**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**
**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

# Αναγνώριση ανθρώπινης δραστηριότητας από Δεδομένα Αισθητήρων Κίνησης με χρήση Αρχιτεκτονικών Νευρωνικών Δικτύων Προσοχής

**Νικόλαος Κ. Κουτσάκης**

**ΑΘΗΝΑ**


**NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS**

**SCHOOL OF SCIENCES**
**DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS**

**PROGRAM OF POSTGRADUATE STUDIES**


**MSc THESIS**

# Human Activity Recognition by Motion Sensor Data using Neural Networks of Attention Architecture


**Nikolaos K. Koutsakis**


**ATHENS**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

Αναγνώριση ανθρώπινης δραστηριότητας από Δεδομένα Αισθητήρων Κίνησης με χρήση Αρχιτεκτονικών Νευρωνικών Δικτύων Προσοχής

**Νικόλαος Κ. Κουτσάκης**
**Α.Μ.:** cs1180006

**ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: Περαντώνης Στάυρος**, Διευθυντής Έρευνας Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών Εθνικό Κέντρο Επιστημονικής Έρευνας "ΔΗΜΟΚΡΙΤΟΣ"

**ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ:**

**Περαντώνης Στάυρος**, Διευθυντής Έρευνας Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών Εθνικό Κέντρο Επιστημονικής Έρευνας "ΔΗΜΟΚΡΙΤΟΣ"

**Σπύρου Ευάγγελος**, Επίκουρος Καθηγητής Τμήμα Πληροφορικής και Τηλεπικοινωνιών Πανεπιστήμιο Θεσσαλίας

**Σταματόπουλος Παναγιώτης**, Επίκουρος Καθηγητής Τμήμα Πληροφορικής και Τηλεπικοινωνιών Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

**Ιούνιος 2023**

**MSc THESIS**

Human Activity Recognition by Motion Sensor Data using Neural Networks of Attention Architecture

**Nikolaos K. Koutsakis**
**A.M.:** cs1180006

**SUPERVISOR: Perantonis Stavros**, Research Director Institute of Informatics and Telecommunications National Center for Scientific Research ”DEMOKRITOS”

**THREE-MEMBER ADVISORY COMMITTEE:**

**Perantonis Stavros**, Research Director Institute of Informatics and Telecommunications National Center for Scientific Research ”DEMOKRITOS”

**Spyrou Evaggelos**, Assistant Professor Department of Informatics and Telecommunications University of Thessaly

**Stamatopoulos Panagiotis**, Assistant Professor Department of Informatics and Telecommunication National and Kapodistrian University of Athens

**June 2023**

# ΠΕΡΙΛΗΨΗ

Η αναγνώριση ανθρώπινης δραστηριότητας έχει απασχολήσει αισθητά το ερευνητικό ενδιαφέρον την τελευταία δεκαετία. Συγκεκριμένα, η ταξινόμηση χρονοσειρών με δεδομένα από αισθητήρες κίνησης αποτελεί τον πυρήνα για αρκετές έρευνες οι οποίες κυρίως χρησιμοποιούν βαθιά νευρωνικά δίκτυα συνέλιξης και ανατροφοδότησης. Η χρήση τέτοιων δικτύων όμως δεν φαίνεται να είναι επαρκής για μεγάλου μήκους ακολουθίες, καθώς τα χαρακτηριστικά που μαθαίνονται στα αρχικά στάδια δεν διατηρούνται, με αποτέλεσμα την απώλεια πληροφορίας. Η εμφάνιση νευρωνικών δικτύων προσοχής όμως, παρουσιάζει την ικανότητα να διαχειρίζεται τέτοιες αδυναμίες και με κύριο αντιπρόσωπο το μοντέλο βαθιάς μάθησης Transformer, όπως ονομάζεται, να πετυχαίνει υψηλές επιδόσεις σε προβλήματα επεξεργασίας φυσικής γλώσσας και computer vision, καθιστώντας αναπόφευκτη την εφαρμογή του και σε άλλους τομείς, όπως η ταξινόμηση χρονοσειρών. Στην έρευνα που ακολουθεί, αναλύεται σχολαστικά ο μηχανισμός προσοχής που αποτελεί θεμέλιο του μοντέλου, καθώς και η εφαρμογή του σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας με δεδομένα χρονοσειρών από αισθητήρες κίνησης. Συγκρίνεται με μοντέλα νευρωνικών δικτύων συνέλιξης και ανατροφοδότησης παρουσιάζοντας καλύτερα αποτελέσματα στην ταξινόμηση δραστηριοτήτων και τέλος εξετάζεται κατά πόσο είναι επαρκής για την αποτελεσματική επίλυση τέτοιων προβλημάτων.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ**: Αναγνώριση Ανθρώπινης Δραστηριότητας

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ**: μηχανισμός προσοχής, ακολουθία-προς-ακολουθία, χρονοσειρές, αισθητήρες κίνησης, ταξινόμηση δραστηριοτήτων

# ABSTRACT

Human activity recognition has attracted considerable research interest in the last decade. In particular, the classification of time series with motion sensor data is the core of several researches which mainly use deep convolution and recurrent neural networks. However, the use of such networks does not seem to be sufficient for long sequences, as the features learned at the initial stages are not preserved, resulting in information loss. The emergence of attention neural networks, however, shows the ability to handle such weaknesses and, with the deep learning Transformer model, as it is called, as its main representative, to achieve high performance in natural language and computer vision processing problems, making its application in other areas, such as time series classification, inevitable. In the following research, the attention mechanism that is the foundation of the model is thoroughly analyzed, as well as its application to problems of recognizing human activity with time series data from motion sensors. It is compared with convolution and recurrent neural networks models showing better results in activity classification and finally it is examined whether it is sufficient to effectively solve such problems.



**SUBJECT AREA**: Human Activity Recognition

**KEYWORDS**: attention mechanism, sequence-to-sequence, timeseries, motion sensors, activities classification

# ΠΕΡΙΕΧΟΜΕΝΑ

# ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

# ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

# 1. ΕΙΣΑΓΩΓΗ

Οι ανθρώπινες δραστηριότητες έχουν χρησιμοποιηθεί εκτενέστερα για τον ορισμό ανθρώπινων πρότυπων συμπεριφοράς και εδώ, με τον όρο δραστηριότητα εννοείται η κίνηση ολόκληρου του ανθρώπινου σώματος ή διαφορετικών θέσεων που αυτό παίρνει σε σχέση με τον χρόνο και ενάντια στη βαρύτητα. Η αναγνώριση ανθρώπινης δραστηριότητας για ταξινόμηση χρονοσειρών, ουσιαστικά, στοχεύει στην αναγνώριση της ανθρώπινης συμπεριφοράς με την χρήση δεδομένων από αισθητήρες οι οποίοι βρίσκονται σε φορητές συσκευές όπως smartphones, tablets ή smartwatches. Η μελέτη για την φύση της δραστηριότητας και οι πληροφορίες που απορρέουν, προέρχονται από αυτούς τους αισθητήρες. Οι αισθητήρες αδράνειας ή οι αισθητήρες κίνησης περιλαμβάνονται σε αυτές τις συσκευές και χρησιμοποιούνται για την αναγνώριση φυσικών δραστηριοτήτων. Το επιταχυνσιόμετρο είναι ένας αισθητήρας ο οποίος συλλέγει το σήμα της επιτάχυνσης ενός σώματος και είναι χρήσιμος στην αναγνώριση δραστηριοτήτων όπως το περπάτητα, το τρέξιμο ή τα άλματα. Το γυροσκόπιο είναι ένας αισθητήρας ο οποίος συλλέγει τις κινήσεις περιστροφής ενός σώματος και χρησιμεύει στην αναγνώριση δραστηριοτήτων όπως, αιώρηση, περιστροφή και επανατοποθέτηση.

Τα δεδομένα σημάτων που προέρχονται από αυτές τις συσκευές ταξινομούνται στις αντίστοιχες δραστηριότητες με την χρήση μεθόδων μηχανικής μάθησης. Συνεπώς, η αναγνώριση ανθρώπινης δραστηριότητας επιλύει προβλήματα τα οποία έχουν αντιμετωπιστεί ως τυπικά προβλήματα αναγνώρισης μοτίβων και ειδικότερα, προβλήματα ταξινόμησης, δηλαδή, την αναγνώριση της δραστηριότητας που εκτελείται από ένα άτομο σε μια δεδομένη στιγμή. Οι περισσότερες λύσεις έχουν αναπτυχθεί χρησιμοποιώντας μεθόδους τεχνητής νοημοσύνης, μέσω διάφορων τεχνικών μηχανικής μάθησης, οι οποίες περιλαμβάνουν ρηχούς (π.χ. Support Vector Machines (SVM), Naive Bayes και K-Nearest Neighbors) και βαθύς αλγόριθμους (π.χ. νευρωνικά δίκτυα συνέλιξης, νευρωνικά δίκτυα ανατροφοδότησης) και υβριδικές παραλλαγές τους.

Οι τεχνικές μηχανικής μάθησης για προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας χρειάζονται το σχεδιασμό και την επιλογή σχετικών χαρακτηριστικών. Αυτή η διαδικασία περιλαμβάνει αρκετή ανθρώπινη παρέμβαση και ειδική γνώση στο βάθος του προβλήματος και τα ειδικά σχεδιασμένα χαρακτηριστικά μπορεί και πάλι να μην πετύχουν τα επιθυμητά αποτελέσματα.

Για την αποφύγη του σχεδιασμού χαρακτηριστικών με χειροκίνητο τρόπο, έχουν προταθεί διάφορες μέθοδοι βαθιάς μάθησης [8] [5] [17]. Οι μέθοδοι βαθιάς μάθησης είναι εξαιρετικά χρήσιμες για την αναγνώριση ανθρώπινης δραστηριότητας. Πρώτον, δεν χρειάζεται ο χειροκίνητος σχεδιασμός χαρακτηριστικών ο οποίος συχνά χρειάζεται ειδική γνώση. Δεύτερον, καλύτερα αποτελέσματα του δείκτη accuracy έχουν αναφερθεί σε σύγκριση με συμβατικές μεθόδους μηχανικής μάθησης. Τρίτον, έχουν την ιδιότητα να μαθαίνουν από μη επισημασμένα δεδομένα (unlabeled data), το οποίο είναι σημαντικό και χρήσιμο, καθώς είναι πρακτικά αδύνατο ο μεγάλος όγκος των datasets να περιέχει μόνο επισημασμένα δεδομένα. Τέταρτον, κατέχουν την ισχυρή ικανότητα εκμάθησης χρήσιμων χαρακτηριστικών από ακατέργαστα δεδομένα και μπορούν να αντιμετωπίσουν την ανίχνευση χαρακτηρι-

στικών από δεδομένα δραστηριοτήτων διαφορετικών ανθρώπων, διαφορετικών φορητών συσκευών και διαφορετικών στάσεων των συσκευών αυτών.

Τα δίκτυα βαθιάς μάθησης είναι τεχνητά νευρωνικά δίκτυα με παραπάνω από ένα κρυφό επίπεδο, επομένως ονομάζονται και βαθιά νευρωνικά δίκτυα. Μια μορφή κατηγοριοποίησης είναι σε βαθιά δίκτυα για εποπτευόμενη μάθηση και σε βαθιά δίκτυα για μη εποπτευόμενη μάθηση. Τα βαθιά δίκτυα για εποπτευόμενη μάθηση χρειάζονται μια προ-εκπαίδευση με datasets τα οποία περιέχουν τους στόχους του προβλήματος, ενώ τα βαθιά δίκτυα για μη εποπτευόμενη μάθηση ακολουθούν ένα σύνολο κανόνων κατά την αναπτυξή τους.

Μια άλλη μορφή κατηγοριοποίησης είναι σε βαθιά παραγωγικά μοντέλα (deep generative models), σε βαθιά διακριτικά μοντέλα (deep discriminative models) και σε υβριδικά μοντέλα. Τα βαθιά παραγωγικά μοντέλα στοχεύουν στην εκμάθηση χρήσιμων αναπαραστάσεων των δεδομένων μέσω μη εποπτευόμενης μάθησης, ή αλλιώς στην εκμάθηση της κοινής κατανομής των δεδομένων και των σχετικών κλάσεων τους. Τέτοια μοντέλα είναι τα Restricted Boltzmann Machines, τα Generative Advesarial Networks καθώς και παραλλαγές τους. Τα βαθιά διακριτικά μοντέλα στοχεύουν στην εκμάθηση της υπό συνθήκη κατανομής πιθανότητας των κλάσεων των δεδομένων, τα οποία περιέχουν έμμεσα ή άμεσα τους στόχους του προβλήματος. Τέτοια μοντέλα είναι τα νευρωνικά δίκτυα συνέλιξης, τα νευρωνικά δίκτυα ανατροφοδότησης καθώς και παραλλαγές τους. Τα υβριδικά μοντέλα με την σειρά τους, συνδυάζουν τα δυο παραπάνω μοντέλα και στοχεύουν στην εκπαίδευση, με συχνό τρόπο, η έξοδος του παραγωγικού μοντέλου να χρησιμοποιείται ως την είσοδο του διακριτικού μοντέλου, τα οποία και χρησιμοποιούνται για ταξινόμηση ή παλινδρόμηση.

Η χρήση κλασικών αλγόριθμων μηχανικής μάθησης όπως SVM, Random Forest, κ.α., περιέχει την πρόκληση ότι αρκετά από τα χαρακτηριστικά χρειάζονται εντοπισμό και εξαγωγή με χειροκίνητο τρόπο, ενέργεια η οποία απαιτεί αρκετό χρόνο. Οι τεχνικές βαθιάς μάθησης όμως, μπορούν αυτόματα να εντοπίσουν τα χαρακτηριστικά που απαιτούνται για το εκάστοτε πρόβλημα, καθιστώντας την ταξινόμηση ανθρώπινων δραστηριοτήτων αποτελεσματικότερη. Νευρωνικά δίκτυα συνέλιξης μιας διάστασης, έχουν την ικανότητα να εντοπίζουν και να εξάγουν τοπικά χαρακτηριστικά στα δεδομένα με αυτόματο τρόπο και νευρωνικά δίκτυα ανατροφοδότησης, λόγο της φύσης τους, μπορούν να αποθηκεύουν πληροφορία που αποκτάται από δεδομένα χρονοσειρών, με αποτέλεσμα τον εντοπισμό και την εξαγωγή χρονικών χαρακτηριστικών τους. Συχνά, συναντάται ο συνδυασμός των παραπάνω νευρωνικών δικτύων σε προβλήματα αναγνώρισης ανθρώπινων δραστηριοτήτων.

Η χρήση τέτοιων υβριδικών μοντέλων αν και έχει προσφέρει σημαντικά αποτελέσματα σε προβλήματα ταξινόμησης ανθρώπινων δραστηριοτήτων, παρουσιάζει μειονεκτήματα. Τα νευρωνικά δίκτυα συνέλιξης τα οποία χρησιμοποιούνται ως βασικά δομικά στοιχεία, υπολογίζουν με παράλληλο τρόπο τις κρυφές αναπαραστάσεις των χαρακτηριστικών για όλες τις θέσεις εισόδου εξόδου. Όμως, ο αριθμός των λειτουργιών που απαιτούνται για το συσχετισμό σημάτων δυο αυθαίρετων θέσεων εισόδου ή εξόδου μεγαλώνει ανάλογα την απόσταση τους, αυξάνοντας έτσι την συνολική πολυπλοκότητα του μοντέλου και καθιστά δυσκολότερη την εκμάθηση εξαρτήσεων των χαρακτηριστικών ανάμεσα σε θέσεις με μεγάλη απόσταση.

Τα νευρωνικά δίκτυα ανατροφοδότησης με τη σειρά τους, παραμετροποιούν τον υπολογισμό συμβόλων κατά μήκος των θέσεων τους στις εισόδους και εξόδους ακολουθιών. Ουσιαστικά ευθυγραμμίζουν τις θέσεις με τα βήματα που απαιτούνται στο χρόνο υπολογισμού της ακολουθίας, δημιουργώντας μια ακολουθία κρυφών καταστάσεων ως συνάρτηση της προηγούμενης κρυφής κατάστασης. Έτσι, η εγγενώς διαδοχική φύση αυτών των δικτύων αποκλείει την παραλληλοποίηση στα στάδια της εκπαίδευσης και για μεγαλύτερα μήκη ακολουθιών περιορίζει την ομαδοποίηση ανάμεσα σε αυτά. Αν και έχουν προταθεί αρκετοί τρόποι βελτίωσης της συνολικής απόδοσης τέτοιων μοντέλων, το πρόβλημα του χρονικού υπολογισμού των ακολουθιών παραμένει.

Για την αντιμετώπιση αυτών των περιορισμών, έχουν προταθεί μοντέλα τα οποία συνδυάζουν δίκτυα όπως τα παραπάνω με μηχανισμούς προσοχής, οι οποίοι επιτρέπουν την μοντελοποίηση εξαρτήσεων των χαρακτηριστικών χρονοσειρών, χωρίς να λαμβάνεται υπ' όψιν η απόσταση τους στις ακολουθίες εισόδου εξόδου. Η "αυτο-προσοχή", που μερικές φορές ονομάζεται και "ενδο-προσοχή", είναι ένας μηχανισμός "προσοχής", ο οποίος σχετίζει τις διαφορετικές θέσεις μια μεμονωμένης ακολουθίας, με σκοπό τον υπολογισμό της αναπαράστασης της ακολουθίας και θα αναλυθεί εκτενώς στο κεφάλαιο 3.

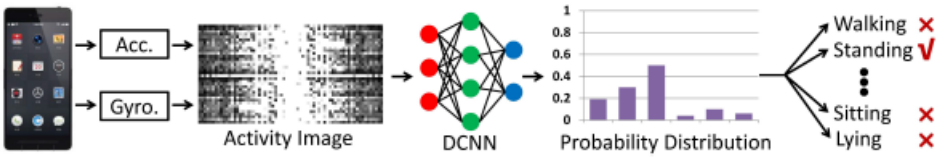
Η ταξινόμηση ως πρόβλημα αναγνώρισης ανθρώπινης δραστηριότητας αντιμετωπίζεται με την εφαρμογή ενός βαθέος νευρωνικού δικτύου το οποίο δέχεται ως εισόδο απευθείας κανονικοποιήμενες χρονοσειρές σημάτων που προέρχονται από τους αισθητήρες και χρησιμοποιεί μόνο μηχανισμούς "αυτο-προσοχής". Οι μηχανισμοί εντοπίζουν συσχετίσεις χαρακτηριστικών ανάμεσα στις χρονοσειρές και σε αντίθεση με τα νευρωνικά δίκτυα ανατροφοδότησης, επιτρέπουν παραλληλισμό στον υπολογισμό τους. Μεγαλύτερα μήκη χρονοσειρών εισέρχονται ως είσοδος με αποτέλεσμα η εκμάθηση των χαρακτηριστικών να καθίσταται πιο ακριβής. Η υπολογιστική ταχύτητα σε συνδυασμό με την ακρίβεια προβλέψεων είναι τα βασικά στοιχεία που συνθέτουν νευρωνικά δίκτυα τα οποία αντιμετωπίζουν προβλήματα ταξινόμησης δραστηριοτήτων αναγνώρισης ανθρώπινης δραστηριότητας. Η μέθοδος sequence-to-sequence που εφαρμόζεται για τις προβλέψεις των δραστηριοτήτων αντιστοιχίζει όλα τα χρονοβήματα που εξέρχονται από το μοντέλο σε ονομασίες(κωδικοποιήσεις) δραστηριοτήτων. Με αυτόν τον τρόπο, είναι πιθανόν η δραστηριότητα που εκτελείται να ταξινομηθεί από τον ίδιο τον χρήστη μέσω φορητής συσκευής, καθιστώντας αποδοτική την εφαρμογή του μοντέλου, ακόμα και σε συσκευές όπως, smartphones, tablets και smartwatches.

# 2. ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΑΙ ΣΧΕΤΙΚΕΣ ΕΡΓΑΣΙΕΣ

## 2.1 Εργασίες που στοχεύουν στο ίδιο ή σε παρόμοιο πρόβλημα

Εκτεταμένες έρευνες έχουν λάβει μέρος σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας. Οι προσεγγίσεις, αν και ποικίλουν, επικεντρώνονται κυρίως γύρω από νευρωνικά δίκτυα συνέλιξης, ανατροφοδότησης, και τώρα τελευταία από μηχανισμούς προσοχής.

Ξεκινώντας από εφαρμογές νευρωνικών δικτύων συνέλιξης, στο [10], οι ακολουθίες σημάτων από αισθητήρες όπως επιταχυνσιόμετρο και γυροσκόπιο ενσωματώνονται σε μια εικόνα δραστηριότητας, αντί για χειροκίνητη εξαγωγή τους. Η εικόνα δραστηριότητας τροφοδοτείται σε ένα νευρωνικό δίκτυο συνέλιξης για την αυτόματη εκμάθηση των χαρακτηριστικών τους, όπου θα οδηγήσουν στην αναγνώριση των δραστηριοτήτων που ορίζει το πρόβλημα.

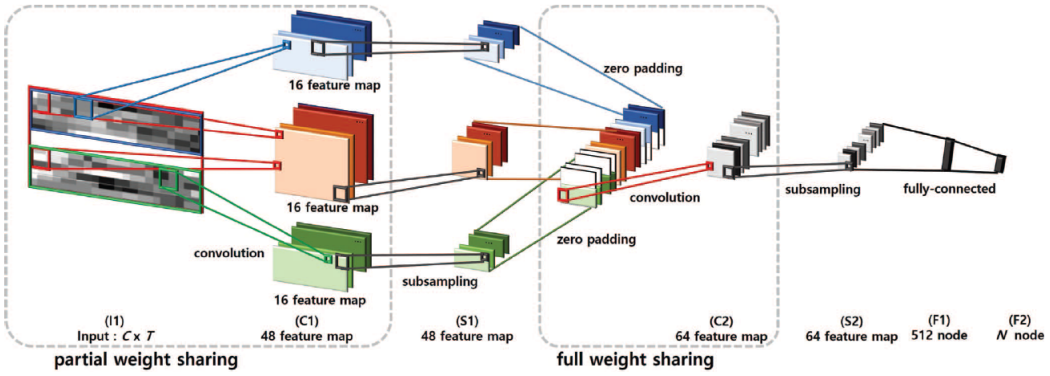



**Σχήμα 2.1: Επισκόπηση αναγνώρισης δραστηριοτήτων μέσω σημάτων**

Τα datasets όπου χρησιμοποιήθηκαν είναι τα UCI, USC και SHO, με αισθητήρες κίνησης στο γοφό και στον καρπό. Τα αποτελέσματα για την παραπάνω μέθοδο πετυχαίνουν $95.18\%$ για το UCI, $97.01\%$ για το USC και $99.93\%$ για το SHO, του δείκτη accuracy. Καταλήγοντας, ανταποκρίνεται επιτυχώς στο πρόβλημα της αναγνώρισης ανθρώπινης δραστηριότητας με αισθητήρες κίνησης. Το χαμηλό υπολογιστικό κόστος σε συνδυασμό με την υψηλή απόδοση του μοντέλου, ανοίγουν τους ορίζοντες για την αποτελεσματική εξερεύνηση σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας.

Στα πλαίσια αυτής της εξερεύνησης εντάσσεται και το [7]. Εδώ, παρουσιάζονται νευρωνικά δίκτυα συνέλιξης τα οποία αντιμετωπίζουν δυσκολίες για πολυτροπικά δεδομένα από διαφορετικούς αισθητήρες κίνησης. Δυο μοντέλα νευρωνικών δικτύων συνέλιξης με δυο επίπεδα εφαρμόζονται. Το πρώτο περιέχει μερική κατανομή από βάρη στο πρώτο επίπεδο και καθολική κατανομή στο δεύτερο, ενώ το επόμενο μοντέλο περιέχει μερική και καθολική κατανομή από βάρη στο πρώτο επίπεδο και μόνο καθολική στο δεύτερο.

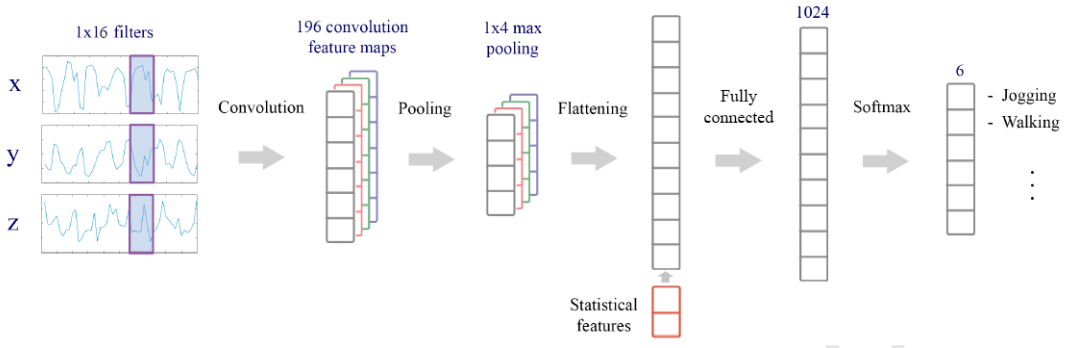
Τα datasets τα οποία χρησιμοποιούνται είναι το Mhealth, το οποίο περιέχει δεδομένα από τέσσερις διαφορετικούς αισθητήρες, το επιταχυνσιόμετρο, το γυροσκόπιο, το μαγνητόμετρο αλλά και ηλεκτροκαρδιογράφημα. Τα δυο μοντέλα συγκριτικά με υπάρχουσες μεθόδους ταξινόμησης, όπως Hidden Markov model, Support Vector Machines, 1-D CNN, κ.α., πετυχαίνουν συνολικά $91.94\%$ του δείκτη accuracy, $1.51\%$ παραπάνω από τη δεύτερη μέθοδο. Για προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας σε χρονοσειρές πολλών μεταβλητών από δεδομένα ετερογενών αισθητήρων, τα μοντέλα όπου παρουσιά-




**Σχήμα 2.2: Τα δύο μοντέλα όπου εφαρμόζονται. Αριστερά: Μερική κατανομή. Δεξιά: Καθολική κατανομή.**

ζονται, καταφέρνουν υψηλό score στο δείκτη accuracy για προσεγγίσεις ταξινόμησης και με λιγότερες παραμέτρους.

Σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας, οι προσεγγίσεις ταξινόμησης υπολογίζονται συνήθως μέσω του τρόπου της εποπτευόμενης μάθησης. Στο [9] όμως, η προσέγγιση που προτείνουν οι συγγραφείς περιλαμβάνει ταξινόμηση δραστηριοτήτων ανεξάρτητη από τον χρήστη και σε πραγματικό χρόνο με την χρήση νευρωνικών δικτύων συνέλιξης. Η μορφή και το μέγεθος από τις χρονοσειρές των δεδομένων σχετίζεται με την ανίχνευση του είδους της δραστηριότητας, επομένως η χρήση νευρωνικών δικτύων συνέλιξης είναι ιδανική για την εκμάθηση των χαρακτηριστικών τους. Με αυτό τον μη εποπτευόμενο τρόπο, η μάθηση κληρονομείται στα βαθύτερα επίπεδα συνέλιξης και στη συνέχεια περνάει σε ένα πλήρες-ενωμένο επίπεδο, όπου η ταξινόμηση παίρνει μέρος. Μέσω του αλγόριθμου backpropagation, το νευρωνικό δίκτυο εκπαιδεύεται σαν ένα σύνολο, ελαχιστοποιώντας έτσι το συνολικό σφάλμα ταξινόμησης.

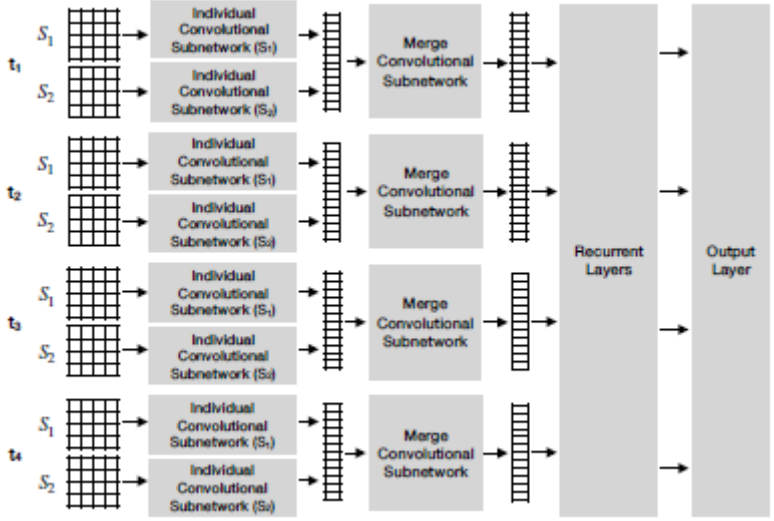



**Σχήμα 2.3: Η αρχιτεκτονική του προτεινόμενου νευρωνικού δικτύου συνέλιξης.**

Τα datasets τα οποία χρησιμοποιούνται είναι τα WISDM και UCI, τα οποία περιέχουν δεδομένα από επιταχυνσιόμετρο από 36 και 30 χρήστες αντίστοιχα. Τα αποτελέσματα που αναφέρονται είναι, για το WISDM, $93.32\%$ του δείκτη accuracy, $11\%$ υψηλότερα από την προηγούμενη μέθοδο και για το UCI dataset, $97.62\%$ του δείκτη accuracy. Η παραπάνω μέθοδο παρουσιάζει υψηλή απόδοση για προβλήματα αναγνώρισης ανθρώπινης δραστη-

ριότητας, με παράλληλα χαμηλό υπολογιστικό κόστος και καθόλου χειροκίνητη εξαγωγή χαρακτηριστικών.

Πέρα όμως από νευρωνικά δίκτυα συνέλιξης σε προσεγγίσεις προβλημάτων αναγνώρισης ανθρώπινης δραστηριότητας, εμφανίζονται και νευρωνικά δίκτυα ανατροφοδότησης. Στο [18], το DeepSense χρησιμοποιεί χρονοσειρές με δεδομένα από πολλαπλούς αισθητήρες κίνησης. Είναι ένα framework βαθιάς μάθησης το οποίο προσπαθεί να επιλύσει προβλήματα όπως ο θόρυβος από τα δεδομένα των αισθητήρων και να ξεπεράσει τα εμπόδια προσαρμογής των χαρακτηριστικών τους με έναν ενιαίο τρόπο. Ενσωματώνει νευρωνικά δίκτυα συνέλιξης και ανατροφοδότησης. Τα πρώτα χρησιμοποιούνται για τον υπολογισμό των συντεταγμένων από όλους τους αισθητήρες εντός των χρονικών διαστημάτων, από τα οποία εξάγονται τα τοπικά χαρακτηριστικά, και στη συνέχεια συνδυάζονται σε καθολικά. Τα δεύτερα χρησιμοποιούνται για τον υπολογισμό των συντεταγμένων από τους αισθητήρες κατά μήκος των χρονικών διαστημάτων από των οποίων και εξάγονται οι χρονικές εξαρτήσεις τους.




**Σχήμα 2.4: Το framework DeepSense.**

Τα datasets τα οποία χρησιμοποιούνται είναι τα CarTrack, HHAR, και UserID. Τα δυο datasets δεν αναφέρονται, καθώς βρίσκονται εκτός πλαισίου της συγκεκριμένης εργασίας. Για το HHAR dataset, αναφέρονται αποτελέσματα από τρεις δείκτες: $0.942 \pm 0.032$ accuracy, $0.931 \pm 0.041$ macro F1-score και $0.942 \pm 0.032$ micro F1-score. Το DeepSense framework φαίνεται υποσχόμενο σε πολλά προβλήματα χρονοσειρών. Αυτή η αρχιτεκτονική βαθιάς μάθησης μπορεί να αντιμετωπίσει προβλήματα αισθητήρων κίνησης όπως θορυβώδεις ενδείξεις, σε ένα ενοποιημένο framework.

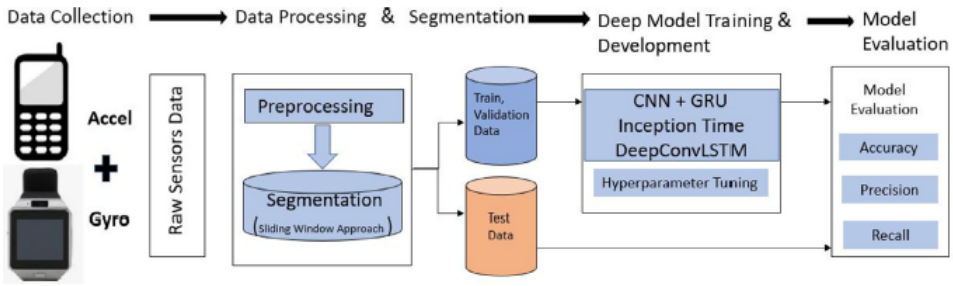
Ο συνδυασμός νευρωνικών δικτύων συνέλιξης και ανατροφοδότησης εμφανίζει υποσχόμενα αποτελέσματα σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας. Στο [15], οι συγγραφείς του άρθου προτείνουν ένα υβριδικό νευρωνικό δίκτυο, το οποίο συνδυάζει δίκτυα συνέλιξης και μακροπρόθεσμης-βραχυπρόθεσμης μνήμης για εφαρμογές υγειονομικής περίθαλψής με την χρήση αισθητήρων κίνησης. Μέσω της βαθιάς εκμάθησης μπο-

ρούν και αναγνωρίζονται συγκεκριμένες δραστηριότητες καθώς και οι μεταβάσεις μεταξύ τους, τόσο σε μικρή διάρκεια όσο και σε χαμηλή συχνότητα. Τα δίκτυα συνέλιξης βοηθούν στην εξαγωγή των χαρακτηριστικών τα οποία προέρχονται από τα δεδομένα των αισθητήρων και στην συνέχεια, τα δίκτυα ανατροφοδότησης χρησιμοποιούνται για να εντοπίσουν μακροπρόθεσμες εξαρτήσεις μεταξύ δυο δραστηριοτήτων ώστε να βοηθήσουν στην καλύτερη αναγνώριση των μεταβάσεων τους.

Το dataset το οποίο χρησιμοποιείται είναι το HAPT dataset, το οποίο περιέχει 12 δραστηριότητες οι οποίες συλλέχθηκαν από αισθητήρες κίνησης (επιταχυνσιόμετρο και γυροσκόπιο) και με την χρήση της τεχνικής batch normalization, επιτυγχάνεται ποσοστό $95.87\%$ του δείκτη accuracy, το οποίο είναι υψηλότερο από μεθόδους όπως Random Forrest και K-Nearest Neighbor, τόσο στην αυτούσια αναγνώριση των δραστηριοτήτων, όσο και στις μεταβάσεις μεταξύ τους. Συμπερασματικά, η παραπάνω μέθοδος εξετάζει την αναγνώριση απλών δραστηριοτήτων και των μεταβάσεων τους, αφήνοντας χώρο για μελλοντική επεκτασή της, η οποία θα αφορά περισσότερο πολύπλοκες δραστηριότητες εξατομικεύοντας και τις μεταβάσεις μεταξύ τους, οι οποίες θα διαφέρουν ανάλογα τον εκάστοτε χρήστη.

Η προσέγγιση δικτύων συνέλιξης και ανατροφοδότησης εξετάζεται και με διαφορετικούς τρόπους. Στο [6] οι συγγραφείς του άρθρου προτείνουν και εδώ ένα υβριδικό μοντέλο νευρωνικού δικτύου βαθιάς μάθησης, το οποίο όμως, αποτελείται από ένα δίκτυο συνέλιξης σε συνδυασμό με στοιχεία ανατροφοδότησης (gated recurrent units), για την επίλυση προβλημάτων ταξινόμησης στον τομέα της αναγνώρισης ανθρώπινης δραστηριότητας. Η χρήση μονοδιάστατων νευρωνικών δικτύων συνέλιξης σε συνδυασμό με στοιχεία ανατροφοδότησης, συντελούν στην δημιουργία μιας αρχιτεκτονικής, η οποία συνδυάζει την ικανότητα των μονοδιάστατων δικτύων συνέλιξης να εντοπίζουν τοπικά χαρακτηριστικά από τα δεδομένα των αισθητήρων και τη δυνατότητα των δυο στοιχείων ανατροφοδότησης να εντοπίζουν χρονικά χαρακτηριστικά.

Το dataset το οποίο χρησιμοποιήθηκε είναι το WISDM, το οποίο αποτελείται από δραστηριότητες οι οποίες συλλέχθηκαν μέσω αισθητήρων κίνησης, όπως το επιταχυνσιόμετρο και το γυροσκόπιο, οι οποίοι βρίσκονται συσκευές smartphone και smartwatch. Το προτεινόμενο υβριδικό μοντέλο συγκρίθηκε με δυο άλλα μοντέλα, το Inception Time και το DeepConvLSTM. Για την απόκτηση των υπερ-παραμέτρων των παραπάνω μοντέλων χρησιμοποιήθηκε η ανοιχτού λογισμικού AutoML βιβλιοθήκη McFly γραμμένη σε γλώσσα python. Τα αποτελέσματα όπου σημειώθηκαν για τον δείκτη accuracy στη συσκευή smartwatch είναι $96.54\%$ για το μοντέλο CNN-GRU, $95.79\%$ για το Inception Time και $87.65$ για το DeepConvLSTM. Για τη συσκευή smartphone $90.44\%$ για το μοντέλο CNN-GRU, $88.50\%$ για το Inception Time και $75.31\%$ για το DeepConvLSTM. Η εφαρμογή του προτεινόμενου μοντέλου παρουσιάζει ότι υβριδικά μοντέλα βαθιάς μάθησης, μπορούν αποτελεσματικά και με αυτόματο τρόπο να εντοπίζουν χαρακτηριστικά από δεδομένα αισθητήρων κίνησης και να τα ταξινομούν σε σύνθετες ανθρώπινες δραστηριότητες. Όπως αναφέρεται και από τους ίδιους τους συγγραφείς του άρθρου, μελλοντικές επεκτάσεις αφορούν την χρήση μοντέλων Transformer για προβλήματα ταξινόμησης χρονοσειρών σε ανθρώπινες δραστηριότητες του WISDM dataset.




**Σχήμα 2.5: Framework για αναγνώριση ανθρώπινης δραστηριότητας με την χρήση υβριδικού δικτύου.**

## 2.2 Εργασίες με attention-based δίκτυα/αρχιτεκτονικές

Η χρήση νευρωνικών δικτύων συνέλιξης και ανατροφοδότησης έχει θέσει ψηλά τον πήχη σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας. Σημαντικές βελτιώσεις έχουν σημειωθεί, τόσο στην υπολογιστική αποτελεσματικότητα των μοντέλων, όσο και στην γενικότερη απόδοση τους. Παρ' όλα αυτά, ο θεμελιώδης περιορισμός για την ελάττωση του χρόνου υπολογισμού ακολουθιών παραμένει. Ο μηχανισμός self-attention συσχετίζει τις διαφορετικές θέσεις μιας ενιαίας ακολουθίας με σκοπό να υπολογίσει μια αναπαράσταση της.

Στο [14] εμφανίζεται για πρώτη φορά ένα μοντέλο μεταγωγής ακολουθιών, το οποίο απαρτίζεται εξ' ολοκλήρου από μηχανισμούς self-attention για τον υπολογισμό αναπαραστάσεων τους. Το κύριο πρόβλημα όπου καλείται να επιλύσει το άρθρο είναι η αντικατάσταση σύνθετων δικτύων ανατροφοδότησης όπου ήδη χρησιμοποιούν encoder, decoder αλλά και attention μηχανισμούς με απλούστερες αρχιτεκτονικές δικτύων, οι οποίες χρησιμοποιούν attention μηχανισμούς αποκλειστικά.

Για προβλήματα sequence-to-sequence η αρχιτεκτονική encoder, decoder των RNNs παρουσιάζει σημαντικούς περιορισμούς. Η ικανότητα τους να διατηρούν πληροφορίες για τα αρχικά στοιχεία χάνεται όταν καινούργια στοιχεία ενσωματώνονται στην ακολουθία. Στον encoder, το hidden state για κάθε χρονοβήμα σχετίζεται με μια συγκεκριμένη λέξη στην αρχική πρόταση, συνήθως στην πιο πρόσφατη. Συνεπώς, ο decoder αποκτά πρόσβαση μόνο στο τελευταίο hidden state του, με αποτέλεσμα να χάνονται πληροφορίες από τα αρχικά στοιχεία της ακολουθίας.

Για να αντιμετωπιστεί αυτός ο περιορισμός, εισάγεται η έννοια του attention μηχανισμού. Σε κάθε βήμα του decoder, αναζητούνται όλα τα states του encoder παρέχοντας πρόσβαση σε όλα τα στοιχεία της αρχικής πρότασης. Αυτή είναι η βασική λειτουργία του attention μηχανισμού, εξάγει πληροφορίες από ολόκληρη την πρόταση, ένα άθροισμα από βάρη για όλα τα παρελθοντικά states του encoder. Με αυτόν τον τρόπο, ο decoder αναθέτει μεγαλύτερα βάρη ή περισσότερη "προσοχή" σε ένα συγκεκριμένο στοιχείο της εισόδου για κάθε στοιχείο της εξόδου. Η μάθηση σε κάθε βήμα εξασφαλίζει ότι για το σωστό στοιχείο της εισόδου πραγματοποιείται πρόβλεψη για το επόμενο στοιχείο της εξόδου.

Αυτη η προσέγγιση όμως περιέχει περιορισμούς, καθώς κάθε πρόταση αντιμετωπίζεται

σαν ένα στοιχείο την φορά. Αμφότεροι encoder και decoder πρέπει να περιμένουν την ολοκλήρωση $t-1$ βημάτων για να επεξεργαστούν τα επόμενα $t-th$ βήματα. Συνεπώς, για την επεξεργασία μεγάλων datasets ο χρόνος εκτέλεσης αλλά και η πολυπλοκότητα της προσέγγισης πρέπει να αντιμετωπιστούν με άλλες τεχνικές, οι οποίες θα συζητηθούν παρακάτω.

Τα datasets τα οποία χρησιμοποιούνται είναι, το WMT 2014 English-German dataset, το οποίο περιέχει περίπου 4.5 εκατομμύρια ζευγάρια προτάσεων και το WMT 2014 English-French dataset, το οποίο περιέχει περίπου 36 εκατομμύρια ζευγάρια προτάσεων. Για το WMT 2014 English-to-German dataset αναφέρονται αποτελέσματα της τάξεως $28.4$ BLEU score, δύο μονάδες πάνω από το προηγούμενο επίσημο μοντέλο. Για το WMT 2014 English-to-French dataset αναφέρονται αποτελέσματα της τάξεως 41.0 BLEU score, ξεπερνώντας όλα τα προηγούμενα επίσημα μοντέλα σε λιγότερο του $1/4$ του κόστους εκπαίδευσης τους. Ο Transformer, όπως ονομάζεται, είναι το πρώτο μοντέλο μεταγωγής ακολουθίας το οποίο βασίζεται εξ' ολοκλήρου στο μηχανισμό attention, αντικαθιστώντας τα επίπεδα ανατροφοδότησης τα οποία χρησιμοποιούνται σε αρχιτεκτονικές encoder-decoder, με το μηχανισμό multi-head self-attention. Η ικανότητα γενίκευσης σε συνδυασμό με το γεγονός ότι το όριο απόδοσης του μοντέλου εξαρτάται από τις δυνατότητες της επεξεργαστικής ισχύς του υλικού (hardware), καθιστούν τον Transformer ικανό για την επίλυση προβλημάτων πέρα από γλωσσικών, όπως computer vision ή προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας.

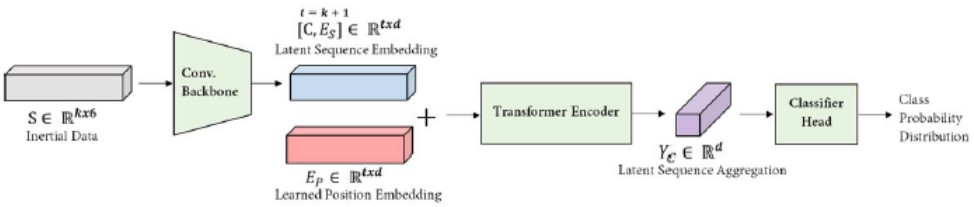
Η εφαρμογή του μοντέλου Transformer σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας, ξεκινά με συνδυασμό της αρχιτεκτονικής self-attention και αρχιτεκτονικών συνέλιξης. Στο [11] προτείνεται ένα νευρωνικό μοντέλο βασισμένο σε αρχιτεκτονική self-attention, το οποίο παράγει υψηλότερης διάστασης χαρακτηριστικά τα οποία χρησιμοποιούνται για προβλήματα ταξινόμησης. Επίσης, παρατηρείται ότι, η αρχιτεκτονική self-attention βοηθάει στην ενίσχυση του τρόπου λειτουργίας των αισθητήρων, οδηγώντας σε πιο ακριβή αποτελέσματα μεταξύ διαφορετικών δραστηριοτήτων για προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας.

Τα δεδομένα από τους αισθητήρες εισέρχονται με τη μορφή χρονοσειρών σε ένα μπλοκ το οποίο χρησιμοποιεί αρχιτεκτονική self-attention για να συλλέξει τις αναπαραστάσεις από τα βάρη των αισθητήρων και να εξακριβώσει τον τρόπο λειτουργίας τους. Στη συνέχεια μετασχηματίζονται σε διανύσματα διάστασης $d$, και μονοδιάστατος μηχανισμός συνέλιξης εφαρμόζεται σε αυτά. Η κωδικοποίηση θέσης και τα μπλοκ self-attention εφαρμόζονται ακριβώς όπως και στο [αττενσιον]. Η έξοδος που προκύπτει τροφοδοτείται σε ένα μπλοκ, το οποίο "μαθαίνει" τις παραμέτρους από κάθε χρονοβήμα, ώστε να βοηθήσουν στην ταξινόμηση της δραστηριότητας για την τρέχουσα χρονοσειρά. Για τη τελική ταξινόμηση των δραστηριοτήτων χρησιμοποιείται ένα πλήρως ενωμένο επίπεδο με την χρήση της συνάρτησης softmax.

Τα datasets τα οποία χρησιμοποιούνται είναι τα PAMAP2, OPPORTUNITY, USC-HAD και SKODA, τα οποία περιέχουν δραστηριότητες από αισθητήρες κίνησης τοποθετημένους σε καίρια σημεία του ανθρώπινου σώματος. Το προτεινόμενο μοντέλο πετυχαίνει $0.96$ του δείκτη macro F1-score για το PAMAP2, $0.67$ για το OPPORTUNITY, $0.55$ για το USC-HAD, και $0.97$ για το SKODA και επιτυγχάνει υψηλότερο score από τα μοντέλα με τα οποία συ-

γκρίνεται, τα οποία περιέχουν αρχιτεκτονικές ανατροφοδότησης. Συνολικά, η εφαρμογή του παραπάνω μοντέλου παρουσιάζει υποσχόμενα αποτελέσματα για ταξινόμηση δραστηριοτήτων σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας, διευρύνοντας έτσι τους ορίζοντες για περαιτέρω έρευνα.

Ακολουθώντας την τάση για συνδυασμό αρχιτεκτονικών self-attention, στο [13] επιχειρείται η προσέγγιση του μοντέλου Transformer, όπως το παραπάνω, με συνδυασμό στοιχείων συνέλιξης. Η βασική ιδέα της έρευνας είναι η αναγνώριση ανθρώπινης δραστηριότητας με την χρήση του μοντέλου IMU-Transformer για δεδομένα από αισθητήρες με βάση την αδράνεια. Αντικαθιστούν αρχιτεκτονικές μακροπρόθεσμης μνήμης με το μοντέλο Transformer, το οποίο αποτελείται από αρχιτεκτονικές encoder, self-attention, αλλά και δικτύων συνέλιξης. Η βασική αρχιτεκτονική όπου ακολουθείται είναι όπως παρουσιάστηκε στο [14], με την προσθήκη τεσσάρων μονοδιάστατων δικτύων συνέλιξης μετά την είσοδο των δεδομένων, όπου παράγουν ενσωματώσεις ακολουθιών και μια ακολουθία για τη συλλογή των χρονικών θέσεων μετά την έξοδο από το μπλοκ του Transformer.




**Σχήμα 2.6: Το προτεινόμενο framework για ταξινόμηση δραστηριοτήτων με την χρήση μοντέλων Transformers.**

Τα datasets τα οποία χρησιμοποιήθηκαν είναι τα SLR, HAR και SHAR, τα οποία περιέχουν 27.76 ώρες καταγραφής δραστηριοτήτων από 91 ανθρώπους. Τα αποτελέσματα του δείκτη accuracy, τα οποία αναφέρονται από την σύγκριση του παραπάνω μοντέλου με ένα μοντέλο το οποίο αποτελείται απο νευρωνικά δίκτυα συνέλιξης, είναι: για το SLR dataset, $97.4\%$ για το μοντέλο IMU-Transformer και $96.5\%$ για το μοντέλο IMU-CNN. Για το HAR dataset, $89.6\%$ για το μοντέλο IMU-Transformer και $86.2\%$ για το IMU-CNN. Για το SHAR dataset, $90.7\%$ για το IMU-Transformer και $88.7\%$ για το IMU-CNN. Συνολικά, το μοντέλο IMU-Transformer πέτυχε $90.7\%$ του δείκτη accuracy, $2\%$ υψηλότερα από το IMU-CNN. Είναι η πρώτη φορά όπου εφαρμόζεται το μοντέλο Transformer για αναγνώριση ανθρώπινης δραστηριότητας με βάση την αδράνεια. Αν και λίγο πιο αργό όσο αφορά τον χρόνο εκτέλεσης, πετυχαίνει καλύτερα αποτελέσματα στο δείκτη accuracy από άλλα μοντέλα βασισμένα σε δίκτυα συνέλιξης ή ανατροφοδότησης. Μια μελλοντική επέκταση του παραπάνω μοντέλου θα ήταν η διευρεύνηση τεχνικών επιτάχυνσης χρόνου εκτέλεσης, καθώς και η περαιτέρω μελέτη σχετικά με τη μεταφορά μάθησης.

Μια προσέγγιση σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας είναι το [2]. Το κύριο πρόβλημα όπου καλείται να επιλύσει και εδώ το άρθρο, είναι η αντικατάσταση αρχιτεκτονικών ανατροφοδότησης με μηχανισμούς βασισμένους σε αρχιτεκτονική self-attention και μόνο, για την άμεση ανίχνευση χρονικών εξαρτήσεων σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας με την χρήση αισθητήρων κίνησης. Οι συγγραφείς προτείνουν ένα framework βαθιάς μάθησης, το Trasend, το οποίο χρησιμοποιεί αρχιτε-

κτονική attention για πολυτροπικά χρονικά δεδομένα (multimodal temporal data). Ένας μηχανισμός πρόσβασης-μνήμης ξεπερνά τα εμπόδια των RNNs και αποκτά την ικανότητα εκμάθησης από μεγάλου μήκους ακολουθίες εισόδου. Η εφαρμογή ενός προσωρινού μπλοκ εξαγωγής πληροφοριών, το οποίο ακολουθεί τη βασική αρχιτεκτονική του μοντέλου Transformer, όπως παρουσιάστηκε στο [14], επιτυγχάνει καλύτερα αποτελέσματα συγκριτικά με state-of-the-art μοντέλα σε τρία διαφορετικά datasets αναγνώρισης ανθρώπινης δραστηριότητας με προβλήματα ταξινόμησης. Επίσης, σε συνδυασμό με την χρήση μεθόδων μεταφοράς μάθησης, επιτυγχάνεται εξατομίκευση του μοντέλου σε κάθε χρήστη, εξυπηρετώντας έτσι τον κύριο στόχο των προβλημάτων αναγνώρισης ανθρώπινης δραστηριότητας.

Τα datasets τα οποία χρησιμοποιούνται είναι τα HHAR, PAMAP2, και USC-HAD. Για το HHAR dataset αναφέρονται αποτελέσματα για το δείκτη F1-score της τάξεως $0.848$. Για το PAMAP2 dataset $0.723$ και για το USC-HAD dataset $0.702$, τα οποία είναι κατά μέσο όρο $7\%$ υψηλότερα από το προηγούμεο μοντέλο με τις καλύτερες επιδόσεις. Το Trasend είναι ένα framework βαθιάς μάθησης για πολυτροπικές χρονικές ακολουθίες (multimodal time series), το οποίο σε συνδυασμό με τη μέθοδο μεταφοράς μάθησης εξατομικεύει το μοντέλο προς τις ανάγκες κάθε χρήστη. Παρ' όλα αυτά, η διαδικασία εξατομίκευσης μπορεί να επηρεάσει τη συνολική εμπειρία του χρήστη από την χρήση της εφαρμογής, καθώς οι επαναλαμβανόμενες ερωτήσεις προς τον χρήστη για την αξιολόγηση των προβλέψεων, μπορεί να καταστεί μη εφικτή. Η αρχιτεκτονική attention έχει χρησιμοποιηθεί μόνο ως αντικαταστάτης των RNNs και όχι ως ένα μέσο για την άμεση αποτύπωση χρονικών εξαρτήσεων, συνεπώς προκύπτει η ανάγκη για την χρήση περισσότερων frameworks όπως το παραπάνω.

Η επιτυχία των αρχιτεκτονικών self-attention σε προβλήματα επεξεργασίας φυσικής γλώσσας δεν άργησε να εφαρμοστεί και σε άλλους τομείς. Αν και αφορά επεξεργασία εικόνας, μια σημαντική αναφορά αρχιτεκτονικής είναι το [4], στο οποίο, οι συγγραφείς του άρθρου προτείνουν μια αρχιτεκτονική για προβλήματα computer vision η οποία αποτελείται αποκλειστικά από μηχανισμούς self-attention, ονομάζοντας το μοντέλο Vision Transformer(ViT). Η εικόνα που εισέρχεται στο μοντέλο χωρίζεται σε κομμάτια και στη συνέχεια αυτά τα κομμάτια μέσω γραμμικών ενσωματώσεων τροφοδοτούνται στην ακολουθία.

Όπως εφαρμόζεται η ενσωμάτωση θέσης στο [14], έτσι συμβαίνει και εδώ, με την διαφορά ότι προστίθεται και η ενσωμάτωση των κομματιών που χωρίστηκε η εικόνα. Σε αντίθεση με την επεξεργασία φυσικής γλώσσας όπου οι προτάσεις μεταφράζονται από μια γλώσσα σε κάποια άλλη εδώ, ο ορισμός του προβλήματος αφορά ταξινόμηση εικόνων, επομένως η αρχιτεκτονική του μοντέλου περιέχει μόνο μπλοκ encoder. Το μοντέλο λειτουργεί με επίβλεψη και εκπαιδεύεται για την ταξινόμηση εικόνων με την χρήση Multilayer Perceptron(MLP) με ένα κρυφό επίπεδο για την προ-εκπαίδευση και ένα γραμμικό επίπεδο για το fine-tuning.

Τα datasets που χρησιμοποιούνται είναι τα ILSVRC-2012 ImageNet με 1000 κλάσεις και 1.3M εικόνες, το ImageNet-21K με 21k κλάσεις και 14M εικόνες και το JFT με 18k κλάσεις και 303M εικόνες υψηλής ανάλυσης. Τα αποτελέσματα όπου αναφέρονται είναι ο μέσος όρος και η τυπική απόκλιση του δείκτη accuracy και μετά από fine-tuning, το μοντέλο Vision Transformer υπερτερεί παγιωμένων μοντέλων όπως το ResNet σε όλα τα datasets (2%

παραπάνω του δείκτη accuracy) πετυχαίνοντας και μείωση σε υπολογιστικούς πόρους κατά την εκπαίδευση του. Η εφαρμογή του μοντέλου Transformer και σε προβλήματα computer vision θέτει τα θεμέλια για την εξευρεύνηση προβλημάτων και σε άλλους τομείς, αφού με την χρήση του, πληθώρα προβλημάτων μπορεί να αντιμετωπιστεί επιτυχώς αλλά και με χαμηλότερο κόστος, αρκεί η είσοδος που τροφοδοτείται σε αυτό να είναι σε μορφή ακολουθίας.

Στο [3] η προσέγγιση είναι λίγο διαφορετική. Οι συγγραφείς του άρθρου παρουσιάζουν το μοντέλο Transformer, ακριβώς όπως παρουσιάστηκε στο [14], όμως αντί για επεξεργασία φυσικής γλώσσας, εφαρμόζεται για ανάλυση χρονοσειρών από σήματα κίνησης. Εδώ δεν εφαρμόζεται απλά ο μηχανισμός self-attention όπως στο [2], αλλά προσαρμόζεται ολόκληρο το μοντέλο Transformer σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας με την χρήση αισθητήρων κίνησης. Περνώντας μόνο από ένα επίπεδο normalization, τα σήματα από τις χρονοσειρές τροφοδοτούνται ως είσοδο στο νευρωνικό δίκτυο. Με αυτόν τον τρόπο δεν χρειάζεται να γίνει χειροκίνητη ρύθμιση σήματος και το μοντέλο είναι έτοιμο για εκτέλεση σε οποιαδήποτε κινητή συσκευή. Το επίπεδο εξόδου είναι γραμμικό για να προσφέρει μεγαλύτερη υπολογιστική ταχύτητα για την ταξινόμηση των δραστηριοτήτων.

Το dataset που χρησιμοποιείται είναι το KU-HAR, για το οποίο μια ολόκληρη χρονοσειρά σήματος περιέχει μια μόνο δραστηριότητα και μια επισήμανση για αυτήν. Οι αισθητήρες που χρησιμοποιούνται είναι επιταχυνσιόμετρα και γυροσκόπια τοποθετημένα σε μια τσάντα μέσης. Τα αποτελέσματα όπου σημειώνονται είναι $99.2\%$ του δείκτη accuracy, $9.53\%$ υψηλότερα από τη μέθοδο Random Forest όπου συγκρίνεται για αυτό το dataset. Η εφαρμογή αυτού του μοντέλου Transformer για προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας μέσω ανάλυσης σήματος, προσφέρει υψηλά αποτελέσματα για προβλήματα ταξινόμησης δραστηριοτήτων, ακολουθώντας μια αυτοματοποιημένη διαδικασία και θα αναλυθεί εκτενέστερα παρακάτω.

# 3. ΜΗΧΑΝΙΚΗ ΚΑΙ ΒΑΘΙΑ ΜΑΘΗΣΗ

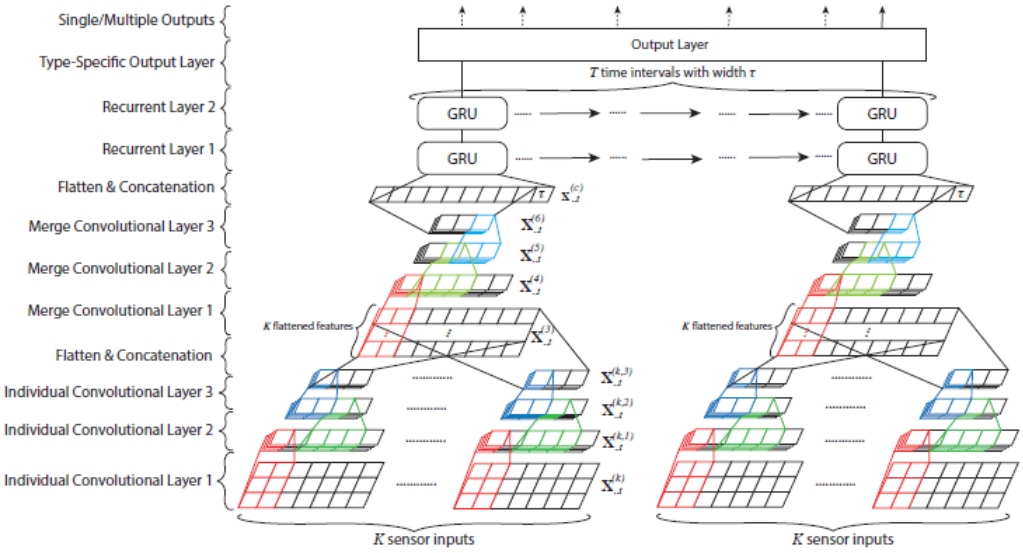
## 3.1 Μηχανική μάθηση και νευρωνικά δίκτυα

Για προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας μέσω αισθητήρων κίνησης, έχουν προταθεί ένας ικανοποιητικός αριθμός από αρχιτεκτονικές, οι οποίες περιλαμβάνουν τόσο κλασικές τεχνικές μηχανικής μάθησης, όσο και ανάπτυξη νευρωνικών δικτύων με αρκετά επίπεδα βάθους. Η παρακολούθηση της διάρκειας αλλά και ίδιου του χρόνου των δραστηριοτήτων είναι ζωτικής σημασίας, αφού τα δεδομένα προέρχονται από αισθητήρες κίνησης. Οι προσεγγίσεις όπου ξεχωρίζουν συνήθως, περιλαμβάνουν συνδυασμό νευρωνικών δικτύων συνέλιξης και ανατροφοδότησης, καθώς τα δεύτερα έχουν αποδεικτεί αρκετά αποτελεσματικά στον εντοπισμό εξαρτήσεων ανάμεσα στις χρονοσειρές δεδομένων.

Αρχιτεκτονικές όπου ακολουθούν τις παραπάνω προσεγγίσεις περιλαμβάνουν συνήθως εκτός από δίκτυα συνέλιξης διάφορες παραλλαγές και δικτύων ανατροφοδότησης, όπως μακροπρόθεσμης-βραχυπρόθεσμης μνήμης και ανατροφοδότησης με πύλη. Η δεύτερη περίπτωση εμφανίζεται και στο [18], όπου εφαρμόζεται το framework DeepSense.

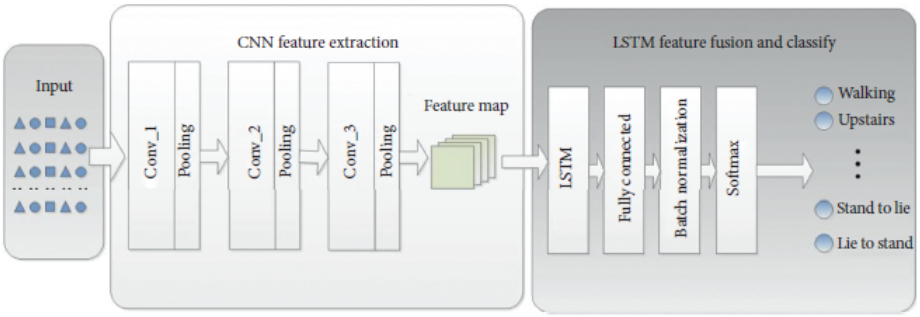
Αποτελείται από επίπεδα συνέλιξης, ανατροφοδότησης με πύλη καθώς και ένα επίπεδο εξόδου. Τα επίπεδα συνέλιξης αποτελούνται από τα μεμονωμένα επίπεδα συνέλιξης και τα συνελικτικά επίπεδα συγχώνευσης. Τα δεδομένα από τους αισθητήρες εισέρχονται ως είσοδο στο δίκτυο μέσω του μεμονωμένου υποδικτύου συνέλιξης με τρία επίπεδα $((k,1),(k,2),(k,3))$. Με την εφαρμογή δυσδιάστατων φίλτρων, το δίκτυο μέσω του μηχανισμού συνέλιξης έχει την ικανότητα να μάθει αλληλεπιδράσεις και τοπικά μοτίβα μεταξύ των χαρακτηριστικών των δεδομένων. Στη συνέχεια, οι πίνακες των διανυσμάτων όπου παράγονται ως έξοδο αθροίζονται και αναδιαμορφώνονται σε διανύσματα τα οποία χρησιμοποιούνται ως είσοδο στα συνελικτικά επίπεδα συγχώνευσης. Εφαρμόζεται και εδώ ένα δυσδιάστατο φίλτρο για την τελική εκμάθηση των αλληλεπιδράσεων όλων των δεδομένων των αισθητήρων, καθώς και η τεχνική batch normalization σε κάθε επίπεδο για την ελάττωση της εσωτερικής μεταβλητής μετατόπισης. Τα διανύσματα όπου παράγονται αθροίζονται και αναδιαμορφώνονται πάλι σε διανύσματα τα οποία θα τροφοδοτοτηθούν ως είσοδο στο δίκτυο ανατροφοδότησης όπου ακολουθεί.

Το δίκτυο ανατροφοδότησης αποτελείται από δυο στοιβαγμένες μονάδες ανατροφοδότησης με πύλη, δύο επιπέδων βάθους, καθώς και με την χρήση της τεχνικής dropout ανάμεσα στις συνδέσεις τους. Η τεχνική batch normalization εφαρμόζεται και εδώ ώστε να ελαττωθεί η εσωτερική μεταβλητή μετατόπιση των χρονοσειρών όπου προκύπτουν από τα δεδομένα των αισθητήρων. Το τελευταίο επίπεδο όπου εφαρμόζεται σε αυτό το μοντέλο είναι το επίπεδο εξόδου. Μέχρι αυτό το σημείο, το δίκτυο περιέχει ένα σύνολο διανυσμάτων για κάθε χρονοβήμα. Πλεόν, ανάλογα το πρόβλημα μπορεί να μπει σε ισχύ μια διαφορετική λειτουργία. Για προβλήματα παλινδρόμησης, χρειάζεται να οριστεί ένα dictionary μέσω του οποίου θα γίνει η μάθηση. Για προβλήματα ταξινόμησης, τα χαρακτηριστικά αθροίζονται ως προς τον χρόνο και μέσω της συνάρτησης softmax παράγονται οι πιθανότητες πρόβλεψης των δραστηριοτήτων του μοντέλου.



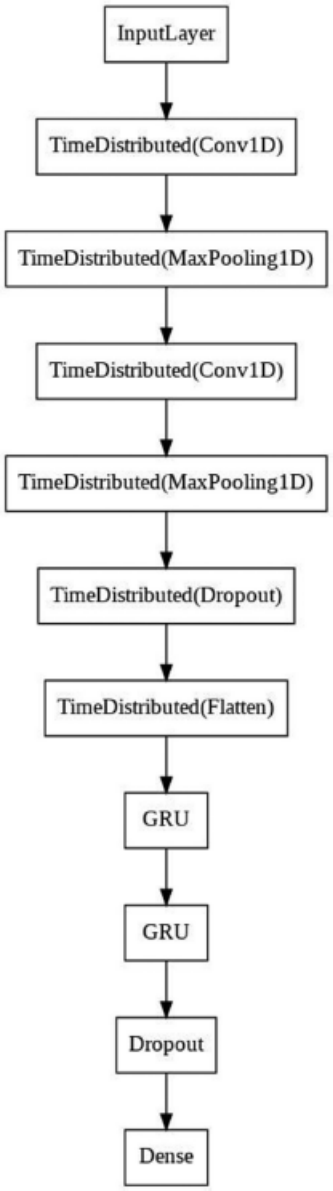

**Σχήμα 3.1: Η αρχιτεκτονική του μοντέλου DeepSense.**

Πέρα από τις προσεγγίσεις οι οποίες περιλαμβάνουν δίκτυα ανατροφοδότησης με πύλη, για προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας εφαρμόζονται και δίκτυα ανατροφοδότησης με μακροπρόθεσμη-βραχυπρόθεσμη μνήμη. Στο [15] η μέθοδος που προτείνεται αποτελείται από τρία μέρη. Το πρώτο μέρος περιλαμβάνει την προεργασία και τον μετασχηματισμό των αρχικών δεδομένων από τους αισθητήρες κίνησης σε ένα δυσδιάστατο πίνακα διανυσμάτων. Το δεύτερο μέρος περιλαμβάνει την αυτόματη εξαγωγή χαρακτηριστικών από τον πίνακα διανυσμάτων μέσω νευρωνικού δικτύου συνέλιξης τριών επιπέδων με τα χαρακτηριστικά όπου συλλέγονται να διαμορφώνουν τον χάρτη χαρακτηριστικών. Το τρίτο μέρος αποτελείται από ένα μοντέλο μακροπρόθεσμης-βραχυπρόθεσμης μνήμης, το οποίο δέχεται ως είσοδο ένα διάνυσμα χαρακτηριστικών ώστε να εντοπίσει τις σχέσεις ανάμεσα σε χρονικές ακολουθίες και ακολουθίες που προκύπτουν από τις δραστηριότητες. Στη συνέχεια, με την χρήση ενός πλήρους ενωμένου επιπέδου πραγματοποιείται ο συνδυασμός πολλαπλών χαρακτηριστικών και με τη βοήθεια της τεχνικής batch normalization αυτά περνούν μέσω της συνάρτησης softmax στην τελική ταξινόμηση δραστηριοτήτων.




**Σχήμα 3.2: Η αρχιτεκτονική ολόκληρης της μεθόδου με την χρήση δικτύων CNN-LSTM.**

Μια προσέγγιση η οποία ενισχύει περαιτέρω την αποτελεσματικότητα των υβριδικών μοντέλων συνέλιξης και ανατροφοδότησης είναι το [6]. Η μέθοδος όπου προτείνεται είναι ένα υβρίδιο νευρωνικών δικτύων συνέλιξης και ανατροφοδότησης με πύλη. Η διαφορά του με το [18] είναι ότι εδώ εφαρμόζεται το μοντέλο βαθιάς μάθησης InceptionTime, το οποίο χρησιμοποιείται για προβλήματα ταξινόμησης χρονοσειρών και η εφαρμογή της αρχιτεκτονικής του γίνεται μέσω της ανοιχτού λογισμικού βιβλιοθήκης AutoML, McFly.




**Σχήμα 3.3: Το προτεινόμενο υβριδικό βαθύ νευρωνικό δίκτυο CNN-GRU.**

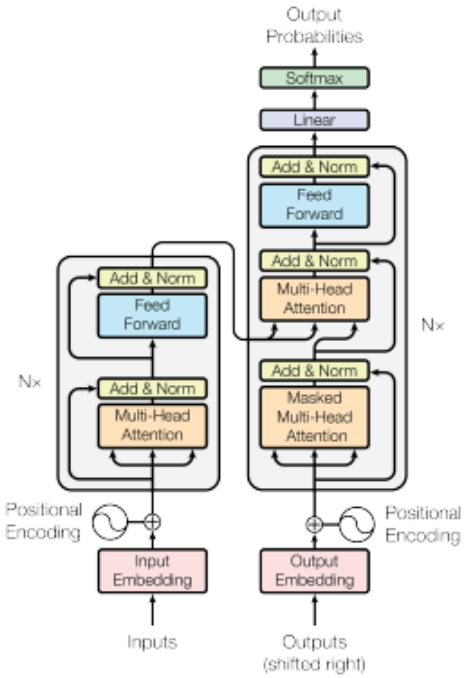
Πιο συγκεκριμένα, το μοντέλο αποτελείται από δυο μονοδιάστατα νευρωνικά δίκτυα συνέλιξης περιτυλιγμένα σε έναν Time Distributed (wrapper). Κάθε μονοδιάστατο επίπεδο συνέλιξης έχει αριθμό φίλτρων όσο και το μέγεθος του πυρήνα του, όπου είναι και ο αριθμός των υπερ-παραμέτρων, οι οποίες ρυθμίζονται κατά τη δημιουργία του μοντέλου με συνάρτηση ενεργοποίησης την ReLU. Ακολουθείται από επίπεδα max pooling, τα οποία προστατεύουν από το overfitting και τον θόρυβο από τα δεδομένα των αισθητήρων. Στη συνέχεια εφαρμόζεται η τεχνική dropout και η έξοδος αυτού μετατρέπεται σε διάνυσμα μιας διάστασης για να τροφοδοθεί με τη σειρά του στα επίπεδα ανατροφοδότησης. Δύο επίπεδα ανατροφοδότησης με πύλη αναλαμβάνουν δράση, τα οποία ακολουθούνται πάλι από την τεχνική dropout και ένα πυκνό επίπεδο το οποίο οδηγεί στην τελική ταξινόμηση των δραστηριοτήτων. Το πλεονέκτημα των δικτύων ανατροφοδότησης με πύλη σε σχέση με τα μακροπρόθεσμης-βραχυπρόθεσμης μνήμης είναι ότι έχουν μικρότερο αριθμό παραμέτρων, με αποτέλεσμα να έχουν μικρότερο χρόνο εκτέλεσης, παρέχοντας ταυτόχρονα ικανοποιητικά αποτελέσματα.

## 3.2 Βαθιά μάθηση και βασικές αρχιτεκτονικές attention

Πληθώρα αρχιτεκτονικών έχει εφαρμοστεί για προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας. Από τα νευρωνικά δίκτυα συνέλιξης μέχρι στα δίκτυα ανατροφοδότησης, καθώς και σε συνδυασμούς τους, οι προσεγγίσεις σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας ευδοκιμούν σε ολόκληρο το φάσμα τους. Ωστόσο, η εμφάνιση εξ' ολοκλήρου attention αρχιτεκτονικών σε προβλήματα επεξεργασίας φυσικής γλώσσας φαίνεται να πετυχαίνει ικανοποιητικότερα αποτελέσματα, συνεπώς σημαντικές έρευνες έχουν πραγματοποιηθεί με σκοπό την εφαρμογή αυτών των αρχιτεκτονικών σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας.

Τα νευρωνικά δίκτυα ανατροφοδότησης με αρχιτεκτονικές encoder-decoder εφαρμόζονται σε προβλήματα sequence-to-sequence με ικανοποιητικά αποτελέσματα. Παρουσιάζουν όμως ένα σημαντικό περιορισμό, για μεγάλου μήκους ακολουθίες, η ικανότητα τους να διατηρούν πληροφορία από τα πρώτα στοιχεία χάνεται καθώς νεα στοιχεία εισέρχονται στην ακολουθία. Οι μηχανισμοί attention όμως ήρθαν για να το αλλάξουν αυτό, καθώς μπορούν να αποκτήσουν πρόσβαση σε όλα τα στοιχεία της ακολουθίας εισόδου και με ένα άθροισμα από βάρη να εξάγουν πληροφορία για ολόκληρη την ακολουθία. Το επιστημονικό άρθρο "Attention is all you need" εισήγαγε το μοντέλο Transformer, το οποίο αποτελείται αποκλειστικά από δίκτυα με attention αρχιτεκτονικές και παρουσιάζοντας κορυφαία αποτελέσματα σε προβλήματα επεξεργασίας φυσικής γλώσσας, δεν άργησε να εφαρμοστεί και σε άλλους τομείς, όπως computer vision, αναγνώριση ανθρώπινης δραστηριότητας, ανάλυση σήματος κ.α.

Για την κατανόηση του μοντέλου Transformer και των δικτύων με attention αρχιτεκτονικές σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας, χρειάζεται να ξεκινήσουμε από την αρχή.




**Σχήμα 3.4: Η αρχιτεκτονική του μοντέλου Transformer.**

### 3.2.1 Τι είναι (self)-Attention

Ο μηχανισμός self-attention είναι μια λειτουργία sequence-to-sequence: μια ακολουθία διανυσμάτων εισέρχεται και μια ακολουθία διανυσμάτων εξέρχεται. Έστω $x_1, x_2, ..., x_t$ τα διανύσματα εισόδου και $y_1, y_2, ..., y_t$ τα αντίστοιχα διανύσματα εξόδου. Όλα τα διανύσματα έχουν διάσταση $k$. Για να πάρουμε διάνυσμα εξόδου $y_i$, ο μηχανισμός self-attention χρειάζεται ένα ζυγισμένο μέσο από όλα τα διανύσματα εισόδου.

$$y_i = \sum_j w_{ij} x_j \tag{3.1}$$

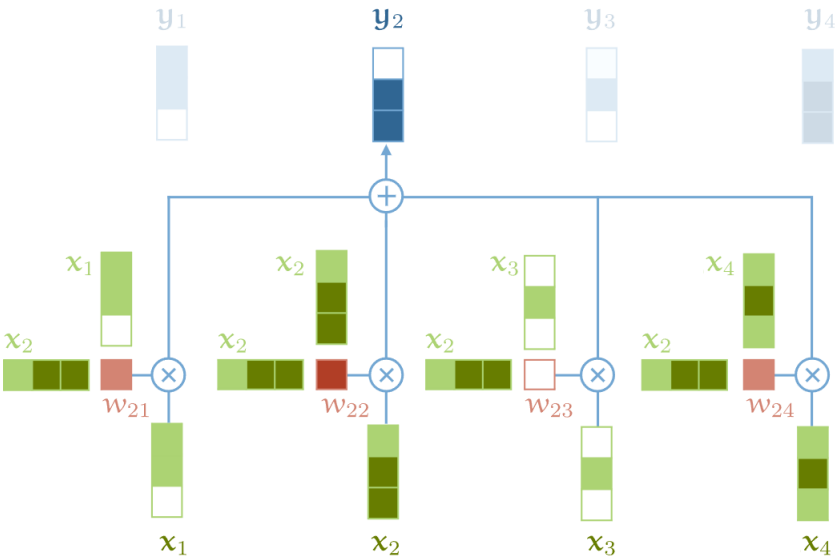
όπου το $j$ ταξινομείται σε ολόκληρη την ακολουθία και τα βάρη αθροίζονται πάνω από όλα τα $j$. Το βάρος $w_{ij}$ δεν είναι μια παράμετρος, αλλά προκύπτει από μια συνάρτηση του $x_i$ και του $x_j$. Η απλούστερη επιλογή για αυτή τη συνάρτηση είναι το εσωτερικό γινόμενο.

$$w'_{ij} = x_i^T x_j \tag{3.2}$$

Το εσωτερικό γινόμενο δίνει τιμές μεταξύ $\pm$ άπειρου, γι' αυτό εφαρμόζεται η συνάρτηση softmax, ώστε να περιορίσει τις τιμές στο διάστημα $[0, 1]$ και να διασφαλίσει ότι ολόκληρη η ακολουθία αθροίζει στο 1, επίσης, η χρήση του εσωτερικού γινομένου ως πράξη μεταξύ δυο διανυσμάτων, εκφράζει πόσο αυτά τα διανυσμάτα σχετίζονται μεταξύ τους.

$$w_{ij} = \frac{\exp w'_{ij}}{\sum_j \exp w'_{ij}} \tag{3.3}$$

Από μαθηματική σκοπιά αυτή είναι η θεμελιώδης λειτουργία του μηχανισμού (self)-attention. Είναι η μόνη λειτουργία σε ολόκληρη την αρχιτεκτονική, η οποία μεταφέρει πληροφορία μεταξύ των διανυσμάτων.




**Σχήμα 3.5: Οπτικοποίηση ενός απλού μηχανισμού self-attention.**

### 3.2.2 Γιατί λειτουργεί

Η ιδέα του μηχανισμού self-attention, όπως παρουσιάστηκε παραπάνω, είναι αρκετά απλή, όμως η κατανόηση γιατί λειτουργεί μπορεί να προβεί σε σύγχυση. Γι' αυτό η βοήθεια ενός παραδείγματος θα βοηθήσει στην εξήγηση του τρόπο με τον οποίο λειτουργεί. Όπως ορίστηκε και παραπάνω, έστω μια ακολουθία εισόδου, η οποία περιλαμβάνει την πρόταση "το παιδί παίζει στο πάρκο". Για να εφαρμόσουμε το μηχανισμό self-attention χρειάζεται απλά να αναθέσουμε στην κάθε λέξη $t$ της πρότασης, ένα διάνυσμα ενσωμάτωσης $v_t$, τις τιμές του οποίου και θα μάθουμε. Με την χρήση λοιπόν του επιπέδου ενσωμάτωσης, όπως ονομάζεται, μετατρέπουμε την ακολουθία της πρότασης το,παιδί,παίζει,στο,πάρκο στην ακολουθία διανυσμάτων

$$v_{to}, v_{paidi}, v_{paizei}, v_{sto}, v_{parko}$$

της οποίας η έξοδος θα είναι μια άλλη ακολουθία διανυσμάτων

$y_{to}, y_{paidi}, y_{paizei}, y_{sto}, y_{parko}$

όπου το διάνυσμα $y_{paidi}$ είναι ένα ζυγισμένο άθροισμα από όλα τα διανύσματα ενσωμάτωσης της πρώτης ακολουθίας, και το βάρος αυτό προκύπτει από την πράξη του εσωτερικού γινομένου του διάνυσματος $v_{paidi}$ με όλα τα υπόλοιπα.

Όπως είναι επόμενο, το πόσο σχετίζονται δυο λέξεις εξαρτάται από το πρόβλημα. Γενικά, ο μηχανισμός self-attention φαίνεται να λειτουργεί με τρόπο τέτοιο ώστε, π.χ. αν ζητήσουμε τα βάρη της λέξης $to$ πιθανόν τα είναι πολύ χαμηλά ή μήδεν μέσω της softmax, καθώς η βαρύτητα που έχει αυτή η λέξη μέσα στην πρόταση δεν είναι μεγάλη. Αν όμως ζητήσουμε τα βάρη της λέξης $paizei$ θα είναι μεγαλύτερα αφού αυτή η λέξη έχει μεγαλύτερη βαρύτητα.

Όπως αναφέρθηκε και παραπάνω το εσωτερικό γινόμενο εκφράζει πόσο σχετίζονται μεταξύ τους δυο διανύσματα, και η συσχέτιση αυτή ορίζεται μέσω της μάθησης όπου αποκτάται για το εκάστοτε πρόβλημα. Τα διανύσματα εξόδου όπου προκύπτουν είναι αθροίσματα από τα βάρη των διανυσμάτων ολόκληρης της αρχικής πρότασης-ακολουθίας και τα βάρη αυτά διαμορφώνονται από τα αποτελέσματα των πράξεων των εσωτερικών γινομένων της λέξης με ολόκληρη την πρόταση.

Μετα το πέρας του ορισμού του μηχανισμού self-attention, μερικές παρατηρήσεις είναι οι εξής:

1. Η παραπάνω είναι η μόνη λειτουργία σε ολόκληρη την αρχιτεκτονική όπου μεταφέρει πληροφορία ανάμεσα στα διανύσματα. Όπως θα δούμε αναλυτικότερα και στο μοντέλο Transformer παρακάτω, οποιαδήποτε άλλη λειτουργία εφαρμόζεται σε κάθε διάνυσμα μέσω της ακολουθίας εισόδου, χωρίς περαιτέρω αλληλεπιδράσεις μεταξύ των διανυσμάτων.

2. Μέχρι στιγμής δεν υπάρχουν παράμετροι. Αυτό που κάνει ένας βασικός μηχανισμός self-attention καθορίζεται πλήρως από το μηχανισμό όπου δημιουργεί την ακολουθία εισόδου.

3. Ο μηχανισμός self-attention αντιμετωπίζει την είσοδο ως ένα σύνολο και όχι ως ακολουθία. Για την ανάπτυξη του μοντέλου Transformer, όπως θα δούμε παρακάτω, εφαρμόζονται μερικές τεχνικές για την αντιμετώπιση αυτού, αλλά γενικά ο μηχανισμός self-attention αγνοεί την φυσική σειρά στις θέσεις εισόδου (permutation equivariant).

### 3.2.3 Το μοντέλο Transformer

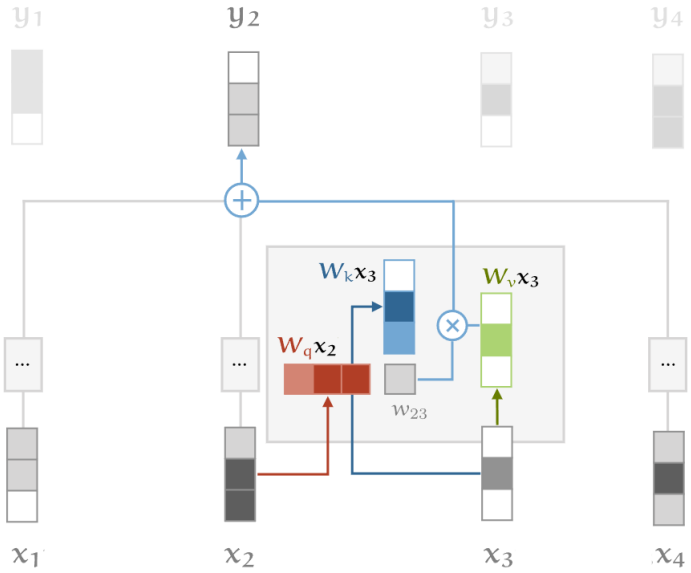
Έχοντας ορίσει τον τρόπο με τον οποίο λειτουργεί ένας βασικός μηχανισμός self-attention, είμαστε σε θέση να αναλύσουμε το μοντέλο Transformer, όπως παρουσιάζεται στο [14]. Πέρα από τη λειτουργία του ίδιου του μηχανισμού, μια βασική τεχνική που εφαρμόζεται στο μοντέλο είναι η εισαγωγή τριών νέων στοιχείων, του Query, του Key και του Value. Βάσει των νέων στοιχείων λοιπόν, κάθε διάνυσμα εισόδου $x_i$ χρησιμοποιείται με τρεις διαφορετικούς τρόπους:

- Συγκρίνεται με όλα τα υπόλοιπα διανύσματα για να διαμορφώσει τα βάρη για τη δική του έξοδο $y_i$
- Συγκρίνεται με όλα τα υπόλοιπα διανύσματα για να διαμορφώσει τα βάρη για την έξοδο του $j$-οστού διανύσματος $y_j$
- Αποτελεί ένα μέρος του αθροίσματος από βάρη μόλις αυτά διαμορφωθούν, για τον υπολογισμό κάθε διανύσματος εξόδου

Συνεπώς, κάθε διάνυσμα εισόδου χρειάζεται να ακολουθήσει και τα τρία αυτά ”μονοπάτια”, και στη συνέχεια πολλαπλασιάζεται με τους τρεις πίνακες $W_q, W_k, W_v$ από τα βάρη που έχουν διαμορφωθεί, διάστασης $k \times k$ και υπολογίζονται οι παρακάτω τρεις γραμμικοί μετασχηματισμοί για το καθένα.

$$q_i = W_q x_i \qquad k_i = W_k x_i \qquad v_i = W_v x_i \tag{3.4}$$

Σε αυτό το σημείο, σύμφωνα με την δεύτερη παρατήρηση παραπάνω, επιτυγχάνεται μια ελεγχόμενη παραγωγή παραμέτρων στα επίπεδα self-attention επιτρέποντας την μετατροπή των διανυσμάτων εισόδου ώστε να μπορούν να ακολουθήσουν καθένα από τα τρία ”μονοπάτια”.




**Σχήμα 3.6: Οπτικοποίηση του μηχανισμού self-attention με τους μετασχηματισμούς Query, Key, Value.**

### 3.2.4 Κλιμακωτό εσωτερικό γινόμενο attention

Η συνάρτηση softmax για μεγάλες τιμές εισόδου μπορεί να γίνει αρκετά ”ευαίσθητη”, και με αυτό εννοείται ότι η εφαπτομένη της παραγώγου της συνάρτησης μικραίνει την κλίση

της και επιβραδύνει τη μάθηση. Αφού ο μέσος όρος του εσωτερικού γινομένου της τιμής Value μεγαλώνει καθώς αυξάνεται η διάσταση ενσωμάτωσης $k$, βοηθάει να ελαττωθεί η τιμή του ώστε η τιμή της εισόδου της συνάρτησης softmax να μην γίνει πολύ μεγάλη:

$$w'_{ij} = \frac{q_i^T k_j}{\sqrt{k}} \tag{3.5}$$

γιατί επιλέγουμε να διαιρέσουμε με το $\sqrt{k}$. Έστω διάνυσμα στον $R^k$ με όλες τις τιμές του μια σταθερά $c$. Η ευκλείδεια απόσταση του είναι $\sqrt{kc}$. Επομένως, με αυτή τη διαίρεση, η αύξηση στην διάσταση δεν αυξάνει και το μήκος του μέσου όρου της απόστασης των διανυσμάτων της ακολουθίας.

Οπότε, για τον υπολογισμό των εσωτερικών γινομένων του Query με όλα τα Keys, το γινόμενο τους διαιρείται με $\sqrt{k}$ και εφαρμόζεται σε αυτό η συνάρτηση softmax, μέσω της οποίας αποκτούνται τα βάρη των Values. Επομένως, από τις (3.2) και (3.3) παίρνουμε ότι:

$$w_{ij} = softmax(w'_{ij}) \tag{3.6}$$

όπου η (3.6) μέσω της (3.5) γίνεται:

$$w_{ij} = softmax(\frac{q_i^T k_j}{\sqrt{k}}) \tag{3.7}$$

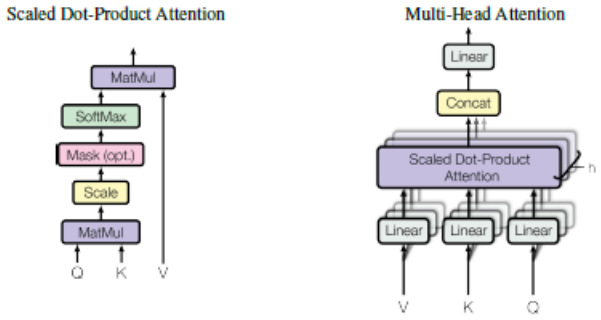
και η (3.1) για διάνυσμα εισόδου $v_j$ γίνεται:

$$y_i = \sum_j w_{ij} v_j \tag{3.8}$$

Όπως αναφέρεται από τους συγγραφείς του [14], το σύνολο των Queries υπολογίζεται ταυτόχρονα και στοιβάζεται σε έναν πίνακα $Q$, ομοίως με τα Keys και Values στους πίνακες $K$ και $V$ αντίστοιχα. Συνεπώς, ο μηχανισμός self-attention, μέσω της (3.8) παράγει ως έξοδο τον πίνακα της μορφής:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{k}})V \tag{3.9}$$

Η παραπάνω σχέση εκφράζει το κλιμακωτό εσωτερικό γινόμενο attention, το οποίο είναι ο πυρήνας του self-attention μηχανισμού [14]. Πέρα όμως από αυτό εφαρμόζονται και άλλες τεχνικές όπως θα δούμε παρακάτω.




**Σχήμα 3.7: (αριστερά) Κλιμακωτό εσωτερικό γινόμενο attention. (δεξιά) Μηχανισμός attention πολλαπλών κεφαλών με παράλληλα επίπεδα.**

### 3.2.5 Μηχανισμός attention πολλαπλών κεφαλών

Ο μηχανισμός self-attention μπορεί να αποκτήσει καλύτερη ικανότητα στο να διακρίνει χαρακτηριστικά, συνδυάζοντας αρκετές κεφαλές τέτοιων μηχανισμών, καθεμία με τους δικούς της πίνακες $W_r^q$, $W_r^k$ $W_r^v$ (για δείκτη $r$) και ονομάζονται κεφαλές attention. Για κάθε διάνυσμα εισόδου $x_i$ κάθε κεφαλή attention παράγει ένα διαφορετικό διάνυσμα εξόδου $y_i^r$. Οι έξοδοι από κάθε κεφαλή αθροίζονται και μετά από ένα γραμμικό μετασχηματισμό, η διάσταση τους επιστρέφει στην τιμή $k$. Ο μηχανισμός attention πολλαπλών κεφαλών επιτρέπει στο μοντέλο να συλλέγει πληροφορίες από διαφορετικούς υπόχωρους αναπαραστάσεων των χαρακτηριστικών και από διαφορετικές θέσεις. Για μια κεφαλή από το [14] δίνεται ο παρακάτω τύπος:

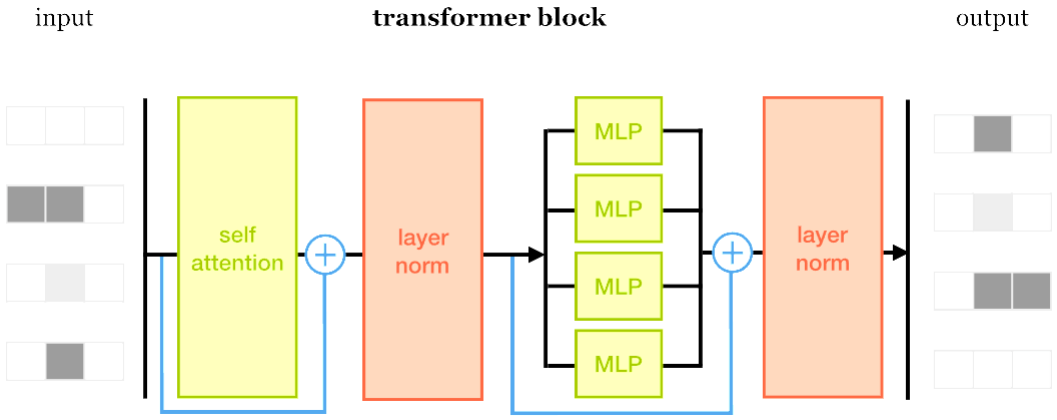
$$MultiHead(Q, K, V) = Concat(head_1, ..., head_h)W^O \tag{3.10}$$

όπου $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$.

Ο ευκολότερος τρόπος για να κατανοήσει κανείς τον τρόπο με τον οποίο λειτουργούν οι κεφαλές attention, είναι να εξετάσει τον μηχανισμό τους σε μικρή κλίμακα και με παράλληλη λειτουργία. Όπως αναφέρθηκε και παραπάνω, κάθε κεφαλή πραγματοποιεί τους δικούς της μετασχηματισμούς Query, Key και Value. Ένα σημαντικό μειονέκτημα που απορρέει είναι, ότι κατά κανόνα, για αριθμό $R$ κεφαλών, ο μηχανισμός self-attention είναι $R$ φορές πιο αργός. Παρ' όλα αυτά, υπάρχει τρόπος ώστε ο μηχανισμός attention με $R$ αριθμό κεφαλών να έχει ταχύτητα εκτέλεσης περίπου ίση με μια κεφαλή, και ταυτόχρονα να διατηρεί τα πλεονεκτήματα που προσφέρει η παράλληλη επεξεργασία. Ο τρόπος για να επιτευχθεί αυτό είναι τα διανύσματα εισόδου να μοιραστούν σε κομμάτια. Εάν το διάνυσμα εισόδου έχει διάσταση $256$, τότε κόβεται σε $8$ κομμάτια με διάσταση $32$ το καθένα. Για κάθε κομμάτι παράγονται πίνακες Query, Key και Value διάστασης $32$, άρα οι πίνακες $W_r^q$, $W_r^k$, $W_r^v$ είναι όλοι διάστασης $32 \times 32$.

### 3.2.6 Το μπλοκ Transformer

Έχοντας ορίσει τις βασικότερες λειτουργίες του μηχανισμού self-attention όπως πρωτοεμφανίστηκε στο [14], μπορούμε να περάσουμε στην ανάλυση ολόκληρου του μπλοκ ενός μοντέλου Transformer, όπως το συναντάμε στις περισσότερες προσεγγίσεις.




**Σχήμα 3.8: Ένα "κλασικό" μπλοκ Transformer.**

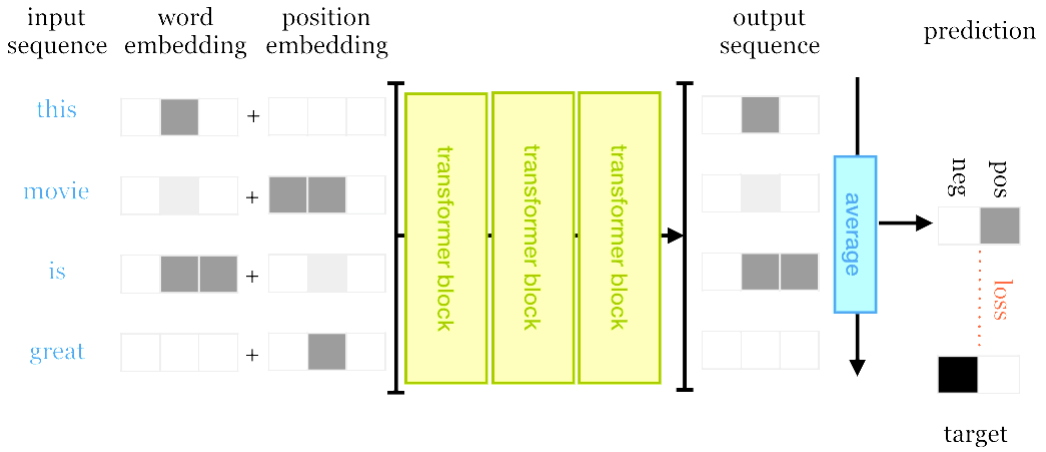
Όπου, από τα αριστερά εφαρμόζονται: ένα επίπεδο self-attention, ένα normalization επίπεδο, ένα επίπεδο τροφοδοσίας προς τα εμπρος (εφαρμόζεται ένας Multi-Layer Perceptron ανεξάρτητα σε κάθε διάνυσμα), και άλλο ένα επίπεδο normalization. Υπολειπόμενες ενώσεις (residual connections) προστίθενται γύρω από τα επίπεδα self-attention και τροφοδοσίας προς τα εμπρος, καθώς και πριν το επίπεδο normalization. Οι υπολειπόμενες ενώσεις και το normalization είναι παγιωμένες τεχνικές και βοηθούν ένα βαθύ νευρωνικό δίκτυο να εκπαιδευτεί γρηγότερα και αποδοτικότερα. Το επίπεδο του normalization εφαρμόζεται μόνο στη διάσταση ενσωμάτωσης.

### 3.2.7 Transformer για προβλήματα ταξινόμησης

Όπως έχει αναφερθεί και παραπάνω, το μοντέλο Transformer και η αρχιτεκτονική self-attention, αρχικά εφαρμόστηκαν για προβλήματα επεξεργασίας φυσικής γλώσσας. Τα προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας όμως, πάντα περιέχουν την λειτουργία της ταξινόμησης, με σκοπό την αναγνώριση των δραστηριοτήτων για το εκάστοτε πρόβλημα. Είναι σημαντικό επομένως, να αναλυθεί ο τρόπος με τον οποίο το μοντέλο Transformer μπορεί να εφαρμοστεί και για τέτοια προβλήματα. Συμφωνα με όσα έχουν αναλυθεί παραπάνω για ταξινόμηση ακολουθιών, στον πυρήνα της αρχιτεκτονικής θα βρίσκεται μια μεγάλη "στίβα" από μπλοκ Transformer. Το μόνο που χρειάζεται να διαμορφωθεί είναι ο τρόπος με τον οποίο θα τροφοδοτηθούν οι ακολουθίες εισόδου στο δίκτυο και η μετατροπή της τελικής ακολουθίας εξόδου σε μια μόνο κατηγορία ταξινόμησης.

Για την παραγωγή της ακολουθίας εξόδου με στόχο την ταξινόμηση, ο πιο κοινός τρόπος, ο οποίος περιλαμβάνει επίπεδα sequence-to-sequence, είναι η εφαρμογή μιας καθολικά μέσης ομαδοποίησης στην τελική ακολουθία εξόδου και στη συνέχεια η αντιστοίχιση των

αποτελεσμάτων της σε ένα διάνυσμα ταξινόμησης, στο οποίο έχει εφαρμοστεί η συνάρτηση softmax.




**Σχήμα 3.9: Επισκόπηση ενός απλού μοντέλου Transformer για προβλήματα ταξινόμησης ακολουθιών.**

Όπως βλέπουμε στο σχήμα 3.9, η ακολουθία εξόδου αθροίζεται και ένας μέσος όρος παράγεται ως ένα μοναδικό διάνυσμα, το οποίο αντιπροσωπεύει ολόκληρη την ακολουθία. Αυτό το διάνυσμα συγκρίνεται με ένα άλλο διάνυσμα, το οποίο περιέχει ένα στοιχείο ανά κατατηγορία δραστηριότητας και μέσω της συνάρτησης softmax παράγονται οι πιθανότητες ομοιότητας τους.

Από την πλευρά της ακολουθίας εισόδου, όπως έχει αναφερθεί και παραπάνω, η στοίχιση των επιπέδων στο δίκτυο επιδέχεται μεταθέσεις, ενώ η τελική καθολική μέση ομαδοποίηση όχι, επομένως το δίκτυο στο σύνολο του δεν επιδέχεται μεταθέσεις. Πιο απλά, αν ανακατέψουμε τις λέξεις σε μια πρόταση, θα πάρουμε ακριβώς την ίδια έξοδο, όποια βάρη και αν προκύψουν. Μια λύση σε αυτό το πρόβλημα είναι να δημιουργηθεί ένα δεύτερο διάνυσμα (ίσου μήκους με το διάνυσμα εισόδου), το οποίο θα αντιπροσωπεύει τις θέσεις των λέξεων στην συγκεκριμένη πρόταση και στη συνέχεια να προστεθεί στην ενσωμάτωση της λέξης. Υπάρχουν δυο επιλογές.

### 3.2.8 Ενσωμάτωση θέσης

Απλά εισάγονται διανύσματα θέσεων μήκους τόσου όσου περιμένουμε ότι θα είναι η ακολουθία. Το μειονέκτημα είναι ότι χρειάζεται να εκπαιδευτούν διαφορετικών μηκών ακολουθίες, αλλιώς οι σχετικές ενσωματώσεις θέσεων δεν θα εκπαιδευτούν. Το πλεονέκτημα είναι ότι λειτουργεί αποτελεσματικά και είναι εύκολο στην εφαρμογή.

### 3.2.9 Κωδικοποίηση θέσης

Η κωδικοποίηση θέσης λειτουργεί με παρόμοιο τρόπο με την ενσωμάτωση. Η μόνη διαφορά είναι ότι τα διανύσματα θέσεων δεν είναι γνωστά, αλλά επιλέγονται μέσω μιας συ-

νάρτησης $f : N \rightarrow R^k$ η οποία αντιστοιχίζει τις θέσεις των διανυσμάτων με πραγματικές τιμές και ”αφήνει” το δίκτυο να βρει τον τρόπο που θα ερμηνεύσει τις κωδικοποιήσεις. Το πλεονέκτημα είναι ότι μέσω μιας σωστά επιλεγμένης συνάρτησης, το δίκτυο μπορεί να αντιμετωπίσει ακολουθίες μεγαλύτερου μήκους από αυτές με τις οποίες εκπαιδεύτηκε. Τα μειονεκτήματα είναι ότι η επιλογή της συνάρτησης κωδικοποίησης είναι μια σύνθετη υπερ-παράμετρος και περιπλέκει κάπως τα πράγματα.

### 3.2.10 Γιατί ονομάστηκε έτσι

Πριν την εμφάνιση αρχιτεκτονικών self-attention, τα μοντέλα ακολουθιών απαρτιζόταν κυρίως από δίκτυα συνέλιξης ή ανατροφοδότησης. Κάποια στιγμή ανακαλύφθηκε ότι θα βοηθούσε να προστεθούν σε αυτά μηχανισμοί attention, δηλαδή αντί να τροφοδοτείται η ακολουθία εξόδου των προηγούμενων επιπέδων απευθείας στην είσοδο του επόμενου επιπέδου, ένας ενδιάμεσος μηχανισμός εμφανίστηκε, ο οποίος ”αποφασίζει” ποια στοιχεία της εισόδου είναι σχετικά με μια συγκεκριμένη λέξη της εξόδου.

Πέρα λοιπόν από τον πολλαπλασιασμό διανυσμάτων, εισάγονται και τρία νέα στοιχεία στον μηχανισμό λειτουργίας του, τα οποία περιγράφονται με τον εξής τρόπο: Η είσοδος ονομάζεται Values. Ο μηχανισμός όπου περιγράφθηκε (ο οποίος εκπαιδεύεται με τη συνεχή ανανέωση από τα βάρη) αντιστοιχίζει ένα Key σε κάθε Value. Στη συνέχεια κάποιος άλλος μηχανισμός, αντιστοιχίζει ένα Query. Με αυτή τη δομή δεδομένων του ζευγαριού Key-Value αναμένεται μόνο ένα στοιχείο να έχει ένα Key που αντιστοιχεί σε ένα Query, το οποίο και επιστρέφεται ως έξοδο όταν εκτελεστεί αυτό το Query. Ο μηχανισμός self-attention ακολουθεί μια εκδοχή του παραπάνω μηχανισμού. Κάθε Key αντιστοιχεί σε κάποιο βαθμό στο Query. Επιστρέφονται όλα τα Keys ως έξοδο και παίρνοντας το μέσο όρο από τα βάρη τους, υπολογίζεται πόσο το κάθε Key ταιριάζει με το Query.

Η μεγάλη ανακάλυψη του μηχανισμού self-attention είναι ότι ο μηχανισμός από μόνος του είναι αρκετά ισχυρός ώστε να εκτελέσει ολόκληρη την εκπαίδευση, όπως ακριβώς αναφέρει ο τίτλος του [14]. Τα Key, Query και Value είναι και τα τρία τα ίδια διανύσματα, με μικρές διαφορές γραμμικών μετασχηματισμών. ”Φροντίζουν” τον εαυτό τους και στοιβάζονται με τρόπο τέτοιο, ώστε ο μηχανισμός self-attention να προσφέρει αρκετή μη γραμμική και αντιπροσωπευτική δύναμη για να ”μάθει” αρκετά πολύπλοκες συναρτήσεις.

# 4. ΜΕΘΟΔΟΛΟΓΙΑ

Η μεθοδολογία όπου εφαρμόστηκε για την υλοποίηση του μοντέλου βασίστηκε στο [3] και περιγράφει τον τρόπο με τον οποίο το μοντέλο Transformer προσαρμόζεται για προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας με δεδομένα από αισθητήρες κίνησης. Αρχικά, το μοντέλο δημιουργήθηκε για προβλήματα επεξεργασίας φυσικής γλώσσας, όμως η προσαρμοστικότητα του μηχανισμού του δεν άργησε να εφαρμοστεί και σε άλλους τομείς. Η γενική μεθοδολογία όπου εφαρμόζεται για τη ροή πληροφορίας που εισέρχεται σε ένα νευρωνικό δίκτυο περιγράφεται σε τρία βήματα: προ-επεξεργασία, είσοδος των δεδομένων στο μοντέλο και μετα-επεξεργασία.

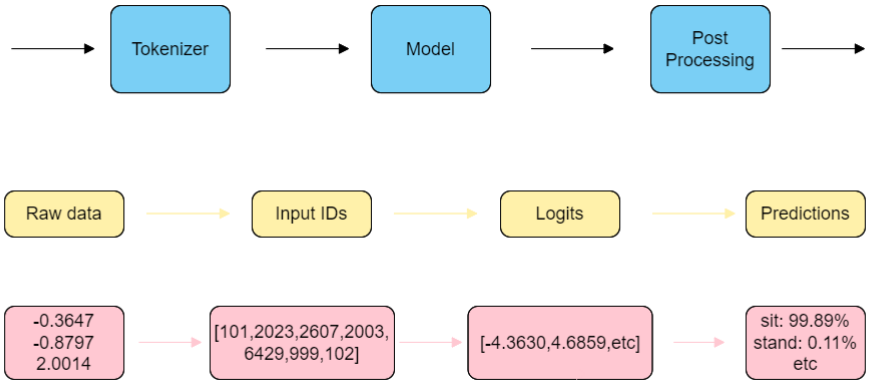
## 4.1 Προ-επεξεργασία

Όπως συμβαίνει και σε άλλα νευρωνικά δίκτυα, τα μοντέλα Transformer δεν μπορούν να επεξεργαστούν κατευθείαν δεδομένα σε μορφή κειμένου, εικόνων, ή στην περίπτωση μας μετρήσεων από αισθητήρες κίνησης, συνεπώς χρειάζεται να μετατραπούν σε μια ακολουθία αριθμών (κωδικών εισόδου), η οποία μπορεί να κατανοηθεί από το μοντέλο. (σχήμα 4.1).

Η διαδικασία αυτή είναι υπεύθυνη για τον χωρισμό αυτών των δεδομένων σε μικρότερα κομμάτια ή ακόμα και σε σύμβολα, τα οποία ονομάζονται tokens (ενδείξεις), την αντιστοίχιση τους σε ακέραιους όπου κατανοούνται από το μοντέλο και τέλος, την ενδεχόμενη εισαγωγή παραπάνω πληροφορίας η οποία θα είναι χρήσιμη για το εκάστοτε πρόβλημα που επιλύει το μοντέλο. Η διαδικασία αυτή χρειάζεται να γίνει με ακριβώς τον ίδιο τρόπο με τον οποίο το μοντέλο προ-εκπαιδεύτηκε και στη δική μας περίπτωση αυτό συνέβη μέσω του προ-εκπαιδευμένου μοντέλου BEiT [1]. Τα μοντέλα Transformer δέχονται ως είσοδο μόνο τένσορες, συνεπώς το τελευταίο βήμα είναι η μετατροπή αυτών των κωδικών εισόδου, οι οποίοι τροφοδοτούνται κατευθείαν στο μοντέλο. Με αυτόν τον τρόπο παράγεται μια λίστα η οποία περιέχει μοναδικά αναγνωριστικά για τα tokens κάθε εισόδου.

## 4.2 Είσοδος των δεδομένων στο μοντέλο

Το μοντέλο Transformer που εφαρμόστηκε παρουσιάζεται στην παρακάτω εικόνα(σχήμα 4.2). Σε αντίθεση με προγενέστερα μοντέλα Transformer όπου βασίστηκε [14][4], εδώ τα δεδομένα σημάτων από τις χρονοσειρές των αισθητήρων κίνησης διοχετεύονται κατευθείαν στο μπλοκ του encoder μαζί με την πληροφορία η οποία αποφασίζει την θέση των χαρακτηριστικών στις χρονοσειρές σημάτων (ενσωμάτωση θέσης). Σε αυτό το σημείο γίνεται η χρήση του προ-εκπαιδευμένου μοντέλου Transformer, το οποίο δέχεται την είσοδο των δεδομένων, παράγει τα χαρακτηριστικά από την επεξεργασία τους και καθορίζει τα βάρη τους. Για κάθε είσοδο στο μοντέλο παράγεται ένα διάνυσμα υψηλής διάστασης το οποίο αποτελεί την είσοδο στην κεφαλή του μηχανισμού attention. Η έξοδος της κεφα-

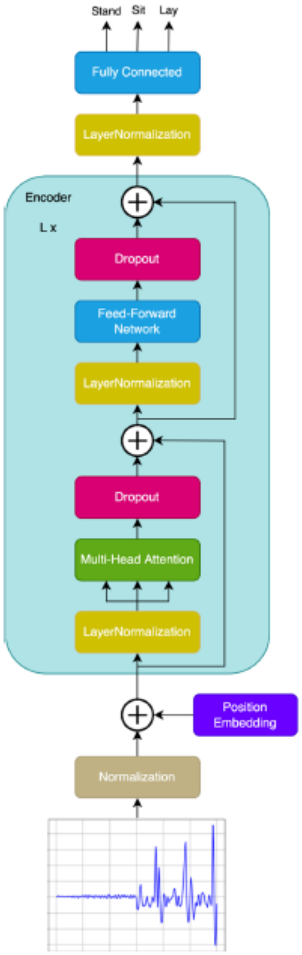



**Σχήμα 4.1: Ροή πληροφορίας του μοντέλου Transformer.**

λής αυτής επομένως, χρησιμοποιώντας το ήδη προ-εκπαιδευμένο μοντέλο Transformer, μπορεί πλέον να προσαρμοστεί για την επίλυση πληθώρας προβλημάτων. Για προβλήματα ταξινόμησης δραστηριοτήτων, η έξοδος αυτή χρειάζεται να μετατραπεί σε κατανομή πιθανοτήτων ώστε να οδηγήσει στην αναγνώριση τους.

## 4.3 Μετα-επεξεργασία

Όπως αναφέρθηκε και παραπάνω, η κάθε κεφαλή του μοντέλου παίρνοντας ως είσοδο ένα διάνυσμα υψηλής διάστασης, παράγει ως έξοδο διανύσματα τα οποία περιέχουν τιμές όσες και το πλήθος των δραστηριοτήτων που χρειάζεται να αναγνωριστούν για το εκάστοτε dataset. Οι τιμές αυτές όμως δεν αναπαριστούν πιθανότητες, συνεπώς χρειάζεται η χρήση συνάρτησης ενεργοποίησης π.χ. softmax. Στην περίπτωση του συγκεριμένου μοντέλου όμως, το επίπεδο εξόδου του είναι γραμμικό ώστε να επιτυγχάνεται μεγαλύτερη υπολογιστική ταχύτητα ακόμα και από συσκευές χαμηλότερης υπολογιστικής ισχύς, Επίσης, η μέγιστη τιμή που αντιστοιχεί στη προβλεπόμενη δραστηριότητα, αφού χρησιμοποιείται μόνο κατά τη διάρκεια της εκπαίδευσης μέσω της συνάρτησης απώλειας, μπορεί να χρησιμοποιηθεί και πριν την εφαρμογή της softmax. Έτσι, εκτός από το πλεονέκτημα της υπολογιστικής ταχύτητας, οι λογαριθμικοί υπολογισμοί που θα προέκυπταν από την χρήση της softmax δεν θα περιείχαν αρνητικούς αριθμούς, σε αντίθεση με το γραμμικό επίπεδο όπου εφαρμόζεται εδώ.




**Σχήμα 4.2: Η αρχιτεκτονική του μοντέλου Transformer.**

# 5. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Το μοντέλο Transformer που υλοποιήθηκε, εξετάστηκε αρχικά αν ικανοποιεί τα αποτελέσματα όπου αναφέρουν οι συγγραφείς του [3] και στη συνέχεια συγκρίθηκε με μοντέλα διαφορετικών αρχιτεκτονικών, τα οποία εφαρμόζονται ευρέως σε προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας. Τα πειραματικά αποτελέσματα που θα ακολουθήσουν περιλαμβάνουν τη σύγκριση με το υβριδικό νευρωνικό δίκτυο CNN-GRU όπως ακριβώς εφαρμόστηκε στο [6]. Μια σύντομη περιγραφή των datasets και της διαδικασίας της προεπεξεργασίας θα πάρει μέρος και στη συνέχεια θα ακολουθήσει η σύγκριση των μεθόδων και τα συμπεράσματα που προκύπτουν.

## 5.1 Datasets

Δυο datasets χρησιμοποιήθηκαν για τις πειραματικές μετρήσεις των μεθόδων.Το dataset KU-HAR [12] περιέχει 18 δραστηριότητες από 90 συμμετέχοντες. Το αρχείο csv περιέχει $20.750$ υποδείγματα δραστηριοτήτων συνολικής διάρκειας 3 δευτερολέπτων με 100 μετρήσεις ανά δευτερόλεπτο για καθένα από τους δυο αισθητήρες κίνησης. Συνεπώς, κάθε γραμμή του αρχείου αντιπροσωπεύει και ένα δείγμα δραστηριότητας, η οποία αποτελείται από 300 μετρήσεις για καθένα από τους δυο τρισδιάστατους αισθητήρες κίνησης.

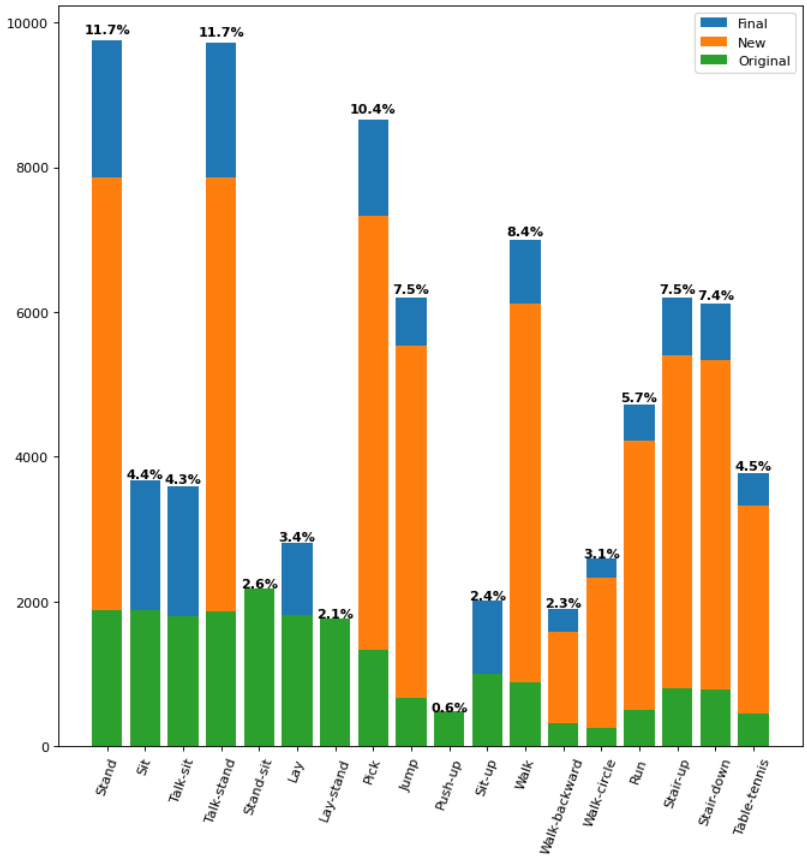
Η μέθοδος data augmentation χρησιμοποιήθηκε ώστε να αυξηθεί ο αριθμός των υποδειγμάτων για τη διαδικασία της εκπαίδευσης και ο τρόπος με τον οποίο αυτό συνέβη, είναι ο συνδυασμός δραστηριοτήτων σε ζευγάρια τα οποία εμφανίζονται στην καθημερινότητα ενός ατόμου. Δημιουργήθηκαν όλοι οι δυνατοί συνδυασμοί από δραστηριότητες εκτός από όμοιους συνδυασμούς τους. Τα δεδομένα από τους αισθητήρες κίνησης συνδυάζονται στα ζευγάρια δραστηριοτήτων με αριθμό ζευγαριών-δειγμάτων ίσο με 100. Στη συνέχεια τα δεδομένα αποθηκεύονται σε δυο πίνακες signals και labels αντίστοιχα, καθώς επίσης υπολογίζονται η μέση και η τυπική τους απόκλιση.

Τα δεδομένα από τους αισθητήρες κίνησης που προκύπτουν από τους συνδυασμούς τους σε ζευγάρια δραστηριοτήτων, αποθηκεύονται σε δυο νέους πίνακες new signals και new labels αντίστοιχα, με πλήθος ίσο με το μικρότερο αριθμό υποδειγμάτων της εκάστοτε δραστηριότητας. Στη συνέχεια τα δεδομένα που προκύπτουν από τους αρχικούς πίνακες και από τους πίνακες των συνδυασμών σε ζευγάρια, συγχωνεύονται σε δυο τελικούς πίνακες final signals και final labels αντίστοιχα και υπολογίζεται η μέση και τυπική τους απόκλιση. Τα τελικά δεδομένα αποθηκεύονται σε ένα νέο dataset. Παράχθηκαν $83.129$ παραδείγματα από τα αρχικά $20.750$ και στη συνέχεια το dataset χωρίστηκε σε σύνολα training, validation και testing με αναλογία $70:15:15$ αντίστοιχα (σχήμα 5.1).

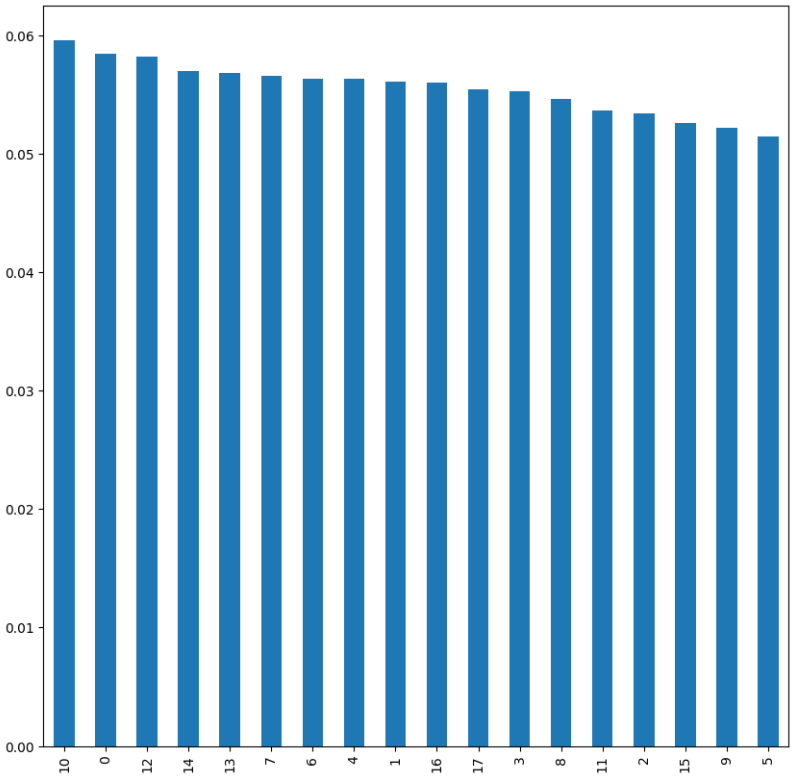
Το dataset WISDM [16] περιέχει και αυτό 18 δραστηριότητες από 51 συμμετέχοντες με συνολικό αριθμό υποδειγμάτων $91.995$ και η συνολική διάρκεια κάθε δραστηριότητας είναι 3 λεπτά. Η ιδιαιτερότητα αυτού του dataset είναι ότι περιέχει δραστηριότητες όπως βούρτσισμα δοντιών, δίπλωμα ρούχων, ή κατανάλωση διάφορων ειδών τροφής, ευρίνοντας το πεδίο εφαρμόγης του πέρα από δραστηριότητες άσκησης ή κίνησης που συνήθως

χρησιμοποιούνται. Επίσης, το πλήθος των μετρήσεων κάθε δευτερόλεπτο διαφέρει ανα δραστηριότητα, για το λόγο ότι ορισμένες δραστηριότητες δεν εκτελέστηκαν από κάποιους συμμετέχοντες στον προβλεπόμενο χρόνο.

Το dataset περιέχει μετρήσεις από συσκευές smartphone και smartwatch, όμως μόνο τα δεδομένα των αισθητήρων από τα smartphones χρησιμοποιήθηκαν και στη συνέχεια συγχωνεύτηκαν σε ένα dataframe. Στο συγκεκριμένο dataset δεν χρειάζεται η εξισορρόπηση των κλάσεων των δραστηριοτήτων, αφού τα δεδομένα είναι ομοιόμορφα κατανεμημένα σε αυτές. Στη συνέχεια, τα δεδομένα χωρίζονται σε χρονικά διαστήματα των 10 δευτερολέπτων με επικάλυψη $50\%$, ώστε η μορφή τους να είναι κατάλληλη για την εισαγωγή σε μοντέλα βαθιάς μάθησης. Το dataset χωρίστηκε σε σύνολα training, validation και testing με αναλογία $60 : 20 : 20$ αντίστοιχα (σχήμα 5.2).




**Σχήμα 5.1: Η κατανομή των δραστηριότητων του dataset KU-HAR μετά την επαύξηση**




**Σχήμα 5.2: Η κατανομή των δραστηριότητων του dataset WISDM**

## 5.2 Αποτελέσματα Transformer

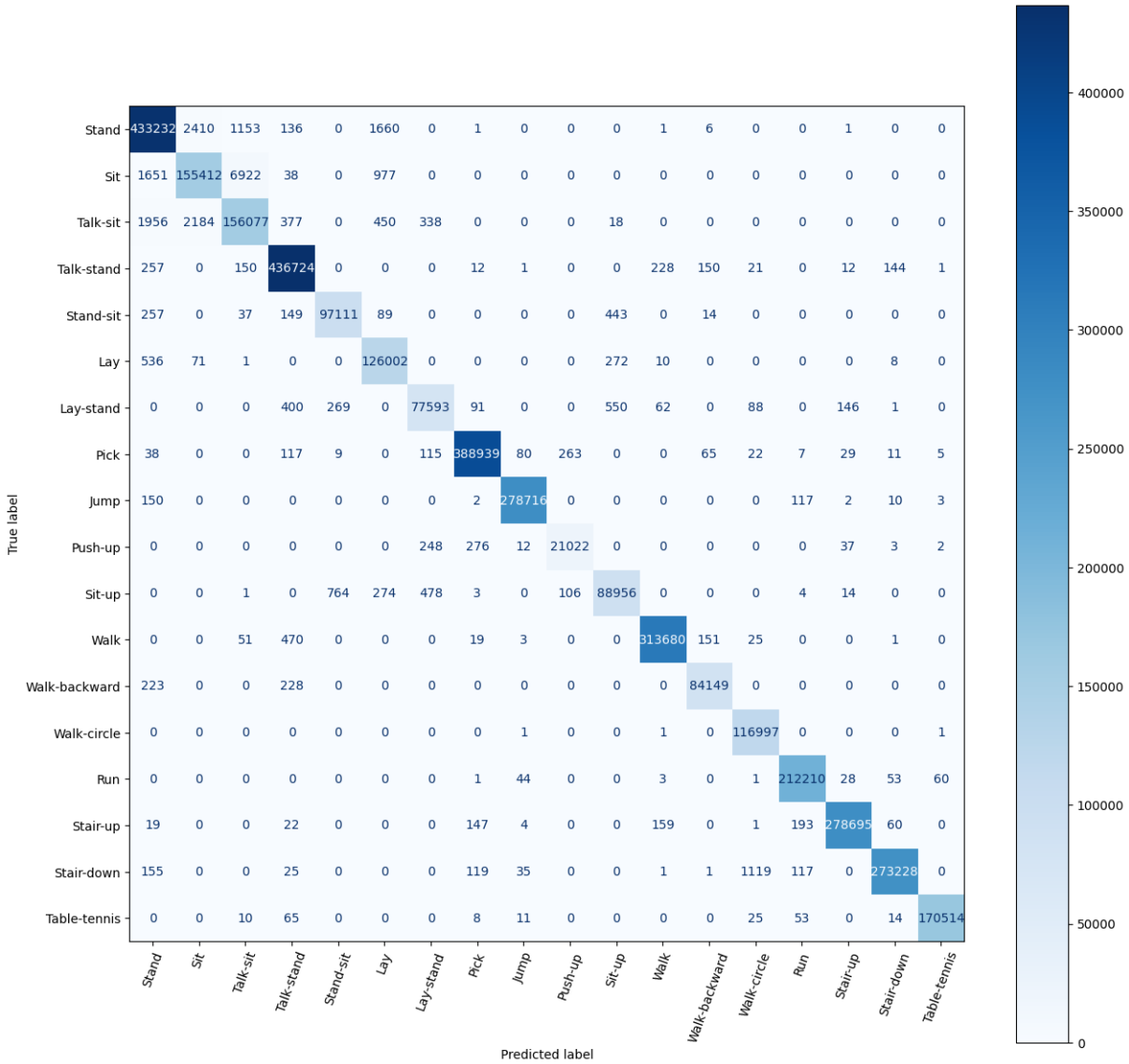
Για τη διαδικασία της εκπαίδευσης, η βιβλιοθήκη Wandb.ai χρησιμοποιήθηκε για την οπτικοποίηση και τη συλλογή αποτελεσμάτων που προκύπτουν από την εφαρμογή του μοντέλου και μέσω αυτής αποθηκεύονται όλες οι παράμετροι του. Η μέθοδος Transformer [3] εφαρμόστηκε και για τα δυο datasets και τα βάρη από τη διαδικασία της εκπαίδευσης χρησιμοποιήθηκαν για τη διαδικασία της δοκιμής. Τα αποτελέσματα αναφέρονται παρακάτω.

**Πίνακας 5.1: Πληροφορίες μοντέλου Transformer**

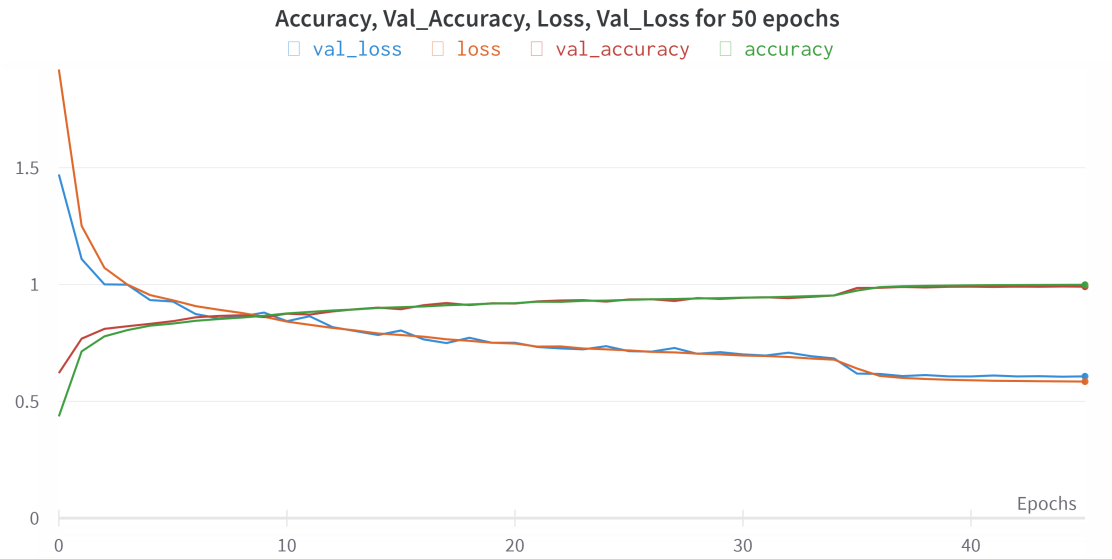
| Model: Transformer | | |
|---|---|---|
| Layer (type) | Output Shape | Parameters # |
| Normalization | Multiple | 13 |
| Positional Embedding | Multiple | 39296 |
| Encoder | Multiple | 462080 |
| Encoder 1 | Multiple | 462080 |
| Encoder 2 | Multiple | 462080 |
| Layer Normalization | Multiple | 256 |
| Dense | Multiple | 2322 |
| Total Parameters: 1.428.127<br>Trainable Parameters: 1.428.114<br>Non-trainable Parameters: 13 | | |

**Πίνακας 5.2: Αποτελέσματα της μεθόδου Transformer για το dataset KU-HAR**

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Stand | 0.988 | 0.988 | 0.988 | 438600 |
| Sit | 0.971 | 0.942 | 0.956 | 165000 |
| Talk-sit | 0.949 | 0.967 | 0.958 | 161400 |
| Talk-stand | 0.995 | 0.998 | 0.997 | 437700 |
| Stand-sit | 0.989 | 0.990 | 0.990 | 98100 |
| Lay | 0.973 | 0.993 | 0.983 | 126900 |
| Lay-stand | 0.985 | 0.980 | 0.982 | 79200 |
| Pick | 0.998 | 0.998 | 0.998 | 389700 |
| Jump | 0.999 | 0.999 | 0.999 | 279000 |
| Push-up | 0.983 | 0.973 | 0.978 | 21600 |
| Sit-up | 0.986 | 0.982 | 0.984 | 90600 |
| Walk | 0.999 | 0.998 | 0.998 | 314400 |
| Walk-backward | 0.995 | 0.995 | 0.995 | 84600 |
| Walk-circle | 0.989 | 1.000 | 0.994 | 117000 |
| Run | 0.998 | 0.999 | 0.998 | 212400 |
| Stair-up | 0.999 | 0.998 | 0.998 | 279300 |
| Stair-down | 0.999 | 0.994 | 0.997 | 274800 |
| Table-tennis | 1.000 | 0.999 | 0.999 | 170700 |
| Accuracy | | | 0.992 | 3741000 |
| Macro avg | 0.989 | 0.988 | 0.989 | 3741000 |
| Weighted avg | 0.992 | 0.992 | 0.992 | 3741000 |




**Σχήμα 5.3: Ο πίνακας confusion matrix για τις 18 δραστηριότητες του dataset KU-HAR**

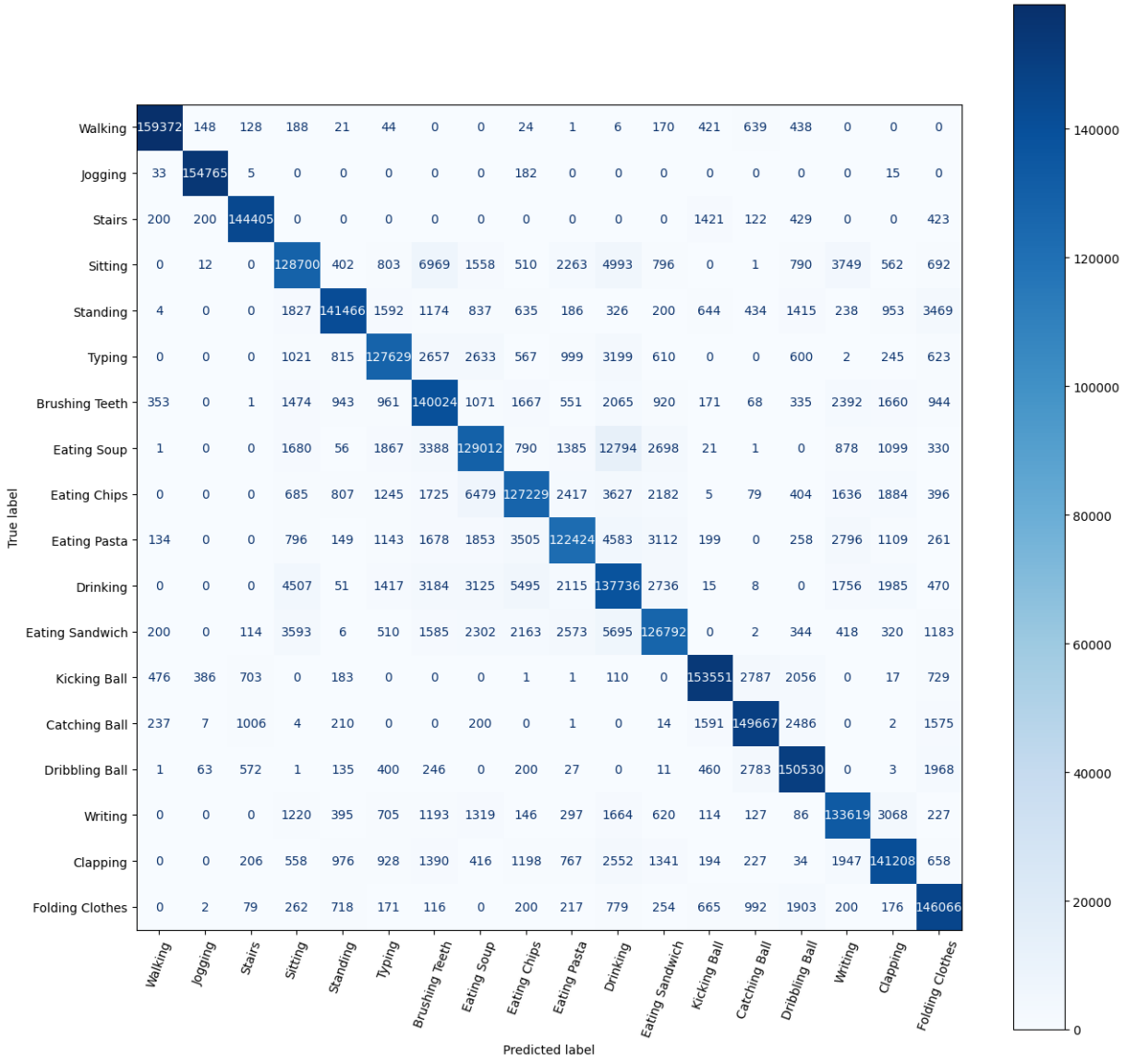



**Σχήμα 5.4: Συγκεντρωτικός πίνακας των παραπάνω δεικτών για το dataset KU-HAR**

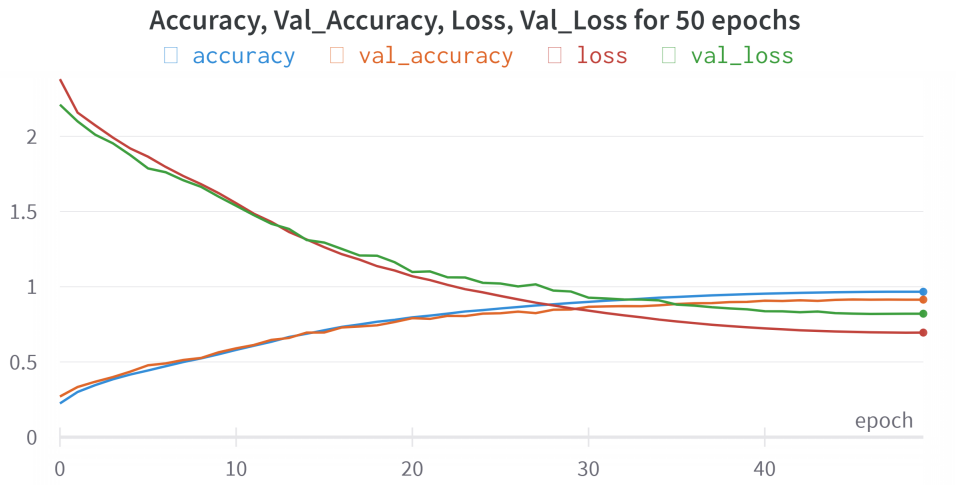
Στο dataset KU-HAR εξετάζεται η ικανότητα της μεθόδου Transformer να ανιχνεύει την εναλλαγή δραστηριοτήτων μέσα σε μια χρονοσειρά. Για το λόγο αυτό οι συγγραφείς του [3] επιλέγουν να επαυξήσουν τα δεδομένα στο στάδιο της προ-επεξεργασίας και με ζευγάρια δραστηριοτήτων, με τα αποτελέσματα που προκύπτουν από τον πίνακα 5.2 να καταδεικνύουν ότι το μοντέλο Transformer καταφέρνει να διακρίνει και τις επιμέρους δραστηριότητες μέσα σε μια χρονοσειρά. Μια παρατήρηση που προκύπτει είναι ότι το dataset είναι μη-ισορροπημένο με αποτέλεσμα να υπάρχει κίνδυνος overfit, όμως από τα σχήματα 5.1, 5.2 και όπως θα δειχθεί και παρακάτω, η μέθοδος διαθέτει ισχυρή ικανότητα εκμάθησης. Οι δείκτες precision, recall και F1-score χρησιμοποιούνται για να ενισχύσουν την παραπάνω παρατήρηση. Ο δείκτης precision δείχνει με υψηλό ποσοστό ότι οι θετικές ταυτοποιήσεις είναι πράγματι σωστές, και ο δείκτης recall, επίσης με μεγάλο ποσοστό, δείχνει ότι το μεγαλύτερο πλήθος των πραγματικών θετικών αποτελεσμάτων εντοπίστηκε σωστά. Η σύγκλιση των παραπάνω δεικτών γύρω από τον σχεδόν ίδιο αριθμό δείχνει ότι ο αρμονικός μέσος των δεδομένων του μοντέλου διατηρείται, ισχυρισμό που ενισχύει περισσότερο ο συνδυασμός των δεικτών macro-averaged F1-score και weighted F1-score. Τέλος, οι καμπύλες του σχήματος 5.2 δείχνουν ότι οι συναρτήσεις απώλειας και accuracy του μοντέλου στα σύνολα εκπαίδευσης και επικύρωσης δεν αποκλίνουν και παραμένούν σταθερές μετά από 30 επαναλήψεις, με αποτέλεσμα το μοντέλο να αποφεύγει τον κίνδυνο overfit και να έχει ισχυρή ικανότητα γενίκευσης.

**Πίνακας 5.3: Αποτελέσματα της μεθόδου Transformer για το dataset WISDM**

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Walking | 0.990 | 0.986 | 0.988 | 161600 |
| Jogging | 0.995 | 0.998 | 0.997 | 155000 |
| Stairs | 0.981 | 0.981 | 0.981 | 147200 |
| Sitting | 0.878 | 0.842 | 0.860 | 152800 |
| Standing | 0.960 | 0.910 | 0.935 | 155400 |
| Typing | 0.915 | 0.901 | 0.908 | 141600 |
| Brushing Teeth | 0.847 | 0.900 | 0.873 | 155600 |
| Eating Soup | 0.855 | 0.827 | 0.841 | 156000 |
| Eating Chips | 0.880 | 0.844 | 0.862 | 150800 |
| Eating Pasta | 0.899 | 0.850 | 0.874 | 144000 |
| Drinking | 0.765 | 0.837 | 0.799 | 164600 |
| Eating Sandwich | 0.890 | 0.858 | 0.874 | 147800 |
| Kicking Ball | 0.963 | 0.954 | 0.958 | 161000 |
| Catching Ball | 0.948 | 0.953 | 0.950 | 157000 |
| Dribbling Ball | 0.929 | 0.956 | 0.942 | 157400 |
| Writing | 0.893 | 0.923 | 0.908 | 144800 |
| Clapping | 0.915 | 0.913 | 0.914 | 154600 |
| Folding Clothes | 0.913 | 0.956 | 0.934 | 152800 |
| Accuracy | | | 0.911 | 2760000 |
| Macro avg | 0.912 | 0.911 | 0.911 | 2760000 |
| Weighted avg | 0.912 | 0.911 | 0.911 | 2760000 |




**Σχήμα 5.5: Ο πίνακας confusion matrix για τις 18 δραστηριότητες του dataset WISDM**




**Σχήμα 5.6: Συγκεντρωτικός πίνακας των παραπάνω δεικτών για το dataset WISDM**

Σε αντίθεση με το KU-HAR, το dataset WISDM είναι ισορροπημένο σχετικά με την κατανομή των δραστηριοτήτων του. Αν και αυτό συνήθως δεν παρατηρείται σε δεδομένα πραγματικών συνθηκών, η μέθοδος Transformer εφαρμόζεται και εδώ για να ενισχύσει την αποτελεσματικότητα της και σε αυτή την κατηγορία. Η επαύξηση των δεδομένων δεν χρησιμοποιήθηκε σε αυτό το dataset καθώς ούτε ο συνδυασμός των δραστηριοτήτων σε ζευγάρια, αλλά εφαρμόστηκε η μέθοδος στο στάδιο προ-επεξεργασίας του [6]. Να σημειωθεί ότι το μέγεθος του dataset είναι επαρκές για την εφαρμογή του σε μεθόδους βαθιάς μάθησης. Τα αποτελέσματα από τον πίνακα 5.3 και το σχήμα 5.3 καταδεικνύουν ότι το μοντέλο Transformer επιτυγχάνει άριστα αποτελέσματα στην ταξινόμηση δραστηριοτήτων, με τους δείκτες precision, recall και F1-score να μην αποκλίνουν μεταξύ τους, διατηρώντας τον αρμονικό μέσο των δεδομένων του μοντέλου, και οδηγώντας στην υψηλή αποτελεσματικότητα του. Τέλος, και εδώ παρατηρείται ότι οι καμπύλες του σχήματος 5.4 δείχνουν ότι οι συναρτήσεις απώλειας και accuracy του μοντέλου στα σύνολα εκπαίδευσης και επικύρωσης δεν αποκλίνουν, συντελώντας στην συνολική ικανότητα γενίκευσης του μοντέλου.

## 5.3 Αποτελέσματα CNN-GRU

Σε αντίθεση με την ταξινόμηση όπου εφαρμόζουν οι συγγραφείς του [6] σε τρείς κατηγορίες δραστηριοτήτων, με βάση την βάδιση, με βάση την γενική κίνηση του χεριού και με βάση την κίνηση του χεριού για τροφή, η ταξινόμηση που εφαρμόστηκε εδώ αφορά τις 18 αρχικές δραστηριότητες του dataset. Για τη διαδικασία της εξομάλυνσης (regularization), χρησιμοποιήθηκε η βιβλιοθήκη Optuna για την επιλογή των κατάλληλων υπερ-παραμέτρων. Τα αποτελέσματα που προκύπτουν αναφέρονται παρακάτω.

**Πίνακας 5.4: Πληροφορίες μοντέλου CNN-GRU**

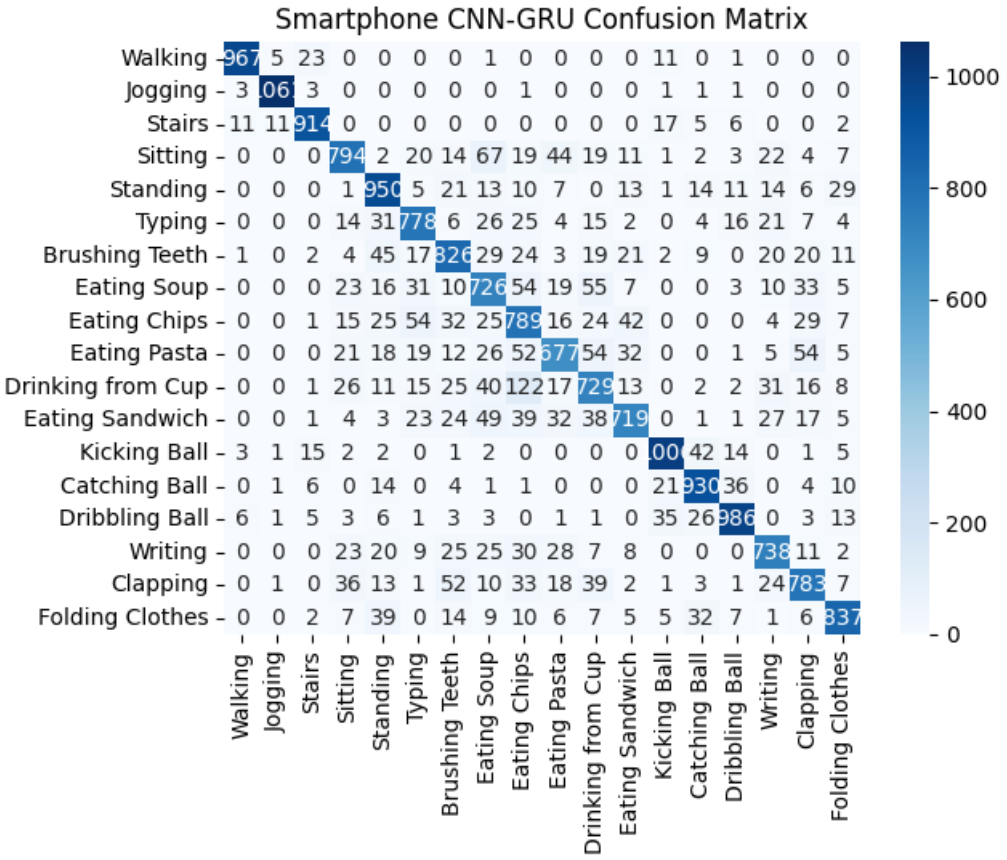
| Model: CNN-GRU | | |
|---|---|---|
| Layer (type) | Output Shape | Parameters # |
| Time distributed 1 | (None, None, 48, 32) | 608 |
| Time distributed 2 | (None, None, 24, 32) | 0 |
| Time distributed 3 | (None, None, 22, 128) | 12416 |
| Time distributed 4 | (None, None, 11, 128) | 0 |
| Time distributed 5 | (None, None, 11, 128) | 0 |
| Time distributed 6 | (None, None, 1408) | 0 |
| GRU 1 | (None, None, 64) | 283088 |
| GRU 2 | (None, 64) | 24960 |
| Dropout | (None, 64) | 0 |
| Dense | (None, 18) | 1170 |
| Total Parameters: 322.162<br>Trainable Parameters: 322.162<br>Non-trainable Parameters: 0 | | |

**Πίνακας 5.5: Αποτελέσματα του Validation συνόλου για το dataset WISDM**

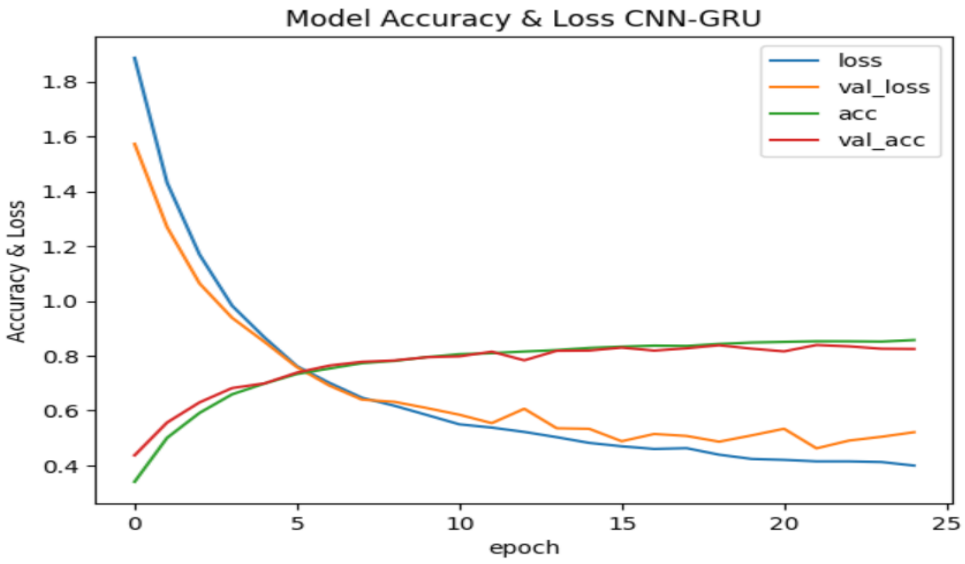
| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Brushing teeth | 0.8074 | 0.7536 | 0.7796 | 1124 |
| Catching ball | 0.9157 | 0.8729 | 0.8938 | 1070 |
| Clapping | 0.7677 | 0.7998 | 0.7834 | 1004 |
| Dribbling ball | 0.9145 | 0.8969 | 0.9056 | 1038 |
| Drinking | 0.6900 | 0.7298 | 0.7093 | 1040 |
| Eating chips | 0.7141 | 0.6042 | 0.6546 | 1104 |
| Eating pasta | 0.6868 | 0.7555 | 0.7195 | 859 |
| Eating sandwich | 0.7255 | 0.8061 | 0.7637 | 892 |
| Eating soup | 0.7390 | 0.7115 | 0.7250 | 1130 |
| Folding clothes | 0.8364 | 0.8824 | 0.8588 | 1020 |
| Jogging | 0.9980 | 0.9851 | 0.9915 | 1010 |
| Kicking ball | 0.9363 | 0.9168 | 0.9265 | 1106 |
| Sitting | 0.7799 | 0.7885 | 0.7842 | 993 |
| Stairs | 0.9497 | 0.9537 | 0.9517 | 951 |
| Standing | 0.8661 | 0.8014 | 0.8325 | 1138 |
| Typing | 0.7867 | 0.8011 | 0.7938 | 930 |
| Walking | 0.9565 | 0.9706 | 0.9635 | 1088 |
| Writing | 0.7503 | 0.8226 | 0.7848 | 902 |
| Accuracy | | | 0.8243 | 18399 |
| Macro avg | 0.8234 | 0.8251 | 0.8234 | 18399 |
| Weighted avg | 0.8258 | 0.8243 | 0.8243 | 18399 |

**Πίνακας 5.6: Αποτελέσματα του Testing συνόλου για το dataset WISDM**

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Brushing teeth | 0.7844 | 0.7727 | 0.7785 | 1069 |
| Catching ball | 0.9047 | 0.8683 | 0.8861 | 1071 |
| Clapping | 0.7646 | 0.7877 | 0.7760 | 994 |
| Dribbling ball | 0.9021 | 0.9054 | 0.9038 | 1089 |
| Drinking | 0.6890 | 0.7239 | 0.7061 | 1007 |
| Eating chips | 0.7422 | 0.6526 | 0.6945 | 1209 |
| Eating pasta | 0.6936 | 0.7764 | 0.7327 | 872 |
| Eating sandwich | 0.7314 | 0.8217 | 0.7740 | 875 |
| Eating soup | 0.7319 | 0.6901 | 0.7104 | 1052 |
| Folding clothes | 0.8480 | 0.8746 | 0.8611 | 957 |
| Jogging | 0.9907 | 0.9815 | 0.9861 | 1081 |
| Kicking ball | 0.9196 | 0.9137 | 0.9166 | 1101 |
| Sitting | 0.7716 | 0.8160 | 0.7932 | 973 |
| Stairs | 0.9462 | 0.9394 | 0.9428 | 973 |
| Standing | 0.8676 | 0.7950 | 0.8297 | 1195 |
| Typing | 0.8164 | 0.7996 | 0.8079 | 973 |
| Walking | 0.9593 | 0.9758 | 0.9675 | 991 |
| Writing | 0.7970 | 0.8048 | 0.8009 | 917 |
| Accuracy | | | 0.8267 | 18399 |
| Macro avg | 0.8256 | 0.8277 | 0.8260 | 18399 |
| Weighted avg | 0.8280 | 0.8267 | 0.8267 | 18399 |




**Σχήμα 5.7: Ο πίνακας confusion matrix για τις 18 δραστηριότητες του dataset WISDM**




**Σχήμα 5.8: Συγκεντρωτικός πίνακας των παραπάνω δεικτών για το dataset WISDM**

Τα αποτελέσματα της μεθόδου για τις αρχικές 18 δραστηριότητες του dataset WISDM αναφέρονται παραπάνω. Τα αποτελέσματα για την ταξινόμηση σε 3 κατηγορίες, όπως προτείνουν και οι συγγραφείς του [6], αναφέρονται παρακάτω.

**Πίνακας 5.7: Πληροφορίες μοντέλου CNN-GRU για τις αρχικές 3 κατηγορίες**

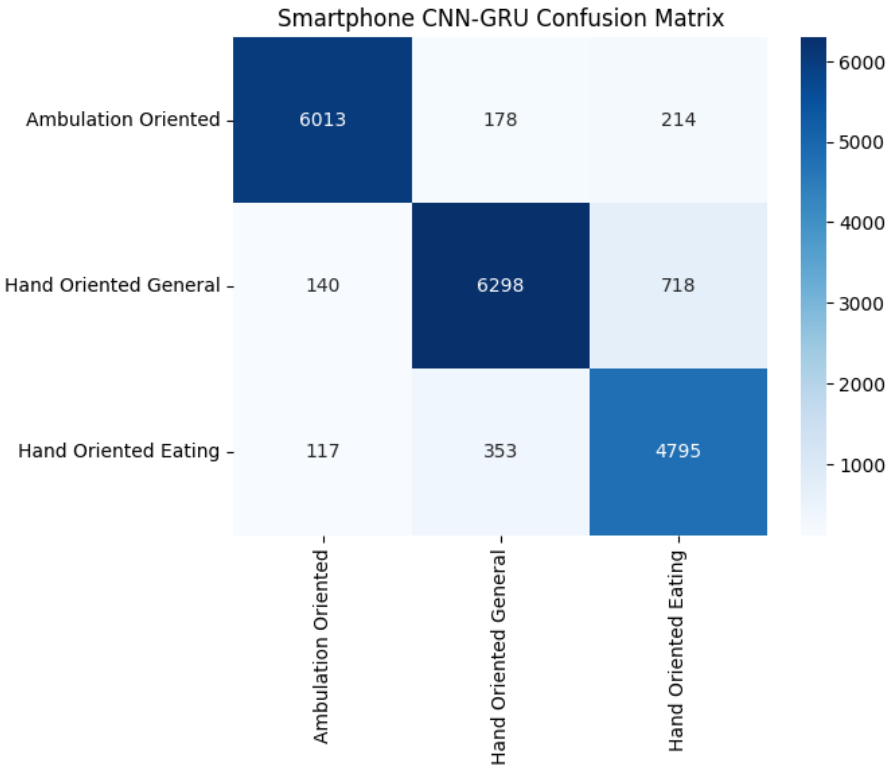
| Model: CNN-GRU | | |
|---|---|---|
| Layer (type) | Output Shape | Parameters # |
| Time distributed 1 | (None, None, 48, 32) | 608 |
| Time distributed 2 | (None, None, 24, 32) | 0 |
| Time distributed 3 | (None, None, 22, 128) | 12416 |
| Time distributed 4 | (None, None, 11, 128) | 0 |
| Time distributed 5 | (None, None, 11, 128) | 0 |
| Time distributed 6 | (None, None, 1408) | 0 |
| GRU 1 | (None, None, 64) | 283088 |
| GRU 2 | (None, 64) | 24960 |
| Dropout | (None, 64) | 0 |
| Dense | (None, 3) | 195 |
| Total Parameters: 321.187<br>Trainable Parameters: 321.187<br>Non-trainable Parameters: 0 | | |

**Πίνακας 5.8: Αποτελέσματα του Validation συνόλου για το dataset WISDM με τρεις κατηγορίες**

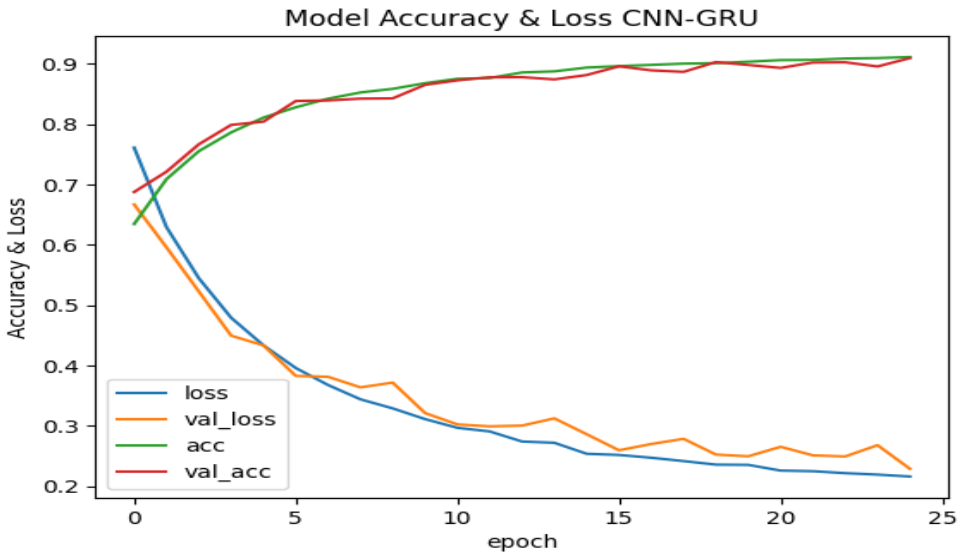
| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Με βάση τη βάδιση | 0.9397 | 0.9537 | 0.9467 | 6247 |
| Κίνηση χεριού για τροφή | 0.9149 | 0.8386 | 0.8751 | 5668 |
| Γενική κίνηση χεριού | 0.8789 | 0.9272 | 0.9024 | 6911 |
| Accuracy | | | 0.9093 | 18826 |
| Macro avg | 0.9112 | 0.9065 | 0.9081 | 18826 |
| Weighted avg | 0.9099 | 0.9093 | 0.9089 | 18826 |

**Πίνακας 5.9: Αποτελέσματα του Testing συνόλου για το dataset WISDM με τρεις κατηγορίες**

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Με βάση τη βάδιση | 0.9388 | 0.9590 | 0.9488 | 6270 |
| Κίνηση χεριού για τροφή | 0.9107 | 0.8373 | 0.8725 | 5727 |
| Γενική κίνηση χεριού | 0.8801 | 0.9222 | 0.9007 | 6829 |
| Accuracy | | | 0.9086 | 18826 |
| Macro avg | 0.9099 | 0.9062 | 0.9073 | 18826 |
| Weighted avg | 0.9090 | 0.9086 | 0.9081 | 18826 |




**Σχήμα 5.9: Ο πίνακας confusion matrix για τις 3 κατηγορίες του dataset WISDM**




**Σχήμα 5.10: Συγκεντρωτικός πίνακας των παραπάνω δεικτών για τις 3 κατηγορίες του dataset WISDM**

Η υβριδική μέθοδος CNN-GRU η οποία εφαρμόζεται για το dataset WISDM παρουσιάζει πολύ καλά αποτελέσματα στην αναγνώριση και ταξινόμηση ανθρώπινων δραστηριοτήτων με δεδομένα χρονοσειρών από αισθητήρες κίνησης. Συγκεκριμένα, η ικανότητα εκμάθησης χωρικών ή τοπικών χαρακτηριστικών του μονοδιάστατου νευρωνικού δικτύου συνέλιξης, σε συνδυασμό με την ικανότητα εκμάθησης χρονικών χαρακτηριστικών του νευρωνικού δικτύου ανατροφοδότησης, καθιστούν το μοντέλο παραπάνω από ικανό στην ανίχνευση χωροχρονικών χαρακτηριστικών. Τα σχήματα 5.5, 5.7 δείχνουν ότι η μέθοδος διαθέτει ισχυρή ικανότητα στην ταξινόμηση δραστηριοτήτων και οι καμπύλες των συναρτήσεων απώλειας και accuracy των σχημάτων 5.6, 5.8, ότι έχει ισχυρή ικανότητα γενίκευσης. Η μέθοδος εφαρμόζεται σε ισορροπημένα δεδομένα, και παρουσιάζει πολύ καλά αποτελέσματα στην ταξινόμηση δραστηριοτήτων και ταυτόχρονα γρήγορο χρόνο εκτελέσης της διαδικασίας της εκπαίδευσης.

## 5.4 Σύγκριση μεθόδων

Οι μέθοδοι Transformer και το υβριδικό νευρωνικό δίκτυο συνέλιξης και ανατροφοδότησης υλοποιήθηκαν για την ταξινόμηση και αναγνώριση ανθρώπινων δραστηριοτήτων με δεδομένα από αισθητήρες κίνησης. Τα datasets τα οποία χρησιμοποιήθηκαν για τις συγκρίσεις, περιέχουν και τα δυο 18 δραστηριότητες στο συνολό τους.

Η μέθοδος Transformer όπως παρουσιάστηκε στο [3] εφαρμόστηκε και για τα δυο datasets, παρουσιάζοντας $0.992$ του δείκτη accuracy για το KU-HAR και $0.911$ για το WISDM, παρουσιάζοντας σχεδόν τέλεια αποτελέσματα τόσο σε ισορροπημένα όσο και σε μη-ισορροπημένα δεδομένα. Από τους πίνακες 5.3 και 5.6 η μέθοδος Transformer παρουσιάζει αύξηση $9$% του δείκτη accuracy συγκριτικά με τη μέθοδο CNN-GRU για το dataset WISDM. Τα αποτελέσματα που προκύπτουν δείχνουν ότι ο μηχανισμός προσοχής, που είναι η θεμελιώδης λειτουργία της μεθόδου, διαθέτει ισχυρή ικανότητα εκμάθησης των χαρακτηριστικών των datasets, το οποίο προκύπτει και από τον αριθμό παραμέτρων του μοντέλου Transformer συγκριτικά με αυτών του μοντέλου CNN-GRU (πίνακες 5.1 και 5.4). Επίσης, στο dataset KU-HAR εξετάζεται και η ικανότητα του μηχανισμού προσοχής να ανιχνεύει την εναλλαγή δραστηριοτήτων μέσα σε μια χρονοσειρά παρουσιάζοντας άριστα αποτελέσματα. Το στάδιο της προεπεξεργασίας του dataset WISDM εφαρμόστηκε ακριβώς όπως στο [6], επομένως δεν δημιουργήθηκαν και ζευγάρια δραστηριοτήτων για να τροφοδοτηθούν στο μοντέλο Transformer όπως συνέβει στο στάδιο προ-επεξεργασίας του dataset KU-HAR, αφενός για να ελεγχθεί η ικανότητα του μοντέλου και σε αυτή την περίπτωση, και αφετέρου γιατί η φύση των περισσότερων δραστηριοτήτων του dataset WISDM δεν ενδείκνυται για τέτοιο είδους συνδυασμό.

Η μέθοδος CNN-GRU εφαρμόστηκε όπως παρουσιάστηκε στο [6] με τη μοναδική αλλαγή της κατηγοριοποίησης των δραστηριοτήτων στις αρχικές 18 του dataset WISDM και όχι στις 3 που προτείνουν οι συγγραφείς του. Ο λόγος που τροποποιήθηκε είναι για να γίνει καλύτερη σύγκριση ανάμεσα στα μοντέλα Transformer και CNN-GRU. Αν και αυτό το βήμα εφαρμόζεται μετά την λειτουργία του μοντέλου, βάσει αποτελεσμάτων, επηρεάζει τη συνολική του απόδοση κατά $8$% του δείκτη accuracy από τους πίνακες 5.6 και 5.9. Επί-

σης, η μέθοδος CNN-GRU εφαρμόζεται σε ισορροπημένα δεδομένα, μια κατάσταση που δεν εμφανίζεται συχνά σε δεδομένα πραγματικών συνθηκών, περιορίζοντας σε ένα βαθμό το πεδίο εφαρμογής της σε καθολικής φύσης προβλήματα αναγνώρισης ανθρώπινης δραστηριότητας.

Παρατηρείται ότι, η μέθοδος CNN-GRU δεν αποδίδει το ίδιο ικανοποιητικά όσο η μέθοδος Transformer. Από τον αριθμό των παραμέτρων των μοντέλων προκύπτει ότι η μέθοδος Transformer είναι περισσότερο πολύπλοκη και έχει μεγαλύτερο χρόνο εκτέλεσης για τη διαδικασία της εκπαίδευσης, γεγονός που καταδεικνύει ότι είναι αποτελεσματική με μεγάλο όγκο δεδομένων, συμπέρασμα που σημειώνεται και στο [14]. Επίσης, μια σημαντική διαφορά είναι ότι οι μέθοδοι χρησιμοποιούν διαφορετικές συναρτήσεις απώλειας για την τελική ταξινόμηση των δραστηριοτήτων, με αποτέλεσμα την διαφοροποίηση των αποτελεσμάτων τους. Τέλος, και για τις δυο μεθόδους χρησιμοποιήθηκαν έτοιμες διαφορετικές βιβλιοθήκες για την επιλογή των υπερ-παραμέτρων τους, οι οποίες και χρησιμοποιήθηκαν χωρίς αλλαγή στις υλοποιήσεις των μοντέλων.

# 6. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ

Η αναγνώριση ανθρώπινης δραστηριότητας με την χρήση αισθητήρων κίνησης έχει προσφέρει πλεονεκτήματα και διευκολύνσεις στην ποιότητα της καθημερινής ζωής. Σημαντικό ρόλο σε αυτό έχουν η τεχνητή νοημοσύνη και η μηχανική μάθηση όπου με την ραγδαία εξέλιξη τους έχουν θέσει ψηλά τον πήχη, επιτρέποντας η αναγνώριση των ανθρώπινων δραστηριοτήτων να πραγματοποιείται με τρόπο αυτοματοποιημένο και βατό. Τα τελευταία χρόνια, με την χρήση φορητών συσκευών και εφαρμογών, η βαθιά μάθηση έχει καταφέρει υψηλή απόδοση στην αναγνώριση δραστηριοτήτων και αυτό οφείλεται τόσο σε εκτεταμένη έρευνα που έχει λάβει μέρος, όσο και στην τεχνολογική εξέλιξη των αισθητήρων κίνησης που περιλαμβάνονται στις συσκευές αυτές. Από την εφαρμογή εξ' ολοκλήρου νευρωνικών δικτύων συνέλιξης έως το συνδυασμό τους με δίκτυα ανατροφοδότησης, η αναγνώριση ανθρώπινων δραστηριοτήτων και η ταξινόμηση τους παρουσιάζει ολοένα και περισσότερο καλύτερα αποτελέσματα, κυρίως με την χρήση έτοιμων βιβλιοθηκών οι οποίες ρυθμίζουν αυτόματα τις παραμέτρους που χρειάζονται.

Το μοντέλο Transformer που εφαρμόζεται εδώ, χρησιμοποιεί όλα τα παραπάνω με κύριο στόχο να αποδείξει την καταλληλότητα του ως μια βιώσιμη εναλλακτική σε διάφορες μορφές νευρωνικών δικτύων συνέλιξης και ανατροφοδότησης. Τα αποτελέσματα δείχνουν ότι αποδίδει καλύτερα, καθώς παρουσιάζει καλύτερη αντιπροσωπευτική δύναμη από αρχιτεκτονικές βραχυπρόθεσμης μνήμης, έχει την ικανότητα να εντοπίζει παραμέτρους σε μεγάλη κλίμακα ενώ ταυτόχρονα μπορεί να εφαρμοστεί σε φορητές συσκευές. Η εισαγωγή δεδομένων από σήματα χρονοσειρών κατευθείαν στο μοντέλο, μόνο μετά από κανονικοποίηση, δεν απαίτει την ανάγκη προ-μετασχηματισμού τους και η ικανότητα παραλληλισμού του το καθιστούν εξαιρετική επιλογή για την χρήση του σε μονάδες GPU. Σύμφωνα με τα αποτελέσματα που παρουσιάστηκαν, καταφέρνει επιτυχώς την αναγνώριση και ταξινόμηση τόσο μιας δραστηριότητας μέσα σε μια χρονοσειρά, όσο και το συνδυασμό με άλλες και σε ισορροπημένα αλλα και μη ισορροπημένα δεδομένα. Τα αποτελέσματα της υλοποίησης δείχνουν ότι οι μηχανισμοί προσοχής εντοπίζουν συσχετίσεις σε χρονοσειρές μεγάλου μήκους δίνοντας βαρύτητα στις σημαντικές από αυτές, σε αντίθεση με τα δίκτυα ανατροφοδότησης, όπου παρουσιάζουν σχετική αδυναμία. Παρ΄όλα αυτά, το μοντέλο Transformer δεν διαθέτει κάποιες από τις επαγωγικές τάσεις που είναι εγγενείς στα δίκτυα συνέλιξης, όπως η τοπικότητα, και επομένως δεν γενικεύει καλά όταν εκπαιδεύεται σε ανεπαρκή όγκο δεδομένων.

Τα δίκτυα προσοχής σε πολλούς τομείς της βαθιάς μάθησης κατέχουν ήδη σημαντική θέση. Ο μηχανισμός λειτουργίας τους επιτρέπει τον αποτελεσματικό εντοπισμό συσχετίσεων στα χαρακτηριστικά των δεδομένων και με την αύξηση της υπολογιστικής δύναμης, ο χρόνος εκτέλεσης τους είναι πλεόν πρακτικά διαχειρίσιμος. Μελλοντικά, το μοντέλο Transformer μπορεί να εφαρμοστεί και σε μεγαλύτερα datasets, με περισσότερο αριθμό δραστηριοτήτων και ενδεχομένως με δεδομένα από πολλαπλούς αισθητήρες, εξυπηρετώντας ακόμη περισσότερο τις καθημερινές δραστηριότητες των ανθρώπων.

# ΒΙΒΛΙΟΓΡΑΦΙΑ


[1] Hangbo Bao, Li Dong, Songhao Piao, and Furu Wei. Beit: Bert pre-training of image transformers, 2021.

[2] Davide Buffelli and Fabio Vandin. Attention-based deep learning framework for human activity recognition with user adaptation. *IEEE Sensors Journal*, 21(12):13474–13483, jun 2021.

[3] Iveta Dirgová Luptáková, Martin Kubovčík, and Jiří Pospíchal. Wearable sensor-based human activity recognition with transformer model. *Sensors*, 22(5), 2022.

[4] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2020.

[5] Hristijan Gjoreski, Jani Bizjak, Martin Gjoreski, and Matja Gams. Comparing deep and classical machine learning methods for human activity recognition using wrist accelerometer. 2016.

[6] Saurabh Gupta. Deep learning based human activity recognition (har) using wearable sensor data. *International Journal of Information Management Data Insights*, 1(2):100046, 2021.

[7] Sojeong Ha and Seungjin Choi. Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 381–388, 2016.

[8] Nils Y. Hammerla, Shane Halloran, and Thomas Ploetz. Deep, convolutional, and recurrent models for human activity recognition using wearables, 2016.

[9] Andrey Ignatov. Real-time human activity recognition from accelerometer data using convolutional neural networks. *Applied Soft Computing*, 62:915–922, 2018.

[10] Wenchao Jiang and Zhaozheng Yin. Human activity recognition using wearable sensors by deep convolutional neural networks. In *Proceedings of the 23rd ACM International Conference on Multimedia*, MM '15, page 1307–1310, New York, NY, USA, 2015. Association for Computing Machinery.

[11] Saif Mahmud, M Tanjid Hasan Tonmoy, Kishor Kumar Bhaumik, A K M Mahbubur Rahman, M Ashraful Amin, Mohammad Shoyaib, Muhammad Asif Hossain Khan, and Amin Ahsan Ali. Human activity recognition from wearable sensor data using self-attention, 2020.

[12] Abdullah Nahid, Niloy Sikder, and Ibrahim Rafi. Ku-har: An open dataset for human activity recognition (dataset). 07 2020.

[13] Yoli Shavit and Itzik Klein. Boosting inertial-based human activity recognition with transformers. *IEEE Access*, 9:53540–53547, 2021.

[14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[15] Huaijun Wang, Jing Zhao, Junhuai li, Ling Tian, Ting Cao, Yang An, Kan Wang, and Shancang Li. Wearable sensor-based human activity recognition using hybrid deep learning techniques. *Security and Communication Networks*, 2020:1–12, 07 2020.

[16] Gary M. Weiss, Kenichi Yoneda, and Thaier Hayajneh. Smartphone and smartwatch-based biometrics using activities of daily living. *IEEE Access*, 7:133190–133202, 2019.

[17] Li Xue, Si Xiandong, Nie Lanshun, Li Jiazhen, Ding Renjie, Zhan Dechen, and Chu Dianhui. Understanding and improving deep neural network for activity recognition, 2018.

[18] Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. Deepsense: A unified deep learning framework for time-series mobile sensing data processing, 2016.