



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

**Ανάπτυξη ερμηνεύσιμου μοντέλου εκτίμησης του κινδύνου  
εμφάνισης υπεργλυκαιμικών και υπογλυκαιμικών επεισοδίων με  
χρήση τεχνικών μηχανικής μάθησης**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

της

**Νεφέλης Μυροπούλου**

**Επιβλέπουσα:** Κωνσταντίνα Νικήτα  
Καθηγήτρια Ε.Μ.Π.

Αθήνα, Μάρτιος 2023





**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ  
ΥΛΙΚΩΝ

**Ανάπτυξη ερμηνεύσιμου μοντέλου εκτίμησης του κινδύνου  
εμφάνισης υπεργλυκαιμικών και υπογλυκαιμικών επεισοδίων με  
χρήση τεχνικών μηχανικής μάθησης**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

**Νεφέλης Μυροπούλου**

**Επιβλέπουσα:** Κωνσταντίνα Νικήτα  
Καθηγήτρια Ε.Μ.Π.

Εγκρίθηκε από την τριμελή επιτροπή την 27η Μαρτίου 2023.

.....  
Κωνσταντίνα Νικήτα  
Καθηγήτρια Ε.Μ.Π.

.....  
Αθανάσιος Βουλόδημος  
Επίκουρος Καθηγητής Ε.Μ.Π.

.....  
Γεώργιος Στάμου  
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2023

.....

**Νεφέλη Μυροπούλου**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Μυροπούλου Νεφέλη, 2023.

Με επιφύλαξη κάθε δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

# Περίληψη

Ο Σακχαρώδης Διαβήτης (ΣΔ) αποτελεί χρόνια ασθένεια με παγκόσμια εξάπλωση και σοβαρές επιπλοκές για τους πάσχοντες. Σύμφωνα με την Παγκόσμια Ομοσπονδία Διαβήτη, ο συνολικός αριθμός νοσούντων με ΣΔ αναμένεται να ανέλθει ως και τα 643 εκατομμύρια έως το 2030. Με στόχο την επίτευξη ορθού γλυκαιμικού ελέγχου για την πρόληψη ή και την καθυστέρηση της εμφάνισης των μακροπρόθεσμων επιπλοκών της ασθένειας, στην παρούσα εργασία αναπτύχθηκε ερμηνεύσιμο μοντέλο εκτίμησης του κινδύνου εμφάνισης υπεργλυκαιμικών και υπογλυκαιμικών επεισοδίων σε άτομα με Σακχαρώδη Διαβήτη Τύπου 1 βασισμένο στην εφαρμογή υβριδικής προσέγγισης που συνδυάζει τη χρήση διαμερισματικών μοντέλων και μεθόδων μηχανικής μάθησης. Για την ανάπτυξη και την αξιολόγηση του μοντέλου χρησιμοποιήθηκαν δεδομένα πραγματικών ασθενών που συλλέχθηκαν στο πλαίσιο του ερευνητικού προγράμματος ENDORSE (T1EDK-03695) και περιλάμβαναν μετρήσεις γλυκοζης, τιμές χορηγημένης ινσουλίνης, καθώς και πληροφορίες λαμβανόμενων γευμάτων και σωματικής δραστηριότητας. Με στόχο την ακριβέστερη αποτύπωση των μηχανισμών απορρόφησης της ινσουλίνης από τον υποδόριο χώρο στην αιματική κυκλοφορία, της ενεργοποίησης των ενδοκυτταρικών μονοπατιών μεταβίβασης σήματος ινσουλίνης και της απορρόφησης της γλυκοζης γεύματος από το έντερο στην αιματική κυκλοφορία, αξιοποιήθηκαν κατάλληλα διαμερισματικά μοντελα που συνδυάστηκαν με ένα αναδρομικό νευρωνικό δίκτυο μακράς βραχείας μνήμης (Long Short Term Memory Neural Networks - LSTMs). Για την αποτελεσματική διαχείριση της μη ισορροπημένης φύσης των δεδομένων διερευνήθηκαν ποικίλες τεχνικές εξισορρόπησης των δεδομένων βασισμένες στην υπερδειγματοληψία. Για την εξαγωγή επεξηγήσεων επί των αποφάσεων του μοντέλου εφαρμόστηκε η μέθοδος LIME (Local Interpretable Model-agnostic Explanations). Το μοντέλο αξιολογήθηκε ως προς τη διακριτική του ικανότητα, καθώς και ως προς την ικανότητά του να εκτιμά ακριβείς πιθανότητες διακινδύνευσης στις περιπτώσεις υπεργλυκαιμικών και υπογλυκαιμικών επεισοδίων, ενώ η εφαρμογή της LIME μεθόδου προσέφερε πολύτιμες πληροφορίες για τον τρόπο λειτουργίας του μοντέλου και φανέρωσε πιθανούς τρόπους περαιτέρω βελτίωσής του.

**Λέξεις Κλειδιά:** σακχαρώδης διαβήτης, μηχανική μάθηση, βαθιά μάθηση, μοντέλα εκτίμησης κινδύνου, υβριδικά μοντέλα, ερμηνευσιμότητα, ερμηνεύσιμη τεχνητή νοημοσύνη, προσομοίωση φυσιολογικών συστημάτων, διαμερισματικά μοντέλα

# Abstract

Diabetes Mellitus (DM) is a chronic disease characterized by increasing prevalence and serious complications. According to the International Diabetes Federation, the total number of DM patients is expected to reach 643 million by 2030. To facilitate optimal glycemic control and prevent or delay the onset of the long-term complications of the disease, the present thesis aims at the design, development, and evaluation of an interpretable model for assessing the risk of hyperglycemic and hypoglycemic episodes in people with Type 1 Diabetes Mellitus. The model leverages a hybrid approach based on the combined use of compartmental models and machine learning methods. Real-life patient data collected within the framework of the ENDORSE research program (T1EDK-03695) were used for development and evaluation purposes. The considered data included glucose measurements, previously administered insulin values, as well as meal and physical activity information. In order to capture the physiological mechanisms of insulin absorption from the subcutaneous space into the blood circulation, the activation of the insulin signaling pathway and the absorption of meal glucose from the intestine into the blood circulation, appropriate compartmental models were combined with Long Short Term Memory Neural Networks (LSTMs). Various techniques based on oversampling were explored towards addressing the unbalanced nature of the used dataset. The LIME (Local Interpretable Model-agnostic Explanations) method was deployed to derive explanations on the model's decisions. The model was assessed in terms of its discrimination ability as well as its ability to produce accurate risk scores in the case of hyperglycemic and hypoglycemic episodes, whereas the application of the LIME method provided valuable insights on the most influential features and revealed potential ways for improving the model's performance.

**Keywords:** diabetes mellitus, machine learning, deep learning, risk prediction models, hybrid models, explainability, explainable artificial intelligence, physiological models, compartmental models

# Ευχαριστίες

Σε αυτό το σημείο θα ήθελα αρχικά να ευχαριστήσω την καθηγήτρια κα. Κωνσταντίνα Νικήτα, για την επίβλεψη και την καθοδήγησή της, καθώς και για την μοναδική ευκαιρία που μου προσέφερε αναθέτοντας μου ένα θέμα που χαρακτηρίζεται από έντονο ερευνητικό ενδιαφέρον και άμεσες κοινωνικές προεκτάσεις. Παράλληλα, θα ήθελα να ευχαριστήσω θερμά την μεταδιδάκτορα κα. Κωνσταντία Ζαρκογιάννη και την υποψήφια διδάκτορα κα. Μαρία Αθανασίου, που μέσω του πολύτιμου χρόνου που αφιέρωσαν και των εύστοχων παρατηρήσεων τους, με βοήθησαν να φέρω εις πέρας την παρούσα εργασία και να εμπλουτίσω τις γνώσεις μου.

Ευχαριστώ ακόμα τους φίλους μου, παλιούς και νέους, για την αγάπη και υποστήριξη τους. Ειδικά προς τους φίλους που υπήρξαν και συμφοιτητές, σας ευχαριστώ για την πάντοτε κρατημένη θέση στο αμφιθέατρο και την ενθάρρυνση και το γέλιο που μου χαρίσατε απλόχερα, καθώς σπουδάσαμε μαζί.

Τέλος, ένα εγκάρδιο ευχαριστώ οφείλεται στους γονείς και στις δύο αδερφές μου, για την ανιδιοτελή αγάπη και αμέριστη κατανόησή τους από την αρχή. Δίχως τη στήριξη τους, δεν θα ήμουν σε θέση να επιδιώξω τα προσωπικά μου όνειρα και επιθυμίες και για αυτό τους οφείλω πολλά.

# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b>	<b>1</b>
1.1	Το ερευνητικό πρόβλημα . . . . .	1
1.2	Κίνητρο της εργασίας . . . . .	1
1.3	Δομή της εργασίας . . . . .	2
<b>2</b>	<b>Σακχαρώδης Διαβήτης</b>	<b>4</b>
2.1	Γενικά στοιχεία και ταξινόμηση . . . . .	4
2.1.1	Σακχαρώδης Διαβήτης Τύπου 1 . . . . .	4
2.1.2	Σακχαρώδης Διαβήτης Τύπου 2 . . . . .	5
2.1.3	Σακχαρώδης Διαβήτης Κυήσεως . . . . .	5
2.2	Μεταβολισμός και ομοιόσταση της γλυκόζης . . . . .	5
2.2.1	Μεταβολισμός γλυκόζης . . . . .	6
2.2.2	Ομοιόσταση γλυκόζης . . . . .	6
2.3	Συμπτώματα και επιπλοκές . . . . .	10
2.3.1	Συμπτώματα . . . . .	10
2.3.2	Οξείες επιπλοκές . . . . .	10
2.3.3	Μακροπρόθεσμες επιπλοκές . . . . .	11
2.4	Διάγνωση και αντιμετώπιση . . . . .	13
2.4.1	Διάγνωση . . . . .	13
2.4.2	Αντιμετώπιση . . . . .	14
<b>3</b>	<b>Τεχνητά Νευρωνικά Δίκτυα</b>	<b>19</b>
3.1	Παραλληλισμός με βιολογικά συστήματα . . . . .	19
3.2	Νευρώνας (Perceptron) . . . . .	20
3.2.1	Μοντέλο λειτουργίας . . . . .	20
3.2.2	Συναρτήσεις ενεργοποίησης . . . . .	22
3.3	Perceptron πολλών επιπέδων (Multilayer Perceptron) . . . . .	24
3.4	Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks) . . . . .	26
3.4.1	Μοντέλο λειτουργίας και εκπαίδευση αναδρομικών νευρωνικών δικτύων . . . . .	26
3.4.2	Long Short-Term Memory (LSTM) και Gated Recurrent Unit (GRU) δίκτυα . . . . .	29
3.5	Εκπαίδευση και αξιολόγηση τεχνητών νευρωνικών δικτύων . . . . .	31
3.5.1	Ταξινόμηση ειδών μάθησης . . . . .	31
3.5.2	Τεχνικές εκπαίδευσης στην επιβλεπόμενη μάθηση . . . . .	33
3.5.3	Προεπεξεργασία δεδομένων . . . . .	34
3.5.4	Μετρικές αξιολόγησης μοντέλων μηχανικής μάθησης . . . . .	36



<b>4</b>	<b>Ερμηνευσιμότητα</b>	<b>38</b>
4.1	Η αξία της ερμηνευσιμότητας στον κλάδο της τεχνητής νοημοσύνης . . .	38
4.2	Ταξινόμηση μεθόδων ερμηνευσιμότητας . . . . .	40
4.2.1	Διαφανή και αδιαφανή μοντέλα μηχανικής μάθησης . . . . .	40
4.2.2	Model-specific και model-agnostic μέθοδοι . . . . .	40
4.2.3	Τοπικές (local) και καθολικές (global) μέθοδοι . . . . .	41
4.3	Post-hoc μέθοδοι ερμηνευσιμότητας: LIME και SHAP . . . . .	41
4.3.1	Local Interpretable Model-agnostic Explanations (LIME) . . . .	41
4.3.2	SHapley Additive exPlanation (SHAP) Values . . . . .	43
<b>5</b>	<b>Βιβλιογραφική Επισκόπηση</b>	<b>45</b>
5.1	Μοντέλα προσομοίωσης φυσιολογικών συστημάτων . . . . .	45
5.2	Μοντέλα πρόβλεψης επιπέδων γλυκόζης . . . . .	46
5.3	Μοντέλα πρόβλεψης υπογλυκαιμικών και υπεργλυκαιμικών επεισοδίων .	47
5.4	Τεχνικές ερμηνευσιμότητας σε μοντέλα πρόβλεψης . . . . .	48
<b>6</b>	<b>Ανάπτυξη ερμηνεύσιμου μοντέλου εκτίμησης του κινδύνου εμφάνισης υπεργλυκαιμικών και υπογλυκαιμικών επεισοδίων</b>	<b>50</b>
6.1	Δεδομένα και επεξεργασία . . . . .	50
6.1.1	Τα αρχικά δεδομένα . . . . .	50
6.1.2	Σχηματισμός και labeling χρονοσειρών γλυκόζης . . . . .	51
6.2	Δομή υβριδικού μοντέλου . . . . .	52
6.2.1	Μοντέλο απορρόφησης γλυκόζης γεύματος από το έντερο στο αίμα	53
6.2.2	Μοντέλο απορρόφησης ινσουλίνης από τον υποδόριο χώρο στο αίμα	56
6.2.3	Μοντέλο ενεργοποίησης του ενδοκυτταρικού μονοπατιού μεταβίβασης σήματος ινσουλίνης . . . . .	57
6.2.4	Χρήση δεδομένων σωματικής δραστηριότητας . . . . .	59
6.2.5	Μοντέλο μηχανικής μάθησης . . . . .	60
6.3	Εκπαίδευση, αξιολόγηση και ερμηνεία του μοντέλου μηχανικής μάθησης	61
6.3.1	Εκπαίδευση μοντέλου μηχανικής μάθησης . . . . .	61
6.3.2	Αξιολόγηση μοντέλου μηχανικής μάθησης . . . . .	65
6.3.3	Ερμηνευσιμότητα μοντέλου μηχανικής μάθησης . . . . .	65
<b>7</b>	<b>Αποτελέσματα και συζήτηση</b>	<b>67</b>
7.1	Πρόβλεψη υπεργλυκαιμικών επεισοδίων . . . . .	67
7.1.1	Ορίζοντας πρόβλεψης 0-30 λεπτά . . . . .	67
7.1.2	Ορίζοντας πρόβλεψης 30-60 λεπτά . . . . .	70
7.2	Πρόβλεψη υπογλυκαιμικών επεισοδίων . . . . .	72
7.2.1	Ορίζοντας πρόβλεψης 0-30 λεπτά . . . . .	72
7.2.2	Ορίζοντας πρόβλεψης 30-60 λεπτά . . . . .	75
7.3	Ερμηνευσιμότητα μοντέλου . . . . .	77
7.3.1	Πρόβλεψη υπεργλυκαιμικών επεισοδίων . . . . .	77
7.3.2	Πρόβλεψη υπογλυκαιμικών επεισοδίων . . . . .	81
7.3.3	Υπολογισμός μέσων βαρών . . . . .	82
<b>8</b>	<b>Συμπεράσματα και μελλοντικές προεκτάσεις</b>	<b>85</b>

# Κατάλογος Σχημάτων

2.1	Ο φυσιολογικός ρόλος της ινσουλίνης σε διαδικασίες του παγκρέατος, του λιπώδους ιστού και των σκελετικών μυών [5] . . . . .	7
2.2	Ομοίωση της γλυκόζης: οι ρόλοι της ινσουλίνης, γλυκαγόνης, αμυλίνης και GLIP-1 [6] . . . . .	9
2.3	Συστήματα αντλιών έγχυσης ινσουλίνης [17] . . . . .	16
3.1	Μοντέλο ενός νευρώνα $k$ [20] . . . . .	20
3.2	Μοντέλο νευρώνα όπου $w_{k0} = b_k$ [20] . . . . .	21
3.3	Διαχωριστική ευθεία για την ταξινόμηση δισδιάστατου προβλήματος σε δύο κλάσεις $C_1$ και $C_2$ . [20] . . . . .	22
3.4	Πλήρως συνδεδεμένο δίκτυο πρόσθιας τροφοδότησης με ένα ενδιάμεσο επίπεδο [20] . . . . .	24
3.5	Αφαιρετική απεικόνιση ενός αναδρομικού νευρωνικού δικτύου [27] . . .	26
3.6	Ανεπτυγμένη μορφή ενός αναδρομικού νευρωνικού δικτύου [27] . . . .	27
3.7	Αφαιρετική απεικόνιση των απλών, επαναλαμβανόμενων δομών που συνθέτουν ένα ανεπτυγμένο, κανονικό RNN δίκτυο [33] . . . . .	29
3.8	Αφαιρετική απεικόνιση των επαναλαμβανόμενων δομών που συνθέτουν ένα ανεπτυγμένο LSTM δίκτυο. [33] . . . . .	30
3.9	Μέσω των μεταβλητών $c_i$ είναι δυνατή η ροή πληροφορίας καθόλη την αλυσίδα, όπως φαίνεται και από τη (σχεδόν) αδιάκοπτη γραμμή που διατρέχει το πάνω μέρος του διαγράμματος [33] . . . . .	30
3.10	Εννοιολογική απεικόνιση της επιβλεπόμενης μάθησης [20] . . . . .	32
3.11	Η $k$ -fold cross-validation μέθοδος για $k$ ίσο με 5 [44] . . . . .	34
4.1	Στόχοι της Explainable AI [54] . . . . .	39
4.2	AI σε σύγκριση με XAI [54] . . . . .	39
4.3	Ενδεικτική χρήση της model-specific Grad-CAM τεχνικής για την επεξήγηση CNN ταξινομητή δύο κλάσεων (γάτες και σκύλοι) [57] . . . . .	41
4.4	Η πολύπλοκη, άγνωστη συνάρτηση απόφασης $f$ (που αναπαρίσταται από τα μπλε/ροζ χρώματα) δεν μπορεί να περιγραφεί από κάποιο γραμμικό μοντέλο. Υπολογίζονται τα $f(z)$ για τα $z$ κοντά στην περιοχή του σημείου ενδιαφέροντος (έντονος κόκκινος σταυρός) και εκτιμάται το γραμμικό μοντέλο (διακεκομμένη γραμμή) που εξηγεί τη συμπεριφορά του $f$ τοπικά [59] . . . . .	43
6.1	Διαγραμματική απεικόνιση του αναπτυσσόμενου (υβριδικού) μοντέλου πρόβλεψης υπογλυκαιμικών και υπεργλυκαιμικών επεισοδίων . . . . .	53

6.2 Υπολογισμός των μεγεθών $k_{empt}, Q_{sto}, Q_{sto1}, Q_{sto2}, Q_{gut}, R_a$ για διάστημα 1 ώρας, πρόσληψη υδατανθράκων 5g και σωματικό βάρος 70kg, με χρήση του Intestinal Glucose Absorption CM . . . . .	54
6.3 Υπολογισμός των μεγεθών $k_{empt}, Q_{sto}, Q_{sto1}, Q_{sto2}, Q_{gut}, R_a$ για τρεις διαφορετικές προσλήψεις υδατανθράκων 3g, 4g και 5g εντός ενός διαστήματος 12 ωρών. Ο υπολογισμός συντελέστηκε χρησιμοποιώντας το Intestinal Glucose Absorption CM και πραγματοποιώντας υπέρθεση των αντίστοιχων καμπυλών εξόδου . . . . .	55
6.4 Υπολογισμός των μεγεθών $I_{sc1}, I_{sc2}, R_i$ για τρεις διαφορετικές χορηγήσεις ινσουλίνης 1.5, 4 και 2.5 μονάδων εντός ενός διαστήματος 12 ωρών και σωματικό βάρος 50kg, με χρήση του Insulin Kinetics CM. . . . .	57
6.5 Υπολογισμός των αναγραφόμενων μεγεθών με χρήση του μαθηματικού μοντέλου προσομοίωσης Sedaghat, έπειτα από την εφαρμογή ινσουλίνης σταθερής συγκέντρωσης $10^{-5}M$ , από $t = 0 \text{ min}$ έως $t = 15 \text{ min}$ . . . . .	58
6.6 Υπολογισμός των αναγραφόμενων μεγεθών με χρήση του μαθηματικού μοντέλου προσομοίωσης Sedaghat, έπειτα από την εφαρμογή ινσουλίνης σταθερής συγκέντρωσης $10^{-5}M$ , από $t = 0 \text{ min}$ έως $t = 15 \text{ min}$ . . . . .	59
6.7 Η ανεπτυγμένη δομή ενός Bidirectional LSTM (BiLSTM) για χρονοσειρά τριών βημάτων [107] . . . . .	60
6.8 Παράδειγμα εφαρμογής της nested cross-validation τεχνικής, όπου για το εξωτερικό cross-validation χρησιμοποιούνται 5 folds και για το εσωτερικό, 2 folds. [111] . . . . .	64
7.1 Αποτελέσματα LIME τεχνικής - Υπεργλυκαιμία, 0-30min . . . . .	77
7.2 Αποτελέσματα LIME τεχνικής - Υπεργλυκαιμία, 0-30min . . . . .	78
7.3 Αποτελέσματα LIME τεχνικής - Υπεργλυκαιμία, 0-30min . . . . .	79
7.4 Αποτελέσματα LIME τεχνικής - Υπεργλυκαιμία, 30-60min . . . . .	79
7.5 Αποτελέσματα LIME τεχνικής - Υπεργλυκαιμία, 30-60min . . . . .	80
7.6 Αποτελέσματα LIME τεχνικής - Υπεργλυκαιμία, 30-60min . . . . .	80
7.7 Αποτελέσματα LIME τεχνικής - Υπογλυκαιμία, 0-30min . . . . .	81
7.8 Αποτελέσματα LIME τεχνικής - Υπογλυκαιμία, 0-30min . . . . .	82
7.9 Αποτελέσματα LIME τεχνικής - Υπογλυκαιμία, 30-60min . . . . .	82
7.10 Μέση τιμή βαρών - Υπεργλυκαιμία . . . . .	83
7.11 Μέση τιμή βαρών - Υπογλυκαιμία . . . . .	83

# Κατάλογος Πινάκων

6.1	Η κατανομή των κλάσεων για κάθε σύνολο δεδομένων . . . . .	52
6.2	Τιμές που ελέγχθηκαν για την βελτιστοποίηση των υπερπαραμέτρων . .	61
7.1	Αποτελέσματα αξιολόγησης για οριζόντια πρόβλεψης 0-30 λεπτά - Υπεργλυκαιμία . . . . .	68
7.2	Αποτελέσματα αξιολόγησης για οριζόντια πρόβλεψης 0-30 λεπτά - Υπεργλυκαιμία . . . . .	68
7.3	Αποτελέσματα Wilcoxon-signed rank test και paired t-test για την μετρική AUC και οριζόντια πρόβλεψης 0-30 λεπτών - Υπεργλυκαιμία . . . . .	69
7.4	Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική F1-score και οριζόντια πρόβλεψης 0-30 λεπτών - Υπεργλυκαιμία . . .	69
7.5	Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική ECE και οριζόντια πρόβλεψης 0-30 λεπτών - Υπεργλυκαιμία . . . . .	69
7.6	Αποτελέσματα αξιολόγησης για οριζόντια πρόβλεψης 30-60 λεπτά - Υπεργλυκαιμία . . . . .	70
7.7	Αποτελέσματα αξιολόγησης για οριζόντια πρόβλεψης 30-60 λεπτά - Υπεργλυκαιμία . . . . .	70
7.8	Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική AUC για οριζόντια πρόβλεψης 30-60 λεπτών - Υπεργλυκαιμία . . . . .	71
7.9	Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική F1-score για οριζόντια πρόβλεψης 30-60 λεπτών - Υπεργλυκαιμία . .	71
7.10	Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική ECE για οριζόντια πρόβλεψης 30-60 λεπτών - Υπεργλυκαιμία . . . . .	71
7.11	Αποτελέσματα αξιολόγησης για οριζόντια πρόβλεψης 0-30 λεπτά - Υπογλυκαιμία . . . . .	72
7.12	Αποτελέσματα αξιολόγησης για οριζόντια πρόβλεψης 0-30 λεπτά - Υπογλυκαιμία . . . . .	73
7.13	Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική AUC για οριζόντια πρόβλεψης 0-30 λεπτών - Υπογλυκαιμία . . . . .	74
7.14	Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική F1-score για οριζόντια πρόβλεψης 0-30 λεπτών - Υπογλυκαιμία . . .	74
7.15	Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική ECE για οριζόντια πρόβλεψης 0-30 λεπτών - Υπογλυκαιμία . . . . .	74
7.16	Αποτελέσματα αξιολόγησης για οριζόντια πρόβλεψης 30-60 λεπτά - Υπογλυκαιμία . . . . .	75
7.17	Αποτελέσματα αξιολόγησης για οριζόντια πρόβλεψης 30-60 λεπτά - Υπογλυκαιμία . . . . .	75

7.18	Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική AUC για οριζόντια πρόβλεψη 30-60 λεπτών - Υπογλυκαιμία . . . . .	76
7.19	Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική F1-score για οριζόντια πρόβλεψη 30-60 λεπτών - Υπογλυκαιμία . . . . .	76
7.20	Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική ECE για οριζόντια πρόβλεψη 30-60 λεπτών - Υπογλυκαιμία . . . . .	76

# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Το ερευνητικό πρόβλημα

Ο Σακχαρώδης Διαβήτης αποτελεί χρόνια πάθηση η οποία χαρακτηρίζεται από υπερβολικά υψηλές τιμές γλυκόζης του αίματος, είτε λόγω περιορισμένης έκκρισης της ορμόνης ινσουλίνης, είτε λόγω μειωμένης δράσης της, είτε λόγω συνδυασμού των δύο αυτών παραγόντων. Η ασθένεια σχετίζεται άμεσα με μακροχρόνια βλάβη, δυσλειτουργία και τελική ανεπάρκεια πληθώρας ζωτικών οργάνων, όπως των ματιών, των νεφρών, των νεύρων, της καρδιάς και αιμοφόρων αγγείων. Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας, τις τελευταίες δεκαετίες ο διαβήτης έχει λάβει χαρακτηριστικά επιδημίας, καθώς τα αντίστοιχα περιστατικά ανήλθαν από 108 εκατομμύρια το 1980, σε 422 εκατομμύρια το 2014.

Για την αντιμετώπισή του, πέρα από την υιοθέτηση μιας ισορροπημένης διατροφής και την πραγματοποίηση φυσικής άσκησης, η ρύθμιση της γλυκόζης του αίματος μέσω τακτικών μετρήσεων και χορήγησης ινσουλίνης, είναι επίσης καίριας σημασίας. Παρόλα αυτά, η γλυκαιμική ρύθμιση αποτελεί μια απαιτητική και σύνθετη διαδικασία, καθώς οι τιμές γλυκόζης επηρεάζονται από πληθώρα παραγόντων, με αποτέλεσμα αυτή τελικά να γίνεται εμπειρικά και να είναι επιρρεπής σε σφάλματα. Τα τελευταία χρόνια, η εφαρμογή μοντέλων μηχανικής μάθησης έχει συμβάλει στην ανάπτυξη ακριβέστερων μεθόδων γλυκαιμικού ελέγχου που λαμβάνουν υπόψιν σε μεγαλύτερο βαθμό τις ατομικές ιδιαιτερότητες της κάθε περίπτωσης, οδηγώντας τελικά σε αποδοτικότερη και πιο εξατομικευμένη αντιμετώπιση. Η περαιτέρω εξέλιξη τέτοιων τεχνικών τεχνητής νοημοσύνης ενισχύει τις ελπίδες για μετρίαση των επιπλοκών της πάθησης και βελτίωση της γενικότερης ποιότητας ζωής των ασθενών.

### 1.2 Κίνητρο της εργασίας

Στόχος της παρούσας εργασία αποτελεί η ανάπτυξη ενός ερμηνεύσιμου μοντέλου εκτίμησης του κινδύνου εμφάνισης υπεργλυκαιμικών και υπογλυκαιμικών επεισοδίων, με χρήση τεχνικών μηχανικής μάθησης. Η πρόβλεψη αυτή δεν θα συνίσταται στην άμεση εκτίμηση των μελλοντικών τιμών γλυκόζης, αλλά στην εκτίμηση της πιθανότητας εμφάνισης ενός υπεργλυκαιμικού ή υπογλυκαιμικού επεισοδίου, εντός ενός ορισμένου μελλοντικού χρονικού διαστήματος. Ακόμα, κρίνεται σημαντική η ενσωμάτωση τεχνικών ερμηνευσιμότητας στο τελικό σύστημα, με στόχο την βαθύτερη κατανόηση του

τρόπου λειτουργίας του και της διαδικασίας εξαγωγής των αποτελεσμάτων του. Το χαρακτηριστικό της ερμηνευσιμότητας αποτελεί ολόένα και περισσότερο ζητούμενο για πολλά σύγχρονα μοντέλα μηχανικής μάθησης, ειδικά όταν αυτά επιστρατεύονται στον τομέα της υγείας και χρησιμοποιούνται συμβουλευτικά κατά τη λήψη ιατρικών αποφάσεων. Η παράθεση εξηγήσεων για τις εξόδους ενός μοντέλου ενισχύει την αξιοπιστία του, δρα ως μέσο επαλήθευσης της ορθής λειτουργίας του και συμβάλλει στον πιο εύκολο εντοπισμό σφαλμάτων που πιθανώς να εμφανιστούν κατά την εφαρμογή του μοντέλου σε πραγματικές, μη πειραματικές συνθήκες.

Τέλος, η διαθεσιμότητα νέων δεδομένων που δεν έχουν χρησιμοποιηθεί σε κάποια άλλη αντίστοιχη έρευνα, έδρασε ως επιπλέον κίνητρο για την διεκπεραίωση της παρούσας εργασίας. Το νέο αυτό σύνολο δεδομένων παρέχει πληροφορίες σχετικά με τα καταγεγραμμένα επίπεδα γλυκόζης του αίματος, τη φυσική δραστηριότητα, την πρόσληψη ινσουλίνης και υδατανθράκων των συμμετεχόντων. Η συμπερίληψη τέτοιων επιπρόσθετων χαρακτηριστικών αναμένεται να βελτιώσει την ακρίβεια του μοντέλου πρόβλεψης γλυκόζης, καθώς οι παραπάνω παράγοντες επηρεάζουν την εμφάνιση ή μη υπεργλυκαιμικών και υπογλυκαιμικών επεισοδίων.

### 1.3 Δομή της εργασίας

Το παρόν έγγραφο διαρθρώνεται σε 8 κεφάλαια. Τα πρώτα κεφάλαια αποσκοπούν στην θεωρητική θεμελίωση της εργασίας, μέσω βαθύτερης κατανόησης της φύσης του σακχαρώδη διαβήτη, αλλά και των μηχανισμών που διέπουν τα τεχνητά νευρωνικά δίκτυα και τις τεχνικές ερμηνευσιμότητας που εφαρμόζονται. Αντίστοιχα, το 2ο σκέλος της εργασίας αφορά κυρίως στην πειραματική διαδικασία που ακολουθείται. Στο τμήμα αυτό παρατίθενται οι αναπτυσσόμενες μέθοδοι, τα αποτελέσματα που προέκυψαν, αλλά και τελικά συμπεράσματα και παρατηρήσεις.

Αναλυτικότερα :

- Το παρόν κεφάλαιο (*Κεφάλαιο 1*) αποτελεί μια σύντομη εισαγωγή, μέσω της οποίας επιχειρείται η αποσαφήνιση του ερευνητικού προβλήματος και του κινήτρου της έρευνας, παράλληλα με την παρουσίαση της γενικότερης δομής της εργασίας.
- Στο *Κεφάλαιο 2* γίνεται μια γενική επισκόπηση της ασθένειας του σακχαρώδη διαβήτη. Συγκεκριμένα, παρουσιάζονται οι συχνότεροι τύποι της ασθένειας, τα αίτια εμφάνισης και τα συμπτώματά της, καθώς και βραχυπρόθεσμες και μακροπρόθεσμες επιπλοκές της. Τέλος, γίνεται αναφορά στους τρόπους διάγνωσης και ορθής διαχείρισής της.
- Στο *Κεφάλαιο 3* διεξάγεται μια εισαγωγή σε βασικές έννοιες των τεχνητών νευρωνικών δικτύων, με έμφαση στα αναδρομικά νευρωνικά δίκτυα, τα οποία χρησιμοποιούνται συχνά σε εφαρμογές σαν και αυτή που μελετάται. Ακόμα, αναφέρονται κύριες μέθοδοι εκπαίδευσης και αξιολόγησης μοντέλων μηχανικής μάθησης, πολλές εκ των οποίων εφαρμόζονται και στη συνέχεια, κατά το πειραματικό σκέλος.
- Στο *Κεφάλαιο 4* παρουσιάζεται η έννοια της ερμηνεύσιμης τεχνητής νοημοσύνης, γίνεται ταξινόμησή της και επισημαίνεται η αξία της στο πεδίο της βιοϊατρικής. Επιπλέον, γίνεται αναφορά σε δύο δημοφιλείς τεχνικές ερμηνευσιμότητας, LIME και SHAP.

- Στο *Κεφάλαιο 5* επιχειρείται η βαθύτερη κατανόηση και η εξοικείωση με το ερευνητικό πρόβλημα που αντιμετωπίζεται, μέσω επισκόπησης της σχετιζόμενης, πιο πρόσφατης βιβλιογραφίας. Έμφαση δίνεται σε σύγχρονα μοντέλα πρόβλεψης των επιπέδων γλυκόζης ή υπογλυκαιμικών/υπεργλυκαιμικών επεισοδίων με τη χρήση μηχανικής μάθησης. Ειδική αναφορά γίνεται ακόμα σε συστήματα που ενσωματώνουν τεχνικές ερμηνευσιμότητας και μοντέλα προσομοίωσης φυσιολογικών συστημάτων.
- Στο *Κεφάλαιο 6* παρουσιάζονται οι τεχνικές που χρησιμοποιήθηκαν για την επεξεργασία των διαθέσιμων δεδομένων, καθώς και για την ανάπτυξη και αξιολόγηση του εφαρμοζόμενου μοντέλου μηχανικής μάθησης.
- Στο *Κεφάλαιο 7* παρέχονται και σχολιάζονται τα αποτελέσματα των πειραμάτων από την εφαρμογή των προτεινόμενων μεθόδων και αλγορίθμων για τις περιπτώσεις της υπογλυκαιμίας και της υπεργλυκαιμίας.
- Τέλος, στο *Κεφάλαιο 8* συνοψίζονται τα συμπεράσματα της διπλωματικής εργασίας και προτείνονται ιδέες για μελλοντική ερευνητική μελέτη.



# Κεφάλαιο 2

## Σακχαρώδης Διαβήτης

### 2.1 Γενικά στοιχεία και ταξινόμηση

Ο Σακχαρώδης Διαβήτης (ή ΣΔ για συντομία) αποτελεί χρόνια σύνδρομο με παγκόσμια κατανομή και σοβαρές επιπλοκές για τους πάσχοντες. Σύμφωνα με την Παγκόσμια Ομοσπονδία Διαβήτη (International Diabetes Federation), το 2021 οι συνολικές περιπτώσεις ενηλίκων με ΣΔ ήταν 537 εκατομμύρια, ενώ οι σχετιζόμενοι θάνατοι 6.7 εκατομμύρια. Αξίζει να σημειωθεί ακόμα ότι ο συνολικός αριθμός νοσούντων αναμένεται να ανέλθει σε 643 εκατομμύρια έως το 2030. Τρεις στους τέσσερις ενηλίκους με διαβήτη ζουν σε χώρες χαμηλού ή μεσαίου εισοδήματος, ενώ οι μισοί εξ αυτών δεν έχουν διαγνωστεί. [1] Σε όλες τις μορφές του, ο ΣΔ χαρακτηρίζεται από απορρύθμιση του μεταβολισμού υδατανθράκων, λιπών και πρωτεϊνών, με αποτέλεσμα την εμφάνιση αυξημένων τιμών γλυκόζης του αίματος (ή υπεργλυκαιμικών επεισοδίων). Διακρίνεται στο σακχαρώδη διαβήτη τύπου 1, στο σακχαρώδη διαβήτη τύπου 2, στο διαβήτη κύησης και σε άλλους ειδικούς, πιο σπάνιους τύπους.

#### 2.1.1 Σακχαρώδης Διαβήτης Τύπου 1

Ο ΣΔ τύπου 1 ευθύνεται περίπου για το 5%-10% όλων των περιπτώσεων της νόσου. Στον τύπο αυτό, ο ίδιος ο οργανισμός του ασθενή καταστρέφει σταδιακά τα παγκρεατικά κύτταρα που παράγουν την ορμόνη ινσουλίνη, μέσω αυτοάνοσου μηχανισμού. Δίχως τη συγκεκριμένη ορμόνη δεν είναι δυνατή η ρύθμιση των επιπέδων γλυκόζης στο αίμα, επομένως η καθημερινή, εξωτερική χορήγηση ινσουλίνης κρίνεται απαραίτητη από την αρχή της ασθένειας. Ο σακχαρώδης διαβήτης τύπου 1 συνήθως εμφανίζεται σε παιδιά, εφήβους και νεαρούς ενήλικες, αλλά πιο σπάνια μπορεί να εμφανιστεί και σε μεγαλύτερες ηλικίες. Οι αιτίες εκδήλωσής του δεν είναι σαφώς ορισμένες, παρόλα αυτά η ύπαρξη περιπτώσεων ΣΔ τύπου 1 ή άλλων αυτοάνοσων νοσημάτων στο οικογενειακό ιστορικό υγείας, αποτελεί αποδεδειγμένος παράγοντας κινδύνου. Το βάρος και η φυσική κατάσταση δεν σχετίζονται άμεσα με την εμφάνιση του ΣΔ τύπου 1, αν και συνήθειες όπως η τακτική φυσική άσκηση και η υιοθέτηση μιας ισορροπημένης διατροφής συμβάλλουν καθοριστικά στην αντιμετώπιση του. [2]

### **2.1.2 Σακχαρώδης Διαβήτης Τύπου 2**

Ο ΣΔ τύπου 2 συνιστά την σημαντικότερη πλειοψηφία περιστατικών διαβήτη παγκοσμίως (90% - 95%) και προκαλείται λόγω αυξημένης αντίστασης του οργανισμού στη δράση της ινσουλίνης. Τα επίπεδα έκκρισης της ορμόνης μπορεί να είναι φυσιολογικά ή και υψηλότερα στα αρχικά στάδια της νόσου, παρόλα αυτά με την εξέλιξη της, παύουν να επαρκούν. Στη μεταγενέστερη αυτή φάση τα επίπεδα παραγωγής ινσουλίνης φθίνουν και επιβάλλεται η εξωτερική χορήγησή της. Τα συνηθέστερα κλινικά χαρακτηριστικά ασθενών με ΣΔ τύπου 2 είναι η παχυσαρκία, η μέση ή και προχωρημένη ηλικία, η ύπαρξη αντίστοιχου περιστατικού σε συγγενή πρώτου βαθμού, καθώς και η προγενέστερη, σταδιακά ενισχυόμενη υπεργλυκαιμία, με ήπια ή και καθόλου συμπτώματα. Ειδικά ο καθιστικός τρόπος ζωής και η αδράνεια, σε συνδυασμό με τη μη ισορροπημένη διατροφή που συναντώνται σε πολλές χώρες, ευθύνονται για σημαντικό ποσοστό των περιπτώσεων διαβήτη τύπου 2 σήμερα, καθώς και για τη συχνότερη εμφάνιση της ασθένειας σε ολοένα και νεότερες ηλικίες. Παρά τα παραπάνω συχνά κοινά κλινικά χαρακτηριστικά, οι ασθενείς με ΣΔ τύπου 2 διακατέχονται από αξιοσημείωτη ετερογένεια: η νόσος μπορεί να εμφανιστεί σε παιδιά και εφήβους, αλλά και σε πιο προχωρημένες ηλικίες, σε παχύσαρκα ή φυσιολογικού βάρους άτομα, ενώ παράλληλα οι περιπτώσεις κυμαίνονται από σχεδόν ασυμπτωματικές, σε σοβαρές και απειλητικές για την ζωή. Αυτή η ποικιλομορφία δυσχεραίνει ακόμα περισσότερο τη διαχείριση του διαβήτη από ειδικούς υγείας και πάσχοντες και βαθαίνει την ανάγκη περαιτέρω εξατομίκευσής της. [2]

### **2.1.3 Σακχαρώδης Διαβήτης Κυήσεως**

Ο ΣΔ κυήσεως (ή εγκυμοσύνης) είναι ένα τύπος διαβήτη που εμφανίζεται μόνο κατά τη διάρκεια της κύησης. Αυτή η μορφή διαβήτη επηρεάζει το 3%-9% των εγκυμονούντων και σχετίζεται με ορμονικές μεταβολές που συντελούνται κατά τη διάρκεια της εγκυμοσύνης, ειδικά κατά το δεύτερο και τρίτο εξάμηνο αυτής. Το επιπλέον βάρος και η κληρονομικότητα επίσης αποτελούν παράγοντες κινδύνου. Αν και τα υπεργλυκαιμικά συμπτώματα παύουν μετά τη γέννα, σημαντική μερίδα των γυναικών που εμφανίζουν ΣΔ κυήσεως στη συνέχεια θα εμφανίσουν και ΣΔ τύπου 2, επομένως η πρόληψη και ο συστηματικός έλεγχος είναι κρίσιμης σημασίας για τα χρόνια που ακολουθούν. Σημειώνουμε ακόμα ότι στην περίπτωση που η νόσος δεν αντιμετωπιστεί αποτελεσματικά κατά τη διάρκεια της κύησης, υπάρχουν αυξημένες πιθανότητες πρόωρης γέννας ή γέννας υπέρβαρου βρέφους, ενώ το ίδιο το παιδί βρίσκεται σε κίνδυνο εκδήλωσης παχυσαρκίας ή σακχαρώδη διαβήτη τύπου 2 μεγαλώνοντας. [2]

## **2.2 Μεταβολισμός και ομοιοστάση της γλυκόζης**

Από τα παραπάνω ίσως να έχει γίνει ήδη κατανοητή η επίδραση που έχει η περιορισμένη παραγωγή ή/και απορρόφηση της ορμόνης ινσουλίνης στα επίπεδα γλυκόζης του αίματος. Παρόλα αυτά, ο μεταβολισμός της γλυκόζης αποτελεί μια σύνθετη διαδικασία, στην οποία εμπλέκονται και άλλες ορμόνες, πέρα από την ινσουλίνη. Με στόχο την πληρότητα του παρόντος κεφαλαίου και τη βαθύτερη κατανόηση της παθολογίας και πολυπλοκότητας του σακχαρώδους διαβήτη, στη συνέχεια παρουσιάζονται γενικά στοιχεία για την ομοιοστάση της γλυκόζης, καθώς και για τις ορμόνες που συμμετέχουν στη διαδικασία αυτή.

### 2.2.1 Μεταβολισμός γλυκόζης

Η γλυκόζη αποτελεί τον πιο συνήθη μονοσακχαρίτη και την πιο σημαντική πηγή ενέργειας για φυτά και ζώα. Στους ανθρώπους, η γλυκόζη παράγεται μέσω της πρόσληψης τροφής και των επιμέρους διαδικασιών της *γλυκογονόλυσης* και της *γλυκονεογένεσης*. Όσον αφορά τη διαδικασία πρόσληψης τροφής, οι υδατάνθρακες αποτελούν τη βασική εξωτερική πηγή γλυκόζης. Οι υδατάνθρακες είναι πολυσακχαρίτες που συναντώνται σε πληθώρα τροφών και κατά το πλείστον στο άμυλο φυτικών τροφίμων, όπως τα δημητριακά, τις πατάτες και το ρύζι. Λίγα μόλις λεπτά μετά την κατανάλωση υδατανθράκων έχει ήδη αρχίσει η διαδικασία υδρόλυσής τους σε απλούστερους μονοσακχαρίτες (με κυριότερο εξ αυτών τη γλυκόζη), πρώτα από το στομάχι κι έπειτα από το μικρό έντερο. Μετά την απορρόφηση της από το μικρό έντερο, η απορρέουσα γλυκόζη εισέρχεται στην ηπατική πυλαία φλέβα και λαμβάνεται απευθείας από το ήπαρ. Εκεί απορροφάται απευθείας ένα τμήμα της γλυκόζης για την εξυπηρέτηση ηπατικών λειτουργιών, ενώ το μεγαλύτερο μέρος της διέρχεται στην κυκλοφορία του αίματος για να διανεμηθεί στον εγκέφαλο, τους μύς και τους υπόλοιπους ιστούς.

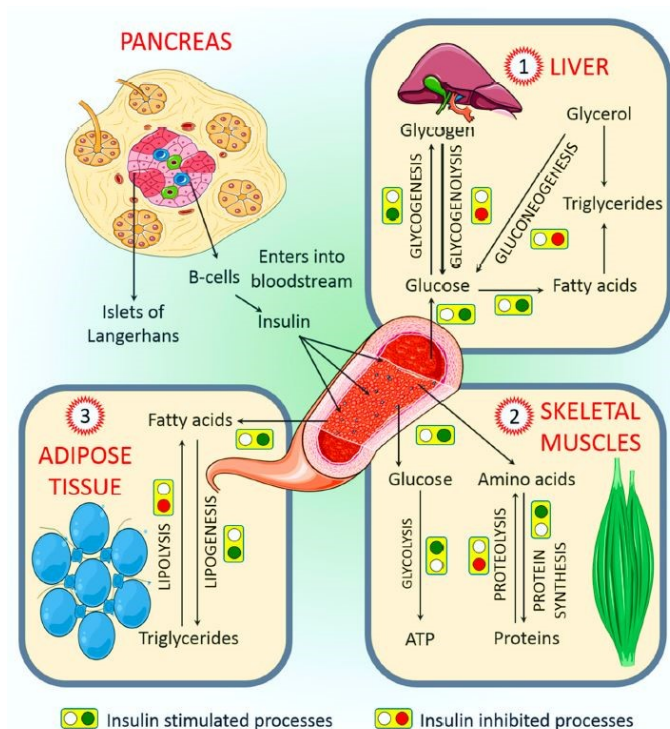
Η γλυκόζη κινείται από το διάμεσο υγρό στο εσωτερικό των κυττάρων με τη βοήθεια της κατηγορίας μεμβρανικών φορέων που ονομάζεται GLUT. Στα κύτταρα, το μεγαλύτερο μέρος της εισερχόμενης γλυκόζης καταναλώνεται άμεσα για τη σύνθεση ATP, της οργανικής ένωσης που χρησιμοποιείται ως το κυτταρικό «ενεργειακό νόμισμα» και τροφοδοτεί τις περισσότερες κυτταρικές διεργασίες. Η γλυκόζη που δεν χρησιμοποιείται για ενεργειακή σύνθεση αποθηκεύεται για μελλοντική χρήση, είτε ως το πολυμερές *γλυκογόνο*, είτε ως λίπος. Το γλυκογόνο αποτελεί ένα μεγάλο πολυσακχαρίτη και προκύπτει ως η σύνδεση πολλών επιμέρους μορίων γλυκόζης σε μια διακλαδιζόμενη αλυσίδα. Όταν είναι απαραίτητο, αποσυντίθεται για να παραχθεί γλυκόζη, μέσω της οδού που ονομάζεται *γλυκογονόλυση*. Οι εκτενέστερες αποθήκες γλυκογόνου εντοπίζονται στο ήπαρ, με υψηλές συγκεντρώσεις να περιέχονται και στους σκελετικούς μύς. Αν και οι υδατάνθρακες αποτελούν το πρωτεύον συστατικό για τη δημιουργία γλυκόζης, τα αμινοξέα των πρωτεϊνών και τα τριγλυκερίδια των λιπαρών οξέων μπορούν επίσης να χρησιμοποιηθούν για την παρασκευή της. Συγκεκριμένα, εάν η πρόσληψη υδατανθράκων είναι χαμηλή, τα αμινοξέα μετατρέπονται σε γλυκόζη μέσω της οδού που είναι γνωστή ως *γλυκονεογένεση*. Τέλος, και η γλυκερίνη των τριγλυκεριδίων μπορεί να χρησιμοποιηθεί για τη σύνθεση γλυκόζης σε περιόδους νηστείας. [3], [4]

### 2.2.2 Ομοίωση γλυκόζης

Από τα παραπάνω ίσως να έχει ήδη διαφανεί η πολυπλοκότητα της διαδικασίας ομοίωσης της γλυκόζης, αφού οι αντιδράσεις στις οποίες αυτή συμμετέχει ποικίλλουν και εμπλέκουν πολλούς διαφορετικούς ιστούς και όργανα. Παράλληλα, η οριοθέτηση των επιπέδων της γλυκόζης του αίματος σε ένα περιορισμένο εύρος είναι κρίσιμης σημασίας, καθώς υπό φυσιολογικές συνθήκες η γλυκόζη αποτελεί την κυριότερη πηγή ενέργειας των περισσότερων ιστών και επίσης τη μόνη θρεπτική ουσία που μπορεί να μεταβολίσει ο εγκέφαλος. [3] Πράγματι, τα επίπεδα γλυκόζης αίματος ενός υγιούς ατόμου ρυθμίζονται στενά έτσι ώστε να παραμένουν εντός ενός προκαθορισμένου φάσματος, παρά τα γεύματα που λαμβάνονται κατά τη διάρκεια της ημέρας. Ο έλεγχος αυτός πραγματοποιείται με τη χρήση ορμονών, οι σημαντικότερες εκ των οποίων είναι η ινσουλίνη και η γλυκαγόνη, ενώ και άλλοι τύποι σχετιζόμενων ορμονών έχουν ανακαλυφθεί τις τελευταίες δεκαετίες. Οι ορμόνες αυτές θα μας απασχολήσουν στη συνέχεια, αρχίζοντας από την ινσουλίνη.

## Ινσουλίνη

Η ινσουλίνη αποτελεί μια τυπική πεπτιδική ορμόνη. Συγκροτείται από δύο πολυπεπτιδικές αλυσίδες και 51 διαφορετικούς τύπους αμινοξέων, ενώ απομονώθηκε για πρώτη φορά το 1921 από τον γιατρό Frederick Banting. Παράγεται από τα *β-κύτταρα* του παγκρέατος, έναν από τους τέσσερις διακριτούς τύπους κυττάρων που συγκροτούν τις *νησίδες Langerhans*, συστάδες κυττάρων διάσπαρτων στο πάγκρεας υπεύθυνων για την έκκριση σημαντικών ορμονών. Οι νησίδες Langerhans συνιστούν λιγότερο από το 2% του παγκρέατος, ενώ τα β-κύτταρα αποτελούν περίπου το 75% των νησίδων αυτών. Ο πρωταρχικός ρόλος της ινσουλίνης είναι η μείωση των επιπέδων γλυκόζης του αίματος έπειτα από κάθε γεύμα και η επαναφορά τους στις αρχικές, φυσιολογικές τιμές. Επιπρόσθετες λειτουργίες της περιλαμβάνουν τη βελτίωση της κυτταρικής χρήσης και αποθήκευσης της γλυκόζης, την ενεργοποίηση ενζύμων σύνθεσης γλυκογόνου, πρωτεϊνών και λίπους, καθώς επίσης και την αναστολή παραγωγής νέας γλυκόζης, μέσω διαδικασιών όπως η γλυκονεογένεση.



Σχήμα 2.1: Ο φυσιολογικός ρόλος της ινσουλίνης σε διαδικασίες του παγκρέατος, του λιπώδους ιστού και των σκελετικών μυών [5]

Αναλυτικότερα, η απορρόφηση των θρεπτικών συστατικών ενός γεύματος αυξάνει τη συγκέντρωση γλυκόζης και αμινοξέων στο πλάσμα, ενεργοποιώντας τα ευαίσθητα στη γλυκόζη, β-κύτταρα. Με την ενεργοποίηση των β-κυττάρων απελευθερώνεται ινσουλίνη, η οποία στη συνέχεια στοχεύει κυρίως σε κύτταρα του ήπατος, του εγκεφάλου και των λιπωδών και μυϊκών ιστών. Η πρόσδεση της ινσουλίνης στους μεμβρανικούς υποδοχείς αυτών των κυττάρων-στόχων πυροδοτεί τη διαδοχική φωσφορύλιωση σειράς πρωτεϊνών εντός τους, ωστόσο τελικά ενεργοποιηθούν οι μεταφορείς GLUT4. Οι μεταφορείς αυτοί εξέρχονται από τα κυτταροπλασματικά κυστίδια όπου είναι αποθηκευμένοι και πλησιάζουν την κυτταρική μεμβράνη, επιτρέποντας τελικά σε σημαντικές ποσότητες γλυκόζης

να εισέλθουν, μέσω διευκολυνόμενης διάχυσης. Έτσι, τα επίπεδα γλυκόζης του αίματος μειώνονται και επενέρχονται σταδιακά στα φυσιολογικά τους επίπεδα. [3], [4], [5]

### **Γλυκαγόνη**

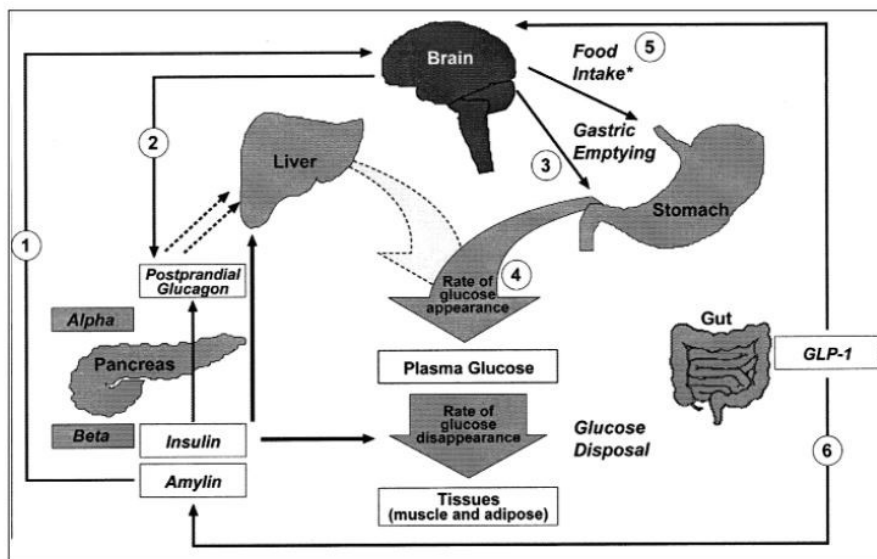
Η γλυκαγόνη είναι μία ακόμη σημαντική πεπτιδική ορμόνη για τη γλυκαιμική ρύθμιση που ανιχνεύθηκε για πρώτη φορά το 1922. Αποτελείται από 29 αμινοξέα και παράγεται από τα *α-κύτταρα* των νησίδων Langerhans του παγκρέατος. Δρα ανταγωνιστικά σε σχέση με την ινσουλίνη κι εκκρίνεται όταν οι τιμές γλυκόζης του πλάσματος είναι χαμηλές, διεγείροντας τις διεργασίες σύνθεσης γλυκόζης, γλυκονόλυση και γλυκονεογένεση. Συγκεκριμένα, η διάσπαση γλυκογόνου (γλυκογονόλυση) από το ήπαρ είναι η κυριότερη πηγή γλυκόζης κατά τις 8-12 πρώτες ώρες νηστείας, ενώ για μεγαλύτερα διαστήματα σημαντικές ποσότητες γλυκόζης παράγονται και μέσω ηπατικής γλυκονεογένεσης, με τη χρήση αμινοξέων. [3]

Συμπερασματικά, η αναλογία ινσουλίνης-γλυκαγόνης είναι αυτή που καθορίζει κυρίως το ποια ορμόνη θα επικρατήσει τελικά και το πως θα κινηθούν τα επίπεδα γλυκόζης του αίματος. Στη φάση νηστείας η μεταβολική ρύθμιση αποτρέπει την εμφάνιση υπογλυκαιμικών επεισοδίων, μέσω έκκρισης γλυκαγόνης. Όταν κυριαρχεί η γλυκαγόνη, το ήπαρ χρησιμοποιεί κυρίως το γλυκογόνο για να απελευθερώσει γλυκόζη στο αίμα. Σε ένα άτομο με φυσιολογικό μεταβολισμό, η γλυκόζη νηστείας του πλάσματος διατηρείται περίπου στα 90mg/dL, η έκκριση ινσουλίνης είναι χαμηλή και τα επίπεδα γλυκαγόνης είναι σχετικά υψηλά. Μετά την απορρόφηση των θρεπτικών συστατικών ενός γεύματος, η γλυκόζη του πλάσματος αυξάνεται. Η άνοδος αυτή αναστέλλει την έκκριση γλυκαγόνης και διεγείρει την απελευθέρωση ινσουλίνης. Η ινσουλίνη με τη σειρά της ενισχύει τη μεταφορά γλυκόζης στα κύτταρα όπως προαναφέρθηκε, με αποτέλεσμα οι συγκεντρώσεις γλυκόζης στο πλάσμα να αρχίζουν να πέφτουν σύντομα πίσω στα επίπεδα νηστείας. [3]

### **Άλλες σχετιζόμενες ορμόνες**

Αν και η ινσουλίνη και η γλυκαγόνη αποτελούν τις κυριότερες ορμόνες για την οριοθέτηση των επιπέδων της γλυκόζης στο αίμα, κι άλλες σχετιζόμενες ορμόνες έχουν ανιχνευθεί τα τελευταία χρόνια. Η *αμυλίνη* για παράδειγμα αποτελεί μια πεπτιδική ορμόνη, με συμπληρωματική ως προς την ινσουλίνη, δράση. Αποτελείται από 37 αμινοξέα και ανιχνεύθηκε μόλις το 1987. Όπως και η ινσουλίνη, παράγεται κι εκκρίνεται από τα β-κύτταρα του παγκρέατος, ενώ με βάση έρευνες που έχουν πραγματοποιηθεί σε ανθρώπους, οι δύο αυτές ορμόνες απελευθερώνονται παράλληλα, ενεργοποιούμενες έπειτα από κάθε γεύμα. Η αναλογία απελευθέρωσης ινσουλίνης-αμυλίνης είναι αρχικά 50:1, ενώ σταδιακά διαμορφώνεται σε 20:1. Μία από τις βασικές χρήσεις της αμυλίνης είναι η επιβράδυνση του ρυθμού με τον οποίο το λεπτό έντερο απορροφά τη γλυκόζη από το στομάχι. Καθυστερώντας τη μεταφορά της γλυκόζης από το στομάχι, στο λεπτό έντερο και έπειτα στο ήπαρ και την κυκλοφορία του αίματος, τα επίπεδα γλυκόζης του πλάσματος αυξάνονται με πιο σταδιακό και αργό ρυθμό, συμβάλλοντας έτσι στην αποτροπή υπεργλυκαιμικών επεισοδίων. Μια ακόμα λειτουργία της αμυλίνης είναι η καταστολή της παραγωγής γλυκαγόνης από τα α-κύτταρα του παγκρέατος. Όπως αναφέρθηκε και πρωτύτερα, η ορμόνη γλυκαγόνη προωθεί τη σύνθεση γλυκόζης, επομένως ο περιορισμός της έκκρισης της είναι απαραίτητος για τη επιθυμητή μείωση των επιπέδων γλυκόζης. Επιπλέον, η αμυλίνη ενισχύει το αίσθημα του κορεσμού, με στόχο τον περιορισμό της πρόσληψης τροφής και τη διατήρηση ενός φυσιολογικού βάρους. [3]

Τέλος, δύο ακόμα σχετιζόμενες πεπτιδικές ορμόνες είναι οι *GIP* (Glucose-dependent Insulinotropic Polypeptide) και *GLP-1* (Glucagon-like Peptide-1). Παράγονται από τα L-κύτταρα του εντέρου και όπως η ινσουλίνη και η αμυλίνη συνεισφέρουν στην αποτροπή εμφάνισης υπεργλυκαιμικών επεισοδίων. Η *GLP-1* αποτελείται από 30 αμινοξέα και όπως η αμυλίνη συμβάλλει στη σταδιακή διάχυση της γλυκόζης προς το αίμα, καθώς και στην αναστολή παραγωγής γλυκαγόνης. Επιπρόσθετα, η απελευθέρωση *GLP-1* ορμόνης προκαλεί έκκριση ινσουλίνης από το πάγκρεας, πολλαπλασιασμό των β-κυττάρων, περιορισμό της γλυκονεογένεσης στο ήπαρ και ενίσχυση του αισθήματος του κορεσμού. Παρόμοια, η ορμόνη *GIP* επίσης προωθεί τη παραγωγή ινσουλίνης από το πάγκρεας και την ταχεία αναπαραγωγή των β-κυττάρων. Και οι δύο παραπάνω ορμόνες (*GIP*, *GLP-1*) απελευθερώνονται έπειτα από την πρόσληψη γευμάτων, ειδικά όταν αυτά είναι πλούσια σε σάκχαρα και λίπη. [6]



Σχήμα 2.2: Ομοίωση της γλυκόζης: οι ρόλοι της ινσουλίνης, γλυκαγόνης, αμυλίνης και *GLP-1* [6]

Στην περίπτωση του σακχαρώδη διαβήτη, η παραπάνω μεταβολική διαδικασία έχει διαταραχθεί με έναν ή περισσότερους τρόπους, με αποτέλεσμα ο οργανισμός να αδυνατεί να ασκήσει αποτελεσματικά γλυκαιμικό έλεγχο. Ειδικότερα, κατά τον ΣΔ τύπου 1, συντελείται σταθερή καταστροφή των β-κυττάρων του παγκρέατος από τον ίδιο τον οργανισμό, έχοντας ως συνέπεια την μερική ή και ολική ανεπάρκεια ινσουλίνης, αλλά και αμυλίνης. Ως αποτέλεσμα, τα μεταγευματικά επίπεδα γλυκόζης αυξάνονται λόγω της ανεπαρκούς απομάκρυνσης της γλυκόζης από το αίμα, της κακώς ρυθμιζόμενης παραγωγής γλυκόζης από το ήπαρ και τον επιταχυμένο ρυθμό γαστρικής εκκένωσης. Στο ΣΔ τύπου 2, η δράση των β-κυττάρων επίσης παρουσιάζεται ελαττωματική, ενώ παράλληλα οι περιφερειακοί ιστοί εμφανίζουν ανθεκτικότητα απέναντι στη δράση της ινσουλίνης, με αποτέλεσμα και πάλι ανεπαρκή γλυκαιμική ρύθμιση. Τέλος, σε όλους τους τύπους σακχαρώδη διαβήτη εμφανίζονται διαταραχές κατά την παραγωγή γλυκαγόνης και μειωμένη απελευθέρωση της ορμόνης *GLP-1*. [3], [6]

## 2.3 Συμπτώματα και επιπλοκές

Σε αυτό το σημείο έχει ολοκληρωθεί η επισκόπηση σημαντικών μηχανισμών που συνεισφέρουν στην ομοιόσταση της γλυκόζης και η παρουσίαση της βασικών αιτιών εμφάνισης του σακχαρώδους διαβήτη. Στην παρούσα υποενότητα θα εξεταστούν τα συμπτώματα της ασθένειας, καθώς και συχνές επιπλοκές που εμφανίζονται κατά την εξέλιξή της.

### 2.3.1 Συμπτώματα

Στα πρώτα στάδια της νόσου οι ασθενείς πιθανώς να είναι πλήρως ασυμπτωτικοί, με αποτέλεσμα να μην γνωρίζουν ότι πάσχουν. Σε αυτήν την περίπτωση η διάγνωση του σακχαρώδη διαβήτη μπορεί να γίνει έπειτα από την εμφάνιση οξείων ή και μακροχρόνιων επιπλοκών, που αναλύονται εκτενέστερα παρακάτω. Παρόλα αυτά, συνήθη αρχικά συμπτώματα περιλαμβάνουν πολυδιψία, ξηροστομία, πολυφαγία, ανεξήγητη απώλεια βάρους, κόπωση και θαμπή όραση. [7]

### 2.3.2 Οξείες επιπλοκές

Η παρατεταμένη έκθεση διαφόρων σημείων του σώματος στις υπεργλυκαιμίες που συναντώνται στο σακχαρώδη διαβήτη, συνδέονται με την εμφάνιση ποικίλων οξείων και μακροπρόθεσμων επιπλοκών. Στη συνέχεια αναφέρουμε τα κύρια χαρακτηριστικά της διαβητικής κετοξέωσης (ΔΚΟ), της υπεροσμωτικής υπεργλυκαιμικής κατάστασης (ΥΥΚ) και της υπογλυκαιμίας, οι οποίες και συγκαταλέγονται στις βασικότερες οξείες επιπλοκές της ασθένειας.

#### Διαβητική κετοξέωση (ΔΚΟ)

Η διαβητική κετοξέωση αποτελεί οξεία μεταβολική διαταραχή η οποία εμφανίζεται σε συνθήκες μεγάλης έλλειψης ινσουλίνης και υπερέκκρισης ανταγωνιστικών της ορμονών, όπως της γλυκαγόνης. Η ελαττωμένη χρησιμοποίηση της γλυκόζης από τους περιφερειακούς ιστούς, σε συνδυασμό με την αύξηση της γλυκονεογένεσης και της γλυκονόλυσης προκαλούν υπεργλυκαιμία. Όταν τα επίπεδα γλυκόζης του αίματος υπερβούν το λεγόμενο νεφρικό ουδό (180 mg/dl), προκαλείται γλυκοζουρία (αποβολή γλυκόζης με τα ούρα), ωσμωτική διούρηση και πολυουρία, με αποτέλεσμα την απώλεια ύδατος και ηλεκτρολυτών. Η απώλεια ύδατος οδηγεί σε αφυδάτωση και σε υποξία των ιστών, δηλαδή σε ανεπαρκή οξυγόνωση τους. Παράλληλα, η ανεπάρκεια γλυκόζης στα κύτταρα απορρυθμίζει τον μεταβολισμό των λιπαρών οξέων και έχει ως αποτέλεσμα την απελευθέρωση κετονικών σωμάτων. Τα οξέα αυτά, αν και αβλαβή σε μικρές συγκεντρώσεις, όταν συσσωρεύονται μπορούν να διαταράξουν την οξεοβασική ισορροπία του οργανισμού και να οδηγήσουν σε μεταβολική οξέωση.

Συμπτώματα της διαβητικής κετοξέωσης αποτελούν η πολυδιψία, η αφυδάτωση, η πολυουρία, η ναυτία, ο εμετός, η κόπωση, καθώς και η φρουτώδης οσμή στην αναπνοή. Η ΔΚΟ συναντάται συχνότερα σε ασθενείς με σακχαρώδη διαβήτη τύπου 1, αλλά μπορεί να εμφανιστεί και σε αυτούς με ΣΔ τύπου 2. Αιτίες εμφάνισης της μπορεί να είναι η ύπαρξη λοίμωξης, η παράλειψη χορήγησης της απαιτούμενης ινσουλίνης, καταστάσεις στρες, χειρουργικές επεμβάσεις, αλλά και η χρήση ναρκωτικών ουσιών. Αν δεν υπάρξει άμεση θεραπεία η διαβητική κετοξέωση μπορεί να αποβεί μοιραία. Στους τρόπους

αντιμετώπισης περιλαμβάνονται η λήψη ινσουλίνης, υγρών και ηλεκτρολυτών, τα οποία χορηγούνται με ενδοφλέβια ένεση. Οι ασθενείς συνήθως είναι αφυδατωμένοι και η άμεση ενυδάτωση παίζει σημαντικό ρόλο στην βελτίωση της κατάστασης τους. Τέλος, είναι απαραίτητη η παρακολούθηση της αρτηριακής πίεσης, αλλά και της γενικότερης υγείας του ασθενούς, κατά τις πρώτες ώρες που ακολουθούν τη χορήγηση υγρών για την αντιμετώπιση της ΔΚΟ. [7], [8]

### **Υπερωσμωτική υπεργλυκαιμική κατάσταση (ΥΥΚ)**

Η υπερωσμωτική υπεργλυκαιμική κατάσταση (ΥΥΚ) αποτελεί μια σχετικά συχνή, επικίνδυνη για τη ζωή, ενδοκρινική διαταραχή. Μαζί με τη διαβητική κετοξέωση αποτελούν τις πιο σοβαρές οξείες μεταβολικές επιπλοκές του σακχαρώδους διαβήτη. Αν και η ΥΥΚ μπορεί να εμφανιστεί και σε νοσούντες με ΣΔ τύπου 1, είναι συνηθέστερη σε ασθενείς μεγαλύτερης ηλικίας με ΣΔ τύπου 2. Χαρακτηρίζεται από ακραία υπεργλυκαιμία, υπερωσμωτικότητα (παθολογική κατάσταση όπου η σχέση διαλυμένων σωματιδίων προς το ολικό ύδωρ του οργανισμού είναι αυξημένη) και ήπια ή καθόλου κέτωση, ενώ οι λοιμώξεις αποτελούν την κυριότερη αιτία εμφάνισης της. Κλινικά εκδηλώνεται κυρίως με έντονη αφυδάτωση και ποικίλες νευρολογικές διαταραχές. Αν και η υπερωσμωτική υπεργλυκαιμική κατάσταση διακρίνεται από την διαβητική κετοξέωση ως προς τον χρόνο έναρξης, τον βαθμό αφυδάτωσης και τη σοβαρότητα της κέτωσης, οι δύο αυτές διαταραχές μπορεί να συνυπάρξουν. Όπως και η ΔΚΟ, η ΥΥΚ αντιμετωπίζεται μέσω διόρθωσης της αφυδάτωσης, των ηλεκτρολυτικών διαταραχών και της υπεργλυκαιμίας. Σημαντικές είναι ακόμη η αναγνώριση και η αντιμετώπιση συνοδών παθολογικών καταστάσεων, η πρόληψη πιθανών επιπλοκών και η εντατική παρακολούθηση του ασθενή. [7]

### **Υπογλυκαιμία**

Η υπογλυκαιμία είναι μία ακόμα συνήθης, οξεία επιπλοκή του σακχαρώδους διαβήτη που επηρεάζει κυρίως άτομα στα οποία γίνεται εξωτερική χορήγηση ινσουλίνης. Χαρακτηρίζεται από παθολογικά χαμηλές τιμές γλυκόζης αίματος (< 70mg/dl) και εκδηλώνεται με συμπτώματα δυσφορίας, αδυναμίας, εφίδρωσης, πείνας, σύγχυσης, ευερεθιστότητας, ταχυκαρδίας και κεφαλγίας, μεταξύ άλλων. Σε σοβαρές περιπτώσεις υπογλυκαιμίας ο ασθενής χρήζει βοήθειας από κάποιο άλλο άτομο για την αντιμετώπισή της, ενώ στις κρισιμότερες επιπλοκές συγκαταλέγονται οι καρδιακές αρρυθμίες, οι σπασμοί, το κώμα και σπανιότερα, ο θάνατος. Σημειώνουμε ότι είναι πιθανή και η εκδήλωση ασυμπτωματικών υπογλυκαιμικών επεισοδίων, ένα φαινόμενο που περιγράφεται με τον όρο *ανεπίγνωστη υπογλυκαιμία*. Τα κυριότερα αίτια εμφάνισης υπογλυκαιμίας είναι η λήψη μεγαλύτερης δόσης αντιδιαβητικών φαρμάκων από την συνιστώμενη, η παράλειψη γευμάτων, η μη προγραμματισμένη φυσική άσκηση και η κατανάλωση αλκοόλ επί νηστείας. Υπογλυκαιμικά περιστατικά αντιμετωπίζονται μέσω της χορήγησης γλυκόζης είτε άμεσα, είτε έμμεσα, μέσω πρόσληψης τροφής. Για την αποτροπή τους είναι κρίσιμης σημασίας η συνεχής παρακολούθηση και η συνεπής ρύθμιση των επιπέδων γλυκόζης του αίματος. [7]

### **2.3.3 Μακροπρόθεσμες επιπλοκές**

Οι μακροπρόθεσμες επιπλοκές αποτελούν την κύρια αιτία θνησιμότητας ατόμων με σακχαρώδη διαβήτη, αν και στις περισσότερες περιπτώσεις αυτές μπορούν να αποτραπούν ή και να επιβραδυνθούν μέσω της συνεπούς διαχείρισης του και της υιοθέτησης



ενός υγιεινού τρόπου ζωής. Παρακάτω επικεντρωνόμαστε στις επιπλοκές που σχετίζονται με παθήσεις της καρδιάς και των μεγάλων αγγείων, παθήσεις των νεφρών, των οφθαλμών και των περιφερειακών άκρων. Αξίζει όμως να επισημανθεί ότι ο σακχαρώδης διαβήτης μπορεί να συμβάλει ακόμα στην εκδήλωση πεπτικών, ουρολογικών, σεξουαλικών αλλά και ψυχικών διαταραχών, μεταξύ άλλων. [7], [9]

### **Καρδιαγγειακές νόσοι**

Παθήσεις της καρδιάς και των μεγάλων αγγείων όπως η στεφανιαία νόσος, το έμφραγμα του μυοκαρδίου και το αγγειακό εγκεφαλικό επεισόδιο είναι συχνότερες σε διαβητικούς ασθενείς και αποτελούν τη βασικότερη αιτία θνησιμότητας ατόμων με ΣΔ τύπου 2. Πράγματι, το 10%-11% όλων των θανάτων από καρδιαγγειακές ασθένειες οφείλονται σε άτομα με ΣΔ τύπου 2, ενώ περίπου το 70% των ατόμων αυτών με ηλικία μεγαλύτερη των 65 ετών πεθαίνει από καρδιαγγειακή νόσο. Μάλιστα, οι ασθενείς με ΣΔ τύπου 2 χωρίς ιστορικό στεφανιαίας νόσου έχουν ίσο καρδιαγγειακό κίνδυνο με τους ασθενείς με προηγούμενο έμφραγμα του μυοκαρδίου, ενώ γενικότερα οι διαβητικοί έχουν διπλή έως τετραπλάσια πιθανότητα να εμφανίσουν κάποια καρδιαγγειακή πάθηση. Ως αποτέλεσμα των παραπάνω, ο σακχαρώδης διαβήτης συγκαταλέγεται ανάμεσα στους υπόλοιπους παράγοντες κινδύνου εμφάνισης παθήσεων σχετιζόμενων με την καρδιά και τα μεγάλα αγγεία. Όλοι οι ασθενείς που πάσχουν από διαβήτη πρέπει να ελέγχονται με καρδιογράφημα ετησίως, ενώ αν έχουν συμπτώματα που υποδηλώνουν καρδιακές παθήσεις πρέπει να υποβάλλονται σε ειδικές εξετάσεις και να λαμβάνουν τακτική καρδιολογική παρακολούθηση. [7], [9]

### **Διαβητική νεφρική νόσος**

Η διαβητική νεφρική νόσος είναι επίσης μια συχνά εμφανιζόμενη μακροπρόθεσμη επιπλοκή του σακχαρώδη διαβήτη. Συγκεκριμένα, περίπου οι μισοί των ασθενών με ΣΔ τύπου 2 και το ένα τρίτο των νοσούντων με ΣΔ τύπου 1, θα την εκδηλώσουν. Η εμφάνιση της οφείλεται κυρίως στη βαθμιαία φθορά που υφίστανται τα μικρά αιμοφόρα αγγεία των νεφρών, λόγω των παρατεταμένα υψηλών επιπέδων γλυκόζης του αίματος. Αυτή η σταδιακή βλάβη των νεφρικών αγγείων επιδρά αρνητικά στις διηθητικές τους ικανότητες, με αποτέλεσμα ένα από τα πρώτα συμπτώματα της διαβητικής νεφρικής νόσου να είναι η αποβολή μεγάλων ποσοτήτων της πρωτεΐνης αλβουμίνης από το αίμα στα ούρα. Άλλα συμπτώματα μπορεί να περιλαμβάνουν αύξηση του βάρους, οίδημα στα άκρα, συχνουρία, αυξημένη αρτηριακή πίεση, κόπωση και ναυτία. Τέλος, στην περίπτωση που δεν αντιμετωπιστεί, η νεφρική νόσος μπορεί να οδηγήσει σε νεφρική ανεπάρκεια. Παρόλα αυτά, η διαβητική νεφρική νόσος μπορεί να αποτραπεί ή και να επιβραδυνθεί, μέσω συνεπούς διαχείρισης του σακχαρώδη διαβήτη και υιοθέτησης υγιεινών συνηθειών. Εξίσου σημαντικός είναι ο τακτικός έλεγχος της νεφρικής λειτουργίας στα πλαίσια της παρακολούθησης του διαβήτη, με τη μορφή απλών αιματολογικών εξετάσεων και συλλογής ούρων. [7], [9]

### **Παθήσεις των οφθαλμών**

Παθήσεις των ματιών όπως η θολή όραση, η αμφιβληστροειδοπάθεια, ο καταρράκτης και το γλαύκωμα συχνά εκδηλώνονται κατόπιν της παρατεταμένης έκθεσης των αιμοφόρων αγγείων των ματιών σε υψηλές τιμές γλυκόζης. Πράγματι, η διάγνωση με ΣΔ αυξάνει τη πιθανότητα τύφλωσης κατά 25 φορές, ενώ ο διαβήτης αποτελεί την πρωταρχι-

κή αιτία τύφλωσης ενηλίκων 20 έως 74 ετών. Η προοδευτική αλλοίωσή των αγγείων των οφθαλμών συνήθως αρχίζει πριν από τη διάγνωση με ΣΔ, όταν οι τιμές γλυκόζης είναι αυξημένες, αλλά όχι αρκετά υψηλές για την κατάταξη του ατόμου ως διαβητικού. Αυτή η κλιμακωτή φθορά μπορεί να οδηγήσει στη διαρροή υγρών από τα αγγεία, στην εκδήλωση οιδημάτων και στην ανάπτυξη νέων, βλαβερών αγγείων στην περιοχή των οφθαλμών, με αποτέλεσμα σοβαρά προβλήματα όρασης. Για την αποτροπή τέτοιων χρόνιων επιπλοκών είναι σημαντική η αποτελεσματική αντιμετώπιση του ΣΔ, η αποχή των νοσούντων από το κάπνισμα, καθώς και η υιοθέτηση ενός ισορροπημένου τρόπου ζωής. Τέλος, οι ασθενείς που πάσχουν από σακχαρώδη διαβήτη πρέπει να υποβάλλονται σε ετήσιο προληπτικό οφθαλμολογικό έλεγχο, ενώ σε περίπτωση που εμφανίσουν συμπτώματα, είναι απαραίτητη η πιο τακτική οφθαλμολογική παρακολούθησή τους. [7], [9]

### **Περιφερειακή νευροπάθεια**

Η περιφερειακή νευροπάθεια είναι μια σοβαρή μακροχρόνια επιπλοκή του σακχαρώδους διαβήτη που προσβάλλει περίπου το 50% των ασθενών. Αν και ο ΣΔ επηρεάζει τη νευρική λειτουργία στο σύνολό της, οι συχνότερες νευροπάθειες αφορούν τα άνω και τα κάτω άκρα. Συχνές ενδείξεις περιφερειακής νευροπάθειας αποτελούν ο καυστικός πόνος στα πόδια (κυρίως την νύχτα), η μειωμένη αισθητικότητα, το μούδιασμα και το μυρμήγκιασμα των άκρων. Η περιφερειακή απώλεια αισθήσεων μπορεί επίσης να οδηγήσει στην αγνόηση και την ελλιπή φροντίδα μικροτραυματισμών, με αποτέλεσμα την εμφάνιση λοιμώξεων και ελκών που μπορεί να επιβάλλουν ακόμα και αναγκαστικό ακρωτηριασμό. Παρόλα αυτά, μια σημαντική μερίδα ασθενών με χρόνια περιφερειακή νευροπάθεια είναι ασυμπτωματικοί. Οι σοβαρές επιπλοκές που προαναφέρονται, καθώς και το ενδεχόμενο ασυμπτωματικών περιπτώσεων, καθιστούν τον ενδελεχή, τακτικό νευρολογικό έλεγχο των άκρων απαραίτητο στα πλαίσια διαχείρισης του ΣΔ. Τέλος, όπως και στις παραπάνω περιπτώσεις μακροπρόθεσμων επιπλοκών, η επιμελής παρακολούθηση και ρύθμιση των επιπέδων γλυκόζης του αίματος συνεισφέρουν στην επιβράδυνση εξέλιξης της περιφερειακής νευροπάθειας ή και στην αποτροπή της. [7], [9]

## **2.4 Διάγνωση και αντιμετώπιση**

### **2.4.1 Διάγνωση**

Η εκδήλωση συμπτωμάτων ή και ορισμένων επιπλοκών εξ αυτών που αναφέρονται παραπάνω, συχνά συνιστά την αφορμή διάγνωσης ενός ατόμου με σακχαρώδη διαβήτη. Παρόλα αυτά, είναι πάντοτε απαραίτητη η επαλήθευση της νόησης με κάποιου είδους διαγνωστικού τεστ. Οι κλινικοί διαγνωστικοί αυτοί έλεγχοι ποικίλλουν και πρέπει να διεξάγονται πάντοτε από κάποιον ειδικό. Στη συνέχεια αναφέρονται οι κυριότεροι έλεγχοι που εφαρμόζονται στην πράξη για την διάγνωση ενός ανθρώπου με κάποιο τύπο ΣΔ. Επισημαίνουμε ότι στις περισσότερες περιπτώσεις, για την οριστική διάγνωση, είναι απαραίτητη η υποβολή του πιθανού ασθενή σε πάνω από έναν από τους παρακάτω ελέγχους, με την απόκλιση κάποιων ημερών.

### **Μέτρηση γλυκοζυλιωμένης αιμοσφαιρίνης (Hb1C έλεγχος)**

Το λεγόμενο HbA1C τεστ ελέγχει το ποσοστό γλυκοζυλιωμένης αιμοσφαιρίνης που εντοπίζεται στο φλεβικό αίμα. Σημειώνεται ότι η πρωτεΐνη αιμοσφαιρίνη βρίσκεται στα

ερυθρά αιμοσφαίρια και είναι υπεύθυνη για την μεταφορά οξυγόνου στους ιστούς, ενώ η γλυκοζυλιωμένη αιμοσφαιρίνη (ή HbA1c) προκύπτει από τη χημική ένωση της αιμοσφαιρίνης με τη γλυκόζη. Το ποσοστό HbA1c επηρεάζεται από τις τιμές σακχάρου όλου του 24ώρου των τελευταίων 120 ημερών, επομένως μπορεί να χρησιμοποιηθεί για τον καθορισμό της μέσης τιμής γλυκόζης αίματος, τους τελευταίους 3 μήνες περίπου. Ποσοστό γλυκοζυλιωμένης γλυκόζης μεγαλύτερο από ή ίσο με 6.5% μεταφράζεται σε διάγνωση με ΣΔ, ενώ αντίστοιχα τιμές 5.7%-6.4% σηματοδοτούν την ύπαρξη προδιαβήτη, διαταραχής που προηγείται του διαβήτη και αυξάνει τις πιθανότητες εμφάνισής του. [10]

### **Μέτρηση γλυκόζης πλάσματος νηστείας (FPG έλεγχος)**

Ο FPG έλεγχος συνίσταται στη μέτρηση της γλυκόζης πλάσματος έπειτα από νηστεία τουλάχιστον 8 ωρών. Αποτελεί έναν απλό, οικονομικό και ακριβή τρόπο διάγνωσης της μεταβολικής ανισορροπίας που σχετίζεται με το σακχαρώδη διαβήτη. Τιμές γλυκόζης μικρότερες από 100mg/dl αποτελούν τις φυσιολογικές τιμές, τιμές 100mg/dl ως και 125mg/dl υποδηλώνουν την νόσηση με προδιαβήτη, ενώ τέλος, μετρήσεις γλυκόζης μεγαλύτερες από 126mg/dl σηματοδοτούν την ύπαρξη σακχαρώδους διαβήτη. [10]

### **Έλεγχος ανοχής γλυκόζης (OGTT έλεγχος)**

Το συγκεκριμένο τεστ συνίσταται στην μέτρηση της γλυκόζης πλάσματος αμέσως πριν και δύο ώρες έπειτα από την κατανάλωση ενός διαλύματος γλυκόζης. Στόχος είναι ο έλεγχος της ανοχής του οργανισμού του υποβληθέντος ατόμου στη γλυκόζη. Όπως και στην περίπτωση του FPG τεστ, ο ασθενής πρέπει να υποβληθεί στον έλεγχο αυτό έπειτα από 8 ώρες νηστείας. Με βάση την Αμερικανική Ένωση Διαβήτη, έπειτα από τις δύο ώρες, γλυκόζη πλάσματος μικρότερη από 140mg/dl αντιστοιχεί σε φυσιολογικό αποτέλεσμα, τιμές από 140mg/dl έως 199mg/dl αποτελούν ένδειξη προδιαβήτη, και μέτρηση υψηλότερη από 200mg/dl υποδηλώνει νόσηση με ΣΔ. [10]

## **2.4.2 Αντιμετώπιση**

Η αντιμετώπιση του σακχαρώδους διαβήτη αποτελεί μια περίπλοκη και χρόνια διαδικασία, η οποία απαιτεί τη στενή παρακολούθηση της προόδου της ασθένειας, τόσο από τον ίδιο τον ασθενή, όσο και από κάποιον ειδικό. Όπως έχει ήδη αναφερθεί, η εξέλιξη της νόσου επηρεάζεται καθημερινά από μια ποικιλία παραγόντων, όπου μεταξύ άλλων περιλαμβάνονται ο τρόπος ζωής και η διατροφή που ακολουθεί ο ασθενής, η χορήγηση ινσουλίνης ή άλλων αντιδιαβητικών φαρμάκων, η παρουσία ή μη ψυχολογικού στρες, η ύπαρξη άλλων υποκείμενων νοσημάτων, καθώς και περιβαλλοντικές συνθήκες όπως η θερμοκρασία. Παρόλα αυτά, η συνεπής και ενημερωμένη διαχείριση του ΣΔ μπορεί να συμβάλει στη διατήρηση μιας ικανοποιητικής ποιότητας ζωής κι ακόμα να αποτρέψει την εμφάνιση οξείων και μακροχρόνιων επιπλοκών της ασθένειας. Παρακάτω παρουσιάζονται οι σημαντικότερες όψεις στην αντιμετώπιση του διαβήτη.

### **Παρακολούθηση γλυκόζης αίματος**

Η μέτρηση της γλυκόζης του αίματος ανά τακτικά χρονικά διαστήματα αποτελεί αναγκαιότητα και αναπόσπαστο τμήμα της καθημερινότητας ενός ατόμου με ΣΔ. Η διαδικασία αυτή φέρει πολλαπλά οφέλη, καθώς επιτρέπει την ορθότερη ρύθμιση των γευμάτων

και της φαρμακευτικής αγωγής του ασθενούς, ενώ ακόμα συνεισφέρει στην αποτροπή σοβαρών υπεργλυκαιμικών ή υπογλυκαιμικών επεισοδίων. Η καταγραφή και στη συνέχεια, η μελέτη των μετρήσεων γλυκόζης αίματος σε βάθος χρόνου παρέχουν πολύτιμες πληροφορίες για την εξέλιξη της ασθένειας και επιτρέπουν τη λήψη αποφάσεων σχετικά με απαιτούμενες αλλαγές στο τρόπο ζωής του ασθενούς, από το μέρος του συνεργαζόμενου γιατρού.

Ο συμβατικός τρόπος μέτρησης της γλυκόζης περιλαμβάνει τη λήψη μιας μικρής ποσότητας αίματος (1-2 σταγόνες), συνήθως από τα δάχτυλα του ασθενούς, κάποιες φορές την ημέρα. Η λήψη αίματος και η μέτρηση πραγματοποιούνται με χρήση ειδικών συσκευών και μπορούν να γίνουν από τον ίδιο τον ασθενή. Η συχνότητα μέτρησης μπορεί να διαφέρει ανάλογα με τον τύπο ΣΔ που αντιμετωπίζεται, τις ιδιομορφίες του εκάστοτε ατόμου, καθώς και το είδος φαρμακευτικής αγωγής που λαμβάνεται. Για παράδειγμα, στην περίπτωση του ΣΔ τύπου 1, μπορεί να απαιτείται μέτρηση 4-10 φορές την ημέρα, ενώ σε κάθε περίπτωση είναι σημαντική η λήψη μετρήσεων πριν από κάθε γεύμα και πριν την ώρα του ύπνου. Για ένα άτομο διαγνωσμένο με ΣΔ, τα επίπεδα γλυκόζης φυσιολογικά αναμένεται να είναι κάτω από 180mg/dl δύο ώρες μετά από ένα γεύμα και στο διάστημα 80mg/dl - 130mg/dl κατά διάστημα νηστείας. [11], [12]

Τα τελευταία χρόνια, νέες συσκευές μέτρησης που επιτρέπουν την αυτόματη καταγραφή γλυκόζης (Continuous Glucose Monitoring - CGM) έχουν γίνει εμπορικά διαθέσιμες, βελτιώνοντας την ποιότητα γλυκαιμικού ελέγχου, σύμφωνα με πρόσφατες, σχετικές έρευνες. [13] Η μέτρηση πραγματοποιείται με τη χρήση αισθητήρων που εφαρμόζονται κυρίως στον βραχίονα ή την κοιλιακή χώρα. Οι αισθητήρες αυτοί μετρούν τη γλυκόζη στο μεσοκυττάριο υγρό, δηλαδή τη γλυκόζη στο υγρό μεταξύ των κυττάρων, ανά μερικά λεπτά. Οι τιμές αυτές, αν και μπορούν να διαφέρουν ελαφρώς από τις τιμές γλυκόζης του πλάσματος του αίματος, είναι αντιπροσωπευτικές αυτών και των διακυμάνσεών τους. Οι αισθητήρες πρέπει να αντικαθίστανται ανά μερικές ημέρες (ανάλογα με το μοντέλο), είναι κατά κύριο λόγο αδιάβροχοι και οι περισσότεροι εξ αυτών δεν απαιτούν βαθμονόμηση. Η ακρίβεια των παραπάνω CGM συσκευών αξιολογείται κυρίως με την μετρική Mean Absolute Relative Difference (MARD), με τα εξαρτήματα που εμφανίζουν MARD τιμή κοντά στο 10% να θεωρούνται αξιόπιστα και κατάλληλα για χρήση. [13], [14]

Ένα από τα μεγαλύτερα πλεονεκτήματα των CGM τεχνολογιών είναι η δυνατότητά τους να παρακολουθούν τη γλυκόζη των χρηστών καθόλη τη διάρκεια της νύχτας, κάτι που δεν είναι δυνατό μέσω συμβατικών μετρήσεων. Ακόμα, οι περισσότεροι αισθητήρες διαθέτουν την ικανότητα ακουστικής ειδοποίησης των ασθενών σε περίπτωση που τα επίπεδα γλυκόζης τους ξεπεράσουν κάποιο υπερβολικά υψηλό ή χαμηλό κατώφλι. Τα όρια αυτά μπορούν να καθοριστούν από τον χρήστη με βάση τις οδηγίες του επιβλέποντα γιατρού, με στόχο τη μεγαλύτερη εξατομίκευση της θεραπευτικής διαδικασίας. Επιπλέον, εφόσον γίνει αποθήκευση των μετρήσεων για ένα ικανό χρονικό διάστημα, είναι δυνατή η μελέτη των δεδομένων από κάποιο ειδικό για την εξαγωγή χρήσιμων, γενικών συμπερασμάτων, αλλά και η χρήση των πληροφοριών αυτών για την ανάπτυξη κάποιου συστήματος πρόβλεψης σαν και αυτό που μελετάται στην παρούσα εργασία. [13], [14] Σημειώνουμε τέλος, ότι ανάλογα με το εάν οι καταγραφόμενες τιμές αναγράφονται σε πραγματικό χρόνο στην ειδική συσκευή που χρησιμοποιείται για την παρουσίαση των αποτελεσμάτων (ή εναλλακτικά στο κινητό του χρήστη), ή μόνο κατ'επίκληση του ασθενούς, γίνεται διαχωρισμός των CGM αισθητήρων σε rtCGM (real time CGM) και isCGM (intermittently scanned CGM), αντίστοιχα. Παραδείγματα rtCGM μοντέλων είναι οι Dexcom G6, Dexcom G5, Medtrum Touch Care Nano και Medtronic Guardian 3/4 αισθητήρες, ενώ τα μοντέλα FreeStyle Libre, FreeStyle Libre 2 αποτελούν ενδεικτικές

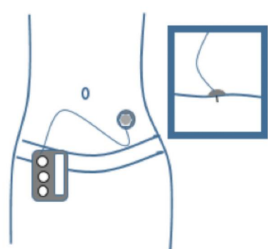
isCGM συσκευές. Τα isCGM μοντέλα δεν έχουν την δυνατότητα επίδειξης των real-time δεδομένων, εκτός και αν γίνει σάρωση του αισθητήρα από τον χρήστη, συνήθως με τη χρήση κινητού τηλεφώνου. [15]

Τα άτομα που επωφελούνται κυρίως από την εφαρμογή των CGM τεχνολογιών είναι αυτά που εμφανίζουν συχνές υπογλυκαιμίες (ειδικά όταν τα περιστατικά αυτά αντιστοιχούν σε ανεπίγνωστη υπογλυκαιμία) και μεγάλες διακυμάνσεις στα επίπεδα γλυκόζης. Μειονεκτήματα των continuous glucose monitoring συσκευών είναι η ανάγκη συνεχούς εφαρμογής αισθητήρα και εξοικείωσης του ασθενούς με τη μέθοδο, καθώς και η επιπρόσθετη οικονομική επιβάρυνσή του. [13], [14], [15]

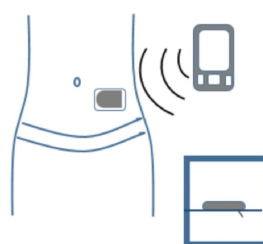
### **Χορήγηση ινσουλίνης και αντιδιαβητικών φαρμάκων**

Η εξωτερική χορήγηση ινσουλίνης αποτελεί μια ακόμα καίρια όψη της αντιμετώπισης του σακχαρώδους διαβήτη. Η εξωτερική παροχή ινσουλίνης κρίνεται εξαρχής απαραίτητη στις περιπτώσεις ΣΔ τύπου 1, όπου παρατηρείται ολική έλλειψη της. Αντίθετα, στις περιπτώσεις ΣΔ τύπου 2, συνήθως στην αρχή χορηγούνται αντιδιαβητικά φάρμακα (όπως στοματικά χορηγούμενα δισκία), ενώ ινσουλίνη παρέχεται αφού η φαρμακευτική αγωγή παύει να επαρκεί, και η ποσότητα παραγόμενης από τον ίδιο τον οργανισμό αρχίζει να στερεύει. Μια μέθοδος χορήγησης ινσουλίνης είναι η θεραπεία πολλαπλών καθημερινών ενέσεων (Multiple Daily Injection - MDI). Με βάση αυτήν την αντιμετώπιση, ινσουλίνη χορηγείται υποδόρια στην κοιλιά, τα πόδια ή τα χέρια, ορισμένες φορές την μέρα, μέσω σύριγγας ή πέννας ινσουλίνης. Οι βελόνες που προσαρμόζονται στην πένα συγκεκριμένα είναι πολύ μικρές και πρακτικά η ένεση είναι σχεδόν ανώδυνη. Όσον αφορά το ακολουθούμενο θεραπευτικό σχήμα, αυτό αποφασίζεται από τον γιατρό, με βάση την περίπτωση σακχαρώδους διαβήτη και το καθημερινό πρόγραμμα του ασθενή. Συνήθως, λαμβάνεται ινσουλίνη μακράς διάρκειας δράσης πριν από την ώρα του ύπνου και ινσουλίνη ταχείας δράσης πριν από κάθε γεύμα. [16]

Εναλλακτικά, είναι δυνατή η χρήση αντλίας έγχυσης γλυκόζης (insulin pump), με την οποία γίνεται σταδιακή, συνεχής χορήγηση ινσουλίνης ταχείας δράσης, σε μικρές δόσεις, καθόλη τη διάρκεια των 24 ωρών. Μέσω αυτής της σταθερής, σταδιακής έγχυσης, στόχος είναι η προσομοίωση της λειτουργίας του παγκρέατος ενός υγιούς ατόμου, ενώ ο ρυθμός έγχυσης (basal rate) μπορεί να προγραμματιστεί από τον χρήστη. Ένα σύστημα ανλίας έγχυσης αποτελείται από έναν μικρό, υποδόριο καθετήρα που συνδέεται με την ανλία ινσουλίνης μέσω ενός σωληναρίου. Διαφορετικά, είναι δυνατή η χρήση των λεγόμενων "patch pumps", όπου δεν χρησιμοποιείται σωληνάριο, αλλά ο καθετήρας και η δεξαμενή ινσουλίνης είναι προσκολλημένα μαζί πάνω στο δέρμα. Ο προγραμματισμός τους γίνεται με μια απομακρυσμένη συσκευή ασύρματα.



(α) Αντλία έγχυσης με σωληνάριο



(β) Αντλία έγχυσης "patch pump"

Σχήμα 2.3: Συστήματα αντλιών έγχυσης ινσουλίνης [17]

Σε περίπτωση που απαιτείται λόγω επακόλουθου γεύματος ή υψηλής καταγεγραμμένης τιμής γλυκόζης, είναι δυνατός ο προγραμματισμός συμπληρωματικών, μεμονωμένων δόσεων ινσουλίνης (insulin boluses). Οι περισσότερες αντλίες διαθέτουν ενσωματωμένη τη δυνατότητα υπολογισμού της επιπρόσθετης, απαιτούμενης δόσης, με βάση την τρέχουσα τιμή γλυκόζης, τον εισαγόμενο από τον χρήστη αριθμό υδατανθράκων που πρόκειται να καταναλωθεί και την παραμένουσα ινσουλίνη από προηγούμενες δόσεις ("insulin on board"). Η υπολογιζόμενη αυτή τιμή προτείνεται στον ασθενή, ο οποίος έχει τη δυνατότητα να εισάγει την ακριβή αυτή δόση ή να την προσαρμόσει. Επιπρόσθετα, σε περίπτωση που χρησιμοποιείται CGM συσκευή καταγραφής γλυκόζης είναι δυνατή η σύζευξη της αντλίας ινσουλίνης και της συσκευής καταγραφής. Σε ορισμένα μοντέλα μάλιστα (π.χ. Minimed 530G) η έγχυση της ινσουλίνης διακόπτεται σε περίπτωση που τα επίπεδα γλυκόζης λαβουν ή τείνουν να λάβουν υπογλυκαιμικές τιμές. Σε νεότερα προϊόντα επιχειρείται η αυτόματη χορήγηση ινσουλίνης με βάση τα προκύπτοντα CGM δεδομένα, με στόχο την κατασκευή ενός "τεχνητού παγκρέατος". Ένα σύστημα αυτόματης χορήγησης ινσουλίνης αποτελείται από μια αντλία έγχυσης, ένα σύστημα CGM καταγραφής και έναν αλγόριθμο ελέγχου, που προσαρμόζει τον ρυθμό έγχυσης σε πραγματικό χρόνο, ανάλογα με τις CGM τιμές και διακυμάνσεις. Ορισμένα τέτοια συστήματα (π.χ. Minimed 670G), όπου γίνεται αυτόματη προσαρμογή του basal rate, έχουν εγκριθεί από τον Αμερικανικό Οργανισμό Τροφίμων και Φαρμάκων (Food and Drug Administration - FDA). Παρόλα αυτά, στα περισσότερα από αυτά, προς το παρόν, απαιτείται η προσαρμογή των μεμονωμένων δόσεων από τους ίδιους τους ασθενείς. Όσον αφορά πιθανά μειονεκτήματα των αντλιών ινσουλίνης, σε αυτά συγκαταλέγονται η ανάγκη συνεχούς εφαρμογής του καθετήρα και η ανάγκη εκπαίδευσης του ασθενούς, έτσι ώστε αυτός να κάνει ορθή χρήση της συσκευής και να μπορεί να την ρυθμίσει αποτελεσματικά, ανάλογα με τις ανάγκες του. [17]

### **Διατροφή και σωματική δραστηριότητα**

Η διατροφή και η σωματική δραστηριότητα αποτελούν ακρογωνιαίοι λίθοι στη ρύθμιση της ζωής με ΣΔ. Συγκεκριμένα, η διατροφή επιτρέπει τη σωστή προσαρμογή της φαρμακευτικής αγωγής, με στόχο την επίτευξη καλύτερων επιπέδων μεταγευματικού σακχάρου, κι επομένως, καλύτερου γλυκαιμικού ελέγχου. Σύμφωνα με τις σύγχρονες συστάσεις, άτομα με διαβήτη μπορούν να καταναλώσουν μεγάλο εύρος τροφών μέσα στο πλαίσιο μιας υγιεινής διατροφής, η οποία άλλωστε θα πρέπει να ακολουθείται και από το γενικό πληθυσμό. Με βάση τις συστάσεις της Ελληνικής Διαβητολογικής Εταιρείας, η διατροφή που προτείνεται σε άτομα με ΣΔ περιλαμβάνει καθημερινή κατανάλωση φρούτων και λαχανικών, συχνή κατανάλωση οσπρίων και δημητριακών και εβδομαδιαία κατανάλωση ψαριού. Η κατανάλωση κόκκινου κρέατος καλό είναι να περιορίζεται σε μία φορά την εβδομάδα. Ακόμα, σημαντική είναι η αποφυγή των απλών υδατανθράκων, (π.χ. ζάχαρη, λευκό ψωμί, μακαρόνια κ.α.) και η αντικατάστασή τους με υδατάνθρακες με υψηλή περιεκτικότητα σε ίνες. Τα προϊόντα αυτά επιβραδύνουν τη διαδικασία της πέψης κι ενισχύουν το αίσθημα του κορεσμού. Τέλος, σημειώνουμε ότι για τον ακριβέστερο υπολογισμό της δόσης ινσουλίνης ή της αντιδιαβητικής αγωγής που θα πρέπει να ληφθεί για την κάλυψη της καταναλωμένης ποσότητας τροφής, συνήθως ζητείται από τους ασθενείς να μετρήσουν τους ισοδύναμους υδατάνθρακες του κάθε γεύματος, πριν την πρόσληψή του. [18]

Η σωματική δραστηριότητα επίσης φέρει αξιοσημείωτα οφέλη για τη ρύθμιση του ΣΔ. Συγκεκριμένα, η απώλεια βάρους στην περίπτωση υπέρβαρων ή παχύσαρκων ατόμων

με ΣΔ αποδεδειγμένα συνδράμει στην πολύ αποτελεσματικότερη αντιμετώπιση της ασθένειας, και ακόμα και στην αποτροπή της, στην περίπτωση του ΣΔ τύπου 2. Αναλυτικότερα, σε συνδυασμό με μια ισορροπημένη διατροφή, συνιστάται μέτριας ή έντονης έντασης αερόβια άσκηση διάρκειας τουλάχιστον 30 λεπτών την ημέρα (συνεχώς ή διακεκομμένα), το λιγότερο 5 φορές την εβδομάδα. Τα άτομα με ΣΔ πρέπει να αποφεύγουν την καθιστική ζωή, και όταν είναι αναγκαία λόγω επαγγελματικών υποχρεώσεων, να κάνουν τακτικά διαλείμματα. Αξίζει να σημειωθεί βέβαια ότι η άσκηση πρέπει να γίνεται με προσοχή, ειδικά στην περίπτωση ατόμων με ΣΔ τύπου 1, και να αποφεύγεται όταν η γλυκόζη είναι υπερβολικά υψηλή ή χαμηλή. Ακόμα, εάν η άσκηση είναι προγραμματισμένη, πρέπει να γίνεται μείωση της δόσης της ινσουλίνης που προηγείται της άσκησης, ενώ αν είναι παρατεταμένη, συνιστάται λήψη υδατανθράκων και κατά τη διάρκεια της άσκησης. [19]

## Κεφάλαιο 3

# Τεχνητά Νευρωνικά Δίκτυα

### 3.1 Παραλληλισμός με βιολογικά συστήματα

Ο βιολογικός εγκέφαλος αποτελεί μία περίπλοκη, μη-γραμμική υπολογιστική μονάδα παράλληλης επεξεργασίας. Τα επιμέρους δομικά στοιχεία του (*νευρικά κύτταρα* ή *νευρώνες*) είναι έτσι οργανωμένα ώστε να πραγματοποιούν συγκεκριμένους υπολογισμούς (π.χ. αναγνώριση προτύπων, έλεγχο κινήσεων) με εξαιρετικά υψηλή ταχύτητα. Ενδεικτικά αναφέρουμε ότι η αναγνώριση ενός γνωστού προσώπου από τον ανθρώπινο εγκέφαλο, σε ένα κατά τα άλλα μη γνώριμο περιβάλλον, απαιτεί μόλις 100-200ms, ενώ άλλες απλούστερες διαδικασίες συντελούνται ακόμα ταχύτερα. Οι εντυπωσιακές αυτές επιδόσεις οφείλονται ως ένα βαθμό στη δομική πολυπλοκότητά του εγκεφάλου· υπολογίζεται ότι αυτός διαθέτει περίπου 10 δισεκατομμύρια νευρώνες, οι οποίοι με τη σειρά τους συνδέονται με περίπου 60 τρισεκατομμύρια διασυνδέσεις (ή *συνάψεις*). Παράλληλα, υπάρχει δυνατότητα δημιουργίας νέων συνάψεων, ή μεταβολής ήδη υπάρχουσών, λόγω του σημαντικού χαρακτηριστικού της πλαστικότητας. Η ιδιότητα αυτή είναι που επιτρέπει στον εγκέφαλο να προσαρμόζεται στις συνθήκες του εξωτερικού περιβάλλοντος και να συνεχίζει την ανάπτυξή του.

Ο τρόπος λειτουργίας του εγκεφάλου αποτέλεσε σημαντική πηγή έμπνευσης για το σχεδιασμό των λεγόμενων τεχνητών νευρωνικών δικτύων (ή ΤΝΔ για συντομία), μιας σημαντικής υποκατηγορίας του πεδίου της μηχανικής μάθησης (Machine Learning ή ML). Τα δίκτυα αυτά συντίθενται από υπολογιστικές μονάδες μικρής κλίμακας, τους τεχνητούς νευρώνες. Στην πιο γενική του μορφή, ένα μοντέλο τεχνητών νευρωνικών δικτύων σχεδιάζεται έτσι ώστε να μιμείται τον τρόπο με τον οποίο ο εγκέφαλος εκτελεί μια συγκεκριμένη λειτουργία. Συνολικά, ένα τέτοιο σύστημα κατανεμημένης επεξεργασίας χαρακτηρίζεται από υψηλό παραλληλισμό και διαθέτει την εγγενή ικανότητα αποθήκευσης, και στη συνέχεια χρήσης, εμπειρικής γνώσης. Αναλυτικότερα, ένα ΤΝΔ προσομοιάζει τον τρόπο λειτουργίας ενός βιολογικού εγκεφάλου με τους εξής δύο τρόπους: και αυτό ανακτά γνώση από το περιβάλλον του μέσω μιας διαδικασίας εκμάθησης, ενώ η γνώση αυτή αποθηκεύεται στα βάρη που χαρακτηρίζουν τις ενώσεις μεταξύ των νευρώνων (ή αλλιώς στα συναπτικά βάρη). Η διαδικασία διαμόρφωσης των συναπτικών βαρών, ωστόσο αυτά να προσδώσουν το επιθυμητό αποτέλεσμα, ονομάζεται *εκπαίδευση* του ΤΝΔ και πραγματοποιείται με τη χρήση ενός αλγορίθμου εκπαίδευσης. Γενικά, τα τεχνητά νευρωνικά δίκτυα αντλούν την υπολογιστική τους ισχύ από την παράλληλη και κατανεμημένη λειτουργία τους, καθώς και από την ικανότητα γενίκευσης που διαθέτουν. Το χαρακτηριστικό αυτό επιτρέπει σε αυτά τα συστήματα να εξάγουν

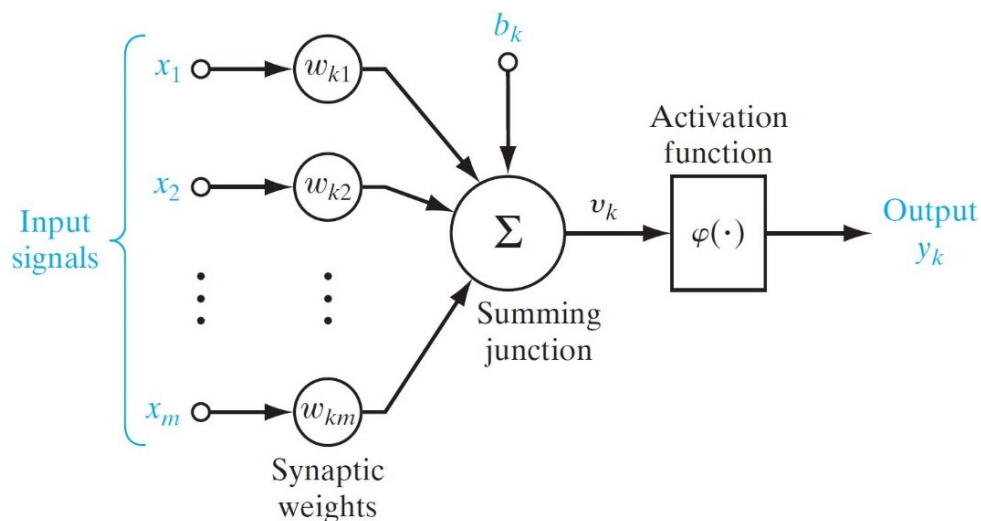


λογικά αποτελέσματα για δείγματα εισόδου τα οποία δεν έχουν συναντήσει κατά την εκπαίδευσή τους, κι επομένως τους επιτρέπει να παρέχουν αποδεκτές προσεγγιστικές λύσεις σε προβλήματα μεγαλύτερης κλίμακας. Έτσι διάφορες παραλλαγές μοντέλων ΤΝΔ χρησιμοποιούνται για: αναγνώριση ανθρωπίνων προσώπων, πρόβλεψη δεδομένων (π.χ. πρόβλεψη τιμών του χρηματιστηρίου ή του καιρού), διαχωρισμό των ανεπιθύμητων e-mails από τα κύρια, υποβοήθηση λήψης ιατρικών αποφάσεων, ανίχνευση ψευδών ειδήσεων σε μέσα κοινωνικής δικτύωσης, αυτοματοποιημένη ανακάλυψη νέων φαρμάκων. Τεχνητά νευρωνικά δίκτυα συναντώνται σε εφαρμογές των πεδίων της όρασης υπολογιστών, της επεξεργασίας φυσικής γλώσσας και της αναγνώρισης λόγου. Οι παραπάνω χρήσεις είναι ενδεικτικές· παρόλα αυτά ποικίλλουν και αποτυπώνουν με σαφήνεια την υψηλή αποτελεσματικότητα και τη μεγάλη εισχώρηση των παραπάνω τεχνολογιών στην ατομική και κοινωνική ζωή των ανθρώπων. [20], [21]

## 3.2 Νευρώνας (Perceptron)

### 3.2.1 Μοντέλο λειτουργίας

Ο νευρώνας αποτελεί το κύριο δομικό στοιχείο ενός τεχνητού νευρωνικού δικτύου, καθώς αποτελεί τη βάση σύνθεσης των πιο περίπλοκων και εκτεταμένων συστημάτων που μελετώνται στη συνέχεια. Παρακάτω φαίνεται το μοντέλο που περιγράφει ένα νευρώνα:



Σχήμα 3.1: Μοντέλο ενός νευρώνα  $k$  [20]

Ο παραπάνω νευρώνας  $k$  διαθέτει  $m$  διασυνδέσεις (ή συνάψεις), ενώ σε κάθε του διασύνδεση  $i$  προσδίδεται ένα βάρος  $w_{ki}$ . Κάθε σήμα εισόδου πολλαπλασιάζεται με το αντίστοιχο βάρος  $w_{ki}$ , ενώ στη συνέχεια, τα πολλαπλασιασμένα με τα βάρη σήματα εισόδου αθροίζονται, οπότε και προκύπτει το μέγεθος  $u_k$ :

$$u_k = \sum_{j=1}^m w_{kj} \cdot x_j \quad (3.1)$$

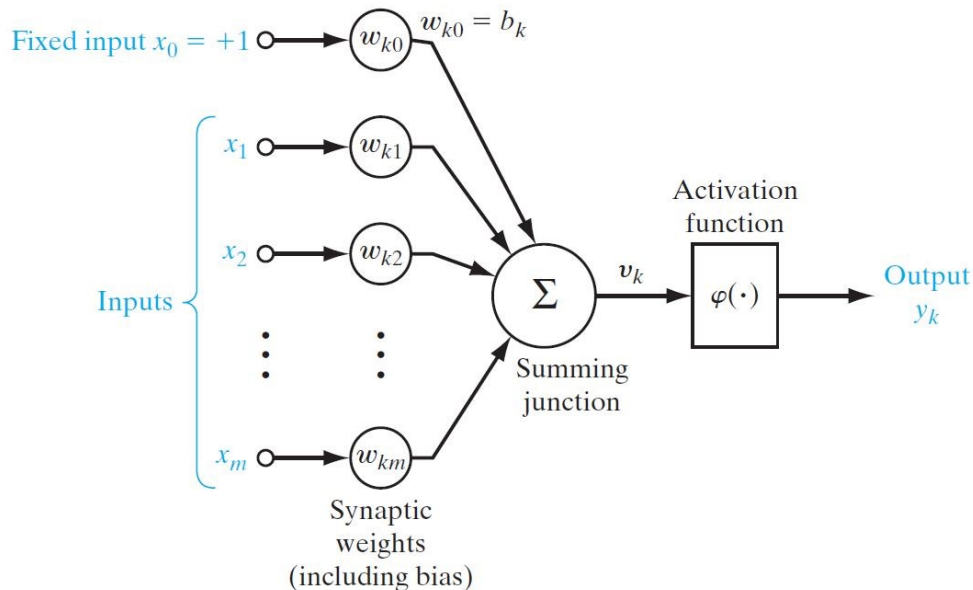
Το μοντέλο περιλαμβάνει ακόμα μια εξωτερικά εφαρμοζόμενη μεταβλητή, το bias ( $b_k$ ), με στόχο την αύξηση ή την μείωση της εξόδου  $u_k$ , ανάλογα με το αν αυτή είναι θετική ή

αρνητική αντίστοιχα. Τέλος, η λεγόμενη *συνάρτηση ενεργοποίησης*  $\phi(\cdot)$  (activation function) περιορίζει την έξοδο του νευρώνα, οπότε προκύπτει το τελικό αποτέλεσμα :

$$y_k = \phi(u_k + b_k) \quad (3.2)$$

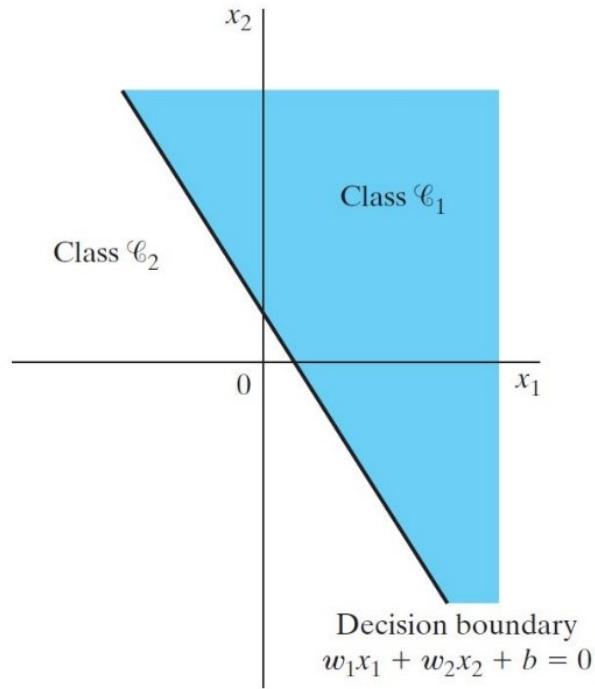
Θεωρώντας μια νέα σύναψη  $x_0 = +1$  και  $v_k = u_k + b_k$ , μπορούμε να γράψουμε:  $v_k = \sum_{j=0}^m w_{kj} \cdot x_j$  όπου  $w_{k0} = b_k$ . Τελικά προκύπτει:

$$y_k = \phi(v_k) \quad (3.3)$$



Σχήμα 3.2: Μοντέλο νευρώνα όπου  $w_{k0} = b_k$  [20]

Σημειώνουμε ότι τα ΤΝΔ τα οποία αποτελούνται μόνο από έναν νευρώνα ονομάζονται Perceptrons. Τα Perceptrons μπορούν να χρησιμοποιηθούν για την ταξινόμηση γραμμικά διαχωρίσιμων μοτίβων, σε δύο κλάσεις. Για δυαδική (binary) ταξινόμηση, εάν η ποσότητα  $v_k = \sum_{j=0}^m w_{kj} \cdot x_j$  είναι θετική, τότε η έξοδος του νευρώνα είναι +1, ενώ αν είναι αρνητική, τότε είναι -1. Έτσι, το σημείο που ορίζεται από τις εισόδους  $x_1, x_2, \dots, x_m$  κατατάσσεται στην κλάση  $C_1$  για έξοδο perceptron +1, και στην κλάση  $C_2$  για έξοδο perceptron -1. Στην απλούστερη μορφή ταξινόμησης, οι δύο περιοχές ταξινόμησης μπορούν να διαχωριστούν από ένα υπερεπίπεδο, το οποίο ορίζεται από τη σχέση:  $\sum_{i=0}^m w_i \cdot x_i + b = 0$ . Για μόνο δύο σήματα εισόδου  $x_1$  και  $x_2$ , το υπερεπίπεδο αυτό λαμβάνει τη μορφή μιας ευθείας ( $\epsilon$ ):  $w_1 \cdot x_1 + w_2 \cdot x_2 + b = 0$ . Επομένως, ένα σημείο  $(x_1, x_2)$  του επιπέδου κατατάσσεται στην κλάση  $C_1$ , εάν βρίσκεται πάνω από την ευθεία διαχωρισμού και στην κλάση  $C_2$ , εάν βρίσκεται κάτω από την ευθεία.



Σχήμα 3.3: Διαχωριστική ευθεία για την ταξινόμηση δισδιάστατου προβλήματος σε δύο κλάσεις  $C_1$  και  $C_2$ . [20]

Τα συναπτικά βάρη  $w_1, w_2, \dots, w_m$  ανανεώνονται μέσω μιας απλής επαναληπτικής διαδικασίας, ωστόσο ισχύει  $w^T \cdot x > 0$  για κάθε διάνυσμα εισόδου  $x \in C_1$  και  $w^T \cdot x < 0$  για κάθε διάνυσμα εισόδου  $x \in C_2$ . Ο χρησιμοποιούμενος κανόνας ανανέωσης βαρών είναι ο κάτωθι:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta \cdot (d(n) - y(n)) \cdot \mathbf{x}(n), \quad (3.4)$$

όπου  $n$  ο αριθμός της τρέχουσας επανάληψης,  $\eta \in (0, 1)$  ο ρυθμός εκπαίδευσης (learning rate),  $y(n) = \phi(w^T(n) \cdot x(n))$  η έξοδος του Perceptron για την επανάληψη  $n$  και  $d(n)$  η επιθυμητή έξοδος, δηλαδή:

$$d(n) = \begin{cases} +1, & x(n) \in C_1 \\ -1, & x(n) \in C_2 \end{cases} \quad (3.5)$$

Αποδεικνύεται ότι εφόσον οι κλάσεις  $C_1$  και  $C_2$  διαχωρίζονται από ένα υπερεπίπεδο (δηλαδή είναι γραμμικά διαχωρίσιμες), η παραπάνω διαδικασία συγκλίνει, οπότε μπορεί να υπολογιστεί το ζητούμενο διάνυσμα  $\mathbf{w}$  έπειτα από έναν πεπερασμένο αριθμό επαναλήψεων. Τέλος, στη συνέχεια παρατίθενται διάφορα είδη συχνά χρησιμοποιούμενων συναρτήσεων ενεργοποίησης. Η συνάρτηση ενεργοποίησης  $\phi(\cdot)$  είναι αυτή που καθορίζει την τελική έξοδο ενός νευρώνα.

### 3.2.2 Συναρτήσεις ενεργοποίησης

#### 1. Συνάρτηση κατωφλίου

$$\phi(v) = \begin{cases} 1, & v > 0 \\ 0, & v < 0 \end{cases} \quad (3.6)$$

Η συνάρτηση κατωφλίου (*threshold function*) αποτελεί την απλούστερη συνάρτηση ενεργοποίησης. Με βάση αυτήν, για αρνητική είσοδο, ο αντίστοιχος νευρώνας δεν ενεργοποιείται, ή ισοδύναμα, δίνει μηδενική έξοδο.

## 2. Γραμμική συνάρτηση

$$\phi(v) = v \quad (3.7)$$

Χρησιμοποιώντας τη γραμμική συνάρτηση ως συνάρτηση ενεργοποίησης, η έξοδος ενός νευρώνας είναι ανάλογη της εισόδου του. Όπως και η συνάρτηση κατωφλίου, η γραμμική συνάρτηση δεν χρησιμοποιείται συχνά, λόγω της απλοϊκότητάς της.

## 3. Σιγμοειδής (Sigmoid) συνάρτηση

$$\phi(v) = \frac{1}{1 + e^{-v}} \quad (3.8)$$

Η σιγμοειδής συνάρτηση αποτελεί μία από τις πιο κοινές συναρτήσεις ενεργοποίησης, λόγω της μη γραμμικότητας και παραγωγισιμότητάς της. Η έξοδος της σιγμοειδούς συνάρτησης περιορίζεται στο διάστημα  $(0, 1)$ , ενώ η ίδια διαθέτει χαρακτηριστικό σχήμα που μοιάζει με το λατινικό γράμμα  $S$ .

## 4. Συνάρτηση υπερβολικής εφαπτομένης $\tanh$

$$\phi(v) = \tanh(v) = \frac{2}{1 + e^{-2v}} - 1 \quad (3.9)$$

Η συνάρτηση υπερβολικής εφαπτομένης είναι επίσης συνεχής και παραγωγίσιμη, ενώ η έξοδος της κυμαίνεται στο διάστημα  $(-1, 1)$ . Λόγω της συμμετρικότητας της γύρω από την αρχή των αξόνων, συχνά προτιμάται σε σχέση με τη σιγμοειδή.

## 5. Rectified Linear Unit (ReLU) συνάρτηση

$$\phi(v) = \max(0, v) \quad (3.10)$$

Με βάση τη μη γραμμική συνάρτηση ReLU, ένας νευρώνας ενεργοποιείται μόνο όταν η είσοδος του είναι θετική, οπότε και η έξοδος του είναι γραμμική. Ένα από τα σημαντικότερα οφέλη της ReLU είναι η ετεροχρονισμένη ενεργοποίηση των νευρώνων ενός ΤΝΔ. Σημειώνουμε ότι μια βελτιωμένη παραλλαγή της ReLU, η λεγόμενη *Leaky ReLU*, ορίζεται ως  $0.01v$  για  $v \leq 0$  και  $v$  για  $v > 0$ . Με αυτήν την μικρή τροποποίηση, αποφεύγεται η νέκρωση υπερβολικά πολλών νευρώνων.

## 6. Softmax συνάρτηση

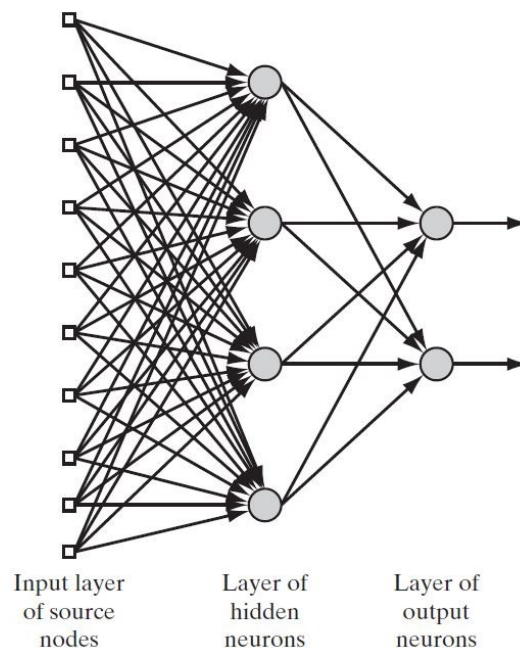
$$\phi(v)_j = \frac{e^{v_j}}{\sum_{k=1}^K e^{v_k}}, \quad (3.11)$$

όπου  $k$  το πλήθος των κλάσεων προς ταξινόμηση και  $j$  από 1 έως  $k$ .

Η softmax συνάρτηση ενεργοποίησης αποτελεί συνδυασμό πολλών σιγμοειδών συναρτήσεων και μπορεί να χρησιμοποιηθεί για την ταξινόμηση πολλών κλάσεων. Το πλήθος των απαιτούμενων σιγμοειδών συναρτήσεων ταυτίζεται με το πλήθος των κλάσεων ταξινόμησης, ενώ, καθώς η έξοδος κάθε σιγμοειδούς συνάρτησης ανήκει στο διάστημα  $(0, 1)$ , μπορεί να θεωρηθεί ως η πιθανότητα το δεδομένο εισόδου να ανήκει στην αντίστοιχη κατηγορία. [20], [23]

### 3.3 Perceptron πολλών επιπέδων (Multilayer Perceptron)

Για την αντιμετώπιση πιο σύνθετων προβλημάτων που περιλαμβάνουν την ταξινόμηση μη γραμμικά διαχωρίσιμων κλάσεων, χρησιμοποιούνται επίπεδα νευρώνων, ή Perceptrons πολλών επιπέδων (Multilayer Perceptrons ή MLPs). Τα μοντέλα αυτά χαρακτηρίζονται από την ύπαρξη ενδιάμεσων επιπέδων (hidden layers), δηλαδή επιπέδων νευρώνων που δεν είναι ορατά από την είσοδο και την έξοδο του δικτύου. Τυπικά, κάθε νευρώνας ενός επιπέδου δέχεται ως είσοδο τις εξόδους των νευρώνων του αμέσως προηγούμενου επιπέδου, ενώ αντίστοιχα, η έξοδος κάθε νευρώνα ενός επιπέδου γίνεται είσοδος των νευρώνων του αμέσως επόμενου. Λόγω αυτής της διαρρύθμισης τα MLPs καλούνται πλήρως συνδεδεμένα (fully connected) ΤΝΔ. Καθώς η πληροφορία ρέει από την είσοδο προς την έξοδο, τέτοια δίκτυα ονομάζονται επίσης πρόσθιας τροφοδότησης ή feedforward δίκτυα. Τέλος, ένα ακόμα χαρακτηριστικό των multilayer perceptrons αποτελεί η χρήση μη γραμμικών συναρτήσεων ενεργοποίησης, με στόχο την απεικόνιση τόσο των γραμμικών, όσο και των μη γραμμικών συναρτήσεων που περιγράφουν τη σχέση εισόδου-εξόδου. Μάλιστα, έχει αποδειχθεί ότι ένα απλό πολυεπίπεδο δίκτυο πρόσθιας τροφοδότησης μπορεί να προσομοιώσει οποιαδήποτε μαθηματική συνάρτηση, με την επιθυμητή ακρίβεια, εφόσον διαθέτει τον κατάλληλο αριθμό ενδιάμεσων επιπέδων. [24]



Σχήμα 3.4: Πλήρως συνδεδεμένο δίκτυο πρόσθιας τροφοδότησης με ένα ενδιάμεσο επίπεδο [20]

Αυτή η κατανομημένη μη γραμμικότητα και η υψηλή συνδεσιμότητα που περιγράφει τα MLPs καθιστά την εκπαίδευσή τους πολύ πιο σύνθετη, σε σχέση με αυτή των απλών perceptrons. Η πιο συχνά χρησιμοποιούμενη μέθοδος για την εκπαίδευση τέτοιων δικτύων αποτελεί ο αλγόριθμος διάδοσης σφάλματος προς τα πίσω (backpropagation algorithm), ο οποίος εδραιώθηκε το 1986. [25] Με βάση τον συγκεκριμένο αλγόριθμο, η εκπαίδευση πραγματοποιείται σε δύο φάσεις: στη φάση διάδοσης προς τα εμπρός

(*forward phase*), και στη φάση διάδοσης προς τα πίσω (*backward phase*). Στην πρώτη, τα συναπτικά βάρη του δικτύου παραμένουν αμετάβλητα και η πληροφορία μεταβιβάζεται από επίπεδο σε επίπεδο, από τα αριστερά προς τα δεξιά. Η έξοδος του ΤΝΔ διαμορφώνεται με βάση τη χρησιμοποιούμενη συνάρτηση ενεργοποίησης και την αρχική είσοδο. Αντίθετα, στο δεύτερο στάδιο, στο δίκτυο μεταδίδεται ένα σήμα σφάλματος το οποίο προκύπτει από τη σύγκριση της υπολογιζόμενης εξόδου με την επιθυμητή, αυτήν την φορά με κατεύθυνση από την έξοδο προς την είσοδο. Σε αυτήν την φάση γίνεται διαδοχική προσαρμογή των βαρών κάθε επιπέδου, με στόχο τη διόρθωση της αντίστοιχης απόκλισης. Όπως μπορεί να παρατηρηθεί, η εκπαίδευση με τον backpropagation αλγόριθμο προϋποθέτει την εκ των προτέρων γνώση της επιθυμητής εξόδου  $d(n)$  για κάθε δεδομένο εισόδου  $x(n)$ , οπότε κατηγοριοποιείται στον λεγόμενο επιβλεπόμενο τρόπο μάθησης (*supervised learning*).

Αναλυτικότερα, στόχος είναι ο υπολογισμός των ποσοτήτων  $\Delta w_{ji}(n)$ , που με τη σειρά τους είναι ανάλογες με τις αντίστοιχες μερικές παραγώγους  $\partial \mathcal{E}(n)/\partial w_{ji}(n)$ , για κάθε δεδομένη είσοδο  $x(n)$ . Οι παράγωγοι αυτές φανερώνουν πόσο πρέπει να μεταβληθούν τα συναπτικά βάρη  $w_{ji}(n)$  για να ελαχιστοποιηθεί η ποσότητα  $\mathcal{E}(n)$ , δηλαδή το συνολικό σφάλμα του δικτύου. Ορίζουμε:

$$\mathcal{E}(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n), \quad (3.12)$$

όπου  $e_j(n) = d_j(n) - y_j(n)$  το σήμα σφάλματος για τον νευρώνα  $j$  και  $C$  το σύνολο που περιλαμβάνει όλους τους νευρώνες του επιπέδου εξόδου.

Για τον υπολογισμό κάθε ποσότητας  $\partial \mathcal{E}(n)/\partial w_{ji}(n)$  μπορεί να χρησιμοποιηθεί ο κανόνας αλυσίδας (*chain rule*):

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = \frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial w_{ji}(n)} \quad (3.13)$$

Λαμβάνοντας ακόμα υπόψιν ότι  $y_j(n) = \sum_{i=0}^m w_{ji}(n)x_i(n)$  και  $y_j(n) = \phi_j(y_j(n))$ , κι εφόσον η χρησιμοποιούμενη συνάρτηση ενεργοποίησης  $\phi(\cdot)$  είναι παραγωγίσιμη, καταλήγουμε στη σχέση:

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = -e_j(n)\phi'_j(y_j(n))y_i(n) \quad (3.14)$$

Η αλλαγή  $\Delta w_{ji}(n)$  που εφαρμόζεται στο αντίστοιχο βάρος  $w_{ji}(n)$  καθορίζεται από τον κανόνα δέλτα (*delta rule*):

$$\Delta w_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} \quad (3.15)$$

όπου  $\eta$  ο ρυθμός εκπαίδευσης του backpropagation αλγορίθμου.

Το αρνητικό πρόσημο στην παραπάνω σχέση προκύπτει διότι αναζητείται η μεταβολή του αντίστοιχου συναπτικού βάρους που επιφέρει μείωση της ποσότητας  $\mathcal{E}(n)$ , όπως ορίζει η μέθοδος κατάβασης δυναμικού (*gradient descent*). Οπότε τελικά έχουμε:

$$\Delta w_{ji}(n) = \eta \delta_j(n)y_i(n), \quad (3.16)$$

όπου  $\delta_j(n) = -e_j(n)\phi'_j(y_j(n))$ .

Παρατηρούμε επομένως ότι μένει μόνο να υπολογιστεί το μέγεθος  $e_j(n)$ , οπότε και διακρίνουμε περιπτώσεις. Στην πιο απλή περίπτωση όπου ο  $j$  αποτελεί νευρώνας του

επιπέδου εξόδου, η τιμή  $e_j(n)$  είναι γνωστή και δίνεται από την προαναφερθείσα σχέση  $e_j(n) = d_j(n) - y_j(n)$ . Αντίθετα, στην πιο σύνθετη περίπτωση όπου ο  $j$  αποτελεί νευρώνας ενδιάμεσου επιπέδου, αποδεικνύεται ότι ισχύει η αναδρομική σχέση:

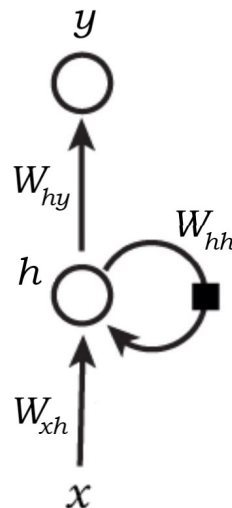
$$\delta_j(n) = \phi_j'(v_j(n)) \sum_{k \in C} \delta_k(n) w_{kj}(n), \quad (3.17)$$

όπου  $C$  το σύνολο που περιλαμβάνει όλους τους νευρώνες του αμέσως επόμενου επιπέδου με τους οποίους συνδέεται ο κρυμμένος νευρώνας  $j$ . [20]

## 3.4 Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks)

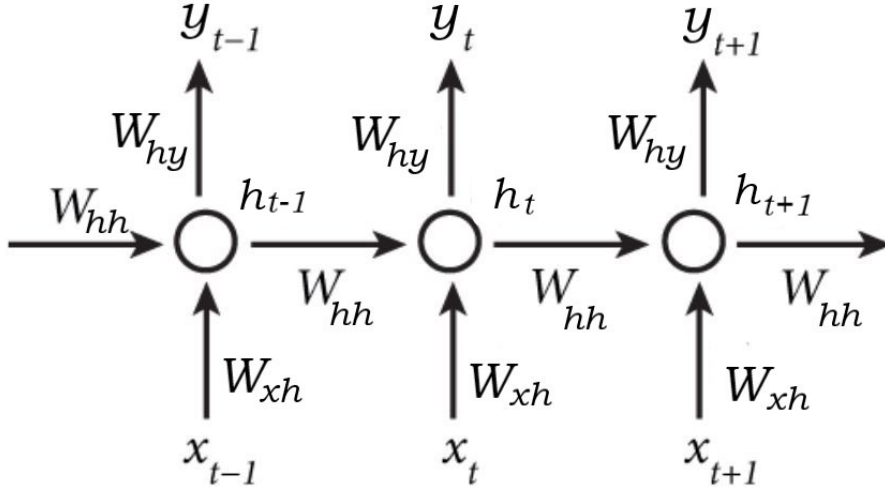
### 3.4.1 Μοντέλο Λειτουργίας και Εκπαίδευση αναδρομικών νευρωνικών δικτύων

Τα αναδρομικά νευρωνικά δίκτυα (Recurrent Neural Networks, ή εν συντομία RNNs) αποτελούν μια κατηγορία νευρωνικών δικτύων, γνωστά για την ικανότητά τους να διαχειρίζονται αποτελεσματικά ακολουθιακά δεδομένα. Συναντώνται συχνά σε εφαρμογές επεξεργασίας φωνής και φυσικής γλώσσας (natural language and speech processing), καθώς και για τη προσθήκη υπομνημάτων σε εικόνες (image captioning). [26] Η δομή ενός αναδρομικού νευρωνικού δικτύου είναι παρόμοια με αυτή ενός κανονικού πολυεπίπεδου Perceptron, με τη διαφορά ότι επιτρέπονται διασυνδέσεις μεταξύ ενδιάμεσων μονάδων, οι οποίες σχετίζονται με χρονικές συσχετίσεις. Έτσι, παρελθοντικές εισοδοι μπορούν να επηρεάζουν την τρέχουσα είσοδο και έξοδο του δικτύου, προσομοιάζοντας ένα είδος μνήμης.



Σχήμα 3.5: Αφαιρετική απεικόνιση ενός αναδρομικού νευρωνικού δικτύου [27]

“Ξετυλίγοντας” τον αναδρομικό βρόγχο του παραπάνω διαγράμματος έτσι ώστε η λειτουργία να απεικονίζεται πλήρως για μια ακολουθία εισόδου  $\mathbf{x} = (x_1, \dots, x_T)$  καταλήγουμε σε ένα νευρωνικό δίκτυο με  $T$  επίπεδα, με μορφή σαν και αυτή που απεικονίζεται στη συνέχεια.



Σχήμα 3.6: Ανεπτυγμένη μορφή ενός αναδρομικού νευρωνικού δικτύου [27]

Το διάνυσμα  $\mathbf{h} = (h_1, \dots, h_T)$  περιέχει τις μεταβλητές *κρυφής κατάστασης*, με τη χρήση των οποίων είναι δυνατή η διατήρηση πληροφορίας από το παρελθόν. Σε αντίθεση με τα νευρωνικά δίκτυα πρόσθιας τροφοδότησης, τα αναδρομικά μοιράζονται τις παραμέτρους βαρών ( $W_{xh}$ ,  $W_{hh}$ ,  $W_{hy}$ ) σε κάθε επίπεδο, μειώνοντας έτσι σημαντικά το συνολικό πλήθος εκπαιδευσίμων παραμέτρων. Αυτή η ιδιότητα επιτρέπει στα RNNs να γενικεύουν τα συμπεράσματά τους και για ακολουθίες με διαφορετικό μήκος από αυτό που έχουν συναντήσει κατά την εκπαίδευσή τους. Τέλος, οι εξισώσεις που περιγράφουν την κατάσταση των μεταβλητών του κάθε επιπέδου δίνονται παρακάτω:

$$\begin{aligned} h_t &= g_1(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \\ y_t &= g_2(W_{hy}h_t + b_y) \end{aligned} \quad (3.18)$$

όπου  $W$  οι πίνακες βαρών,  $b$  τα διανύσματα bias και  $g_1, g_2$  μη γραμμικές συναρτήσεις ενεργοποίησης (όπως για παράδειγμα η  $\tanh$ ). [28].

Η εκπαίδευση των αναδρομικών δικτύων γίνεται μέσω του αλγορίθμου διάδοσης του σφάλματος προς το πίσω δια μέσου χρόνου (*backpropagation through time* ή BPTT), μιας παραλλαγής του παραδοσιακού backpropagation αλγορίθμου. Στον BPTT, κατά την ενημέρωση των βαρών για την ελαχιστοποίηση του σφάλματος  $E_t = \mathcal{L}(y_t)$  που έχει υπολογιστεί με βάση μια είσοδο  $x_t$ , πρέπει να ληφθούν υπόψιν και οι προηγούμενες είσοδοι  $x_1, \dots, x_t - 1$ , σε αντίθεση με την απλή εκδοχή του αλγορίθμου, όπου η διαδικασία διαμόρφωσης βαρών πραγματοποιείται ανεξάρτητα για κάθε είσοδο.

Πιο συγκεκριμένα, έστω ότι επιθυμούμε να εκτελέσουμε τον BPTT αλγόριθμο για μια είσοδο  $x_3$ , μέρος μιας ακολουθίας  $(x_1, x_2, x_3)$ . Υπολογίζουμε το σφάλμα  $E_3$ , συναρτήσει της (γνωστής) επιθυμητής εξόδου  $\hat{y}_3$  και της υπολογιζόμενης από το δίκτυο εξόδου,  $y_3$ . Στη συνέχεια, για τη προσαρμογή των βαρών  $W_{xh}$ ,  $W_{hh}$ ,  $W_{hy}$  με στόχο την ελαχιστοποίηση του σφάλματος, υπολογίζουμε τις αντίστοιχες μερικές παραγώγους, με χρήση του κανόνα αλυσίδας.

Για τον πίνακα βαρών  $W_{hy}$ :

$$\frac{\partial E_3}{\partial W_{hy}} = \frac{\partial E_3}{\partial y_3} \frac{\partial y_3}{\partial W_{hy}}, \quad (3.19)$$

καθώς το μέγεθος  $E_3$  αποτελεί συνάρτηση του  $y_3$  και το  $y_3$  εξαρτάται άμεσα από το  $W_{hy}$ , όπως φαίνεται και σε προαναφερθείσα εξίσωση.



Παρόλα αυτά, η περίπτωση προσαρμογής του πίνακα βαρών  $W_{hh}$  είναι πιο σύνθετη. Διατυπώνουμε αντίστοιχα τον κανόνα αλυσίδας  $\frac{\partial \mathcal{E}_3}{\partial W_{hh}} = \frac{\partial \mathcal{E}_3}{\partial y_3} \frac{\partial y_3}{\partial h_3} \frac{\partial h_3}{\partial W_{hh}}$  και παρατηρούμε ότι τώρα, η μεταβλητή  $h_3$  εξαρτάται από την  $h_2$ , η οποία με τη σειρά της εξαρτάται από την τιμή του  $W_{hh}$ . Επομένως, κατά τον υπολογισμό της μερικής παραγώγου  $\frac{\partial \mathcal{E}_3}{\partial W_{hh}}$  δεν μπορούμε να αντιμετωπίσουμε το  $h_2$  ως σταθερά και οφείλουμε να συνεχίσουμε να εφαρμόζουμε τον κανόνα αλυσίδας επαναληπτικά. [28] Τελικά προκύπτει:

$$\frac{\partial \mathcal{E}_3}{\partial W_{hh}} = \frac{\partial \mathcal{E}_3}{\partial y_3} \frac{\partial y_3}{\partial h_3} \frac{\partial h_3}{\partial W_{hh}} + \frac{\partial \mathcal{E}_3}{\partial y_3} \frac{\partial y_3}{\partial h_3} \frac{\partial h_3}{\partial h_2} \frac{\partial h_2}{\partial W_{hh}} + \frac{\partial \mathcal{E}_3}{\partial y_3} \frac{\partial y_3}{\partial h_3} \frac{\partial h_3}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial W_{hh}} \quad (3.20)$$

Ενώ γενικότερα μπορούμε να γράψουμε:

$$\frac{\partial \mathcal{E}_N}{\partial W_{hh}} = \sum_{i=1}^N \frac{\partial \mathcal{E}_N}{\partial y_N} \frac{\partial y_N}{\partial h_N} \frac{\partial h_N}{\partial h_k} \frac{\partial h_k}{\partial W_{hh}} \quad (3.21)$$

Κάθε  $i$ -στος όρος του παραπάνω αθροίσματος αποτυπώνει το πως επηρεάζει η τιμή βάρους  $W_{hh}$  που αντιστοιχεί στο βήμα  $i$ , το υπολογιζόμενο σφάλμα  $\mathcal{E}_N$  της χρονικής στιγμής  $N > i$ . [29]

Ομοίως για  $W_{xh}$ :

$$\frac{\partial \mathcal{E}_3}{\partial W_{xh}} = \frac{\partial \mathcal{E}_3}{\partial y_3} \frac{\partial y_3}{\partial h_3} \frac{\partial h_3}{\partial W_{xh}} + \frac{\partial \mathcal{E}_3}{\partial y_3} \frac{\partial y_3}{\partial h_3} \frac{\partial h_3}{\partial h_2} \frac{\partial h_2}{\partial W_{xh}} + \frac{\partial \mathcal{E}_3}{\partial y_3} \frac{\partial y_3}{\partial h_3} \frac{\partial h_3}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial W_{xh}} \quad (3.22)$$

$$\frac{\partial \mathcal{E}_N}{\partial W_{xh}} = \sum_{i=1}^N \frac{\partial \mathcal{E}_N}{\partial y_N} \frac{\partial y_N}{\partial h_N} \frac{\partial h_N}{\partial h_k} \frac{\partial h_k}{\partial W_{xh}} \quad (3.23)$$

Αν και τα RNNs αποδίδουν ικανοποιητικά σε πλήθος εφαρμογών, αποτυγχάνουν να αποτυπώσουν συσχετίσεις με μεγάλη χρονική απομάκρυνση, με αποτέλεσμα η αποτελεσματικότητά τους να περιορίζεται σημαντικά. Το φαινόμενο αυτό περιγράφεται ως το πρόβλημα "εξαφανιζόμενων κλίσεων" (*vanishing gradients problem*) και παρουσιάστηκε για πρώτη φορά από πλήθος ανεξάρτητων ερευνητών. [30], [31]. Το *vanishing gradients* πρόβλημα μπορεί να γίνει καλύτερα κατανοητό λαμβάνοντας υπόψιν ότι η μερική παράγωγος  $\partial h_N / \partial h_k$  που εμφανίζεται στους παραπάνω υπολογισμούς επίσης υπολογίζεται με χρήση του κανόνα αλυσίδας:

$$\frac{\partial h_N}{\partial h_k} = \prod_{i=k+1}^N \frac{\partial h_i}{\partial h_{i-1}} \quad (3.24)$$

Αποδεικνύεται [29] ότι ισχύει:

$$\frac{\partial h_N}{\partial h_k} = \prod_{i=k+1}^N \frac{\partial h_i}{\partial h_{i-1}} = \prod_{i=k+1}^N \mathbf{W}_{hh}^T \text{diag}(g'_1(h_{i-1})) \quad (3.25)$$

όπου η συνάρτηση *diag* μετατρέπει ένα διάνυσμα σε διαγώνιο πίνακα και η  $g'_1$  υπολογίζει ανά στοιχείο την παράγωγο της συνάρτησης ενεργοποίησης  $g_1$ .

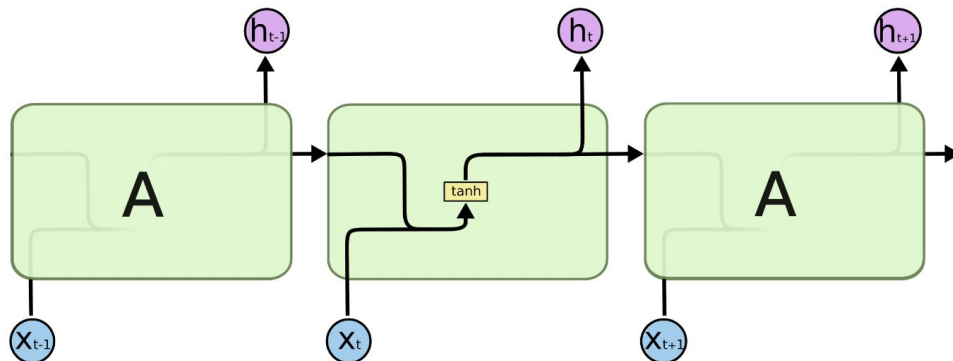
Για μη γραμμικές συναρτήσεις ενεργοποίησης  $g_1$  όπου η παράγωγος  $|g'_1(x)|$  είναι φραγμένη (όπως συμβαίνει στις περιπτώσεις των συχνά χρησιμοποιούμενων *tanh* και *sigmoid*), η τιμή  $\|\text{diag}(g'_1(h_{i-1}))\|$  είναι επίσης φραγμένη, και στις περισσότερες περιπτώσεις προκύπτει:  $\|\mathbf{W}_{hh}^T\| \cdot \|\text{diag}(g'_1(h_{i-1}))\| < 1$ .

Όπως ο συνεχής πολλαπλασιασμός πραγματικών αριθμών μικρότερων της μονάδας οδηγεί σε συρρίκνωση τους, έτσι συμβαίνει και στην περίπτωση πολλαπλασιασμού πινάκων, με αποτέλεσμα οι όροι που αποτυπώνουν μεγαλύτερες χρονικές συσχετίσεις να πλησιάζουν εκθετικά σε μηδενικές τιμές. Ως αποτέλεσμα, η διαδικασία της εκπαίδευσης επιβραδύνεται και η αποδοτικότητα των RNN δικτύων πλήττεται ήδη για ακολουθίες με μόλις 10 - 20 στοιχεία.

Αξίζει τέλος να σημειωθεί ότι στις πιο σπάνιες περιπτώσεις όπου ισχύσει  $\|W_{hh}^T\| \cdot \|diag(g_1'(h_{i-1}))\| > 1$ , εμφανίζεται το αντίστροφο πρόβλημα (*exploding gradients problem*), κατά το οποίο οι όροι μακροπρόθεσμης συσχέτισης αυξάνονται εκθετικά, σε σχέση με τους υπόλοιπους. Το exploding gradients problem επίσης προκαλεί πρόβλημα στη διαδικασία της εκπαίδευσης, αλλά ανιχνεύεται πιο εύκολα, σε σχέση με το vanishing gradients πρόβλημα. [28], [29]

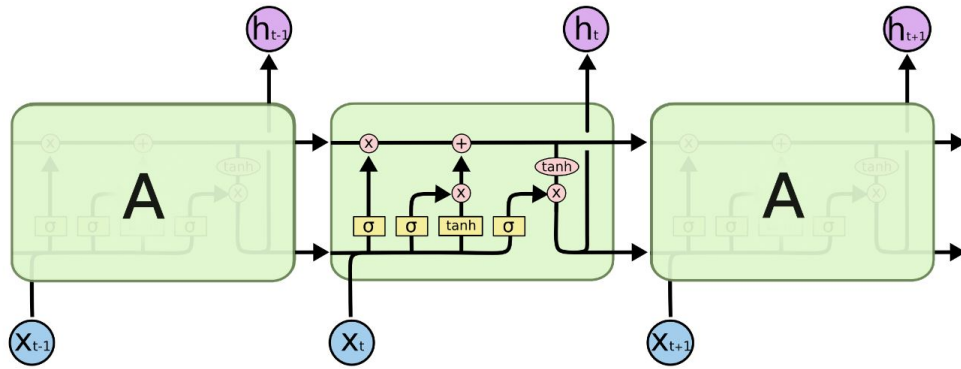
### 3.4.2 Long Short-Term Memory (LSTM) και Gated Recurrent Unit (GRU) δίκτυα

Για την συμπερίληψη των μακροπρόθεσμων εξαρτήσεων και την αντιμετώπιση του προβλήματος των vanishing gradients, το 1997 προτάθηκαν τα Long Short-Term Memory (LSTM) δίκτυα. [32] Αν και τα LSTMs επίσης διακατέχονται από την αλυσιδωτή δομή που χαρακτηρίζει όλα τα αναδρομικά νευρωνικά δίκτυα, μέσω αυτών συστήνεται ένας πιο σύνθετος τρόπος κατασκευής των επαναλαμβανόμενων μονάδων (*κελιών*) που συνθέτουν την εν λόγω αλυσίδα. Συγκεκριμένα, ένα κανονικό RNN δίκτυο μπορεί να περιγραφεί αφαιρετικά από την παρακάτω εικόνα:



Σχήμα 3.7: Αφαιρετική απεικόνιση των απλών, επαναλαμβανόμενων δομών που συνθέτουν ένα ανεπτυγμένο, κανονικό RNN δίκτυο [33]

Αντίστοιχα, μέσω των LSTMs προτείνεται μια πιο σύνθετη, πολυεπίπεδη δομή, για κάθε κατασκευαστική μονάδα του αναδρομικού δικτύου:



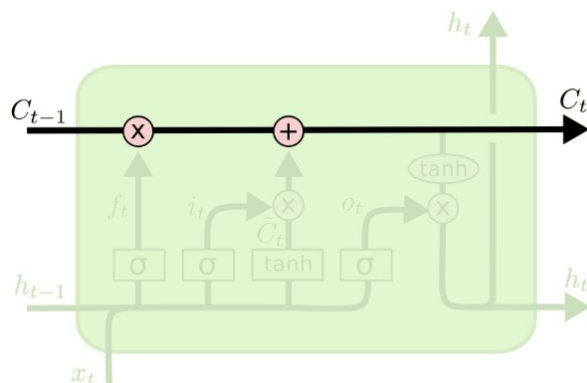
Σχήμα 3.8: Αφαιρετική απεικόνιση των επαναλαμβανόμενων δομών που συνθέτουν ένα ανεπτυγμένο LSTM δίκτυο. [33]

Αν και υφίστανται διάφορες παραλλαγές LSTM δικτύων με παρόμοια, μεταξύ τους, επίδοση [34], η επικρατέστερη στη βιβλιογραφία εκδοχή περιγράφεται από τις παρακάτω σχέσεις: [35]

$$\begin{aligned}
 f_t &= \sigma(W_{fh}h_{t-1} + W_{fx}x_t + b_f) \\
 i_t &= \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \\
 o_t &= \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \\
 \tilde{c}_t &= \tanh(W_{ch}h_{t-1} + W_{cx}x_t + b_c) \\
 c_t &= f_t * c_{t-1} + i_t * \tilde{c}_t \\
 h_t &= o_t * \tanh(c_t)
 \end{aligned}
 \tag{3.26}$$

όπου με  $\sigma$  συμβολίζεται η σιγμοειδής συνάρτηση και με  $W$  και  $b$  οι αντίστοιχοι πίνακες βαρών και τα biases.

Βασική έννοια των LSTM δικτύων αποτελεί η λεγόμενη κατάσταση κελιού (cell state), η οποία για ένα συγκεκριμένο κελί  $t$  συμβολίζεται ως  $c_t$ . Μέσω της κατάστασης κελιού διευκολύνεται η διατήρηση και η απευθείας μεταφορά πληροφορίας από το παρελθόν, με λίγες μόνο παρεμβαλλόμενες γραμμικές τροποποιήσεις.



Σχήμα 3.9: Μέσω των μεταβλητών  $c_t$  είναι δυνατή η ροή πληροφορίας καθόλη την αλυσίδα, όπως φαίνεται και από τη (σχεδόν) αδιάκοπτη γραμμή που διατρέχει το πάνω μέρος του διαγράμματος [33]

Η κατάσταση ενός κελιού  $t$ ,  $c_t$ , διαμορφώνεται μέσω των λεγόμενων *πυλών* (gates)  $f_t$  (forget gate),  $i_t$  (input gate) και  $o_t$  (output gate). Όπως φαίνεται και από τις παραπάνω σχέσεις, οι εξισώσεις που καθορίζουν τις τιμές των μεταβλητών  $f_t$ ,  $i_t$  και  $o_t$  είναι σχεδόν πανομοιότυπες, με μόνη διαφορά τις τιμές των αντίστοιχων πινάκων βαρών και biases, οι οποίες διαμορφώνονται κατά τη διαδικασία εκπαίδευσης. Μέσω της εφαρμοζόμενης σιγμοειδούς συνάρτησης  $\sigma$  οι έξοδοι των πυλών περιορίζονται στο διάστημα  $[0, 1]$ , όπου για έξοδο ίση με 0 η ροή πληροφορίας διακόπτεται, ενώ για έξοδο ίση με 1 η πληροφορία ρέει πλήρως. Σημειώνουμε επίσης ότι η μεταβλητή  $\tilde{c}_t$  αποτελεί την *υποψήφια* νέα τιμή της κατάστασης κελιού. Τελικά, το μέγεθος  $c_t$  προκύπτει από την προηγούμενη αντίστοιχη τιμή  $c_{t-1}$  (ένα τμήμα της οποίας μπορούμε να αγνοήσουμε με χρήση της forget gate  $f_t$ ) και της υποψήφιας νέας τιμής  $\tilde{c}_t$  (η επίδραση της οποίας καθορίζεται από την  $i_t$ ). Η τελική κρυφή κατάσταση ενός κελιού  $t$ ,  $h_t$ , αποτελεί συνάρτηση της υπολογιζόμενης cell state  $c_t$  και της output gate  $o_t$ . Παρατηρούμε ότι η συμπεριφορά ενός απλού RNN μπορεί να προσομοιαστεί θέτοντας τις εξόδους των πυλών  $i_t$  και  $o_t$  ίσες με 1 και την έξοδο της πύλης  $f_t$  ίση με 0. [35]

Ένας τύπος δικτύων παρεμφερής με τα LSTM δίκτυα, αποτελεί αυτός των λεγόμενων *Gated Recurrent Unit* ή GRU δικτύων. Τα δίκτυα αυτά προτάθηκαν το 2014 [36] κι έκτοτε έχουν χρησιμοποιηθεί σε πολλές εφαρμογές ως μια πιο απλή εκδοχή των LSTMs, για την αντιμετώπιση του vanishing gradients προβλήματος. Στην περίπτωση των GRUs οι πύλες είναι δύο αντί για τρεις, ενώ δεν γίνεται διάκριση μεταξύ cell και hidden κατάστασης, όπως φαίνεται στη συνέχεια :

$$\begin{aligned} u_t &= \sigma(W_{uh}h_{t-1} + W_{ux}x_t + b_u) \\ r_t &= \sigma(W_{rh}h_{t-1} + W_{rx}x_t + b_r) \\ \tilde{h}_t &= \tanh(W_{hh}(r_t * h_{t-1}) + W_{ht}x_t + b_h) \\ h_t &= u_t * h_{t-1} + (1 - u_t) * \tilde{h}_t \end{aligned} \tag{3.27}$$

όπου  $W$  και  $b$  τα αντίστοιχα βάρη και biases,  $\sigma$  η σιγμοειδής συνάρτηση, και  $u_t$ ,  $r_t$  η *update* και η *reset* πύλη αντίστοιχα. [35]

Με βάση εμπειρικές συγκρίσεις που έχουν πραγματοποιηθεί [37] [38], τα LSTM και GRU δίκτυα εμφανίζουν συγκρίσιμες επιδόσεις στις περισσότερες εφαρμογές. Παρόλα αυτά, οι λιγότερες εκπαιδευσιμες παράμετροι που περιλαμβάνουν τα GRUs πιθανόν να καθιστούν την εκπαίδευσή τους συντομότερη και αποδοτικότερη. Αντίστοιχα, η μεγαλύτερη εκφραστικότητα των LSTMs θα μπορούσε να αξιοποιηθεί σε περιπτώσεις εκτενέστερων και περιπλοκότερων συνόλων δεδομένων.

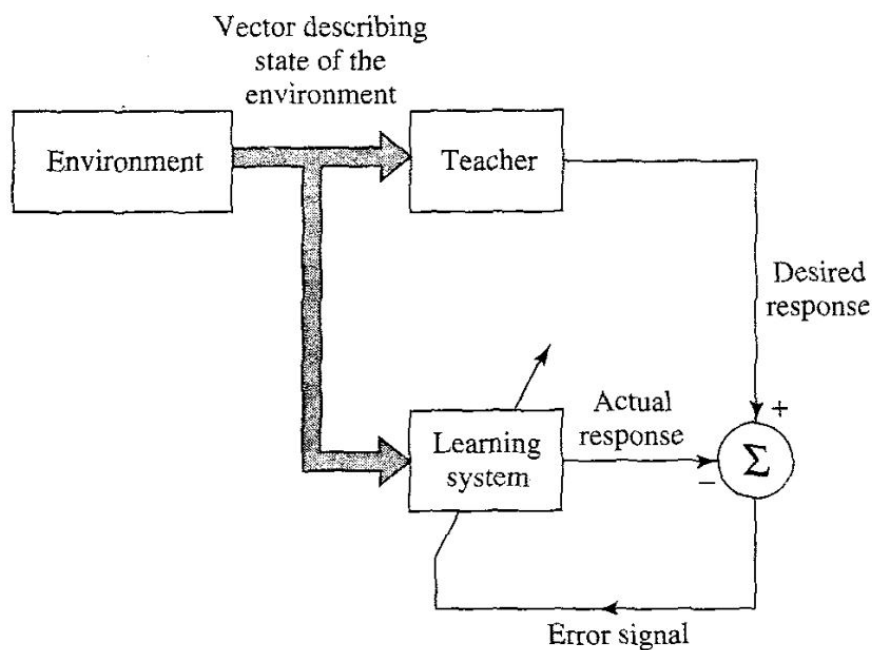
## 3.5 Εκπαίδευση και αξιολόγηση τεχνητών νευρωνικών δικτύων

### 3.5.1 Ταξινόμηση ειδών μάθησης

Τα είδη μηχανικής μάθησης ποικίλλουν και ένα από τα βασικότερα κριτήρια ταξινόμησης τους αποτελεί το εάν αυτές πραγματοποιούνται υπό την επίβλεψη κάποιου "δασκάλου" ή όχι. Στην πρώτη περίπτωση, δηλαδή στην περίπτωση *επιβλεπόμενης* (supervised) μάθησης, υποθέτουμε ότι υπάρχει ήδη κάποια γνώση στο περιβάλλον, η οποία παρουσιάζεται στο εκπαιδευόμενο μοντέλο υπό την μορφή ενός συνόλου εισόδων και των αντίστοιχων, επιθυμητών εξόδων. Οι παράμετροι του δικτύου προσαρμόζονται υπό την συνδυασμένη επίδραση αυτού του συνόλου εκπαίδευσης (*training set*) και ενός

σήματος σφάλματος (*error signal*), το οποίο ποσοτικοποιεί τη διαφορά μεταξύ υπολογιζόμενων και επιθυμητών εξόδων. Καθώς το σφάλμα αυτό αποτελεί συνάρτηση όλων των ελεύθερων παραμέτρων του δικτύου, μπορεί να οριστεί μια επιφάνεια, η λεγόμενη *επιφάνεια σφάλματος*. Μέσω της προσαρμογής των βαρών του μοντέλου και μεθόδων βελτιστοποίησης *gradient descent* [39], αναζητείται ένας συνδυασμός παραμέτρων ο οποίος αντιστοιχεί σε τοπικό ή ολικό ελάχιστο της προαναφερθείσας επιφάνειας.

Η διαδικασία εκπαίδευσης πραγματοποιείται με επαναληπτικό τρόπο, καθώς τα ζεύγη εισόδων-επιθυμητών εξόδων (ή *πρότυπα*) παρουσιάζονται στο δίκτυο με κυκλική σειρά και όταν εξαντληθούν επαναλαμβάνονται από την αρχή. Ένας πλήρης κύκλος χρήσης όλων των προτύπων ονομάζεται *εποχή* (*epoch*). Όταν το υπολογιζόμενο σφάλμα έχει ελαχιστοποιηθεί ικανοποιητικά, θεωρούμε ότι η γνώση από το περιβάλλον έχει μεταφερθεί επαρκώς στο υπό εκπαίδευση μοντέλο. [20], [22]



Σχήμα 3.10: Ενοιολογική απεικόνιση της επιβλεπόμενης μάθησης [20]

Η επιβλεπόμενη μάθηση χρησιμοποιείται συχνά σε εφαρμογές ταξινόμησης, όπου με βάση μοτίβα που έχουν ήδη εξαχθεί από το μοντέλο μέσω της εκπαίδευσής του, οι νέοι εισόδοι ταξινομούνται σε προκαθορισμένες, διακριτές κατηγορίες. Ανάλογα με το πλήθος των κλάσεων αυτών, η ταξινόμηση μπορεί να διακριθεί σε *δυαδική* (*binary*) ή *multi-class*. Στην πρώτη περίπτωση, οι διαθέσιμες κλάσεις εξόδου είναι δύο, ενώ στη δεύτερη, τα αντικείμενα εισόδου κατηγοριοποιούνται σε μία κλάση ανάμεσα σε πάνω από δύο διαθέσιμες. Για περιπτώσεις ταξινόμησης όπου οι εισόδοι μπορούν να ανήκουν σε πάνω από μία κατηγορία, λέμε ότι έχουμε *multi-label* ταξινόμηση. Δημοφιλείς ταξινομητές αποτελούν τα τεχνητά νευρωνικά δίκτυα, τα δέντρα απόφασης (*decision trees*), τα *Support Vector Machines* (*SVMs*) και αλγόριθμοι όπως ο *Naïve Bayes* και ο *k-Nearest Neighbours* (*k-NN*).

Αντίστοιχα, σε εφαρμογές όπου η επιθυμητή έξοδος του δικτύου αποτελεί μια συνεχή, αριθμητική τιμή, ως μέθοδος πρόβλεψης χρησιμοποιείται συχνά η στατιστική μέθοδος της *παλινδρόμησης* (*regression*). Μέσω της συγκεκριμένης τεχνικής, επιχειρείται να γίνει κατανοητός ο τρόπος με τον οποίο ένα πλήθος ανεξάρτητων μεταβλητών επηρεάζει

κάποιες άλλες, εξαρτημένες μεταβλητές. Στο πεδίο της μηχανικής μάθησης ανεξάρτητες μεταβλητές μπορούν να θεωρηθούν τα χαρακτηριστικά (features) των δεδομένων εισόδου, τα οποία καθορίζουν την τελική έξοδο του δικτύου, η οποία αποτελεί την εξαρτημένη μεταβλητή. [40]

Αντίθετα, στην περίπτωση της *μη επιβλεπόμενης* (unsupervised) μάθησης τα δεδομένα εκπαίδευσης που τροφοδοτούν ένα δίκτυο δεν είναι επισημασμένα, οπότε το δίκτυο καλείται να εξάγει από μόνο του αναπαραστάσεις και πρότυπα που τα περιγράφουν σε ικανοποιητικό βαθμό. Συνήθεις εφαρμογές μη επιβλεπόμενης μάθησης περιλαμβάνουν την ομαδοποίηση (clustering) των δεδομένων εισόδου, αλλά και την αναπαράστασή τους με μειωμένα χαρακτηριστικά, σε χώρους λιγότερων διαστάσεων (dimensionality reduction). [41] Μια ακόμη μορφή μάθησης χωρίς εκπαιδευτή είναι η λεγόμενη *ενισχυμένη μάθηση* (reinforcement learning). Στόχος της ενισχυμένης μάθησης είναι η ανάπτυξη ενός αυτοσυντηρούμενου συστήματος το οποίο σε συνεχείς αλληλουχίες προσπαθεί και αποτυγχάνει. Το σύστημα βελτιώνεται παρατηρώντας τις συνέπειες των προηγούμενων δράσεών του, οι οποίες γίνονται καλύτερα αντιληπτές μέσω σημάτων ανταμοιβής που παρέχονται από το περιβάλλον. Τα σήματα ανταμοιβής χρησιμεύουν ως εργαλείο πλοήγησης για τους αλγόριθμους ενίσχυσης, αφού συμβάλλουν στην κατανόηση της ορθής και λανθασμένης πορείας δράσης. [42]

### 3.5.2 Τεχνικές εκπαίδευσης στην επιβλεπόμενη μάθηση

Καθώς η παρούσα εργασία αποτελεί μια εφαρμογή ταξινόμησης με χρήση επιβλεπόμενης μάθησης, στη συνέχεια παρουσιάζονται με μεγαλύτερη λεπτομέρεια οι συνήθεις τεχνικές που επιστρατεύονται για την εκπαίδευση και την αξιολόγηση ML μοντέλων με επισημασμένα (labeled) δεδομένα.

Στην πιο απλή εκδοχή εκπαίδευσης, το διαθέσιμο σύνολο δεδομένων διασπάται σε δύο ξένα μεταξύ τους υποσύνολα, στο σύνολο εκπαίδευσης (training set) και στο σύνολο ελέγχου (test set). Η συγκεκριμένη μέθοδος ονομάζεται hold-out μέθοδος, και με βάση αυτήν, μόνο το training set χρησιμοποιείται για την προσαρμογή των βαρών του υπό εκπαίδευση δικτύου. Αντίστοιχα, το test set (με συνήθως μέγεθος 20% - 30% του αρχικού dataset) χρησιμοποιείται αποκλειστικά για την αξιολόγησή του μοντέλου με βάση κάποιες επιλεγμένες μετρικές, αφού έχει ολοκληρωθεί η διαδικασία της εκπαίδευσης. Η τεχνική αυτή χρησιμοποιείται συχνά λόγω της υπολογιστικής της απλότητας και τα αποτελέσματά της θεωρούνται σχετικά αξιόπιστα, ειδικά στην περίπτωση πολύ εκτενών συνόλων δεδομένων. Παρόλα αυτά, στην περίπτωση πιο περιορισμένων σε μέγεθος datasets, είναι συχνά θεμιτή η αξιοποίηση όλων των αρχικών δεδομένων για την εκπαίδευση του δικτύου, με στόχο την καλύτερη συμπερίληψη όλων των περιπτώσεων που μπορούν να συναντηθούν στο εξεταζόμενο πρόβλημα. Ένα ακόμα μειονέκτημα της hold-out μεθόδου αποτελεί το γεγονός ότι τα συμπεράσματά της εξαρτώνται σε μεγάλο βαθμό από το πώς γίνεται ο διαμοιρασμός των δεδομένων σε train και test sets, με αποτέλεσμα να πλήττεται η αξιοπιστία της.

Για την αντιμετώπιση των παραπάνω προβλημάτων, μια δημοφιλής εναλλακτική αποτελεί η μέθοδος *διασταυρωμένης επικύρωσης* (k-fold cross-validation) ή k-fold CV. Με βάση την τεχνική αυτή, τα διαθέσιμα δεδομένα του προβλήματος διαχωρίζονται σε  $k$  ίσα (ή σχεδόν ίσα) διαφορετικά υποσύνολα (ή folds) και η hold-out μέθοδος εκτελείται  $k$  φορές. Σε κάθε επανάληψη, ένα διαφορετικό υποσύνολο χρησιμοποιείται ως σύνολο ελέγχου, ενώ τα υπόλοιπα  $k - 1$  folds χρησιμοποιούνται ως ένα εννιαίο training set. Το σφάλμα του υπό εξέταση μοντέλου υπολογίζεται τελικά ως ο μέσος όρος των επιμέρους

σφαλμάτων που προκύπτουν από κάθε εκτέλεση. Έτσι, το μοντέλο εκπαιδεύεται και αξιολογείται σε  $k$  διαφορετικά σύνολα, μειώνοντας την πιθανότητα εμφάνισης ικανοποιητικής απόδοσης λόγω του τρόπου διαμοιρασμού των δεδομένων. Επίσης, το αρχικό dataset συμμετέχει στη διαδικασία προσαρμογής βαρών στο σύνολό του, κάτι που είναι ιδιαίτερα σημαντικό στην περίπτωση περιορισμένων δεδομένων. Παρόλα αυτά, αξίζει να σημειωθεί ότι λόγω των περισσότερων επαναλήψεων που απαιτούνται, η μέθοδος  $k$ -fold cross-validation διακρίνεται από υψηλότερη υπολογιστική πολυπλοκότητα. [43]



Σχήμα 3.11: Η  $k$ -fold cross-validation μέθοδος για  $k$  ίσο με 5 [44]

Μια ειδική, ακραία περίπτωση της  $k$ -fold cross-validation, είναι η τεχνική *leave-one-out* cross-validation (ή LOOCV) όπου η τιμή  $k$  τίθεται ίση με το συνολικό πλήθος των δειγμάτων (έστω  $N$ ). Έτσι το υπό μελέτη μοντέλο εκπαιδεύεται με  $N$  διαφορετικά σύνολα, όπου σε κάθε περίπτωση το training set έχει μέγεθος  $N - 1$ , ενώ η πρόβλεψη και η αξιολόγηση γίνονται μόνο για ένα μεμονωμένο δείγμα. Το κυριότερο πλεονέκτημα της LOOCV είναι ότι κάθε φορά το αρχικά διαθέσιμο dataset αξιοποιείται σχεδόν ολόκληρο, με άμεση βέβαια επιβάρυνση στις χρονικές και υπολογιστικές απαιτήσεις της μεθόδου. Τέλος, μια ακόμα σημαντική μορφή της  $k$ -fold cross-validation είναι η *stratified*  $k$ -fold cross-validation. Η τεχνική αυτή είναι πανομοιότυπη με την απλή  $k$ -fold cross-validation μέθοδο, μόνο που πριν τη διαίρεση σε  $k$  υποσύνολα, τα δεδομένα αναδιατάσσονται, έτσι ώστε το κάθε τελικό υπόσυνολο να είναι όσο το δυνατόν περισσότερο αντιπροσωπευτικό του αρχικού συνόλου. Για παράδειγμα, αν το αρχικό dataset αποτελείται από στοιχεία δύο κλάσεων, με κατανομή κλάσεων 50%, γίνεται προσπάθεια έτσι ώστε το κάθε fold να διακρίνεται από κατανομή κλάσεων της ίδιας περιόδου τιμής. Η τεχνική stratified  $k$ -fold cross-validation είναι ιδιαίτερος χρήσιμη για περιπτώσεις συνόλων που χαρακτηρίζονται από ανισορροπία κλάσεων. [43], [45]

### 3.5.3 Προεπεξεργασία δεδομένων

Η διαδικασία συλλογής και προεπεξεργασίας των δεδομένων εκπαίδευσης αποτελεί ένα από τα σημαντικότερα στάδια σε εφαρμογές μηχανικής μάθησης. Η ποσότητα, αλλά και ποιότητα των διαθέσιμων data καθορίζουν σε μεγάλο βαθμό τα μελλοντικά, τελικά αποτελέσματα και για αυτό, ένα σημαντικό μερίδιο χρόνου αφιερώνεται στην εξόρυξη και επεξεργασία τους. Το στάδιο προεπεξεργασίας συχνά περιλαμβάνει διαδικασίες

καθαρισμού (εξαιρέση ή διόρθωση μη έγκυρων τιμών), κανονικοποίησης, μείωσης χαρακτηριστικών (feature extraction/selection) και παραγωγής νέων χαρακτηριστικών από τα υπάρχοντα (feature construction).

### Κανονικοποίηση δεδομένων

Ειδικότερα, η *κανονικοποίηση* (normalization) αποτελεί μια εδραιωμένη μέθοδος προεπεξεργασίας σε εφαρμογές τεχνητών νευρωνικών δικτύων, καθώς έχει φανεί ότι βελτιώνει την απόδοσή τους και επιταχύνει τη διαδικασία της εκπαίδευσης. Ο περιορισμός των δεδομένων εισόδου σε ένα προκαθορισμένο εύρος σταθμίζει την επιρροή που διαθέτουν τυχούσες τιμές-outliers, με αποτέλεσμα τιμές που δεν είναι χαρακτηριστικές του συνόλου εκπαίδευσης, και πιθανώς να αντιστοιχούν σε λανθασμένες μετρήσεις, να επηρεάζουν σε μικρότερο βαθμό τα τελικά αποτελέσματα. Επιπλέον, η κανονικοποίηση είναι ιδιαίτερως σημαντική σε περιπτώσεις όπου τα χαρακτηριστικά εισόδου βρίσκονται σε διαφορετική κλίμακα. Μέσω του μετασχηματισμού τους, οι υψηλότερες αριθμητικές τιμές ορισμένων χαρακτηριστικών παύουν να επικρατούν έναντι άλλων, χαμηλότερων τιμών, κι επόμενως τα χαρακτηριστικά αυτά λαμβάνονται εξίσου υπόψιν κατά την προσαρμογή των βαρών, από την αρχή της εκπαίδευσης. Καθώς οι μέθοδοι κανονικοποίησης ποικίλλουν, στη συνέχεια παραθέτουμε δύο από τις πιο συχνά χρησιμοποιούμενες τεχνικές:

#### 1. Z-score κανονικοποίηση

$$x'_{i,n} = \frac{x_{i,n} - \mu_i}{\sigma_i} \quad (3.28)$$

όπου  $\mu_i$  και  $\sigma_i$  η μέση τιμή και η τυπική απόκλιση του  $i$ -οστού χαρακτηριστικού εισόδου και  $x_{i,n}, x'_{i,n}$  η αρχική και τελική τιμή του  $n$ -οστού στοιχείου του  $i$ -οστού feature.

#### 2. Min-Max κανονικοποίηση

$$x'_{i,n} = \frac{x_{i,n} - \min(x_i)}{\max(x_i) - \min(x_i)}(nMax - nMin) + nMin \quad (3.29)$$

όπου με  $\min(x_i)$  και  $\max(x_i)$  συμβολίζονται η ελάχιστη και η μέγιστη τιμή του  $i$ -οστού χαρακτηριστικού. Με τους όρους  $nMin$  και  $nMax$  συμβολίζονται τα κάτω και άνω όρια του διαστήματος όπου θα ανήκουν όλα τα δεδομένα έπειτα από την κανονικοποίησή τους. Συνήθως, ως διαστήματα επιλέγονται τα  $[0, +1]$  ή  $[-1, +1]$ . [46]

### Αντιμετώπιση ανισορροπίας κλάσεων δεδομένων

Ένα πρόβλημα το οποίο συναντάται συχνά σε εφαρμογές του πραγματικού κόσμου και μπορεί να αντιμετωπιστεί μερικώς μέσω προεπεξεργασίας των δεδομένων εισόδου, είναι το πρόβλημα της *ανισορροπίας κλάσεων* (class imbalance problem). Σε αυτό το πρόβλημα, τα περισσότερα διαθέσιμα δείγματα ανήκουν στην μία κλάση (κλάση πλειοψηφίας ή majority class), ενώ πολύ λιγότερα δείγματα ανήκουν σε μια δεύτερη κλάση, η οποία αποτελεί την κλάση μειοψηφίας (minority class). Η κλάση μειοψηφίας είναι συνήθως και η κλάση της οποίας η πρόβλεψη έχει τη μεγαλύτερη ερευνητική και επιχειρηματική αξία, καθώς αυτή αποτυπώνει ένα σπάνιο, αξιοσημείωτο ή και επιζήμιο γεγονός. Ενδεικτικές εφαρμογές που χαρακτηρίζονται από ανισορροπία κλάσεων είναι



προβλήματα ανίχνευσης εξαπάτησης (fraud detection), πρόβλεψης φυσικών καταστροφών και ιατρικές εφαρμογές. Σε περιπτώσεις ανισορροπίας δεδομένων έχει αποδειχθεί ότι οι συμβατικοί ταξινομητές δεν αποδίδουν εξίσου καλά, καθώς η κλάση πλειοψηφίας τείνει να κυριαρχεί κατά την διαδικασία προσαρμογής βαρών και η κλάση μειοψηφίας τελικά υποεκπροσωπείται.

Για τη διαχείριση της παραπάνω προβληματικής κατάστασης επιστρατεύονται μέθοδοι προεπεξεργασίας, αλλά και πιο σύνθετοι αλγόριθμοι εκπαίδευσης και ταξινόμησης. Οι συνηθέστερες τεχνικές προεπεξεργασίας επιχειρούν να μεταβάλλουν την κατανομή των δεδομένων εισόδου, μέσω υποδειγματοληψίας (undersampling) ή/και υπερδειγματοληψίας (oversampling) των κλάσεων πλειοψηφίας και μειοψηφίας αντίστοιχα. Στην πρώτη προσέγγιση (περίπτωση υποδειγματοληψίας), αφαιρούνται δείγματα που ανήκουν στη majority class με βάση κάποιο προκαθορισμένο κριτήριο, ή μέσω τυχαίας επιλογής. Η μέθοδος αυτή εφαρμόζεται κυρίως στην περίπτωση όπου ο όγκος των αρχικών δεδομένων είναι επαρκής και ο φόβος απώλειας χρήσιμης πληροφορίας είναι αμελητέος.

Αντίστοιχα, η τεχνική της υπερδειγματοληψίας επαυξάνει τα δεδομένα της minority class μέσω τυχαίου πολλαπλασιασμού ορισμένων εξ αυτών (random oversampling) ή πιο σύνθετων υπολογισμών. Η τυχαία υπερδειγματοληψία, αν και απλή διαδικασία, ενισχύει τον κίνδυνο για *overfitting*, δηλαδή την πιθανότητα το υπό εκπαίδευση μοντέλο να απομνημονεύσει ιδιαιτερότητες και θόρυβο που περιέχονται στο training set, με αποτέλεσμα την υποβάθμιση της ικανότητας γενίκευσής του. Για τον λόγο αυτό, μέθοδοι όπως η SMOTE (Synthetic Minority Oversampling TEchnique) [47], η παραλλαγή της Bordeline SMOTE [48] και ο Adaptive Synthetic (ADASYN) [49] αλγόριθμος αποτελούν δημοφιλείς εναλλακτικές, καθώς παράγουν νέα συνθετικά δεδομένα από τα υπάρχοντα. Η σύνθεση δεδομένων συντελείται μέσω παρεμβολών σε ήδη κοντινά δείγματα της κλάσης μειοψηφίας και πραγματοποιείται σε όλο τον δειγματικό χώρο, ή για την περίπτωση της Bordeline SMOTE, μόνο γύρω από τη διαχωριστική επιφάνεια των κλάσεων. Πέρα από τη μεταβολή της κατανομής των δεδομένων εισόδου, για την αντιμετώπιση της ανισορροπίας κλάσεων εφαρμόζονται ακόμα η μάθηση με ευαισθησία κόστους (cost-sensitive learning), πιο σύνθετα ML μοντέλα (όπως ensemble μοντέλα), ενώ τέλος χρησιμοποιούνται πολλαπλές, πιο ενδεικτικές μετρικές αξιολόγησης. [50], [51]

### 3.5.4 Μετρικές αξιολόγησης μοντέλων μηχανικής μάθησης

Έπειτα από την εκπαίδευση ενός δικτύου, είναι απαραίτητη η αξιολόγηση του με βάση κάποιες μετρικές. Ανάλογα με το αν η υπό εξέταση εφαρμογή αποτελεί classification ή regression πρόβλημα, οι χρησιμοποιούμενες μετρικές μπορεί να διαφέρουν. Επειδή στην περίπτωση μας το πρόβλημα που μελετάται αντιστοιχεί σε πρόβλημα δυαδικής ταξινόμησης, στη συνέχεια περιγράφονται μετρικές που εφαρμόζονται για αυτόν τον σκοπό, αφού αρχικά γίνει μια παρουσίαση των σχετιζόμενων όρων TP (True Positive), TN (True Negative), FP (False Positive) και FN (False Negative). Για τους παρακάτω ορισμούς υποθέτουμε ότι μία από τις δύο κατηγορίες ταξινόμησης ονομάζεται *θετική* κλάση, ενώ η δεύτερη ονομάζεται *αρνητική*.

1. *True Positive (TP)*: Αναφέρεται στο σύνολο των δειγμάτων για τα οποία η πρόβλεψη του μοντέλου είναι θετική, ενώ η πραγματική τιμή είναι επίσης θετική και άρα η ταξινόμηση πραγματοποιείται ορθά.
2. *True Negative (TN)*: Αναφέρεται στο σύνολο των δειγμάτων για τα οποία η προ-

βλεπόμενη κλάση από το μοντέλο είναι η αρνητική, ενώ η πραγματική τιμή είναι επίσης αρνητική και άρα η ταξινόμηση πραγματοποιείται ορθά.

3. *False Positive (FP)*: Περιγράφει τις περιπτώσεις δειγμάτων τα οποία κατατάσσονται στην θετική κλάση, ενώ στην πραγματικότητα ανήκουν στην αρνητική, κι επομένως η ταξινόμηση πραγματοποιείται λανθασμένα.
4. *False Negative (FN)*: Περιγράφει τις περιπτώσεις δειγμάτων τα οποία κατατάσσονται στην αρνητική κλάση, ενώ στην πραγματικότητα ανήκουν στην θετική, και άρα η ταξινόμηση πραγματοποιείται λανθασμένα.

Με βάση τις παραπάνω έννοιες μπορούμε να ορίσουμε τις μετρικές αξιολόγησης accuracy, recall, precision και F1-score:

1. *Accuracy*: ποσοστό επιτυχίας του μοντέλου κατά την ταξινόμηση όλων των δειγμάτων και στις δύο κλάσεις.

$$\frac{TP + TN}{TP + TN + FP + FN} \quad (3.30)$$

2. *Recall*: ποσοστό των συνολικών θετικών δειγμάτων τα οποία ταξινομήθηκαν σωστά, καλείται ισοδύναμα True Positive Rate (TPR) ή και Sensitivity.

$$\frac{TP}{TP + FN} \quad (3.31)$$

3. *Precision*: ποσοστό των θετικών προβλέψεων που αντιστοιχούν σε πραγματικά θετικές τιμές.

$$\frac{TP}{TP + FP} \quad (3.32)$$

4. *F1-score*: ο αρμονικός μέσος των precision και recall, χρησιμοποιείται για μια πιο ισορροπημένη αξιολόγηση του μοντέλου, λαμβάνοντας υπόψιν και τις δύο αυτές μετρικές.

$$2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (3.33)$$

Μια ακόμα χρήσιμη μετρική αξιολόγησης ταξινομητών αποτελεί η AUC (Area Under the Curve), η οποία προκύπτει από το λεγόμενο ROC (Receiver Operating Characteristic) διάγραμμα. Το διάγραμμα αυτό προκύπτει σχεδιάζοντας το μέγεθος recall (ή True Positive Rate) σε σχέση με το False Positive Rate (FPR), που ορίζεται αντίστοιχα ως  $\frac{FP}{TP+FP}$ , για διάφορα κατώφλια ταξινόμησης. Μέσω του ROC διαγράμματος και της περιοχής κάτω από την αντίστοιχη ROC καμπύλη (AUC) αποτυπώνεται η διαχωριστική ικανότητα του εξεταζόμενου ταξινομητή. Πιο αναλυτικά, υψηλότερες τιμές AUC αντιστοιχούν σε καλύτερη ικανότητα ταξινόμησης, με τιμή ίση με 1 να αντιστοιχεί σε ορθή πρόβλεψη όλων των θετικών και αρνητικών δειγμάτων. Αντίστοιχα, για τιμές AUC μεγαλύτερες από 0.5, το εξεταζόμενο μοντέλο καταλήγει σε περισσότερα TP και TN αποτελέσματα, από ό,τι σε FP και FN. [52]

Σημειώνουμε τέλος ότι οι παραπάνω μετρικές αναφέρονται στην περίπτωση της δυαδικής ταξινόμησης, παρόλα αυτά ο ορισμός τους μπορεί να επεκταθεί για να καλύπτει και την περίπτωση ταξινόμησης με πολλαπλές κλάσεις (multi-class classification). [53]

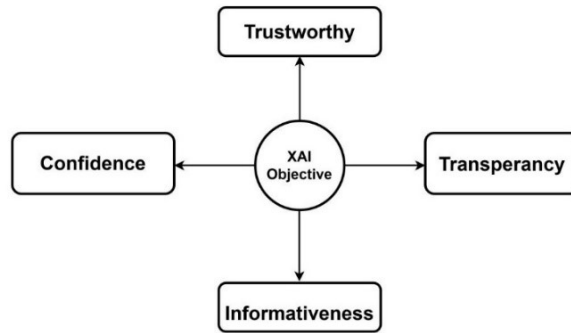
# Κεφάλαιο 4

## Ερμηνευσιμότητα

### 4.1 Η αξία της ερμηνευσιμότητας στον κλάδο της τεχνητής νοημοσύνης

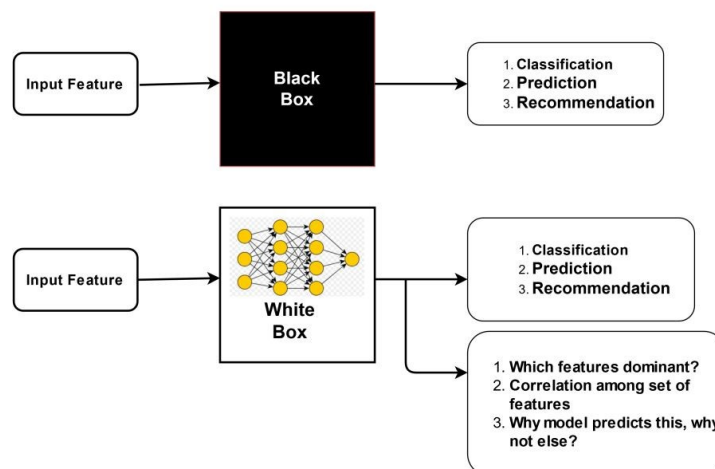
Τα τελευταία χρόνια, τα εφαρμοζόμενα συστήματα τεχνητής νοημοσύνης γίνονται όλο και πιο σύνθετα και οι χρήσεις τους επεκτείνονται σε όλο κι ευρύτερο φάσμα ανθρωπίνων δραστηριοτήτων. Ενδεικτικά αναφέρουμε ότι στις μέρες μας, διάφορα μοντέλα μηχανικής μάθησης συναντώνται σε τραπεζικά συστήματα, συστήματα ιατρικής περίθαλψης, κοινωνικής δικτύωσης, αλλά και άμυνας, μεταξύ άλλων. Παράλληλα, τα περισσότερα από αυτά τα ΑΙ μοντέλα χαρακτηρίζονται από σημαντική πολυπλοκότητα. Αν και η πολυπλοκότητα αυτή συνήθως συμβάλλει στη υψηλότερη αποτελεσματικότητα τους, καθιστά την πλήρη κατανόηση και ερμηνεία του τρόπου λειτουργίας τους απροσπέλαστες ακόμα και για τους ίδιους τους κατασκευαστές τους. Τέτοια συστήματα δρουν ως αδιαφανείς λειτουργικές μονάδες ή «μαύρα κουτιά» (black-boxes) και η σύνθεσή τους πραγματοποιείται μέσω μιας *trial-and-error* διαδικασίας, κατά την οποία ο εκάστοτε μηχανικός πειραματίζεται με τη δομή και τις υπερπαραμέτρους του δικτύου, με στόχο την βελτιστοποίηση του.

Τα παραπάνω έχουν ως άμεσο αποτέλεσμα να πλήττεται η αξιοπιστία και η εμπιστοσύνη προς τα σύνθετα αυτά μοντέλα και να εκφράζονται επιφυλάξεις για την εφαρμογή τους σε προβλήματα του πραγματικού κόσμου και κατά τη λήψη καίριων αποφάσεων. Έτσι, η γενικότερη εφαρμοσιμότητα της τεχνητής νοημοσύνης περιορίζεται, ενώ στις περιπτώσεις που αυτή χρησιμοποιείται δεν είναι πάντοτε εύκολη η επαλήθευση της ακεραιότητας και της ορθότητας των αποτελεσμάτων της. Για την αντιμετώπιση των προαναφερθέντων προκλήσεων, επιστρατεύονται μέθοδοι του κλάδου της *Ερμηνεύσιμης Τεχνητής Νοημοσύνης* (Explainable AI ή XAI). Αυτό το ταχέως αναπτυσσόμενο πεδίο επιχειρεί, μέσω ενός συνόλου διαδικασιών, να παρέχει εξηγήσεις (explanations) για την έξοδο ΑΙ συστημάτων. Στόχος είναι η βαθύτερη κατανόηση και η εξάλειψη τυχόν συστηματικών σφαλμάτων και προκαταλήψεων (bias) ML συστημάτων, πριν αυτά κυκλοφορήσουν στην αγορά. Τελικά, μέσω των XAI μεθόδων επιτυγχάνεται ο σχεδιασμός ορθότερων, πιο δίκαιων (fair) και πιο διαφανών μοντέλων, αυξημένης αξιοπιστίας. [54]



Σχήμα 4.1: Στόχοι της Explainable AI [54]

Ειδικότερα στον ιατρικό τομέα, το χαρακτηριστικό της ερμηνευσιμότητας διαθέτει ακόμα μεγαλύτερη σημασία, λόγω της κρισιμότητας των αποφάσεων που καλούνται να λάβουν τα συστήματα αποφάσεων τεχνητής νοημοσύνης. Σε αυτές τις περιπτώσεις, η κατανόηση της συλλογιστικής πορείας και του τρόπου εξαγωγής αποτελεσμάτων των μοντέλων μπορεί να αποτρέψει τη λήψη λανθασμένων αποφάσεων και να διαδραματίσει καθοριστικό ρόλο όταν οι κρίσεις των ΑΙ συστημάτων δεν συμφωνούν με αυτές των επαγγελματιών υγείας. Η αξία της ερμηνεύσιμης τεχνητής νοημοσύνης τα τελευταία χρόνια εδραιώνεται και νομικά, με χαρακτηριστικό παράδειγμα τον ευρωπαϊκό Γενικό Κανονισμό για τη Προστασία των Δεδομένων (General Data Protection Regulation ή GDPR) του 2018. Με βάση το GDPR, κατοχυρώνεται το δικαίωμα στην εξήγηση (*right of explanation*), σύμφωνα με το οποίο οι ασθενείς δικαιούνται να λάβουν ουσιαστικές πληροφορίες σχετικά με το πως λήφθηκε μια ιατρική απόφαση που τους αφορά. Αν και ο νομικά απαιτούμενος βαθμός ερμηνευσιμότητας δεν καθορίζεται σαφώς, είναι βέβαιο ότι παρόμοιες νομοθετικές προσθήκες θα συνεχίσουν να γίνονται στο κοντινό μέλλον, καθώς όλο και πιο πολύπλοκα ΑΙ συστήματα γίνονται μέρος της ανθρώπινης καθημερινότητας. Τέλος, σημειώνουμε ότι η χρήση τεχνικών ερμηνευσιμότητας συνάδει με τις patient-centered προσεγγίσεις ιατρικής περίθαλψης, με τις οποίες δίνεται έμφαση στην ενημέρωση και στη συμπερίληψη του ασθενούς καθ' όλη τη θεραπευτική διαδικασία. [55]



Σχήμα 4.2: ΑΙ σε σύγκριση με XAI [54]

## 4.2 Ταξινόμηση μεθόδων ερμηνευσιμότητας

### 4.2.1 Διαφανή και αδιαφανή μοντέλα μηχανικής μάθησης

Οι τεχνικές ερμηνευσιμότητας, ανάλογα με το τρόπο εφαρμογής, τα μοντέλα στα οποία εφαρμόζονται και τα αποτελέσματά τους, μπορούν να διακριθούν σε κατηγορίες. Ένα συχνά εμφανιζόμενο κριτήριο ταξινόμησης βασίζεται στα χαρακτηριστικά των μοντέλων μηχανικής μάθησης που μελετώνται. Συγκεκριμένα, τα μοντέλα μπορούν να διακριθούν σε *εγγενώς ερμηνεύσιμα* (ή διαφανή) και σε *μη-εγγενώς ερμηνεύσιμα* (ή black-boxes). Αν και τα όρια αυτού του διαχωρισμού δεν είναι πάντα σαφή, συνήθως, για να θεωρηθεί ένα σύστημα διαφανές θα πρέπει να διαθέτει κάποια (ή κάποιες) από τις ιδιότητες *simulatability*, *decomposability* ή *algorithmic transparency*.

Όσον αφορά τα *simulatable* συστήματα, η λειτουργία αυτών μπορεί να προσομοιωθεί λογικά πλήρως από κάποιον άνθρωπο. Για μεγαλύτερα και πιο πολύπλοκα συστήματα όπου τέτοια προσομοίωση δεν είναι πια δυνατή, σε μερικές περιπτώσεις είναι δυνατή η τμηματοποίησή με στόχο την μετέπειτα ερμηνεία του κάθε μέρους. Τα μοντέλα που ανήκουν σε αυτήν την κατηγορία χαρακτηρίζονται από *decomposability*. Τέλος, τα *εγγενώς ερμηνεύσιμα* συστήματα τα οποία δεν είναι *simulatable* ή *decomposable*, μπορούν να γίνουν κατανοητά μέσω της *algorithmic transparency* ιδιότητας. Η ιδιότητα αυτή προσδίδει την πιο ασθενή μορφή εγγενούς ερμηνευσιμότητας και με βάση αυτή, η λειτουργία ενός μοντέλου μπορεί να επεξηγηθεί μέσω κατανόησης του αλγορίθμου που χρησιμοποιείται από αυτό. Χαρακτηριστικό παράδειγμα *algorithmic transparent* συστημάτων είναι τα *linear/logical regression* μοντέλα. Μερικά ακόμη δημοφιλή διαφανή μοντέλα μηχανικής μάθησης αποτελούν τα δέντρα αποφάσεων (*decision trees*), οι *rule-based learners*, τα *Bayesian* μοντέλα και αλγοριθμικές προσεγγίσεις κοντινότερων γειτόνων, όπως ο αλγόριθμος *k-nearest neighbors (k-NN)*. Παρόλα αυτά, πολλά ευρέως χρησιμοποιούμενα *ML* συστήματα δεν είναι εγγενώς ερμηνεύσιμα, λόγω της υψηλής περιπλοκότητάς και των πολλών παραμέτρων τους. Συγκεκριμένα, στα αδιαφανή μοντέλα κατατάσσονται συνήθως τα *tree ensembles*, τα *support vector machines* (για τις πιο κοινές συναρτήσεις πυρήνα), καθώς και μοντέλα βαθιάς μάθησης, όπως τα πολυεπίπεδα, τα συνελκτικά και τα αναδρομικά νευρωνικά δίκτυα. Για την πληρέστερη κατανόηση των προαναφερθέντων συστημάτων μπορεί να γίνει χρήση *post-hoc* μεθόδων ερμηνευσιμότητας, που εφαρμόζονται μετά την κατασκευή του υπό μελέτη δικτύου. [56]

### 4.2.2 Model-specific και model-agnostic μέθοδοι

Ένα από τα κυρίαρχα στη βιβλιογραφία, κριτήρια ταξινόμησης τεχνικών ερμηνευσιμότητας αποτελεί το κριτήριο της εφαρμοσιμότητας. Μέθοδοι που μπορούν να χρησιμοποιηθούν σε οποιοδήποτε μοντέλο μηχανικής μάθησης, ανεξάρτητα από την αρχιτεκτονική του, υπάγονται στην *model-agnostic* κατηγορία. Για να εφαρμοστεί κάποια *model-agnostic* τεχνική απαιτείται γνώση μόνο των δεδομένων εισόδου και εξόδου. Αντίθετα, κάθε μέθοδος ερμηνευσιμότητας που ανήκει στη λεγόμενη *model-specific* κατηγορία, εφαρμόζεται σε συγκεκριμένο τύπο *ML* συστημάτων ή υπό συγκεκριμένες συνθήκες. Ενδεικτικά, αναφέρουμε τις δημοφιλείς *model-specific* τεχνικές ερμηνευσιμότητας *Integrated Gradients*, *SmoothGrad* και *Grad-CAM*, οι οποίες οπτικοποιούν την αξία που προσδίδεται από μοντέλα βαθιάς μάθησης σε κάθε διαφορετικό χαρακτηριστικό εισόδου κατά τις προβλέψεις τους. Για παράδειγμα, η *Grad-CAM* τεχνική μπορεί να χρησιμοποιηθεί για τη διάκριση των εικονοστοιχείων στα οποία δίνει τη μεγαλύτερη

έμφαση ένα συνελκτικό νευρωνικό δίκτυο κατά τις προβλέψεις του. [55] Αντίστοιχα, σημαντικές model-agnostic προσεγγίσεις αποτελούν οι τεχνικές LIME (Local Interpretable Model-agnostic Explanations) και SHAP (Shapley Additive exPlanations). [56]



Σχήμα 4.3: Ενδεικτική χρήση της model-specific Grad-CAM τεχνικής για την επεξήγηση CNN ταξινομητή δύο κλάσεων (γάτες και σκύλοι) [57]

### 4.2.3 Τοπικές (local) και καθολικές (global) μέθοδοι

Τέλος, ένας ακόμη δυνατός τρόπος κατηγοριοποίησης τεχνικών ερμηνευσιμότητας αποτελεί ο διαχωρισμός τους σε *τοπικές* (local) και *καθολικές* (global). Συγκεκριμένα, οι καθολικές μέθοδοι επιχειρούν να κατανοήσουν τον τρόπο με τον οποίο το υπό μελέτη μοντέλο πραγματοποιεί προβλέψεις, λαμβάνοντας ολιστικά υπόψιν τα χαρακτηριστικά και τις εκπαιδευσιμες παραμέτρους του. Έτσι, οι παραγόμενες καθολικές εξηγήσεις είναι εφαρμόσιμες σε όλο το σύνολο δεδομένων εισόδου. Αντίθετα, οι τοπικές τεχνικές ερμηνευσιμότητας χρησιμοποιούνται για την κατανόηση του τρόπου λειτουργίας του ML συστήματος για μεμονωμένα δείγματα ή υποομάδες του συνόλου εισόδου. Τοπικά, η συμπεριφορά του μοντέλου είναι συχνά απλούστερη (ακόμα και γραμμική), οπότε είναι δυνατή η πραγματοποίηση παραδοχών που οδηγούν σε ακριβέστερα συμπεράσματα για το εξεταζόμενο τμήμα δεδομένων. Γενικά, οι τοπικές μέθοδοι εφαρμόζονται πιο εύκολα σε αδιαφανή, μη-εγγενώς ερμηνεύσιμα μοντέλα, ενώ οι καθολικές σε διαφανή, εγγενώς ερμηνεύσιμα συστήματα. [54], [58]

## 4.3 Post-hoc μέθοδοι ερμηνευσιμότητας: LIME και SHAP

### 4.3.1 Local Interpretable Model-agnostic Explanations (LIME)

Η μέθοδος LIME (Local Interpretable Model-agnostic Explanations) [59] αποτελεί μία model-agnostic, post-hoc μέθοδο ερμηνευσιμότητας που παρέχει τοπικές (local) εξηγήσεις. Στόχος είναι η επεξήγηση του black-box μοντέλου  $f$  για μεμονωμένα δείγματα του συνόλου εισόδου  $x$ , με χρήση ενός υποκατάστατου, ερμηνεύσιμου μοντέλου  $g$ . Πιο συγκεκριμένα, στόχος είναι η παραγωγή των εξηγήσεων  $\xi(x)$  μέσω της παρακάτω σχέσης:

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \mathcal{L}(f, g, \pi_x) + \Omega(g) \quad (4.1)$$

Δηλαδή, με βάση τη LIME τεχνική, για την επίτευξη ερμηνευσιμότητας και τοπικής πιστότητας, αναζητείται το ερμηνεύσιμο μοντέλο  $g$  για το οποίο ελαχιστοποιούνται οι

ποσότητες  $\mathcal{L}(f, g, \pi_x)$  και  $\Omega(g)$ . Σημειώνουμε ότι η ποσότητα  $\mathcal{L}(f, g, \pi_x)$  αποτυπώνει την απόκλιση του μοντέλου  $g$  από το εξεταζόμενο  $f$ , για τη γειτονιά γύρω από το σημείο ενδιαφέροντος  $x$  (έστω  $\pi_x$ ), ενώ η ποσότητα  $\Omega(g)$  αποτελεί μέτρο της πολυπλοκότητας του  $g$ . Το  $\Omega(g)$  προκύπτει ανάλογα με το είδος του χρησιμοποιούμενου ερμηνεύσιμου μοντέλου. Για παράδειγμα, στην περίπτωση όπου αυτό ανήκει στην κλάση των δέντρων απόφασης, το προαναφερθέν μέγεθος εξαρτάται από το βάθος του χρησιμοποιούμενου δέντρου. Ανίσοιχα, για τον υπολογισμό του  $\mathcal{L}(f, g, \pi_x)$  στο αρχικό paper χρησιμοποιείται η σχέση:

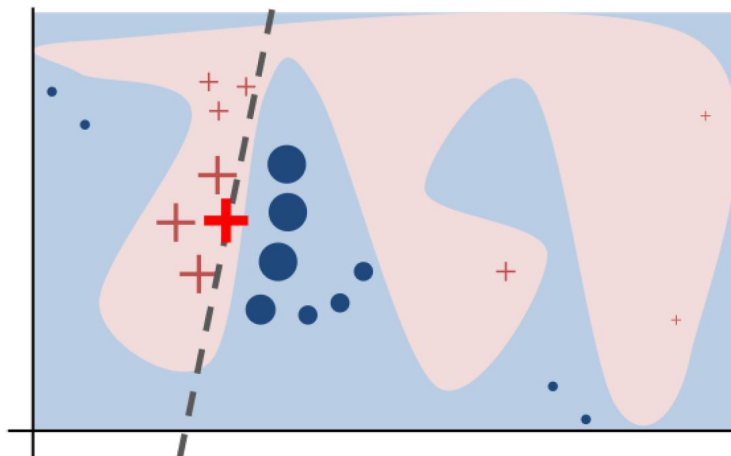
$$\mathcal{L}(f, g, \pi_x) = \sum_{z, z' \in \mathcal{Z}} \pi_x(z)(f(z) - g(z'))^2, \quad (4.2)$$

όπου  $\mathcal{Z}$  σύνολο που περιέχει σημεία κοντά στο σημείο ενδιαφέροντος  $x$ ,  $\pi_x(z) = \exp(-D(x, z)^2 / \sigma^2)$  και  $D$  κάποια συνάρτηση απόστασης.

Έτσι, ο αλγόριθμος που ακολουθεί η μέθοδος LIME για την προσαρμογή του απλού μοντέλου γύρω από το εξεταζόμενο δείγμα  $x$  είναι ο κάτωθι:

1. Λήψη του δείγματος  $x$  και παραγωγή νέων "διαταραγμένων" δειγμάτων με βάση το σύνολο εκπαίδευσης.
2. Πρόβλεψη των πιθανοτήτων για όλα τα διαταραγμένα δείγματα χρησιμοποιώντας το σύνθετο/γενικό μοντέλο.
3. Υπολογισμός της απόστασης του δείγματος  $x$  από όλα τα διαταραγμένα δείγματα και μετατροπή αυτής της απόστασης σε βαθμό ομοιότητας.
4. Επιλογή των  $m$  χαρακτηριστικών των διαταραγμένων δεδομένων που περιγράφουν καλύτερα το αποτέλεσμα του σύνθετου μοντέλου.
5. Υπολογισμός του απλού μοντέλου με βάση τις προβλέψεις του σύνθετου μοντέλου, αξιοποιώντας τους βαθμούς ομοιότητας των διαταραγμένων δεδομένων και τα  $m$  επιλεγμένα χαρακτηριστικά τους.
6. Εξαγωγή και χρήση των βαρών χαρακτηριστικών του απλού μοντέλου για την επεξήγηση της συμπεριφοράς του σύνθετου μοντέλου τοπικά, στην περιοχή  $\pi_x$ .

Από τα παραπάνω παρατηρούμε ότι υπάρχει μεγάλο περιθώριο βελτιστοποίησης της παραγόμενης εξήγησης. Τα πιο σημαντικά μεταξύ των επιλογών που επηρεάζουν την ποιότητα της εξήγησης είναι ο τρόπος δημιουργίας των διαταραγμένων δεδομένων, ο τρόπος υπολογισμού του βαθμού ομοιότητας, ο αριθμός και η μέθοδος επιλογής των  $m$  χαρακτηριστικών και τέλος το εφαρμοζόμενο, απλό μοντέλο  $g$ .



Σχήμα 4.4: Η πολύπλοκη, άγνωστη συνάρτηση απόφασης  $f$  (που αναπαρίσταιται από τα μπλε/ροζ χρώματα) δεν μπορεί να περιγραφεί από κάποιο γραμμικό μοντέλο. Υπολογίζονται τα  $f(z)$  για τα  $z$  κοντά στην περιοχή του σημείου ενδιαφέροντος (έντονος κόκκινος σταυρός) και εκτιμάται το γραμμικό μοντέλο (διακεκομμένη γραμμή) που εξηγεί τη συμπεριφορά του  $f$  τοπικά [59]

### 4.3.2 SHapley Additive exPlanation (SHAP) Values

Η μέθοδος SHAP (SHapley Additive exPlanation) παρουσιάστηκε για πρώτη φορά το 2017 [60] και κάνει χρήση των λεγόμενων *Shapley* τιμών από την θεωρία των παιγνίων, για την εκτίμηση της συνεισφοράς κάθε χαρακτηριστικού κατά την πραγματοποίηση προβλέψεων. Όπως και τις υπόλοιπες μεθόδους προσθετικής συμβολής των χαρακτηριστικών, θεωρείται ότι το αποτέλεσμα μιας πρόβλεψης εκφράζεται σαν μια γραμμική συνάρτηση των χαρακτηριστικών. Έτσι, εστω ότι  $f$  είναι το αρχικό, σύνθετο μοντέλο,  $g$  το απλό, ερμηνεύσιμο μοντέλο και  $x$  μια είσοδος. Ακόμα έστω ότι το ερμηνεύσιμο μοντέλο χρησιμοποιεί μία mapping συνάρτηση  $x = h_x(x')$  για να δημιουργήσει μία νέα αναπαράσταση δεδομένων, που είναι κατανοητή από το  $g$ . Τότε η γραμμική συνάρτηση της εξόδου του ερμηνεύσιμου μοντέλου  $g$  θα είναι η:

$$g(x') = \varphi_0 + \sum_{i=1}^M \varphi_i x'_i \quad (4.3)$$

όπου  $M$  το πλήθος των χαρακτηριστικών,  $x' \in \{0, 1\}^M$ ,  $g(x') \approx f(h_x(x'))$  και  $\varphi_i$  πραγματικοί αριθμοί. Είναι επιθυμητή ακόμα η ικανοποίηση των παρακάτω ιδιοτήτων:

1. Τοπική Ακρίβεια (Local Accuracy): η ιδιότητα αυτή απαιτεί από την μέθοδο εξήγησης  $g$  με είσοδο  $x'$  να έχει την ίδια έξοδο με την  $f$  (για είσοδο  $x$ ), δηλαδή απαιτεί να ισχύει:  $g(x') = f(x)$ .
2. Απουσία (Missingness): η ιδιότητα missingness ορίζει ότι η συμβολή ενός χαρακτηριστικού που δεν περιλαμβάνεται πρέπει να είναι μηδενική:  $x'_i = 0 \Rightarrow \varphi_i = 0$ .
3. Συνέπεια (Consistency): με βάση αυτήν την ιδιότητα, αν η συμβολή ενός απλουστευμένου χαρακτηριστικού  $x'_i$  αυξηθεί ή παραμείνει η ίδια ανεξάρτητα από τα υπόλοιπα χαρακτηριστικά, η τιμή  $\varphi_i$  δεν μπορεί να μειωθεί.



Αποδεικνύεται ότι η μοναδική λύση δίνεται από τις τιμές Shapley, που προέρχονται από την Συνεργατική Θεωρία Παιγνίων και περιγράφουν τη συνεισφορά κάθε παίκτη σε ένα εξεταζόμενο τελικό αποτέλεσμα. Οι τιμές αυτές υπολογίζονται μέσω της παρακάτω σχέσης:

$$\phi_i(v) = \sum_{S \subset N \setminus \{i\}} \frac{|S|!(n - |S|)!}{n!} (v(S \cup \{i\}) - v(S)) \quad (4.4)$$

Οι τιμές Shapley συνδυάζονται με την επεξήγηση μοντέλων μηχανικής μάθησης αν ως παίκτης θεωρηθούν τα χαρακτηριστικά του συνόλου δεδομένων, η συνάρτηση κέρδους είναι το μοντέλο της μηχανικής μάθησης που εξετάζεται και μια πρόβλεψη του μοντέλου αποτελεί το τελικό αποτέλεσμα. Έτσι οι τιμές Shapley φανερώνουν τη συμβολή κάθε χαρακτηριστικού, και άρα και την εξήγηση για μια συγκεκριμένη απόφαση του μοντέλου:

$$\phi_i(v) = \sum_{S \in F \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} (f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)) \quad (4.5)$$

όπου  $M$  είναι το πλήθος των χαρακτηριστικών,  $S$  ένα υποσύνολο χαρακτηριστικών,  $f_{S \cup \{i\}}(x_{S \cup \{i\}})$  η έξοδος του μοντέλου όταν χρησιμοποιείται το  $S$  υποσύνολο χαρακτηριστικών μαζί με το  $i$  χαρακτηριστικό του οποίου η Shapley τιμή αναζητείται και  $f_S(x_S)$  η έξοδος του μοντέλου όταν χρησιμοποιείται μόνο το  $S$  υποσύνολο χαρακτηριστικών, χωρίς το  $i$ .

## Κεφάλαιο 5

### Βιβλιογραφική Επισκόπηση

Με στόχο τη βαθύτερη κατανόηση και την εξοικείωση με το ερευνητικό πρόβλημα που αντιμετωπίζεται, πραγματοποιήθηκε μελέτη της υπάρχουσας, σχετιζόμενης βιβλιογραφίας. Η ραγδαία εξάπλωση του σακχαρώδους διαβήτη και η πολυπλοκότητα της ασθένειας, σε συνδυασμό με την ταχεία ανάπτυξη των υπολογιστικών μεθόδων και τη μεγαλύτερη διαθεσιμότητα σχετικών δεδομένων, έχουν στρέψει το ενδιαφέρον πολλών ερευνητών στο συγκεκριμένο πεδίο, καθιστώντας την εν λόγω βιβλιογραφία πλούσια και πολυδιάστατη. Στο πλαίσιο αυτής της επισκόπησης εστιάζουμε σε μελέτες του πρόσφατου παρελθόντος, οι οποίες εφαρμόζουν τεχνικές μηχανικής μάθησης, έναντι άλλων κλασικών στατιστικών μεθόδων. Καθώς ενδιαφερόμαστε για την εκτίμηση του κινδύνου εμφάνισης υπογλυκαιμικών και υπεργλυκαιμικών επεισοδίων, επικεντρώναμε στα ML μοντέλα που προβλέπουν τα επίπεδα γλυκόζης του αίματος και σχετιζόμενα με αυτά γεγονότα (υπογλυκαιμία/υπεργλυκαιμία). Παρόλα αυτά, στη βιβλιογραφία συναντώνται ακόμα πολλές μελέτες που έχουν υλοποιηθεί με στόχο την αξιολόγηση του κινδύνου εμφάνισης [61], [62], τη διάγνωση [63], τη πρόγνωση [64], [65], [66], [67], [68] και την διαχείριση της νόσου [69], [70], [71], [72], [73]. Επιδίωξη μας είναι να διακρίνουμε τις σύγχρονες τάσεις όσον αφορά τους χρησιμοποιούμενους αλγόριθμους και τις τεχνικές εκπαίδευσης, βελτιστοποίησης και αξιολόγησης. Ειδική μνεία γίνεται ακόμα σε μοντέλα προσομοίωσης φυσιολογικών συστημάτων που μπορούν να ενσωματωθούν σε συστήματα πρόβλεψης, καθώς και σε περιπτώσεις μελετών που περιέχουν μεθόδους ερμηνευσιμότητας.

#### 5.1 Μοντέλα προσομοίωσης φυσιολογικών συστημάτων

Τις τελευταίες δεκαετίες, αρκετοί συγγραφείς έχουν προτείνει μοντέλα δράσης της ινσουλίνης και της κινητικής γλυκόζης χρησιμοποιώντας πειραματικά δεδομένα για τη μέτρηση της παραγωγής γλυκόζης, της χρησιμοποίησης γλυκόζης και της απορρόφησης της ινσουλίνης και των γευμάτων. Πολλά από αυτά τα μοντέλα είναι μοντέλα διαμερισμάτων, τα οποία εφαρμόζονται με κυρίαρχο στόχο την δυναμική μελέτη της κινητικής ουσιών σε φυσιολογικά συστήματα. Οι πιο δημοφιλείς προτάσεις για την προσομοίωση της δράσης και της κινητικής της ινσουλίνης και της γλυκόζης είναι το Dalla Man [74] μοντέλο, το Hovorka [75] και το Bergman [76] μοντέλο. Τα διάφορα αυτά συστήματα επιτρέπουν τον υπολογισμό διαφορετικών παραμέτρων, όπως η υποδόρια απορρόφηση της ινσουλίνης, ο ρυθμός γαστρικής εκκένωσης, η πέψη και η

απορρόφηση υδατανθράκων, η κινητική της ινσουλίνης και ο μεταβολισμός της γλυκόζης. Αναλυτικότερα, το Dalla Man μοντέλο συνίσταται από ένα υποσύνολο γλυκόζης και από ένα υποσύνολο ινσουλίνης τα οποία συνδέονται από τον έλεγχο της ινσουλίνης στη χρησιμοποίηση γλυκόζης και την ενδογενή παραγωγή της. Σε αντίθεση, το Bergman μοντέλο χρησιμοποιεί ένα διαμερισματικό μοντέλο τριών διαμερισμάτων για την αναπαράσταση των συγκεντρώσεων της ινσουλίνης του πλάσματος, της εξωγενούς ινσουλίνης και της γλυκόζης του πλάσματος. Τέλος, η προσέγγιση Hovorka κάνει χρήση ενός διαμερισματικού μοντέλου δύο διαμερισμάτων για την περιγραφή της απορρόφησης της υποδόριας ινσουλίνης και της απορρόφησης γλυκόζης από το έντερο. [77] Για την εκτίμηση της απορρόφησης γλυκόζης και σε κυτταρικό επίπεδο, είναι δυνατή η προσομοίωση της διαδικασίας ενεργοποίησης του ενδοκυτταρικού μονοπατιού μεταβίβασης σήματος ινσουλίνης. Ένα από τα πιο δημοφιλή τέτοια συστήματα προσομοίωσης είναι αυτό του Sedaghat κ.α. [78], το οποίο, με δεδομένη τη συγκέντρωση ινσουλίνης, και χρησιμοποιώντας ένα σύστημα 20 διαφορικών εξισώσεων, εκτιμά το ποσοστό GLUT4 μεταφορέα στο κυτταρικό τοίχωμα. Το μοντέλο αυτό έχει επεκταθεί και εμπλουτιστεί και με νεότερες, σχετικές μελέτες. [79] Μια ακόμα έρευνα που μελετά την μετανάστευση του GLUT4 μεταφορέα από την ενδοκυττάρια δεξαμενή όπου αρχικά εντοπίζεται στην κυτταρική μεμβράνη, με βάση την ινσουλίνη στην αιματική κυκλοφορία, είναι η [80]. Η μελέτη αυτή αναπτύσσει διάφορα πιθανά μοντέλα και καταλήγει σε αυτά τα οποία συμφωνούν στο βέλτιστο βαθμό με τα διαθέσιμα, πραγματικά, πειραματικά δεδομένα. Πολλά μοντέλα πρόβλεψης [81], [82] επιστρατεύουν τεχνικές προσομοίωσης φυσιολογικών συστημάτων σαν και τις προαναφερθείσες, με στόχο να βελτιώσουν την ποιότητα των αποτελεσμάτων τους. Τα μοντέλα αυτά ονομάζονται υβριδικά (hybrid) και διαφοροποιούνται από τα λεγόμενα data-driven συστήματα πρόβλεψης, τα οποία δεν χρησιμοποιούν physiological modelling.

## 5.2 Μοντέλα πρόβλεψης επιπέδων γλυκόζης

Σε μελέτη του 2020, ο Kim κ.α. [83] ανέπτυξαν ένα εξατομικευμένο μοντέλο πρόβλεψης γλυκόζης με χρήση αναδρομικών νευρωνικών δικτύων. Τα διαθέσιμα δεδομένα (CGM μετρήσεις 20 υπο νοσηλεία ασθενών με ΣΔ τύπου 2), χρησιμοποιήθηκαν για την εκπαίδευση μεμονωμένων μοντέλων για κάθε νοσούντα, με στόχο τη μεγαλύτερη εξατομίκευση των αποτελεσμάτων. Πραγματοποιήθηκαν πειράματα με απλά RNN, LSTM και GRU δίκτυα και τις bi-directional εκδοχές τους, από όπου τελικά επιλέχθηκε ένα μοντέλο με ένα GRU και δύο πυκνά επίπεδα ως το βέλτιστο. Στόχος ήταν η εκτίμηση της τιμής γλυκόζης 30 λεπτά έπειτα από ένα κυλιόμενο διάστημα 7 παρελθοντικών καταγραφών, που χρησιμοποιείται κάθε φορά ως είσοδος για την πρόβλεψη. Τέλος, για την αξιολόγηση εφαρμόστηκαν οι Root Mean Square Error (RMSE) και Root Mean Square Error (RMSE) μετρικές, αλλά και η Clarke Error Grid ανάλυση.

Ο Sun κ.α. το 2018 [84] χρησιμοποίησαν ένα Autoregressive integrated moving average (ARIMA), ένα Support Vector Machine (SVM) και ένα τεχνητό νευρωνικό μοντέλο για την πρόβλεψη των επιπέδων γλυκόζης έπειτα από 15, 30, 45 και 60 λεπτά. Για την εκπαίδευση των παραπάνω μοντέλων αξιοποιήθηκαν CGM καταγραφές που αντιστοιχούσαν σε 20 πραγματικούς ασθενείς με ΣΔ τύπου 1, και 11 εικονικούς ασθενείς του UVA/PADOVA προσομοιωτή. Το κάθε μοντέλο, για την εκπαίδευσή του, τροφοδοτήθηκε με δεδομένα από όλους τους ασθενείς, έτσι ώστε να προκύψουν τρία γενικευμένα (global) μοντέλα πρόβλεψης. Το νευρωνικό δίκτυο, το οποίο αποτελείτο από ένα bi-directional και ένα απλό LSTM επίπεδο, σε συνδυασμό με κάποια πλήρως-συνδεδεμένα

επίπεδα, εμφάνισε τις καλύτερες αποδόσεις για όλους τους οριζόντες πρόβλεψης. Σημειώνουμε ότι ως μετρικές αξιολόγησης χρησιμοποιήθηκαν οι Root mean square error (RMSE), correlation coefficient (CC), time lag (TL) και fit.

Ο Tena κ.α. σε έρευνα του 2021 [85] επίσης εξέτασαν ML δίκτυα πρόβλεψης των επιπέδων γλυκόζης, πραγματοποιώντας μια συνολική σύγκριση μοντέλων που είχαν υλοποιηθεί σε προγενέστερες μελέτες. Αναλυτικότερα, συγκρίνονται μοντέλα αναδρομικών νευρωνικών δικτύων σε διάφορες παραλλαγές και δομές [84], [86], [87], [89], [90], [91], [92], [93], ένα συνελκτικό νευρωνικό δίκτυο CNN [88] και μοντέλα συλλογικής μάθησης, με base learners δίκτυα αποτελούμενα από LSTM, dense και Partial Least Square Regression (PLSR) επίπεδα [94]. Η εκπαίδευση και το testing των συστημάτων έγινε με 10-fold cross-validation υπό τις ίδιες συνθήκες, με χρήση δεδομένων από το OhioT1DM dataset. Το συγκεκριμένο dataset περιέχει CGM καταγραφές γλυκόζης και πληροφορίες για την έγχυση ινσουλίνης και την πρόσληψη υδατανθράκων νοσούτων με ΣΔ τύπου 1. Ως είσοδος χρησιμοποιήθηκαν χρονοσειρές 24 τιμών από όλα τα παραπάνω χαρακτηριστικά, ενώ η πρόβλεψη γλυκόζης έγινε για οριζόντες πρόβλεψης 30, 60 και 120 λεπτά. Η αξιολόγηση και η σύγκριση των μοντέλων πραγματοποιήθηκε με μετρικές όπως οι RMSE και CC και μεθόδους όπως η Parkes error grid. Τελικά, το LSTM [86], το bi-directional LSTM [84] και τα μοντέλα συλλογικής μάθησης [94] κρίνονται ως οι αποδοτικότερες επιλογές.

Σε μια ακόμη μελέτη ο Martinsson κ.α. [95] ανέπτυξαν ένα δίκτυο βασισμένο σε LSTMs για την πρόβλεψη των επιπέδων γλυκόζης. Χρησιμοποίησαν ως δεδομένα CGM μετρήσεις που αντιστοιχούσαν σε έξι ασθενείς του Ohio T1DM dataset. Στα τελικά πειράματα, ως είσοδος εφαρμόζονταν κάθε φορά καταγραφές που αντιστοιχούσαν σε διάρκεια μιας ώρας, ενώ η πρόβλεψη γινόταν για τα επόμενα 30 και 60 λεπτά. Ιδιαίτερο ενδιαφέρον έχει ο υπολογισμός της εκτιμώμενης απόκλισης για κάθε τιμή πρόβλεψης με βάση την έξοδο του τελευταίου πυκνού επιπέδου, με στόχο την εξαγωγή συμπερασμάτων για το πόσο βέβαιο είναι το μοντέλο για τα αποτελέσματά του. Διεξήχθησαν πειράματα με την Negative log-likelihood (NLL) και μια glucose-specific loss συνάρτηση, ενώ για την προσαρμογή των υπερπαραμέτρων χρησιμοποιήθηκε και η grid search τεχνική. Τέλος, η αξιολόγηση πραγματοποιήθηκε με την μετρική RMSE και την μέθοδο Surveillance Error Grid.

### **5.3 Μοντέλα πρόβλεψης υπογλυκαιμικών και υπεργλυκαιμικών επεισοδίων**

Στις προαναφερθείσες μελέτες η πρόβλεψη της γλυκόζης αντιμετωπίζεται ως ένα regression πρόβλημα, όπου κάθε φορά ζήτημα είναι η πρόβλεψη συνεχών, αριθμητικών τιμών. Χαρακτηρίζοντας τις τιμές γλυκόζης ανάλογα με το αν αυτές αντιστοιχούν σε υπογλυκαιμία, ευγλυκαιμία ή υπεργλυκαιμία, το παραπάνω πρόβλημα μπορεί να μετατραπεί σε πρόβλημα ταξινόμησης, όπου στόχος είναι η πρόβλεψη υπογλυκαιμικών και υπεργλυκαιμικών επεισοδίων αντίστοιχα. Στη συνέχεια εξετάζονται κάποιες έρευνες που μελετούν την συγκριμένη προσέγγιση.

Σε έρευνα του 2019, οι Seo et al. [96] χρησιμοποίησαν τεχνικές μάθησης για την πρόβλεψη μεταγευματικής υπογλυκαιμίας. Τέτοια περιστατικά θα μπορούσαν να παρατηρηθούν σε περιπτώσεις όπου το προσλαμβανόμενο γεύμα περιέχει περιορισμένη ποσότητα υδατανθράκων ή η αντίστοιχη προγευματική δόση ινσουλίνης χορηγείται λανθασμένα. Για την πρόβλεψη χρησιμοποιήθηκαν CGM δεδομένα 104 ασθενών που έχουν

βιώσει τουλάχιστον ένα υπογλυκαιμικό επεισόδιο σε περίοδο τριήμερης συνεχούς παρακολούθησης, ενώ ως χαρακτηριστικά εισόδου θεωρήθηκαν ο ρυθμός μεταβολής της γλυκόζης (Glucose Rate of Change, GRC) και ένα ακόμα χαρακτηριστικό που αποτύπωνε την διακύμανση των επιπέδων γλυκόζης έπειτα από ένα γνωστοποιημένο γεύμα (*rate of increase in glucose*). Ως ορίζοντας πρόβλεψης χρησιμοποιήθηκαν τα 30 λεπτά, ενώ πραγματοποιήθηκαν δοκιμές με τους αλγορίθμους λογιστικής παλινδρόμησης και kNN, καθώς και με Random Forest και Support Vector Machine μοντέλα. Η εκπαίδευση των δικτύων περιλάμβανε 5-fold cross-validation, τεχνικές αναζήτησης πλέγματος (grid search) και εξισορρόπησης δεδομένων. Από τις διαδικασίες αξιολόγησης και σύγκρισης που πραγματοποιήθηκαν με τη χρήση των μετρικών specificity, sensitivity, AUC και F1-score, αλλά και με στατιστικές μεθόδους, προέκυψε ότι το Random Forest δίκτυο παρουσίασε την βέλτιστη απόδοση.

Ο Witte κ.α. με μελέτη του 2022 [97] προσέγγισαν το εξεταζόμενο ερευνητικό ζήτημα ως πρόβλημα ταξινόμησης πολλών κλάσεων. Αναλυτικότερα, μία κλάση πρόβλεψης χρησιμοποιείται για την πρόβλεψη υπογλυκαιμικών περιστατικών, ενώ τρεις ακόμη εφαρμόζονται για την ταξινόμηση υπεργλυκαιμικών επεισοδίων ανάλογα με την κλινική τους σημασία (ήπια, μέτρια και σοβαρή υπεργλυκαιμία). Ως δεδομένα χρησιμοποιήθηκαν πληροφορίες από 38,250 ασθενείς που είχαν νοσηλευτεί τουλάχιστον μία φορά, λόγω κάποιου τύπου ΣΔ ή και γενικότερης παρατηρούμενης υπογλυκαιμίας ή υπεργλυκαιμίας. Τα χρησιμοποιούμενα χαρακτηριστικά εισόδου περιλάμβαναν στοιχεία όπως η μέση τιμή, η τυπική απόκλιση, η μέγιστη τιμή κ.α. αναλυτέων ουσιών από διαγνωστικές εξετάσεις, καθώς και δεδομένα όπως το βάρος και η ηλικία των ασθενών, τα οποία ήταν διαθέσιμα μέσω των αντίστοιχων ηλεκτρονικών φακέλων υγείας. Για την πρόβλεψη υλοποιήθηκαν XGBoost δυαδικοί ταξινομητές, ορισμένοι εκ των οποίων στη συνέχεια συνδυάστηκαν για τον σχηματισμό ενός συλλογικού ταξινομητή πολλαπλών κλάσεων. Οι μεμονωμένοι δυαδικοί ταξινομητές εκπαιδεύτηκαν με τη χρήση τεχνικών αναζήτησης πλέγματος και εξισορρόπησης δεδομένων, ενώ αξιολογήθηκαν με μετρικές όπως οι AUC, precision, sensitivity και informedness. Ο βέλτιστος συνδυασμός binary ταξινομητών για την παραγωγή του ensemble classifier προέκυψε με την εφαρμογή γενετικών αλγορίθμων, ανάλογα με την υπολογιζόμενη ακρίβεια ανά περίπτωση. Τέλος, αξίζει να σημειωθεί ότι κάθε φορά χρησιμοποιούνται παρελθοντικά δεδομένα 1.9, 1.0, 1.5 και 1.8 ημερών για την πρόβλεψη της υπογλυκαιμίας, της ήπιας, της μέτριας και της σοβαρής υπεργλυκαιμίας αντίστοιχα, ενώ η πρόβλεψη κάθε φορά γίνεται για τις επόμενες 7 ώρες για την υπογλυκαιμία, και για τις επόμενες 4 ώρες για όλες τις περιπτώσεις υπεργλυκαιμίας.

## **5.4 Τεχνικές ερμηνευσιμότητας σε μοντέλα πρόβλεψης**

Πρόσφατη μελέτη του 2022 [98], ενσωμάτωσε την SHapley Additive exPlanations (SHAP) τεχνική για μετέπειτα ερμηνεία των αποτελεσμάτων της. Η έρευνα είχε ως στόχο την πρόβλεψη υπογλυκαιμικών και υπεργλυκαιμικών επεισοδίων, χρησιμοποιώντας δεδομένα εφήβων και νεαρών ενηλίκων, νοσούντων με σακχαρώδη διαβήτη τύπου 1. Συγκεκριμένα, χρησιμοποιήθηκαν CGM μετρήσεις 153 ασθενών, οι οποίες είχαν ληφθεί επί χρονικά διαστήματα διάρκειας από 6 έως 12 μήνες. Από τα δεδομένα αυτά εξήχθησαν χαρακτηριστικά όπως η μέγιστη αύξηση, η μέγιστη μείωση και η μέση τιμή των τιμών γλυκόζης που παρατηρούνται εντός ενός οριζόμενου χρονικού διαστήματος, οι οποίες

σε συνδυασμό με την πιο πρόσφατη τιμή γλυκόζης και βασικές πληροφορίες όπως η ηλικία, χρησιμοποιούνται για την πρόβλεψη υπογλυκαιμίας ή υπεργλυκαιμίας εντός του επόμενου χρονικού διαστήματος διάρκειας 60 λεπτών. Για την αντιμετώπιση του προβλήματος γίνονται δοκιμές με ένα μοντέλο λογιστικής παλινδρόμησης, ένα baseline και ένα XGBoost μοντέλο. Έπειτα από αξιολόγησή τους με χρήση των μετρικών AUC, sensitivity, specificity και precision προκύπτει ότι το σύστημα XGBoost εμφανίζει τις καλύτερες επιδόσεις. Τέλος, η επιδραστικότητα του καθενός από τα συνολικά 30 χαρακτηριστικά εισόδου στις τελικές προβλέψεις, παρουσιάζεται μέσω ραβδογραμμμάτων με χρήση της SHAP μεθόδου, και για τις δύο περιπτώσεις, υπογλυκαιμίας και υπεργλυκαιμίας.

Η SHAP τεχνική εφαρμόστηκε σε μια ακόμα έρευνα του 2020 [99], όπου γίνεται πρόβλεψη των επιπέδων γλυκόζης του αίματος έπειτα από 30 και 60 λεπτά. Για τη μελέτη χρησιμοποιήθηκαν δεδομένα 6 ασθενών με ΣΔ τύπου 2 από το OhioT1DM Dataset. Τα δεδομένα αυτά περιλαμβάναν CGM μετρήσεις γλυκόζης, πληροφορίες για την χορήγηση ινσουλίνης, τα προσλαμβανόμενα γεύματα και τη φυσική δραστηριότητα των ασθενών. Αξίζει να σημειωθεί ότι τα παραπάνω data μετασχηματίστηκαν μέσω φίλτρων για καλύτερη προσομοίωση των συντελούμενων φυσιολογικών διεργασιών. Όσον αφορά το εφαρμοζόμενο μοντέλο μηχανικής μάθησης, αυτό αποτελείτο από ένα bi-directional και δύο απλά LSTM επίπεδα, καθώς και από ένα πλήρως συνδεδεμένο επίπεδο. Η εκπαίδευση του μοντέλου έγινε για διάφορα υποσύνολα των χαρακτηριστικών εισόδου, με στόχο την διάκριση των πιο χρήσιμων εξ αυτών, ενώ η τελική αξιολόγηση του δικτύου πραγματοποιήθηκε με τις μετρικές RMSE, Mean Absolute Error (MAE), Time Gain και delay. Τέλος, τα αποτελέσματα της SHAP τεχνικής παρουσιάστηκαν με τη μορφή beeswarm διαγραμμμάτων.

## **Κεφάλαιο 6**

# **Ανάπτυξη ερμηνεύσιμου μοντέλου εκτίμησης του κινδύνου εμφάνισης υπεργλυκαιμικών και υπογλυκαιμικών επεισοδίων**

Τα προηγούμενα κεφάλαια είχαν ως στόχο την πληροφόρηση του αναγνώστη σχετικά με το βασικό θεωρητικό υπόβαθρο των εργαλείων που χρησιμοποιούνται κατά την πρακτική εφαρμογή της παρούσας εργασίας. Στη συνέχεια ακολουθεί η παρουσίαση του πειραματικού σκέλους. Στην περίπτωση μας εξετάζεται ένα πρόβλημα ταξινόμησης, όπου με είσοδο τιμές γλυκόζης και άλλων χαρακτηριστικών, εκτιμάται η πιθανότητα εμφάνισης υπογλυκαιμικών και υπεργλυκαιμικών επεισοδίων στο κοντινό μέλλον. Στο παρόν κεφάλαιο περιγράφονται: α) τα διαθέσιμα δεδομένα και οι τεχνικές που εφαρμόστηκαν για την επεξεργασία τους και β) το μοντέλο και οι μέθοδοι που χρησιμοποιήθηκαν για την πραγματοποίηση εκτιμήσεων και την ερμηνεία των εξαγόμενων αποτελεσμάτων. Σημειώνουμε ότι για την υλοποίηση έγινε χρήση της γλώσσας Python και συγκεκριμένα η 3.8.10 έκδοσή της. Επιμέρους βιβλιοθήκες και πακέτα όπως τα Keras, scikit-learn, NumPy, pandas, GEKKO κ.α. χρησιμοποιήθηκαν επίσης για τον σχηματισμό των συνόλων εκπαίδευσης, καθώς και για την ανάπτυξη και αξιολόγηση του εφαρμοζόμενου μοντέλου.

### **6.1 Δεδομένα και επεξεργασία**

#### **6.1.1 Τα αρχικά δεδομένα**

Για την ανάπτυξη και την αξιολόγηση του μοντέλου χρησιμοποιήθηκαν δεδομένα πραγματικών ασθενών με ΣΔ τύπου 1 που συλλέχθηκαν στο πλαίσιο του ερευνητικού προγράμματος ENDORSE (Τ1ΕΚ-03695) [100], [101] και περιελάμβαναν μετρήσεις γλυκόζης, τιμές χορηγημένης ινσουλίνης, καθώς και πληροφορίες λαμβανόμενων γευμάτων και σωματικής δραστηριότητας. Συγκεκριμένα, τα αρχικά δεδομένα ανήκαν σε συνολικά 48 ασθενείς, 17 εκ των οποίων στη συνέχεια απορρίφθηκαν λόγω ελλειπών πληροφοριών μετρήσεων γλυκόζης, με αποτέλεσμα τα τελικά δεδομένα να προέρχονται από 31 διαφορετικά άτομα.

Τα ανεπεξέργαστα (raw) δεδομένα ενός ασθενούς αποτελούνται από πλήθος γραμμών (καταγραφών), κάθε μία εκ των οποίων αντιστοιχεί σε μια χρονική στιγμή που αποτυπώνεται μέσω ενός Unix timestamp. Η κάθε γραμμή δύναται να διαθέτει τιμές που αντιστοιχούν σε μέτρηση γλυκόζης, πρόσληψη υδατανθράκων, χορήγηση ενέσιμης ινσουλίνης, καθώς και σε καταγραφή φυσικής δραστηριότητας με τη μορφή βημάτων. Σε περίπτωση που κάποια από τις προαναφερθείσες πληροφορίες δεν είναι διαθέσιμη για τη συγκεκριμένη χρονική στιγμή, αυτό δήλωναται μέσω μηδενικής τιμής.

Η καταγραφή της γλυκόζης του αίματος πραγματοποιήθηκε από τους ασθενείς με χρήση της εφαρμογής FreeStyle LibreLink, μέσω της οποίας είναι δυνατή η σάρωση τοποθετημένου αισθητήρα με χρήση κινητού τηλεφώνου, για την καταγραφή των τρέχοντων επιπέδων γλυκόζης. Ο αντίστοιχος αισθητήρας διαθέτει μικρό μέγεθος και εφαρμόζεται στο πίσω μέρος του άνω βραχίονα για έως και 14 ημέρες. Μέσω της παραπάνω διαδικασίας η ανάγκη για μέτρηση της γλυκόζης μέσω τρυπήματος στα δάκτυλα περιορίζεται σημαντικά, αν δεν εξαλείφεται πλήρως. Ακόμα, όσον αφορά την χορήγηση ινσουλίνης, αυτή πραγματοποιήθηκε με χρήση του Humalog Junior KwikPen, μιας συσκευής τύπου πέννας, προγεμισμένης με ενέσιμο διάλυμα ινσουλίνης με ταχεία δράση και συγκέντρωση 100 μονάδες/ml. Τέλος, η καταγραφή φυσικής δραστηριότητας συντελέστηκε μέσω smartwatch τύπου fitbit. Οι συσκευές αυτές, ανάλογα με τον τύπο τους, συμβάλλουν στην παρακολούθηση της φυσικής κατάστασης αυτού που τις φέρει, μέσω καταγραφής παραμέτρων όπως το πλήθος βημάτων και καρδιακών παλμών που σημειώνονται, το πλήθος θερμίδων που καταναλώνονται, αλλά και μέσω παρακολούθησης του ύπνου και των αναπνοών.

Στην περίπτωση μας θεωρήσαμε ωφέλιμη τη συμπερίληψη όλων των διαθέσιμων δεδομένων γλυκόζης, υδατανθράκων, βημάτων και ινσουλίνης, ως χαρακτηριστικών εισόδου, καθώς από τις υποενότητες 2.2.2 και 2.4.2 έχει ήδη γίνει φανερή η μεγάλη επιδραστικότητα της διατροφής, της σωματικής δραστηριότητας και της ορμόνης ινσουλίνης στη διαμόρφωση των επιπέδων γλυκόζης του αίματος και στη γενικότερη εξέλιξη του σακχαρώδη διαβήτη.

### **6.1.2 Σχηματισμός και labeling χρονοσειρών γλυκόζης**

Καθώς οι διαθέσιμες καταγραφές γλυκόζης έχουν προκύψει μέσω σάρωση αισθητήρα από τους ίδιους τους ασθενείς και αποτελούν αποτέλεσμα Intermittently Scanned Continuous Glucose Monitoring (isCGM), είναι σημαντική η εύρεση μετρήσεων που έχουν πραγματοποιηθεί με μικρές χρονικές αποστάσεις και θα μπορούσαν να χρησιμοποιηθούν για τον σχηματισμό χρονοσειρών γλυκόζης. Επιλέγουμε να σχηματίσουμε χρονοσειρές με μέση διάρκεια 1 ώρας, οπότε διατρέχουμε τον χώρο των δεδομένων, και με την λογική ενός κυλιόμενου παραθύρου, ομαδοποιούμε τις καταγραφές γλυκόζης που έχουν πραγματοποιηθεί εντός της ίδιας μίας ώρας. Καθώς επιθυμούμε οι παραγόμενες χρονοσειρές γλυκόζης να διαθέτουν το ίδιο μήκος, ανεξαρτήτως από το πλήθος μετρήσεων που έχει πραγματοποιήσει ένας ασθενής εντός της αντίστοιχης περιόδου, πράττουμε το εξής: για κάθε χρονοσειρά γλυκόζης που αντιπροσωπεύει ένα διάστημα  $[t_{start}, t_{end}]$ , εξετάζουμε τα υποδιαστήματα  $[t_{start} + i, t_{start} + i + 15min]$  για  $i = 0, 15 min, 30 min, 45 min$  και σε περίπτωση όπου για κάποιο από αυτά τα υποδιαστήματα ο ασθενής έχει καταγράψει πάνω από μία τιμή γλυκόζης, υπολογίζεται η μέση τιμή των μετρήσεων αυτών. Τελικά, μέσω αυτής της διαδικασίας, έχουμε σχηματίσει χρονοσειρές γλυκόζης με μήκος 4 μετρήσεις, που περιγράφουν το γλυκαιμικό προφίλ ενός ασθενή για περίοδο 1 ώρας, με βάση καταγραφές που έχει πραγματοποιήσει ο ίδιος.



Για μια χρονοσειρά πρόβλεψης που περιέχει μετρήσεις γλυκόζης που έχουν πραγματοποιηθεί εντός ενός διαστήματος  $[t_{start}, t_{end}]$ , στόχος μας είναι η εκτίμηση της πιθανότητας εμφάνισης υπογλυκαιμικών και υπεργλυκαιμικών επεισοδίων εντός των διαστημάτων  $[t_{end}, t_{end} + 30min]$  και  $[t_{end} + 30min, t_{end} + 60min]$ . Λέμε επομένως ότι στην περίπτωση μας, οι εξεταζόμενοι *ορίζοντες πρόβλεψης* (prediction horizons, PH) είναι τα 0-30 και 30-60 λεπτά. Για να χαρακτηριστεί η προαναφερθείσα χρονοσειρά ως θετική για την περίπτωση της υπογλυκαιμίας για τον ορίζοντα πρόβλεψης 0-30 λεπτών, θα πρέπει να έχει καταγραφεί μέσω σάρωσης μία υπογλυκαιμική τιμή γλυκόζης εντός του  $[t_{end}, t_{end} + 30min]$  διαστήματος, ενώ για να χαρακτηριστεί ως θετική για την περίπτωση της υπεργλυκαιμίας, αρκεί να έχει σημειωθεί κάποια υπεργλυκαιμική τιμή εντός του ίδιου διαστήματος. Αντίστοιχα πράττουμε για τον ορίζοντα πρόβλεψης 30 έως 60 λεπτά. Σημειώνουμε ότι θεωρούμε υπογλυκαιμικές τις τιμές γλυκόζης που είναι μικρότερες από 70mg/dl και υπεργλυκαιμικές αυτές που είναι υψηλότερες από 180mg/dl. Επειδή το πλήθος των παραγόμενων χρονοσειρών γλυκόζης ανά ασθενή είναι περιορισμένο, επιλέγουμε να αναπτύξουμε ένα *global*, γενικευμένο μοντέλο, χρησιμοποιώντας δεδομένα από όλους τους ασθενείς. Με βάση το αν κάθε φορά γίνεται πρόβλεψη υπογλυκαιμικών ή υπεργλυκαιμικών συμβάντων, αλλά και τον εφαρμοζόμενο ορίζοντα πρόβλεψης, προκύπτουν τέσσερα διαφορετικά datasets. Η κατανομή των δύο κλάσεων για κάθε σύνολο δεδομένων φαίνεται στον πίνακα που ακολουθεί:

Πίνακας 6.1: Η κατανομή των κλάσεων για κάθε σύνολο δεδομένων

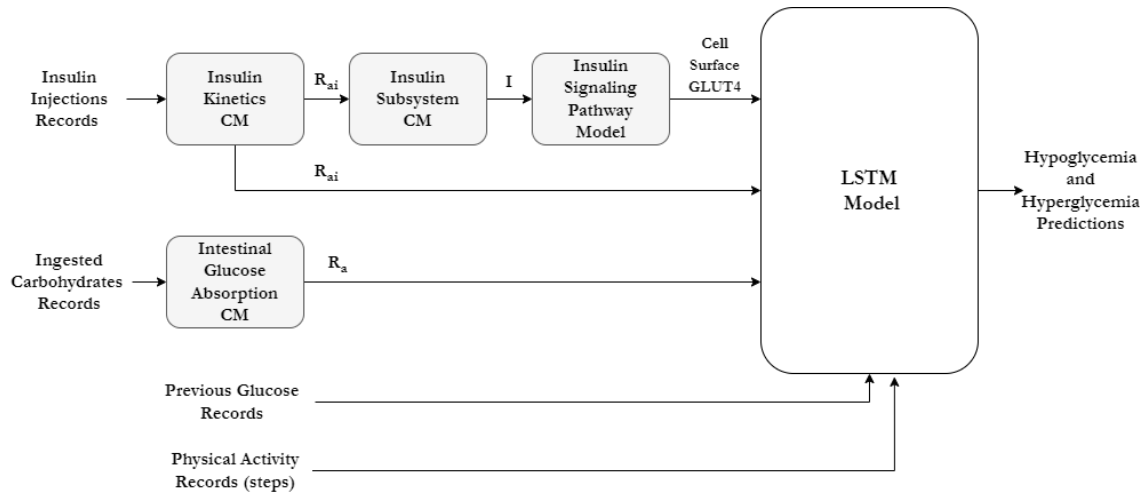
Σύνολο δεδομένων	Κατανομή κλάσεων
Υπογλυκαιμία, PH 0-30 min	3%
Υπογλυκαιμία, PH 30-60 min	4%
Υπεργλυκαιμία, PH 0-30 min	35%
Υπεργλυκαιμία, PH 30-60 min	34%

Παρατηρούμε ότι και τα τέσσερα διαφορετικά προβλήματα ταξινόμησης χαρακτηρίζονται από ανισορροπία κλάσεων, καθώς τα επίπεδα γλυκόζης που αντιστοιχούν σε υπογλυκαιμικά και υπεργλυκαιμικά συμβάντα εμφανίζονται λιγότερο συχνά από φυσιολογικές τιμές. Η πιο ακραία εκδήλωση ανισορροπίας κλάσεων αφορά την περίπτωση της υπογλυκαιμίας, λόγω της μεγαλύτερης σπανιότητας των υπογλυκαιμικών επεισοδίων, σε σχέση με τα υπεργλυκαιμικά.

## 6.2 Δομή υβριδικού μοντέλου

Για την εκτίμηση του κινδύνου εμφάνισης υπεργλυκαιμικών και υπογλυκαιμικών επεισοδίων χρησιμοποιήθηκε μια υβριδική προσέγγιση που συνδυάζει τη χρήση μοντέλων προσομοίωσης φυσιολογικών συστημάτων και μεθόδων μηχανικής μάθησης. Αναλυτικότερα, με στόχο την ακριβέστερη αποτύπωση των μηχανισμών απορροφησης της ινσουλίνης από τον υποδόριο χώρο στην αιματική κυκλοφορία, της ενεργοποίησης των ενδοκυτταρικών μονοπατιών μεταδίβασης σήματος ινσουλίνης και της απορρόφησης της γλυκόζης γεύματος από το έντερο στην αιματική κυκλοφορία, αξιοποιήθηκαν κατάλληλα διαμερισματικά μοντέλα ΔΜ (compartmental models - CMs) που συνδυάστηκαν με ένα αναδρομικό νευρωνικό δίκτυο μακράς βραχείας μνήμης LSTM.

Σημειώνουμε ότι τα διαμερισματικά μοντέλα εφαρμόζονται με κυρίαρχο στόχο την δυναμική μελέτη της κινητικής ουσιών σε φυσιολογικά συστήματα. Οι ουσίες αυτές μπορεί να είναι εξωγενείς (π.χ. φάρμακα) ή και ενδογενείς (ορμόνες κ.λπ.). Τα εν λόγω συστήματα περιγράφουν διαδικασίες παραγωγής, κατανομής, μεταφοράς, χρησιμοποίησης των ουσιών αυτών, καθώς και μηχανισμούς αλληλεπίδρασης με, και ελέγχου από άλλες ουσίες. Ως διαμέρισμα ορίζεται η ποσότητα μιας ουσίας, η οποία συμπεριφέρεται σαν να είναι καλά αναμειγμένη και κινητικά ομογενής. Ένα διαμερισματικό μοντέλο αποτελείται από έναν πεπερασμένο αριθμό διαμερισμάτων που συνδέονται με καθορισμένο τρόπο μεταξύ τους, όπου η κάθε σύνδεση αντιπροσωπεύει τη ροή ουσίας από ένα διαμέρισμα σε άλλο. [102]



Σχήμα 6.1: Διαγραμματική απεικόνιση του αναπτυσσόμενου (υβριδικού) μοντέλου πρόβλεψης υπογλυκαιμικών και υπεργλυκαιμικών επεισοδίων

### 6.2.1 Μοντέλο απορρόφησης γλυκόζης γεύματος από το έντερο στο αίμα

Για την αξιοποίηση των διαθέσιμων πληροφοριών υδατανθράκων με πιο ακριβή τρόπο, προχωρούμε στην χρήση του διαμερισματικού μοντέλου Dalla Man [103], [74]. Το μοντέλο περιγράφει τη ροή της γλυκόζης εντός του στομαχίου και του εντέρου και αποτελείται από τρία διαμερίσματα, δύο για την αναπαράσταση του στομαχίου και ένα για την αναπαράσταση του εντέρου. Το αντίστοιχο σύστημα διαφορικών εξισώσεων δίνεται παρακάτω:

$$\begin{aligned}
 Q_{sto}(t) &= Q_{sto1}(t) + Q_{sto2}(t) \\
 Q_{sto}(0) &= 0 \\
 \dot{Q}_{sto1}(t) &= -k_{gri} \cdot Q_{sto1}(t) + D \cdot \delta(t) \\
 Q_{sto1}(0) &= 0 \\
 \dot{Q}_{sto2}(t) &= -k_{empt} \cdot Q_{sto2}(t) + k_{gri} \cdot Q_{sto1}(t) \\
 Q_{sto2}(0) &= 0 \\
 \dot{Q}_{gut}(t) &= -k_{abs} \cdot Q_{gut}(t) + k_{empt} \cdot Q_{sto2}(t) \\
 Q_{gut}(0) &= 0
 \end{aligned} \tag{6.1}$$

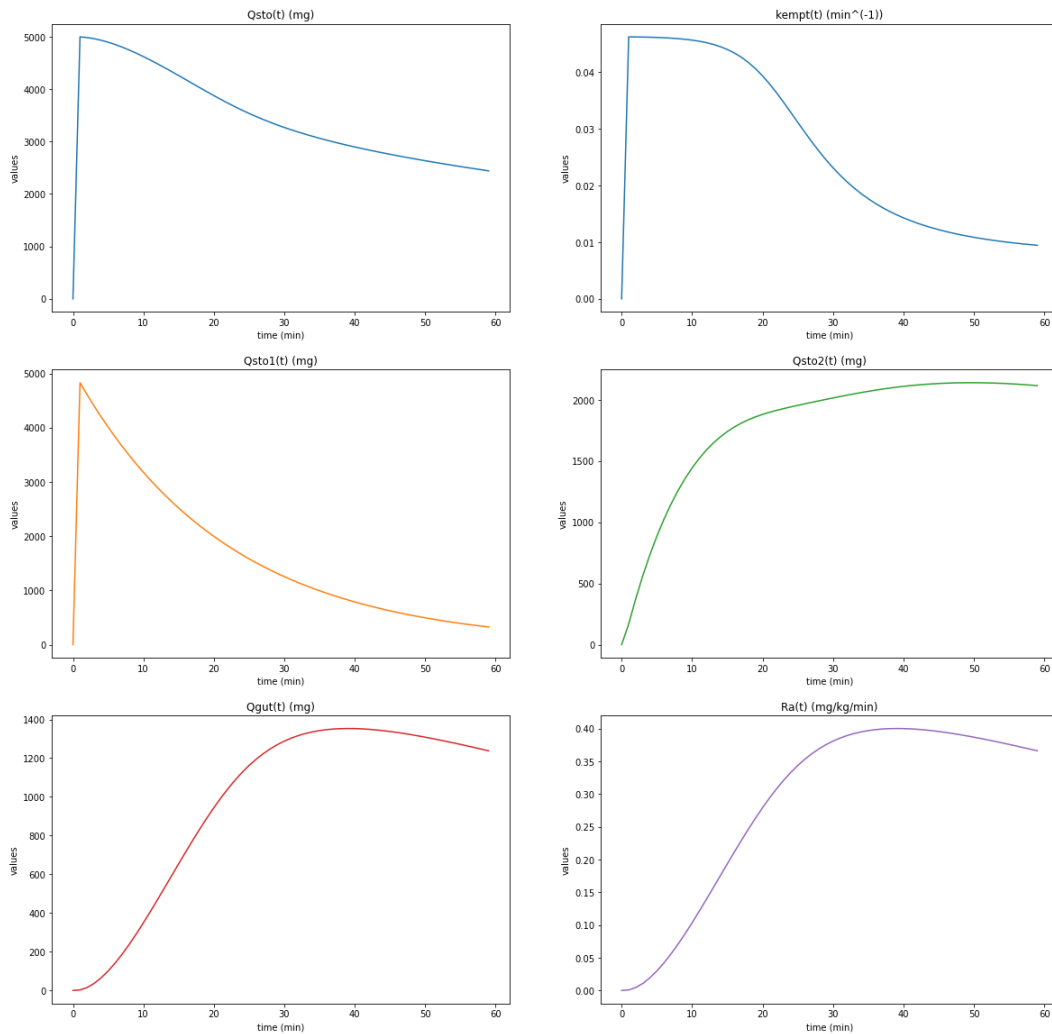
$$R_a(t) = \frac{f \cdot k_{abs} \cdot Q_{gut}(t)}{BW}$$

$$R_a(0) = 0$$

$$k_{empt}(Q_{sto}) = k_{min} + \frac{k_{max} - k_{min}}{2} \cdot (\tanh(a(Q_{sto} - b \cdot D)) + 1)$$

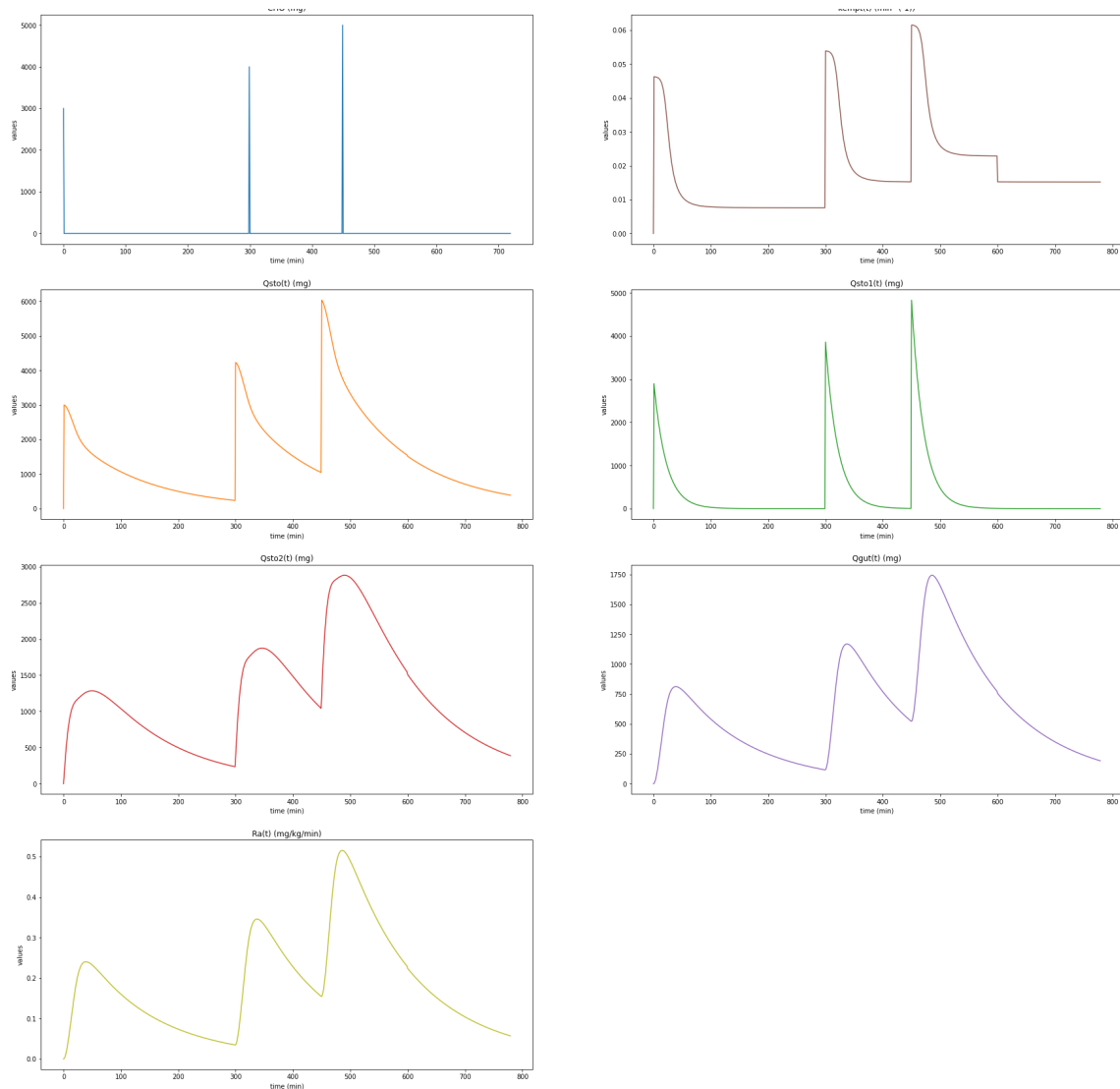
$$a = \frac{5}{2 \cdot D \cdot (1 - b)}$$

όπου  $Q_{sto}(mg)$  η συνολική ποσότητα γλυκόζης στο στομάχι,  $Q_{sto1}(mg)$  η ποσότητα γλυκόζης στο στομάχι σε στερεή μορφή,  $Q_{sto2}(mg)$  η ποσότητα γλυκόζης στο στομάχι σε υγρή μορφή,  $Q_{gut}(mg)$  η ποσότητα γλυκόζης στο έντερο,  $R_a(mg/kg/min)$  ο ρυθμός εμφάνισης γλυκόζης στο αίμα,  $k_{empt}(min^{-1})$  ο ρυθμός γαστρικής εκκένωσης,  $k_{gri}(min^{-1})$  ο ρυθμός άλεσης (grinding) της γλυκόζης,  $k_{abs}(min^{-1})$  ο ρυθμός απορρόφησης γλυκόζης από το έντερο,  $D(mg)$  η προσλαμβανόμενη ποσότητα υδατανθράκων και  $BW(kg)$  το βάρος του ατόμου που εξετάζεται. Τα  $k_{gri}$ ,  $k_{abs}$ ,  $f$ ,  $b$ ,  $k_{min}$ ,  $k_{max}$  αποτελούν σταθερές των οποίων οι τιμές ορίζονται στη βιβλιογραφία.



Σχήμα 6.2: Υπολογισμός των μεγεθών  $k_{empt}$ ,  $Q_{sto}$ ,  $Q_{sto1}$ ,  $Q_{sto2}$ ,  $Q_{gut}$ ,  $R_a$  για διάστημα 1 ώρας, πρόσληψη υδατανθράκων 5g και σωματικό βάρος 70kg, με χρήση του Intestinal Glucose Absorption CM

Επομένως, δεδομένης μιας ποσότητας υδατανθράκων  $D$ , είναι δυνατή η εκτίμηση του ρυθμού εμφάνισης γλυκόζης στο αίμα, μιας πληροφορίας που φέρει μεγαλύτερη κλινική σημασία και θα μπορούσε να συμβάλει στην πραγματοποίηση ακριβέστερων προβλέψεων.



Σχήμα 6.3: Υπολογισμός των μεγεθών  $k_{empt}$ ,  $Q_{sto}$ ,  $Q_{sto1}$ ,  $Q_{sto2}$ ,  $Q_{gut}$ ,  $R_a$  για τρεις διαφορετικές προσλήψεις υδατανθράκων 3g, 4g και 5g εντός ενός διαστήματος 12 ωρών. Ο υπολογισμός συντελέστηκε χρησιμοποιώντας το Intestinal Glucose Absorption CM και πραγματοποιώντας υπέρθεση των αντίστοιχων καμπυλών εξόδου

Παρατηρούμε ότι η καμπύλη του ρυθμού εμφάνισης γλυκόζης στο αίμα  $R_a(t)$  διαμορφώνεται από το σύνολο των γευμάτων που έχουν προηγηθεί στο πρόσφατο παρελθόν, ανάλογα με την ποσότητα υδατανθράκων που προσλαμβάνεται και την χρονική στιγμή κατά την οποία σημειώνεται το κάθε γεύμα. Συμπεραίνουμε επομένως ότι το μέγεθος  $R_a(t)$  αναπαριστά με πιο ουσιαστικό και ακριβή τρόπο τον τρόπο με τον οποίο συντελείται συνολικά η απορρόφηση γλυκόζης, και άρα μπορεί να χρησιμοποιηθεί ως χαρακτηριστικό εισόδου για την πραγματοποίηση βελτιωμένων προβλέψεων.

Στην περίπτωση μας, με δεδομένη μια χρονοσειρά γλυκόζης που περιγράφει το γλυκαιμικό προφίλ ενός ασθενή εντός ενός διαστήματος  $[t_{start}, t_{end}]$  λαμβάνονται υπόψιν

οι καταγεγραμμένες προσλήψεις υδατανθράκων που αυτός έχει πραγματοποιήσει εντός του διαστήματος  $[t_{end} - 12h, t_{end}]$ . Με αυτήν την είσοδο υπολογίζουμε το  $R_a(t)$  για  $t \in [t_{end}, t_{end} + 60min]$ , με χρήση του διαμερισματικού μοντέλου απορρόφησης γλυκόζης κι εφαρμόζοντας υπέρθεση των καμπυλών που προκύπτουν από το μοντέλο για κάθε μεμονωμένη πρόσληψη υδατανθράκων. Έτσι, για τον ορίζοντα πρόβλεψης 0 έως 30 λεπτά, χρησιμοποιούμε ως χαρακτηριστικά εισόδου τις τιμές  $R_a(t)$  για  $t = t_{end}$ ,  $t = t_{end} + 15min$  και  $t = t_{end} + 30min$ . Αντίστοιχα, για τον ορίζοντα πρόβλεψης 30 έως 60 λεπτά, εκτός από τις προαναφερθείσες, χρησιμοποιούμε ακόμα ως τιμές εισόδου, τις τιμές  $R_a(t = t_{end} + 45min)$  και  $R_a(t = t_{end} + 60min)$ .

## 6.2.2 Μοντέλο απορρόφησης ινσουλίνης από τον υποδόριο χώρο στο αίμα

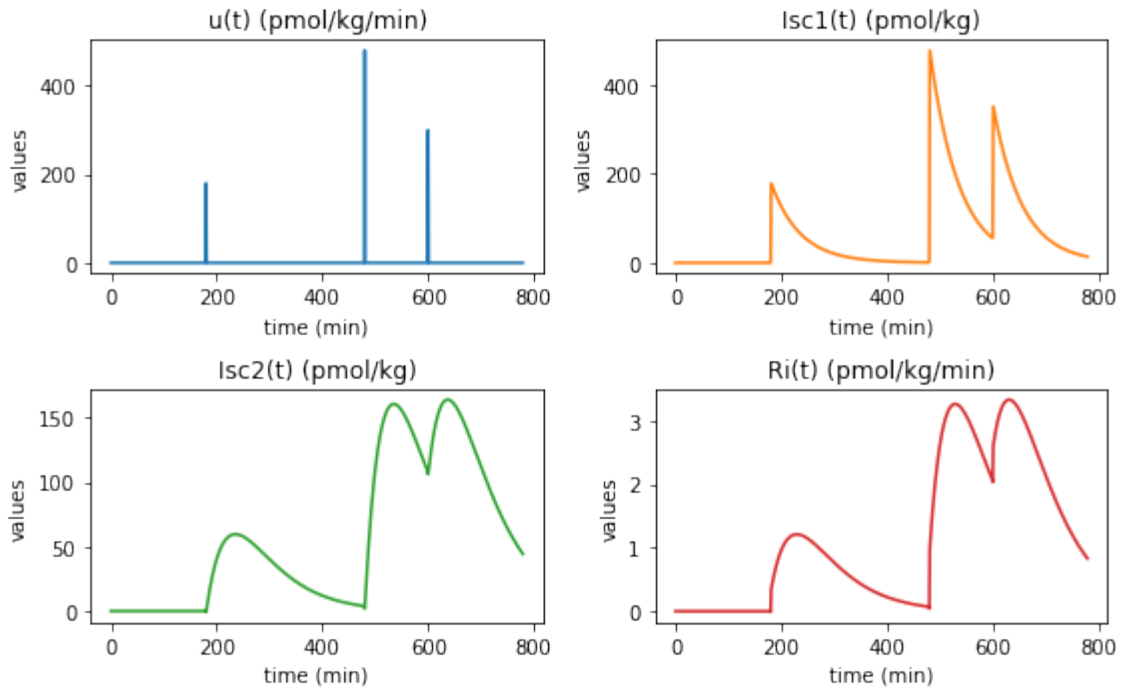
Για την αποτύπωση των περίπλοκων διαδικασιών που σχετίζονται με την απορρόφηση της προσλαμβανόμενης ινσουλίνης από τον υποδόριο χώρο στην αιματική κυκλοφορία, εφαρμόζεται το παρακάτω διαμερισματικό μοντέλο προσομοίωσης της κινητικής της υποδόριας ινσουλίνης: [104]

$$\begin{aligned}
 \dot{I}_{sc1}(t) &= -(k_d + k_{a1}) \cdot I_{sc1}(t) + u(t) \\
 \dot{I}_{sc2}(t) &= k_d \cdot I_{sc1}(t) - k_{a2} \cdot I_{sc2}(t) \\
 R_{ai}(t) &= k_{a1} \cdot I_{sc1}(t) + k_{a2} \cdot I_{sc2}(t) \\
 I_{sc1}(0) &= I_{sc1ss} \\
 I_{sc2}(0) &= I_{sc2ss}
 \end{aligned} \tag{6.2}$$

όπου  $I_{sc1}(pmol/kg)$  η ποσότητα μη μονομερούς ινσουλίνης στον υποδόριο χώρο,  $I_{sc2}(pmol/kg)$  η ποσότητα μονομερούς ινσουλίνης στον υποδόριο χώρο,  $u(t)(pmol/kg/min)$  ο ρυθμός εξωγενούς έγχυσης ινσουλίνης,  $k_d(min^{-1})$  ο σταθερός ρυθμός διάσπασης ινσουλίνης και  $k_{a1}(min^{-1})$ ,  $k_{a2}(min^{-1})$  οι σταθερές απορρόφησης της μη μονομερούς και της μονομερούς ινσουλίνης, αντίστοιχα. Επίσης, το μέγεθος  $R_{ai}(pmol/kg/min)$  περιγράφει το ρυθμό εμφάνισης της ινσουλίνης στο πλάσμα του αίματος.

Συνολικά, οι παραπάνω σχέσεις περιγράφουν τη διαδικασία διάχυσης της εισαγόμενης ινσουλίνης από τον υποδόριο χώρο, στο πλάσμα του αίματος. Αξίζει να σημειωθεί ότι η ινσουλίνη εισάγεται συνήθως στους ασθενείς στην πιο σταθερή, πολυμερή της μορφή και διασπάται στην συνέχεια σε μονομερή ινσουλίνης, τα οποία και απορροφώνται ταχύτερα από τον οργανισμό.

Ομοίως με την περίπτωση των γευμάτων, δεδομένης μιας χρονοσειράς γλυκόζης που εκτείνεται εντός του διαστήματος  $[t_{start}, t_{end}]$ , λαμβάνονται υπόψιν οι χορηγήσεις ινσουλίνης που έχουν πραγματοποιηθεί εντός της περιόδου  $[t_{end} - 12h, t_{end}]$ , με στόχο τον υπολογισμό του ρυθμού εμφάνισης ινσουλίνης στο αίμα  $R_{ai}(t)$  για  $t \in [t_{end}, t_{end} + 60min]$ . Οι τιμές που χρησιμοποιούνται ως χαρακτηριστικά εισόδου είναι οι  $R_{ai}(t_{end})$ ,  $R_{ai}(t_{end} + 15min)$ ,  $R_{ai}(t_{end} + 30min)$  για ορίζοντα πρόβλεψης 0-30 λεπτά, ενώ για τον ορίζοντα πρόβλεψης 30 έως 60 λεπτά, χρησιμοποιούνται επιπλέον οι  $R_{ai}(t_{end} + 45min)$  και  $R_{ai}(t_{end} + 60min)$  τιμές.



Σχήμα 6.4: Υπολογισμός των μεγεθών  $I_{sc1}$ ,  $I_{sc2}$ ,  $R_i$  για τρεις διαφορετικές χορηγήσεις ινσουλίνης 1.5, 4 και 2.5 μονάδων εντός ενός διαστήματος 12 ωρών και σωματικό βάρος 50kg, με χρήση του Insulin Kinetics CM.

### 6.2.3 Μοντέλο ενεργοποίησης του ενδοκυτταρικού μονοπατιού μεταβίβασης σήματος ινσουλίνης

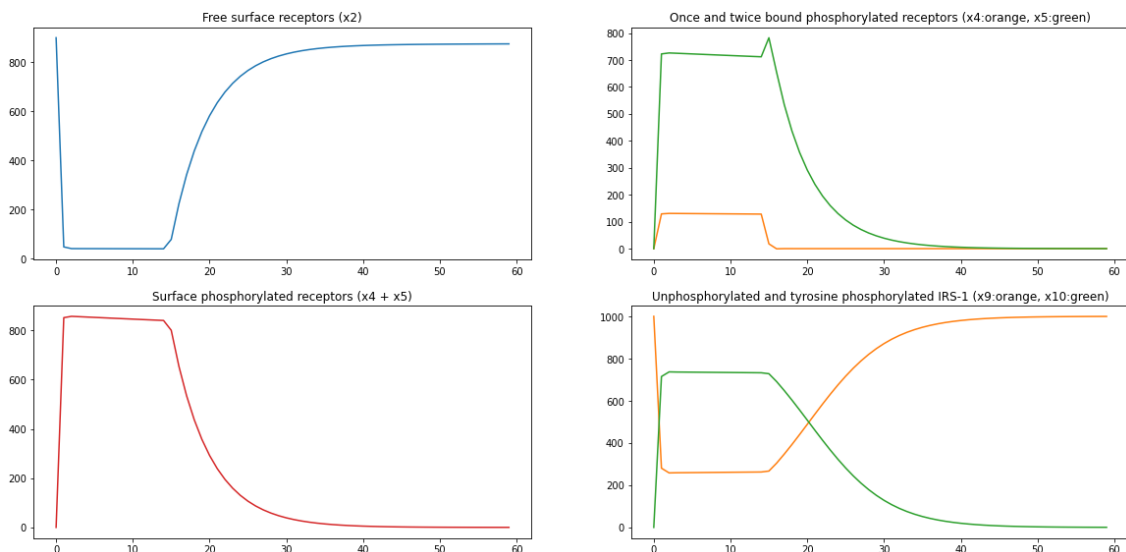
Με στόχο την εξέταση της διαδικασίας απορρόφησης της ινσουλίνης και σε κυτταρικό επίπεδο, γίνεται χρήση δύο ακόμα μοντέλων προσομοίωσης. Το πρώτο εξ αυτών (*Insulin Subsystem CM*) εφαρμόζει διαμερισματική μοντελοποίηση, αποτελείται από δύο διαμερίσματα (ήπαρ και πλάσμα αίματος) κι έχει ως στόχο τον υπολογισμό της συγκέντρωσης της ινσουλίνης στο πλάσμα  $I$ , δεδομένου του αντίστοιχου ρυθμού εμφάνισης  $R_{ai}$ : [74], [105]

$$\begin{aligned}
 \dot{I}_p(t) &= -(m_2 + m_4) \cdot I_p(t) + m_1 \cdot I_l(t) + R_{ai}(t) \\
 \dot{I}_l(t) &= -(m_1 + m_3) \cdot I_l(t) + m_2 \cdot I_p(t) \\
 I(t) &= I_p(t)/V_I \\
 I_l(0) &= I_{lb} \\
 I_p(0) &= I_{pb}
 \end{aligned} \tag{6.3}$$

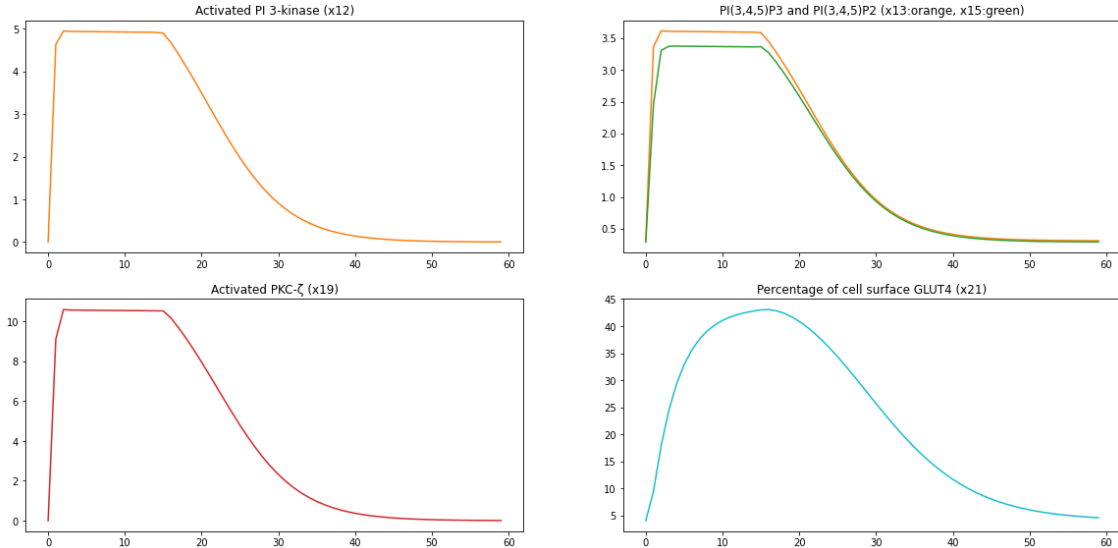
όπου  $I_p(\text{pmol/kg})$ ,  $I_l(\text{pmol/kg})$  οι ποσότητες ινσουλίνης στο πλάσμα και στο ήπαρ αντίστοιχα,  $I(\text{pmol/l})$  η συγκέντρωση ινσουλίνης στο πλάσμα,  $R_{ai}(\text{pmol/kg/min})$  ο ρυθμός εμφάνισης ινσουλίνης στο πλάσμα,  $V_I(\text{l/kg})$  ο όγκος κατανομής ινσουλίνης και  $m_1(\text{min}^{-1})$ ,  $m_2(\text{min}^{-1})$ ,  $m_3(\text{min}^{-1})$ ,  $m_4(\text{min}^{-1})$  σταθερές που αποτυπώνουν τις διαδικασίες απομάκρυνσης ινσουλίνης από το ήπαρ και τους περιφερειακούς ιστούς. Ακόμα, σημειώνουμε ότι τα  $I_{lb}$ ,  $I_{pb}$  αποτελούν τις τιμές των μεγεθών  $I_l$ ,  $I_p$  σε κατάσταση νηστείας (basal state). Στην περίπτωση μας, ο ρυθμός εμφάνισης ινσουλίνης  $R_{ai}$  προκύπτει από το CM κινητικής της ινσουλίνης, με βάση τον ρυθμό εξωγενούς έγχυσης της ουσίας.

Τέλος, το μαθηματικό μοντέλο που χρησιμοποιήθηκε για την προσομοίωση της ενεργοποίησης του ενδοκυτταρικού μονοπατιού μεταβίβασης του σήματος ινσουλίνης είναι αυτό του Sedaghat κ.α. [78] του 2002. Το συγκεκριμένο μοντέλο επιχειρεί να προσομοιώσει βασικά στοιχεία του PI3K/AKT μονοπατιού σήμανσης της ινσουλίνης. Το PI3k/AKT μονοπάτι ενεργοποιείται σε ένα κύτταρο με την πρόσδεση ινσουλίνης στους αντίστοιχους διαμεμβρανικούς υποδοχείς, έπειτα από την οποία πυροδοτείται αλυσιδωτή φωσφορυλίωση μιας σειράς πρωτεϊνών. Οι ενδοκυττάρια αυτές αλληλεπιδράσεις των πρωτεϊνών επιτρέπουν την μετάδοση του σήματος από τον υποδοχέα, προς διάφορες κατευθύνσεις, και έχει ως τελικό αποτέλεσμα την μετανάστευση των κυστιδίων που περιέχουν τους διαμεταφορείς γλυκόζης GLUT4, από την ενδοκυττάρια δεξαμενή, στην πλασματική μεμβράνη του κυττάρου. Οποιαδήποτε διατάραξη του παραπάνω μονοπατιού σήμανσης μπορεί να οδηγήσει στην λεγόμενη *αντίσταση* στην ινσουλίνη και σε μειωμένη εισχώρηση γλυκόζης εντός των κυττάρων.

Το μοντέλο Sedaghat που χρησιμοποιήθηκε για την προσομοίωση αυτής της διαδικασίας κάνει χρήση 21 μεταβλητών ( $x_1, x_2, \dots, x_{21}$ ) που υπολογίζονται από ένα σύστημα 20 διαφορικών εξισώσεων. Η είσοδος  $x_1$  αποτυπώνει τη συγκέντρωση της ινσουλίνης στο διάμεσο υγρό, ενώ η μεταβλητή  $x_{21}$  (την οποία και θα χρησιμοποιήσουμε) περιγράφει το ποσοστό GLUT4 που καταλήγει στο κυτταρικό τοίχωμα. Για τον υπολογισμό της συγκέντρωσης ινσουλίνης στο διάμεσο υγρό χρησιμοποιούμε την συγκέντρωση ινσουλίνης πλάσματος που προκύπτει με χρήση του Insulin Subsystem CM και παράλληλα κάνουμε την παραδοχή ότι η συγκέντρωση ινσουλίνης στο διάμεσο υγρό είναι κατά προσέγγιση 40% χαμηλότερη από ότι είναι στο πλάσμα, όπως αναφέρεται και σε σχετικές δημοσιεύσεις. [106]



Σχήμα 6.5: Υπολογισμός των αναγραφόμενων μεγεθών με χρήση του μαθηματικού μοντέλου προσομοίωσης Sedaghat, έπειτα από την εφαρμογή ινσουλίνης σταθερής συγκέντρωσης  $10^{-5}M$ , από  $t = 0 \text{ min}$  έως  $t = 15 \text{ min}$



Σχήμα 6.6: Υπολογισμός των αναγραφόμενων μεγεθών με χρήση του μαθηματικού μοντέλου προσομοίωσης Sedaghat, έπειτα από την εφαρμογή ινσουλίνης σταθερής συγκέντρωσης  $10^{-5}M$ , από  $t = 0 \text{ min}$  έως  $t = 15 \text{ min}$

Συνοψίζοντας, για τη συμπερίληψη πληροφοριών σχετιζόμενων με το ποσοστό GLUT4 στο κυτταρικό τοίχωμα ( $x_{21}$ ) έπειτα από χορηγήσεις ινσουλίνης, γίνεται χρήση του διαμερισματικού μοντέλου κινητικής ινσουλίνης, του ΔM υπολογισμού της συγκέντρωσης ινσουλίνης στο πλάσμα (Insulin Subsystem CM) και του Sedaghat μοντέλου. Αναλυτικότερα, δεδομένης μιας χρονοσειράς γλυκόζης που περιγράφει ένα διάστημα  $[t_{start}, t_{end}]$ , λαμβάνονται υπόψιν οι χορηγήσεις ινσουλίνης που έχουν πραγματοποιηθεί από τον αντίστοιχο ασθενή εντός της περιόδου  $[t_{end} - 12h, t_{end}]$ . Εφαρμόζοντας το Insulin Kinetics CM υπολογίζεται ο ρυθμός  $R_{ai}$  εντός του αντίστοιχου διαστήματος  $[t_{end} - 12h, t_{end}]$  και στη συνέχεια, με είσοδο τον ρυθμό αυτόν, εκτιμάται η συγκέντρωση ινσουλίνης  $I$  για το ίδιο διάστημα, εφαρμόζοντας το Insulin Subsystem CM. Τελικά, από την συγκέντρωση ινσουλίνης, και με χρήση του Sedaghat μοντέλου, υπολογίζεται το μέγεθος  $x_{21}$  για το διάστημα ενδιαφέροντος  $[t_{end}, t_{end} + 60min]$ , όπως αυτό έχει διαμορφωθεί από τις προσλήψεις ινσουλίνης εντός του παρελθοντικού  $[t_{end} - 12h, t_{end}]$  διαστήματος. Όμοια με προηγουμένως, για τον ορίζοντα πρόβλεψης 0-30 λεπτών επιλέγουμε να χρησιμοποιήσουμε ως χαρακτηριστικά εισόδου τις τιμές  $x_{21}(t_{end})$ ,  $x_{21}(t_{end} + 15min)$  και  $x_{21}(t_{end} + 30min)$ . Αντίστοιχα, για τον ορίζοντα πρόβλεψης από 30 έως 60 λεπτά, μαζί με τις προαναφερθείσες, χρησιμοποιούνται ακόμα ως χαρακτηριστικά εισόδου οι  $x_{21}(t_{end} + 45min)$  και  $x_{21}(t_{end} + 60min)$  τιμές.

#### 6.2.4 Χρήση δεδομένων σωματικής δραστηριότητας

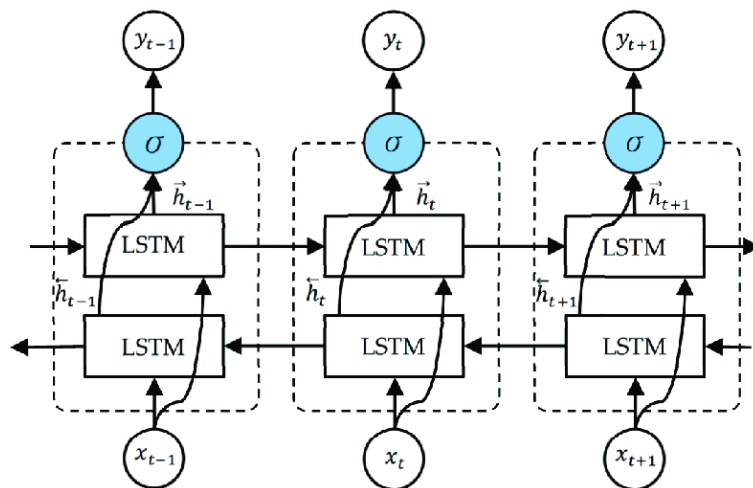
Οι διαδικασίες που περιγράφηκαν παραπάνω αφορούσαν τα δεδομένα γλυκόζης, ινσουλίνης και υδατανθράκων. Στην περίπτωση των διαθέσιμων πληροφοριών βημάτων δεν εφαρμόστηκε φυσιολογική μοντελοποίηση, αλλά έγινε άμεση χρήση τους. Συγκεκριμένα, με δεδομένη μια χρονοσειρά γλυκόζης που ανήκει στο διάστημα  $[t_{start}, t_{end}]$ , χρησιμοποιούμε ως χαρακτηριστικό εισόδου το συνολικό άθροισμα των βημάτων που σημειώθηκαν από τον ασθενή εντός της ίδιας αυτής περιόδου. Τελικά, για την περίπτωση πρόβλεψης με prediction horizon 0-30 λεπτών κάθε δείγμα αποτελείται κάθε φορά από συνολικά 14 τιμές: 4 τιμές γλυκόζης, 3 τιμές του μεγέθους  $R_a$ , 3 σιγμιότυπα



από το μέγεθος  $R_{ai}$  και 3 από το  $x_{21}$  και μία τιμή για τη συμπερίληψη δεδομένων σωματικής δραστηριότητας. Αντίστοιχα, για τον ορίζοντα πρόβλεψης 30 έως 60 λεπτών εφαρμόζονται κάθε φορά συνολικά 20 χαρακτηριστικά, λόγω της συμπερίληψης επιπλέον πληροφοριών από τα  $R_a$ ,  $R_{ai}$  και  $x_{21}$  μεγέθη.

### 6.2.5 Μοντέλο μηχανικής μάθησης

Για τα δυαδικά προβλήματα ταξινόμησης που αντιμετωπίζεται επιλέξαμε να χρησιμοποιήσουμε ένα αναδρομικό νευρωνικό δίκτυο, αποτελούμενο από ένα αμφίδρομο (Bidirectional) LSTM, ένα απλό LSTM και δύο πλήρως συνδεδεμένα επίπεδα. Ο συνδυασμός στοιβαγμένων αναδρομικών επιπέδων, ακολουθούμενων από πυκνά επίπεδα συναντάται συχνά σε εφαρμογές πρόβλεψης γλυκόζης, όπως φαίνεται και από τη βιβλιογραφική επισκόπηση του κεφαλαίου 5. Οι βασικές αρχές λειτουργίας των πυκνών και των απλών LSTMs περιγράφονται στο κεφάλαιο 3. Όσον αφορά το αμφίδρομο LSTM, αυτό αποτελείται από δύο LSTMs: το ένα λαμβάνει και επεξεργάζεται την είσοδο σε μια κατεύθυνση προς τα μπρος και το άλλο προς την αντίθετη κατεύθυνση. Στόχος είναι η αύξηση του πλήθους των πληροφοριών που είναι διαθέσιμες στο δίκτυο. Η έξοδος των δύο αυτών δικτύων τελικά συγχωνεύονται, συνήθως μέσω της concatenation μεθόδου.



Σχήμα 6.7: Η ανεπτυγμένη δομή ενός Bidirectional LSTM (BiLSTM) για χρονοσειρά των βημάτων [107]

Τα δεδομένα δίνονται στο BiLSTM επίπεδο με την τρισδιάστατη μορφή (samples, timesteps, features), όπου samples το συνολικό πλήθος δειγμάτων-χρονοσειρών, timesteps το πλήθος χρονικών βημάτων κάθε χρονοσειράς και features ο αριθμός χαρακτηριστικών από την οποία αποτελείται το κάθε βήμα. Στην περίπτωση μας είναι timesteps = 1 και features = 14 ή 20, ανάλογα με το αν εξετάζεται ο ορίζοντας πρόβλεψης 0-30 ή 30-60 min, αντίστοιχα. Η έξοδος του BiLSTM επιπέδου έχει διάσταση (samples, timesteps, units<sub>1</sub>), όπου units<sub>1</sub> το πλήθος μονάδων που έχει οριστεί για το συγκεκριμένο επίπεδο, δηλαδή το πλήθος μονάδων καθενός από τα δύο LSTMs που συναποτελούν το BiLSTM. Η έξοδος αυτή δίνεται ως είσοδος στο απλό LSTM επίπεδο, το οποίο διαθέτει έξοδο (samples, units<sub>2</sub>), όπου units<sub>2</sub> ο αριθμός μονάδων που έχει επιλεγεί για το απλό LSTM layer. Η τρισδιάστατη αυτή έξοδος εισέρχεται στο πρώτο πυκνό επίπεδο, το οποίο δίνει έξοδο της μορφής (samples, units<sub>3</sub>), όπου units<sub>3</sub> ο καθορισμένος αριθμός μονάδων του επιπέδου αυτού αντίστοιχα. Τέλος, το τελικό πυκνό επίπεδο πραγματοποιεί

την τελική ταξινόμηση. Σημειώνουμε ότι ως συνάρτηση ενεργοποίησης των δύο πρώτων επιπέδων (BiLSTM, LSTM) χρησιμοποιείται η προεπιλεγμένη από το Keras *tanh*, ως συνάρτηση ενεργοποίησης του πρώτου πυκνού επιπέδου εφαρμόζεται η *ReLU*, ενώ για το τελευταίο πλήρως συνδεδεμένο επίπεδο γίνεται χρήση της σιγμοειδούς συνάρτησης, με στόχο την τελική κατηγοριοποίηση των δειγμάτων στις δύο κλάσεις. Για την εκπαίδευση του δικτύου χρησιμοποιήθηκαν η *binary cross-entropy* συνάρτηση κόστους και ο Adam βελτιστοποιητής, με ρυθμό εκπαίδευσης *lr* ίσο με το προκαθορισμένο από το Keras, 0.001. Τέλος, με στόχο την βελτιστοποίηση του δικτύου έγιναν δοκιμές με τις εξής τιμές υπερπαραμέτρων:

Πίνακας 6.2: Τιμές που ελέγχθηκαν για την βελτιστοποίηση των υπερπαραμέτρων

Υπερπαραμέτρος	Τιμές που ελέγχονται
Πλήθος μονάδων bi-directional LSTM επιπέδου ( $units_1$ )	32, 64, 128, 256
Πλήθος μονάδων απλού LSTM επιπέδου ( $units_2$ )	32, 64, 128, 256
Πλήθος μονάδων πρώτου πυκνού επιπέδου ( $units_3$ )	50, 100, 150

Στη συνέχεια αναλύονται εκτενέστερα οι μέθοδοι που εφαρμόστηκαν για την εκπαίδευση, την αξιολόγηση και την ερμηνεία του ML μοντέλου.

## 6.3 Εκπαίδευση, αξιολόγηση και ερμηνεία του μοντέλου μηχανικής μάθησης

### 6.3.1 Εκπαίδευση μοντέλου μηχανικής μάθησης

#### Κανονικοποίηση και εξισορρόπηση δεδομένων

Καθώς τα χαρακτηριστικά εισόδου βρίσκονται σε διαφορετική κλίμακα, απαιτείται κανονικοποίησή τους, με στόχο οι υψηλότερες αριθμητικές τιμές ορισμένων χαρακτηριστικών να πάψουν να επικρατούν έναντι άλλων, χαμηλότερων τιμών και όλα τα χαρακτηριστικά να λαμβάνονται εξίσου υπόψιν κατά την προσαρμογή των βαρών από την αρχή της εκπαίδευσης. Η μέθοδος κανονικοποίησης που επιλέχθηκε ήταν η Min-Max τεχνική κανονικοποίησης και ως διάστημα κανονικοποίησης το  $[0, +1]$ . Έτσι, με βάση και τη σχέση που παρατίθεται στην υποενότητα 3.5.3, οι τιμές κάθε χαρακτηριστικού μετασχηματίζονται ως εξής:

$$x'_{i,n} = \frac{x_{i,n} - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (6.4)$$

όπου με  $\min(x_i)$  και  $\max(x_i)$  συμβολίζονται η ελάχιστη και η μέγιστη τιμή του *i*-οστού χαρακτηριστικού.

Ακόμα, για την αντιμετώπιση της υψηλής ανισορροπίας των κλάσεων που αποτυπώνεται και στον πίνακα 6.1, διερευνήθηκαν οι εξής τέσσερις τεχνικές εξισορρόπησης των δεδομένων βασισμένες στην υπερδειγματοληψία:

1. Τυχαία υπερδειγματοληψία (Random Oversampling): η τεχνική αυτή αποτελεί την πιο απλή προσέγγιση, κατά την οποία τυχαία επιλεγμένα δείγματα της κλάσης

μειοψηφίας αναπαράγονται και στην συνέχεια προστίθενται στο αρχικό σύνολο αυτούσια.

2. Synthetic Minority Oversampling Technique (SMOTE) [45]: με τον αλγόριθμο αυτόν συντίθενται νέα στιγμιότυπα της κατηγορίας μειοψηφίας από τα υπάρχοντα δεδομένα. Ο αλγόριθμος επιλέγει τυχαία ένα δείγμα, έστω διάνυσμα  $X$ , της κλάσης μειοψηφίας. Για αυτό το στιγμιότυπο βρίσκει τους  $k$  πλησιέστερους γείτονές του ( $k$  nearest neighbors), επιλέγει έναν από αυτούς, έστω  $Y$ , και υπολογίζει τη διαφορά των δύο διανυσμάτων  $X, Y$ . Τέλος, πολλαπλασιάζει τη διαφορά με έναν τυχαίο αριθμό στο  $[0, 1]$ , οπότε προκύπτει το διάνυσμα  $Z$ . Το συνθετικό δείγμα προκύπτει από το άθροισμα των διανυσμάτων  $X$  και  $Z$ . Η παραπάνω διαδικασία επαναλαμβάνεται για  $N$  τυχαία δείγματα της μειονότητας, όπου το  $N$  προκύπτει ανάλογα με την τελική αναλογία κλάσεων που επιδιώκεται να επιτευχθεί.
3. Borderline Synthetic Minority Oversampling Technique (BSMOTE) [46]: μια παραλλαγή της SMOTE τεχνικής αποτελεί η Borderline SMOTE. Με την μέθοδο αυτή πραγματοποιείται υπερδειγματοληψία μόνο των δειγμάτων μειοψηφίας που εντοπίζονται κοντά στη διαχωριστική γραμμή των κλάσεων, καθώς τα δείγματα αυτά επηρεάζουν περισσότερο τα τελικά αποτελέσματα ταξινόμησης, από ότι τα στιγμιότυπα που βρίσκονται μακριά από τη διαχωριστική γραμμή.
4. Adaptive Synthetic (ADASYN) υπερδειγματοληψία [47]: ο ADASYN αλγόριθμος παράγει συνθετικά δεδομένα της κλάσης μειονότητας, ανάλογα με τη δυσκολία του κάθε στιγμιότυπου μειονότητας, όπου με τον όρο "δυσκολία" περιγράφεται η απαιτητικότητα ταξινόμησης του συγκεκριμένου δείγματος μειοψηφίας. Για να καθορισθεί ο αριθμός των συνθετικών δειγμάτων που θα δημιουργηθούν για κάθε δείγμα της κλάσης μειοψηφίας χρησιμοποιείται μια κατανομή πυκνότητας. Η κατανομή πυκνότητας  $r_i$  για κάθε δείγμα  $x_i$  της κλάσης μειοψηφίας υπολογίζεται ως  $r_i = \frac{\Delta_i}{K}$  όπου  $\Delta_i$  το πλήθος των  $K$  κοντινότερων γειτόνων του  $x_i$  που ανήκουν στην κλάση πλειοψηφίας. Ο αριθμός των συνθετικών δειγμάτων που δημιουργούνται τελικά για το δείγμα  $x_i$  είναι  $g_i = \hat{r}_i \cdot G$ , όπου  $\hat{r}_i$  η κανονικοποιημένη τιμή του  $r_i$  και  $G = (m_l - m_s) \cdot \beta$  όπου  $m_l, m_s$  το πλήθος των δειγμάτων των κλάσεων πλειοψηφίας και μειοψηφίας αντίστοιχα και  $\beta \in [0, 1]$  παράμετρος που προσδιορίζει το επιθυμητό μέγεθος της ισορροπίας του συνόλου των δεδομένων μετά την επαύξηση. Για  $\beta = 1$  έχουμε πλήρη εξισορρόπηση.

### **Βαθμονόμηση μοντέλου**

Η βαθμονόμηση (calibration) του δικτύου θεωρήθηκε αρκετά σημαντική, καθώς ήταν επιθυμητή η δυνατότητα εκτίμησης της εμφάνισης υπογλυκαιμικών και υπεργλυκαιμικών επεισοδίων και με τη μορφή πιθανότητας. Συνοπτικά, ένα μοντέλο μηχανικής μάθησης θεωρείται βαθμονομημένο όταν η πρόβλεψη του για μια κλάση, με τη μορφή πιθανότητας  $p$ , είναι ορθή σε ποσοστό  $(100 \cdot p)\%$  όλων των συνολικών προβλέψεων. Το χαρακτηριστικό αυτό έχει ιδιαίτερη αξία στις περιπτώσεις όπου δεν μας ενδιαφέρει μόνο το σε ποια κλάση ταξινομήθηκε τελικά ένα δείγμα, αλλά και η σιγουριά (confidence) με την οποία το εξεταζόμενο ML μοντέλο ταξινόμησε το συγκεκριμένο δείγμα στην κλάση αυτή. Στην περίπτωση μας, λόγω της φύσης του προβλήματος, είναι επιθυμητός ο υπολογισμός confidence scores για τη θετική κλάση, τα οποία αντανakλούν την πραγματική προβλεπτική ικανότητα του δικτύου. Αν και ορισμένα μοντέλα μηχανικής μάθησης (π.χ. μοντέλα λογιστικής παλινδρόμησης) είναι εκ φύσεως βαθμονομημένα,

άλλα (όπως οι Naive Bayes ταξινομητές) απαιτούν εξωτερική βαθμονόμηση. [108] Αντίστοιχα, η δυνατότητα ερμηνείας της εξόδου των τεχνητών νευρωνικών δικτύων ως πραγματικές πιθανότητες πρόβλεψης, εξαρτάται σε μεγάλο βαθμό και από τη δομή τους. Συνήθως, απλά νευρωνικά δίκτυα είναι βαθμονομημένα σε ικανοποιητικό βαθμό, με την βαθμονόμηση τους όμως να φθίνει, καθώς το πλήθος των επιπέδων τους και ο αριθμός των παραμέτρων τους αυξάνουν. [109] Στην περίπτωση μας, επιλέξαμε να εφαρμόσουμε επιπρόσθετη εξωτερική βαθμονόμηση με χρήση της μεθόδου Temperature Scaling, μια παραλλαγή της Platt Scaling μεθόδου. [109] Η Temperature Scaling τεχνική χρησιμοποιεί μια βαθμωτή παράμετρο  $T > 0$ , για να αλλάξει την κλίμακα της εξόδου του τελευταίου επιπέδου, πριν αυτή δοθεί ως είσοδος στη σιγμοειδή συνάρτηση. Η τιμή  $T$  προκύπτει έπειτα από την ελαχιστοποίηση της negative log likelihood μετρικής για ένα σύνολο επικύρωσης, δεδομένων των πιθανοτήτων πρόβλεψης του μη βαθμονομημένου μοντέλου και των αντίστοιχων επιθυμητών εξόδων, για το σύνολο αυτό. Για την αξιολόγηση του ποσοστού βαθμονόμησης του μοντέλου, κι επομένως την αξιολόγηση των παραγόμενων confidence scores χρησιμοποιούνται οι μετρικές Brier Score και Expected Calibration Error (ECE). [110] Σημειώνουμε ότι η Brier Score μετρική ορίζεται ως:

$$Brier\ Score = \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2 \quad (6.5)$$

όπου  $N$  το συνολικό πλήθος προβλέψεων και  $f_i, o_i$  η πιθανότητα πρόβλεψης του μοντέλου και η επιθυμητή έξοδος για το δείγμα  $i$  αντίστοιχα. Αν και η Brier μετρική μπορεί να δώσει κάποια ένδειξη για τη βαθμονόμηση ενός δικτύου καθώς κάνει χρήση πιθανοτήτων πρόβλεψης, μια πιο ενδεικτική μετρική αξιολόγησης για το συγκεκριμένο σκοπό, είναι το Expected Calibration Error (ECE):

$$ECE = \sum_{m=1}^M \frac{|B_m|}{n} \cdot |acc(B_m) - conf(B_m)| \quad (6.6)$$

Για τον υπολογισμό του ECE τα δείγματα κατηγοριοποιούνται σε bins, ανάλογα με το εύρος των πιθανοτήτων πρόβλεψής τους. Έτσι, το  $B_m$  αποτελεί το σετ των δειγμάτων των οποίων η πιθανότητα πρόβλεψης ανήκει στο διάστημα  $I_m = (\frac{m-1}{M}, \frac{m}{M}]$ , όπου  $M$  το πλήθος των bins. Οι ποσότητες  $acc(B_m)$  και  $conf(B_m)$  ορίζονται ως:

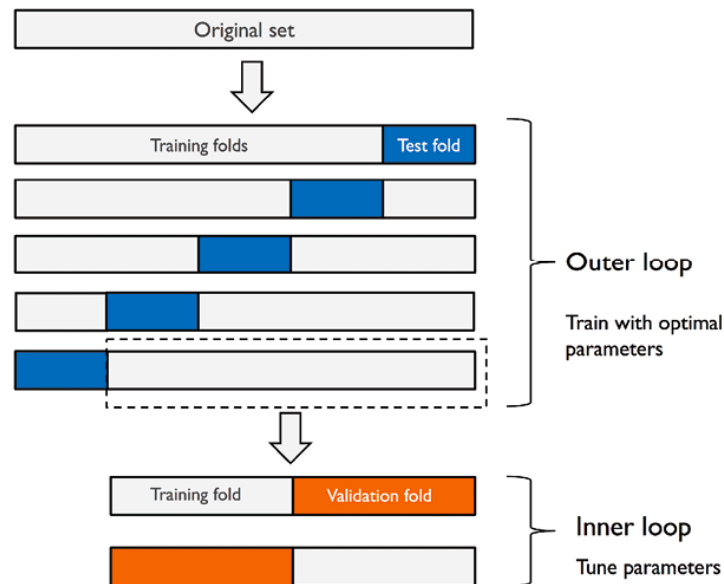
$$\begin{aligned} acc(B_m) &= \frac{1}{|B_m|} \sum_{i \in B_m} 1(y_i = o_i) \\ conf(B_m) &= \frac{1}{|B_m|} \sum_{i \in B_m} (p_i) \end{aligned} \quad (6.7)$$

όπου  $y_i, o_i, p_i$  η προβλεπόμενη κλάση, η πραγματική κλάση και η πιθανότητα πρόβλεψης του δείγματος  $i$ . Σημειώνουμε ότι στην περίπτωση μας, για τους παραπάνω υπολογισμούς χρησιμοποιήθηκαν 10 bins, δηλαδή  $M = 10$ .

### Εμφωλευμένη διασταυρούμενη επικύρωση

Για την εκπαίδευση του δικτύου για κάθε σύνολο δεδομένων, επιστρατεύτηκε η μέθοδος της *εμφωλευμένης διασταυρούμενης επικύρωσης* (nested cross-validation). Η τεχνική αυτή εφαρμόζει ένα εξωτερικό κι ένα εσωτερικό (εμφωλευμένο) k-fold cross-validation, όπου το εξωτερικό cross-validation χρησιμοποιείται για την αξιολόγηση του μοντέλου

και το εσωτερικό χρησιμοποιείται για τον προσδιορισμό των βέλτιστων υπερπαραμέτρων του. Στη μέθοδο nested cross-validation ενσωματώθηκαν ακόμα οι προαναφερθείσες τεχνικές εξισορρόπησης των δεδομένων και βαθμονόμησης, ενώ έγιναν ακόμα κάποιες δοκιμές με προσαρμοσμένο κατώφλι ταξινόμησης.



Σχήμα 6.8: Παράδειγμα εφαρμογής της nested cross-validation τεχνικής, όπου για το εξωτερικό cross-validation χρησιμοποιούνται 5 folds και για το εσωτερικό, 2 folds. [111]

Αναλυτικότερα, η μεθοδολογία που ακολουθείται για κάθε σύνολο δεδομένων είναι η εξής: αρχικά, το εξεταζόμενο σύνολο δεδομένων κανονικοποιείται στο διάστημα  $[0, 1]$  με τη χρήση Min-Max κανονικοποίησης. Στη συνέχεια, στο σύνολο δεδομένων εφαρμόζεται stratified 5-fold cross-validation (εξωτερικό CV), οπότε αυτό διασπάται 5 φορές σε σύνολο εκπαίδευσης και σύνολο ελέγχου, σε αναλογία 4:1 και με διαφορετικό τμήμα να αποτελεί κάθε φορά το σύνολο ελέγχου. Για μεγαλύτερη ευκολία αναφέρουμε τα σύνολα εκπαίδευσης και ελέγχου του  $i$ -οστού fold του εξωτερικού CV, ως  $Train Set_i$  και  $Test Set_i$  αντίστοιχα, όπου  $i = 1, 2, \dots, 5$ . Έπειτα, σε κάθε fold  $i$ , το αντίστοιχο  $Train Set_i$  υφίσταται εξισορρόπηση δεδομένων με κάποια επιλεγμένη μέθοδο υπερδειγματοληψίας, εκ των SMOTE, BSMOTE, Random και ADASYN. Μετά την εξισορρόπησή του, το κάθε  $Train Set_i$  χρησιμοποιείται για την εύρεση των βέλτιστων υπερπαραμέτρων για το συγκεκριμένο εξωτερικό fold. Ο προσδιορισμός αυτός πραγματοποιείται χρησιμοποιώντας stratified 3-fold cross-validation (εσωτερικό CV), σε συνδυασμό με αναζήτηση πλέγματος (grid search). Οι υπερπαραμέτροι που κάθε φορά βελτιστοποιούνται μέσω της εξαντλητικής αναζήτησης είναι αυτές που παρουσιάζονται στον πίνακα 6.2. Ως μετρική αξιολόγησης για το εσωτερικό CV χρησιμοποιείται η AUC μετρική. Μετά από την αξιολόγηση όλων των συνδυασμών που ορίζονται από το hyperparameter space, με χρήση 3-fold cross-validation, έχουν προσδιοριστεί οι βέλτιστες τιμές υπερπαραμέτρων με βάση το  $Train Set_i$ . Στη συνέχεια προχωρούμε σε βαθμονόμηση αυτού του προκύπτοντος, βέλτιστου μοντέλου με χρήση της Temperature Scaling τεχνικής. Για το σκοπό αυτό, ένα 10% του  $Train Set_i$  χρησιμοποιείται ως σύνολο επικύρωσης για τον προσδιορισμό της αντίστοιχης παραμέτρου  $T$ . Η εκπαίδευση του, βαθμονομημένου μοντέλου βέλτιστων υπερπαραμέτρων γίνεται με χρήση του υπολοίπου 90%

$Train Set_i$  (έστω  $Train Set'_i$ ), ενώ στη συνέχεια η αξιολόγησή του γίνεται με βάση το  $Test Set_i$ . Προχωρούμε στην αξιολόγηση αυτή με δύο διαφορετικά κατώφλια ταξινόμησης: με το προκαθορισμένο κατώφλι ίσο με 0.5 και με ένα ακόμα, προσαρμοσμένο κατώφλι ταξινόμησης. Η προσαρμογή του κατωφλίου γίνεται με χρήση ενός validation set που προκύπτει ως υποσύνολο των δεδομένων που χρησιμοποιούνται τελικά για την εκπαίδευση ( $Train Set'_i$ ), με αναλογία training και validation set 70:30 αντίστοιχα. Δοκιμάζονται κατώφλια ταξινόμησης από το 0 έως το 1, με βήμα 0.001, ενώ κριτήριο επιλογής του τελικού κατωφλίου αποτελεί η μεγιστοποίηση της F1-score μετρικής για αυτό το σύνολο επικύρωσης. Για κάθε επανάληψη για το μοντέλο αποθηκεύονται τα αποτελέσματα και των δύο αξιολογήσεων, με ή χωρίς προσαρμοσμένο κατώφλι. Στόχος μας είναι να συμπεράνουμε κατά πόσο η τεχνική threshold tuning βελτιώνει τα τελικά αποτελέσματα στην περίπτωση μας.

### 6.3.2 Αξιολόγηση μοντέλου μηχανικής μάθησης

Η αξιολόγηση του μοντέλου κατά το εξωτερικό CV της nested cross-validation πραγματοποιείται με τις μετρικές: Accuracy, Precision, Recall, AUC, F1-score, Brier Score και Expected Calibration Error (ECE). Παράλληλα, μαζί με αυτές τις μετρικές, υπολογίζεται και το confidence score του εξεταζόμενου μοντέλου για τη θετική κλάση, ως η μέση τιμή των πιθανοτήτων πρόβλεψης που αντιστοιχούν σε δείγματα που έχουν κατηγοριοποιηθεί στην κλάση αυτή. Οι μετρικές Accuracy, Precision, Recall, AUC και F1-score (οι οποίες παρουσιάστηκαν στην υποενότητα 3.5.4) χρησιμοποιούνται για την αξιολόγηση της προβλεπτικής ικανότητας του δικτύου, με τις AUC και F1-score να θεωρούνται ως οι πλέον αντιπροσωπευτικές, λόγω της παρατηρούμενης υψηλής ανισορροπίας των κλάσεων. Αντίστοιχα, οι Brier Score, ECE μετρικές συμπεριλαμβάνονται με στόχο τη συμπληρωματική αξιολόγηση του ποσοστού βαθμονόμησης του μοντέλου.

Τέλος, για τη σύγκριση των αποτελεσμάτων διαφορετικών μεθόδων υπερδειγματοληψίας γίνεται χρήση στατιστικών tests. Συγκεκριμένα, χρησιμοποιήθηκε το paired t-test και το Wilcoxon signed rank test. Καθώς το paired t-test έχει ως προϋπόθεση εφαρμογής την κανονικότητα των διαφορών των δύο groups που συγκρίνονται κάθε φορά, σε περίπτωση που αυτή καταπατάται, γίνεται εναλλακτική χρήση του Wilcoxon signed-rank test. Ο έλεγχος αυτός εφαρμόζεται για τον ίδιο σκοπό με το paired t-test, όμως δεν διαθέτει την προαναφερθείσα προϋπόθεση εφαρμογής. Σημειώνουμε ότι ο έλεγχος της κανονικότητας πραγματοποιείται με χρήση του Shapiro-Wilk test.

### 6.3.3 Ερμηνευσιμότητα μοντέλου μηχανικής μάθησης

Για την ερμηνεία του μοντέλου επιστρατεύεται η LIME τεχνική, η οποία προσφέρει εξηγήσεις σε τοπικό επίπεδο, μέσω επεξήγησης επιλεγμένων στιγμιοτύπων. Για κάθε σύνολο δεδομένων, οι δοκιμές με τη LIME μέθοδο πραγματοποιούνται με χρήση του ομώνυμου Python πακέτου και του αντίστοιχου Recurrent Tabular Explainer, ο οποίος εφαρμόζεται για την επεξήγηση αναδρομικών δικτύων που κάνουν χρήση δεδομένων σε μορφή πίνακα. Όπως φάνηκε και από το κεφάλαιο 4, η ποιότητα των τελικών εξηγήσεων εξαρτάται σε μεγάλο βαθμό από τον τρόπο κατασκευής των διαταραγμένων δειγμάτων, τη μέθοδο υπολογισμού του βαθμού ομοιότητας, τον αριθμό και τη μέθοδο επιλογής των  $m$  πιο χρήσιμων χαρακτηριστικών και το εφαρμοζόμενο, απλό μοντέλο  $g$ . Σε αυτήν την υποενότητα μιλάμε για τις σχετικές επιλογές που έγιναν στην παρούσα εργασία.

Όσον αφορά στη διατάραξη ενός δείγματος, για tabular δεδομένα σαν και αυτά που μεταχειριζόμαστε εμείς, η επιλεγμένη μέθοδος από το LIME πακέτο που χρησιμοποιείται είναι η εξής: αρχικά εξάγονται τα στατιστικά στοιχεία των τιμών που λαμβάνει κάθε χαρακτηριστικό στο σύνολο εκπαίδευσης. Στη συνέχεια, λαμβάνονται δείγματα από μια κανονική κατανομή (0,1) και ύστερα, πραγματοποιώντας την αντίστροφη διαδικασία του κεντραρίσματος γύρω από τη μέση τιμή (mean-centering) και της κλιμάκωσης (scaling), με χρήση των μέσων τιμών και των τυπικών αποκλίσεων που υπολογίστηκαν από την στατιστική ανάλυση κάθε συνόλου τιμών των χαρακτηριστικών στο σύνολο εκπαίδευσης, παράγονται νέες τιμές χαρακτηριστικών και κατ' επέκταση τα διαταραγμένα δείγματα. Αυτό σημαίνει ότι τα διαταραγμένα δείγματα είναι στην πραγματικότητα ανεξάρτητα από το δείγμα ενδιαφέροντος, καθιστώντας τον υπολογισμό του βαθμού ομοιότητας ακόμη πιο κρίσιμο, καθώς τελικά αυτός ο βαθμός είναι το μόνο πράγμα που εξασφαλίζει την τοπικότητα γύρω από το δείγμα που εξετάζεται. Σχετικά με την εκτίμηση του βαθμού ομοιότητας, αυτή υπολογίστηκε με βάση την προεπιλεγμένη μετρική απόστασης (ευκλείδια απόσταση), η οποία μετατρέπεται σε βαθμό ομοιότητας χρησιμοποιώντας έναν εκθετικό πυρήνα εξομάλυνσης, με προεπιλεγμένο πλάτος ίσο με 0.75 φορές την τετραγωνική ρίζα του πλήθους των χαρακτηριστικών. Για το απλό μοντέλο  $g$ , εφαρμόστηκε μέθοδος παλινδρόμησης τύπου ridge (ridge regression). Ακόμα, όσον αφορά τα χαρακτηριστικά που χρησιμοποιούνται για την εξήγηση, δίνεται η επιλογή στον χρήστη να ορίσει ο ίδιος το πλήθος  $m$  που επιθυμεί να συμπεριλάβει. Για την μέθοδο επιλογής των  $m$  χαρακτηριστικών, σε περίπτωση όπου το  $m$  είναι μικρότερο από το συνολικό πλήθος διαθέσιμων features, προσφέρονται οι εξής αλγόριθμοι:

1. Επιλογή προς τα εμπρός: Τα χαρακτηριστικά προστίθενται ένα-ένα ανάλογα με τη βελτίωση που προσφέρουν στην προσέγγιση του σύνθετου γενικού μοντέλου από το απλό.
2. Επιλογή προς τα πίσω: Ξεκινώντας από το μοντέλο που περιέχει όλα τα χαρακτηριστικά εξετάζεται ένα-ένα κάθε χαρακτηριστικό ως προς το εάν θα φέρει μεγαλύτερη βελτίωση η αφαίρεση του, μέχρι να επιτευχθεί ένα μοντέλο με  $m$  χαρακτηριστικά.
3. Υψηλότερα βάρη: Επιλέγονται τα  $m$  χαρακτηριστικά με το υψηλότερο απόλυτο βάρος στην ridge παλινδρόμηση που έχει προκύψει ύστερα από προσαρμογή στα αποτελέσματα-προβλέψεις του σύνθετου μοντέλου.
4. LASSO: Επιλέγονται τα  $m$  χαρακτηριστικά που είναι λιγότερο επιρρεπή σε συρρίκνωση βάσει της διαδρομής κανονικοποίησης μιας LASSO regression εφαρμοσμένης στα αποτελέσματα του σύνθετου μοντέλου.

Στην δική μας υλοποίηση, για την περίπτωση πρόβλεψης με ορίζοντα 0-30 min επιλέξαμε να χρησιμοποιήσουμε όλα τα χαρακτηριστικά εισόδου (δηλαδή τίθεται  $m = 14$ ), ενώ για την περίπτωση ορίζοντα πρόβλεψης 30-60 min χρησιμοποιούνται  $m = 15$  από τα 20 χαρακτηριστικά, τα οποία επιλέγονται με την μέθοδο υψηλότερων βαρών.

# Κεφάλαιο 7

## Αποτελέσματα και συζήτηση

Στο παρόν κεφάλαιο παρουσιάζονται και σχολιάζονται τα αποτελέσματα της αξιολόγησης του αναπτυχθέντος μοντέλου, για κάθε σύνολο δεδομένων. Στη συνέχεια, στο δεύτερο τμήμα, παρέχονται τα αποτελέσματα εφαρμογής της επιλεγμένης τεχνικής ερμηνευσιμότητας LIME.

### 7.1 Πρόβλεψη υπεργλυκαιμικών επεισοδίων

Για την πρόβλεψη υπεργλυκαιμικών επεισοδίων πραγματοποιήθηκαν δοκιμές με δύο διαφορετικά σύνολα δεδομένων, ανάλογα με το αν στόχος ήταν η πρόβλεψη υπεργλυκαιμιών για prediction horizon 0-30 ή 30-60 λεπτών. Και τα δύο αυτά εξεταζόμενα datasets εμφάνιζαν μέτρια ανισορροπία κλάσεων ίση με 35% και 34% αντίστοιχα. Από τα αποτελέσματα που προέκυψαν έπειτα από την αξιολόγηση του αναπτυχθέντος, αναδρομικού μοντέλου με τη χρήση των μεθόδων του κεφαλαίου 6, φάνηκε ότι η προσαρμογή του κατωφλίου ταξινόμησης δεν ευνόησε την πρόβλεψη για κανένα από τα δύο σύνολα δεδομένων. Για το λόγο αυτό, στο παρόν κεφάλαιο, παραθέτουμε και σχολιάζουμε τα αποτελέσματα που προέκυψαν εφαρμόζοντας το προκαθορισμένο κατώφλι ταξινόμησης, ίσο με 0.5. Στο Παράρτημα Α' παρέχονται συμπληρωματικά οι τιμές των μετρικών αξιολόγησης, έπειτα την προσαρμογή του κατωφλίου με χρήση συνόλου επικύρωσης.

#### 7.1.1 Ορίζοντας πρόβλεψης 0-30 λεπτά

		Accuracy	Precision	Recall	AUC	F1-score
SMOTE	mean	83.643%	78.892%	76.015%	81.902%	76.703%
	std	4.499%	11.861%	6.971%	4.059%	5.483%
BSMOTE	mean	79.12%	65.786%	87.03%	80.922%	74.628%
	std	5.071%	6.698%	7.635%	4.781%	5.475%
Random	mean	83.327%	75.523%	78.309%	82.174%	76.761%
	std	2.87%	5.412%	4.147%	2.813%	3.672%



Πίνακας 7.1: Αποτελέσματα αξιολόγησης για οριζόντια πρόβλεψη 0-30 λεπτά - Υπεργλυκαιμία

ADASYN	mean	80.25%	67.374%	85.18%	81.365%	75.085%
	std	2.799%	3.406%	7.503%	3.453%	3.968%

Πίνακας 7.2: Αποτελέσματα αξιολόγησης για οριζόντια πρόβλεψη 0-30 λεπτά - Υπεργλυκαιμία

		Mean Confidence	ECE	Brier Score
SMOTE	mean	83.892%	9.417%	0.1265
	std	2.08%	1.927%	0.0283
BSMOTE	mean	80.34%	14.003%	0.1432
	std	4.198%	4.2%	0.0342
Random	mean	83.907%	10.581%	0.1248
	std	3.275%	1.245%	0.0258
ADASYN	mean	81.301%	14.276%	0.1379
	std	3.592%	2.329%	0.0278

Στους παραπάνω πίνακες (πίνακας 7.1, πίνακας 7.2) μπορούμε να δούμε τα αποτελέσματα των μετρικών αξιολόγησης που εξετάζονται στα πλαίσια της παρούσας διπλωματικής, για το διάστημα πρόβλεψης 0 έως 30 λεπτά. Οι μετρικές οι οποίες μας ενδιαφέρουν περισσότερο είναι οι AUC και F1-score. Εφόσον τα δεδομένα είναι μη ισορροπημένα, η μετρική της ακρίβειας δεν μπορεί να δώσει εξίσου σημαντικές πληροφορίες για την απόδοση του μοντέλου. Αντίθετα, η F1 μετρική λαμβάνει υπόψιν στον ίδιο βαθμό τα δείγματα της θετικής κλάσης, τα οποία σε αυτήν την περίπτωση αντιστοιχούν σε υπεργλυκαιμίες. Παρατηρούμε ότι οι βέλτιστες τιμές AUC και F1-score (AUC = 82.174%, F1-score = 76.761%) αντιστοιχούν στην ίδια μέθοδο εξισορρόπησης δεδομένων και συγκεκριμένα στη μέθοδο της τυχαίας (Random) υπερδειγματοληψίας. Κατά τη τεχνική αυτή, δείγματα της κλάσης μειονότητας επαναλαμβάνονται αυτούσια, με στόχο την άμβλυση των συνεπειών της ανισορροπίας των κλάσεων. Όσον αφορά την αξιολόγηση της βαθμονόμησης του δικτύου, οι μετρικές που διαθέτουν τη μεγαλύτερη αξία είναι η Expected Calibration Error (ECE) και δευτερευόντως, η Brier Score μετρική. Παρατηρούμε ότι η μικρότερη τιμή ECE, ίση με 9.417%, παρατηρείται για τη Synthetic Minority Oversampling τεχνική (SMOTE), όπου τα δείγματα της κλάσης μειονότητας αυξάνονται μέσω της σύνθεσης νέων θετικών δειγμάτων, με βάση τα υπάρχοντα. Αντίστοιχα, το βέλτιστο Brier Score (0.1248) σημειώνεται για την περίπτωση εφαρμογής τυχαίας υπερδειγματοληψίας, αν και αυτό εμφανίζει μικρή απόκλιση από το αντίστοιχο αποτέλεσμα για τη SMOTE τεχνική. Αξίζει να επισημανθεί ότι οι διαφορές αυτές θα μπορούσαν να οφείλονται στη στοχαστικότητα των αλγορίθμων εκπαίδευσης, καθώς οι μετρικές αξιολόγησης κατά μήκος των διαφορετικών τεχνικών εξισορρόπησης κυμαίνονται σε παρόμοια επίπεδα με μικρές αποκλίσεις. Όσον αφορά τη μέση σιγουριά (mean confidence) του μοντέλου κατά τη πρόβλεψη της θετικής κλάσης, παρατηρούμε ότι αυτή κινείται σε σχετικά υψηλά επίπεδα, τα οποία προσεγγίζουν τις τιμές ακρίβειας. Για την αξιολόγηση της αξιοπιστίας των mean confidence τιμών μπορεί να

ληφθεί υπόψιν το ποσοστό βαθμονόμησης του δικτύου. Καθώς στη συγκεκριμένη περίπτωση οι τιμές ECE δεν κυμαίνονται σε υπερβολικά υψηλά επίπεδα, συμπεραίνουμε ότι η σιγουριά αυτή δεν αποκλίνει πολύ από την πραγματικότητα.

Για τη σύγκριση των διαφόρων μοντέλων με πιο αξιόπιστο τρόπο πραγματοποιήθηκαν ακόμα στατιστικά tests για κάθε συνδυασμό. Τα πειράματα αυτά διεξήχθησαν για τις κύριες μετρικές ενδιαφέροντος, F1-score, AUC και ECE και τα αποτελέσματα τους φαίνονται παρακάτω.

Πίνακας 7.3: Αποτελέσματα Wilcoxon-signed rank test και paired t-test για την μετρική AUC και ορίζοντα πρόβλεψης 0-30 λεπτών - Υπεργλυκαιμία

<b>AUC</b>			
	SMOTE	BSMOTE	Random
BSMOTE	1.0		
Random	0.8125	0.625	
ADASYN	0.5210	0.8125	0.625

Πίνακας 7.4: Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική F1-score και ορίζοντα πρόβλεψης 0-30 λεπτών - Υπεργλυκαιμία

<b>F1-score</b>			
	SMOTE	BSMOTE	Random
BSMOTE	0.8125		
Random	0.8125	0.4375	
ADASYN	0.4375	1.0	0.1441

Πίνακας 7.5: Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική ECE και ορίζοντα πρόβλεψης 0-30 λεπτών - Υπεργλυκαιμία

<b>ECE</b>			
	SMOTE	BSMOTE	Random
BSMOTE	0.0625		
Random	0.125	0.1073	
ADASYN	0.0625	1.0	0.0625

Για τη σύγκριση κάθε συνδυασμού μοντέλων με στατιστικές μεθόδους χρησιμοποιήθηκαν τα επιμέρους αποτελέσματα που είχαν προκύψει από κάθε fold για κάθε μία από τις εξεταζόμενες μετρικές. Το paired t-test εφαρμόστηκε στις περιπτώσεις όπου ίσχυε η υπόθεση περί κανονικότητας των διαφορών των δύο συγκρινόμενων ομάδων, η οποία και ελέγχθηκε με χρήση του Shapiro-Wilk test. Στην περίπτωση όπου αυτή η υπόθεση δεν ευσταθούσε προχωρήσαμε σε εφαρμογή του αντίστοιχου Wilcoxon signed-rank ελέγχου. Από τα αποτελέσματα που προκύπτουν για τις μετρικές AUC, F1-score και ECE καταλήγουμε στο ότι οι παρατηρούμενες αποκλίσεις δεν διαθέτουν στατιστική σημασία, οπότε δεν είναι δυνατή η άμεση εξαγωγή συμπερασμάτων σχετικά με τη βέλτιστη τεχνική εξισορρόπησης για τις μετρικές αυτές. Με στόχο την πιο αξιόπιστη σύγκριση των εφαρμοζόμενων τεχνικών υπερδειγματοληψίας, είναι δυνατή η εφαρμογή repeated (nested) cross-validation τεχνικών, έτσι ώστε τα δείγματα των ομάδων που συγκρίνονται κάθε φορά να είναι πιο πολυπληθή.

### 7.1.2 Ορίζοντας πρόβλεψης 30-60 λεπτά

Όπως και προηγουμένως, για την εξαγωγή συμπερασμάτων για τον ορίζοντα πρόβλεψης 30-60 λεπτών λαμβάνουμε υπόψιν κυρίως τις μετρικές AUC και F1-score για την αξιολόγηση της προβλεπτικής ικανότητας του μοντέλου και τις μετρικές ECE και Brier Score για την αξιολόγηση του ποσοστού βαθμονόμησής του. Στη συνέχεια παρουσιάζονται τα αποτελέσματα όλων των εφαρμοζόμενων μετρικών.

Πίνακας 7.6: Αποτελέσματα αξιολόγησης για ορίζοντα πρόβλεψης 30-60 λεπτά - Υπεργλυκαιμία

		Accuracy	Precision	Recall	AUC	F1-score
SMOTE	mean	71.153%	58.186%	68.303%	70.453%	62.235%
	std	5.726%	9.126%	4.432%	3.85%	4.214%
BSMOTE	mean	67.604%	52.567%	75.405%	69.455%	61.468%
	std	4.519%	4.098%	9.523%	2.752%	2.882%
Random	mean	73.405%	60.921%	65.03%	71.39%	62.745%
	std	3.501%	5.956%	3.536%	2.929%	3.65%
ADASYN	mean	70.414%	55.536%	72.162%	70.844%	62.534%
	std	1.69%	2.768%	5.377%	0.977%	0.995%

Πίνακας 7.7: Αποτελέσματα αξιολόγησης για ορίζοντα πρόβλεψης 30-60 λεπτά - Υπεργλυκαιμία

		Mean Confidence	ECE	Brier Score
SMOTE	mean	75.383%	15.551%	0.1932
	std	2.359%	5.757%	0.0209
BSMOTE	mean	67.108%	15.568%	0.1995
	std	2.232%	3.333%	0.0172
Random	mean	72.835%	11.067%	0.1807
	std	2.173%	1.511%	0.0094
ADASYN	mean	69.266%	15.243%	0.1925
	std	1.986%	3.347%	0.0091

Πραγματοποιώντας μια απλή επισκόπηση όλων των μετρικών για όλες τις τεχνικές εξισορρόπησης δεδομένων, γίνεται φανερό ότι η γενικότερη απόδοση του μοντέλου είναι χειρότερη σε σχέση με αυτή που είχε σημειωθεί για τον ορίζοντα πρόβλεψης 0-30 λεπτών, κάτι που υποδηλώνει την ικανότητα του μοντέλου να πραγματοποιεί καλύτερες προβλέψεις για το κοντινότερο μέλλον. Όπως και στην περίπτωση πρόβλεψης υπεργλυκαιμιών με ορίζοντα πρόβλεψης 0-30 λεπτά, η παρατηρούμενη F1 μετρική είναι χαμηλότερη από την ακρίβεια για όλες τις μορφές υπερδειγματοληψίας, κάτι που μπορεί να αποδοθεί στην αρχική ανισοροπία των κλάσεων (34%). Παρόλα αυτά, αξίζει να επισημανθεί ότι αυτές οι διαφορές, αν και υπαρκτές, κινούνται σε σχετικά περιορισμένα επίπεδα. Οι υψηλότερες τιμές AUC και F1-score (AUC = 71.39%, F1-score = 62.745%)

παρατηρούνται, όπως και προηγουμένως, για την μέθοδο της τυχαίας υπερδειγματοληψίας. Όσον αφορά το ποσοστό βαθμονόμησης, οι βέλτιστες (δηλαδή χαμηλότερες) τιμές ECE και Brier Score μετρικών αντιστοιχούν επίσης στην Random Oversampling τεχνική. Συγκεκριμένα, με χρήση τυχαίας υπερδειγματοληψίας σημειώνεται ECE ίσο με 11.067% και Brier Score ίσο με 0.1807. Η μέση σιγουριά του δικτύου κατά την πρόβλεψη της κλάσης μειονότητας κυμαίνεται σε χαμηλότερα επίπεδα από ότι στην περίπτωση ορίζοντα πρόβλεψης 0-30 λεπτών, κάτι που είναι αναμενόμενο κι επιθυμητό, λόγω του πιο απομακρυσμένου χρονικά ορίζοντα πρόβλεψης των 30-60 λεπτών. Και σε αυτήν την περίπτωση πάντως, τα εκτιμώμενα επίπεδα σιγουριάς προσεγγίζουν τις αντίστοιχες τιμές ακρίβειας. Στη συνέχεια παρατίθενται τα αποτελέσματα των στατιστικών tests που πραγματοποιήθηκαν για την εξαγωγή συμπερασμάτων για το κατά πόσο οι παρατηρούμενες αποκλίσεις μεταξύ των μοντέλων διαθέτουν στατιστική σημασία.

Πίνακας 7.8: Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική AUC για ορίζοντα πρόβλεψης 30-60 λεπτών - Υπεργλυκαιμία

<b>AUC</b>			
	SMOTE	BSMOTE	Random
BSMOTE	1.0		
Random	0.625	0.0625	
ADASYN	1.0	0.3125	1.0

Πίνακας 7.9: Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική F1-score για ορίζοντα πρόβλεψης 30-60 λεπτών - Υπεργλυκαιμία

<b>F1-score</b>			
	SMOTE	BSMOTE	Random
BSMOTE	1.0		
Random	0.4375	0.4375	
ADASYN	1.0	0.8125	1.0

Πίνακας 7.10: Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική ECE για ορίζοντα πρόβλεψης 30-60 λεπτών - Υπεργλυκαιμία

<b>ECE</b>			
	SMOTE	BSMOTE	Random
BSMOTE	1.0		
Random	0.4375	0.0625	
ADASYN	1.0	1.0	0.125

Από τους διεξαχθέντες στατιστικούς ελέγχους συμπεραίνουμε ότι για τις εξεταζόμενες μετρικές δεν υπάρχει στατιστική ανεξαρτησία μεταξύ των διαφορετικών προσεγγίσεων, καθώς δεν είναι δυνατή η απόρριψη της μηδενικής υπόθεσης (null hypothesis). Επομένως, οριστικά συμπεράσματα όσον αφορά τη βέλτιστη μέθοδο είναι επισφαλή. Συνοψίζοντας, για την περίπτωση της υπεργλυκαιμίας, τα αποτελέσματα των μοντέλων μας κρίνονται ικανοποιητικά, αν και η προϋπάρχουσα ανισορροπία των κλάσεων πλήττει την γενικότερη απόδοση σε όλες τις περιπτώσεις, όπως φαίνεται και από τη διαφορά

μεταξύ των accuracy και F1-score τιμών. Και στις δύο περιπτώσεις ορίζοντα πρόβλεψης, η μέθοδος random oversampling εμφάνισε τα καλύτερα αποτελέσματα για τις περισσότερες από τις μετρικές ενδιαφέροντος. Παρόλα αυτά, λόγω μη στατιστικής ανεξαρτησίας, η οριστική εξαγωγή συμπεράσματος δεν κρίνεται δυνατή. Όσον αφορά τη μέση σιγουριά πρόβλεψης της κλάσης μειονότητας, αυτή κυμαίνεται σε κοντινά επίπεδα με τα εξαγόμενα αποτελέσματα των υπόλοιπων μετρικών αξιολόγησης accuracy, AUC και F1-score και κυρίως, σε κοντινά επίπεδα με τα αποτελέσματα της accuracy μετρικής. Τέλος, η παρατηρούμενη μειωμένη απόδοση του μοντέλου για τον ορίζοντα πρόβλεψης 30-60 λεπτών, σε σχέση με αυτή που είχε σημειωθεί για ορίζοντα 0-30min, φανερώνει την μεγαλύτερη ευκολία του συστήματος να πραγματοποιεί πιο κοντινές χρονικά προβλέψεις.

## 7.2 Πρόβλεψη υπογλυκαιμικών επεισοδίων

Για την πρόβλεψη υπογλυκαιμικών επεισοδίων, όπως και προηγουμένως, πραγματοποιήθηκαν δοκιμές με δύο διαφορετικά σύνολα δεδομένων, ανάλογα με τον ορίζοντα πρόβλεψης (0-30min ή 30-60min). Και τα δύο αυτά εξεταζόμενα datasets εμφάνιζαν ακραία ανισορροπία κλάσεων ίση με 3% και 4% αντίστοιχα. Χαρακτηριστικό των συνόλων αυτών ήταν ακόμα η ύπαρξη περιορισμένων αριθμητικά δειγμάτων της θετικής κλάσης. Από τα αποτελέσματα που προέκυψαν έπειτα από την αξιολόγηση του αναδρομικού μοντέλου, φάνηκε ότι η προσαρμογή του κατωφλίου ταξινόμησης δεν ευνόησε την πρόβλεψη για κανένα από τα δύο σύνολα δεδομένων. Για το λόγο αυτό, στο παρόν κεφάλαιο, παρουσιάζονται τα αποτελέσματα που προέκυψαν εφαρμόζοντας το default κατώφλι ταξινόμησης, ίσο με 0.5. Στο Παράρτημα Β' παρέχονται συμπληρωματικά οι τιμές των μετρικών αξιολόγησης, έπειτα την προσαρμογή του κατωφλίου με χρήση validation set.

### 7.2.1 Ορίζοντας πρόβλεψης 0-30 λεπτά

Πίνακας 7.11: Αποτελέσματα αξιολόγησης για ορίζοντα πρόβλεψης 0-30 λεπτά - Υπογλυκαιμία

		Accuracy	Precision	Recall	AUC	F1-score
SMOTE	mean	74.249%	7.643%	63.333%	68.973%	13.498%
	std	6.496%	1.614%	11.304%	3.482%	2.55%
BSMOTE	mean	80.409%	3.547%	21.667%	51.981%	6.047%
	std	3.122%	1.907%	11.304%	4.851%	3.175%
Random	mean	72.311%	7.798%	68.333%	70.386%	13.755%
	std	7.575%	2.61%	18.559%	7.845%	4.043%
ADASYN	mean	77.804%	7.928%	53.333%	65.974%	13.268%
	std	8.1%	5.533%	28.186%	12.071%	8.39%

Πίνακας 7.12: Αποτελέσματα αξιολόγησης για ορίζοντα πρόβλεψης 0-30 λεπτά - Υπογλυκαιμία

		Mean Confidence	ECE	Brier Score
SMOTE	mean	74.813%	16.396%	0.1542
	std	4.54%	7.905%	0.0266
BSMOTE	mean	76.749%	14.938%	0.1325
	std	4.515%	5.796%	0.0313
Random	mean	67.839%	19.519%	0.1483
	std	4.468%	3.783%	0.0357
ADASYN	mean	72.138%	18.346%	0.1345
	std	4.605%	4.98%	0.0407

Πραγματοποιώντας μια συνολική επισκόπηση των αποτελεσμάτων μπορούμε εύκολα να παρατηρήσουμε ότι η απόδοση του δικτύου κατά την πρόβλεψη των υπογλυκαιμιών είναι χειρότερη σε σχέση με αυτή που σημειώθηκε για την περίπτωση των υπεργλυκαιμιών, κάτι που γίνεται άμεσα αντιληπτό και από τις τιμές της F1-score μετρικής. Το γεγονός αυτό μπορεί να αιτιολογηθεί από την ακραία ανισορροπία των κλάσεων, αλλά και από τον περιορισμένο αριθμό διαθέσιμων δειγμάτων θετικής κλάσης. Οι εφαρμοζόμενες τεχνικές εξισορρόπησης καθιστούν το μοντέλο ικανό να αναγνωρίζει τα δείγματα μειωμένης όταν τα συναντά (όπως φαίνεται και από τις σχετικά υψηλές τιμές recall), όμως δεν του δίνουν τη δυνατότητα να διακρίνει τις ουσιαστικές διαφορές μεταξύ αρνητικών και θετικών δειγμάτων, κάτι που οδηγεί σε μεγάλο αριθμό false positive κατηγοριοποιήσεων (όπως προκύπτει από το χαμηλό precision). Πράγματι, η ισχύς των μεθόδων υπερδειγματοληψίας που χρησιμοποιούνται περιορίζεται καθώς το πλήθος των διαθέσιμων δεδομένων μειωμένης μειώνεται, αφού η παραγωγή νέων δεδομένων πραγματοποιείται με βάση τα υπάρχοντα. Ο μικρός αριθμός θετικών δειγμάτων αυξάνει ακόμα περισσότερο τη διακύμανση των αποτελεσμάτων ανά fold, κάτι που υποδηλώνεται από το υψηλότερο standard deviation σε ορισμένες περιπτώσεις, παρά τη χρήση stratified cross-validation. Συνολικά συμπεραίνουμε ότι οι μέθοδοι υπερδειγματοληψίας που εφαρμόστηκαν, αν και σίγουρα ωφελούν, δεν επαρκούν για να βελτιώσουν ικανοποιητικά τη διακριτική ικανότητα του δικτύου σε περιπτώσεις όπου εμφανίζεται μεγάλη ανισορροπία κλάσεων, σε συνδυασμό με μικρό αριθμητικά πλήθος θετικών δειγμάτων.

Όσον αφορά την τεχνική threshold tuning, αν και από τη δυσανάλογη σχέση precision και recall φαίνεται ότι πολύ πιθανώς σε αυτήν την περίπτωση να διαθέτει αξία, τα αποτελέσματά μας από την εφαρμογή προσαρμοσμένου κατωφλίου (παράρτημα Β) δεν φαίνεται να διαφέρουν σημαντικά. Αυτό θα μπορούσε να ωφείλεται στο γεγονός ότι σε κάθε fold το προσαρμοσμένο κατώφλι επιλέγεται με βάση ενός μόνο validation set, που λόγω της ανισορροπίας των κλάσεων μπορεί να διαθέτει σημαντικές διαφορές από το test set, με βάση το οποίο γίνεται τελικά η αξιολόγηση. Συμπεραίνουμε επομένως ότι η χρήση πιο σύνθετων τεχνικών threshold tuning, όπως για παράδειγμα η χρήση cross-validation για την αξιολόγηση σε πολλά validation sets, αλλά και η άμεση μείωση του κατωφλίου με στόχο την εξισορρόπηση της σχέσης precision-recall, θα μπορούσαν να ωφελήσουν στη συγκεκριμένη περίπτωση. Όσον αφορά τη σιγουριά και τη βαθμονόμηση του δικτύου, παρατηρούμε ότι το δίκτυο είναι υπεραισιόδοξο κατά τις προβλέψεις

του για τη θετική κλάση, κάτι που υποδεικνύει ότι η βαθμονόμησή του με το 10% των δεδομένων του training set σε κάθε fold δεν επαρκεί. Οι τιμές της ECE μετρικής, οι οποίες μπορούν να χρησιμοποιηθούν ως ένα κριτήριο αξιολόγησης της αξιοπιστίας των εξαγόμενων confidence scores, είναι υψηλότερες από αυτές της υπεργλυκαιμίας, αναδεικνύοντας τη μειωμένη ικανότητα του δικτύου να εξάγει αξιόπιστες πιθανότητες πρόβλεψης. Παρόλα αυτά, η απόκλιση τους από τις τιμές που είχαν παρατηρηθεί για την περίπτωση της υπεργλυκαιμίας, δεν είναι ανάλογη της απόκλισης των τιμών F1-score στις δύο αυτές περιπτώσεις. Το γεγονός αυτό δημιουργεί ερωτήματα για το κατά πόσο η ECE μετρική, η οποία υπολογίζεται με βάση τη διαφορά των παρατηρούμενων accuracy και confidence μετρικών ανά bin [86], είναι αρκετή για να αποτυπώσει τη βαθμονόμηση δικτύων σε περιπτώσεις μεγάλης ανισορροπίας των κλάσεων, όπου η μετρική της ακρίβειας παύει να είναι αντιπροσωπευτική.

Τέλος επισημαίνουμε ότι οι βέλτιστες τιμές AUC και F1-score (AUC = 70.386%, F1-score = 13.755%) σημειώνονται και πάλι για την περίπτωση τυχαίας υπερδειγματοληψίας. Για αυτήν την τεχνική εξισορρόπησης παρατηρούνται και ο μικρότερος βαθμός mean confidence και η υψηλότερη τιμή ECE, που σε αυτήν την περίπτωση είναι πιο αντιπροσωπευτικά. Παρόλα αυτά, και με βάση και τους στατιστικούς έλεγχους που πραγματοποιήθηκαν, δεν είναι δυνατή η οριστική εξαγωγή πορισμάτων.

Πίνακας 7.13: Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική AUC για ορίζοντα πρόβλεψης 0-30 λεπτών - Υπογλυκαιμία

<b>AUC</b>			
	SMOTE	BSMOTE	Random
BSMOTE	0.0625		
Random	0.8125	0.0625	
ADASYN	0.6046	0.125	0.465208

Πίνακας 7.14: Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική F1-score για ορίζοντα πρόβλεψης 0-30 λεπτών - Υπογλυκαιμία

<b>F1-score</b>			
	SMOTE	BSMOTE	Random
BSMOTE	0.0625		
Random	1.0	0.0625	
ADASYN	0.625	0.06788	1.0

Πίνακας 7.15: Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική ECE για ορίζοντα πρόβλεψης 0-30 λεπτών - Υπογλυκαιμία

<b>ECE</b>			
	SMOTE	BSMOTE	Random
BSMOTE	0.625		
Random	0.4375	0.125	
ADASYN	0.5918	0.0625	0.4375

## 7.2.2 Ορίζοντας πρόβλεψης 30-60 λεπτά

Πίνακας 7.16: Αποτελέσματα αξιολόγησης για ορίζοντα πρόβλεψης 30-60 λεπτά - Υπογλυκαιμία

		Accuracy	Precision	Recall	AUC	F1-score
SMOTE	mean	67.796%	7.122%	59.0%	63.59%	12.663%
	std	4.037%	3.525%	19.209%	14.599%	3.341%
BSMOTE	mean	81.46%	6.812%	36.0%	59.706%	11.401%
	std	5.496%	4.556%	19.369%	13.521%	4.795%
Random	mean	67.418%	7.227%	59.0%	63.388%	12.853%
	std	4.535%	3.358%	16.907%	13.34%	2.97%
ADASYN	mean	69.104%	7.583%	59.0%	64.261%	13.418%
	std	4.362%	3.805%	19.209%	15.449%	3.762%

Πίνακας 7.17: Αποτελέσματα αξιολόγησης για ορίζοντα πρόβλεψης 30-60 λεπτά - Υπογλυκαιμία

		Mean Confidence	ECE	Brier Score
SMOTE	mean	70.192%	25.378%	0.1806
	std	2.081%	2.832%	0.0241
BSMOTE	mean	72.662%	16.589%	0.1252
	std	3.262%	5.764%	0.0289
Random	mean	70.278%	24.729%	0.1849
	std	2.686%	4.357%	0.0326
ADASYN	mean	74.157%	24.943%	0.1908
	std	1.666%	4.161%	0.031

Με μια απλή επισκόπηση των αποτελεσμάτων παρατηρούμε ότι και σε αυτήν την περίπτωση ορίζοντα πρόβλεψης, η διακριτική ικανότητα του δικτύου δεν επαρκεί για την ορθή πρόβλεψη των δειγμάτων μειονότητας. Η ανεπάρκεια αυτή αντανακλάται στα χαμηλά αποτελέσματα της F1-score και μπορεί να αποδοθεί και πάλι στην υψηλή ανισορροπία των κλάσεων και στο περιορισμένο πλήθος διαθέσιμων θετικών δειγμάτων. Μην έχοντας αρκετά υψηλό και ποικιλόμορφο πληθυσμό δεδομένων μειοψηφίας, οι εφαρμοζόμενες τεχνικές υπερδειγματοληψίας δεν επαρκούν για την όξυνση της διαχωριστικής ικανότητας του δικτύου. Η δυσανάλογη σχέση precision και recall τιμών φανερώνει επίσης ότι η χρήση ενός προσαρμοσμένου κατωφλίου ταξινόμησης πιθανώς να φέρει καλύτερα και πιο ισορροπημένα αποτελέσματα. Στην περίπτωση μας, η προσαρμογή του κατωφλίου με χρήση ενός validation set δεν έφερε καλύτερα αποτελέσματα, κάτι που θα μπορούσε να αποδοθεί σε ποιοτική διαφορά του χρησιμοποιούμενου συνόλου επικύρωσης και του test set για το οποίο τελικά γίνεται η αξιολόγηση, πιθανώς λόγω διαφορετικής ποιοτικής και ποσοτικής κατανομής των θετικών δειγμάτων στα σύνολα αυτά. Συμπεραίνουμε επομένως ότι η εφαρμογή πιο εξεζητημένων threshold tuning τεχνικών (π.χ. χρήση cross-validation με στόχο τη χρήση πολλών συνόλων επικύρωσης)



αποτελεί ζητούμενο. Όσον αφορά τα επίπεδα σιγουριάς του δικτύου, παρατηρούμε ότι αυτά παραμένουν υψηλά και επομένως, δεν είναι αξιόπιστα. Η έλλειψη αξιοπιστίας των confidence levels επισημαίνεται και από τις υψηλές τιμές expected calibration error που υποδηλώνουν ότι η βαθμονόμηση του δικτύου με το 10% των δεδομένων του train set άνα fold δεν επαρκεί για το συγκεκριμένο dataset. Τέλος, σημειώνουμε ότι οι βέλτιστες τιμές AUC και F1-score (AUC = 64.261%, F1-score = 13.418%) προκύπτουν με την ADASYN μέθοδο υπερδειγματοληψίας, παρόλα αυτά η οριστική εξαγωγή συμπερασμάτων δεν είναι δυνατή λόγω αδυναμίας απόρριψης της μηδενικής υπόθεσης, όπως προκύπτει και από τους στατιστικούς ελέγχους.

Συνοψίζοντας, τα παραπάνω αποτελέσματα για την περίπτωση της υπογλυκαιμίας αναδεικνύουν την πρόκληση ανάπτυξης βαθμονομημένων και αξιόπιστων μοντέλων σε περιπτώσεις όπου η ανισορροπία των κλάσεων είναι υψηλή και το πλήθος των δεδομένων της θετικής κλάσης δεν επαρκούν πάντα για τον σχηματισμό ομοιόμορφων συνόλων εκπαίδευσης, επικύρωσης και ελέγχου.

Πίνακας 7.18: Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική AUC για ορίζοντα πρόβλεψης 30-60 λεπτών - Υπογλυκαιμία

<b>AUC</b>			
	SMOTE	BSMOTE	Random
BSMOTE	0.0625		
Random	1.0	0.4375	
ADASYN	1.0	0.1875	0.715

Πίνακας 7.19: Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική F1-score για ορίζοντα πρόβλεψης 30-60 λεπτών - Υπογλυκαιμία

<b>F1-score</b>			
	SMOTE	BSMOTE	Random
BSMOTE	0.625		
Random	1.0	0.8125	
ADASYN	0.60702	0.8125	0.715

Πίνακας 7.20: Αποτελέσματα Wilcoxon signed-rank test και paired t-test για την μετρική ECE για ορίζοντα πρόβλεψης 30-60 λεπτών - Υπογλυκαιμία

<b>ECE</b>			
	SMOTE	BSMOTE	Random
BSMOTE	0.1875		
Random	1.0	0.1875	
ADASYN	1.0	0.0625	0.8125

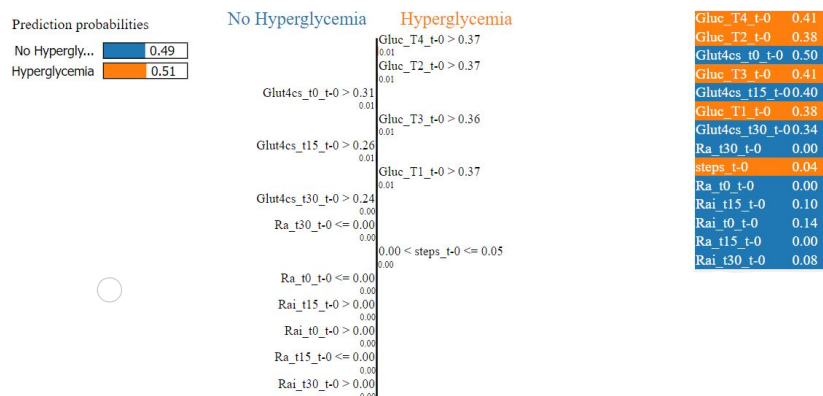
## 7.3 Ερμηνευσιμότητα μοντέλου

Για την ερμηνεία του μοντέλου έγινε χρήση της post-hoc, model-agnostic, local τεχνικής ερμηνευσιμότητας LIME. Η δημοφιλής αυτή τεχνική ερμηνευσιμότητας βασίζει τη λειτουργία της στην υπόθεση ότι κάθε περίπλοκο μοντέλο είναι γραμμικό σε τοπική κλίμακα. Έτσι, η μέθοδος επιχειρεί να προσαρμόσει ένα απλό μοντέλο γύρω από μόνο ένα δείγμα, το οποίο θα μιμείται τη συμπεριφορά του σύνθετου μοντέλου στη περιοχή αυτή. Το απλό αυτό ερμηνεύσιμο μοντέλο μπορεί να χρησιμοποιηθεί για να εξηγήσει τις προβλέψεις του σύνθετου μοντέλου στην περιορισμένη περιοχή που εξετάζεται.

Για τη συγκεκριμένη εργασία χρησιμοποιήθηκε το ομώνυμο lime πακέτο υλοποιημένο σε Python, το οποίο προσφέρει τη δυνατότητα επεξήγησης αναδρομικών μοντέλων με είσοδο αριθμητικά δεδομένα. Για κάθε περίπτωση συνόλου δεδομένων πραγματοποιήθηκε ένα απλό train-test split με αναλογία 80:20, ίση με την αναλογία του εξωτερικού cross-validation του nested cross-validation που εφαρμόστηκε στην κύρια πειραματική διαδικασία. Σε κάθε dataset επιλέξαμε να εφαρμόσουμε τη μέθοδο εξισορρόπησης που είχε αποδώσει καλύτερα με βάση τα αποτελέσματα των κύριων πειραμάτων, ενώ οι υπερπαραμέτροι του δικτύου επιλέχθηκαν με βάση τους βέλτιστους συνδυασμούς υπερπαραμέτρων που είχαν παρατηρηθεί κατά την εκτέλεση της εμφωλευμένης διασταυρούμενης επικύρωσης για το συγκεκριμένο dataset και τεχνική υπερδευγματοληψίας. Στη συνέχεια παρατίθενται κάποιες χρήσιμες παρατηρήσεις που προέκυψαν από την εφαρμογή της LIME τεχνικής.

### 7.3.1 Πρόβλεψη υπεργλυκαιμικών επεισοδίων

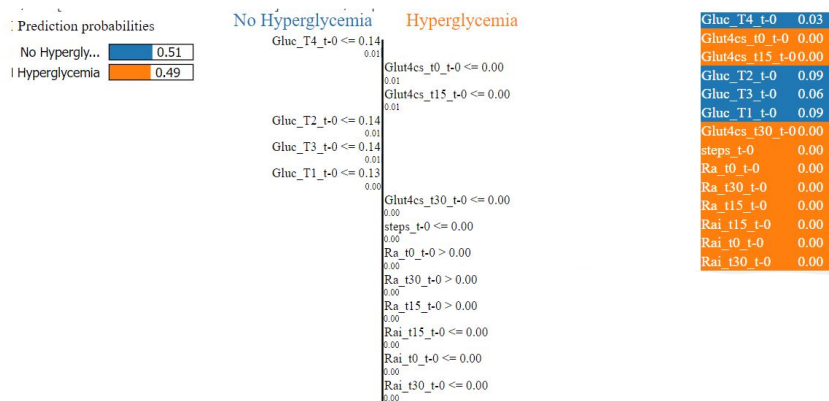
Εκκινούμε την εξέταση μας για την περίπτωση του ορίζοντα πρόβλεψης 0-30 λεπτά, για τον οποίο το μοντέλο μας επέφερε σχετικά ικανοποιητικά αποτελέσματα για όλες τις μετρικές αξιολόγησης. Από τα εξαγόμενα αυτά αποτελέσματα αναμένουμε το δίκτυο να έχει καταφέρει να αποτυπώσει τις πιο σημαντικές συσχετίσεις μεταξύ των χαρακτηριστικών εισόδου και την εμφάνιση ή μη υπεργλυκαιμιών. Στη συνέχεια παρουσιάζονται τα αποτελέσματα εξέτασης ενός θετικού δείγματος που έχει κατηγοριοποιηθεί ορθά (TP):



Σχήμα 7.1: Αποτελέσματα LIME τεχνικής - Υπεργλυκαιμία, 0-30min

Στο δεξί πίνακα παρουσιάζονται τα χρησιμοποιούμενα χαρακτηριστικά εισόδου, συνδεδεμένα από την υπολογιζόμενη "βαρύτητα" (feature importance) που αυτά είχαν

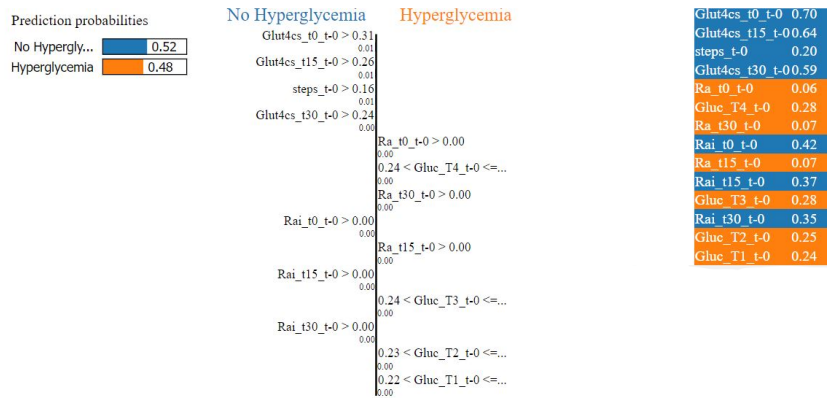
κατά την πραγματοποίηση της πρόβλεψης για το συγκεκριμένο δείγμα. Ανάλογα με τον χρωματισμό τους προκύπτει το εάν αυτά συνέφεραν στην χαρακτηρισμό του δείγματος αυτού ως υπεργλυκαιμία (πορτοκαλί χρώμα) ή μη υπεργλυκαιμία (μπλε χρώμα). Παρατηρούμε ότι στη συγκεκριμένη περίπτωση, οι τιμές γλυκόζης για όλες τις χρονικές τιμές, το πλήθος βημάτων και οι περισσότερες τιμές ρυθμού εμφάνισης γλυκόζης ( $R_a$ ) έχουν συμβάλει στην κατηγοριοποίηση του δείγματος στη θετική κλάση. Αντίθετα, οι τιμές ποσοστού Glut4 στο κυτταρικό τοίχωμα  $Glut4_{cs}$  και οι τιμές του ρυθμού εμφάνισης ινσουλίνης στο αίμα  $R_i$  μειώνουν την πιθανότητα χαρακτηρισμού ως υπεργλυκαιμία. Πράγματι, εξετάζοντας το συγκεκριμένο δείγμα παρατηρούμε ότι οι τιμές γλυκόζης που έχουν καταγραφεί είναι υπεργλυκαιμικές, οι τιμές  $R_i$  και  $Glut4_{cs}$  είναι μη μηδενικές (κάτι που υποδεικνύει πρόσφατη χορήγηση ινσουλίνης), οι τιμές  $R_a$  είναι μηδενικές (κι επομένως δεν έχει καταγραφεί κάποιο πρόσφατο γεύμα) και το πλήθος σημειωμένων βημάτων είναι αρκετό χαμηλό. Όλα αυτά συμφωνούν λογικά με το labeling αυτής της περίπτωσης ως υπεργλυκαιμία, όπως και είναι το ορθό.



Σχήμα 7.2: Αποτελέσματα LIME τεχνικής - Υπεργλυκαιμία, 0-30min

Στην παραπάνω περίπτωση φαίνεται ένα δείγμα το οποίο έχει κατηγοριοποιηθεί ορθώς στην αρνητική κλάση (TN). Παρατηρούμε ότι όλα τα χαρακτηριστικά εισόδου εκτός των μετρήσεων γλυκόζης συνεισφέρουν στο labeling ως υπεργλυκαιμία. Όντως, οι μη μηδενικές  $R_a$  τιμές που υποδηλώνουν πρόσφατα καταγεγραμμένο γεύμα, οι μηδενικές  $Glut_{cs}$ ,  $R_{ai}$  τιμές και το μικρό πλήθος βημάτων συμβάλλουν στον χαρακτηρισμό ως hyperglycemia, παρόλα αυτά οι καταγεγραμμένες μετρήσεις γλυκόζης είναι αρκετά χαμηλές για να 'σπρώξουν' την πρόβλεψη στην αρνητική κλάση.

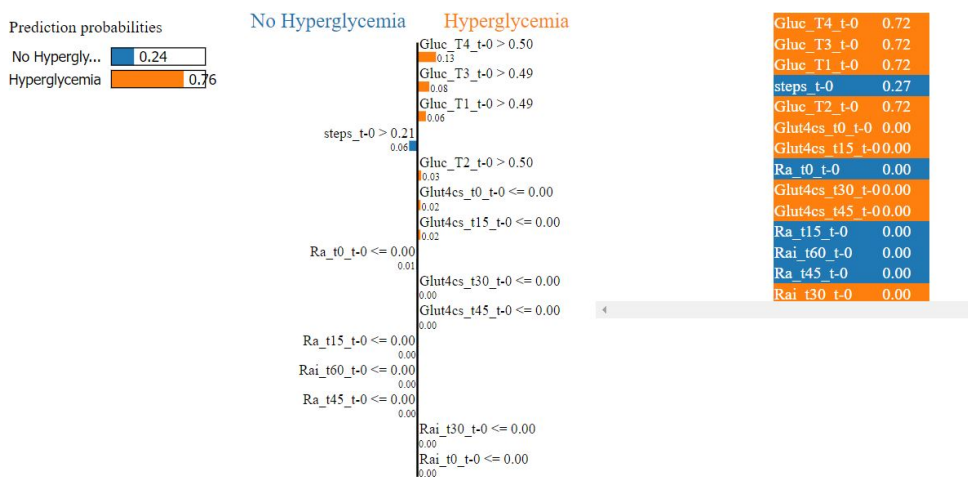
Αντίστοιχα, στη συνέχεια φαίνεται μια false negative περίπτωση, όπου έχει καταγραφεί πρόσφατο γεύμα, πρόσφατη χορήγηση ινσουλίνης, το αντίστοιχο πλήθος βημάτων είναι σχετικά υψηλό, όπως και οι τιμές γλυκόζης. Παρατηρούμε ότι αν και η συμβολή κάθε χαρακτηριστικού είναι προς τη σωστή κατεύθυνση, το μοντέλο έχει υπερεκτιμήσει την ικανότητα της εισαχθείσας ινσουλίνης να μειώσει τα επίπεδα γλυκόζης, όπως φαίνεται και από τις υψηλές feature importance τιμές των σχετιζόμενων με την ινσουλίνη χαρακτηριστικών. Έτσι, έχει προχωρήσει σε (λανθασμένο) labeling του δείγματος ως αρνητικό.



Σχήμα 7.3: Αποτελέσματα LIME τεχνικής - Υπεργλυκαιμία, 0-30min

Πραγματοποιώντας και άλλες τέτοιες παρεμφερείς εξετάσεις μεμονωμένων δειγμάτων για διάφορα αποτελέσματα κατηγοριοποιήσεων, παρατηρούμε ότι για την περίπτωση υπεργλυκαιμίας και ορίζοντα πρόβλεψης 0-30min, τα χαρακτηριστικά εισόδου έχουν συμβάλει στην πρόβλεψη με τον λογικά αναμενόμενο τρόπο. Τα λανθασμένα classifications οφείλονται κυρίως στην ανικανότητα του δικτύου να διακρίνει τα πιο σημαντικά χαρακτηριστικά σε περιπτώσεις όπου πολλοί παράγοντες αλληλεπιδρούν (π.χ. πρόσληψη γεύματος με χορήγηση ινσουλίνης και μέτρια κυμαινόμενες τιμές γλυκόζης εισόδου).

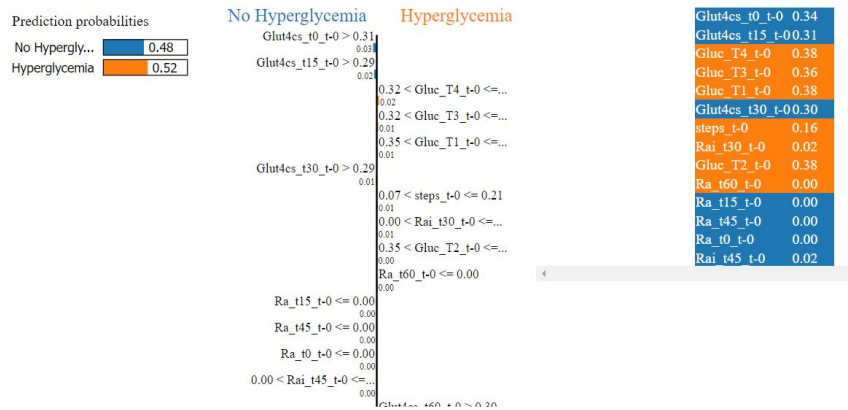
Συνεχίζουμε εξετάζοντας τον ορίζοντα πρόβλεψης 30-60 min όπου τα χαρακτηριστικά εισόδου που εφαρμόζονται κάθε φορά είναι περισσότερα (20 έναντι 14). Για κάθε επεξήγηση εμφανίζουμε την επίδραση των 15 πιο σημαντικών χαρακτηριστικών με βάση την εκτίμηση του απλού LIME regressor. Αρχικά, μελετούμε κάποια δείγματα της θετικής κλάσης που έχουν γίνει labeled σωστά (TP):



Σχήμα 7.4: Αποτελέσματα LIME τεχνικής - Υπεργλυκαιμία, 30-60min

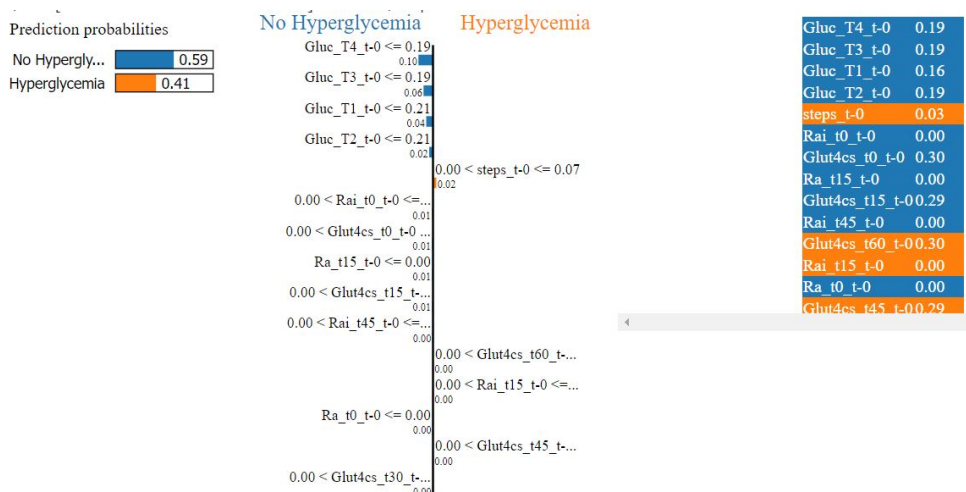
Παρατηρούμε ότι σε αυτή την περίπτωση, οι τιμές γλυκόζης και οι περισσότερες τιμές χαρακτηριστικών σχετιζόμενων με την ινσουλίνη συνεισφέρουν στο classification του δείγματος ως υπεργλυκαιμία, ενώ οι τιμές βημάτων και ο ρυθμός εμφάνισης γλυκόζης στο αίμα μειώνουν την πιθανότητα κατηγοριοποίησής του ως θετικό. Εξετάζοντας το

συγκεκριμένο δείγμα βλέπουμε ότι οι υψηλές τιμές γλυκόζης και πλήθους βημάτων, σε συνδυασμό με την έλλειψη πρόσφατου γεύματος, ή χορήγησης ινσουλίνης συμφωνούν λογικά με τις προαναφερθείσες συνεισφορές. Αντίστοιχα, στη συνέχεια φαίνεται η ερμηνεία ενός ακόμα TP instance, όπου όμως τώρα η πλειοψηφία των περισσότερων χαρακτηριστικών σχετιζόμενων με την ινσουλίνη συμβάλλουν στο χαρακτηρισμό ως μη υπεργλυκαιμία, λόγω πρόσφατης πρόσληψης ινσουλίνης:



Σχήμα 7.5: Αποτελέσματα LIME τεχνικής - Υπεργλυκαιμία, 30-60min

Συνεχίζοντας με την εξέταση μεμονωμένων δειγμάτων του συνόλου ελέγχου, παρατηρούμε ότι κάθε φορά η πλειοψηφία των (5) εισαγόμενων χαρακτηριστικών ανά περίπτωση  $R_a$ ,  $R_{ai}$ ,  $Glut4_{cs}$  συμβάλλουν κατά κύριο λόγο προς την ίδια κατεύθυνση, με τον λογικά αναμενόμενο τρόπο. Όσον αφορά το πλήθος των βημάτων, αυτό φαίνεται να μειώνει το κίνδυνο της υπεργλυκαιμίας όσο μεγαλύτερο είναι. Παρόλα αυτά, αξίζει να σημειωθεί ότι η πρόβλεψη για ορίζοντα 30-60 λεπτών φέρει αντικειμενικά μεγαλύτερη δυσκολία, από την περίπτωση πρόβλεψης για prediction horizon 0-30 min. Αυτό φαίνεται και από το παρακάτω false negative instance:



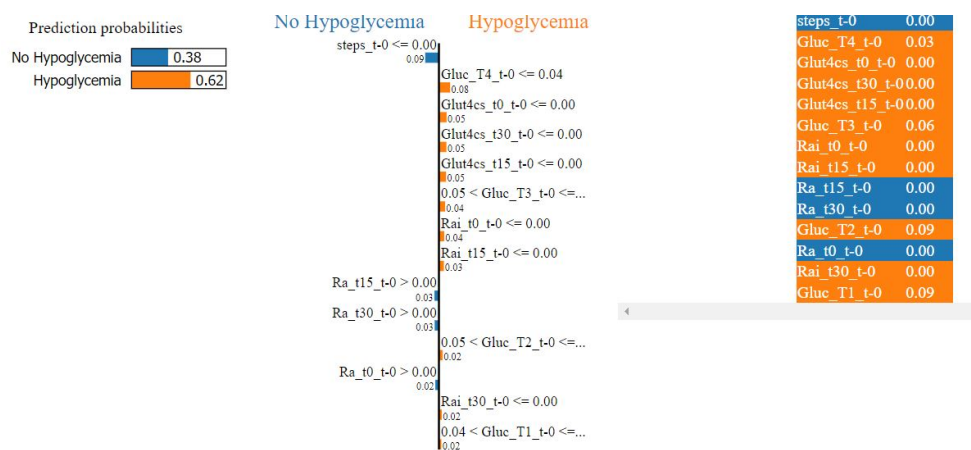
Σχήμα 7.6: Αποτελέσματα LIME τεχνικής - Υπεργλυκαιμία, 30-60min

Οι σχετικά χαμηλές μετρήσεις γλυκόζης, σε συνδυασμό με την καταγεγραμμένη πρόσληψη ινσουλίνης, και την έλλειψη καταγεγραμμένου γεύματος ωθεί σε classification ως

μη υπεργλυκαιμία, αν και ο αντίστοιχος ασθενής τελικά εμφάνισε υπεργλυκαιμία. Συνοψίζοντας, για τον ορίζοντα πρόβλεψης 30-60 λεπτών, οι περισσότερες συσχετίσεις των χαρακτηριστικών εισόδου με την επιθυμητή έξοδο αποτυπώνονται ορθώς από το δίκτυο, αν και πολλές φορές χαρακτηριστικά που αντιστοιχούν σε διάφορες χρονικές στιγμές δεν ερμηνεύονται πάντα σωστά, κάτι που πιθανώς θα μπορούσε να επιλυθεί μέσω περαιτέρω διερεύνησης των χρονικών στιγμιοτύπων των  $R_a$ ,  $R_{ai}$ ,  $Glut4_{cs}$  που είναι καλύτερο να συμπεριληφθούν ως είσοδος.

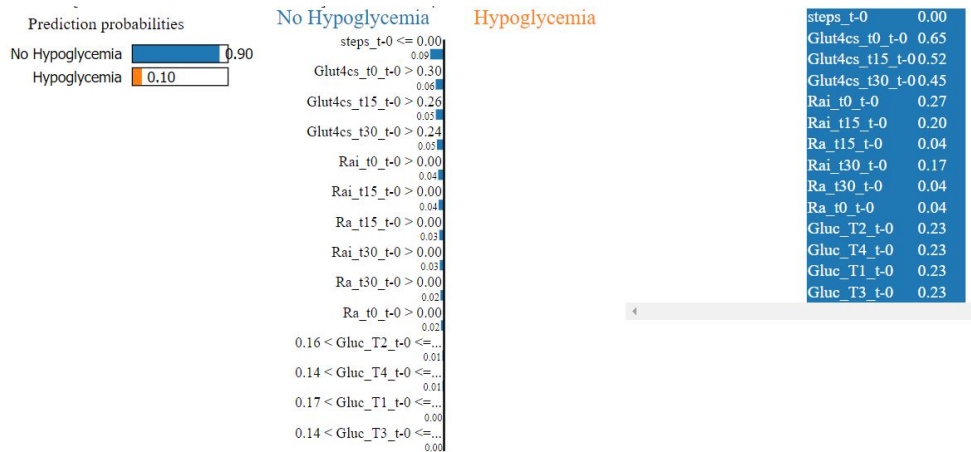
### 7.3.2 Πρόβλεψη υπογλυκαιμικών επεισοδίων

Από τα εξαχθέντα αποτελέσματα για την πρόβλεψη υπογλυκαιμικών επεισοδίων αναμένουμε το μοντέλο να μην έχει καταφέρει να ερμηνεύσει σωστά το ρόλο κάθε χαρακτηριστικού εισόδου. Πράγματι, με βάση τοπικές επεξηγήσεις αντιλαμβανόμαστε ότι το μοντέλο δεν έχει καταφέρει να συνδέσει ορθά τις τιμές των σχετιζόμενων με την ινσουλίνη χαρακτηριστικών, με την εμφάνιση υπογλυκαιμίας. Για παράδειγμα, στη συνέχεια φαίνεται ένα true positive δείγμα. Οι χαμηλές τιμές γλυκόζης συνεισφέρουν με ορθό τρόπο στον χαρακτηρισμό ως υπογλυκαιμία, όπως και οι μη μηδενικές τιμές  $R_a$  που ωθούν στην μείωση της πιθανότητας εμφάνισής της. Παρόλα αυτά, έχει θεωρηθεί ότι τα περισσότερα χαρακτηριστικά σχετιζόμενα με την ινσουλίνη συμβάλλουν στην εμφάνιση υπογλυκαιμίας, ενώ είναι μηδενικά.



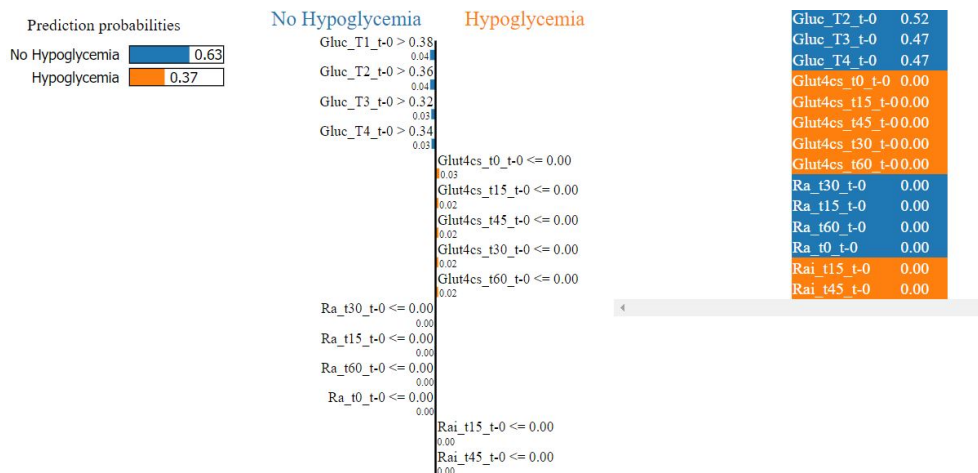
Σχήμα 7.7: Αποτελέσματα LIME τεχνικής - Υπογλυκαιμία, 0-30min

Η λανθασμένη αυτή σχέση αποτυπώνεται και στην παρακάτω περίπτωση true negative. Αν και οι υψηλές τιμές γλυκόζης, μη μηδενικές τιμές  $R_a$  και η χαμηλή τιμή βημάτων ορθά συμβάλλουν στη κατηγοριοποίηση του δείγματος ως μη υπογλυκαιμία, τα χαρακτηριστικά σχετιζόμενα με την ινσουλίνη έχουν ερμηνευθεί λανθασμένα, καθώς στην πραγματικότητα έχει σημειωθεί πρόσφατη χορήγηση ινσουλίνης.



Σχήμα 7.8: Αποτελέσματα LIME τεχνικής - Υπογλυκαιμία, 0-30min

Αντίστοιχα, και για τον ορίζοντα πρόβλεψης 30-60, στην παρακάτω περίπτωση τα χαρακτηριστικά σχετιζόμενα με την ινσουλίνη και την πρόσληψη γευμάτων δεν έχουν ερμηνευθεί σωστά :



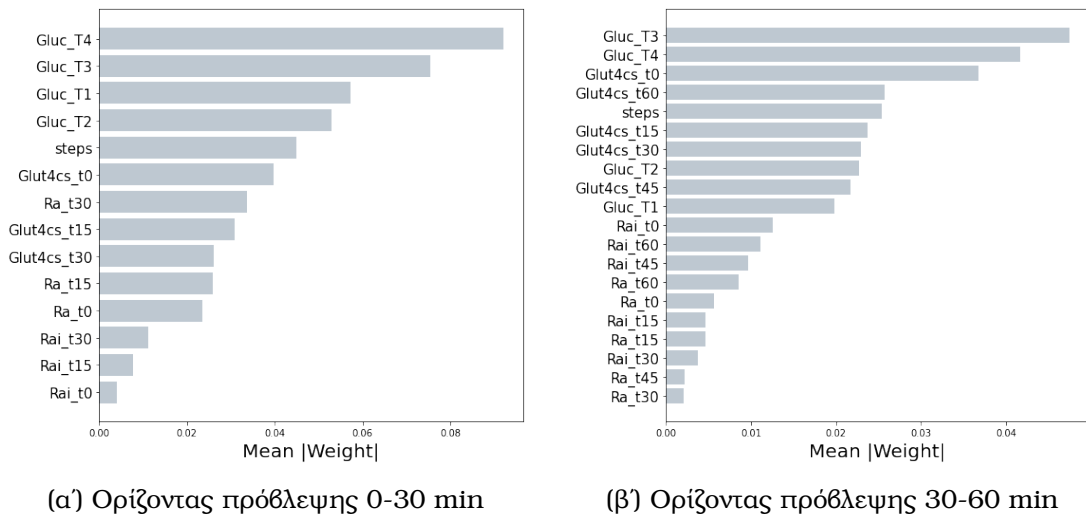
Σχήμα 7.9: Αποτελέσματα LIME τεχνικής - Υπογλυκαιμία, 30-60min

Συνοψίζοντας, για την περίπτωση της υπογλυκαιμίας, η επιθυμητή επίδραση των δεδομένων σχετιζόμενων με την ινσουλίνη και τη γλυκόζη δεν γίνεται πάντα κατανοητή, με τελικό αποτέλεσμα τη μείωση, αντί τη βελτίωση, της γενικότερης ακρίβειας του μοντέλου. Συμπεραίνουμε επομένως ότι είναι επιθυμητή η συμπερίληψη περισσότερων υπογλυκαιμικών δειγμάτων, με στόχο την καλύτερη κατανόηση των σχέσεων εισόδου και επιθυμητών εξόδων από το δίκτυο, ενώ επιθυμητή είναι ακόμα η περαιτέρω διερεύνηση πιθανών χαρακτηριστικών εισόδου που θα μπορούσαν να χρησιμοποιηθούν εναλλακτικά για την περίπτωση της υπογλυκαιμίας.

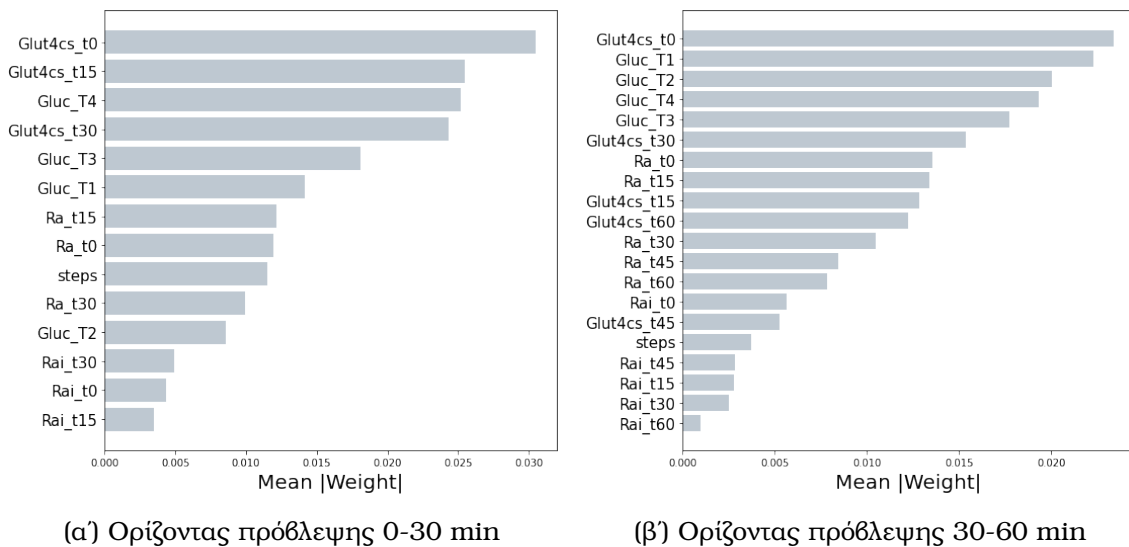
### 7.3.3 Υπολογισμός μέσων βαρών

Επιπλέον, προχωρήσαμε στον υπολογισμό του μέσου βάρους που προσδίδεται από το απλό, ερμηνεύσιμο ridge regression μοντέλο σε κάθε χαρακτηριστικό εισόδου. Ο υπολογισμός αυτός πραγματοποιήθηκε παράγοντας εξηγήσεις για όλα τα δείγματα του

συνόλου αξιολόγησης, κι έπειτα αθροίζοντας τα αντίστοιχα βάρη ανά χαρακτηριστικό εισόδου, με τελικό στόχο τον υπολογισμό των μέσων βαρών. Οι τιμές αυτές είναι ενδεικτικές της βαρύτητας που προσδίδεται στα αντίστοιχα χαρακτηριστικά κατά στις προβλέψεις, κι επομένως μπορούν να χρησιμοποιηθούν για μια προσεγγιστική εκτίμηση του feature importance. Στη συνέχεια φαίνονται τα αποτελέσματα για την περίπτωση της υπογλυκαιμίας και της υπεργλυκαιμίας, με τη μορφή ραβδογραμμάτων.



Σχήμα 7.10: Μέση τιμή βαρών - Υπεργλυκαιμία



Σχήμα 7.11: Μέση τιμή βαρών - Υπογλυκαιμία

Παρατηρούμε ότι οι μετρήσεις γλυκόζης βρίσκονται σταθερά στα χαρακτηριστικά με το υψηλότερο βάρος, για τις περισσότερες περιπτώσεις. Συνήθως, οι καταγραφές γλυκόζης που έχουν τη μεγαλύτερη αξία είναι οι μεταγενέστερες μετρήσεις, δηλαδή αυτές που έχουν πραγματοποιηθεί κατά το δεύτερο ήμισυ του lookback διαστήματος που εφαρμόζεται για τη γλυκόζη. Σε υψηλές θέσεις εντοπίζονται ακόμα στιγμιότυπα του χαρακτηριστικού  $GLUT4_{cs}$  (ποσοστό  $GLUT4$  μεταφορέα στο κυτταρικό τοίχωμα), το οποίο είναι ενδεικτικό των καταγεγραμμένων χορηγήσεων ινσουλίνης. Ειδικά κατά την



πρόβλεψη της υπογλυκαιμίας τέτοια στιγμιότυπα παρουσιάζουν το μέγιστο μέσο βάρος. Παρατηρούμε ότι στα χαρακτηριστικά αυτά προσδίδεται μεγαλύτερη βαρύτητα από ότι στα επίσης σχετιζόμενα με την ινσουλίνη  $R_{ai}$  (ρυθμός εμφάνισης ινσουλίνης στο αίμα), κάτι που πιθανόν να υποδηλώνει μεγαλύτερη αξία των χαρακτηριστικών αυτών και μπορεί να αποτελέσει αντικείμενο περαιτέρω ερευνητικής μελέτης. Όσον αφορά τα υπόλοιπα χαρακτηριστικά, αυτά λαμβάνουν χαμηλότερες-μέτριες θέσεις ανάλογα με την περίπτωση, ενώ το χαρακτηριστικό των βημάτων παρουσιάζει σχετικά υψηλό μέσο βάρος κατά την πρόβλεψη της υπεργλυκαιμίας. Τέλος, αξίζει να σημειωθεί ότι η παραπάνω εκτίμηση του feature importance είναι προσεγγιστική, και ότι άλλες τεχνικές ερμηνευσιμότητας όπως για παράδειγμα η SHAP [60], παρέχουν ακόμα πιο αξιόπιστες καθολικές (global) εκτιμήσεις.

## Κεφάλαιο 8

# Συμπεράσματα και μελλοντικές προεκτάσεις

Στην παρούσα εργασία πραγματοποιήθηκε η ανάπτυξη ενός ερμηνεύσιμου μοντέλου εκτίμησης του κινδύνου εμφάνισης υπογλυκαιμικών και υπεργλυκαιμικών επεισοδίων στο σακχαρώδη διαβήτη. Ως δεδομένα εκπαίδευσης και αξιολόγησης του δικτύου χρησιμοποιήθηκαν πραγματικά δεδομένα ασθενών, τα οποία περιλάμβαναν πληροφορίες σχετικά με μετρήσεις γλυκόζης του αίματος, διεξαχθείσα σωματική δραστηριότητα με τη μορφή βημάτων, χορηγήσεις ενέσιμης ινσουλίνης και προσλήψεις υδατανθράκων. Για την αποτύπωση των περίπλοκων διεργασιών απορρόφησης της ινσουλίνης και της προσλαμβανόμενης γλυκόζης από τον οργανισμό, χρησιμοποιήθηκαν μοντέλα προσομοίωσης φυσιολογικών συστημάτων που μετασχηματίζουν τα δεδομένα εισόδου. Συγκεκριμένα, με χρήση των διαθέσιμων πληροφοριών γευμάτων και ινσουλίνης υπολογίστηκε ο ρυθμός εμφάνισης γλυκόζης και ινσουλίνης στο αίμα, καθώς και το ποσοστό GLUT4 μεταφορέα στο κυτταρικό τοίχωμα, για ορισμένες χρονικές στιγμές ενδιαφέροντος. Για την πραγματοποίηση των προβλέψεων εφαρμόστηκε ένα αναδρομικό νευρωνικό δίκτυο, αποτελούμενο από δύο στοιβαγμένα LSTM επίπεδα, ακολουθούμενα από πλήρως συνδεδεμένα επίπεδα. Η βελτιστοποίηση των υπερπαραμέτρων του μοντέλου πραγματοποιήθηκε παράλληλα με την αξιολόγησή του, με χρήση εμφωλευμένης διασταυρούμενης επικύρωσης. Στην τεχνική αυτή ενσωματώθηκαν μέθοδοι υπερδειγματοληψίας (SMOTE, BSMOTE, Random, ADASYN) για την αντιμετώπιση της ανισορροπίας των κλάσεων, καθώς και η Temperature Scaling μέθοδος βαθμονόμησης, με στόχο την εξαγωγή πιο αξιόπιστων πιθανοτήτων πρόβλεψης. Ακόμα, πραγματοποιήθηκαν κάποιες δοκιμές με προσαρμοσμένο κατώφλι ταξινόμησης. Τέλος, το δίκτυο κατέστη τοπικά ερμηνεύσιμο με εφαρμογή της LIME τεχνικής.

Τα αποτελέσματα που προέκυψαν από την αξιολόγηση του μοντέλου κατά την πρόβλεψη των υπεργλυκαιμιών ήταν σε γενικές γραμμές ικανοποιητικά για όλες τις χρησιμοποιούμενες μετρικές αξιολόγησης. Η σιγουριά πρόβλεψης του δικτύου κρίθηκε σχετικά αξιόπιστη, με βάση τα αντίστοιχα επίπεδα βαθμονόμησης του. Τέλος, παρατηρήθηκε μείωση της ακρίβειας του μοντέλου κατά την πραγματοποίηση εκτιμήσεων για τον πιο απομακρυσμένο χρονικά ορίζοντα πρόβλεψης. Στην περίπτωση της πρόβλεψης των υπογλυκαιμιών, το μοντέλο δεν πραγματοποιήσει εξίσου ακριβείς προβλέψεις, όπως φάνηκε από τις πιο χαμηλές τιμές των πιο αντιπροσωπευτικών μετρικών. Η χαμηλή αυτή απόδοση απεδόθη σε μεγάλο βαθμό στην ακραία ανισορροπία των κλάσεων, καθώς και στην διαθεσιμότητα αριθμητικά λίγων θετικών δειγμάτων. Οι πιθανότητες

πρόβλεψης του μοντέλου σε αυτήν την περίπτωση κρίθηκαν υπεραισιόδοξες, κάτι που φανέρωνε μικρό ποσοστό βαθμονόμησής του, όπως προκύπτει και από τις αντίστοιχες μετρικές αξιολόγησης του ποσοστού βαθμονόμησης. Αν και η εφαρμογή της τεχνικής *threshold tuning* φαίνεται να έχει νόημα σε αυτήν την περίπτωση, η εύρεση του προσαρμοσμένου κατωφλίου ταξινόμησης με χρήση ενός μόνο συνόλου επικύρωσης δεν φάνηκε να επαρκεί. Τέλος, η χρήση της LIME μεθόδου προσέφερε σημαντικές πληροφορίες για τον τρόπο λειτουργίας του μοντέλου για κάθε περίπτωση συνόλου δεδομένων. Στη συνέχεια παραθέτουμε ορισμένες ενδιαφέρουσες προεκτάσεις της παρούσας εργασίας, οι οποίες θα μπορούσαν να μελετηθούν μελλοντικά:

Όσον αφορά τα εφαρμοζόμενα δεδομένα εισόδου, επιθυμητή θα ήταν η αξιολόγηση των παραπάνω μεθόδων με χρήση *real time continuous glucose monitoring* μετρήσεων γλυκόζης, οι οποίες παρέχουν μια πιο ακριβή και λεπτομερή περιγραφή του γλυκαιμικού προφίλ των εξεταζόμενων ασθενών, ενώ συνεισφέρουν στο σχηματισμό περισσότερων χρονοσειρών γλυκόζης. Διαθέτοντας *rtCGM* καταγραφές αναμένεται ακόμα να αυξηθεί το διάθεσιμο πλήθος υπογλυκαιμικών δειγμάτων, κάτι που είναι κρίσιμο για τη βελτίωση της ποιότητας πρόβλεψης υπογλυκαιμικών επεισοδίων. Όσον αφορά την παραμετροποίηση των εφαρμοζόμενων μοντέλων μαθηματικής προσομοίωσης, επιλέχθηκε η χρήση ιστορικού ινσουλίνης και γευμάτων 12 ωρών, παρόλα αυτά αυτό αποτελεί κάτι που μπορεί να μεταβληθεί. Η επιλογή των χρονικών στιγμιотύπων των  $R_a$ ,  $R_{ai}$ ,  $GLUT4_{cs}$  μεγεθών που χρησιμοποιούνται επίσης μπορεί να αποτελέσει αντικείμενο πειραματισμού. Καθώς το μοντέλο που αναπτύσσεται αποτελεί ένα γενικευμένο, *global* μοντέλο, ενδιαφέρον θα είχε επίσης η συμπερίληψη πληροφοριών όπως το βάρος, η ηλικία και το φύλο των ασθενών ως χαρακτηριστικά εισόδου, με στόχο την ενίσχυση του βαθμού εξατομικεύσης. Φυσικά, δεδομένης της επάρκειας των αντίστοιχων δεδομένων η ανάπτυξη πλήρως εξατομικευμένων μοντέλων πρόβλεψης θα διέθετε επίσης αξία. Τέλος, στην περίπτωση μας, για την πρόβλεψη χρησιμοποιήθηκαν όλα τα δεδομένα εισόδου. Ενδιαφέρον θα είχε η διεξαγωγή πειραμάτων με υποσύνολα του *feature space* με στόχο τη διάκριση των πιο χρήσιμων χαρακτηριστικών.

Όσον αφορά την ανάπτυξη του μοντέλου μηχανικής μάθησης, μεγάλη αξία θα είχε ο πιο ενδεδειγμένος πειραματισμός και με άλλες δομές δικτύων, καθώς στην περίπτωση μας παρουσιάζονται τα αποτελέσματα αξιολόγησης μιας δομής. Με βάση και την διεξαχθείσα βιβλιογραφική επισκόπηση (κεφάλαιο 4), οι τεχνικές συλλογικής μάθησης φαίνεται να είναι πολλά υποσχόμενες, ειδικά όταν αυτές κάνουν χρήση αναδρομικών νευρωνικών δικτύων. Σχετικά με τη βαθμονόμηση του μοντέλου, εφαρμόστηκε η *Temperature Scaling* μέθοδος, παρόλα αυτά και άλλες μέθοδοι όπως η *Platt Scaling* και η *Bayesian Binning into Quantiles (BBQ)* [86] είναι διαθέσιμες για τον ίδιο σκοπό. Ακόμα, όπως παρατηρήθηκε για την περίπτωση πρόβλεψης της υπογλυκαιμίας, η τεχνική *threshold tuning* θα μπορούσε να φέρει υποσχόμενα αποτελέσματα, οπότε η δοκιμή πιο εξεζητημένων τεχνικών προσαρμογής του κατωφλίου έχει νόημα. Τέλος, στην περίπτωση μας εφαρμόστηκε η LIME τεχνική για την επεξήγηση των αποτελεσμάτων του μοντέλου, όμως είναι δυνατή η χρήση και άλλων μεθόδων ερμηνευσιμότητας. Ενδεικτικά αναφέρουμε τη SHAP (*SHapley Additive exPlanations*) μέθοδο, η οποία προσφέρει τη δυνατότητα εξαγωγής πιο αξιόπιστων συμπερασμάτων σχετικά με τη συνολική (*global*) ερμηνεία του εξεταζόμενου δικτύου, με βάση τις παραγόμενες, τοπικές εξηγήσεις. [56]

# Βιβλιογραφία

- [1] Diabetes. (2021). World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/diabetes>
- [2] American Diabetes Association. (2014). Diagnosis and classification of diabetes mellitus. *Diabetes care*, 37 Suppl 1, S81–S90. <https://doi.org/10.2337/dc14-S081>
- [3] Silverthort, D. U. (2018). Φυσιολογία του Ανθρώπου. (8η Έκδοση). Broken Hill Publishers Ltd
- [4] Frayn, K. N. (2010). Μεταβολική Ρύθμιση: Μια προοπτική για τον άνθρωπο. (3η έκδοση). Επιστημονικές Εκδόσεις Παρισιανού Α.Ε.
- [5] Rahman, M. S., Hossain, K. S., Das, S., Kundu, S., Adegoke, E. O., Rahman, M. A., Hannan, M. A., Uddin, M. J., Pang, M. G. (2021). Role of Insulin in Health and Disease: An Update. *International journal of molecular sciences*, 22(12), 6403. <https://doi.org/10.3390/ijms22126403>
- [6] Aronoff, S., Berkowitz, KB, Shreiner, B, Want, L. (2004). Glucose Metabolism and Regulation: Beyond Insulin and Glucagon. *Diabetes Spectrum*. 17. <https://doi.org/10.2337/diaspect.17.3.183>
- [7] Cowie, C. C., Casagrande, S. S., Menke, A., Cissell, M. A., Eberhardt, M. S., Meigs, J. B., Gregg, E. W., Knowler, W. C., Barrett-Connor, E., Becker, D. J., Brancati, F. L., Boyko, E. J., Herman, W. H., Howard, B. V., Narayan, K., Rewers, M., Fradkin, J. E. (Eds.). (2018). *Diabetes in America*. (3rd ed.). National Institute of Diabetes and Digestive and Kidney Diseases (US).
- [8] Νικολοπούλου, Α. (2011). Διαβητική Κετοξέωση. *Ελληνικά Διαβητολογικά Χρονικά* 24, 4: 220-234
- [9] Nathan, D. M. (1993). Long-Term Complications of Diabetes Mellitus. *N Engl J Med* 328:1676-1685 doi: 10.1056/NEJM199306103282306
- [10] Diabetes Tests Diagnosis. (2022). National Institute of Diabetes and Digestive and Kidney Diseases. <https://www.niddk.nih.gov/health-information/diabetes/>
- [11] Managing Diabetes. (2016). National Institute of Diabetes and Digestive and Kidney Diseases. <https://www.niddk.nih.gov/health-information/diabetes/>
- [12] Monitoring your Blood Sugar. (2021). Centers for Disease Control and Prevention. <https://www.cdc.gov/diabetes/>

- [13] Rodbard D. (2017). Continuous Glucose Monitoring: A Review of Recent Studies Demonstrating Improved Glycemic Outcomes. *Diabetes technology therapeutics*, 19(S3), S25–S37. <https://doi.org/10.1089/dia.2017.0035>
- [14] Cappon, G., Vettoretti, G. M., Sparacino, G. M. G., Facchinetti, G. M. G. A. (2019). Continuous Glucose Monitoring Sensors for Diabetes Management: A Review of Technologies and Applications. *Diabetes Metabolism Journal*, 43(4), 383–397.
- [15] Soni, A., Wright, N., Agwu, J. C., Timmis, A., Drew, J., Kershaw, M., Moudiotis, C., Regan, F., Williams, E. C., Wan, J., Ng, S. M., ACDC Guideline Development Group (2022). A practical approach to continuous glucose monitoring (rtCGM) and FreeStyle Libre systems (isCGM) in children and young people with Type 1 diabetes. *Diabetes research and clinical practice*, 184, 109196. <https://doi.org/10.1016/j.diabres.2022.109196>
- [16] Ενημέρωση για την ινσουλίνη. (2022). Ελληνική Διαβητολογική Εταιρεία. <https://www.ede.gr/>
- [17] Berget, C., Messer, L. H., Forlenza, G. P. (2019). A Clinical Overview of Insulin Pump Therapy for the Management of Diabetes: Past, Present, and Future of Intensive Therapy. *Diabetes spectrum : a publication of the American Diabetes Association*, 32(3), 194–204. <https://doi.org/10.2337/ds18-0091>
- [18] Βαζαίου, Α. (2013). Οδηγός διατροφής για τη ρύθμιση του διαβήτη. Ελληνική Διαβητολογική Εταιρεία
- [19] Κατευθυντήριες Οδηγίες για τον Σακχαρώδη Διαβήτη. (2022). Ελληνική Διαβητολογική Εταιρεία. <https://www.ede.gr/>
- [20] Haykin, S. S. (2009). *Neural networks and learning machines*. (3rd ed.). Upper Saddle River, NJ: Pearson Education.
- [21] Sarker, I.H. (2021). *Machine Learning: Algorithms, Real-World Applications and Research Directions*. *SN COMPUT. SCI.* 2, 160. <https://doi.org/10.1007/s42979-021-00592-x>
- [22] Διαμαντάρας Κ. (2007). Τεχνητά Νευρωνικά Δίκτυα. Εκδόσεις Κλειδάριθμος
- [23] Sharma, Sid., Sharma, Sim., Athaiya, A. (2020). Activation Functions in Neural Networks. *International Journal of Engineering Applied Sciences and Technology*. 04. 310-316. 10.33564/IJEAST.2020.v04i12.054.
- [24] Hornik, K., Stinchcombe, M.B., White, H.L. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359-366.
- [25] Rumelhart, D. E., Hinton, G. E. Williams, R. J. (1986). Learning Representations by Back-propagating Errors. *Nature*, 323, 533–536. doi: 10.1038/323533a0
- [26] Chen, Y., Li, J. (2021). Recurrent Neural Networks algorithms and applications. 2nd International Conference on Big Data, Artificial Intelligence and Software Engineering (ICBASE). pp. 38-43. doi: 10.1109/ICBASE53849.2021.00015.

- [27] Britz, D. (2015). Recurrent Neural Networks Tutorial - Introduction to RNNs. <https://dennybritz.com/posts/wildml/recurrent-neural-networks-tutorial-part-1/>
- [28] Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. MIT Press.
- [29] Pascanu, R., Mikolov, T., Bengio, Y. (2013). On the difficulty of training recurrent neural networks. International Conference on Machine Learning.
- [30] Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen Netzen. Diplomathesis, T.U. München. 17, 396, 398
- [31] Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157–166.
- [32] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [33] Olah, C. (2015). Understanding LSTM Networks. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [34] Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J. (2015). LSTM: A Search Space Odyssey. IEEE Transactions on Neural Networks and Learning Systems, 28, 2222–2232.
- [35] DiPietro, R., Hager, G. D. (2020). Deep learning: RNNs and LSTM. Handbook of Medical Image Computing and Computer Assisted Intervention, 503–519. doi:10.1016/b978-0-12-816176-0.00026-0
- [36] Cho, K., van Merriënboer, B., Gulcehre, C., Bougares, F., Schwenk, H., Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)
- [37] Józefowicz, R., Zaremba, W. Sutskever, I. (2015). An Empirical Exploration of Recurrent Network Architectures.. In F. R. Bach D. M. Blei (eds.), ICML (p./pp. 2342-2350), : JMLR.org.
- [38] Chung, J., Gulcehre, C., Cho, K., Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In NIPS 2014 Workshop on Deep Learning, December 2014.
- [39] Sun, S., Cao, Z., Zhu, H., Zhao, J. (2020). A Survey of Optimization Methods From a Machine Learning Perspective. IEEE Transactions on Cybernetics, 50(8), 3668–3681. <https://doi.org/10.1109/tcyb.2019.2950779>
- [40] Mohri, M., Rostamizadeh, A., Talwalkar, A. (2018). Foundations of Machine Learning. Cambridge, MA: MIT Press. ISBN: 978-0-262-03940-6
- [41] Becker, S., Plumbley, M. (1996). Unsupervised neural network learning procedures for feature extraction and classification. Applied Intelligence, 6(3), 185–203. doi:10.1007/bf00126625
- [42] Kaelbling, L. P., Littman, M. L. Moore, A. W. (1996). Reinforcement learning: A survey. Journal of Artificial Intelligence Research, 4, 237–285.

- [43] Yadav, S., Shukla, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. 2016 IEEE 6th International Conference on Advanced Computing (IACC). doi:10.1109/iacc.2016.25
- [44] Patro, R. (2021). Cross-Validation: K Fold vs Monte Carlo. Towards Data Science. <https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b>
- [45] Lyashenko, V., Jha, A. (2022). Cross-Validation in Machine Learning: How to Do It Right. MLOps Blog. <https://neptune.ai/blog/cross-validation-in-machine-learning-how-to-do-it-right>
- [46] Singh, D., Singh, B. (2019). Investigating the impact of data normalization on classification performance. Applied Soft Computing, 105524. doi:10.1016/j.asoc.2019.105524
- [47] Chawla, N., Bowyer, K., Hall, L.O., Kegelmeyer, W.P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. ArXiv, abs/1106.1813.
- [48] Han, H., Wang, W.-Y., Mao, B.-H. (2005). Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. Advances in Intelligent Computing, 878-887. doi:10.1007/1153805991
- [49] Haibo He, Yang Bai, Garcia, E. A., Shutao Li. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). doi:10.1109/ijcnn.2008.4633969
- [50] Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications, 73, 220-239. doi:10.1016/j.eswa.2016.12.035
- [51] Guo, X., Yin, Y., Dong, C., Yang, G., Zhou, G. (2008). On the Class Imbalance Problem. 2008 Fourth International Conference on Natural Computation, 4, 192-201.
- [52] M, H., M.N, S. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. International Journal of Data Mining Knowledge Management Process, 5(2), 01-11. doi:10.5121/ijdkp.2015.5201
- [53] Grandini, M., Bagli, E., Visani, G. (2020). Metrics for Multi-Class Classification: an Overview. ArXiv, abs/2008.05756.
- [54] Gohel, Prashant, Singh, Priyanka, Mohanty, Manoranjan. (2021). Explainable AI: current status and future directions.
- [55] Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V. I., Precise4Q consortium (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC medical informatics and decision making, 20(1), 310. <https://doi.org/10.1186/s12911-020-01332-6>
- [56] Schwalbe, G., Finzel, B. (2021). A Comprehensive Taxonomy for Explainable Artificial Intelligence: A Systematic Survey of Surveys on Methods and Concepts.

- [57] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. ICCV (p./pp. 618-626), IEEE Computer Society. ISBN: 978-1-5386-1032-9
- [58] Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd ed.). christophm.github.io/interpretable-ml-book/
- [59] Ribeiro, M., Singh, S., Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- [60] Lundberg, S. M. Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan R. Garnett (ed.), *Advances in Neural Information Processing Systems 30* (pp. 4765–4774) . Curran Associates, Inc.
- [61] Dalakleidi, K., Zarkogianni, K., Thanopoulou, A., Nikita, K. (2017). Comparative assessment of statistical and machine learning techniques towards estimating the risk of developing type 2 diabetes and cardiovascular complications. *Expert Systems*, 34(6), e12214. doi:10.1111/exsy.12214
- [62] Fico, G., Hernanvez, L., Cancela, J., Dagliati, A., Sacchi, L., Martinez-Millana, A., Posada, J., Manero, L., Verdú, J., Facchinetti, A., Ottaviano, M., Zarkogianni, K., Nikita, K., Groop, L., Gabriel-Sanchez, R., Chiovato, L., Traver, V., Merino-Torres, J. F., Cobelli, C., Bellazzi, R., . . . Arredondo, M. T. (2019). What do healthcare professionals need to turn risk models for type 2 diabetes into usable computerized clinical decision support systems? Lessons learned from the MOSAIC project. *BMC medical informatics and decision making*, 19(1), 163. <https://doi.org/10.1186/s12911-019-0887-8>
- [63] Zarkogianni, K., Nikita, K. S. (2015). Personal Health Systems for Diabetes Management, Early Diagnosis and Prevention. In *Handbook of Research on Trends in the Diagnosis and Treatment of Chronic Conditions*. Hershey PA, USA: Medical Information Science Reference.
- [64] Dalakleidi, K.V., Zarkogianni, K., Karamanos, V.G., Thanopoulou, A.C., Nikita, K.S. (2013). A hybrid genetic algorithm for the selection of the critical features for risk prediction of cardiovascular complications in Type 2 Diabetes patients. 13th IEEE International Conference on BioInformatics and BioEngineering, 1-4.
- [65] Skevofilakas, M., Zarkogianni, K., Karamanos, B. G., Nikita, K. S. (2010). A hybrid Decision Support System for the risk assessment of retinopathy development as a long term complication of Type 1 Diabetes Mellitus. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2010, 6713–6716. <https://doi.org/10.1109/IEMBS.2010.5626245>
- [66] Athanasiou, M., Sfrintzeri, K., Zarkogianni, K., Thanopoulou, A.C., Nikita, K.S. (2020). An explainable XGBoost-based approach towards assessing the risk of cardiovascular disease in patients with Type 2 Diabetes Mellitus. 2020 IEEE



20th International Conference on Bioinformatics and Bioengineering (BIBE), 859-864.

- [67] Athanasiou, M., Zarkogianni, K., Monou, S., Kosteria, I., Vasilakis, I., Kanaka-Gantenbein, C., Nikita, K. (2021). An Interpretable LSTM-Based Prediction Model For Assessing The Risk of Hospitalization And Re-Hospitalization in Youth With Type 1 Diabetes Mellitus. *Diabetes Technology Therapeutics*, 23, A27-A28. Mary Ann Liebert, INC 140 Huguenot Street, 3rd FL, New Rochelle, NY 10801 USA.
- [68] Zarkogianni, K., Athanasiou, M., Thanopoulou, A. C. (2018). Comparison of Machine Learning Approaches Toward Assessing the Risk of Developing Cardiovascular Disease as a Long-Term Diabetes Complication. *IEEE journal of biomedical and health informatics*, 22(5), 1637-1647. <https://doi.org/10.1109/JBHI.2017.2765639>
- [69] Mougiakakou, S. G., Nikita, K. S., Protonotarios, E. D., Matsopoulos, G. K. (1998). Neural Network based Decision Support System for Outpatients Diabetes Management. Symposium on COMPUTERS in DIABETES '98 on the Occasion of the EASD '98. Barcelona, Spain.
- [70] Mougiakakou, S. G., Nikita, K. S. (2000). A neural network approach for insulin regime and dose adjustment in type 1 diabetes. *Diabetes technology therapeutics*, 2(3), 381-389. <https://doi.org/10.1089/15209150050194251>
- [71] Zarkogianni, K., Litsa, E., Mitsis, K., Wu, P. Y., Kaddi, C. D., Cheng, C. W., Wang, M. D., Nikita, K. S. (2015). A Review of Emerging Technologies for the Management of Diabetes Mellitus. *IEEE transactions on bio-medical engineering*, 62(12), 2735-2749. <https://doi.org/10.1109/TBME.2015.2470521>
- [72] Iliopoulou, D., Giokas, K., Mougiakakou, S. G., Stoitsis, J. S., Prentza, A., Nikita, K. S. (2005 2005). A Telematic System for Diabetes Management, Reporting and Patient Advice. *The Journal on Information Technology in Healthcare*, 3, 307-313.
- [73] Mougiakakou, S. G., Bartsocas, C. S., Bozas, E., Chaniotakis, N., Iliopoulou, D., Kouris, I., Pavlopoulos, S., Prountzou, A., Skevofilakas, M., Tsoukalis, A., Varotsis, K., Vazeou, A., Zarkogianni, K., Nikita, K. S. (2010). SMART-DIAB: a communication and information technology approach for the intelligent monitoring, management and follow-up of type 1 diabetes patients. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, 14(3), 622-633. <https://doi.org/10.1109/TITB.2009.2039711>
- [74] Dalla Man, C., Rizza, R. A., Cobelli, C. (2007). Meal simulation model of the glucose-insulin system. *IEEE transactions on bio-medical engineering*, 54(10), 1740-1749. <https://doi.org/10.1109/TBME.2007.893506>
- [75] Hovorka, R., Canonico, V., Chassin, L. J., Haueter, U., Massi-Benedetti, M., Orsini Federici, M., Pieber, T. R., Schaller, H. C., Schaupp, L., Vering, T., Wilinska, M. E. (2004). Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiological measurement*, 25(4), 905-920. <https://doi.org/10.1088/0967-3334/25/4/010>

- [76] Bergman, R. N., Ider, Y. Z., Bowden, C. R., Cobelli, C. (1979). Quantitative estimation of insulin sensitivity. *The American journal of physiology*, 236(6), E667–E677. <https://doi.org/10.1152/ajpendo.1979.236.6.E667>
- [77] Oviedo, S., Vehí, J., Calm, R., Armengol, J. (2017). A review of personalized blood glucose prediction strategies for T1DM patients. *International journal for numerical methods in biomedical engineering*, 33(6), 10.1002/cnm.2833. <https://doi.org/10.1002/cnm.2833>
- [78] Sedaghat, A. R., Sherman, A., Quon, M. J. (2002). A mathematical model of metabolic insulin signaling pathways. *American journal of physiology. Endocrinology and metabolism*, 283(5), E1084–E1101. <https://doi.org/10.1152/ajpendo.00571.2001>
- [79] Ho, C. K., Rahib, L., Liao, J. C., Sriram, G., Dipple, K. M. (2015). Mathematical modeling of the insulin signal transduction pathway for prediction of insulin sensitivity from expression data. *Molecular genetics and metabolism*, 114(1), 66–72. <https://doi.org/10.1016/j.ymgme.2014.11.003>
- [80] Nyman, E., Brännmark, C., Palmér, R., Brugård, J., Nyström, F. H., Strålfors, P., Cedersund, G. (2011). A hierarchical whole-body modeling approach elucidates the link between in Vitro insulin signaling and in Vivo glucose homeostasis. *The Journal of biological chemistry*, 286(29), 26028–26041. <https://doi.org/10.1074/jbc.M110.188987>
- [81] Zarkogianni, K., Vazeou, A., Mougiakakou, S.G., Prountzou, A., Nikita, K.S. (2011). An Insulin Infusion Advisory System Based on Autotuning Nonlinear Model-Predictive Control. *IEEE Transactions on Biomedical Engineering*, 58, 2467–2477.
- [82] Zarkogianni, K., Litsa, E.E., Vazeou, A., Nikita, K.S. (2013). Personalized glucose-insulin metabolism model based on self-organizing maps for patients with Type 1 Diabetes Mellitus. *13th IEEE International Conference on BioInformatics and BioEngineering*, 1–4.
- [83] Kim, D. Y., Choi, D. S., Kim, J., Chun, S. W., Gil, H. W., Cho, N. J., Kang, A. R., Woo, J. (2020). Developing an Individual Glucose Prediction Model Using Recurrent Neural Network. *Sensors (Basel, Switzerland)*, 20(22), 6460. <https://doi.org/10.3390/s20226460>
- [84] Sun, Q., Jankovic, M. Bally, L. Mougiakakou, S. (2018). Predicting Blood Glucose with an LSTM and Bi-LSTM Based Deep Neural Network. 1–5. 10.1109/NEUREL.2018.8586990.
- [85] Tena, F., Garnica, O., Lanchares, J. Hidalgo, I. (2021). A Critical Review of the state-of-the-art on Deep Neural Networks for Blood Glucose Prediction in Patients with Diabetes.
- [86] Mirshekarian, S., Bunescu, R., Marling, C., Schwartz, F. (2017). Using LSTMs to learn physiological models of blood glucose behavior. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, 2017, 2887–2891. <https://doi.org/10.1109/EMBC.2017.8037460>

- [87] Meijner, C., Persson, S. (2017) Blood Glucose Prediction for Type 1 Diabetes using Machine Learning Long Short-term Memory based models for blood glucose prediction. [Master's Thesis, University of Gothenburg] <https://publications.lib.chalmers.se/records/fulltext/251317/251317.pdf>
- [88] Gülesir, M.V., Einreichung, T.D., Kauschke, S., Gülesir, G. (2018). Predicting Blood Glucose Levels of Diabetes Patients.
- [89] Idrissi, T.E., Idri, A., Abnane, I., Bakkoury, Z. (2019). Predicting Blood Glucose using an LSTM Neural Network. 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), 35-41.
- [90] Aiello, E.M., Lisanti, G., Magni, L., Musci, M., Toffanin, C. (2020). Therapy-driven Deep Glucose Forecasting. *Eng. Appl. Artif. Intell.*, 87.
- [91] Zhu, T., Li, K., Chen, J., Herrero, P., Georgiou, P. (2020). Dilated Recurrent Neural Networks for Glucose Forecasting in Type 1 Diabetes. *Journal of Healthcare Informatics Research* (Vol. 4, Issue 3, pp. 308-324). <https://doi.org/10.1007/s41666-020-00068-2>
- [92] Mayo, M., Koutny, T. (2020). Neural Multi-class Classification Approach to Blood Glucose Level Forecasting with Prediction Uncertainty Visualisation. *KDH@ECAI*.
- [93] Munoz-Organero M. (2020). Deep Physiological Model for Blood Glucose Prediction in T1DM Patients. *Sensors* (Basel, Switzerland), 20(14), 3896. <https://doi.org/10.3390/s20143896>
- [94] Khadem, H., Nemat, H. Elliott, J. Benaissa, M. (2020). Multi-lag Stacking for Blood Glucose Level Prediction.
- [95] Martinsson, J., Schliep, A., Eliasson, B., Mogren, O. (2020). Blood glucose prediction with variance estimation using recurrent neural networks. *Journal of Healthcare Informatics Research*, 4(1), 1-18. doi:10.1007/s41666-019-00059-y
- [96] Seo, W., Lee, Y.-B., Lee, S., Jin, S.-M., Park, S.-M. (2019). A machine-learning approach to predict postprandial hypoglycemia. *BMC Medical Informatics and Decision Making*, 19(1). doi:10.1186/s12911-019-0943-4
- [97] Witte, H., Nakas, C., Bally, L., Leichtle, A. B. (2022). Machine Learning Prediction of Hypoglycemia and Hyperglycemia From Electronic Health Records: Algorithm Development and Validation. *JMIR formative research*, 6(7), e36176. <https://doi.org/10.2196/36176>
- [98] Duckworth, C., Guy, M. J., Kumaran, A., O'Kane, A. A., Ayobi, A., Chapman, A., Marshall, P., Boniface, M. (2022). Explainable Machine Learning for Real-Time Hypoglycemia and Hyperglycemia Prediction and Personalized Control Recommendations. *Journal of Diabetes Science and Technology*. SAGE Publications. <https://doi.org/10.1177/19322968221103561>
- [99] Cappon, G., Meneghetti, L., Prendin, F., Pavan, J., Sparacino, G., Del Favero, S., Facchinetti, A. (2020). A Personalized and Interpretable Deep Learning Based Approach to Predict Blood Glucose Concentration in Type 1 Diabetes. *CEUR workshop proceedings*, 2675, 75-79.

- [100] Zarkogianni, K., Athanasiou, M., Mitsis, K., Vasilakis, I., Kosteria, I., Perakis, K., Vergeti, D., Antonopoulou, E., Papachristou, V., Chioti, A., Vouterakis, E., Kalafatis, P., Pervanidou, C., Kanaka-Gantenbein Nikita, K. (2021). The Endorse Pilot Trial. *Diabetes Technology Therapeutics*, A68-A68. Mary Ann Liebert, INC 140 Huguenot Street, 3rd FL, New Rochelle, NY 10801 USA.
- [101] Vasilakis, I., Kosteria, I., Mitsis, K., Zarkogianni, K., Kalafatis, L., Athanasiou, M., Perakis, K., Papavasiliou, G., Taliou, A., Nicolaidis, N., Chioti, V., Tokou, I., Vergeti, D., Antonopoulou, D., Papachristou, E., Meklis, V., Nikita, K., Kanaka-Gantenbein, C. (2022). The ENDORSE feasibility pilot trial: Assessing the implementation of serious games strategy and artificial intelligence-based telemedicine in glycemic control improvement. In *Diabetes Technology Therapeutics*, 24(SUPPL 1): A218-A219.
- [102] Νικήτα, Κ. (2011). Προσομοίωση Φυσιολογικών Συστημάτων. Εκδόσεις Τζιόλα
- [103] Man, C. D., Camilleri, M., Cobelli, C. (2006). A System Model of Oral Glucose Absorption: Validation on Gold Standard Data. *IEEE Transactions on Biomedical Engineering*, 53(12), 2472-2478. doi:10.1109/tbme.2006.883792
- [104] Man, C. D., Raimondo, D. M., Rizza, R. A., Cobelli, C. (2007). GIM, simulation software of meal glucose-insulin model. *Journal of diabetes science and technology*, 1(3), 323-330. <https://doi.org/10.1177/193229680700100303>
- [105] Man, C. D., Micheletto, F., Lv, D., Breton, M., Kovatchev, B., Cobelli, C. (2014). The UVA/PADOVA Type 1 Diabetes Simulator: New Features. *Journal of diabetes science and technology*, 8(1), 26-34. <https://doi.org/10.1177/1932296813514502>
- [106] Sjöstrand, M., Holmång, A., Lönnroth, P. (1999). Measurement of interstitial insulin in human muscle. *The American journal of physiology*, 276(1), E151-E154. <https://doi.org/10.1152/ajpendo.1999.276.1.E151>
- [107] Li, Y.-H., Harfiya, L. N., Purwandari, K., Lin, Y.-D. (2020). Real-Time Cuffless Continuous Blood Pressure Estimation Using Deep Learning Model. *Sensors*, 20(19), 5606. doi:10.3390/s20195606
- [108] Ghoshal, B., Tucker, A. (2022). On Calibrated Model Uncertainty in Deep Learning.
- [109] Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q. (2017). On Calibration of Modern Neural Networks. *ArXiv*, abs/1706.04599.
- [110] Huang, Y., Li, W., Macheret, F., Gabriel, R.A., Ohno-Machado, L. (2020). A tutorial on calibration measurements and calibration models for clinical prediction models. *Journal of the American Medical Informatics Association : JAMIA*, 27, 621 - 633.
- [111] Kumar, A. (2020). Nested Cross Validation for Algorithm Selection. *Data Digital*. <https://vitalflux.com/python-nested-cross-validation-algorithm-selection/>

# Παράρτημα Α΄

## Υπεργλυκαιμία

### Μοντέλο με προσαρμοσμένο κατώφλι ταξινόμησης

#### Ορίζοντας πρόβλεψης 0-30 λεπτά

		Accuracy	Precision	Recall	AUC	F1-score
SMOTE	mean	80.89%	74.51%	77.00%	80.04%	74.13%
	std	4.56%	14.30%	8.91%	2.85%	3.82%
BSMOTE	mean	77.33 %	64.31 %	85.70%	79.26%	72.97%
	std	5.56%	8.46%	4.71%	3.92%	4.59%
Random	mean	81.21%	73.15%	77.85%	80.44%	74.60%
	std	3.86%	10.91%	7.22%	2.84%	3.94%
ADASYN	mean	69.76%	56.82%	87.94%	73.92%	67.61%
	std	7.46%	9.47%	11.26%	3.16%	2.89%

		Mean Confidence	ECE	Brier Score
SMOTE	mean	80.62%	9.41%	0.1265
	std	4.45%	1.92%	0.0283
BSMOTE	mean	78.99%	14.00%	0.1432
	std	1.396%	4.2%	0.0342
Random	mean	82.21%	10.58%	0.1248
	std	7.37%	1.24%	0.0258
ADASYN	mean	73.31%	14.27%	0.1379
	std	11.52%	2.32%	0.0278

### Ορίζοντας πρόβλεψης 30-60 λεπτά

		Accuracy	Precision	Recall	AUC	F1-score
SMOTE	mean	55.98%	43.56%	88.52%	63.76%	58.19%
	std	6.85%	3.82%	3.57%	4.59%	2.87%
BSMOTE	mean	60.30%	46.43%	91.21%	67.66%	61.37%
	std	6.04%	3.64%	3.30%	4.87%	2.68%
Random	mean	62.55%	47.69%	87.40%	68.48%	61.62%
	std	3.58%	2.77%	2.35%	2.35%	1.89%
ADASYN	mean	56.0%	44.15%	90.65%	64.38%	59.01%
	std	8.79%	4.38%	6.25%	5.4%	2.84%

		Mean Confidence	ECE	Brier Score
SMOTE	mean	60.29%	15.55%	0.1932
	std	2.85%	5.75%	0.02095
BSMOTE	mean	61.44%	15.56%	0.1995
	std	3.81%	3.33%	0.01727
Random	mean	58.99%	11.06%	0.1807
	std	2.60%	1.51%	0.0094
ADASYN	mean	58.77%	15.24%	0.1925
	std	4.75%	3.35%	0.0091

# Παράρτημα Β΄

## Υπογλυκαιμία

### Μοντέλο με προσαρμοσμένο κατώφλι ταξινόμησης

#### Ορίζοντας πρόβλεψης 0-30 λεπτά

		Accuracy	Precision	Recall	AUC	F1-score
SMOTE	mean	75.87%	7.81%	58.33%	67.39%	13.59%
	std	7.55%	2.05%	10.54%	3.79%	3.19%
BSMOTE	mean	83.80%	4.00%	15.00%	50.48%	6.07%
	std	4.47%	4.08%	12.24%	6.27%	3.76%
Random	mean	77.16%	6.58%	51.67%	64.81%	11.27%
	std	11.37%	4.89%	34.32%	12.52%	6.78%
ADASYN	mean	79.26%	10.43%	53.33%	66.72%	15.75%
	std	8.77%	9.69%	28.19%	12.34%	9.94%

		Mean Confidence	ECE	Brier Score
SMOTE	mean	76.07%	16.4%	0.1542
	std	4.11%	7.9%	0.0266
BSMOTE	mean	81.73%	14.93%	0.1325
	std	4.74%	5.79%	0.0313
Random	mean	70.98%	19.52%	0.14831
	std	2.64%	3.78%	0.0357
ADASYN	mean	74.67%	18.35%	0.1345
	std	4.17%	4.98%	0.0407

**Ορίζοντας πρόβλεψης 30-60 λεπτά**

		Accuracy	Precision	Recall	AUC	F1-score
SMOTE	mean	64.61%	7.06%	63.0%	63.83%	12.66%
	std	2.85%	2.97%	26.94%	12.81%	5.32%
BSMOTE	mean	79.39%	6.61%	40.0%	60.52%	11.34%
	std	3.65%	6.22%	37.41%	16.46%	7.67%
Random	mean	68.37%	6.82%	53.0%	60.97%	12.07%
	std	5.89%	3.54%	29.77%	14.48%	6.33%
ADASYN	mean	66.65%	4.71%	44.0%	55.78%	8.47%
	std	7.43%	3.72%	37.34%	14.91%	6.75%

		Mean Confidence	ECE	Brier Score
SMOTE	mean	67.51%	25.38%	0.1806
	std	3.11%	2.83%	0.0241
BSMOTE	mean	67.09%	16.59%	0.1252
	std	11.16%	5.76%	0.0289
Random	mean	69.91%	24.73%	0.1849
	std	3.67%	4.36%	0.0326
ADASYN	mean	71.86%	24.94%	0.1908
	std	5.38%	4.16%	0.0311