαχωρισμός Δεδομένων:

Δημιουργήθηκαν δέκα επαναλήψεις σετ δεδομένων, όπου κάθε σετ δεδομένων περιελάμβανε τυχαία επιλεγμένους 70 συμμετέχοντες στο σύνολο εκπαίδευσης και επικύρωσης, και 17 συμμετέχοντες στο σύνολο δοκιμών. Η τυχαία ανάθεση συμμετεχόντων στα διάφορα σύνολα διασφάλισε τη διαφορετικότητα των δεδομένων στις δοκιμές, βελτιώνοντας τη γενίκευση και την ανθεκτικότητα των μοντέλων σε διάφορα σενάρια. Εντυπωσιακά, οι συμμετέχοντες στο σύνολο εκπαίδευσης αποκλείονται από το σύνολο δοκιμών, και το αντίστροφο, γεγονός που επιτρέπει τη δημιουργία ενός ισορροπημένου, μη μεροληπτικού και αντιπροσωπευτικού dataset για τα επίπεδα πόνου PA4 και BL1.



Εικόνα 3.3.1.1.2: Στιγμιότυπα από δεδομένα με επίπεδο πόνου PA4.

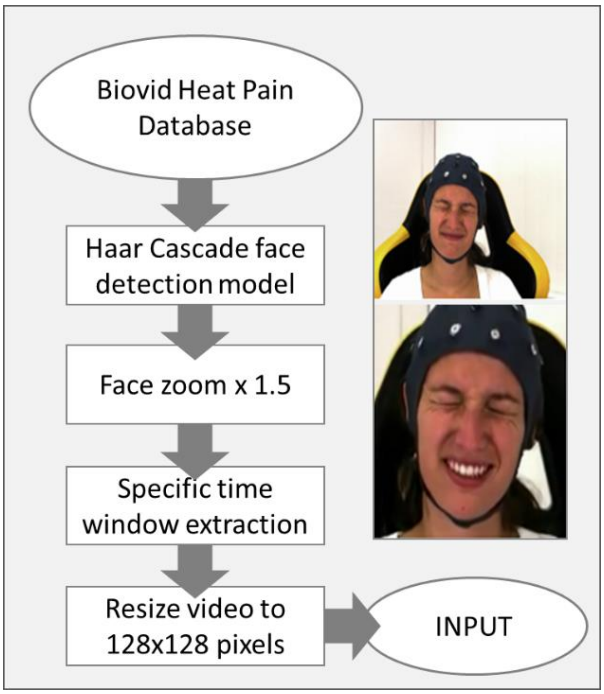


Εικόνα 3.3.1.1.3: Στιγμιότυπα σε γραφικό περιβάλλον χρησιμοποιώντας τη βιβλιοθήκη Matplotlib. Πάνω από κάθε εικόνα δίνεται μια ετικέτα (0 για BL1 και 1 για PA4).

3.3.1.2 Προετοιμασία Δεδομένων για Βίντεο

Η προετοιμασία των βίντεο ξεκίνησε με την επιλογή συγκεκριμένων χρονικών παραθύρων από κάθε δείγμα για την ανάλυση, με διάρκεια 2.5 δευτερόλεπτα, από το 2ο μέχρι και το 4.5 δευτερόλεπτο του αρχικού βίντεο. Η διαδικασία ξεκίνησε με την περικοπή και μεγέθυνση της περιοχής του προσώπου στο κάθε βίντεο κατά παράγοντα 1.5. Ακολούθως, τα βίντεο αναδιαμορφώθηκαν σε διαστάσεις 128x128 pixels για να εναρμονιστούν οι διαστάσεις εισόδου και να μειωθεί το φορτίο επεξεργασίας. Συνολικά, αποκλείστηκαν 24 συμμετέχοντες λόγω ανεπαρκούς ποικιλότητας των προσωπικών εκφράσεων.

Εξαγωγή και Περικοπή Προσώπων:
Για την επεξεργασία βίντεο, χρησιμοποιήθηκε η συνάρτηση VideoCapture της βιβλιοθήκης OpenCV, με την οποία διαβάστηκε κάθε βίντεο καρέ προς καρέ. Για την ανίχνευση των προσώπων, τα καρέ μετατράπηκαν πρώτα σε ασπρόμαυρο μορφότυπο και στη συνέχεια εφαρμόστηκε ο αλγόριθμος Haarcascade Face Detection για την αναγνώριση των προσωπικών χαρακτηριστικών. Τα ανιχνευμένα πρόσωπα περικόπηκαν και μορφοποιήθηκαν σε διαστάσεις 128x128 pixels.



Εικόνα 3.3.1.2.1: Διάγραμμα ροής επεξεργασίας δεδομένων βίντεο.

Σημαντικό είναι το γεγονός ότι για το μοντέλο VideoMAE εγινε μετατροπή στο dataset. Με τα βίντεο να αναπροσαρμόζονται σε διαστάσεις 224x224 pixels, κατάλληλες για τις ανάγκες του συγκεκριμένου μοντέλου.

Διαχωρισμός Δεδομένων:
Το τελικό dataset περιείχε 63 συμμετέχοντες, από τους οποίους το 79% (50 συμμετέχοντες) χρησιμοποιήθηκε για εκπαίδευση και το 21% (13 συμμετέχοντες) για δοκιμές, αντιστοιχώντας σε 2000 και 520 βίντεο αντίστοιχα (με 40 βίντεο ανά συμμετέχοντα, 20 για κάθε επίπεδο πόνου). Για την αξιολόγηση των μοντέλων, δημιουργήθηκαν δέκα διαφορετικά σύνολα δεδομένων, με τα ίδια δεδομένα αλλά με διαφορετική τυχαία κατανομή των συμμετεχόντων στα δεδομένα εκπαίδευσης και στα δεδομένα δοκιμών.

3.4 Πόροι και Υποδομή

Για τη διασφάλιση ισχυρών και συνεπών αποτελεσμάτων σε όλες τις δοκιμαστικές εκτελέσεις, οι υπολογιστικοί πόροι ήταν υψηλών επιδόσεων. Ο υλικός εξοπλισμός περιλάμβανε επεξεργαστή AMD EPYC 7313 16 πυρήνων, από τη σειρά 3ης γενιάς AMD EPYC™ (Zen 3, Milan), με συχνότητα λειτουργίας 3.0 GHz και 16 πυρήνες ανά εργασία. Σε συνδυασμό με 4 TB μνήμης RAM DDR4, αυτός ο επεξεργαστής παρείχε μια σταθερή βάση για την επεξεργασία δεδομένων. Ο σύστημα ήταν εξοπλισμένο με GPU NVIDIA RTX A5000, κάθε μία διαθέτοντας 8,192 πυρήνες CUDA, 64 πυρήνες RT δεύτερης γενιάς, και 256 πυρήνες Tensor τρίτης γενιάς, με ταχύτητες ρολογιού από 1.17 έως 1.66 GHz. Επτά τέτοιες GPU αναπτύχθηκαν ανά εργασία, προσφέροντας τη σημαντική υπολογιστική δύναμη που απαιτείται για την αποτελεσματική εκπαίδευση και αξιολόγηση βαθιάς μάθησης, διατηρώντας έτσι τη συνέπεια και την ακρίβεια διαφόρων πειραμάτων.

Hardware Specifications	
Processor	AMD EPYC 7313 16-Core Processor
Number of CPUs per task	20
Number of GPUs per task	7
Generation	3rd Gen AMD EPYC™, (Zen 3 (Milan))
Num. Cores (CPU)	16 cores
Num. Cores (GPU)	64 Second-generation RT Cores 256 Third-generation Tensor Cores
Frequency	3.0 GHz
RAM	4 TB DDR4
GPU	NVIDIA RTX A5000
Multiprocessor Count	-
CUDA Cores	8,192 CUDA cores
Clock Rate	1.17 – 1.66 GHz (A5000)

Πίνακας 3.4.1: Προδιαγραφές υλικού.

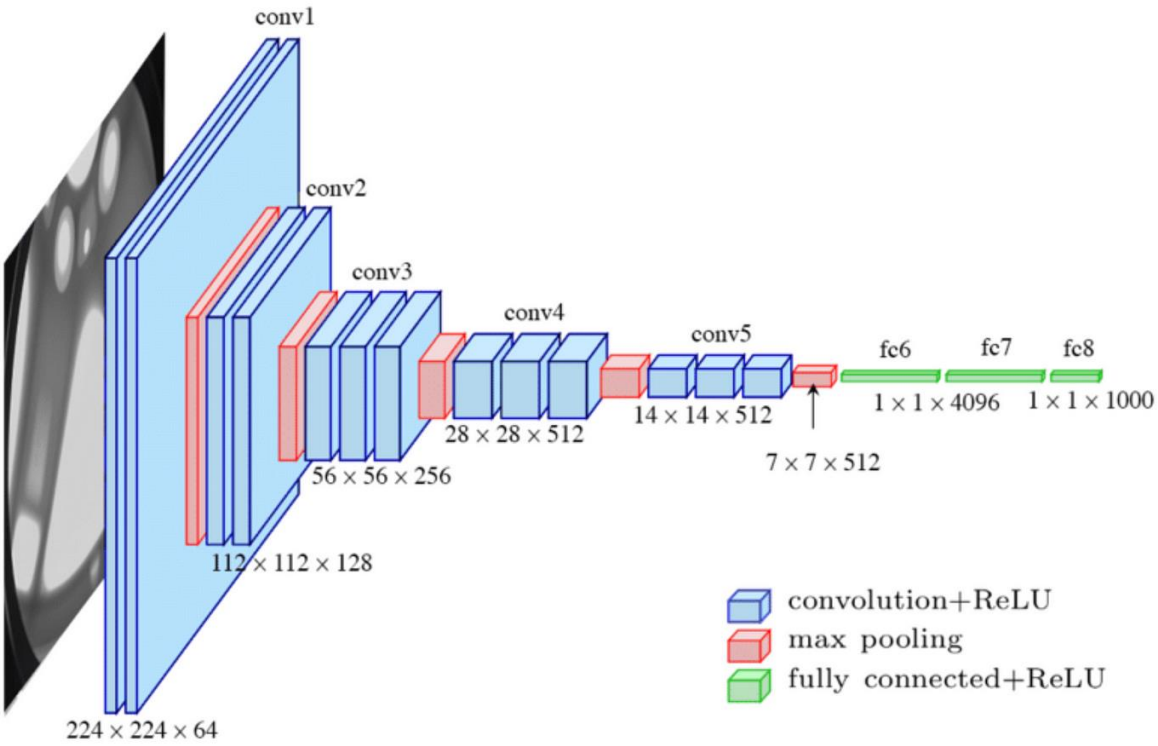
3.5 Αρχιτεκτονική

3.5.1 Ταξινομητές και Υπερπαράμετροι Εκπαίδευσης για Εικόνες

Στην παρούσα έρευνα, χρησιμοποιήθηκαν δύο σύγχρονα μοντέλα για την ταξινόμηση εικόνων με σκοπό την αξιολόγηση της έντασης πόνου: το VGG16 και το MobileViT. Αυτά τα μοντέλα επιλέχθηκαν για τις διακριτές αρχιτεκτονικές τους και την ικανότητά τους να διαχειρίζονται περίπλοκα προβλήματα ταξινόμησης εικόνων με υψηλή ακρίβεια.

VGG16:

Το VGG16, ένα βαθύ συνελικτικό νευρωνικό δίκτυο, χρησιμοποιεί μικρά συνελικτικά φίλτρα 3x3 για την ανίχνευση λεπτομερών χαρακτηριστικών από τις εικόνες, και είναι δομημένο με εναλλαγές συνελικτικών και max-pooling επιπέδων για την μείωση των διαστάσεων των χαρτών χαρακτηριστικών. Οι τελικές ταξινομήσεις πραγματοποιούνται μέσω πλήρως συνδεδεμένων επιπέδων.



Εικόνα 3.5.1.1: Αρχιτεκτονική VGG16.[20]

Το μοντέλο εκπαιδεύτηκε με τις εξής υπερπαραμέτρους:

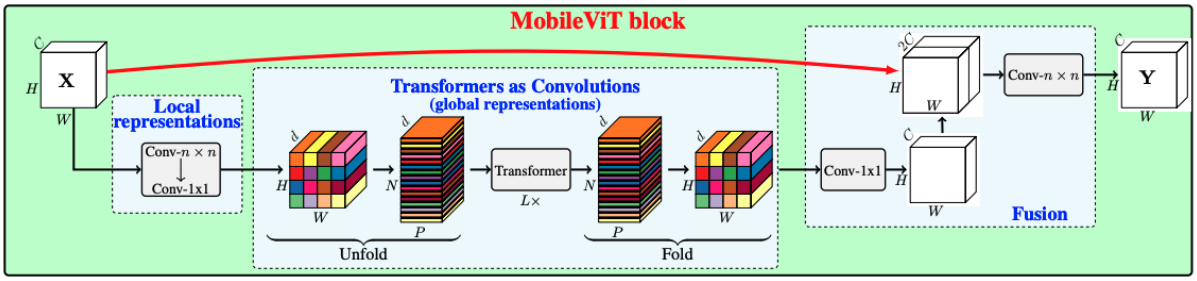
- **Ρυθμός μάθησης:** 0.001
- **Εποχές:** 50
- **Μέγεθος batch:** 32
- **Dropout:** 0.5 στο τελικό πλήρως συνδεδεμένο επίπεδο
- **Activation Function:** ReLU για τα κρυφά στρώματα και Sigmoid για το τελικό στρώμα εξόδου

Parameters for VGG16 architecture		
Network	**Parameter**	**Value**
	Layer 1	224 x 224 RGB
Convolutional	Layers Filters Filter Size Function	13 64, 128, 256, 512 3 ReLU
Max Pooling	Layers Pool Size	5 MaxPooling Layers 2 x 2
Flatten	Layers	1
Fully Connected	Dense Layer Units	2 Dense Layers, 1 Dropout Layer 256, 1
	Activation	ReLU (Dense Layer), Dropout 0.5
	Output Activation Function	Binary Sigmoid
	Solver	Adam

Πίνακας 3.5.1.2: Παράμετροι για το μοντέλο VGG16..

MobileViT:

Το MobileViT, ένα ελαφρύ μοντέλο, ενσωματώνει την αρχιτεκτονική των CNN με τα Vision Transformers, συνδυάζοντας συνελικτικά επίπεδα για την επεξεργασία τοπικών χαρακτηριστικών και Transformer επίπεδα για την επεξεργασία παγκόσμιων πληροφοριών. Το μοντέλο πετυχαίνει εξαιρετική ακρίβεια στην ταξινόμηση εικόνων και είναι κατάλληλο για εφαρμογές σε κινητές συσκευές λόγω της χαμηλής υπολογιστικής του απαιτήσεων.



Εικόνα 3.5.1.3: Αρχιτεκτονική MobileVit.[12]

Υπερπαράμετροι που χρησιμοποιήθηκαν:

- **Ρυθμός μάθησης:** 0.01
- **Εποχές:** 100
- **Μέγεθος batch:** 16
- **Activation Function:** Swish για τα κρυφά στρώματα και Sigmoid για το τελικό στρώμα εξόδου

Parameters for MobileViT architecture		
Network	**Parameter**	**Value**
	Layer 1	256 x 256 RGB
Convolutional	Layers	3
	Filters	16, 24, 48, 64, 80, 96,320
	Filter Size	3, 1
	Function	Swish
Inverted Residual Block	Layers	Multiple blocks with 16 or 80 filters
	Expansion Factor	2x
	Strides	1 or 2
MobileViT Block	Blocks	3 blocks with 64, 80, 96 projection dimensions
	Patch Size	4x4
	Number of Heads	2
Global Pooling	Pooling Type	Global Average Pooling
Fully Connected	Dense Layer	1
	Units	1
	Output	Binary
	Activation Function	Sigmoid
	Solver	Adam

Πίνακας 3.5.1.4: Παράμετροι για το μοντέλο MobileViT..

Και τα δύο μοντέλα έχουν εκπαιδευτεί με χρήση το σετ δεδομένων BioVid Heat Pain, με προσαρμογές στις υπερπαραμέτρους για την αποτελεσματική ταξινόμηση της έντασης του πόνου, παρέχοντας πολύτιμες πληροφορίες για την επίτευξη της ακριβέστερης αξιολόγησης του πόνου.

3.5.2 Ταξινομητές και Υπερπαράμετροι Εκπαίδευσης για Βίντεο

Η έρευνα χρησιμοποιεί τρία προηγμένα μοντέλα για την αξιολόγηση της έντασης του πόνου μέσω βίντεο: TSN, ConvLSTM και VideoMAE. Κάθε μοντέλο είναι σχεδιασμένο να εκμεταλλεύεται τις χρονικές και χωρικές πληροφορίες των βίντεο με διαφορετικό τρόπο.

TSN (Temporal Segment Network):

Το TSN χρησιμοποιεί το ResNet101 ως βάση, λαμβάνοντας είσοδο με μορφή RGB 20x128x128. Το μοντέλο αυτό είναι προ-εκπαιδευμένο στο ImageNet, επιτρέποντας την αξιοποίηση μαθημένων χαρακτηριστικών για αυξημένη ακρίβεια. Ενσωματώνει στρώματα όπως global average pooling και batch normalization, ενώ το dropout

εφαρμόζεται για αποφυγή υπερεκπαίδευσης. Η τελική ταξινόμηση γίνεται με ένα πλήρως συνδεδεμένο στρώμα με δυο εξόδους για δυαδική ταξινόμηση με χρήση softmax.

Το μοντέλο εκπαιδεύτηκε με τις εξής υπερπαραμέτρους:

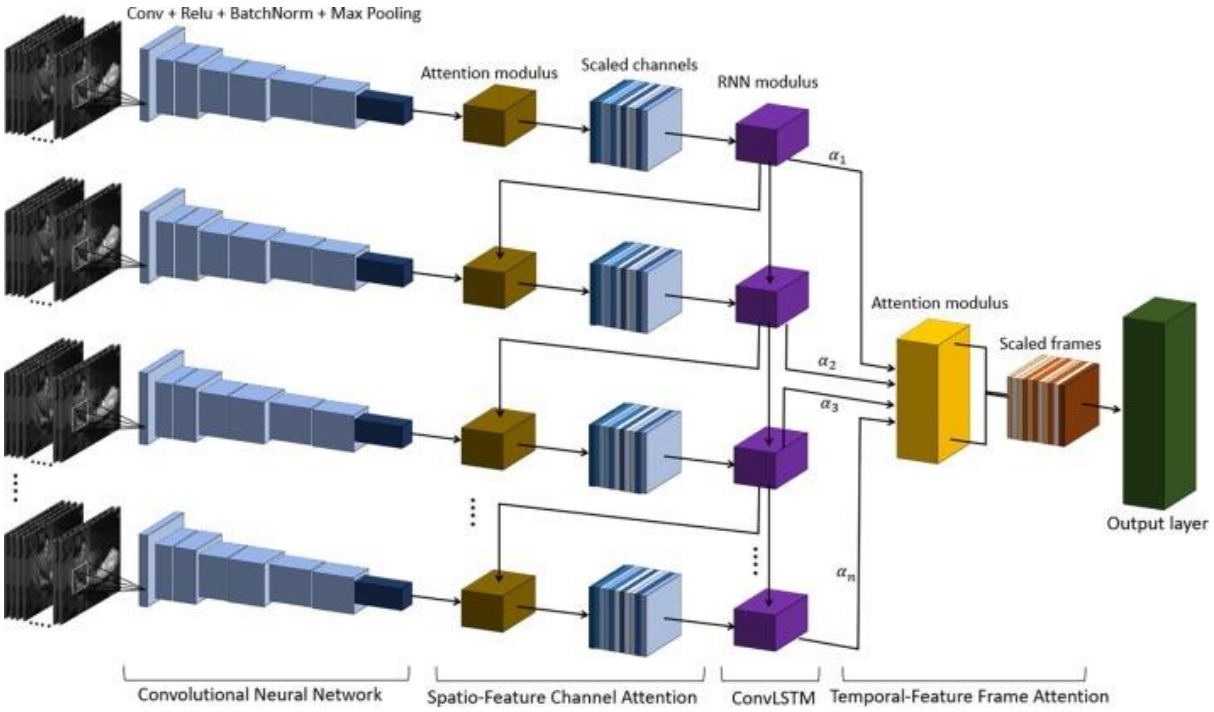
- Ρυθμός μάθησης: 5e-5
- Εποχές: 100
- Μέγεθος batch: 16
- Dropout: 0.5 στο τελικό πλήρως συνδεδεμένο επίπεδο
- Activation Function: ReLU για τα κρυφά στρώματα και softmax για το τελικό στρώμα εξόδου

Network	Parameter	Value
Layer 1	Input Shape	20x128x128 - RGB (20 Segments)
Base Network (ResNet101)	Layers	
	ResNet101 (Backbone)	Pre-trained on ImageNet, Input Shape: (128,128,3)
	Global Average Pooling Layer	Applied after ResNet101 outputs
	Batch Normalization Layer	Applied after Global Average Pooling
	Dropout Layer	Dropout (Rate: 0.5)
Aggregation	Layers	
	Concatenation	Concatenates outputs from 20 segments
Fully Connected Layer 1	Dense Layer	512 units, Activation: ReLU
	Dropout Layer	Dropout (Rate: 0.5)
	Batch Normalization	Applied after Dense Layer
Output Layer	Dense Layer	2 units (for binary classification)
Loss Function		Categorical Cross Entropy
Optimizer		Adam
Learning Rate		5e-5
Number of Segments		20

Πίνακας 3.5.2.1: Παράμετροι για το μοντέλο TSN.

ConvLSTM:

Το ConvLSTM δέχεται εισόδους 20x128x128 RGB καρέ και χρησιμοποιεί τρία επίπεδα ConvLSTM2D. Κάθε επίπεδο συνοδεύεται από batch normalization και χρονικά διανεμημένα στρώματα MaxPooling2D για τη μείωση των χωρικών διαστάσεων, διατηρώντας τις χρονικές πληροφορίες. Το τελικό στρώμα είναι πλήρως συνδεδεμένο με δυο εξόδους για δυαδική ταξινόμηση.



Εικόνα 3.5.2.2: Αρχιτεκτονική ConvLSTM [21].

Υπερπαράμετροι του μοντέλου περιλαμβάνουν:

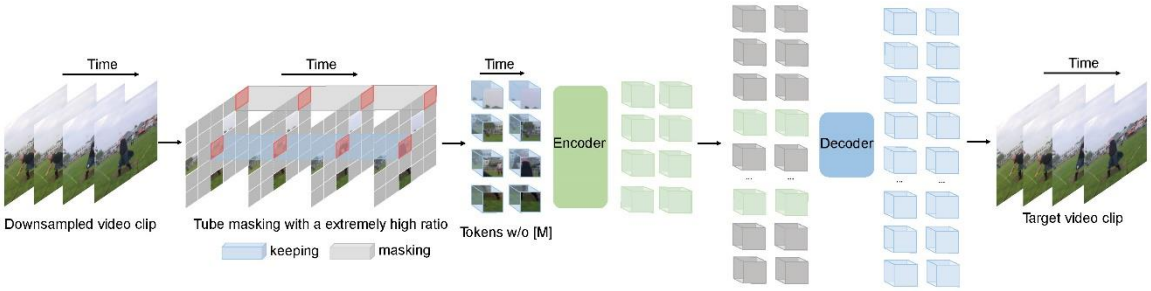
- Ρυθμός μάθησης: 1e-4, προσαρμόζεται δυναμικά βάσει της απόδοσης επικύρωσης
- Εποχές: 50
- Μέγεθος batch: 32
- Dropout: 0.5 στο τελικό πλήρως συνδεδεμένο επίπεδο
- Activation Function: ReLU για τα κρυφά στρώματα και softmax για το τελικό στρώμα εξόδου

Network	Parameter	Value
Layer 1	Input Shape	20x128x128 - RGB (Input Shape)
ConvLSTM	Layers	
	ConvLSTM2D	Filters – 64, Kernel Size: (3x3), Activation: RELU
Batch Normalization	Layer	Applied after ConvLSTM2D
Time Distributed Layer	Layer	Time Distributed MaxPooling2D (Pool Size: (2x2))
ConvLSTM Layer 2	ConvLSTM2D	Filters – 64, Kernel Size: (3x3), Activation: RELU
Batch Normalization	Layer	Applied after ConvLSTM2D
Time Distributed Layer	Layer	Time Distributed MaxPooling2D (Pool Size: (2x2))
ConvLSTM Layer 3	ConvLSTM2D	Filters – 128, Kernel Size: (3x3), Activation: RELU
Batch Normalization	Layer	Applied after ConvLSTM2D
Max Pooling	Layer	Pool Size: (2x2)
Flatten Layer	Layer	Flatten all layers
Fully Connected	Layer	Dense Layer (256 units), Activation: ReLU
Batch Normalization	Layer	Applied after Dense Layer
Dropout	Layer	Dropout (Rate: 0.5)
Output	Dense Layer	2 units (for binary classification)
Loss Function		Categorical Cross Entropy
Optimizer		Adam
Learning Rate		1e-4
Learning Rate Scheduler		ReduceLROnPlateau (Adjusts based on validation loss)

Πίνακας 3.5.2.3: Παράμετροι για το μοντέλο ConvLSTM.

VideoMAE:

VideoMAE χρησιμοποιεί βίντεο που έχουν προσαρμοστεί σε διαστάσεις 224x224 και χρησιμοποιεί μια προσέγγιση masked autoencoder, όπου τμήματα κάθε καρέ μασκαρεύονται κατά τη διάρκεια της εκπαίδευσης για την εκμάθηση σημαντικών αναπαραστάσεων μέσω της ανακατασκευής του μασκαρεμένου περιεχομένου. Το τελικό στρώμα προσφέρει δυαδική ταξινόμηση πόνου με χρήση της softmax.



Εικόνα 3.5.2.4: Αρχιτεκτονική VideoMAE.[19]

Υπερπαράμετροι που χρησιμοποιήθηκαν:

- Ρυθμός μάθησης: Αρχικός ρυθμός 1e-3 με προσαρμογή ανάλογα με την απόδοση επικύρωσης
- Εποχές: 70
- Μέγεθος batch: 16
- Activation Function: Softmax για το τελικό στρώμα εξόδου

Network	Parameter	Value
Layer 1	Input Shape	16x224x224 - RGB (Input Shape)
VideoMAE	Layers	Transformer-based Encoder; Masked Autoencoder Structure
	Frame Embedding	3D patch embedding (patch size: 16x16x16)
Attention Blocks	Number of Layers	12 Transformer blocks
	Number of Heads	12 attention heads per block
	Hidden Size	768
	MLP Layer Size	3072 (four times the hidden size)
	Activation	GELU
Positional Encoding	Type	Learned positional embeddings
Pooling Layer	Type	Global Average Pooling (for classification)
Fully Connected	Layers	Linear layer connected for classification
	Unit	2 (final classification layer for binary output)
Output	Type	Softmax Activation (for final binary classification)
Loss Function		CrossEntropyLoss for classification tasks
Solver	Optimizer	AdamW
Learning Rate	5e-5	

Πίνακας 3.5.2.3: Παράμετροι για το μοντέλο VideoMAE.

Κάθε ένα από αυτά τα μοντέλα εφαρμόζει καινοτόμες τεχνικές για την ανάλυση βίντεο, παρέχοντας έτσι πολύτιμες πληροφορίες για τη βελτιωμένη και ακριβέστερη αξιολόγηση του πόνου.

Κεφάλαιο 4

Αποτελέσματα

4.1 Εισαγωγή

Σε αυτό το κεφάλαιο παρουσιάζονται τα αποτελέσματα της αξιολόγησης των μοντέλων μηχανικής μάθησης που εκπαιδεύτηκαν για την αναγνώριση και ταξινόμηση εικόνων και βίντεο. Συγκεκριμένα, αναλύονται τα μοντέλα VGG16 και MobileViT για την επεξεργασία εικόνων, ενώ τα μοντέλα VideoMAE, TSN και ConvLSTM εστιάζουν στην ανάλυση βίντεο.

Τα μοντέλα για εικόνες, VGG16 και MobileViT, αξιοποιούνται για την ακριβή καταγραφή και ανάλυση επίπεδων πόνου μέσω της αποτελεσματικής επεξεργασίας στατικών εικόνων. Από την άλλη πλευρά, τα μοντέλα VideoMAE, TSN και ConvLSTM είναι σχεδιασμένα να εντοπίζουν και να ερμηνεύουν τις χρονικές συνέπειες και τις αλλαγές μεταξύ διαδοχικών καρέ, προσφέροντας έτσι βαθύτερη κατανόηση σε εφαρμογές όπου η δυναμική των γεγονότων είναι κρίσιμης σημασίας.

Στη συνέχεια, αναλύονται λεπτομερώς τα αποτελέσματα από τις δοκιμές των μοντέλων, εστιάζοντας σε κρίσιμες μετρικές όπως η ακρίβεια, η ανάκληση και το F1 Score. Εξετάζεται επίσης η συνέπεια και η σταθερότητα της απόδοσής τους σε διαφορετικά σετ δεδομένων, προκειμένου να εκτιμηθεί η πραγματική τους αξία στην πρακτική εφαρμογή, ενώ παράλληλα εξάγονται συμπεράσματα για τις προοπτικές τους στην περαιτέρω ανάπτυξη και χρήση.

4.2 Αξιολόγηση Απόδοσης Μοντέλων για Εικόνες

Στο πλαίσιο της αξιολόγησης μοντέλων για την επεξεργασία εικόνων, εξετάστηκαν τα μοντέλα VGG16 και MobileViT. Κάθε μοντέλο αξιολογήθηκε βάσει μετρικών όπως η ακρίβεια, η ανάκληση, η F1 βαθμολογία, και άλλες σχετικές δείκτες, παρέχοντας έτσι μια πλήρη εικόνα της απόδοσής τους σε εργασίες ταξινόμησης εικόνων.

Σημαντικό είναι να αναφερθεί ότι τα αποτελέσματα της αξιολόγησης προέρχονται από τη διαδικασία που περιγράφεται στο κεφάλαιο μεθοδολογίας. Ειδικότερα, δημιουργήθηκαν δέκα επαναλήψεις σετ δεδομένων, κάθε ένα περιλαμβάνοντας τυχαία επιλεγμένους 70 συμμετέχοντες στο σύνολο εκπαίδευσης και επικύρωσης και 17 συμμετέχοντες στο σύνολο δοκιμών. Για την αποφυγή προκατάληψης και τη διασφάλιση μιας δίκαιης αξιολόγησης, οι συμμετέχοντες τοποθετήθηκαν τυχαία σε κάθε σύνολο. Αυτή η στρατηγική τυχαίου δειγματοληψίας εξασφάλισε ότι τα μοντέλα δοκιμάστηκαν σε διάφορους συνδυασμούς συμμετεχόντων, ενισχύοντας τη γενίκευσή τους και την ανθεκτικότητα στην πρόβλεψη αποτελεσμάτων σε διάφορα σενάρια. Επιπλέον, οι συμμετέχοντες που εντάχθηκαν στο σύνολο εκπαίδευσης αποκλείστηκαν από το σύνολο δοκιμών και το αντίστροφο. Αυτή η προσέγγιση επιτρέπει τη δημιουργία ενός ισορροπημένου, αντιπροσωπευτικού και χωρίς προκαταλήψεις dataset για τα επίπεδα PA 4 και BL1.

Τα αποτελέσματα και οι γραφικές παραστάσεις που παρουσιάζονται στη συνέχεια αντιπροσωπεύουν τους μέσους όρους, καθώς και τα μέγιστα και ελάχιστα αποτελέσματα όπου αναφέρονται, από τα δέκα διαφορετικά σύνολα δεδομένων. Αυτή η προσέγγιση προσφέρει μια σταθερή και αξιόπιστη βάση για την αξιολόγηση της αποδοτικότητας των μοντέλων. Η επιλογή να παρουσιαστούν οι μέσοι όροι μαζί με τα άκρα δείχνει τη διακύμανση και τη σταθερότητα των μοντέλων σε διάφορες δοκιμαστικές συνθήκες, εγγυώμενη την αντικειμενικότητα και την αξιοπιστία των αποτελεσμάτων. Αυτό καθιστά τα συμπεράσματα πιο έγκυρα και αντιπροσωπευτικά για περαιτέρω αναλύσεις και συγκρίσεις.

VGG16 αποτελέσματα:

Το μοντέλο VGG16 επέτυχε precision 0,561, που σημαίνει ότι λίγο πάνω από το μισό των θετικών του προβλέψεων ήταν σωστές. Η ανάκληση κατέγραψε κατά μέσο όρο 0,531, αντανακλώντας την ικανότητα του μοντέλου να αναγνωρίζει σωστά περίπου το 53% των πραγματικών θετικών περιπτώσεων. Η συνολική ακρίβεια του μοντέλου ήταν 0,558, υποδηλώνοντας ότι ταξινόμησε σωστά περίπου το 56% όλων των περιπτώσεων. Όσον αφορά την ικανότητά του να διακρίνει διαφορετικές κλάσεις, το ROC AUC ήταν κατά μέσο όρο 0,582, δείχνοντας μέτρια διακριτική δύναμη. Το μέσο F1 Score ήταν 0,612, υποδηλώνοντας αξιοπρεπή απόδοση στη διαχείριση των συμβιβασμών μεταξύ ψευδών θετικών και ψευδών αρνητικών. Η μέση τιμή του Log Loss ήταν 0,756, υποδηλώνοντας σχετικά σταθερή εμπιστοσύνη στις προβλέψεις που πραγματοποιήθηκαν.

Επιπλέον, το τρέχον μοντέλο επέδειξε μεταβλητότητα στην απόδοσή του, όπως φαίνεται από το εύρος των μετρικών. Το precision κυμάνθηκε από ένα ελάχιστο του 0,521 έως ένα μέγιστο του 0,600, με τυπική απόκλιση 0,023, δείχνοντας σχετικά σταθερή προγνωστική ισχύ. Η ανάκληση έδειξε μεγαλύτερη μεταβλητότητα, με τιμές που κυμαίνονταν από 0,430 έως 0,633 και τυπική απόκλιση 0,063. Η ακρίβεια παρέμεινε σταθερή, με χαμηλή τυπική απόκλιση 0,016 και εύρος μεταξύ 0,525 και 0,582. Το ROC AUC είχε κάπως μεγαλύτερο εύρος, από 0,530 έως 0,645, με τυπική απόκλιση 0,039. Το F1 Score έδειξε μεγαλύτερη μεταβλητότητα με τυπική απόκλιση 0,225, κυμαινόμενο από 0,612 έως 1,285. Το Log Loss παρέμεινε σχετικά σταθερό, με μικρή τυπική απόκλιση 0,054.

Συνολικά, το μοντέλο VGG16 επέδειξε σταθερή απόδοση σε όλες τις μετρικές, με μέτρια μεταβλητότητα στον precision, την ανάκληση και το F1 Score. Η χαμηλή τυπική απόκλιση στην ακρίβεια και στο Log Loss υποδηλώνει ότι το μοντέλο διατήρησε συνεπή αποτελέσματα σε διάφορα σύνολα δεδομένων. Ωστόσο, η μεταβλητότητα στην ανάκληση και το F1 Score υποδηλώνει ότι το μοντέλο μπορεί να αντιμετωπίζει κάποιες δυσκολίες στην ισορροπία μεταξύ ψευδώς θετικών και ψευδώς αρνητικών αποτελεσμάτων σε ορισμένα σενάρια.

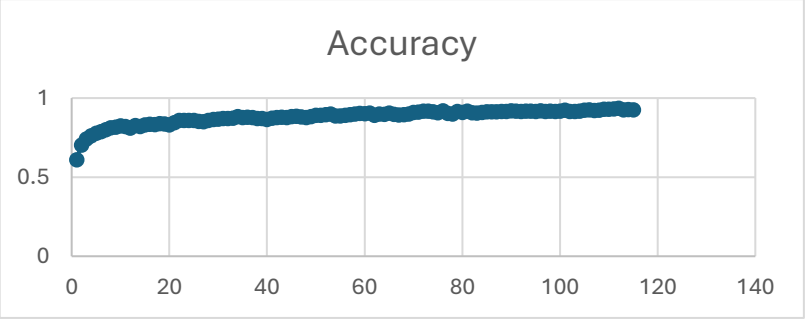
	Precision	Recall	Accuracy	ROC AUC	F1 Score	Log Loss	TP	FP	FN	TN
AVG	0.561	0.531	0.558	0.582	0.612	0.756	191.4	151.2	168.6	208.8
Max	0.6	0.633	0.582	0.645	1.285	0.849	228	199	205	249
Min	0.521	0.43	0.525	0.53	0.612	0.756	155	111	132	161
STDEV.P	0.023	0.063	0.016	0.039	0.225	0.054	22.91	28.44	22.91	28.44
TP: True Positives, FP: False Positives, FN: False Negatives, TN: True Negatives										

Πίνακας 4.2.1: Μετρικές απόδοσης για το μοντέλο VGG16 στα δεδομένα δοκιμών.

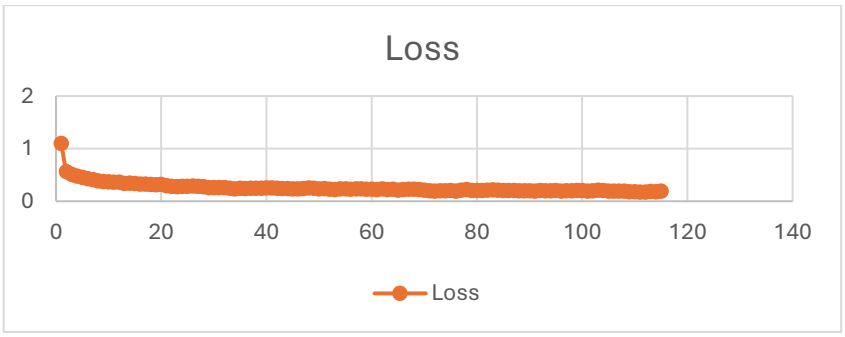
Το μοντέλο VGG16 εκπαιδεύτηκε για 115 εποχές χρησιμοποιώντας τον βελτιστοποιητή Adam, με στόχο την αξιολόγηση της δυναμικής και της τελικής απόδοσης της εκπαίδευσης. Ο μέσος χρόνος εκπαίδευσης ήταν περίπου 9,3 λεπτά. Στο τέλος της εκπαίδευσης, ο μέσος όρος του σφάλματος εκπαίδευσης ήταν 0.18814. Η ακρίβεια εκπαίδευσης στην τελευταία εποχή είχε μέσο όρο 0.92491, δείχνοντας την υψηλή απόδοση του μοντέλου στα δεδομένα εκπαίδευσης. Ωστόσο, η ακρίβεια δοκιμής/επικύρωσης του μοντέλου, η οποία είχε μέσο όρο 0.5558, ήταν χαμηλότερη, υποδηλώνοντας πιθανές προκλήσεις στη γενίκευση σε άγνωστα δεδομένα. Επιπλέον, η μέγιστη καταγεγραμμένη ακρίβεια εκπαίδευσης ήταν 0.9506, με αντίστοιχη ακρίβεια δοκιμής/επικύρωσης 0.5819, δείχνοντας την καλύτερη απόδοση του μοντέλου. Αντίθετα, η ελάχιστη ακρίβεια εκπαίδευσης ήταν 0.8909, με αντίστοιχη ακρίβεια δοκιμής/επικύρωσης 0.525. Αυτή η διακύμανση τονίζεται περαιτέρω από τις τυπικές αποκλίσεις, με το σφάλμα εκπαίδευσης να έχει τυπική απόκλιση 0.0385, την ακρίβεια εκπαίδευσης 0.016345, και την ακρίβεια δοκιμής/επικύρωσης 0.01685. Αυτά τα μετρικά παρέχουν μια λεπτομερή άποψη της διαδικασίας εκμάθησης του μοντέλου, τονίζοντας την ικανότητά του να επιτυγχάνει υψηλή ακρίβεια στα δεδομένα εκπαίδευσης ενώ ταυτόχρονα υποδεικνύει τομείς όπου η γενίκευση θα μπορούσε να βελτιωθεί.

	Epochs	Optimizer	Training Time (min)	Train Error of Last Epoch	Train Accuracy of Last Epoch	Test/Val Accuracy of Last Epoch
AVERAGE	115	Adam	9.3	0.18814	0.92491	0.5558
MAX	115	Adam	9.5	0.2433	0.9506	0.5819
MIN	115	Adam	9.2	0.1033	0.8909	0.525
STDEV.P	115	Adam	9.3	0.0385	0.016345	0.01685

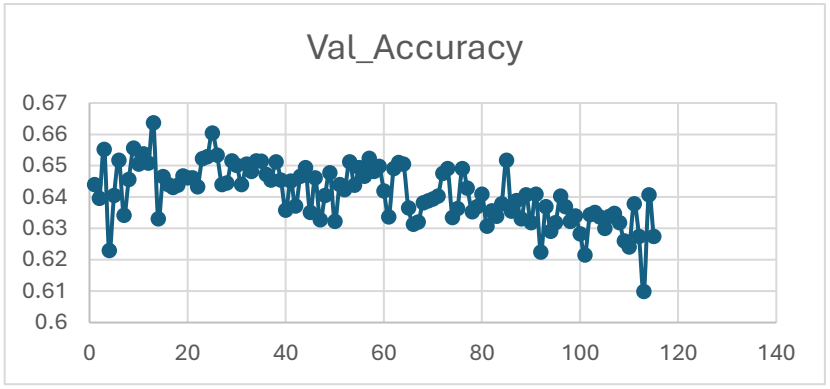
Πίνακας 4.2.2: Μετρικές εποχών για το μοντέλο VGG16.



Εικόνα 4.2.3: Γραφική Παράσταση ακρίβειας (Accuracy) κατά τη διάρκεια της εκπαίδευσης του μοντέλου VGG16.



Εικόνα 4.2.4: Γραφική Παράσταση απώλειας (Loss) κατά τη διάρκεια της εκπαίδευσης του μοντέλου VGG16.



Εικόνα 4.2.5: Γραφική Παράσταση ακρίβειας επικύρωσης (Validation Accuracy) κατά τη διάρκεια της εκπαίδευσης του μοντέλου VGG16.

MobileViT αποτελέσματα:

Το MobileViT πέτυχε Precision 0.494, δείχνοντας ότι δεν είναι αποδοτικό μοντέλο, καθώς λιγότερο από το μισό των θετικών του προβλέψεων ήταν σωστές. Η Ανάκληση είχε μέσο όρο 0.507, αντανακλώντας την ικανότητα του μοντέλου να ταυτοποιεί λίγο περισσότερο από το μισό των πραγματικών θετικών περιπτώσεων. Η γενική Ακρίβεια του μοντέλου ήταν 0.598, υποδηλώνοντας ότι κατάφερε να ταξινομήσει σωστά σχεδόν το 60% όλων των περιπτώσεων. Η μέση τιμή του ROC AUC ήταν 0.490. Η βαθμολογία F1, η οποία ισορροπεί τον Precision και την Ανάκληση, ήταν κατά μέσο όρο 0.579, δείχνοντας μέτρια απόδοση στη διαχείριση του αντιπαλότητας μεταξύ ψευδών θετικών και ψευδών αρνητικών. Η μέση τιμή του Log Loss ήταν 2.753, αναδεικνύοντας κάποια μεταβλητότητα στην εμπιστοσύνη των προβλέψεών του. Κατά μέσο όρο, η απόδοση του MobileViT έδειξε σημαντική διακύμανση μεταξύ διαφορετικών διαδρομών, όπως φαίνεται από τις τυπικές αποκλίσεις των μετρικών. Για παράδειγμα, το Precisionς είχε τυπική απόκλιση 0.064, ενώ η Ανάκληση διέφερε περισσότερο με τυπική απόκλιση

0.092. Η Ακρίβεια και το ROC AUC έδειξαν σχετικά σταθερή απόδοση με τυπικές αποκλίσεις 0.064 και 0.020 αντίστοιχα. Η βαθμολογία F1 είχε υψηλότερη τυπική απόκλιση 0.239, υποδηλώνοντας ασυνέπεια σε κάποιες περιπτώσεις. Οι μέσοι οροι Θετικών Αληθών ήταν 182.8 με τυπική απόκλιση 33.2, και οι μέσοι οροι Αληθών Αρνητικών ήταν 172.4 με τυπική απόκλιση 37.56. Αυτές οι μετρικές τονίζουν τα δυνατά και τα αδύναμα σημεία του μοντέλου MobileViT σε διάφορα σενάρια, προσφέροντας μια ολοκληρωμένη εικόνα της απόδοσής του.

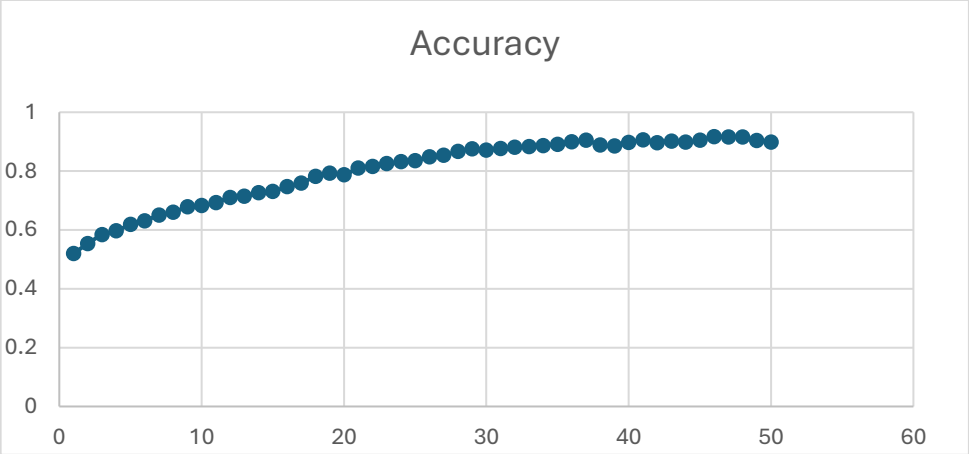
	Precision	**Recall**	**Accuracy**	**ROC AUC**	**F1 Score**	**Log Loss**	**TP**	**FP**	**FN**	**TN**
AVG	0.494	0.507	0.598	0.49	0.579	2.753	182.8	187.6	177.2	172.4
Max	0.519	0.629	0.723	0.531	1.285	4.562	226	239	228	235
Min	0.479	0.366	0.533	0.457	0.418	0.239	132	125	134	121
STDEV.P	0.064	0.092	0.064	0.02	0.239	1.44	33.2	37.5	33.2	37.56
TP: True Positives, FP: False Positives, FN: False Negatives, TN: True Negatives										

Πίνακας 4.2.6: Μετρικές απόδοσης για το μοντέλο MobileViT στα δεδομένα δοκιμών.

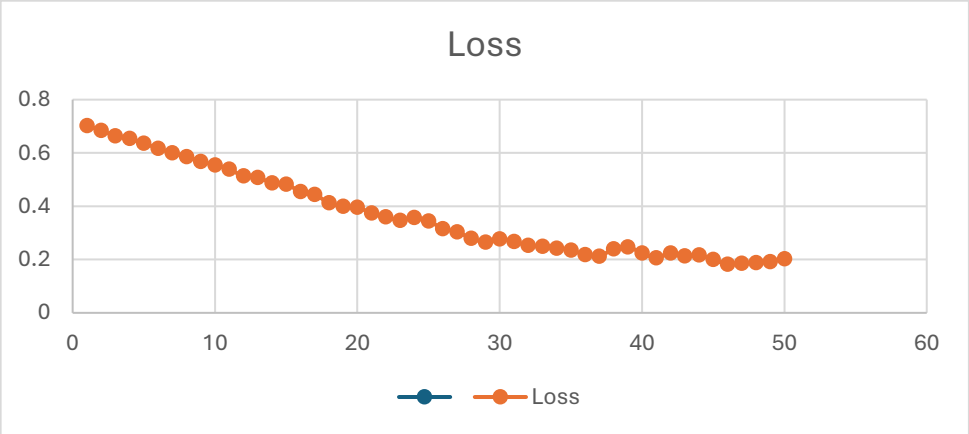
Πέρα από τις μετρικές απόδοσης, η απόδοση του μοντέλου MobileViT κατά τη διάρκεια των εποχών εκπαίδευσης αναλύεται επίσης. Το μοντέλο εκπαιδεύτηκε για 50 εποχές χρησιμοποιώντας τον βελτιστοποιητή Adam, με μέσο χρόνο εκπαίδευσης 10 λεπτά. Το μέσο σφάλμα εκπαίδευσης στην τελευταία εποχή ήταν 0.20292, αντανακλώντας το ποσοστό σφάλματος στο τελικό στάδιο της εκπαίδευσης. Η μέση ακρίβεια εκπαίδευσης της τελευταίας εποχής ήταν 0.8984, δείχνοντας ότι το μοντέλο απέδωσε καλά στα δεδομένα εκπαίδευσης. Ωστόσο, η μέση ακρίβεια δοκιμής/επικύρωσης, η οποία ήταν 0.59859, ήταν αισθητά χαμηλότερη, υποδηλώνοντας ότι το μοντέλο μπορεί να είχε κάποια υπερπροσαρμογή, αποδίδοντας καλύτερα στα δεδομένα εκπαίδευσης παρά σε άγνωστα δεδομένα. Οι μέγιστες τιμές που παρατηρήθηκαν κατά την εκπαίδευση περιελάμβαναν μια ακρίβεια εκπαίδευσης 0.9974 και μια ακρίβεια δοκιμής/επικύρωσης 0.7319, δείχνοντας τις δυνατότητες του μοντέλου υπό ιδανικές συνθήκες. Αντίθετα, η ελάχιστη καταγραφείσα ακρίβεια εκπαίδευσης ήταν 0.6215, με αντίστοιχη ακρίβεια δοκιμής/επικύρωσης 0.5333. Η τυπική απόκλιση της ακρίβειας εκπαίδευσης ήταν 0.1302, υποδηλώνοντας κάποια μεταβλητότητα στην ικανότητα του μοντέλου να γενικεύει σε διάφορες εποχές, ενώ η ακρίβεια δοκιμής/επικύρωσης είχε μια τυπική απόκλιση 0.0639, δείχνοντας σχετικά σταθερή απόδοση κατά την αξιολόγηση.

	Epochs	Optimizer	Training Time (min)	Train Error of Last Epoch	Train Accuracy of Last Epoch	Test/Val Accuracy of Last Epoch
AVERAGE	50	Adam	10	0.20292	0.8984	0.59859
MAX	50	Adam	10	0.6436	0.9974	0.7319
MIN	50	Adam	9.9	0.0105	0.6215	0.5333
STDEV.P	50	Adam	10	0.2360	0.1302	0.0639

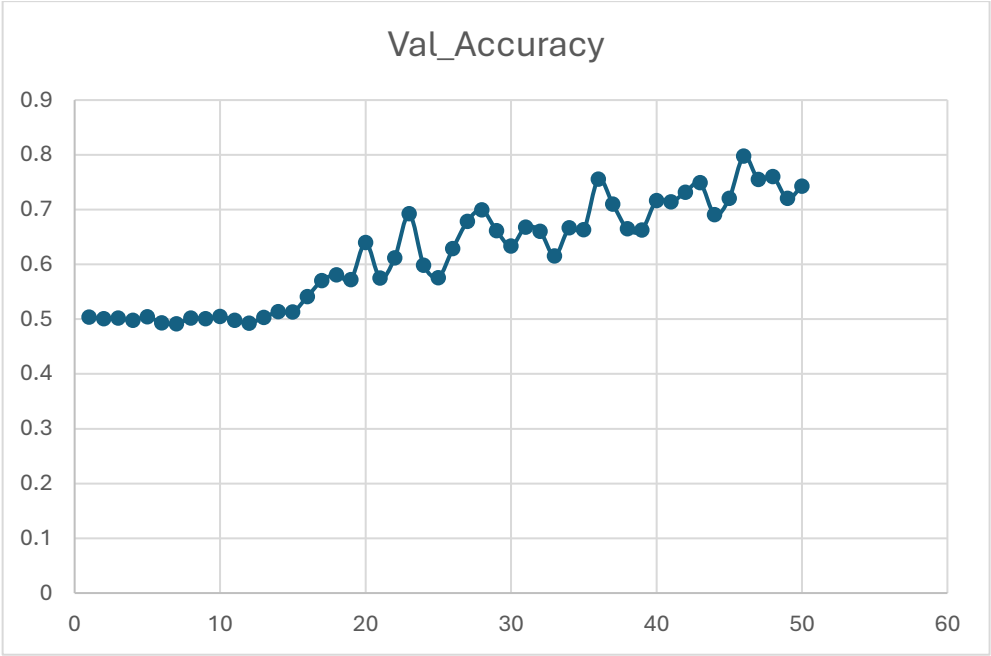
Πίνακας 4.2.7: Μετρικές εποχών για το μοντέλο MobileVit.



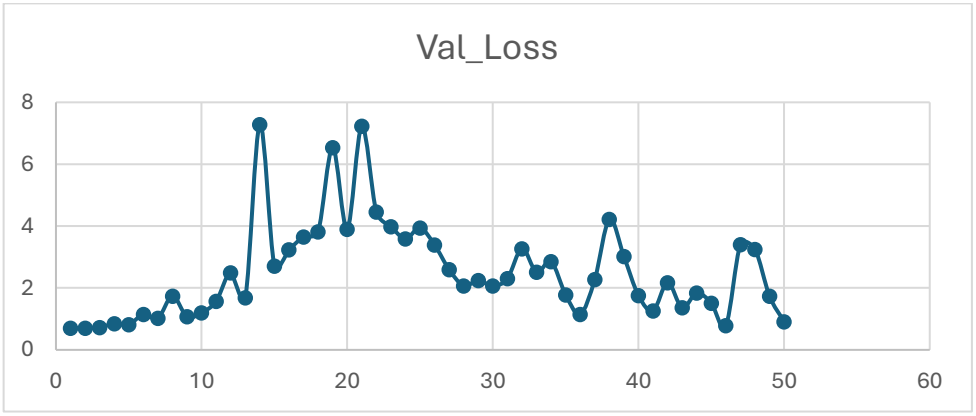
Εικόνα 4.2.8: Γραφική Παράσταση ακρίβειας (Accuracy) κατά τη διάρκεια της εκπαίδευσης του μοντέλου MobileVit.



Εικόνα 4.2.9: Γραφική Παράσταση απώλειας (Loss) κατά τη διάρκεια της εκπαίδευσης του μοντέλου MobileVit.



Εικόνα 4.2.10: Γραφική Παράσταση ακρίβειας επικύρωσης (Validation Accuracy) κατά τη διάρκεια της εκπαίδευσης του μοντέλου MobileVit.



Εικόνα 4.2.11: Γραφική Παράσταση απώλειας επικύρωσης (Validation Loss) κατά τη διάρκεια της εκπαίδευσης του μοντέλου MobileVit.

4.3 Αξιολόγηση Απόδοσης Μοντέλων για Βίντεο

Από τα δέκα αποτελέσματα του κάθε μοντέλου, για τα συνολικά δέκα σετ δεδομένων, επιλέχθηκαν τα πέντε καλύτερα, με έμφαση στις μετρικές ακρίβειας, ανάκλησης, ακρίβειας και βαθμολογίας F1 για την τελική αξιολόγηση της απόδοσης των μοντέλων.

TSN (Temporal Segment Network) αποτελέσματα:

Το μοντέλο TSN επέδειξε συγκρίσιμη σταθερότητα στις μετρήσεις αξιολόγησης πόνου. Κατέγραψε μέσες τιμές 68% για την ακρίβεια, 66% για την ανάκληση, 66% για τη γενική ακρίβεια, 66% για το ROC AUC και 65% για το F1 σκορ, με μέση τιμή λογαριθμικής απώλειας 5.36. Η κατανομή ταξινόμησης έδειξε 181 θετικά αληθή, 95.8 ψευδή θετικά, 79 ψευδή αρνητικά και 164.2 αληθή αρνητικά. Η μέγιστη επίδοση έφτασε το 71% για την ακρίβεια, 69% για την ανάκληση και τη γενική ακρίβεια, με F1 σκορ 69% και μέγιστη λογαριθμική απώλεια 5.67, ενώ οι μέγιστες τιμές για θετικά αληθή, ψευδή θετικά, ψευδή αρνητικά και αληθή αρνητικά ήταν 231, 151, 129 και 223 αντίστοιχα. Οι ελάχιστες μετρήσεις ήταν συνεπείς στο 64% για ακρίβεια, ανάκληση, γενική ακρίβεια και ROC AUC, με F1 σκορ στο 63% και λογαριθμική απώλεια στο 4.96. Οι τυπικές αποκλίσεις παρέμειναν χαμηλές στο 2% για την ακρίβεια και 1.7% για την ανάκληση, τη γενική ακρίβεια και το ROC AUC, με λογαριθμική απώλεια στο 0.27.

	Precision	**Recall**	**Accuracy**	**ROC AUC**	**F1 Score**	**Log Loss**	**TP**	**FP**	**FN**	**TN**
AVG	0.6830	0.6638	0.6638	0.6638	0.6548	5.3591	181	95.8	79	164.2
Max	0.7066	0.6885	0.6885	0.6884	0.6868	5.6718	231	151	129	223
Min	0.6581	0.6442	0.6442	0.6442	0.6337	4.9667	131	37	29	109
STDEV.P	0.0198	0.0174	0.0174	0.0174	0.0206	0.2779	39.18	42.53	39.187	42.527

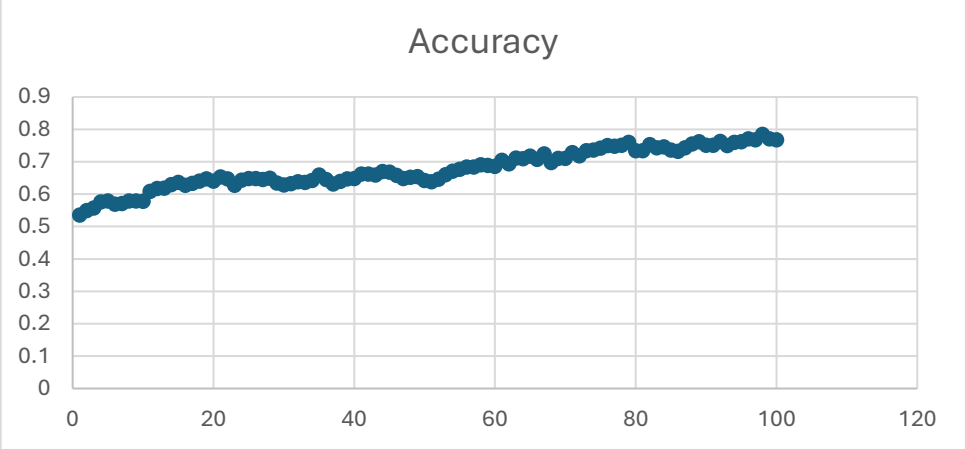
TP: True Positives, FP: False Positives, FN: False Negatives, TN: True Negatives

Πίνακας 4.3.1: Μετρικές απόδοσης για το μοντέλο TSN στα δεδομένα δοκιμών.

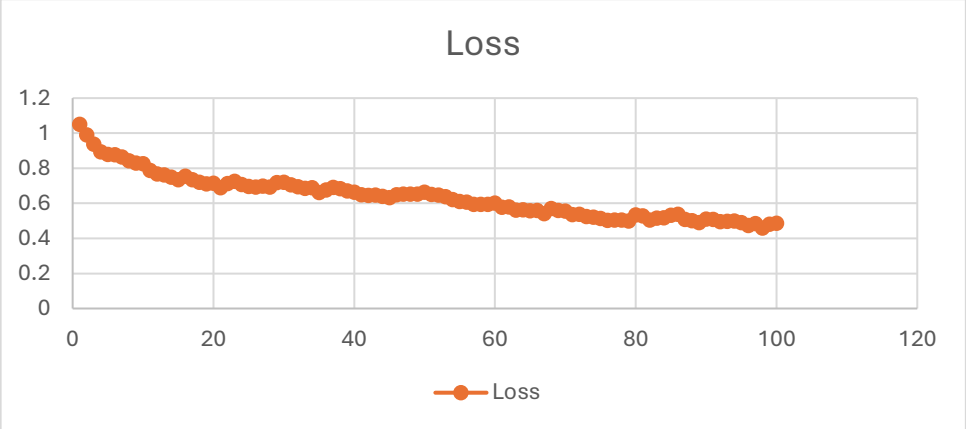
Το μοντέλο TSN υποβλήθηκε σε εκτεταμένη εκπαίδευση για 100 εποχές χρησιμοποιώντας τον βελτιστοποιητή Adam, απαιτώντας 3.7 έως 4 ώρες ανά πλήρη κύκλο. Οι μετρήσεις της τελευταίας εποχής έδειξαν μέσο λάθος εκπαίδευσης 48%, ακρίβεια εκπαίδευσης 76% και ακρίβεια επικύρωσης/δοκιμής 74%. Οι υψηλότεροι δείκτες επίδοσης περιλάμβαναν λάθος εκπαίδευσης 53%, ακρίβεια εκπαίδευσης 82% και ακρίβεια επικύρωσης/δοκιμής 77%. Τα κατώτατα όρια του μοντέλου χαρακτηρίστηκαν από λάθος εκπαίδευσης 41%, ακρίβεια εκπαίδευσης 71% και ακρίβεια επικύρωσης/δοκιμής 71%. Οι ιδιαίτερα μικρές τυπικές αποκλίσεις, 4.3% για το λάθος εκπαίδευσης, 4% για την ακρίβεια εκπαίδευσης και 2% για την ακρίβεια επικύρωσης/δοκιμής, υποδηλώνουν σταθερή απόδοση καθ' όλη τη διάρκεια της εκπαίδευσης.

	Epochs	Optimizer	Training Time (mins)	Train Error of Last Epoch	Train Accuracy of Last Epoch	Test/Val Accuracy of Last Epoch
AVERAGE	100	Adam	222	0.4852	0.7675	0.746
MAX	100	Adam	227	0.5324	0.8263	0.7725
MIN	100	Adam	218	0.4113	0.7163	0.71
STDEV.P	0	Adam	5	0.0431	0.0372	0.0236

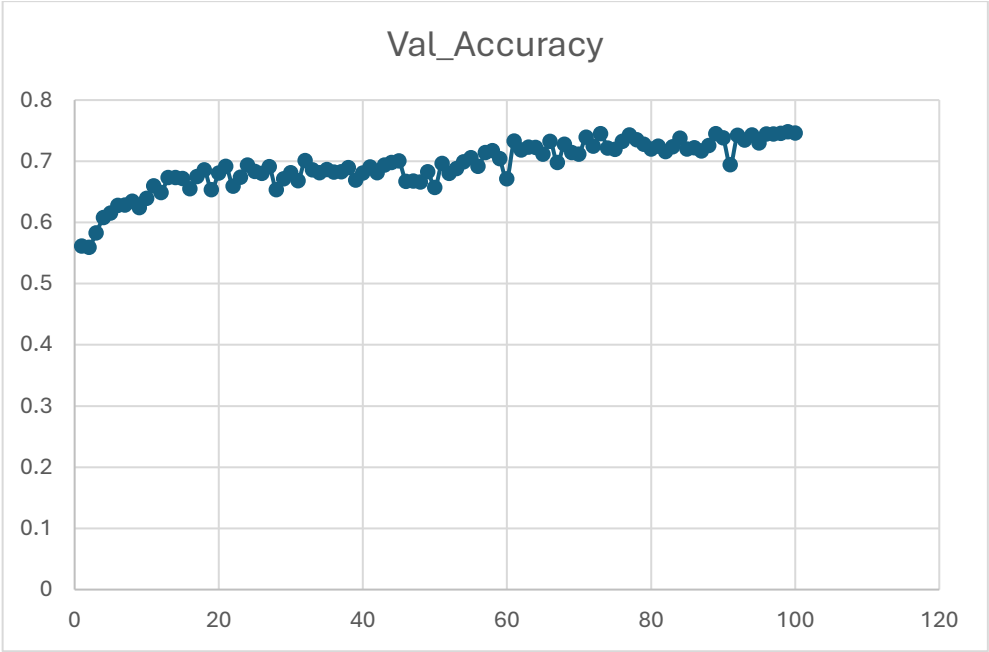
Πίνακας 4.3.2: Μετρικές εποχών για το μοντέλο TSN.



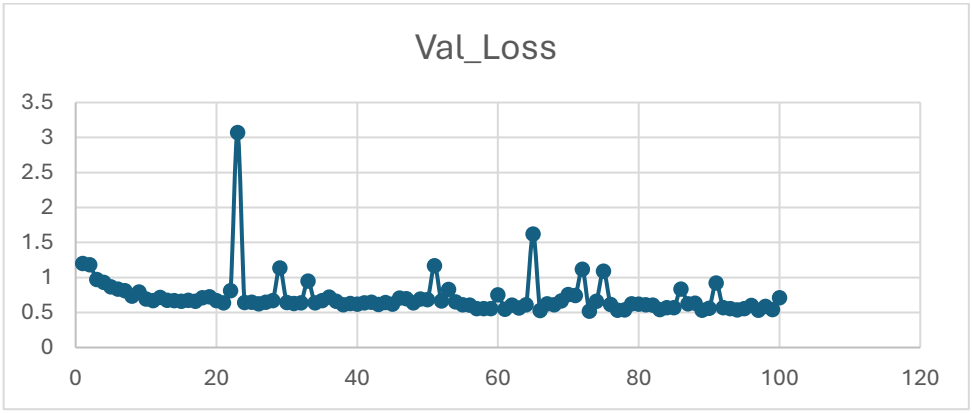
Εικόνα 4.3.3: Γραφική Παράσταση ακρίβειας (Accuracy) κατά τη διάρκεια της εκπαίδευσης του μοντέλου TSN.



Εικόνα 4.3.4: Γραφική Παράσταση απώλειας (Loss) κατά τη διάρκεια της εκπαίδευσης του μοντέλου TSN.



Εικόνα 4.3.5: Γραφική Παράσταση ακρίβειας επικύρωσης (Validation Accuracy) κατά τη διάρκεια της εκπαίδευσης του μοντέλου TSN.



Εικόνα 4.3.6: Γραφική Παράσταση απώλειας επικύρωσης (Validation Loss) κατά τη διάρκεια της εκπαίδευσης του μοντέλου TSN.

ConvLSTM αποτελέσματα:

Η αρχιτεκτονική ConvLSTM επέδειξε ισορροπημένες μετρικές απόδοσης σε εργασίες ανίχνευσης πόνου. Το μοντέλο πέτυχε ομοιόμορφες μέσες τιμές 66% σε ακρίβεια, ανάκληση, ακριβοδικαιοσύνη και ROC AUC, διατηρώντας έναν δείκτη F1 66% και λογαριθμική απώλεια 5.34. Οι αποτελέσματα ταξινόμησης είχαν μέσο όρο 187.6 TP, 102 FP, 72.4 FN και 158 TN. Η κορυφαία απόδοση έδειξε ακρίβεια στο 70%, με 68% για την ανάκληση και ακριβοδικαιοσύνη, ROC AUC στο 68% και δείκτη F1 στο 67%, παράλληλα με μέγιστη λογαριθμική απώλεια 5.54. Οι μέγιστες τιμές ταξινόμησης έφτασαν τα 218 TP, 123 FP, 96 FN και 175 TN. Οι ελάχιστες μετρικές διατηρήθηκαν

σταθερά στο 65% για ακρίβεια, ανάκληση, ακριβοδικαιοσύνη και ROC AUC, με δείκτη F1 στο 65% και λογαριθμική απώλεια 5.05. Οι τυπικές αποκλίσεις παρέμειναν ελάχιστες στο 1.7% για την ακρίβεια και 1.1% για την ανάκληση, ακριβοδικαιοσύνη και ROC AUC, με διακύμανση της λογαριθμικής απώλειας στο 0.17.

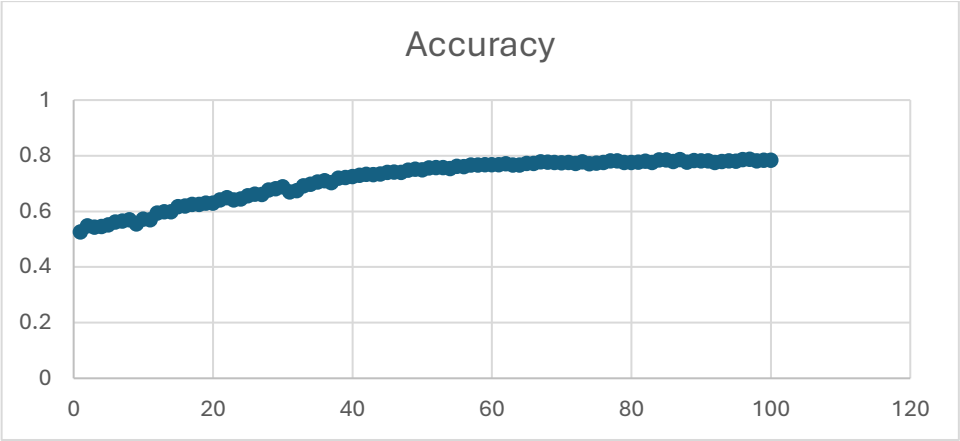
	Precision	Recall	Accuracy	ROC AUC	F1 Score	Log Loss	TP	FP	FN	TN
AVG	0.6696	0.6646	0.6646	0.6646	0.6624	5.3468	187.6	102	72.4	158
Max	0.7023	0.6826	0.6827	0.6827	0.6748	5.5492	218	123	96	175
Min	0.6522	0.6519	0.6519	0.6519	0.6518	5.0586	164	85	42	137
STDEV.P	0.0175	0.0108	0.0108	0.0108	0.0085	0.1729	17.579	12.73	17.58	12.73
TP: True Positives, FP: False Positives, FN: False Negatives, TN: True Negatives										

Πίνακας 4.3.7: Μετρικές απόδοσης για το μοντέλο ConvLSTM στα δεδομένα δοκιμών.

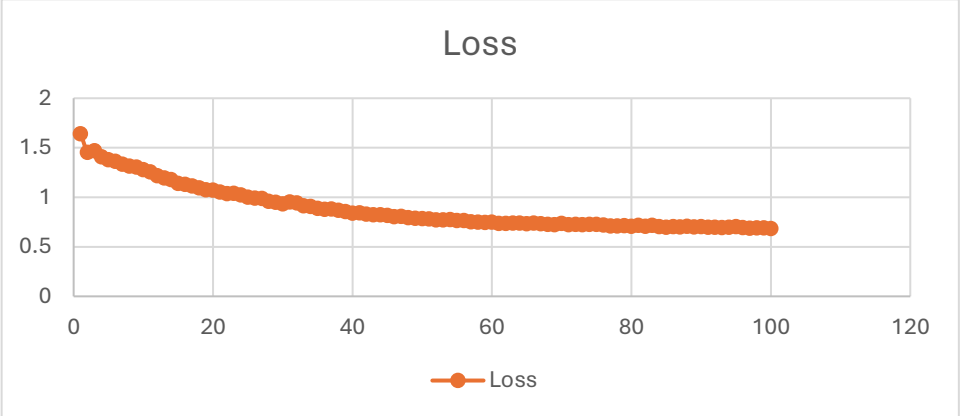
Η εκπαίδευση του ConvLSTM, η οποία υλοποιήθηκε σε 100 εποχές με τη χρήση του βελτιστοποιητή Adam, απαιτούσε περίπου 3,5 ώρες ανά πλήρη κύκλο. Η τελευταία εποχή κατέδειξε μέσο ποσοστό σφάλματος εκπαίδευσης 68%, ακρίβεια εκπαίδευσης 78% και ακριβοδικαιοσύνη στον έλεγχο/επικύρωση 71%. Τα μέγιστα μετρικά στοιχεία περιλάμβαναν ποσοστό σφάλματος εκπαίδευσης 75%, ακρίβεια εκπαίδευσης 84% και ακριβοδικαιοσύνη στον έλεγχο/επικύρωση 76%. Η ελάχιστη απόδοση έδειξε ποσοστό σφάλματος εκπαίδευσης 49%, ακρίβεια εκπαίδευσης 75% και ακριβοδικαιοσύνη στον έλεγχο/επικύρωση 69%. Η σταθερότητα της εκπαίδευσης αποτυπώθηκε στις τυπικές αποκλίσεις, 10% για το σφάλμα εκπαίδευσης, 3.3% για την ακρίβεια εκπαίδευσης και 2.8% για την ακριβοδικαιοσύνη στον έλεγχο/επικύρωση.

	Epochs	Optimizer	Training Time (mins)	Train Error of Last Epoch	Train Accuracy of Last Epoch	Test/Val Accuracy of Last Epoch
AVERAGE	100	Adam	213	0.6878	0.7832	0.7195
MAX	100	Adam	209	0.7497	0.8456	0.7675
MIN	100	Adam	217	0.4894	0.7531	0.6975
STDEV.P	0	Adam	4	**0.0998**	0.0328	0.0273

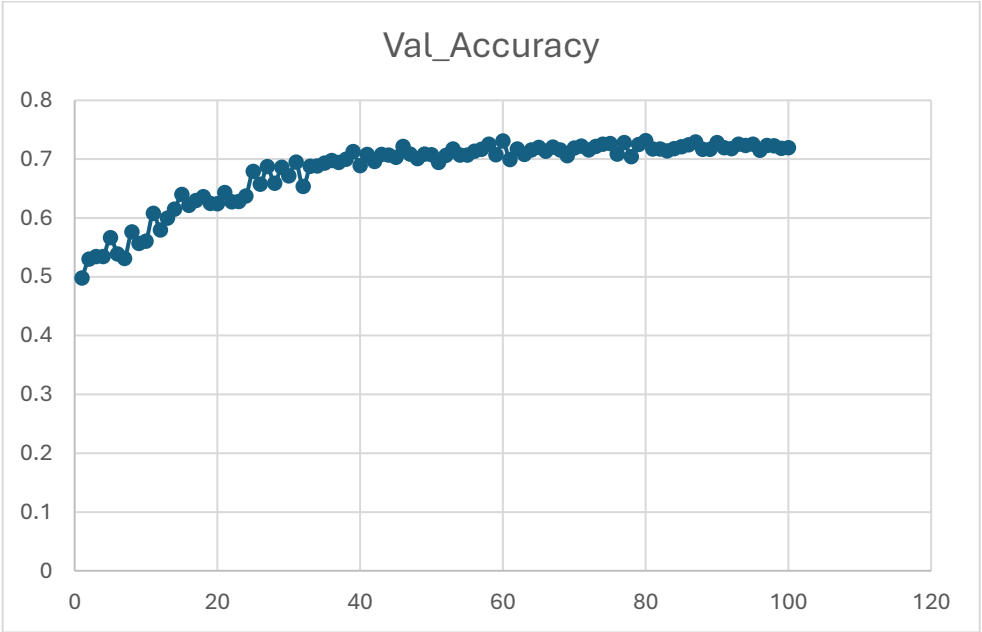
Πίνακας 4.3.8: Μετρικές εποχών για το μοντέλο ConvLSTM.



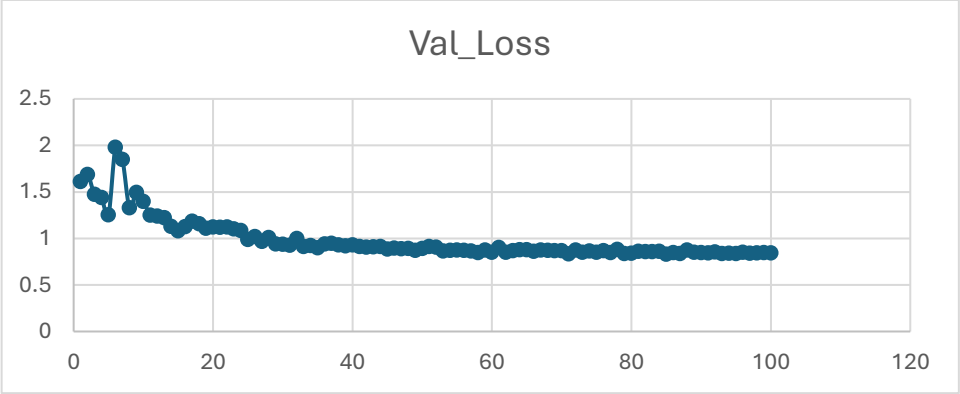
Εικόνα 4.3.9: Γραφική Παράσταση ακρίβειας (Accuracy) κατά τη διάρκεια της εκπαίδευσης του μοντέλου ConvLSTM.



Εικόνα 4.3.10: Γραφική Παράσταση απώλειας (Loss) κατά τη διάρκεια της εκπαίδευσης του μοντέλου ConvLSTM.



Εικόνα 4.3.11: Γραφική Παράσταση ακρίβειας επικύρωσης (Validation Accuracy) κατά τη διάρκεια της εκπαίδευσης του μοντέλου ConvLSTM.



Εικόνα 4.3.12: Γραφική Παράσταση απώλειας επικύρωσης (Validation Loss) κατά τη διάρκεια της εκπαίδευσης του μοντέλου ConvLSTM.

VideoMAE αποτελέσματα:

Το μοντέλο VideoMAE επέδειξε σταθερή απόδοση στην κατηγοριοποίηση επιπέδων πόνου. Το μοντέλο επιτυγχάνει συνεπείς μέσους όρους 71% σε ακρίβεια, ανάκληση, ακριβοδικαιοσύνη, ROC AUC και F1 βαθμολογία, με μέση τιμή log-loss 10.38. Η κατανομή ταξινόμησης έχει μέσους όρους 183.80 TP, 73.60 FP, 76.20 FN, και 186.40 TN. Οι μέγιστες επιδόσεις έφτασαν το 77% για την ακρίβεια, 76% για την ανάκληση και την ακριβοδικαιοσύνη, με το μέγιστο log-loss στο 11.23, αντιστοιχώντας σε τιμές TP, FP, FN, και TN 219, 116, 108, και 214 αντίστοιχα. Οι ελάχιστες τιμές καταγράφηκαν σταθερά στο 69% για ακρίβεια, ανάκληση και ακριβοδικαιοσύνη, με log-loss 8.66, και αντίστοιχες τιμές ταξινόμησης 152 TP, 46 FP, 41 FN, και 144 TN. Οι τυπικές αποκλίσεις διατηρήθηκαν ομοιόμορφες στο 3% για ακρίβεια, ανάκληση, ακριβοδικαιοσύνη, ROC AUC και F1 βαθμολογία, με διακύμανση log-loss στο 0.93.

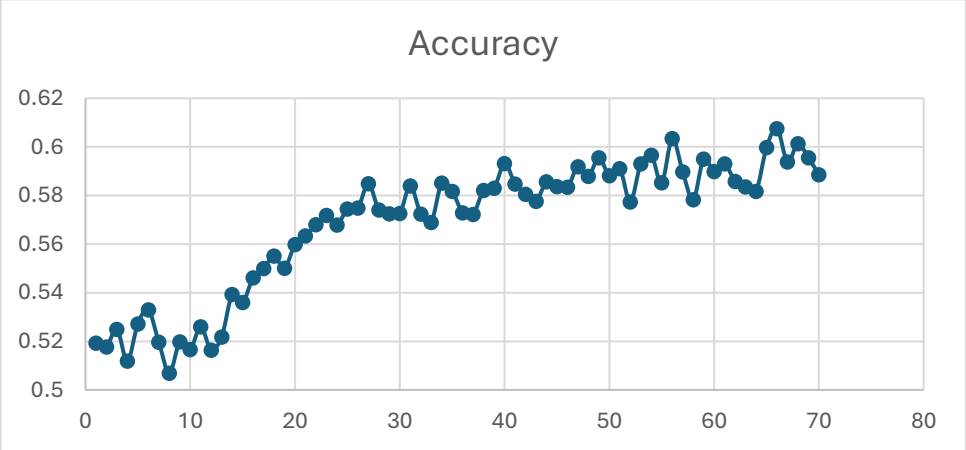
	Precision	Recall	Accuracy	ROC AUC	F1 Score	Log Loss	TP	FP	FN	TN
AVG	0.72078	0.7119	0.71194	0.71194	0.70896	10.3833	183.8	73.6	76.2	186.4
Max	0.765	0.7596	0.7596	0.7596	0.7584	11.229	219	116	108	214
Min	0.6889	0.6885	0.6885	0.6885	0.6883	8.6643	152	46	41	144
STDEV.P	0.025478	0.0256	0.02565	0.02565	0.02636	0.9252	28.35	25.192	28.35	25.192
TP: True Positives, FP: False Positives, FN: False Negatives, TN: True Negatives										

Πίνακας 4.3.13: Μετρικές απόδοσης για το μοντέλο VideoMAE στα δεδομένα δοκιμών.

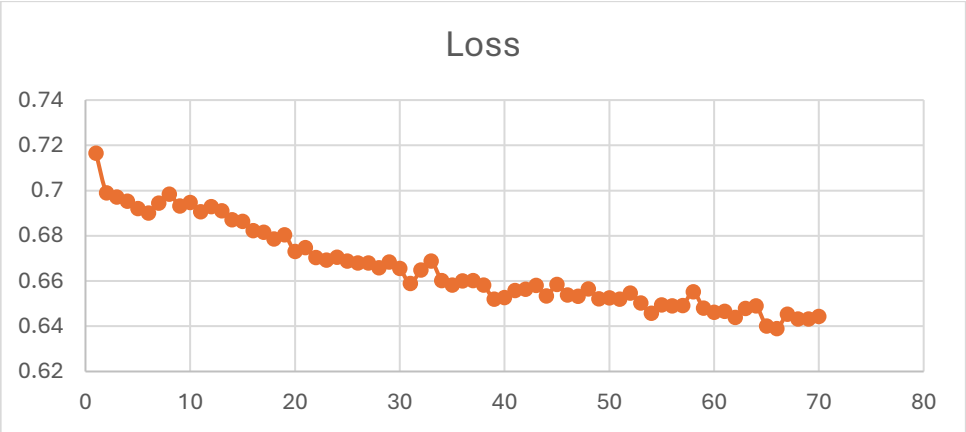
Η διαδικασία εκπαίδευσης του μοντέλου VideoMAE, χρησιμοποιώντας τον βελτιστοποιητή AdamW για 70 εποχές, απαιτούσε κατά μέσο όρο 16.4 ώρες για ολοκλήρωση. Τα μετρικά στοιχεία της τελευταίας εποχής έδειξαν μέσο σφάλμα εκπαίδευσης 64%, ακρίβεια εκπαίδευσης 59% και ακρίβεια δοκιμής/επικύρωσης 71%. Οι μέγιστοι δείκτες επίδοσης περιλάμβαναν σφάλμα εκπαίδευσης 66%, ακρίβεια εκπαίδευσης 62% και ακρίβεια δοκιμής/επικύρωσης 73%, ενώ οι ελάχιστες τιμές καταγράφηκαν σε 62% για σφάλμα εκπαίδευσης, 57% για ακρίβεια εκπαίδευσης και 66% για ακρίβεια δοκιμής/επικύρωσης. Η συνέπεια της εκπαίδευσης αποδείχθηκε με τις χαμηλές τυπικές αποκλίσεις 1% για το σφάλμα εκπαίδευσης, 2% για την ακρίβεια εκπαίδευσης και 3% για την ακρίβεια δοκιμής/επικύρωσης.

	Epochs	Optimizer	Training Time (mins)	Train Error of Last Epoch	Train Accuracy of Last Epoch	Test/Val Accuracy of Last Epoch
AVERAGE	70	AdamW	986	0.64424	0.58862	0.707
MAX	70	AdamW	1013	0.6553	0.6225	0.73
MIN	70	AdamW	922	0.6228	0.5694	0.6575
STDEV.P	0	AdamW	34.46	0.011256	0.01927	0.025855367

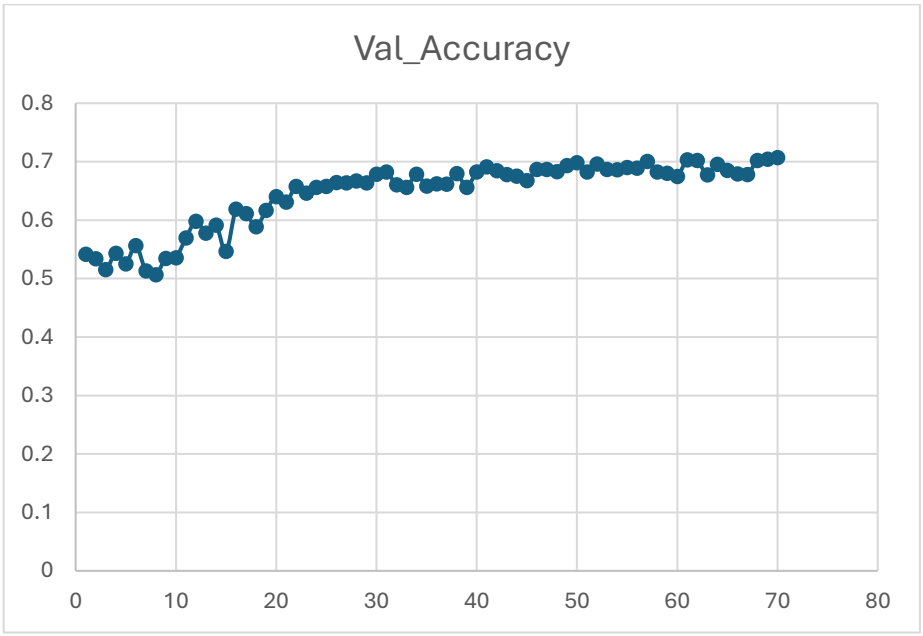
Πίνακας 4.3.14: Μετρικές εποχών για το μοντέλο VideoMAE.



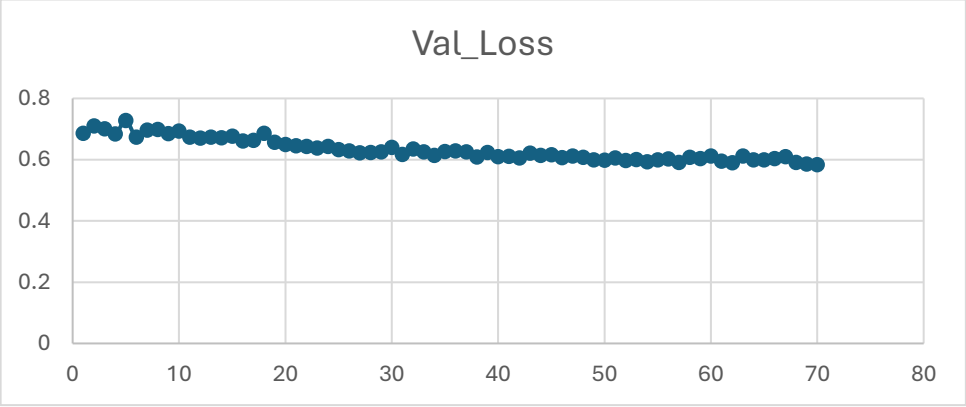
Εικόνα 4.3.15: Γραφική Παράσταση ακρίβειας (Accuracy) κατά τη διάρκεια της εκπαίδευσης του μοντέλου VideoMAE.



Εικόνα 4.3.16: Γραφική Παράσταση απώλειας (Loss) κατά τη διάρκεια της εκπαίδευσης του μοντέλου VideoMAE.



Εικόνα 4.3.17: Γραφική Παράσταση ακρίβειας επικύρωσης (Validation Accuracy) κατά τη διάρκεια της εκπαίδευσης του μοντέλου VideoMAE.



Εικόνα 4.3.18: Γραφική Παράσταση απώλειας επικύρωσης (Validation Loss) κατά τη διάρκεια της εκπαίδευσης του μοντέλου VideoMAE.

4.4 Σύγκριση Μοντέλων

Μετά την ανάλυση των αποδόσεων των μοντέλων VGG16, MobileViT, VideoMAE, TSN, και ConvLSTM, είναι σημαντικό να αναφερθούν τα εξής για τα βίντεο μοντέλα:

- **VideoMAE**: Επέδειξε την πιο ισορροπημένη απόδοση με ομοιόμορφα σταθερές τιμές απόδοσης (71% σε όλες τις μετρικές) και σταθερές τυπικές αποκλίσεις (3%). Παρόλο που παρουσίασε υψηλό log-loss (10.38) και μακρύ χρόνο εκπαίδευσης (16.4 ώρες για 70 εποχές), φαίνεται να είναι ιδανικό για κλινικές εφαρμογές λόγω της συνέπειας των προβλέψεών του.
- **TSN και ConvLSTM**: Παρουσίασαν μέτρια αλλά σταθερή απόδοση γύρω στο 66%, με το ConvLSTM να έχει μικρότερη προβλεπτική μεταβλητότητα (1.1% έναντι 1.7%) και μικρότερο χρόνο εκπαίδευσης (3.5 έναντι 4 ώρες). Αυτά τα μοντέλα προσφέρουν ισορροπημένες λύσεις όταν η σταθερότητα και οι μέτριες υπολογιστικές απαιτήσεις είναι κρίσιμοι παράγοντες.

Από την άλλη πλευρά, για τα μοντέλα εικόνας:

- **VGG16 και MobileViT**: Το VGG16 απέδειξε υψηλότερη συνέπεια στις προβλέψεις με τη χαμηλότερη τυπική απόκλιση στην ακρίβεια (0.016) και το χαμηλότερο log-loss (0.756), υποδεικνύοντας υψηλή εμπιστοσύνη στις προβλέψεις του. Το MobileViT, αν και μοντέρνο αρχιτεκτονικά, παρουσίασε αστάθεια στις προβλέψεις με υψηλότερο log-loss (2.753) και μεγαλύτερη τυπική απόκλιση στην ακρίβεια (0.064), υποδηλώνοντας ανάγκη για βελτίωση, ιδιαίτερα στη συνέπεια και την εμπιστοσύνη των προβλέψεων.

Συνολικά, η σύγκριση υπογραμμίζει τη σημασία της επιλογής του κατάλληλου μοντέλου ανάλογα με τις απαιτήσεις της εφαρμογής, με το VGG16 να ξεχωρίζει για τη συνέπεια και την εμπιστοσύνη των προβλέψεων, ενώ το VideoMAE προτείνεται για κλινικές εφαρμογές λόγω της σταθερής απόδοσής του.

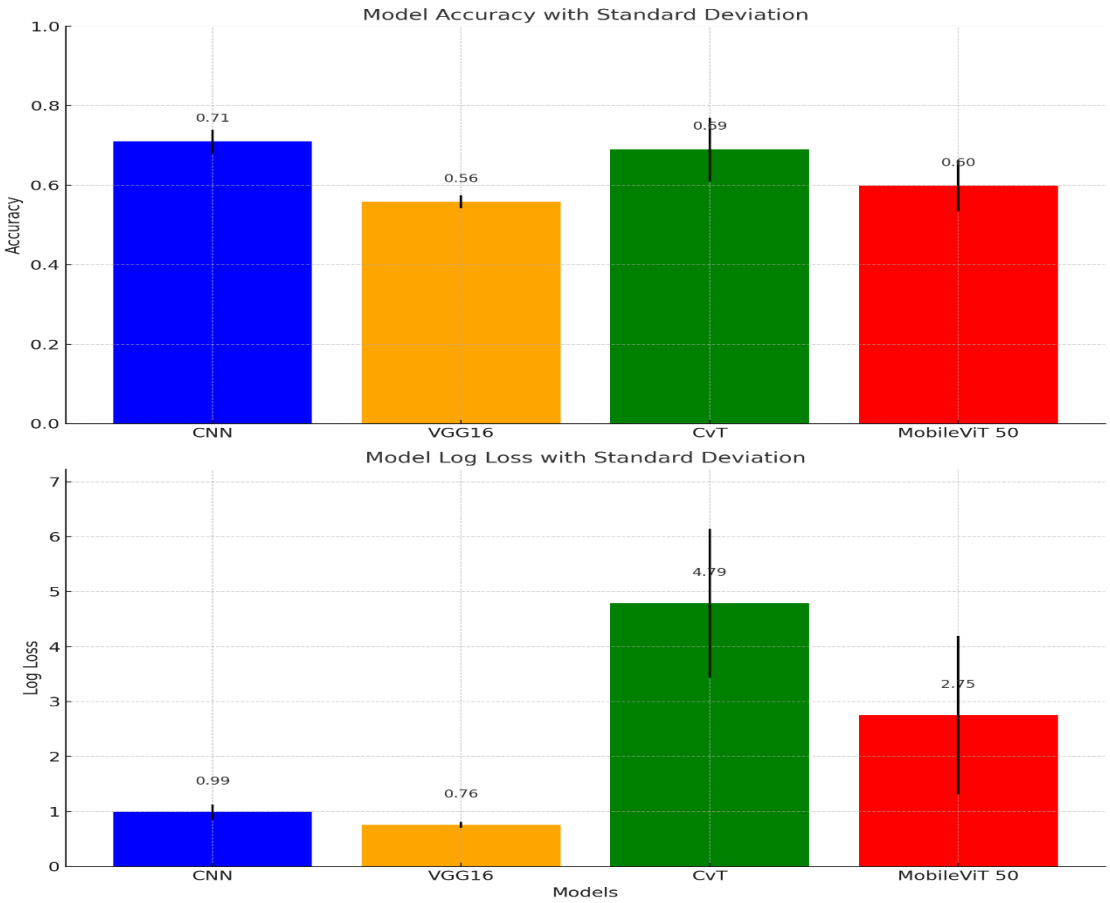
Κεφάλαιο 5

Συζήτηση

5.1 Συζήτηση για Ανάλυση Εικόνων

Η παρούσα ενότητα εξετάζει τα αποτελέσματα της ανάλυσης εικόνων για την αναγνώριση της έντασης πόνου, εστιάζοντας στη σύγκριση των αποτελεσμάτων με άλλες μελέτες από τη βιβλιογραφία. Επιπλέον, συγκρίνονται τα αποτελέσματα με εκείνα του συναδέλφου, Ανδρέα Αναστασίου, ο οποίος χρησιμοποίησε τα μοντέλα Convolutional Neural Network (CNN) και Convolutional Vision Transformer (CvT) για την εκτίμηση της έντασης πόνου.

Σύγκριση με Αποτελέσματα του συναδέλφου Ανδρέα Αναστασίου



Εικόνα 5.1.1: Γραφικές παραστάσεις που παρουσιάζουν την ακρίβεια και την απώλεια των μοντέλων με τις αντίστοιχες τυπικές αποκλίσεις. Συγκρίνονται τα μοντέλα CNN, VGG16, CvT, και MobileViT.

Ο Ανδρέας Αναστασίου χρησιμοποίησε CNN και CvT για την ανάλυση εικόνων της βάσης δεδομένων BioVid Heat Pain, πετυχαίνοντας ακρίβεια 71% και 69% αντίστοιχα. Η έρευνά του διασφάλισε αυστηρό διαχωρισμό συμμετεχόντων σε φάσεις εκπαίδευσης και δοκιμής, όπως και στη δική μου εργασία. Τα αποτελέσματά του δείχνουν την αποτελεσματικότητα των μοντέλων του, με το CvT να προσφέρει ανώτερη ικανότητα σύλληψης τόσο τοπικών όσο και παγκόσμιων χαρακτηριστικών, κάτι που πιθανώς εξηγεί την ανώτερη απόδοσή του σε σχέση με το MobileViT και το VGG16 που χρησιμοποιήσαμε.

Σε σύγκριση, τα MobileViT και VGG16 απέδωσαν ακρίβειες 59.8% και 55.8% αντίστοιχα για τη διάκριση μεταξύ εικόνων χωρίς πόνο (BL1) και εικόνων με μέγιστο πόνο (PA4). Τα αποτελέσματά μας αντανακλούν τη δυνατότητα χρήσης πιο ελαφριών μοντέλων, ενώ ταυτόχρονα υπογραμμίζουν τη σημασία της αρχιτεκτονικής του μοντέλου στην επίτευξη υψηλότερης ακρίβειας.

Το CvT, με την ικανότητα διαχείρισης χωρικών και χρονικών χαρακτηριστικών, υπερέχει στην ανάλυση εικόνων και βίντεο, ενώ το MobileViT παρέχει ελαφριά και φορητή λύση. Τα αποτελέσματά μας δείχνουν ότι το αρχιτεκτονικό σχέδιο των CvT και CNN είναι πιο κατάλληλο για τη συγκεκριμένη εφαρμογή.

Σύγκριση με Βιβλιογραφία

1. **Devi et al., 2019** [31]**:** Χρησιμοποίησαν ένα CNN με ReLU layer, max pooling και dropout, εφαρμόζοντας grayscale εικόνες με μέγεθος 256x256 pixels. Η ακρίβεια των αποτελεσμάτων ήταν πολύ υψηλή, φτάνοντας το 93.24% για τα δεδομένα εκπαίδευσης και 94.15% για τα δεδομένα δοκιμής. Συγκριτικά, η προσέγγισή μας υπερτερεί στη γενίκευση λόγω του αυστηρού διαχωρισμού συμμετεχόντων.
2. **Dragomir et al., 2020** [6]**:** Ανέπτυξαν ένα CNN-ResNet 18 με προεκπαιδευμένο ανιχνευτή χαρακτηριστικών προσώπου και ενσωμάτωσαν data augmentation για βελτίωση της απόδοσης. Παρά τις προσπάθειες, η ακρίβεια της δοκιμής έφτασε μόλις το 36.6% για όλα τα επίπεδα πόνου. Σε σύγκριση, η δική μας προσέγγιση διασφαλίζει τη δοκιμή σε ανεπανάληπτα δεδομένα, γεγονός που ενισχύει την αντικειμενικότητα των αποτελεσμάτων.
3. **Zheng et al., 2022** [40]**:** Χρησιμοποίησαν CNN με input εικόνες 64x64 και τεχνικές κανονικοποίησης προσώπων. Εφάρμοσαν πολλαπλές κλάσεις για την ταξινόμηση πόνου, με ακρίβεια 52% για binary ταξινόμηση και 51.1% για όλα τα επίπεδα πόνου. Η απόδοση αυτή υποδεικνύει τη δυσκολία γενίκευσης με μικρότερα μεγέθη εισόδου. Η μέθοδος μας υπερτερεί στη διαχείριση μεγαλύτερων και περισσότερο ποικιλόμορφων δεδομένων.
4. **Lledo et al., 2023** [3]**:** Το CNN που χρησιμοποιήθηκε εφαρμόστηκε σε δεδομένα που εστίαζαν σε διαφορετικές κατηγορίες, ηλικίες και γένη. Με ακρίβεια κάτω του 50% για pain/no pain και μόλις 20% για όλα τα επίπεδα πόνου, τα αποτελέσματα τονίζουν τις προκλήσεις της μεθοδολογίας τους. Η μεθοδολογία μας, αν και με μικρότερα ποσοστά ακρίβειας, διασφαλίζει μεγαλύτερη αξιοπιστία λόγω της απουσίας bias.
5. **Othman et al., 2019** [26]**:** Χρησιμοποίησαν δύο προσεγγίσεις, RFc με FAD και MobilNetV2, με δεδομένα 96x96 pixels από 87 συμμετέχοντες. Τα μοντέλα αξιολογήθηκαν με 5-fold cross validation, επιτυγχάνοντας ακρίβειες 65.8% και 65.5% αντίστοιχα για binary ταξινόμηση pain/no pain.

Κοινές Παρατηρήσεις

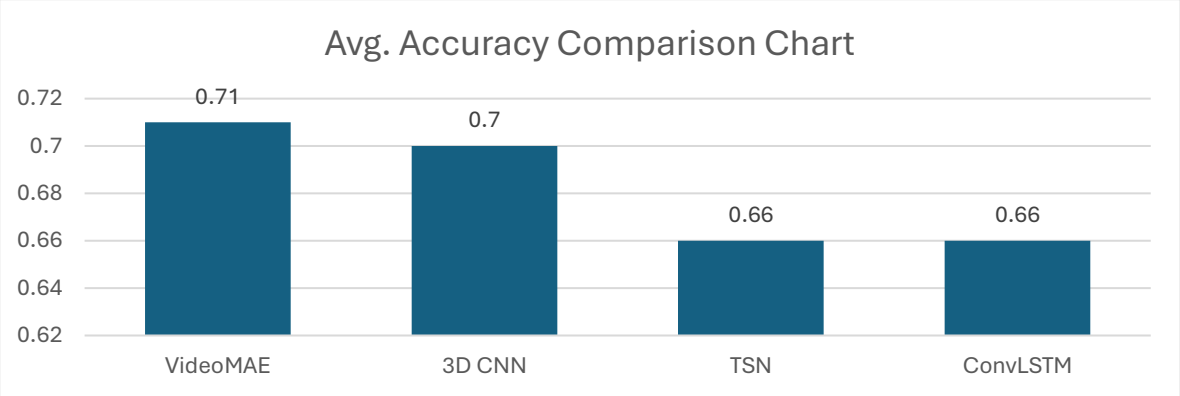
Η σύγκριση αναδεικνύει ότι τα μοντέλα μας, αν και παρουσιάζουν χαμηλότερη ακρίβεια σε ορισμένες περιπτώσεις, υπερτερούν σε αντικειμενικότητα και γενίκευση. Ο διαχωρισμός συμμετεχόντων για εκπαίδευση και δοκιμή ελαχιστοποιεί τη μεροληψία και παρέχει πιο ρεαλιστική αξιολόγηση. Ωστόσο, η ενσωμάτωση προηγμένων χαρακτηριστικών, όπως καλύτερη ανίχνευση χαρακτηριστικών προσώπου και εμπλουτισμός δεδομένων, μπορεί να βελτιώσει περαιτέρω την απόδοση.

Η μελλοντική εργασία θα εστιάσει στην αξιοποίηση αυτών των τεχνικών, με στόχο τη δημιουργία αποτελεσματικών και αξιόπιστων συστημάτων για την αναγνώριση πόνου μέσω μηχανικής μάθησης.

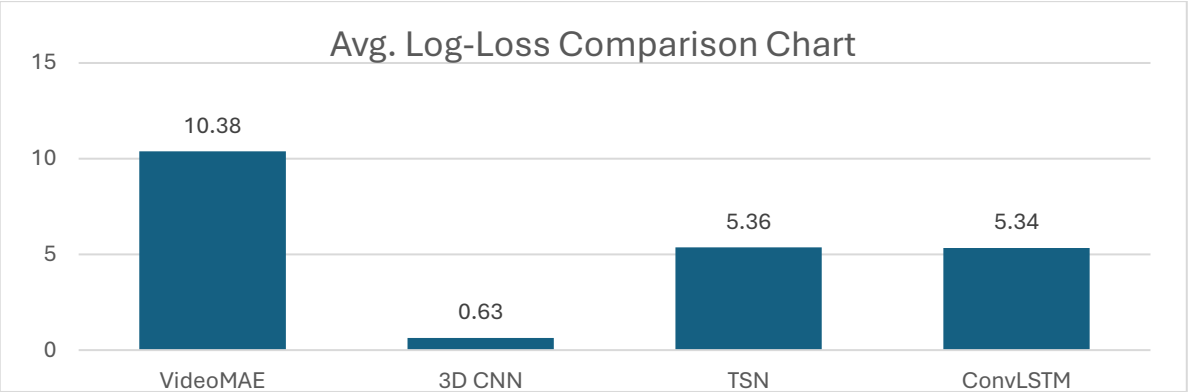
5.2 Συζήτηση για Ανάλυση Βίντεο

Η παρούσα ενότητα εξετάζει τα αποτελέσματα της ανάλυσης βίντεο για την αναγνώριση της έντασης πόνου, εστιάζοντας στη σύγκριση των αποτελεσμάτων με άλλες μελέτες από τη βιβλιογραφία. Επιπλέον, συγκρίνονται τα αποτελέσματα με εκείνα του συναδέλφου Ανδρέα Αναστασίου, ο οποίος χρησιμοποίησε το μοντέλο 3DCNN.

Σύγκριση με Αποτελέσματα του Ανδρέα Αναστασίου



Εικόνα 5.2.1: Διάγραμμα Σύγκρισης Μέσης Ακρίβειας.



Εικόνα 5.2.2: Διάγραμμα Σύγκρισης Μέσης Απώλειας Log-Loss.

Ο Ανδρέας Αναστασίου χρησιμοποίησε το μοντέλο 3DCNN για την ανάλυση βίντεο, πετυχαίνοντας ακρίβεια 70%, την υψηλότερη τιμή ROC AUC (77%), και F1-score 67%. Το μοντέλο βασίστηκε σε Conv2Plus1D layers, τα οποία συνδυάζουν χωρικές και χρονικές διαστάσεις με αποτελεσματικότητα, διατηρώντας παράλληλα μια σχετικά μικρή διάρκεια εκπαίδευσης (1 ώρα και 23 λεπτά για 50 εποχές). Η υψηλή ακρίβεια του 3DCNN υπογραμμίζει την αποδοτικότητα των υπολειμματικών μπλοκ και την ικανότητα εξαγωγής χωροχρονικών χαρακτηριστικών.

Σε σύγκριση, τα δικά μου μοντέλα ConvLSTM και TSN πέτυχαν ακρίβεια 66%, ενώ το VideoMAE υπερείχε με ακρίβεια 71%. Παρότι το VideoMAE σημείωσε την υψηλότερη ακρίβεια, απαιτεί μεγαλύτερη υπολογιστική ισχύ (16.4 ώρες εκπαίδευσης για 70 εποχές). Τα ConvLSTM και TSN, αν και λιγότερο ακριβή, παρέχουν πιο σταθερές επιδόσεις και χαμηλότερη διακύμανση μεταξύ διαφορετικών μετρήσεων.

3DCNN.

Σύγκριση με Βιβλιογραφία

1. **Lledo et al., 2023** [3]**:** Το μοντέλο ViViT εφαρμόστηκε σε υποσύνολο δεδομένων που επιλέχθηκε βάσει διαφορετικών κατηγοριών, ηλικιών και φύλων. Η εκπαίδευση και η δοκιμή πραγματοποιήθηκαν σε 4600 και 1600 βίντεο αντίστοιχα. Παρά τις προσπάθειες, η ακρίβεια της δοκιμής ήταν χαμηλή (30.07%), υποδεικνύοντας τις προκλήσεις γενίκευσης σε τέτοιες αρχιτεκτονικές.
2. **Werner et al., 2014** [38]**:** Χρησιμοποίησαν συνδυασμό IntraFace και ICP για εξαγωγή χαρακτηριστικών προσώπου. Το μοντέλο εκπαιδεύτηκε σε 8700 τυχαία επιλεγμένα δείγματα και αξιολογήθηκε με 10-fold validation και LOOCV. Η ακρίβεια που επιτεύχθηκε ήταν 70% για όλα τα επίπεδα πόνου, παρέχοντας ισχυρό σημείο αναφοράς για πρώιμες μεθόδους αυτόματης ανίχνευσης πόνου.
3. **Werner et al., 2017** [37]**:** Εισήγαγαν ανιχνευτές χαρακτηριστικών Haar-like και ICP, βελτιώνοντας την ακρίβεια σε 72.4% για όλα τα επίπεδα πόνου. Η χρήση 10-fold validation συνέβαλε στην αξιοπιστία των αποτελεσμάτων, αναδεικνύοντας τη σημασία της ανίχνευσης χαρακτηριστικών προσώπου για ανίχνευση πόνου.

4. **Gigkas et al., 2023** [8]**:** Χρησιμοποίησαν MTCNN και FAN για εξαγωγή χαρακτηριστικών προσώπου, εφαρμόζοντας LOSO validation. Η ακρίβεια που επιτεύχθηκε ήταν 73.28% για binary ταξινόμηση pain/no pain και 31.52% για όλα τα επίπεδα πόνου, υπογραμμίζοντας την ανάγκη για βελτίωση της γενίκευσης.
5. **Gigkas et al., 2024** [7]**:** Βελτίωσαν την αρχιτεκτονική ενσωματώνοντας AugmNet, VGGFace2 και RAF-DB, επιτυγχάνοντας ακρίβεια 77.10% για binary ταξινόμηση pain/no pain και 35.39% για όλα τα επίπεδα πόνου. Η χρήση πολυτροπικών δεδομένων ανέδειξε τη δυνατότητα βελτίωσης της ακρίβειας.
6. **Tavakolian et al., 2019** [32]**:** Εφάρμοσαν Spatiotemporal Convolutional Networks (SCN), εκπαιδεύοντας το μοντέλο σε 8700 δείγματα από το CASIA WebFace. Το μοντέλο πέτυχε ακρίβεια 86.02% για binary ταξινόμηση pain/no pain, καταδεικνύοντας την αποτελεσματικότητα στη σύλληψη χρονικών δυναμικών.
7. **Xin et al., 2020** [40]**:** Χρησιμοποίησαν Attention-based Autoencoder, επιτυγχάνοντας ακρίβεια 85.65% για binary ταξινόμηση pain/no pain, αλλά μόνο 40.4% για όλα τα επίπεδα πόνου. Τα αποτελέσματα ανέδειξαν τις προκλήσεις στη μοντελοποίηση διαφορετικών εντάσεων πόνου.
8. **Patania et al., 2022** [27]**:** Εισήγαγαν GNN και DGCNN με landmarks προσώπου και SortPooling layers, πετυχαίνοντας ακρίβεια 73.2% και 83.4% αντίστοιχα για binary ταξινόμηση pain/no pain. Τα αποτελέσματα τονίζουν τη σημασία της γραφηματικής αναπαράστασης για την ανάλυση βίντεο.

Κοινές Παρατηρήσεις

Τα αποτελέσματά μας δείχνουν ότι το VideoMAE παρουσιάζει τη μεγαλύτερη ακρίβεια και σταθερότητα μεταξύ των μοντέλων μας, αν και με υψηλό κόστος σε χρόνο εκπαίδευσης. Το ConvLSTM, παρά τις χαμηλότερες επιδόσεις, προσφέρει μια καλή ισορροπία μεταξύ ακρίβειας και υπολογιστικής αποδοτικότητας, ενώ το TSN προσφέρει ευελιξία μέσω της επεξεργασίας τμημάτων. Συγκριτικά, το 3DCNN του Ανδρέα παρέχει μια αποτελεσματική λύση για ταχεία ανάπτυξη, αν και στερείται της ακρίβειας του VideoMAE.

Η μελλοντική εργασία θα επικεντρωθεί στη βελτίωση της αποτελεσματικότητας του VideoMAE, καθώς και στη διερεύνηση υβριδικών προσεγγίσεων που συνδυάζουν τα πλεονεκτήματα διαφορετικών αρχιτεκτονικών για την ανάλυση βίντεο.

Κεφάλαιο 6

Περιορισμοί και Προκλήσεις

6.1 Περιορισμοί και Προκλήσεις στην Ανάλυση Εικόνων

Η ανάλυση εικόνων για την ανίχνευση της έντασης πόνου παρουσίασε πολυάριθμες προκλήσεις που επηρέασαν την απόδοση των μοντέλων. Ένας από τους σημαντικότερους περιορισμούς ήταν η υποκειμενικότητα του πόνου, η οποία επηρεάζει τις αντιδράσεις των συμμετεχόντων. Κάθε άτομο αντιδρά διαφορετικά στα ίδια ερεθίσματα, γεγονός που καθιστά δύσκολη τη δημιουργία ενός μοντέλου ικανού να γενικεύει αποτελεσματικά. Αυτή η υποκειμενικότητα εκδηλώνεται τόσο στις εκφράσεις του προσώπου όσο και στις συναισθηματικές αντιδράσεις, καθιστώντας τον εντοπισμό ακριβών χαρακτηριστικών πιο απαιτητικό.

Παρά τα υποσχόμενα αποτελέσματα, η έρευνα αντιμετώπισε αρκετές προκλήσεις που επηρέασαν τη συνολική αποτελεσματικότητα και εφαρμοσιμότητα των μοντέλων. Ένα σημαντικό ζήτημα ήταν η υπερπροσαρμογή, ιδιαίτερα σε πιο σύνθετα μοντέλα όπως τα CNN και CvT, τα οποία διαθέτουν μεγάλο αριθμό παραμέτρων. Ενώ τα μοντέλα αυτά παρουσίασαν εξαιρετικές επιδόσεις στα δεδομένα εκπαίδευσης, δυσκολεύτηκαν να γενικεύσουν σε νέους συμμετέχοντες. Αυτό ήταν ιδιαίτερα εμφανές όταν δοκιμάστηκαν σε συμμετέχοντες με εκφράσεις προσώπου που διέφεραν σημαντικά από αυτές του training set. Τεχνικές όπως το dropout, η κανονικοποίηση παρτίδων (batch normalization) και το data augmentation χρησιμοποιήθηκαν για την αντιμετώπιση της υπερπροσαρμογής, αλλά η επίτευξη της σωστής ισορροπίας παρέμεινε πρόκληση.

Ένα μοναδικό χαρακτηριστικό αυτής της έρευνας ήταν η χρήση ξεχωριστών συμμετεχόντων για εκπαίδευση και δοκιμή, μια πρακτική που παρέχει πιο ακριβή εικόνα της απόδοσης των μοντέλων σε πραγματικές εφαρμογές. Ωστόσο, αυτή η προσέγγιση εισήγαγε επίσης προκλήσεις, καθώς τα μοντέλα χρειάστηκε να γενικεύσουν μεταξύ ατόμων με διαφορετική ανοχή στον πόνο και ποικίλες εκφράσεις προσώπου. Ο πόνος είναι μια υποκειμενική εμπειρία, και τα ίδια ερεθίσματα μπορούν να προκαλέσουν πολύ διαφορετικές αντιδράσεις μεταξύ ατόμων. Αυτή η ποικιλομορφία δυσκόλεψε τη γενίκευση, ιδιαίτερα για το μοντέλο MobileViT, το οποίο παρουσίασε χαμηλότερες επιδόσεις από τα CNN και CvT στη διαχείριση της ποικιλίας των συμμετεχόντων.

Ένας άλλος περιορισμός ήταν η ανισορροπία στο σύνολο δεδομένων, το οποίο επικεντρώθηκε κυρίως στα άκρα του πόνου (PA4 και BL1). Αυτή η ανισορροπία έκανε τα μοντέλα πιο αποτελεσματικά στην αναγνώριση των κατηγοριών "χωρίς πόνο", ενώ υπολειτουργούσαν στην ανίχνευση στιγμών έντονου πόνου. Για την αντιμετώπιση αυτού, οι μελλοντικές έρευνες θα μπορούσαν να εξετάσουν τεχνικές data augmentation, όπως η υπερδειγματοληψία (oversampling) ή η συνθετική δημιουργία ενδιάμεσων επιπέδων πόνου για τη δημιουργία ενός πιο ισορροπημένου συνόλου δεδομένων. Επιπλέον, η μετατροπή της εργασίας σε πρόβλημα παλινδρόμησης, όπου ο πόνος προβλέπεται σε μια συνεχή κλίμακα, θα μπορούσε να προσφέρει περισσότερες δυνατότητες εκτίμησης.

Ενώ οι εκφράσεις προσώπου αποτελούν πολύτιμο δείκτη πόνου, η αποκλειστική εξάρτηση από αυτές περιορίζει τη δυνατότητα των μοντέλων να καταγράψουν την πλήρη πολυπλοκότητα της εμπειρίας του πόνου. Πολυτροπικά δεδομένα, όπως φυσιολογικά σήματα (π.χ. καρδιακός ρυθμός, ηλεκτροδερμική δραστηριότητα, ηλεκτρομυογραφία), θα μπορούσαν να βελτιώσουν σημαντικά την ανθεκτικότητα των συστημάτων ανίχνευσης πόνου. Αυτό είναι ιδιαίτερα σημαντικό σε περιπτώσεις όπου οι εκφράσεις προσώπου καταστέλλονται ή απουσιάζουν, αλλά άλλοι φυσιολογικοί δείκτες μπορεί να υποδηλώνουν πόνο.

Συνολικά, οι περιορισμοί που προέκυψαν στην ανάλυση εικόνων αντικατοπτρίζουν την πολυπλοκότητα της εκτίμησης πόνου μέσα από εκφράσεις προσώπου. Παρά τις

προκλήσεις, τα αποτελέσματα ανέδειξαν την αξία της εφαρμογής τεχνικών μηχανικής μάθησης, ενώ υποδεικνύουν την ανάγκη για περαιτέρω βελτίωση των μεθόδων, συμπεριλαμβανομένων τεχνικών ενίσχυσης δεδομένων και ενσωμάτωσης πολυτροπικών δεδομένων.

6.2 Περιορισμοί και Προκλήσεις στην Ανάλυση Βίντεο

Η ανάλυση βίντεο για την εκτίμηση της έντασης πόνου παρουσίασε συγκεκριμένες προκλήσεις και περιορισμούς, που επηρέασαν τις επιδόσεις των μοντέλων. Καταρχάς, η προετοιμασία των δεδομένων βίντεο ήταν ιδιαίτερα απαιτητική. Αρχικά προβλήματα υπερπροσαρμογής απαίτησαν εκτεταμένες τροποποιήσεις, όπως αλλαγή του μεγέθους των δεδομένων, χρήση τεχνικών data augmentation και αρχιτεκτονικές προσαρμογές για τη βελτίωση της απόδοσης της μνήμης. Από τους 87 συμμετέχοντες που αρχικά περιλαμβάνονταν, οι 24 αποκλείστηκαν λόγω ασάφειας στις εκφράσεις του πόνου ή ασυνεπών απαντήσεων, περιορίζοντας το τελικό σύνολο σε 63 συμμετέχοντες. Αυτή η μείωση επηρέασε ιδιαίτερα την απόδοση μοντέλων που βασίζονται σε μετασχηματιστές.

Η ανίχνευση εκφράσεων στα βίντεο αποτέλεσε επίσης πρόκληση, καθώς οι αλγόριθμοι ανίχνευσης προσώπου δεν κατάφεραν πάντα να εντοπίσουν τις εκφράσεις, απαιτώντας συχνά χειροκίνητη παρέμβαση. Επιπλέον, τα χρονικά χαρακτηριστικά της έκφρασης του πόνου αποδείχθηκαν δύσκολο να απομονωθούν, καθώς έπρεπε να επιλεγούν προσεκτικά κρίσιμες στιγμές, γεγονός που οδήγησε ενδεχομένως σε απώλεια σημαντικών πληροφοριών.

Οι υπολογιστικοί περιορισμοί επηρέασαν σημαντικά τη δυνατότητα εξερεύνησης μοντέλων και παραμέτρων εκπαίδευσης. Το VideoMAE, αν και παρουσίασε την υψηλότερη ακρίβεια μεταξύ των μοντέλων, απαιτούσε 16,4 ώρες για εκπαίδευση 70 εποχών, γεγονός που απέκλεισε την εφαρμογή τεχνικών ρύθμισης υπερπαραμέτρων. Αντίστοιχα, οι χρόνοι εκπαίδευσης 3,5-4 ώρες για τα TSN και ConvLSTM περιόρισαν τον αριθμό των επαναλήψεων που μπορούσαν να εξεταστούν. Οι περιορισμοί αυτοί οδήγησαν στον αποκλεισμό πολλά υποσχόμενων αρχιτεκτονικών βασισμένων σε μετασχηματιστές και ensemble μεθόδων που θα μπορούσαν να βελτιώσουν την απόδοση.

Επιπλέον, η έλλειψη διαπολιτισμικής επικύρωσης, οι πιθανές προκαταλήψεις στην ερμηνεία εκφράσεων πόνου και η απουσία συνεκτίμησης της ατομικής ανοχής στον πόνο αποτελούν περιορισμούς που επηρέασαν τα αποτελέσματα. Οι μελλοντικές έρευνες θα πρέπει να αντιμετωπίσουν αυτούς τους περιορισμούς μέσω μεγαλύτερων, πιο ποικιλόμορφων συνόλων δεδομένων, ενσωματώνοντας πολυτροπικές προσεγγίσεις για την εκτίμηση του πόνου.

Κεφάλαιο 7

Συμπεράσματα και Μελλοντικές Κατευθύνσεις

7.1 Συμπεράσματα

Η παρούσα εργασία εστίασε στη χρήση προηγμένων μοντέλων βαθιάς μάθησης για την αυτόματη εκτίμηση της έντασης πόνου μέσω ανάλυσης εικόνων και βίντεο, παρέχοντας μια ολοκληρωμένη σύγκριση μεταξύ των μεθόδων. Τα μοντέλα VGG16 και MobileViT χρησιμοποιήθηκαν για την ανάλυση εικόνων, ενώ τα ConvLSTM, TSN και VideoMAE εφάρμοσαν την ανάλυση βίντεο. Τα αποτελέσματα αναδεικνύουν την αποτελεσματικότητα αυτών των προσεγγίσεων, ενώ παράλληλα υπογραμμίζουν τις προκλήσεις που σχετίζονται με τη γενίκευση, την υπολογιστική αποδοτικότητα και τη διαφοροποίηση των δεδομένων.

Το VGG16 κατέδειξε σταθερή απόδοση στην ανάλυση εικόνων, με ακρίβεια 55.8%, ενώ το MobileViT παρουσίασε καλύτερες επιδόσεις, φτάνοντας το 59.8%. Παρότι το MobileViT είναι ένα υβριδικό μοντέλο που συνδυάζει CNN και Vision Transformer, υπολείπεται σε απόδοση σε σύγκριση με πιο σύνθετες προσεγγίσεις. Στην ανάλυση βίντεο, το VideoMAE παρουσίασε την υψηλότερη ακρίβεια (71%), αλλά το υψηλό υπολογιστικό του κόστος το καθιστά λιγότερο πρακτικό για εφαρμογές πραγματικού χρόνου. Τα ConvLSTM και TSN, αν και λιγότερο ακριβή (66%), προσφέρουν ισορροπημένη απόδοση με μικρότερες απαιτήσεις πόρων.

Η χρήση αυστηρού διαχωρισμού συμμετεχόντων σε φάσεις εκπαίδευσης και δοκιμής αποτέλεσε βασικό χαρακτηριστικό αυτής της μελέτης, εξασφαλίζοντας

αντικειμενικότητα και αξιοπιστία στα αποτελέσματα. Ωστόσο, η ανισορροπία στα δεδομένα, που επικεντρώνονταν κυρίως στα άκρα του πόνου (PA4 και BL1), δημιούργησε προκλήσεις στην ανίχνευση ενδιάμεσων επιπέδων πόνου. Παρά τις προκλήσεις αυτές, τα αποτελέσματα υπογραμμίζουν τη δυναμική της χρήσης βαθιάς μάθησης για την εκτίμηση της έντασης πόνου.

7.2 Μελλοντικές Κατευθύνσεις

Οι μελλοντικές έρευνες θα πρέπει να επικεντρωθούν στη χρήση πολυτροπικών δεδομένων για την ενίσχυση της ακρίβειας και της ανθεκτικότητας των μοντέλων. Ο συνδυασμός εκφράσεων προσώπου και δεδομένων βίντεο με βιοσήματα, όπως καρδιακός ρυθμός και ηλεκτρομυογραφία, μπορεί να προσφέρει πιο ολοκληρωμένα και αξιόπιστα μοντέλα εκτίμησης πόνου, ιδιαίτερα σε περιπτώσεις όπου οι εκφράσεις προσώπου αποκρύπτονται ή είναι ανεπαρκείς. Επιπλέον, η ανάπτυξη ελαφρών μοντέλων που λειτουργούν σε πραγματικό χρόνο είναι κρίσιμη. Τεχνικές όπως η απλοποίηση μοντέλων (model pruning), η κβαντοποίηση (quantization) και η απόσταξη γνώσης (knowledge distillation) μπορούν να μειώσουν τις υπολογιστικές απαιτήσεις διατηρώντας υψηλή απόδοση.

Η διεύρυνση των δεδομένων εκπαίδευσης ώστε να περιλαμβάνουν διαφορετικά δημογραφικά στοιχεία, όπως εθνότητες, ηλικίες και φύλα, είναι ζωτικής σημασίας για τη μείωση των προκαταλήψεων και τη βελτίωση της γενίκευσης των μοντέλων. Τέλος, η ενσωμάτωση τεχνικών εξηγήσιμης τεχνητής νοημοσύνης (explainable AI) θα επιτρέψει στους κλινικούς γιατρούς να κατανοήσουν καλύτερα τις προβλέψεις των μοντέλων, ενισχύοντας την εμπιστοσύνη τους σε κλινικές εφαρμογές.

Η παρούσα μελέτη κατέδειξε τη δυναμική και των δύο προσεγγίσεων, της ανάλυσης εικόνων και βίντεο, ενώ επισήμανε τις περιοχές που απαιτούν βελτιώσεις. Οι μελλοντικές βελτιώσεις θα επικεντρωθούν στη μείωση του υπολογιστικού κόστους, την ενσωμάτωση πολυτροπικών δεδομένων και την ανάπτυξη γενικεύσιμων και ερμηνεύσιμων μοντέλων για την εκτίμηση της έντασης πόνου σε πραγματικές συνθήκες.

Βιβλιογραφία

[1] Adinma C, Gowher F, Kamalinejad E, Mercado J, Soni J, Zhong J. "A Survey of CNN and Facial Recognition Methods in the Age of COVID-19," Proc. ICISDM, Silicon Valley, CA, USA, pp. 11, 2021.

[2] Baltrusaitis T, Robinson P, Morency L-P. "OpenFace: An open source facial behavior analysis toolkit," 2016 IEEE Winter Conf. on Applications of Computer Vision (WACV), Lake Placid, NY, USA, pp. 1-4, 2016.

[3] Benavent-Lledo M, Mulero-Pérez D, Ortiz-Perez D, Rodriguez-Juan J, Berenguer-Agullo A, Psarrou A, Garcia-Rodriguez J. "A Comprehensive Study on Pain Assessment from Multimodal Sensor Data," Sensors, vol. 23, no. 24, 2023, doi: 10.3390/s23249675.

[4] Ben Aoun N. "A Review of Automatic Pain Assessment from Facial Information Using Machine Learning," Technologies, vol. 12, no. 6, 2024, doi: 10.3390/technologies12060092.

[5] Breiman L. "Random Forests," Machine Learning, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[6] Dragomir M-C, Florea C, Pupezescu V. "Automatic Subject Independent Pain Intensity Estimation using a Deep Learning Approach," 2020 Int. Conf. on e-Health and Bioengineering (EHB), Iasi, Romania, pp. 1-4, 2020, doi: 10.1109/EHB50910.2020.9280190.

[7] Gkikas S, Tachos NS, Andreadis S, Pezoulas VC, Zaridis D, Gkois G, Matonaki A, Stavropoulos TG, Fotiadis DI. "Multimodal automatic assessment of acute pain through facial videos and heart rate signals utilizing transformer-based architectures," Front. Pain Res., vol. 5, 1372814, 2024, doi: 10.3389/fpain.2024.1372814.

[8] Gkikas S, Tsiknakis M. "Automatic assessment of pain based on deep learning methods: A systematic review," Comput Methods Programs Biomed., vol. 231, 107365, 2023, doi: 10.1016/j.cmpb.2023.107365.

[9] Haefeli M, Elfering A. "Pain assessment," Eur Spine J., vol. 15, Suppl. 1, pp. S17-24, 2006, doi: 10.1007/s00586-005-1044-x.

[10] He K, Zhang X, Ren S, Sun J. "Deep Residual Learning for Image Recognition," 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778, 2016, doi: 10.1109/CVPR.2016.90.

[11] https://www.geeksforgeeks.org/vgg-16-cnn-model/.

[12] https://keras.io/examples/vision/mobilevit/.

[13] https://www.nit.ovgu.de/nit/en/BioVid-p-1358.html.

[14] https://openreview.net/pdf?id=vh-0sUt8HlG.

[15] https://pypi.org/project/mediapipe/.

[16] https://arxiv.org/abs/1506.04214.

[17] https://arxiv.org/abs/1608.00859.

[18] https://huggingface.co/MCG-NJU/videomae-base-short-ssv2.

[19] https://huggingface.co/docs/transformers/main/model_doc/videomae

[20] https://lekhuyen.medium.com/an-overview-of-vgg16-and-nin-models-96e4bf398484

[21] https://www.sciencedirect.com/science/article/abs/pii/S0888327022010172

[22] Matsangidou M, Liampas A, Pittara M, Pattichi CS, Zis P. "Machine Learning in Pain Medicine: An Up-To-Date Systematic Review," Pain Ther., vol. 10, no. 2, pp. 1067-1084, 2021, doi: 10.1007/s40122-021-00324-2.

[23] Mehta S, Rastegari M. "Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer," arXiv preprint, arXiv:2110.02178, 2021.

[24] Niese R, Werner P, Al-Hamadi A. "Accurate, fast and robust realtime face pose estimation using Kinect camera," Proc. IEEE Int. Conf. Systems, Man, Cybern., pp. 487–490, 2013.

[25] Nichol N. "Image Classification," GitHub repository, ecbc2b2, 2023. [Online]. Available: https://github.com/nicknochnack/ImageClassification/tree/ecbc2b219606c4153bdec47cd07f6ab00652c6d6.

[26] Othman E, Werner P, Saxen F, Al-Hamadi A, Walter S. "Cross-Database Evaluation of Pain Recognition from Facial Video," 2019 11th Int. Symp. on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, pp. 181-186, 2019, doi: 10.1109/ISPA.2019.8868562.

[27] Patania S, Boccignone G, Buršić S, D'Amelio A, Lanzarotti R. "Deep graph neural network for video-based facial pain expression assessment," Proc. of the 37th ACM/SIGAPP Symp. on Applied Computing, New York, NY, USA, pp. 585–591, 2022, doi: 10.1145/3477314.3507094.

[28] Price DD. Psychological mechanisms of pain and analgesia. Seattle: IASP Press; 1999.

[29] PyTorch, "Torchvision Models," PyTorch Documentation, version 0.9, 2021. [Online]. Available: https://pytorch.org/vision/0.9/models.html.

[30] PyTorch, "Vision Transformer (ViT_B_16)," PyTorch Documentation, 2023. [Online]. Available: https://pytorch.org/vision/main/models/generated/torchvision.models.vit_b_16.html.

[31] Subramaniam S. Devi, Dass B. "Automatic Pain Severity Level Classification Using The Convolutional Neural Network," Annals of the Romanian Society for Cell Biology, pp. 20878–20885, 2021.

[32] Tavakolian M, Hadid A. "A Spatiotemporal Convolutional Neural Network for Automatic Pain Intensity Estimation from Facial Dynamics," Int J Comput Vis, vol. 127, pp. 1413–1425, 2019, doi: 10.1007/s11263-019-01191-3.

[33] TensorFlow. ImageDataGenerator, available: https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator.

[34] TensorFlow, "Video classification tutorial with TensorFlow," TensorFlow, 2023. [Online]. Available: https://www.tensorflow.org/tutorials/video/video_classification.

[35] Walter S et al., "The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system," 2013 IEEE Int. Conf. on Cybernetics (CYBCO), Lausanne, Switzerland, pp. 128-131, 2013, doi: 10.1109/CYBConf.2013.6617456.

[36] Werner P, Al-Hamadi A, Limbrecht-Ecklundt K, Walter S, Gruss S, Traue HC. "Automatic Pain Assessment with Facial Activity Descriptors," IEEE Trans. Affective Computing, vol. 8, no. 3, pp. 286-299, 2017, doi: 10.1109/TAFFC.2016.2537327.

[37] Werner P, Al-Hamadi A, Walter S. "Analysis of facial expressiveness during experimentally induced heat pain," 2017 Seventh Int. Conf. on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), San Antonio, TX, USA, pp. 176–180, 2017.

[38] Werner P, Al-Hamadi A, Niese R, Walter S, Gruss S, Traue HC. "Automatic Pain Recognition from Video and Biomedical Signals," 2014 22nd Int. Conf. on Pattern Recognition, Stockholm, Sweden, pp. 4582-4587, 2014, doi: 10.1109/ICPR.2014.784.

[39] Wu H et al., "CvT: Introducing Convolutions to Vision Transformers," 2021 IEEE/CVF Int. Conf. on Computer Vision (ICCV), Montreal, QC, Canada, pp. 22-31, 2021, doi: 10.1109/ICCV48922.2021.00009.

[40] Xin X, Li X, Yang S, Lin X, Zheng X. "Pain expression assessment based on a locality and identity aware network," IET Image Processing, vol. 15, no. 12, pp. 2948-2958, 2021, doi: 10.1049/ipr2.12282.

[41] Xiong X, De la Torre F. "Supervised descent method and its applications to face alignment," IEEE Computer Vision and Pattern Recognition, 2013.

[42] Yamashita R, Nishio M, Do RKG, et al. "Convolutional neural networks: an overview and application in radiology," Insights Imaging, vol. 9, pp. 611–629, 2018, doi: 10.1007/s13244-018-0639-9.

[43] Zheng J, Lin Y. "An Objective Pain Measurement Machine Learning Model through Facial Expressions and Physiological Signals," 2022 28th Int. Conf. on Mechatronics and Machine Vision in Practice (M2VIP), Nanjing, China, pp. 1-4, 2022, doi: 10.1109/M2VIP55626.2022.10041105.

[44] S. M. Nugent, T. I. Lovejoy, S. Shull, S. K. Dobscha, and B. J. Morasco, "Associations of Pain Numeric Rating Scale Scores Collected During Usual Care with Research Administered Patient-Reported Pain Outcomes," Pain Med., vol. 22, no. 10, pp. 2235–2241, Oct. 2021, doi: 10.1093/pm/pnab110.

[45] J.-X. An, Y. Wang, D. K. Cope, and J. P. Williams, "Quantitative Evaluation of Pain with Pain Index Extracted from Electroencephalogram," *Chin. Med. J.*, vol. 130, no. 16, pp. 1926–1931, Aug. 2017, doi: 10.4103/0366-6999.211878.

[46] R. Rojo, J. C. Prados-Frutos, and A. López-Valverde, "Pain assessment using the Facial Action Coding System: A systematic review," *Med. Clin. (Barc.)*, vol. 145, no. 8, pp. 350–355, Aug. 2015, doi: 10.1016/j.medcle.2014.08.002.

[47] Z. Hammal and J. F. Cohn, "Automatic detection of pain intensity," *Proc. ACM Int. Conf. Multimodal Interact.*, vol. 2012, pp. 47–52, Oct. 2012, doi: 10.1145/2388676.2388688.

[48] Otto von Guericke University Magdeburg, "BioVid Pain Database," [Online]. Available: https://www.nit.ovgu.de/BioVid.html. [Accessed: Nov. 18, 2024].

Παράρτημα Α

Ο παρακάτω κώδικας παρέχει μια αυτοματοποιημένη μέθοδο για την ανίχνευση προσώπων και την επεξεργασία βίντεο, εστιάζοντας στην περικοπή και μεγέθυνση της περιοχής του προσώπου.

```
import cv2
import os

# Initialize face detector
face_cascade = cv2.CascadeClassifier(cv2.data.haarcascades +
'haarcascade_frontalface_default.xml')

def crop_and_zoom_face(frame, face_rect, zoom_factor=1.5):
    x, y, w, h = face_rect

    # Calculate the center of the face
    cx, cy = x + w // 2, y + h // 2

    # Calculate new width and height
    new_w, new_h = int(w * zoom_factor), int(h * zoom_factor)

    # Calculate new top-left corner
    new_x, new_y = max(0, cx - new_w // 2), max(0, cy - new_h // 2)

    # Ensure the new bottom-right corner doesn't go out of the image
boundaries
    new_x2, new_y2 = min(frame.shape[1], new_x + new_w),
min(frame.shape[0], new_y + new_h)

    # Crop the face with zoom
    cropped_face = frame[new_y:new_y2, new_x:new_x2]
    return cropped_face

def process_video(input_path, output_path, zoom_factor=1.5, start_time=2,
end_time=4.5):
    try:
        print(f'Starting to process {input_path}')
        cap = cv2.VideoCapture(input_path)
        if not cap.isOpened():
            print(f'Error: Could not open video file {input_path}')
            return

        fps = cap.get(cv2.CAP_PROP_FPS)
        frame_width = int(cap.get(cv2.CAP_PROP_FRAME_WIDTH))
```

```
        frame_height = int(cap.get(cv2.CAP_PROP_FRAME_HEIGHT))
        total_frames = int(cap.get(cv2.CAP_PROP_FRAME_COUNT))
        print(f'Video FPS: {fps}, Frame Width: {frame_width}, Frame
Height: {frame_height}, Total Frames: {total_frames}')

        fourcc = cv2.VideoWriter_fourcc(*'mp4v')  # Using mp4v codec
        out = None

        start_frame = int(start_time * fps)
        end_frame = int(end_time * fps)
        print(f'Trimming video from frame {start_frame} to {end_frame}')

        frame_count = 0
        written_frame_count = 0
        face_rect = None

        while frame_count < start_frame:
            ret, frame = cap.read()
            if not ret:
                print(f'Error: Could not read frame {frame_count}')
                break
            frame_count += 1

        while frame_count < end_frame:
            ret, frame = cap.read()
            if not ret:
                break

            if face_rect is None:
                # Detect face in the first frame within the trimming
range
                gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
                faces = face_cascade.detectMultiScale(gray,
scaleFactor=1.1, minNeighbors=5, minSize=(30, 30))
                if len(faces) > 0:
                    face_rect = faces[0]
                    print(f'Face detected at {face_rect}')
                else:
                    print(f'No face detected in the first frame of
{input_path}')
                    break

            cropped_face = crop_and_zoom_face(frame, face_rect,
zoom_factor)
            if out is None:
                height, width = cropped_face.shape[:2]
                out = cv2.VideoWriter(output_path, fourcc, fps, (width,
height))
```

```
                print(f'Initialized video writer for {output_path} with
frame size {width}x{height} at {fps} FPS')
            out.write(cropped_face)
            written_frame_count += 1
            frame_count += 1

        cap.release()
        if out is not None:
            out.release()

        if frame_count == 0:
            print(f'No frames were read from {input_path}')
        elif written_frame_count == 0:
            print(f'No frames were written for {input_path}')
        else:
            print(f'Successfully processed {input_path} with
{written_frame_count} frames written out of {frame_count} total frames.')
    except Exception as e:
        print(f'Error processing video {input_path}: {e}')

def process_videos_in_folder(input_folder, output_folder,
zoom_factor=1.5, start_time=2, end_time=4.5):
    try:
        if not os.path.exists(output_folder):
            os.makedirs(output_folder)
            print(f'Created output folder: {output_folder}')
        else:
            print(f'Output folder already exists: {output_folder}')

        for filename in os.listdir(input_folder):
            print(f'Found file: {filename}')
            if filename.endswith(('.mp4', '.avi', '.mov', '.mkv')):
                input_path = os.path.join(input_folder, filename)
                output_path = os.path.join(output_folder, filename)
                print(f'Processing {input_path}...')
                process_video(input_path, output_path, zoom_factor,
start_time, end_time)
            else:
                print(f'Skipping non-video file: {filename}')
    except Exception as e:
        print(f'Error processing folder {input_folder}: {e}')

input_folder = 'VideoDataset'
output_folder = 'VideoDatasetFaceDetected'

process_videos_in_folder(input_folder, output_folder)
```

Παράρτημα Β

Ο παρακάτω κώδικας εξάγει όλα τα καρέ από βίντεο που ανήκουν σε προκαθορισμένες κατηγορίες και αποθηκεύει κάθε καρέ σε έναν συγκεκριμένο φάκελο. Αυτή η διαδικασία είναι χρήσιμη για την προεπεξεργασία δεδομένων.

```
import os
import cv2

def extract_all_frames_from_video(video_path, output_dir, video_name):
    os.makedirs(output_dir, exist_ok=True)
    cap = cv2.VideoCapture(video_path)
    frame_id = 0
    while cap.isOpened():
        ret, frame = cap.read()
        if not ret:
            break
        frame_filename = os.path.join(output_dir,
f"{video_name}_frame{frame_id}.jpg")
        cv2.imwrite(frame_filename, frame)
        frame_id += 1
    cap.release()

def process_videos(base_path, output_base_path, categories):
    for category in categories:
        videos_path = os.path.join(base_path, category)
        output_category_path = os.path.join(output_base_path, category)
        os.makedirs(output_category_path, exist_ok=True)

        for video_file in os.listdir(videos_path):
            video_path = os.path.join(videos_path, video_file)
            video_name = os.path.splitext(video_file)[0]
            output_video_path = os.path.join(output_category_path,
video_name)

            extract_all_frames_from_video(video_path, output_video_path,
video_name)
            print(f"Processed {video_file} into {output_video_path}")

# Define paths
base_path =
'/trinity/home/kandreou/biovid/data/BioVidDatasetProcessed/PartADatasetPr
ocessed'
```

```
output_base_path =
'/trinity/home/kandreou/biovid/data/BioVidDatasetProcessed/PartAProcessed
Frames'
categories = ['BL1', 'PA1', 'PA2', 'PA3', 'PA4']

# Process videos
process_videos(base_path, output_base_path, categories)

print("Frame extraction complete.")
```

Παράρτημα Γ

Ο παρακάτω κώδικας δημιουργεί δέκα τυχαία σύνολα δεδομένων (random datasets) από εικόνες προσώπων για τις κατηγορίες "no_pain" και "pain". Κάθε dataset περιλαμβάνει υποσύνολα εκπαίδευσης (train) και δοκιμών (test), με αυστηρό διαχωρισμό των συμμετεχόντων.

```
import os
import shutil
import random
from pathlib import Path

# Define the base directory that already contains the images
destination_base_dir =
'/trinity/home/kandreou/biovid/data/BioVidDatasetProcessed/FaceFrames_256
x256'

# Define the categories
categories = ['no_pain', 'pain']

# Function to get unique participant IDs
def get_participant_ids(directory):
    ids = set()
    for img_name in os.listdir(directory):
        participant_id = img_name.split('_')[0]
        ids.add(participant_id)
    return list(ids)

# Create the 10 random datasets
for i in range(1, 11):
    random_dir =
f'/trinity/home/kandreou/biovid/data/BioVidDatasetProcessed/FaceFrames_25
6x256_random{i}'
    os.makedirs(random_dir, exist_ok=True)

    for category in categories:
        src_path = Path(destination_base_dir) / category
        participant_ids = get_participant_ids(src_path)
        random.shuffle(participant_ids)

        split_point = int(0.795 * len(participant_ids))
        train_ids = set(participant_ids[:split_point])
        test_ids = set(participant_ids[split_point:])
```

```
        for img_name in os.listdir(src_path):
            participant_id = img_name.split('_')[0]
            src_img_path = src_path / img_name
            if participant_id in train_ids:
                dst_path = Path(random_dir) / 'train' / category
            else:
                dst_path = Path(random_dir) / 'test' / category
            dst_path.mkdir(parents=True, exist_ok=True)
            dst_img_path = dst_path / img_name
            shutil.copyfile(src_img_path, dst_img_path)

print("Random datasets created successfully!")
```

Παράρτημα Δ

Ο παρακάτω κώδικας δημιουργεί μια αναφορά δεδομένων για ένα σύνολο δεδομένων εικόνων προσώπων, οργανωμένο σε φακέλους train και test, με κατηγορίες no_pain και pain.

```
import os
from collections import defaultdict

def scan_directory(base_path):
    # Initialize dictionaries for each dataset type
    dataset_info = {
        'train': {'no_pain': defaultdict(int), 'pain': defaultdict(int)},
        'test': {'no_pain': defaultdict(int), 'pain': defaultdict(int)}
    }

    # Traverse each dataset type and category
    for dataset_type in ['train', 'test']:
        for category in ['no_pain', 'pain']:
            category_path = os.path.join(base_path, dataset_type,
category)

            # Verify if the directory exists
            if not os.path.exists(category_path):
                print(f"Directory {category_path} does not exist.")
                continue

            # Loop through each file in the category directory
            for file in os.listdir(category_path):
                if file.endswith(".png"):
                    # Extract participant ID from filename
                    participant_id = file.split('_')[0]
                    # Increment count for this ID in the appropriate
category
                    dataset_info[dataset_type][category][participant_id]
+= 1

    return dataset_info

def aggregate_totals(dataset_info):
    total_counts = defaultdict(int)
```

```
    # Aggregate total image counts across all datasets
    for dataset in dataset_info.values():
        for category in dataset.values():
            for participant_id, count in category.items():
                total_counts[participant_id] += count

    return total_counts

def generate_report(dataset_info, output_file):
    with open(output_file, 'w') as f:
        # Write individual dataset reports
        for dataset in dataset_info:
            f.write(f"{dataset.capitalize()} dataset:\n")
            for category in dataset_info[dataset]:
                # Get sorted list of participant IDs
                sorted_ids =
sorted(dataset_info[dataset][category].keys())

                f.write(f" {category.replace('_', ' ')} total ids:
{len(sorted_ids)}\n")

                # Write each ID and its image count
                for participant_id in sorted_ids:
                    count =
dataset_info[dataset][category][participant_id]
                    f.write(f"  {participant_id} total images:
{count}\n")
            f.write("\n")

        # Calculate and print total counts
        total_counts = aggregate_totals(dataset_info)
        sorted_total_ids = sorted(total_counts.keys())

        f.write("Overall participant totals:\n")
        f.write(f" Total unique ids: {len(sorted_total_ids)}\n")
        for participant_id in sorted_total_ids:
            f.write(f"  {participant_id} total images:
{total_counts[participant_id]}\n")

    print(f"Report generated: {output_file}")

def main():
    # Path to the dataset directory
    dataset_path =
'/trinity/home/kandreou/biovid/data/BioVidDatasetProcessed/FaceFrames_256
x256_random1'
    # Output text file
    output_txt = 'dataset_report.txt'
```

```
    # Scan the directory and gather information
    dataset_info = scan_directory(dataset_path)

    # Generate the report
    generate_report(dataset_info, output_txt)

if __name__ == "__main__":
    main()
```

Παράρτημα Ε

Ο παρακάτω κώδικας ελέγχει αν υπάρχουν ανιχνεύσιμα πρόσωπα σε βίντεο που είναι αποθηκευμένα σε διαφορετικές κατηγορίες (π.χ., BL1, PA1, PA2, PA3, PA4). Χρησιμοποιεί το Haar Cascade της OpenCV για ανίχνευση προσώπων και καταγράφει τα βίντεο όπου δεν ανιχνεύτηκαν πρόσωπα σε ένα αρχείο καταγραφής.

```
import os
import cv2

# Define the paths
video_dir =
'/trinity/home/kandreou/biovid/data/BioVidDatasetProcessed/PartADatasetPr
ocessed'
log_file = 'No_Face_Detected.txt'

# Define the categories
categories = ['BL1', 'PA1', 'PA2', 'PA3', 'PA4']

# Load the Haar cascade for face detection
face_cascade = cv2.CascadeClassifier(cv2.data.haarcascades +
'haarcascade_frontalface_default.xml')

# Function to check for faces in a video
def detect_faces_in_video(video_path):
    cap = cv2.VideoCapture(video_path)
    face_detected = False
    while cap.isOpened():
        ret, frame = cap.read()
        if not ret:
            break
        gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
        faces = face_cascade.detectMultiScale(gray, scaleFactor=1.1,
minNeighbors=5, minSize=(30, 30))
        if len(faces) > 0:
            face_detected = True
            break
    cap.release()
    return face_detected

# Open the log file
with open(log_file, 'w') as log:
    # Iterate over each category
    for category in categories:
        category_dir = os.path.join(video_dir, category)
```

```
        # Iterate over each video in the category
        for video_file in os.listdir(category_dir):
            video_path = os.path.join(category_dir, video_file)
            print(f"Processing {video_path}")
            if not detect_faces_in_video(video_path):
                log.write(f"{video_path}\n")
                print(f"No face detected in {video_path}")

print("Face detection completed.")
```

Παράρτημα ΣΤ

Ο παρακάτω κώδικας είναι κρίσιμος για την επαλήθευση ότι τα βίντεο έχουν υποστεί σωστή επεξεργασία, όπως η χρήση του κατάλληλου χρονικού περιθωρίου (απο το 2ο δευτερόλεπτο εώς το 4.5). Με τον έλεγχο του πλήθους των καρέ σε κάθε βίντεο, διασφαλίζεται ότι τα δεδομένα πληρούν τις προδιαγραφές που τέθηκαν.

```
import cv2
import os

def check_and_delete_videos(folder,
log_file='wrong_video_processing.txt', expected_frames=62):
    wrong_videos = []

    for root, dirs, files in os.walk(folder):
        for filename in files:
            if filename.endswith(('.mp4', '.avi', '.mov', '.mkv')):
                video_path = os.path.join(root, filename)
                print(f'Checking video: {video_path}')
                cap = cv2.VideoCapture(video_path)
                if not cap.isOpened():
                    print(f'Error: Could not open video file
{video_path}')
                    continue

                frame_count = int(cap.get(cv2.CAP_PROP_FRAME_COUNT))
                cap.release()

                if frame_count != expected_frames:
                    wrong_videos.append(video_path)
                    os.remove(video_path)
                    print(f'Deleted video: {video_path}')

    with open(log_file, 'w') as f:
        f.write('Deleted videos are:\n')
        for video in wrong_videos:
            f.write(f'{video}\n')

    if wrong_videos:
        print(f'Videos with incorrect frame count have been logged and
deleted. See {log_file}')
    else:
        print('All videos have the expected frame count.')

output_folder = 'PartATrimmedAndZoom'

check_and_delete_videos(output_folder)
```

Παράρτημα Ζ

Ο παρακάτω κώδικας υλοποιεί τη δημιουργία **10 τυχαίων συνόλων δεδομένων βίντεο** με αναλογίες εκπαίδευσης και δοκιμών (80% - 20%). Τα βίντεο κατηγοριοποιούνται στις κατηγορίες no_pain και pain, βασισμένα στις κατηγορίες BL1 και PA4. Κατά την επεξεργασία, τα βίντεο **αναπροσαρμόζονται σε ανάλυση 128x128 pixel**, με τήρηση αποκλεισμένων συμμετεχόντων και λεπτομερή καταγραφή της διαδικασίας.

```
import os
import shutil
import random
from pathlib import Path
from moviepy.editor import VideoFileClip

# Define the base directory that contains the videos
destination_base_dir =
'/trinity/home/kandreou/biovid/data/BioVidDatasetProcessed/PartADatasetPr
ocessed'

# Define the category mapping
category_mapping = {
    'BL1': 'no_pain',
    'PA4': 'pain'
}

# File paths
log_file_path =
'/trinity/home/kandreou/biovid/data/BioVidDatasetProcessed/EP_Videos224x2
24/failed_videos_log.txt'
train_test_log_file =
'/trinity/home/kandreou/biovid/data/BioVidDatasetProcessed/EP_Videos224x2
24/train_test_ids_log.txt'
excluded_participants_file =
'/trinity/home/kandreou/biovid/data/BioVidDatasetProcessed/ExcludedPartic
ipants/ExcludedParticipants.txt'

# Clear previous logs
with open(log_file_path, 'w') as log_file:
    log_file.write("Failed videos log:\n")

# Clear previous train/test ID logs
with open(train_test_log_file, 'w') as id_log_file:
    id_log_file.write("Train/Test Participants Log:\n\n")

# Function to get unique participant IDs
def get_participant_ids(directory):
```

```
    ids = set()
    for video_name in os.listdir(directory):
        if video_name.endswith('.mp4'):  # Only consider .mp4 files
            participant_id = video_name.split('_')[0]
            ids.add(participant_id)
    return list(ids)

# Function to resize a video and save it
def resize_video(input_path, output_path, new_size=(128, 128)):
    try:
        with VideoFileClip(str(input_path)) as video:
            resized_video = video.resize(new_size)
            resized_video.write_videofile(str(output_path),
codec='libx264', audio_codec='aac')
    except Exception as e:
        # Log error to the file
        with open(log_file_path, 'a') as log_file:
            log_file.write(f"Failed to resize or copy video:
{input_path}. Error: {e}\n")
        print(f"Error resizing video {input_path}: {e}")

# Function to log the participant IDs used for training and testing
def log_train_test_ids(train_ids, test_ids, log_file):
    with open(log_file, 'a') as id_log_file:
        id_log_file.write(f"Participants used for Train: {',
'.join(train_ids)}\n")
        id_log_file.write(f"Total for Train IDs: {len(train_ids)}\n\n")
        id_log_file.write(f"Participants used for Test: {',
'.join(test_ids)}\n")
        id_log_file.write(f"Total for Test IDs: {len(test_ids)}\n\n")

# Function to read excluded participants from the file
def read_excluded_participants(file_path):
    with open(file_path, 'r') as f:
        excluded_ids = f.read().splitlines()
    return set(excluded_ids)

# Load excluded participants
excluded_participants =
read_excluded_participants(excluded_participants_file)

# Create the 10 random datasets
for i in range(6, 11):
    random_dir =
f'/trinity/home/kandreou/biovid/data/BioVidDatasetProcessed/EP_Videos224x
224/Videos__random{i}'
    os.makedirs(random_dir, exist_ok=True)
```

```
    # Collect participant IDs for both categories (BL1 and PA4)
    all_participant_ids = set()
    for folder in category_mapping:
        src_path = Path(destination_base_dir) / folder
        all_participant_ids.update(get_participant_ids(src_path))

    # Remove excluded participants from the list
    all_participant_ids = [pid for pid in all_participant_ids if pid not
in excluded_participants]

    # Shuffle the participant IDs once for the current dataset
    random.shuffle(all_participant_ids)

    # 80% for training, 20% for testing
    split_point = int(0.8 * len(all_participant_ids))
    train_ids = set(all_participant_ids[:split_point])
    test_ids = set(all_participant_ids[split_point:])

    # Log the train and test IDs
    log_train_test_ids(list(train_ids), list(test_ids),
train_test_log_file)

    # Process each category (BL1 and PA4) while using the same train/test
split
    for folder, category in category_mapping.items():
        src_path = Path(destination_base_dir) / folder
        for video_name in os.listdir(src_path):
            if video_name.endswith('.mp4'):  # Only process .mp4 files
                participant_id = video_name.split('_')[0]

                # Skip if participant is excluded or not in train/test
set
                if participant_id not in train_ids and participant_id not
in test_ids:
                    continue

                src_video_path = src_path / video_name
                if participant_id in train_ids:
                    dst_path = Path(random_dir) / 'train' / category
                else:
                    dst_path = Path(random_dir) / 'test' / category
                dst_path.mkdir(parents=True, exist_ok=True)
                dst_video_path = dst_path / video_name

                # Resize and save the video to the destination path
                resize_video(src_video_path, dst_video_path)

print("Random resized video datasets created successfully!")
```