



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών

Τομέας Συστημάτων Μετάδοσης Πληροφορίας &
Τεχνολογίας Υλικών

**Ανάπτυξη ερμηνεύσιμου μοντέλου εκτίμησης του κινδύνου
νοσηλείας και επανανοσηλείας σε άτομα με Σακχαρώδη
Διαβήτη Τύπου 1**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΟΝΟΥ ΣΤΑΜΑΤΙΝΑ

Επιβλέπων : Κωνσταντίνα Νικήτα
Καθηγήτρια Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2021



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών

Τομέας Συστημάτων Μετάδοσης Πληροφορίας &
Τεχνολογίας Υλικών

**Ανάπτυξη ερμηνεύσιμου μοντέλου εκτίμησης του κινδύνου
νοσηλείας και επανανοσηλείας σε άτομα με Σακχαρώδη
Διαβήτη Τύπου 1**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΟΝΟΥ ΣΤΑΜΑΤΙΝΑ

Επιβλέπων : Κωνσταντίνα Νικήτα
Καθηγήτρια Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 26η Φεβρουαρίου 2021.

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

.....
Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

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

Αθήνα, Φεβρουάριος 2021

.....
Μόνου Σταματίνα

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

Copyright © Μόνου Σταματίνα, 2021.

Με επιφύλαξη παντός 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

Αντικείμενο της παρούσας διπλωματικής εργασίας αποτελεί η σχεδίαση, η ανάπτυξη και η αξιολόγηση ενός ερμηνεύσιμου μοντέλου εκτίμησης του κινδύνου νοσηλείας και επανανοσηλείας λόγω διαβητικής κετοξέωσης σε άτομα με Σακχαρώδη Διαβήτη Τύπου 1 (ΣΔΤ1).

Ο ΣΔ αποτελεί αυτοάνοσο νόσημα που χαρακτηρίζεται από αυξημένα επίπεδα γλυκόζης στην κυκλοφορία του αίματος. Ο ΣΔΤ1 διαγιγνώσκεται συχνότερα κατά την παιδική και την εφηβική ηλικία. Η αποτελεσματική διαχείριση της ασθένειας βασίζεται στην επίτευξη ορθού γλυκαιμικού ελέγχου. Τα άτομα με ΣΔΤ1 αντιμετωπίζουν τον κίνδυνο της διαβητικής κετοξέωσης και κατ' επέκταση της έκτακτης νοσηλείας σε νοσοκομείο με κίνδυνο της ζωής τους. Η περιορισμένη χρήση μοντέλων μηχανικής μάθησης για την πρόβλεψη νοσηλείας στα άτομα με ΣΔΤ1 και η ακόμη σπανιότερη δυνατότητα ερμηνείας τους, οδήγησε στην ανάπτυξη της παρούσας διπλωματικής.

Η ανάπτυξη του ερμηνεύσιμου μοντέλου βασίστηκε στη χρήση των Αναδρομικών Νευρωνικών Δικτύων Μακράς-Βραχείας Μνήμης (Long short-term memory - LSTM), τα οποία επιλέχθηκαν λόγω της ικανότητάς τους να διαχειρίζονται αποτελεσματικά σειριακά δεδομένα. Το μοντέλο έλαβε ως είσοδο δημογραφικά, σωματομετρικά και κλινικά δεδομένα, καθώς και δεδομένα θεραπείας, και εξήγαγε την πιθανότητα νοσηλείας και επανανοσηλείας σε βάθος ενός έτους. Για την αποτελεσματική διαχείριση της μη ισορροπημένης φύσης του συνόλου δεδομένων, υιοθετήθηκε προσέγγιση συλλογικής μάθησης με στόχο την εκπαίδευση πολλαπλών αρχικών LSTM μοντέλων, χρησιμοποιώντας ισορροπημένα υποσύνολα του αρχικού συνόλου δεδομένων, τα οποία εξήγαγαν την τελική πρόβλεψη μέσω του συνδυασμού των επιμέρους προβλέψεών τους. Στη συνέχεια οι αποφάσεις του συνδυαστικού μοντέλου ερμηνεύτηκαν μέσω της μεθόδου ερμηνευσιμότητας LIME. Πρόκειται για μία μέθοδο που χρησιμοποιεί τοπικά υποκατάστατα μοντέλα που είναι ερμηνεύσιμα και αξιοποιούνται για να εξηγήσουν μεμονωμένες προβλέψεις συνθετότερων μοντέλων μηχανικής μάθησης. Τα υποκατάστατα μοντέλα εκπαιδεύονται ώστε να μιμηθούν τη συμπεριφορά του σύνθετου-γενικού μοντέλου στην τοπική περιοχή του δείγματος ενδιαφέροντος.

Για την ανάπτυξη και την αξιολόγηση του μοντέλου χρησιμοποιήθηκαν δεδομένα 127 ατόμων με ΣΔΤ1 από τη βάση δεδομένων "SWEET", που παραχωρήθηκε από το Νοσοκομείο Παιδών "Η Αγία Σοφία". Το μοντέλο αξιολογήθηκε ως προς τη διακριτική του ικανότητα, καθώς και ως προς την ικανότητά του να παράγει ακριβείς και ερμηνεύσιμες εκτιμήσεις του κινδύνου νοσηλείας και επανανοσηλείας.

Λέξεις κλειδιά

Σακχαρώδης Διαβήτης Τύπου 1, μοντέλα πρόβλεψης κινδύνου νοσηλείας, μοντέλο LSTM, μη ισορροπημένα δεδομένα, υποδειγματοληψία, συνδιαστικό μοντέλο, ερμηνευσιμότητα, LIME, τοπική ερμηνευσιμότητα

Abstract

The present thesis aims at the design, the development and the evaluation of an interpretable model for the assessment of hospitalization and re-hospitalization risk due to diabetic ketoacidosis in patients with Type 1 Diabetes Mellitus (T1DM).

T1DM seems to be more common in children and adolescents and is characterized by decreased insulin production by the body. People with T1DM face the risk of diabetic ketoacidosis and consequently emergency hospitalization which sets their lives at risk. However, limited research has been devoted to the development of machine learning models able to predict the hospitalization risk in patients with T1DM, while the absence of interpretations on the models' decisions has hampered their adoption in clinical practice.

The development of the interpretable model was based on Long Short-Term Memory Neural Networks (LSTM), which were selected due to their excellent performance when it comes to sequential data. Frequently identified risk factors for recurrent hospital admissions due to diabetic ketoacidosis were considered to compose the input space. The imbalanced nature of the used dataset led to the adoption of an ensemble learning technique based on a subsampling method. In particular, multiple individual LSTM models were trained with balanced subsets of the original dataset, and the final risk scores were calculated through the combination of the individual models' decisions. The LIME interpretability method was leveraged towards the generation of interpretations on the ensemble model's decisions. According to LIME, local interpretable models are developed in order to explain the predictions of more complex machine learning models.

Data collected from a two-year follow-up of 127 T1DM patients at the "Agia Sofia" Children's Hospital, within the framework of the "SWEET" Initiative, were used for development and evaluation purposes. The model's predictive performance was measured in terms of discrimination and calibration, and evidence was provided on the model's ability to produce interpretable hospitalization and re-hospitalization risk scores.

Key words

Type 1 Diabetes Mellitus, Prediction of risk of hospitalization models, LSTM, unbalanced dataset, subsampling, ensemble model, interpretability, LIME, local interpretability

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στο πλαίσιο του ερευνητικού προγράμματος ENDORSE, που χρηματοδοτείται από το ΕΣΠΑ 2014-2020 (Συμφωνία Επιχορήγησης: Τ1ΕΔΚ-03695).

Αρχικά θα ήθελα να ευχαριστήσω την καθηγήτρια κυρία Κωνσταντίνα Νικήτα για την εμπιστοσύνη που μου έδειξε, δίνοντας μου την ευκαιρία να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα.

Επίσης, θα ήθελα να ευχαριστήσω την Δρ. Κωνσταντία Ζαρκογιάννη και κυρίως την υποψήφια διδάκτορα Μαρία Αθανασίου για την προθυμία και την πολύτιμη καθοδήγηση που μου παρείχε σε κάθε βήμα της διπλωματικής μου εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω τον σύζυγό μου, τους γονείς μου και την αδερφή μου, για την απόλυτη στήριξή τους με οποιοδήποτε τρόπο, όλα αυτά τα χρόνια της φοιτητικής μου πορείας καθώς και τους φίλους μου με τους οποίους μοιράστηκα τις δυσκολίες και την ομορφιά της πορείας αυτής.

Μόνου Σταματίνα,

Αθήνα, 26η Φεβρουαρίου 2021

Περιεχόμενα

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Περιεχόμενα	11
Κατάλογος σχημάτων	13
Κατάλογος πινάκων	15
1. Μεταβολισμός γλυκόζης	17
1.1 Εισαγωγή	17
1.2 Ομοιόσταση Γλυκόζης	18
1.2.1 Β-κυτταρικές ορμόνες	19
1.2.2 Α-κυτταρική ορμόνη	20
1.2.3 Ινγκρετινικές ορμόνες	21
1.3 Σακχαρώδης Διαβήτης	21
1.3.1 Διάγνωση του ΣΔ	22
1.3.2 Επιδημιολογία και Στατιστικά Στοιχεία	23
1.3.3 Επιπλοκές	29
2. Χρήση μηχανικής μάθησης για την πρόβλεψη κινδύνου νοσηλείας και επανανοσηλείας	31
2.1 Μηχανική Μάθηση	31
2.1.1 Γραμμική και Λογιστική Παλινδρόμηση	31
2.1.2 Δέντρα Απόφασης	32
2.1.3 Μηχανές Διανυσμάτων Υποστήριξης	33
2.1.4 Τεχνητά Νευρωνικά Δίκτυα	34
2.2 Τεχνικές Διαχείρισης Μη-ισορροπημένων Δεδομένων	37
3. Μηχανική μάθηση και ερμηνευσιμότητα - Η χρήση της για την ανάπτυξη ερμηνεύσιμων μοντέλων στον χώρο της υγείας	39
3.1 Ταξινόμηση των μεθόδων ερμηνευσιμότητας	39
3.2 Μέθοδοι ερμηνευσιμότητας	40
3.2.1 Σημαντικότητα των Χαρακτηριστικών	40
3.2.2 Τοπική μέθοδος ερμηνευσιμότητας ανεξαρτήτως μοντέλου (LIME)	41
3.2.3 Μέθοδος ερμηνευσιμότητας Άγκυρες (Anchors)	42
3.2.4 Ερμηνευσιμότητα μέσω των τιμών Shapley (SHAP)	42

4. Ανάπτυξη ερμηνεύσιμου μοντέλου εκτίμησης του κινδύνου νοσηλείας και επανανοσηλείας σε άτομα με Σακχαρώδη Διαβήτη Τύπου 1	45
4.1 Σύνολο Δεδομένων	45
4.2 Εννοιολογικό Πλαίσιο	46
4.3 Μοντέλο βασισμένο στα Νευρωνικά Δίκτυα Μακράς Βραχείας Μνήμης	46
4.3.1 Ανάλυση μοντέλου	46
4.3.2 Υπερπαράμετροι	49
4.4 Εκπαίδευση πρωταρχικών μοντέλων	49
4.4.1 Συνδυαστικό μοντέλο	50
4.5 Ερμηνευσιμότητα κατά LIME	52
4.5.1 Ανάλυση μηχανισμών μεθόδου	52
4.5.2 LIME και LSTM	56
5. Αποτελέσματα και Μελλοντική Εργασία	59
5.1 Ατομικά Διαγράμματα LIME	61
5.2 Συνολικά διαγράμματα σημαντικότητας χαρακτηριστικών	65
5.3 Συμπεράσματα-Μελλοντική έρευνα	68

Κατάλογος σχημάτων

1	Ομοιόσταση γλυκόζης: ρόλος της ινσουλίνης και της γλυκαγόνης. 1Α. Για υγιή άτομα σε κατάσταση νηστείας, 1Β. Για υγιή άτομα σε κατάσταση σίτισης, 1C. Για άτομα με ΣΔ σε κατάσταση νηστείας, 1D. Για άτομα με ΣΔ σε κατάσταση σίτισης	18
2	Ομοιόσταση γλυκόζης: ρόλος της ινσουλίνης, της γλυκαγόνης, της αμυλίνης και του GLP-1	20
3	Εκτιμώμενος αριθμός ενηλίκων με ΣΔ (σε εκατομμύρια)	24
4	Οι δέκα χώρες ή περιοχές με τον μεγαλύτερο αριθμό ενηλίκων με ΣΔ	25
5	Εκτιμώμενος συνολικός αριθμός ενηλίκων (20-79 ετών) με ΣΔ το 2019	26
6	Αριθμός ενηλίκων (20-79 ετών) με μη-διαγνωσμένο ΣΔ το 2019	26
7	Αριθμός θανάτων λόγω ΣΔ σε ενήλικες (20-79 ετών) ανά ηλικία και φύλο το 2019	27
8	Ποσοστό (%) των ατόμων που πέθαναν από ΣΔ πριν από την ηλικία των 60 ετών σε χώρες σε όλο τον κόσμο το 2019	28
9	Συνολικές δαπάνες υγείας που σχετίζονται με τον ΣΔ (USD) μεταξύ ενηλίκων (20-79 ετών) με ΣΔ το 2019	28
10	Μέσες δαπάνες υγείας που σχετίζονται με τον ΣΔ (USD) ανά άτομο με ΣΔ (20-79 ετών) το 2019	29
11	Νευρωνικό δίκτυο με ένα κρυμμένο επίπεδο	34
12	Βαθιά μάθηση: Νευρωνικό δίκτυο με πολλά κρυμμένα επίπεδα	35
13	Η επαναλαμβανόμενη μονάδα σε ένα LSTM περιλαμβάνει τέσσερα επίπεδα	46
14	Κατάσταση Μνήμης	47
15	Πύλη	47
16	Forget Gate	48
17	Input Gate and tanh layer	48
18	Ενημέρωση κατάστασης μνήμης	49
19	Ενημέρωση κρυφού επιπέδου	49
20	Στάδιο εκπαίδευσης	51
21	Λήψη απόφασης	53
22	Εσωτερική δομή επεξηγητή LIME	56
23	Τελική ερμηνεία	57
24	Ασθενής 1 - Ερμηνεία απόφασης για τον κίνδυνο νοσηλείας (χρονική στιγμή 1)	61
25	Ασθενής 1 - Ερμηνεία απόφασης για τον κίνδυνο επανανοσηλείας (χρονική στιγμή 2)	62
26	Ασθενής 2 - Ερμηνεία απόφασης για τον κίνδυνο νοσηλείας (χρονική στιγμή 1)	63
27	Ασθενής 2 - Ερμηνεία απόφασης για τον κίνδυνο επανανοσηλείας (χρονική στιγμή 2)	64
28	Ασθενής 3 - Ερμηνεία απόφασης για τον κίνδυνο νοσηλείας (χρονική στιγμή 1)	64
29	Ασθενής 3 - Ερμηνεία απόφασης για τον κίνδυνο επανανοσηλείας (χρονική στιγμή 2)	65
30	Διάγραμμα σημαντικότητας χαρακτηριστικών για την εκτίμηση του κινδύνου νοσηλείας σε επίπεδο πληθυσμού	66
31	Διάγραμμα σημαντικότητας χαρακτηριστικών για την εκτίμηση του κινδύνου επανανοσηλείας σε επίπεδο πληθυσμού	67

Κατάλογος πινάκων

1	AUC ανά αλγόριθμο και μοντέλο	37
2	Χαρακτηριστικά	45
3	Υπερπαραμέτροι του LSTM	50
4	Τιμές Υπερπαραμέτρων που ελέγχθηκαν	51
5	Επιλογή τελικών υπερπαραμέτρων	52
6	Κριτήρια αξιολόγησης με χρήση απλού μέσου όρου	60
7	Κριτήρια αξιολόγησης με χρήση σταθμισμένου μέσου όρου	60
8	Κριτήρια αξιολόγησης με χρήση σχήματος ψηφοφορίας	60

Κεφάλαιο 1

Μεταβολισμός γλυκόζης

1.1 Εισαγωγή

Ο Σακχαρώδης Διαβήτης (ΣΔ) ανήκει στην ομάδα των μεταβολικών διαταραχών που χαρακτηρίζονται από υψηλό επίπεδο γλυκόζης στο αίμα για παρατεταμένη χρονική περίοδο. Η γλυκόζη στο αίμα είναι η κύρια πηγή ενέργειας και προέρχεται από την τροφή. Η ινσουλίνη, μια ορμόνη που παράγεται από το πάγκρεας, βοηθά τη γλυκόζη από τα τρόφιμα να εισέλθει στα κύτταρα για να χρησιμοποιηθεί για ενέργεια. Μερικές φορές το σώμα δεν παράγει αρκετή ινσουλίνη ή δεν χρησιμοποιεί αποτελεσματικά την ινσουλίνη. Κατ' επέκταση η γλυκόζη παραμένει στο αίμα και δεν φτάνει στα κύτταρα. Με την πάροδο του χρόνου, η υπερβολική ποσότητα γλυκόζης στο αίμα μπορεί να προκαλέσει προβλήματα υγείας[1]. Οι οξείες επιπλοκές μπορεί να περιλαμβάνουν διαβητική κετοξέωση, υπεροσμωτική υπεργλυκαιμική κατάσταση ή θάνατο. Σοβαρές μακροχρόνιες επιπλοκές περιλαμβάνουν καρδιαγγειακή νόσο, εγκεφαλικό επεισόδιο, χρόνια νεφρική νόσο, έλκη ποδιών, βλάβη στα νεύρα, βλάβη στα μάτια και γνωστική εξασθένηση[2]. Αν και ο ΣΔ δεν έχει θεραπεία, μπορούν να ληφθούν μέτρα για τη διαχείριση του και να παραμείνει το άτομο υγιές.

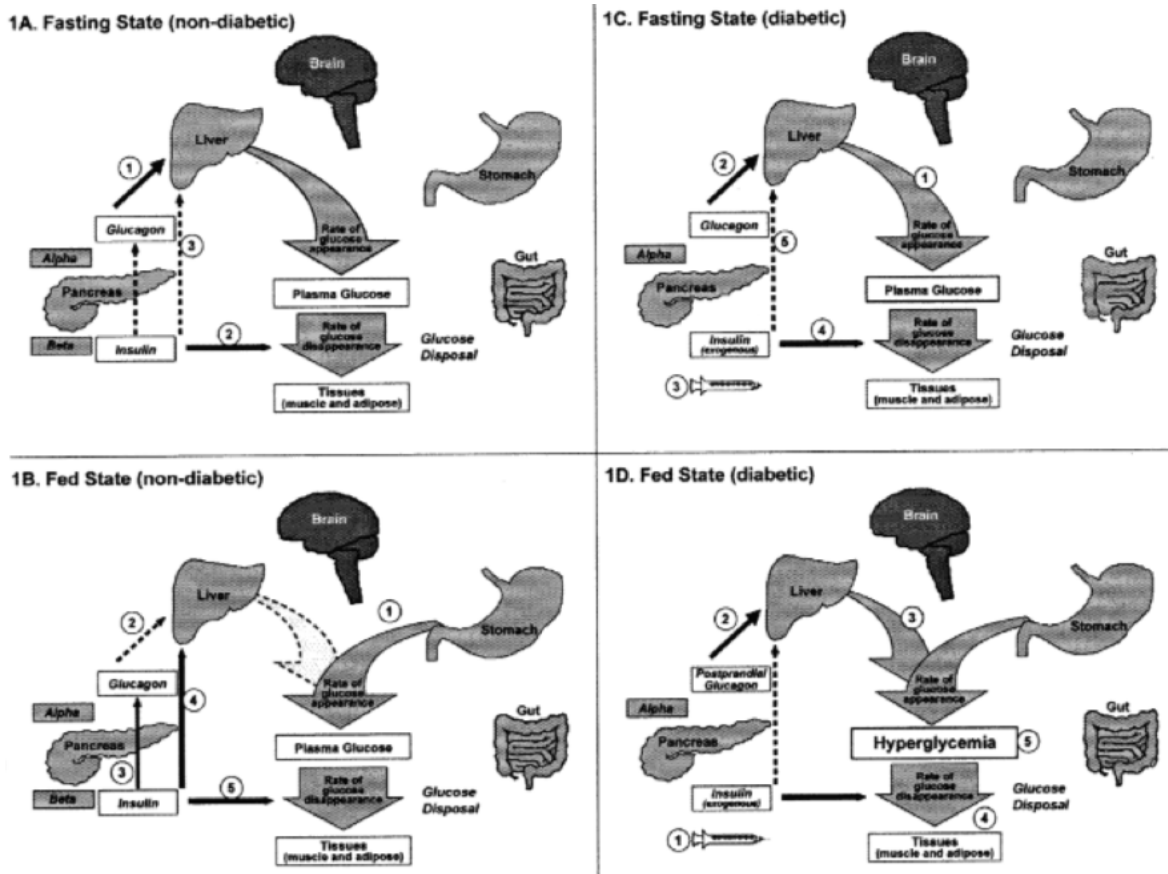
Υπάρχουν τρεις κύριοι τύποι ΣΔ: τύπος 1, τύπος 2 και ΣΔ κύησης:

- Ο ΣΔ τύπου 1 μπορεί να αναπτυχθεί σε οποιαδήποτε ηλικία, αλλά εμφανίζεται συχνότερα σε παιδιά και εφήβους. Σε αυτή την περίπτωση το σώμα παράγει πολύ λίγη ή καθόλου ινσουλίνη, πράγμα που σημαίνει ότι χρειάζονται καθημερινή εξωτερική παροχή ινσουλίνης για να διατηρηθούν τα επίπεδα γλυκόζης στο αίμα υπό έλεγχο.
- Ο ΣΔ τύπου 2 είναι πιο συχνός σε ενήλικες και αντιπροσωπεύει το μεγαλύτερο ποσοστό των περιπτώσεων ΣΔ. Σε αυτή την περίπτωση, το σώμα δεν κάνει καλή χρήση της ινσουλίνης που παράγει. Ο ακρογωνιαίος λίθος της θεραπείας ΣΔ τύπου 2 είναι ο υγιεινός τρόπος ζωής, συμπεριλαμβανομένης της αυξημένης σωματικής δραστηριότητας και της υγιεινής διατροφής. Ωστόσο, με την πάροδο του χρόνου τα περισσότερα άτομα με ΣΔ τύπου 2 θα χρειαστούν από του στόματος φάρμακα ή /και ινσουλίνη για να διατηρήσουν τα επίπεδα γλυκόζης στο αίμα τους υπό έλεγχο.
- Ο ΣΔ κύησης είναι ένας τύπος ΣΔ που χαρακτηρίζεται από υψηλή γλυκόζη στο αίμα κατά τη διάρκεια της εγκυμοσύνης και σχετίζεται με επιπλοκές τόσο στη μητέρα όσο και στο παιδί. Αυτός ο τύπος ΣΔ συνήθως εξαφανίζεται μετά την εγκυμοσύνη, αλλά οι γυναίκες που πλήττονται και τα παιδιά τους διατρέχουν αυξημένο κίνδυνο εμφάνισης ΣΔ τύπου 2 αργότερα στη ζωή τους[3].

1.2 Ομοιόσταση Γλυκόζης

Η συγκέντρωση της γλυκόζης στο πλάσμα είναι συνάρτηση του ρυθμού της γλυκόζης που εισέρχεται στην κυκλοφορία του αίματος και του ρυθμού της απομάκρυνσης της γλυκόζης από την κυκλοφορία του αίματος. Ο μεταβολισμός της γλυκόζης συνδέεται με τρεις πηγές: εντερική απορρόφηση μετά την τροφή, γλυκογονόλυση και γλυκονεογένεση. Ο κύριος παράγοντας που αφορά στο πόσο γρήγορα εμφανίζεται η γλυκόζη στην κυκλοφορία του αίματος κατά τη διάρκεια της τροφής είναι ο ρυθμός εκκένωσης του στομάχου. Άλλοι παράγοντες σχετικοί με τον μεταβολισμό της γλυκόζης προέρχονται κυρίως από ηπατικές διεργασίες: γλυκογονόλυση, διάσπαση γλυκογόνου, πολυμερισμένη μορφή αποθήκευσης γλυκόζης και γλυκονεογένεση, δηλαδή ο σχηματισμός γλυκόζης κυρίως από γαλακτικό οξύ και αμινοξέα κατά τη διάρκεια της νηστείας.

Η γλυκογονόλυση και η γλυκονεογένεση βρίσκονται εν μέρει υπό τον έλεγχο της γλυκαγόνης, μιας ορμόνης που παράγεται στα α-κύτταρα του παγκρέατος. Κατά τη διάρκεια των πρώτων 8-12 ωρών νηστείας, η γλυκογονόλυση είναι ο πρωταρχικός μηχανισμός με τον οποίο διατίθεται η γλυκόζη (Σχήμα 1Α). Η γλυκαγόνη διευκολύνει αυτή τη διαδικασία και έτσι προάγει την εμφάνιση γλυκόζης στην κυκλοφορία. Για μεγαλύτερες περιόδους νηστείας, η γλυκόζη, που παράγεται με γλυκονεογένεση, απελευθερώνεται από το ήπαρ.



Σχήμα 1: Ομοιόσταση γλυκόζης: ρόλος της ινσουλίνης και της γλυκαγόνης. 1Α. Για υγιή άτομα σε κατάσταση νηστείας, 1Β. Για υγιή άτομα σε κατάσταση σίτισης, 1C. Για άτομα με ΣΔ σε κατάσταση νηστείας, 1D. Για άτομα με ΣΔ σε κατάσταση σίτισης

Οι γλυκορυθμιστικές ορμόνες περιλαμβάνουν την ινσουλίνη, την γλυκαγόνη, την αμυλίνη, τα πεπτίδια GLP-1 και GIP, την επινεφρίνη, την κορτιζόλη και την αυξητική ορμόνη. Από αυτές, η ινσουλίνη και η αμυλίνη προέρχονται από τα β-κύτταρα, η γλυκαγόνη από τα α-κύτταρα, και τα GLP-1 και GIP από τα L-κύτταρα του εντέρου.

Οι γλυκορυθμιστικές ορμόνες του σώματος έχουν σχεδιαστεί για να διατηρούν τις συγκεντρώ-

σεις γλυκόζης στην κυκλοφορία του αίματος σε σχετικά στενό εύρος. Στην κατάσταση νηστείας, η γλυκόζη αποχωρεί από την κυκλοφορία του αίματος με σταθερό ρυθμό. Κατ'επέκταση απαιτείται ενδογενής παραγωγή γλυκόζης ώστε να υπάρξει ισορροπία μετά την απώλειά της. Οι νεφροί πραγματοποιούν επίσης γλυκονεογένεση με αποτέλεσμα την παραγωγή μιας μικρής ποσότητας γλυκόζης αλλά αυτό συμβαίνει μόνο σε περιόδους ακραίας νηστείας.

Στο δι-ορμονικό μοντέλο της ομοιόστασης της γλυκόζης, η ινσουλίνη είναι η βασική ρυθμιστική ορμόνη της απορρόφησης της γλυκόζης και η γλυκαγόνη είναι ένας σημαντικός ρυθμιστής για τη δημιουργία της γλυκόζης. Αφού κορυφωθεί η συγκέντρωση γλυκόζης στο αίμα λόγω του γεύματος, τις επόμενες ώρες μειώνεται αργά, επιστρέφοντας τελικά στα επίπεδα νηστείας. Αμέσως μετά το γεύμα, η απορρόφηση της γλυκόζης από τον σκελετικό μυ και τον λιπώδη ιστό οφείλεται κυρίως στην ινσουλίνη. Ταυτόχρονα, η παραγωγή ενδογενούς γλυκόζης καταστέλλεται από 1) την άμεση δράση της ινσουλίνης και 2) την άμεση επικοινωνία εντός του παγκρέατος μεταξύ των α και β κυττάρων, που οδηγεί σε καταστολή της γλυκαγόνης (Σχήμα 1B)[4].

1.2.1 Β-κυτταρικές ορμόνες

Ινσουλίνη

Μέχρι πρόσφατα, η ινσουλίνη ήταν η μόνη παγκρεατική ορμόνη β -κυττάρων που ήταν γνωστό ότι μειώνει τις συγκεντρώσεις γλυκόζης στο αίμα. Η ινσουλίνη, μια μικρή πρωτεΐνη που αποτελείται από δύο πολυπεπτιδικές αλυσίδες που περιέχουν 51 αμινοξέα, είναι μια βασική αναβολική ορμόνη που εκκρίνεται σε απόκριση της αυξημένης γλυκόζης στο αίμα και των αμινοξέων μετά την κατανάλωση ενός γεύματος.

Η ινσουλίνη βοηθά στον έλεγχο της μεταγευματικής γλυκόζης με τρεις τρόπους. Αρχικά, στέλνει σήμα στα κύτταρα των περιφερειακών ιστών, κυρίως σκελετικών μυών, για να αυξήσουν την πρόσληψη γλυκόζης. Δεύτερον, δρα στο ήπαρ για να προάγει τη γλυκογένεση. Ταυτόχρονα, αναστέλλει την έκκριση γλυκαγόνης από τα παγκρεατικά α -κύτταρα, στέλνοντας έτσι σήμα στο ήπαρ να σταματήσει να παράγει γλυκόζη μέσω γλυκογονόλυσης και γλυκονεογένεσης.

Η δράση της ινσουλίνης ρυθμίζεται προσεκτικά σε απόκριση των συγκεντρώσεων γλυκόζης στην κυκλοφορία του αίματος. Η ινσουλίνη δεν εκκρίνεται εάν η συγκέντρωση γλυκόζης στο αίμα είναι $\leq 3,3$ mmol/l, αλλά εκκρίνεται σε αυξανόμενες ποσότητες καθώς οι συγκεντρώσεις γλυκόζης αυξάνονται πάνω από αυτό το κατώφλι. Μετάγευματικά, η έκκριση ινσουλίνης συμβαίνει ως εξής: μια αρχική ταχεία απελευθέρωση ινσουλίνης, που ακολουθείται από αυξημένη παραγωγή ινσουλίνης και απελευθέρωση της αναλόγως με την συγκέντρωση της γλυκόζης στο αίμα. Εάν οι συγκεντρώσεις γλυκόζης παραμείνουν υψηλές, συμβαίνει μακράς δράσης απελευθέρωση ινσουλίνης.

Ενώ η γλυκόζη είναι το πιο ισχυρό ερέθισμα της ινσουλίνης, υπάρχουν και άλλοι παράγοντες που διεγείρουν την έκκριση ινσουλίνης. Αυτά τα πρόσθετα ερεθίσματα περιλαμβάνουν τις αυξημένες συγκεντρώσεις στο πλάσμα ορισμένων αμινοξέων, ειδικά αργινίνη, λευκίνη και λυσίνη, όπως επίσης και η απελευθέρωση από το έντερο των GLP-1 και GIP μετά το γεύμα[4].

Αμυλίνη

Η αμυλίνη είναι ένα πεπτίδιο που αποτελείται από 37 αμινοξέα. Ανακαλύφθηκε το 1987 και εκκρίνεται μαζί με την ινσουλίνη από τα β -κύτταρα του παγκρέατος, σαν απάντηση σε γευματικά ερεθίσματα[5].

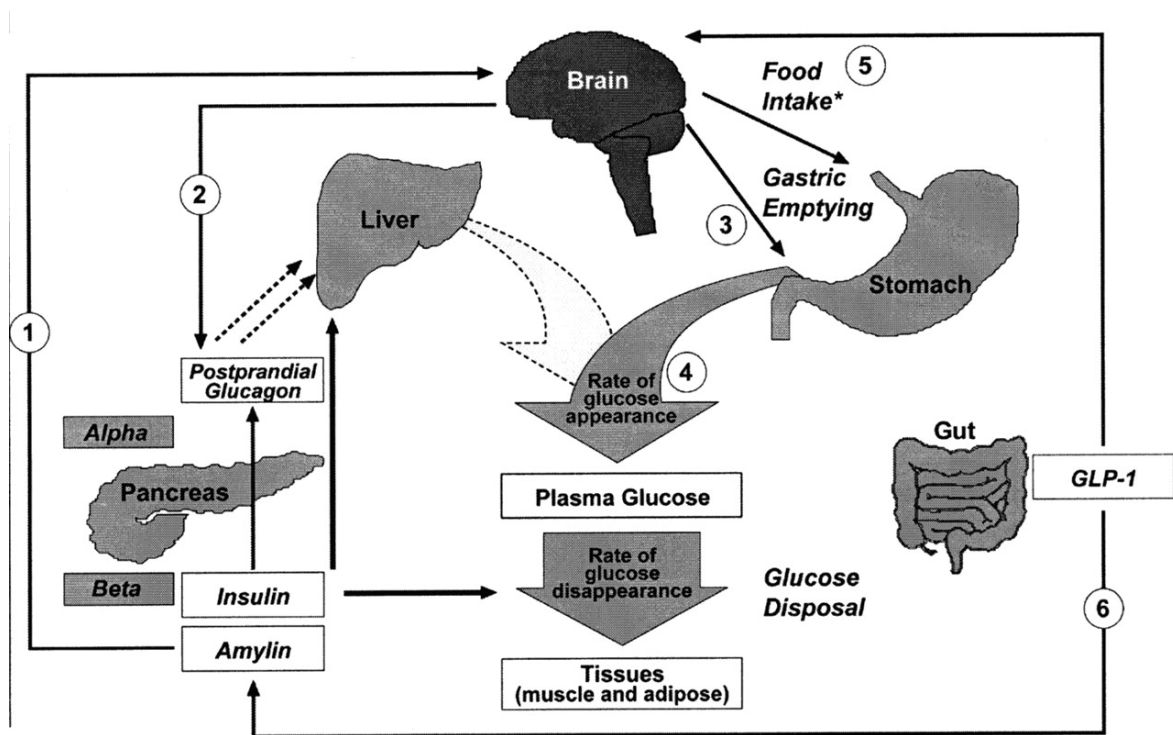
Μελέτες σε ανθρώπους έχουν δείξει ότι τα προφίλ έκκρισης και συγκέντρωσης της ινσουλίνης και της αμυλίνης είναι παρόμοια με χαμηλές συγκεντρώσεις σε περιόδους νηστείας και αυξανόμενες σε απόκριση γευματικών ερεθισμάτων. Σε άτομα με ΣΔ τύπου I υπάρχει ανεπάρκεια αμυλίνης, ενώ σε άτομα με ΣΔ τύπου II οι μεταβολές της συγκέντρωσης της αμυλίνης παραλληλίζονται με εκείνα της ινσουλίνης με διαταραγμένη ανοχή στη γλυκόζη.

Προκλινικά ευρήματα δείχνουν ότι η αμυλίνη συνεργάζεται με την ινσουλίνη για να βοηθήσει στον συντονισμό του ρυθμού εισαγωγής και εξαγωγής της γλυκόζης στην κυκλοφορία του αίματος,

αποτρέποντας έτσι μια ανώμαλη αύξηση των συγκεντρώσεων γλυκόζης.

Η αμυλίνη συμπληρώνει τις επιδράσεις της ινσουλίνης στις συγκεντρώσεις γλυκόζης στην κυκλοφορία μέσω δύο κύριων μηχανισμών (Σχήμα 2). Καταστέλλει την έκκριση μεταγευματικής γλυκαγόνης, μειώνοντας έτσι την ηπατική γλυκόζη που διεγείρεται από τη γλυκαγόνη μετά από πρόσληψη ενός γεύματος. Επίσης επιβραδύνει τον ρυθμό εκκένωσης του γαστρικού συστήματος και, επομένως, τον ρυθμό με τον οποίο τα θρεπτικά συστατικά μεταφέρονται από το στομάχι στο λεπτό έντερο για απορρόφηση.

Συνοπτικά, η αμυλίνη λειτουργεί για να ρυθμίσει τον ρυθμό εμφάνισης γλυκόζης τόσο από ενδογενείς (προερχόμενες από το ήπαρ) όσο και από εξωγενείς (προερχόμενες από γεύματα) πηγές, ενώ η ινσουλίνη ρυθμίζει το ποσοστό απορρόφησης της γλυκόζης[4].



Σχήμα 2: Ομοιόσταση γλυκόζης: ρόλος της ινσουλίνης, της γλυκαγόνης, της αμυλίνης και του GLP-1

1.2.2 Α-κυτταρική ορμόνη

Γλυκαγόνη

Η γλυκαγόνη είναι μια βασική καταβολική ορμόνη που αποτελείται από 29 αμινοξέα και εκκρίνεται από τα παγκρεατικά α-κύτταρα. Περιγράφηκε από τον Roger Unger τη δεκαετία του 1950 και χαρακτηρίστηκε ως αντίθετη στα αποτελέσματα της ινσουλίνης.

Ο Unger ήταν ο πρώτος που περιέγραψε τη διαβητική κατάσταση ως «δι-ορμονική» νόσο που χαρακτηρίζεται από ανεπάρκεια ινσουλίνης και περίσσεια γλυκαγόνης.

Η παραγωγή γλυκόζης από το ήπαρ, η οποία ρυθμίζεται κυρίως από το γλυκαγόνο, διατηρεί τη βασική συγκεντρώση γλυκόζης στο αίμα εντός φυσιολογικού εύρους κατά την κατάσταση νηστείας. Όταν η γλυκόζη στο πλάσμα πέσει κάτω από το φυσιολογικό εύρος, αυξάνεται η έκκριση γλυκαγόνης, με αποτέλεσμα την παραγωγή γλυκόζης από το ήπαρ και την επαναφορά της συγκέντρωσης γλυκόζης στο πλάσμα στο φυσιολογικό εύρος. Αυτή η ενδογενής πηγή γλυκόζης δεν απαιτείται όταν μεσολαβεί γεύμα και έτσι η έκκριση γλυκαγόνης καταστέλλεται. Η καταστολή της γλυκαγόνης οδηγεί σε σχεδόν ολική καταστολή της ηπατικής παραγωγής γλυκόζης καθώς σε περίπτωση γεύματος συνδυάζεται με την άμεση επίδραση της ινσουλίνης στο ήπαρ.

Σε διαβητική κατάσταση, υπάρχει ανεπαρκής καταστολή της έκκρισης γλυκαγόνης στη μεταγευματική κατάσταση) με αποτέλεσμα αυξημένη παραγωγή ηπατικής γλυκόζης[4].

1.2.3 Ινκρετινικές ορμόνες

GLP-1 και GIP

Οι περιπλοκές της ομοιόστασης της γλυκόζης καθίστανται σαφέστερες όταν εξετάσουμε το ρόλο των πεπτιδίων του εντέρου. Μέχρι τα τέλη της δεκαετίας του 1960, οι Perley, Kirpnis και άλλοι, απέδειξαν ότι η κατανάλωση τροφής προκαλούσε μια πιο ισχυρή απελευθέρωση ινσουλίνης από ότι η γλυκόζη που εγχύθηκε ενδοφλεβίως. Αυτό το φαινόμενο, που ονομάζεται «φαινόμενο ινκρετίνης», υποδηλώνει ότι τα σήματα από το έντερο είναι σημαντικά στην ορμονική ρύθμιση της απορρόφησης της γλυκόζης. Επιπλέον, αυτά τα ορμονικά σήματα από το εγγύς έντερο φάνηκαν να βοηθούν στη ρύθμιση της γαστρικής εκκένωσης και της κινητικότητας του εντέρου.

Έχουν εξεταστεί αρκετές ορμόνες ινκρετίνης και οι κυρίαρχες υπεύθυνες για την ομοιόσταση της γλυκόζης είναι οι GIP και GLP-1. Το GIP διεγείρει την έκκριση ινσουλίνης και ρυθμίζει το μεταβολισμό του λίπους, αλλά δεν αναστέλλει την έκκριση γλυκαγόνης ή την εκκένωση του στομάχου. Τα επίπεδα GIP είναι φυσιολογικά ή ελαφρώς αυξημένα σε άτομα με ΣΔ τύπου 2. Το GLP-1 διεγείρει την έκκριση ινσουλίνης όταν οι συγκεντρώσεις γλυκόζης στο πλάσμα είναι υψηλές. Σε αντίθεση με το GIP, το GLP-1 αναστέλλει την έκκριση γλυκαγόνης και επιβραδύνει την γαστρική εκκένωση. Οι συγκεντρώσεις κυκλοφορούντος GLP-1 είναι χαμηλές στην κατάσταση νηστείας. Ωστόσο, τόσο το GIP όσο και το GLP-1 διεγείρονται αποτελεσματικά με την κατάποση ενός μικτού γεύματος ή γευμάτων εμπλουτισμένων με λίπη και υδατάνθρακες. Το GLP-1 μειώνεται σημαντικά στη μεταγευματική κατάσταση σε άτομα με ΣΔ τύπου 2 ή σε άτομα με μειωμένη ανοχή στη γλυκόζη.

Η χορήγηση του GLP-1 έχει συσχετιστεί με τη ρύθμιση της συμπεριφοράς της σίτισης και του σωματικού βάρους. Επιπλέον, έχουν αναφερθεί παρατηρήσεις που υποδηλώνουν ότι το GLP-1 βελτιώνει την ευαισθησία στην ινσουλίνη και ενισχύει την απορρόφηση της γλυκόζης. Ωστόσο να σημειωθεί ότι, ενώ το GLP-1 αναστέλλει την έκκριση γλυκαγόνης μεταγευματικά, δεν φαίνεται να αμβλύνει την απόκριση της γλυκαγόνης στην υπογλυκαιμία[4].

1.3 Σακχαρώδης Διαβήτης

Ο ΣΔ τύπου 1 χαρακτηρίζεται από την καταστροφή των β-κυττάρων του παγκρέατος. Κατ'έπекταση προκύπτει ανεπάρκεια ινσουλίνης που συνεπάγεται επίσης και ανεπάρκεια της άλλης ορμόνης β-κυττάρων, της αμυλίνης. Συνεπώς αυξάνεται η συκέντρωση γλυκόζης μεταγευματικά λόγω της απουσίας της ινσουλίνης που θα διέγειρε την απορρόφηση της από τα κύτταρα. Επίσης παρατηρείται κακή ρύθμιση της ηπατικής παραγωγής γλυκόζης και ανώμαλη γαστρική εκκένωση μετά το γεύμα.

Όσον αφορά στον ΣΔ τύπου 2, η δράση των β-κυττάρων καθίσταται προβληματική, όπως αποδεικνύεται από την απουσία άμεσης απόκρισης στην ινσουλίνη μεταγευματικά. Η περιφερική αντίσταση στην ινσουλίνη σε συνδυασμό με την προοδευτική δυσλειτουργία των β-κυττάρων και τη μειωμένη διαθεσιμότητα ινσουλίνης, αμυλίνης και GLP-1, συμβάλλουν στην κλινική εικόνα της υπεργλυκαιμίας.

Η ανώμαλη γαστρική εκκένωση είναι συχνή τόσο στον ΣΔ τύπου 1 όσο και στον τύπου 2. Ο ρυθμός της γαστρικής εκκένωσης είναι ένας καθοριστικός παράγοντας για τις συγκεντρώσεις γλυκόζης μεταγευματικά. Εάν η γαστρική εκκένωση επιταχυνθεί, τότε η παρουσία γλυκόζης στην κυκλοφορία του αίματος, που προέρχεται από το γεύμα, δεν συγχρονίζεται με την παροχή ινσουλίνης. Σε άτομα με ΣΔ, η απουσία ή καθυστερημένη έκκριση ινσουλίνης επιδεινώνει περαιτέρω τη μεταγευματική υπεργλυκαιμία. Τόσο η αμυλίνη όσο και η GLP-1 ρυθμίζουν την εκκένωση του στομάχου επιβραδύνοντας την μεταφορά των θρεπτικών ουσιών από το στομάχι στο λεπτό έντερο[4].

1.3.1 Διάγνωση του ΣΔ

Τα πρώιμα συμπτώματα του ΣΔ, ειδικά στο ΣΔ τύπου 2, δεν είναι πάντα προφανή. Στην πραγματικότητα, οι ενδείξεις και τα συμπτώματα μπορούν να εμφανιστούν τόσο σταδιακά ώστε ο ασθενής να έχει ΣΔ τύπου 2 για χρόνια πριν διαγνωστεί με την ασθένεια. Παρακάτω παρουσιάζονται ορισμένα συμπτώματά της [6]:

- Αυξημένη δίψα και επιθυμία για ούρηση
- Κούραση
- Θολή όραση
- Απροσδόκητη απώλεια βάρους
- Αυξημένη πείνα
- Αργή επούλωση πληγών και συχνές λοιμώξεις
- Κόκκινα και πρησμένα ούλα
- Μούδιασμα

Οι διαγνωστικές εξετάσεις που γίνονται με σκοπό την αναγνώριση του ΣΔ τύπου 1 και 2 είναι: [6]:

- **Μέτρηση γλυκοζυλιωμένης αιμοσφαιρίνης (A1C):** Αυτή η εξέταση αίματος, η οποία δεν απαιτεί νηστεία, υποδεικνύει το μέσο επίπεδο σακχάρου στο αίμα τους τελευταίους δύο έως τρεις μήνες. Μετρά το ποσοστό του σακχάρου στο αίμα που συνδέεται με την αιμοσφαιρίνη, την πρωτεΐνη που μεταφέρει το οξυγόνο στα ερυθρά αιμοσφαίρια. Όσο υψηλότερα είναι τα επίπεδα σακχάρου, τόσο περισσότερη αιμοσφαιρίνη θα υπάρχει μαζί με το σάκχαρο. Ένα επίπεδο A1C 6,5% (ή υψηλότερο) σε δύο ξεχωριστές εξετάσεις δείχνει ότι ο ασθενής έχει ΣΔ. Κάτω από 5,7 θεωρείται φυσιολογικό.

Εάν τα αποτελέσματα του τεστ-A1C είναι ασαφή ή το τεστ δεν είναι διαθέσιμο ή υπάρχουν συγκεκριμένες συνθήκες που μπορούν να κάνουν το τεστ-A1C ανακριβές - όπως εάν υπάρχει εγκυμοσύνη ή ο ασθενής έχει μια ασυνήθιστη μορφή αιμοσφαιρίνης (γνωστή ως παραλλαγή αιμοσφαιρίνης) - ο γιατρός μπορεί να χρησιμοποιήσει τις ακόλουθες εξετάσεις για τη διάγνωση του ΣΔ [7] :

- **Μέτρηση σακχάρου στο αίμα:** Ένα δείγμα αίματος θα ληφθεί σε τυχαίο χρόνο. Ανεξάρτητα από το πότε λήφθηκε γεύμα τελευταία φορά, ένα τυχαίο επίπεδο σακχάρου στο αίμα 11,1 mmol/L (ή υψηλότερο) υποδηλώνει ΣΔ.
- **Μέτρηση της γλυκόζης πλάσματος νηστείας:** Ένα δείγμα αίματος θα ληφθεί μετά από μια νυχτερινή νηστεία. Αν το επίπεδο σακχάρου στο αίμα μετά τη νηστεία είναι μικρότερο από 5,6 mmol/L είναι φυσιολογικό. Εάν είναι 7 mmol/L (ή υψηλότερο) σε δύο ξεχωριστές εξετάσεις, υπάρχει ΣΔ.
- **Δοκιμασία αντοχής γλυκόζης:** Για αυτό το τεστ, προηγείται νηστεία όλη τη νύχτα και μετράται το επίπεδο σακχάρου στο αίμα. Στη συνέχεια ο ασθενής πίνει ένα ζαχαρούχο υγρό και τα επίπεδα σακχάρου στο αίμα ελέγχονται περιοδικά για τις επόμενες δύο ώρες. Ένα επίπεδο σακχάρου στο αίμα μικρότερο από 7,8 mmol/L είναι φυσιολογικό. Μια ένδειξη μεγαλύτερη από 11,1 mmol/L μετά από δύο ώρες υποδηλώνει ΣΔ.
- Εάν υπάρχει υποψία ΣΔ τύπου 1, θα ελεγχθούν τα ούρα για να αναζητηθεί η παρουσία ενός υποπροϊόντος που παράγεται όταν οι μυϊκοί και λιπώδεις ιστοί χρησιμοποιούνται για ενέργεια επειδή ο οργανισμός δεν έχει αρκετή ινσουλίνη για να χρησιμοποιήσει τη διαθέσιμη γλυκόζη (κετόνες).

1.3.2 Επιδημιολογία και Στατιστικά Στοιχεία

Ο ΣΔ είναι μια νόσος που πληττει όλο και περισσότερους ανθρώπους σε παγκόσμιο επίπεδο. Σύμφωνα με τα στοιχεία της Διεθνούς Ομοσπονδίας Διαβήτη (International Diabetes Federation- IDF) για το 2019 [8]:

- 1 στους 11 ενήλικες ηλικίας 20-79 ετών έχουν ΣΔ ενώ το απόλυτο συνολικό πλήθος ατόμων με ΣΔ παγκοσμίως ανέρχεται στα 463 εκατομμύρια.
- 1 στους 2 ενήλικες με ΣΔ δεν έχουν διαγνωστεί, δηλαδή συνολικά πρόκειται για 232 εκατομμύρια ανθρώπους.
- 1 στους 5 ανθρώπους με ΣΔ είναι πάνω από 65 ετών (136 εκατομμύρια).
- Το 10% των παγκόσμιων δαπανών για την υγεία χρησιμοποιείται για το ΣΔ, δηλαδή πρόκειται για 760 δισεκατομμύρια δολάρια.
- Όσον αφορά τις εγκύους, 1 στις 6 κήσεις παρουσιάζουν υπεργλυκαιμία και το 84% αυτών πάσχουν από ΣΔ κήσης.
- Το 79% των ανθρώπων που πάσχουν από ΣΔ ζουν σε χώρες χαμηλού ή μεσαίου εισοδήματος.
- Πάνω από 1,1 εκατομμύριο παιδιά και έφηβοι κάτω των 20 ετών πάσχουν από ΣΔ τύπου 1.
- 1 στους 13 ενήλικες ηλικίας 20-79 ετών έχουν εξασθενημένη ανοχή στη γλυκόζη (374 εκατομμύρια άνθρωποι)
- 2 στους 3 ανθρώπους με ΣΔ ζουν σε αστικές περιοχές (310,3 εκατομμύρια άνθρωποι)

Αριθμός ενηλίκων με ΣΔ

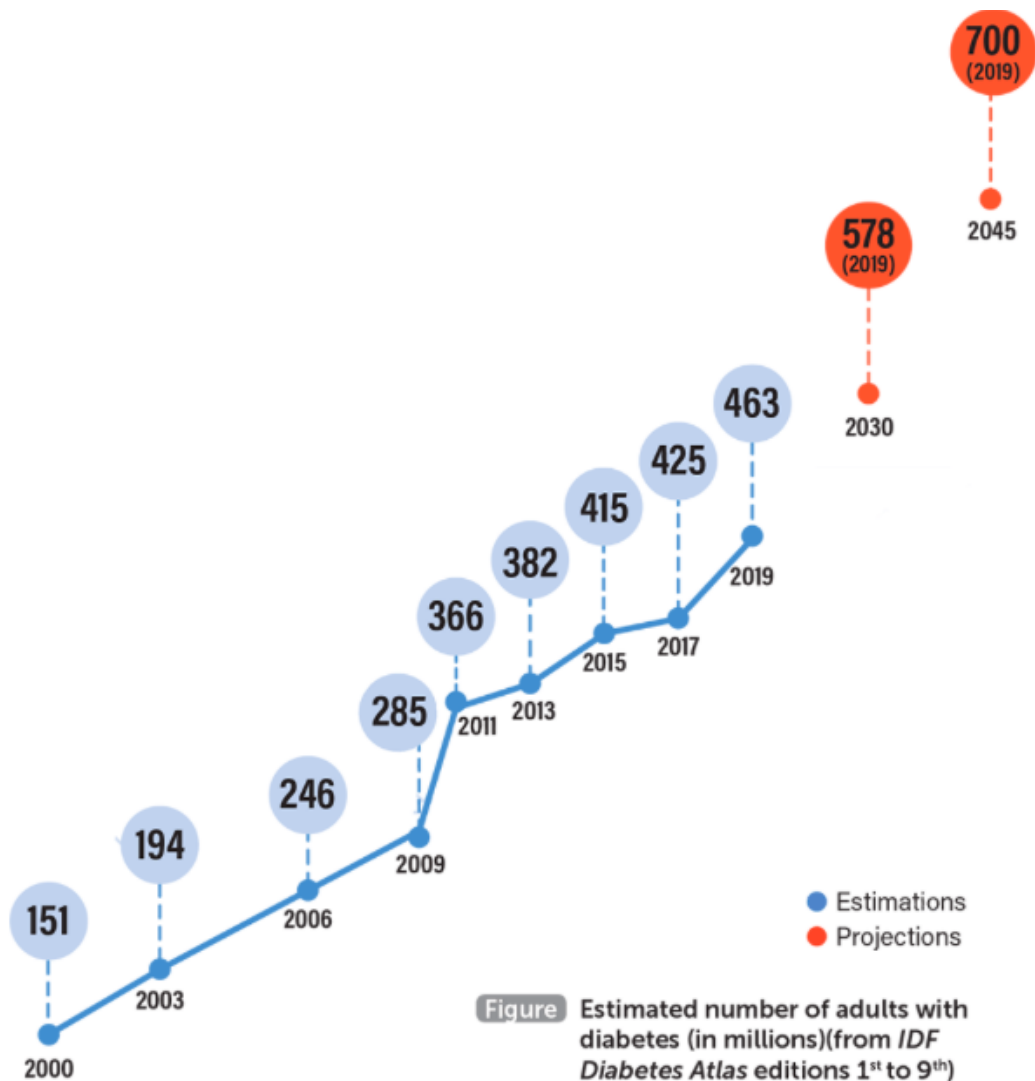
Πιο συγκεκριμένα, ο ΣΔ είναι μια από τις ταχύτερα αναπτυσσόμενες προκλήσεις για την υγεία του 21ου αιώνα, με τον αριθμό των ενηλίκων που ζουν με ΣΔ να έχουν υπερδιπλασιαστεί τα τελευταία 20 χρόνια. Το 2000, η παγκόσμια εκτίμηση των ενηλίκων που ζούσαν με ΣΔ ήταν 151 εκατομμύρια. Μέχρι το 2009 είχε αυξηθεί κατά 88% δηλαδή σε 285 εκατομμύρια. Σήμερα, υπολογίζουμε ότι το 9,3% των ενηλίκων ηλικίας 20-79 ετών - ένα εκπληκτικό 463 εκατομμύρια ανθρώπων - ζουν με ΣΔ. Επιπλέον, 1,1 εκατομμύρια παιδιά και έφηβοι κάτω των 20 ετών, ζουν με ΣΔ τύπου 1.

Πριν από μια δεκαετία, το 2010, η παγκόσμια προβολή για τον ΣΔ το 2025 ήταν 438 εκατομμύρια. Με περισσότερα από πέντε χρόνια να απομένουν, αυτή η πρόβλεψη έχει ήδη ξεπεραστεί κατά 25 εκατομμύρια. Οι αυξανόμενες τιμές νοσούντων από ΣΔ ανά έτος φαίνονται στο σχήμα 3[8].

Ηλικιακό εύρος νοσούντων από ΣΔ

Όσον αφορά στο ηλικιακό εύρος που εμφανίζεται ο ΣΔ γνωρίζουμε ότι:

- 3 στα 4 άτομα που ζουν με ΣΔ (352 εκατομμύρια άτομα) είναι σε ηλικία εργασίας (δηλαδή μεταξύ 20 και 64 ετών). Ο αριθμός αυτός αναμένεται να αυξηθεί στα 417 εκατομμύρια έως το 2030 και στα 486 εκατομμύρια έως το 2045. Αυτό δημιουργεί έναν αυξανόμενο αντίκτυπο στον άνθρωπο και θα προκαλέσει επίσης μια σοβαρή και αυξανόμενη πίεση στην παραγωγικότητα και την οικονομική ανάπτυξη τις επόμενες δεκαετίες.



Σχήμα 3: Εκτιμώμενος αριθμός ενηλίκων με ΣΔ (σε εκατομμύρια)

- Το 2019, ο εκτιμώμενος αριθμός ατόμων άνω των 65 ετών με ΣΔ ήταν 111 εκατομμύρια. 1 στους 5 ενήλικες σε αυτήν την ηλικιακή ομάδα εκτιμάται ότι έχει ΣΔ. Προβλέπεται ότι έως το 2030 ο αριθμός των ατόμων άνω των 65 με ΣΔ θα αυξηθεί περαιτέρω στα 195 εκατομμύρια. Μέχρι το 2045, θα φτάσει τα 276 εκατομμύρια. Αυτά τα δεδομένα δείχνουν σημαντικές αυξήσεις στον πληθυσμό του ΣΔ των γηράσκοντων κοινωνιών τα επόμενα 25 χρόνια, καθώς και τις αναπόφευκτες προκλήσεις που θα φέρει στους τομείς της δημόσιας υγείας και της οικονομίας.
- Όπως προαναφέρθηκε, υπολογίζεται ότι 1,1 εκατομμύρια παιδιά και έφηβοι (ηλικίας κάτω των 20) έχουν ΣΔ τύπου 1. Υπάρχουν κάποιες ενδείξεις ότι ο ΣΔ τύπου 2 μεταξύ παιδιών και εφήβων αυξάνεται σε ορισμένες χώρες, αλλά προς το παρόν δεν είναι δυνατόν να εκτιμηθούν οι αριθμοί λόγω έλλειψης δεδομένων[8].

Γεωγραφικό προφίλ του ΣΔ

Το 2019, οι χώρες με τον μεγαλύτερο αριθμό ενηλίκων με ΣΔ είναι η Κίνα, η Ινδία και οι Ηνωμένες Πολιτείες της Αμερικής και αναμένεται να παραμείνουν μέχρι το 2030. Παρακάτω, στο σχήμα 4, φαίνονται οι 10 χώρες με τους περισσότερους νοσούντες από ΣΔ και στο σχήμα 5 φαίνεται το πλήθος νοσούντων από ΣΔ ανά χώρα σε παγκόσμιο επίπεδο.

Rank	2019		2030		2045	
	Country or territory	No. of people w diabetes (millions)	Country or territory	No. of people w diabetes (millions)	Country or territory	No. of people w diabetes (millions)
1	China	116.4	China	140.5	China	147.2
2	India	77.0	India	101.0	India	134.2
3	United States of America	31.0	United States of America	34.4	Pakistan	37.1
4	Pakistan	19.4	Pakistan	26.2	United States of America	36.0
5	Brazil	16.8	Brazil	21.5	Brazil	26.0
6	Mexico	12.8	Mexico	17.2	Mexico	22.3
7	Indonesia	10.7	Indonesia	13.7	Egypt	16.9
8	Germany	9.5	Egypt	11.9	Indonesia	16.6
9	Egypt	8.9	Bangladesh	11.4	Bangladesh	15.0
10	Bangladesh	8.4	Germany	10.1	Turkey	10.4

Σχήμα 4: Οι δέκα χώρες ή περιοχές με τον μεγαλύτερο αριθμό ενηλίκων με ΣΔ

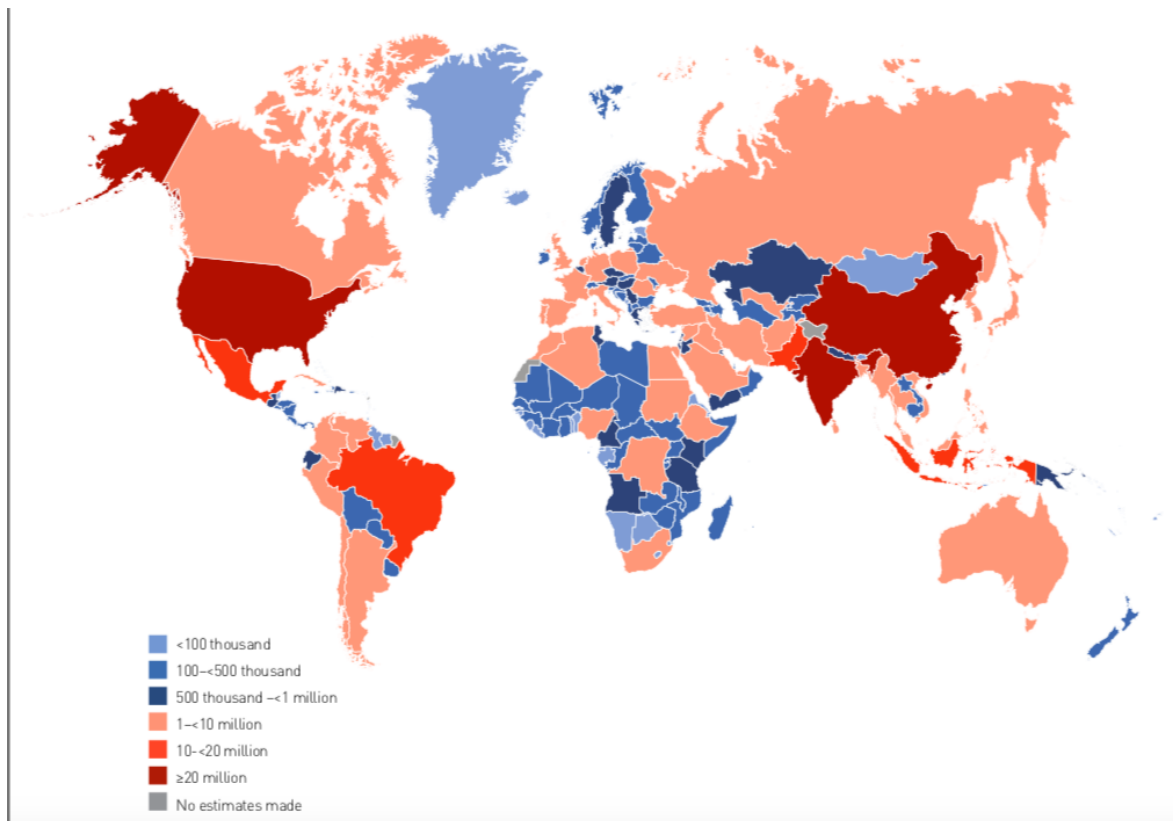
Προβλέπεται ότι ο αριθμός των ατόμων με ΣΔ στο Πακιστάν θα υπερβεί τον αριθμό των Ηνωμένων Πολιτειών της Αμερικής έως το 2045, μεταφέροντας τη χώρα στην τρίτη θέση[8].

Μη διαγνωσμένος ΣΔ

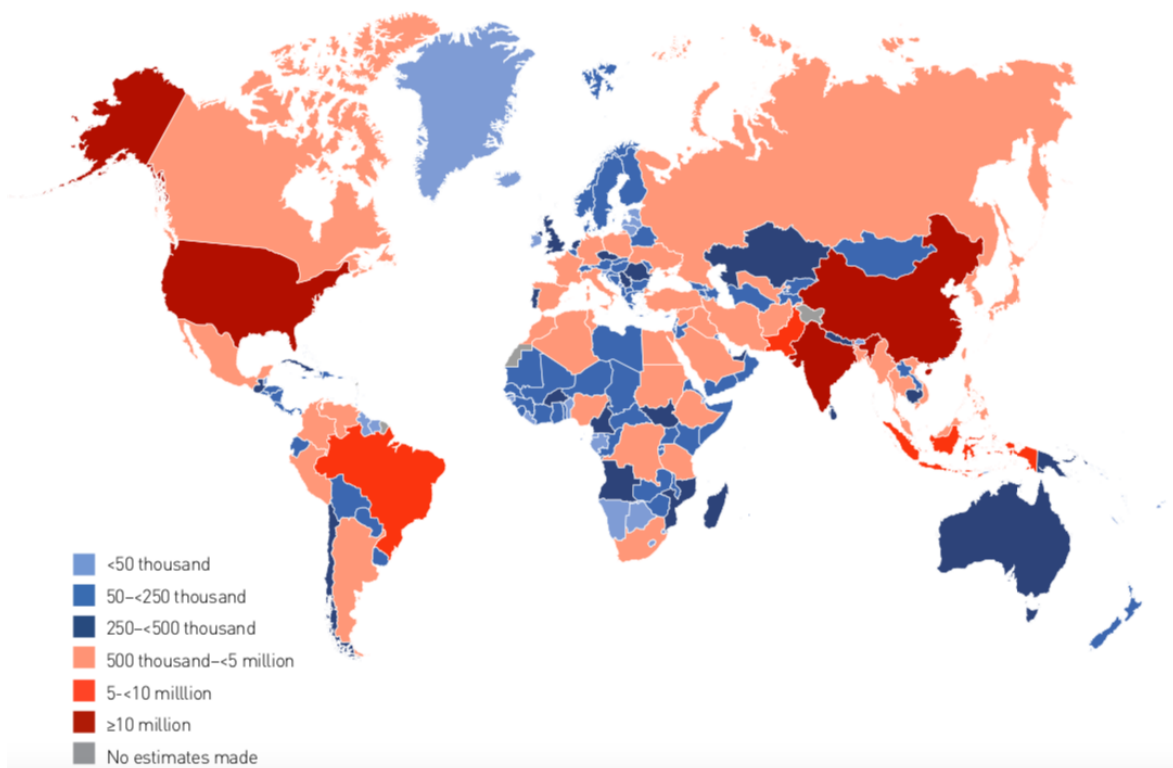
Το 2019, ένας στους δύο (50,1%), ή 231,9 εκατομμύρια από τους 463 εκατομμύρια ενήλικες που ζουν με ΣΔ, (κυρίως ΣΔ τύπου 2, ηλικίας 20-79 ετών) δεν γνωρίζουν ότι έχουν την πάθηση. Αυτές οι εκτιμήσεις δείχνουν μια επείγουσα ανάγκη για έγκαιρη ανίχνευση για βελτιωμένη παγκόσμια εξέταση του ΣΔ. Η έγκαιρη ανίχνευση είναι ζωτικής σημασίας, δεδομένου ότι ο παρατεταμένος μη διαγνωσμένος ΣΔ μπορεί να έχει αρνητικές επιπτώσεις, όπως υψηλότερο κίνδυνο επιπλοκών που σχετίζονται με τον ΣΔ, αυξημένη χρήση της υγειονομικής περίθαλψης και ανάλογο οικονομικό κόστος. Στον χάρτη του σχήματος 6 φαίνεται ο αριθμός των ενηλίκων με μη-διαγνωσμένο ΣΔ το 2019 ανά χώρα παγκοσμίως [8].

ΣΔ και θνησιμότητα

Περίπου 4,2 εκατομμύρια ενήλικες ηλικίας 20-79 ετών εκτιμάται ότι πεθαίνουν ως αποτέλεσμα του ΣΔ και των επιπλοκών του το 2019. Αυτό ισοδυναμεί με έναν θάνατο κάθε οκτώ δευτερόλεπτα. Ο ΣΔ εκτιμάται ότι σχετίζεται με το 11,3% των παγκόσμιων θανάτων από όλες τις αιτίες. Σχεδόν οι



Σχήμα 5: Εκτιμώμενος συνολικός αριθμός ενηλίκων (20-79 ετών) με ΣΔ το 2019



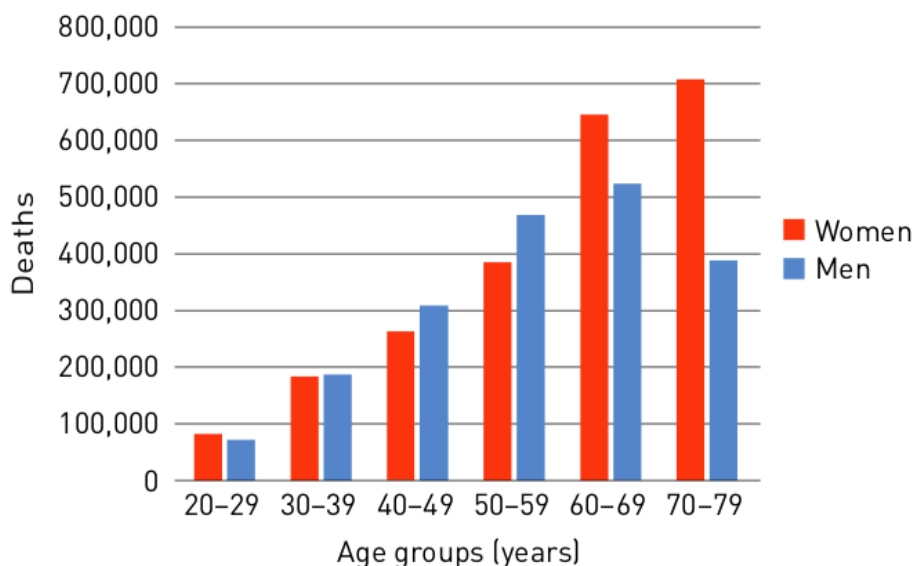
Σχήμα 6: Αριθμός ενηλίκων (20-79 ετών) με μη-διαγνωσμένο ΣΔ το 2019

μισοί (46,2%) των θανάτων που σχετίζονται με ΣΔ μεταξύ της ηλικιακής ομάδας 20-79 ετών είναι σε

άτομα κάτω των 60 ετών. Πιο συγκεκριμένα στο γράφημα του σχήματος 7 παρουσιάζεται ο αριθμός θανάτων λόγω ΣΔ σε ενήλικες (20-79 ετών) ανά ηλικία και φύλο το 2019.

Σε παγκόσμιο επίπεδο, υπάρχουν περισσότεροι θάνατοι που σχετίζονται με ΣΔ στις γυναίκες (2,3 εκατομμύρια) από ό,τι στους άνδρες (1,9 εκατομμύρια). Επίσης ο πρόωρος θάνατος και η αναπηρία λόγω ΣΔ συνδέονται επίσης με αρνητικές οικονομικές επιπτώσεις για τις χώρες, που συχνά αποκαλούνται «έμμεσο κόστος» του ΣΔ. Στις Ηνωμένες Πολιτείες της Αμερικής, εκτιμάται ότι ο πρόωρος θάνατος κοστίζει 19,9 δισεκατομμύρια δολάρια ΗΠΑ για την οικονομία ετησίως και συνολικά 90 δισεκατομμύρια δολάρια χάνονται έμμεσα λόγω ΣΔ.

Όσον αφορά στη γεωγραφική κατανομή των θανάτων από ΣΔ, η Μοζαμβίκη (91,1%), η Κένυα (88,4%), η Ουγκάντα(88,0%), το Εσουατίνι (87,7%) και η Ζιμπάμπουε (86,4%) είναι οι χώρες με το υψηλότερο εκτιμώμενο ποσοστό θάνατοι λόγω ΣΔ πριν από την ηλικία των 60 ετών σε ενήλικες ηλικίας 20-79 ετών, εν μέρει ως συνέπεια της κατανομής ηλικιών των πληθυσμών τους. Η Ιαπωνία (15,8%), η Βόρεια Μακεδονία (15,8%), η Σλοβακία (17,3%), η Σερβία (17,7%) και η Βουλγαρία (17,9%) είναι μεταξύ των χωρών με το χαμηλότερο ποσοστό θανάτων που οφείλονται σε ΣΔ κάτω των 60 ετών μεταξύ ενηλίκων ηλικίας 20-79 ετών (χάρτης σχήματος 8)[8].



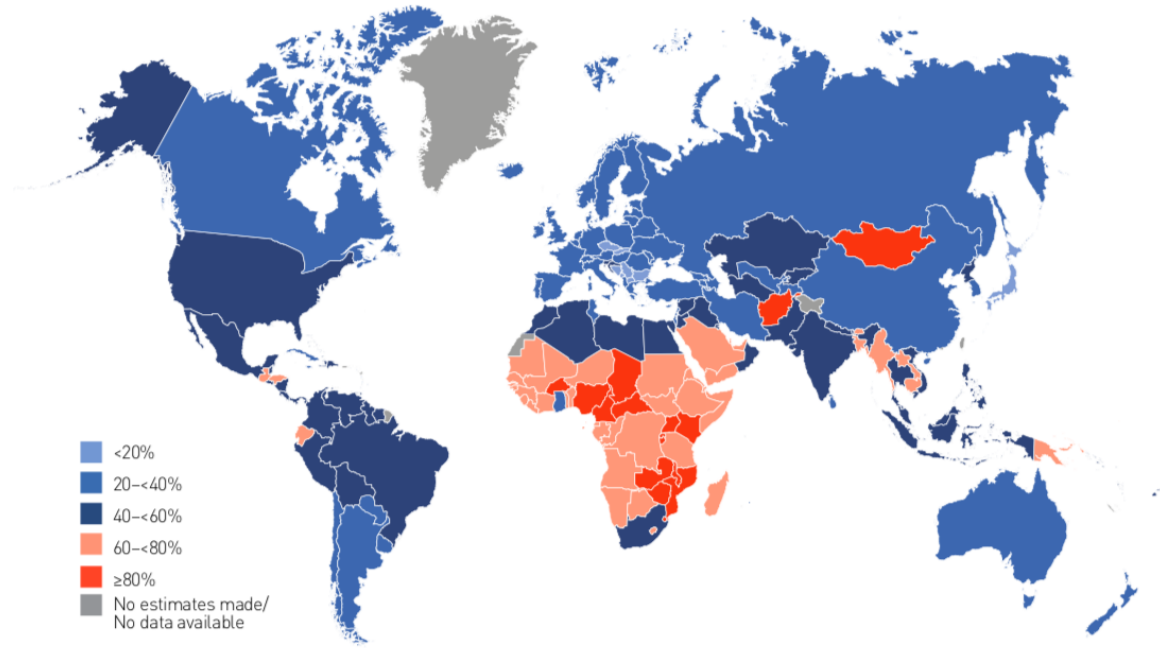
Σχήμα 7: Αριθμός θανάτων λόγω ΣΔ σε ενήλικες (20-79 ετών) ανά ηλικία και φύλο το 2019

Δαπάνες υγείας για τον ΣΔ

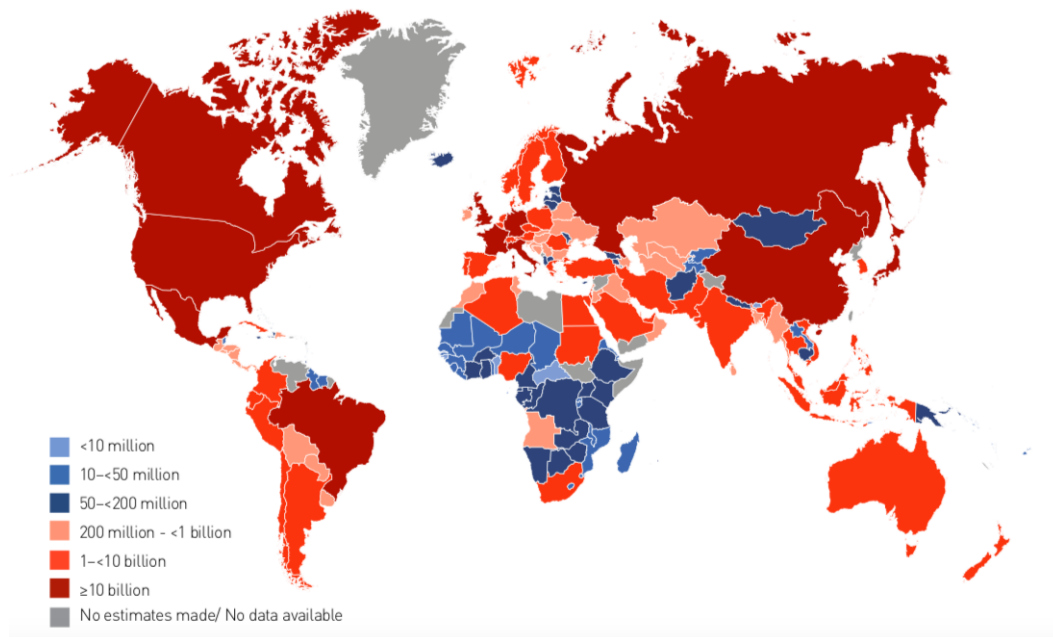
Παρά τον αντίκτυπο του ΣΔ που σχετίζεται με πρόωρη θνησιμότητα και χαμηλότερη ποιότητα ζωής λόγω επιπλοκών, ο ΣΔ επιβάλλει επίσης σημαντικό οικονομικό αντίκτυπο σε χώρες, συστήματα υγείας και στα άτομα με ΣΔ και τις οικογένειές τους, όταν η υγειονομική περίθαλψη πρέπει να χρηματοδοτηθεί από τους ίδιους.

Σε επίπεδο χώρας, οι υψηλότερες δαπάνες υγείας που σχετίζονται με τον ΣΔ εκτιμήθηκαν για τις Ηνωμένες Πολιτείες της Αμερικής με 294,6 δισεκατομμύρια δολάρια, ακολουθούμενες από την Κίνα και τη Βραζιλία, με 109,0 δισεκατομμύρια δολάρια και 52,3 δισεκατομμύρια δολάρια αντίστοιχα. Οι χώρες με τις χαμηλότερες δαπάνες για την υγεία που σχετίζονται με τον ΣΔ ήταν ο Σάο Τομέ και Πρίνσιπε και το Τουβαλού με εκτιμήσεις 1,1 εκατ. USD και 1,8 εκατ. USD (χάρτης σχήματος 9)[8].

Εξετάζοντας τις δαπάνες υγείας που οφείλονται στον ΣΔ ανά άτομο το 2019, υπάρχουν μεγάλες διαφορές μεταξύ των χωρών. Εκείνη με τις υψηλότερες ετήσιες δαπάνες ανά άτομο είναι η Ελβετία

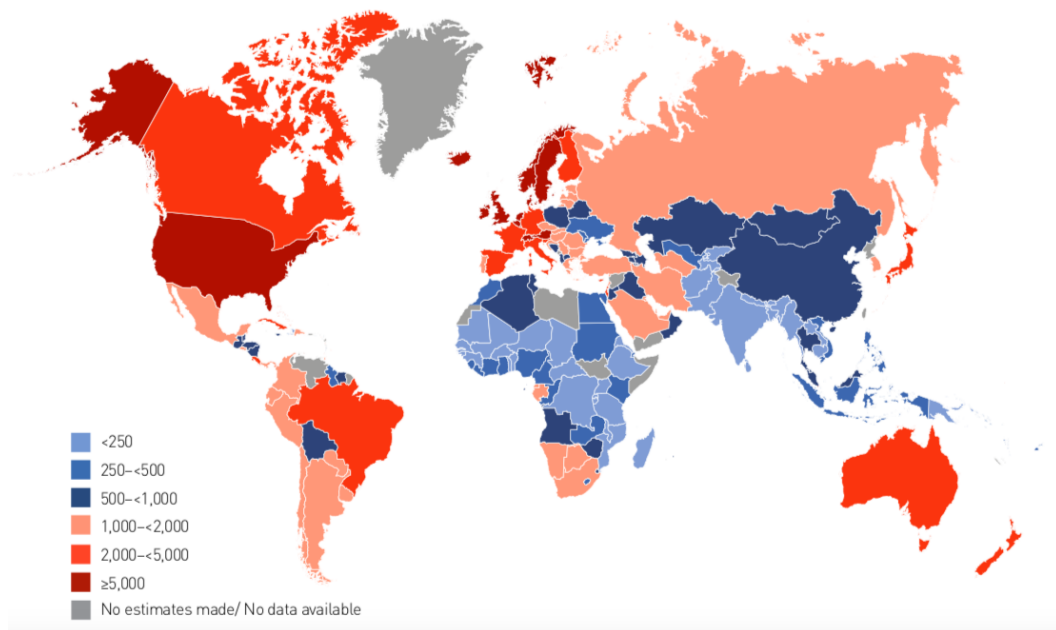


Σχήμα 8: Ποσοστό (%) των ατόμων που πέθαναν από ΣΔ πριν από την ηλικία των 60 ετών σε χώρες σε όλο τον κόσμο το 2019



Σχήμα 9: Συνολικές δαπάνες υγείας που σχετίζονται με τον ΣΔ (USD) μεταξύ ενηλίκων (20-79 ετών) με ΣΔ το 2019

με 11.916 USD, ακολουθούμενη από τις Ηνωμένες Πολιτείες της Αμερικής και τη Νορβηγία με 9.506 USD και 9.061 USD αντίστοιχα. Οι χώρες με τις χαμηλότερες ετήσιες δαπάνες ανά άτομο είναι το Μπαγκλαντές (64 USD), η Κεντροαφρικανική Δημοκρατία (72 USD) και το Νεπάλ (80 USD) (χάρτης σχήματος 10).



Σχήμα 10: Μέσες δαπάνες υγείας που σχετίζονται με τον ΣΔ (USD) ανά άτομο με ΣΔ (20-79 ετών) το 2019

1.3.3 Επιπλοκές

Τα υψηλά επίπεδα σακχάρου στο αίμα μπορούν να βλάψουν σοβαρά τα μέρη του σώματος, συμπεριλαμβανομένων των ποδιών και των ματιών. Πρόκειται για τις επιπλοκές του ΣΔ οι οποίες χωρίζονται σε δύο κατηγορίες: τις οξείες επιπλοκές και τις μακροπρόθεσμες επιπλοκές.

Οι οξείες επιπλοκές του ΣΔ, που προκύπτουν από τα υψηλά επίπεδα γλυκόζης στο αίμα, είναι συχνές στον ΣΔ τύπου 1, μπορεί να εμφανιστούν στον ΣΔ τύπου 2 και σε άλλες μορφές της πάθησης και μπορούν να οδηγήσουν σε μόνιμες νευρολογικές συνέπειες ή θάνατο. Αυτές μπορεί να συμβούν ανά πάσα στιγμή και μπορεί να οδηγήσουν σε χρόνιες επιπλοκές [9] [10]:

- **Υπογλυκαιμία:** όταν τα σάκχαρα στο αίμα είναι πολύ χαμηλά
- **Υπεργλυκαιμία:** τα σάκχαρα στο αίμα είναι πολύ υψηλά
- **Υπεροσμωτική υπεργλυκαιμική κατάσταση (HHS):** μια απειλητική για τη ζωή κατάσταση έκτακτης ανάγκης που συμβαίνει μόνο σε άτομα με ΣΔ τύπου 2 και προκαλείται από σοβαρή αφυδάτωση και πολύ υψηλά σάκχαρα στο αίμα.
- **Διαβητική κετοξέωση (DKA):** Η διαβητική κετοξέωση (DKA) είναι μια σοβαρή πάθηση που μπορεί να οδηγήσει σε διαβητικό κόμα ή ακόμη και σε θάνατο. Όταν τα κύτταρα δεν παίρνουν τη γλυκόζη που χρειάζονται για ενέργεια, ο οργανισμός αρχίζει να καίει λίπος για ενέργεια, η οποία παράγει κετόνες. Οι κετόνες είναι χημικές ουσίες που δημιουργεί το σώμα όταν διασπά το λίπος για να καλύψει τις ενεργειακές του ανάγκες. Ο οργανισμός προβαίνει σε αυτό όταν δεν έχει αρκετή ινσουλίνη ώστε να απορροφήσει τη γλυκόζη, τη φυσιολογική πηγή ενέργειας του. Όταν οι κετόνες συσσωρεύονται στο αίμα, το καθιστούν πιο όξινο. Είναι ένα προειδοποιητικό σημάδι ότι ο ΣΔ είναι εκτός ελέγχου ή ότι ο ασθενής αρρωσταίνει. Τα υψηλά επίπεδα κετονών μπορούν να δηλητηριάσουν το σώμα. Όταν τα επίπεδα γίνουν πολύ υψηλά, μπορεί να αναπτυχθεί DKA. Το DKA μπορεί να συμβεί σε οποιοδήποτε άτομο με ΣΔ, αν και είναι σπάνιο σε άτομα με ΣΔ τύπου 2.

Η θεραπεία για DKA πραγματοποιείται συνήθως στο νοσοκομείο. Αλλά μπορεί να αποτραπεί μαθαίνοντας τα προειδοποιητικά σημάδια και ελέγχοντας τα ούρα και το αίμα τακτικά.

Όσον αφορά στις μακροπρόθεσμες επιπλοκές, που μπορούν να αναπτυχθούν και μπορούν να οδηγήσουν σε σοβαρές βλάβες εάν δεν ελεγχθούν και δεν αντιμετωπιστούν, αναφέρονται τα παρακάτω [9] [10]:

- **Προβλήματα στα μάτια (αμφιβληστροειδοπάθεια):** Μερικά άτομα με ΣΔ αναπτύσσουν μια οφθαλμική νόσο που ονομάζεται διαβητική αμφιβληστροειδοπάθεια η οποία μπορεί να επηρεάσει την όρασή τους. Εάν παρατηρηθεί έγκαιρα η αμφιβληστροειδοπάθεια, μπορεί να αντιμετωπιστεί και να αποφευχθεί η απώλεια όρασης.
- **Προβλήματα στα πόδια:** Τα άτομα με ΣΔ μπορούν να αναπτύξουν πολλά διαφορετικά προβλήματα στα πόδια. Ακόμα και τα συνηθισμένα προβλήματα μπορούν να επιδεινωθούν και να οδηγήσουν σε σοβαρές επιπλοκές. Τα προβλήματα των ποδιών συμβαίνουν συχνότερα όταν υπάρχει βλάβη των νεύρων, που ονομάζεται νευροπάθεια. Αυτό μπορεί να προκαλέσει μυρμηγκιασμα, πόνο (κάψιμο ή τσούξιμο) ή αδυναμία στο πόδι. Μπορεί επίσης να προκαλέσει απώλεια αίσθησης στο πόδι και έτσι να υπάρξει τραυματισμός χωρίς να το γνωρίζει ο ασθενής. Η κακή ροή του αίματος ή οι αλλαγές στο σχήμα των ποδιών ή των δακτύλων μπορεί επίσης να προκαλέσει προβλήματα.
- **Καρδιακή προσβολή και εγκεφαλικό επεισόδιο:** Το υψηλό σάκχαρο στο αίμα μπορεί να βλάψει τα αιμοφόρα αγγεία και αυτό μπορεί μερικές φορές να οδηγήσει σε καρδιακές προσβολές και εγκεφαλικά επεισόδια.
- **Νεφρική Νόσος (Νεφροπάθεια):** Τα νεφρά είναι αξιοσημείωτα όργανα. Μέσα τους υπάρχουν εκατομμύρια μικροσκοπικά αιμοφόρα αγγεία που λειτουργούν ως φίλτρα. Η δουλειά τους είναι να απομακρύνουν τα απόβλητα από το αίμα. Μερικές φορές αυτό το σύστημα φιλτραρίσματος καταρρέει. Ο ΣΔ μπορεί να βλάψει τα νεφρά και να τα οδηγήσει σε δυσλειτουργία. Η δυσλειτουργία των νεφρών τα κάνει να χάνουν την ικανότητά τους να φιλτράρουν τα απόβλητα, με αποτέλεσμα να αναπτυχθεί νεφρική νόσος.
- **Νευρική βλάβη (νευροπάθεια):** Μερικά άτομα με ΣΔ μπορεί να αναπτύξουν κάποια νευρική βλάβη που προκαλείται με την πάροδο του χρόνου από τα υψηλά επίπεδα σακχάρου στο αίμα. Αυτό μπορεί να κάνει πιο δύσκολο για τα νεύρα να μεταφέρουν μηνύματα μεταξύ του εγκεφάλου και κάθε μέρους του σώματός. Έτσι μπορεί να επηρεαστεί ο τρόπος που το άτομο βλέπει, ακούει, αισθάνεται και κινείται.
- **Νόσος των ούλων και άλλα προβλήματα στο στόμα:** Η υπερβολική ποσότητα σακχάρου στο αίμα μπορεί να οδηγήσει σε περισσότερη ζάχαρη στο σάλιο. Αυτό προσελκύει βακτήρια που παράγουν οξύ που προσβάλλει το σμάλτο των δοντιών και βλάπτει τα ούλα. Επίσης τα αιμοφόρα αγγεία στα ούλα μπορεί να φθαρούν, καθιστώντας τα ούλα πιο ευάλωτα σε μία μόλυνση.
- **Καρκίνος:** Οι ασθενείς με ΣΔ διατρέχουν μεγαλύτερο κίνδυνο εμφάνισης ορισμένων καρκίνων. Επιπλέον ορισμένες θεραπείες για τον καρκίνο μπορούν να επηρεάσουν τον ΣΔ και να δυσκολέψουν τον έλεγχο του σακχάρου στο αίμα.

Κεφάλαιο 2

Χρήση μηχανικής μάθησης για την πρόβλεψη κινδύνου νοσηλείας και επανανοσηλείας

2.1 Μηχανική Μάθηση

Η διαδικασία λήψης αποφάσεων σε πολλές πτυχές του συστήματος υγείας είναι συχνά περίπλοκη, για αυτό αναζητούνται τρόποι ώστε οι κλινικοί ιατροί να μπορούν να λαμβάνουν καλύτερες αποφάσεις σχετικά με τις διαγνώσεις των ασθενών και τις επιλογές θεραπείας [11] [12] [13].

Η αξία της μηχανικής μάθησης στην υγειονομική περίθαλψη σχετίζεται με την ικανότητά της να επεξεργάζεται τεράστια σύνολα δεδομένων πέρα από το πεδίο της ανθρώπινης ικανότητας και, στη συνέχεια, να μετατρέπει αξιόπιστα την ανάλυση αυτών των δεδομένων σε κλινικές γνώσεις που βοηθούν τους γιατρούς στο σχεδιασμό και την παροχή φροντίδας [14], οδηγώντας τελικά σε καλύτερα αποτελέσματα, χαμηλότερο κόστος φροντίδας και μεγαλύτερη ικανοποίηση του ασθενούς [15] [16] [17].

Συγκεκριμένα η μηχανική μάθηση είναι μια εφαρμογή της τεχνητής νοημοσύνης που παρέχει στα συστήματα τη δυνατότητα αυτόματης μάθησης και βελτίωσης από την εμπειρία χωρίς να προγραμματίζονται ρητά για αυτό. Η μηχανική μάθηση επικεντρώνεται στην ανάπτυξη προγραμμάτων υπολογιστών που μπορούν να έχουν πρόσβαση σε δεδομένα και να τα χρησιμοποιούν για να μάθουν μόνοι τους. Η διαδικασία της μάθησης ξεκινά από αυτά τα δεδομένα, προκειμένου να αναζητηθούν μοτίβα και να ληφθούν καλύτερες αποφάσεις στο μέλλον. Ας διακρίνουμε δύο γενικές κατηγορίες μηχανικής μάθησης: εποπτευόμενες και μη εποπτευόμενες. Εφαρμόζουμε εποπτευόμενες τεχνικές μηχανικής μάθησης όταν έχουμε ορισμένα δεδομένα για τα οποία θέλουμε να κάνουμε κάποια πρόβλεψη ή να τα εξηγήσουμε. Αυτό το κάνουμε χρησιμοποιώντας προηγούμενα δεδομένα εισόδων και εξόδων για να προβλέψουμε μια έξοδο με βάση μια νέα είσοδο. Αντιθέτως, η μη εποπτευόμενη μηχανική μάθηση εξετάζει τρόπους συσχέτισης και ομαδοποίησης των δεδομένων χωρίς τη χρήση μιας μεταβλητής στόχου για πρόβλεψη. Με άλλα λόγια, αξιολογεί δεδομένα ως προς τα χαρακτηριστικά τους και χρησιμοποιεί τα χαρακτηριστικά αυτά για να σχηματίσει ομάδες αντικειμένων που είναι παρόμοια μεταξύ τους.

Οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούνται σε μια ευρεία ποικιλία εφαρμογών στην υγεία [18], όπως στην διάγνωση ασθενειών [19] [20], στην πρόβλεψη κινδύνου νοσηλείας, στην δημιουργία και κατασκευή φαρμάκων, στην εξατομικευμένη ιατρική [21], στα "έξυπνα" αρχεία υγείας, στην επιδημιολογία για την πρόβλεψη μελλοντικών εστιών μετάδοσης νόσων, και άλλα, όπου είναι δύσκολο ή ανέφικτο να αναπτυχθούν συμβατικοί αλγόριθμοι.

Όσον αφορά στην πρόβλεψη του κινδύνου νοσηλείας και επανανοσηλείας ασθενών, έχουν αναπτυχθεί μοντέλα από διάφορες κατηγορίες μηχανικής μάθησης και, ανάλογα με την περίπτωση, κάποια έχουν διακριθεί για τις ικανότητες πρόβλεψής τους με μεγάλη ακρίβεια. Παρακάτω αναλύονται τα πιο συχνά χρησιμοποιούμενα και παράλληλα πιο αποδοτικά μοντέλα μηχανικής μάθησης [22].

2.1.1 Γραμμική και Λογιστική Παλινδρόμηση

Οι μέθοδοι παλινδρόμησης εμπίπτουν στην κατηγορία της εποπτευόμενης μηχανικής μάθησης και συμβάλλουν στην πρόβλεψη μιας συγκεκριμένης αριθμητικής τιμής με βάση ένα σύνολο προη-

γούμενων δεδομένων.

Η απλούστερη μέθοδος είναι η γραμμική παλινδρόμηση όπου χρησιμοποιούμε τη μαθηματική εξίσωση της γραμμής ($y = m * x + b$) για να μοντελοποιήσουμε ένα σύνολο δεδομένων. Εκπαιδευούμε ένα μοντέλο γραμμικής παλινδρόμησης με πολλά ζεύγη δεδομένων (x, y) υπολογίζοντας τη θέση και την κλίση μιας ευθείας που ελαχιστοποιεί τη συνολική απόσταση μεταξύ όλων των σημείων των δεδομένων και της ευθείας. Συνεπώς, υπολογίζουμε την κλίση (m) και το (b) για μια ευθεία γραμμή που προσεγγίζει καλύτερα τα σημεία των δεδομένων.

Το λογιστικό μοντέλο παλινδρόμησης χρησιμοποιείται για να μοντελοποιήσει την πιθανότητα μιας συγκεκριμένης κλάσης ή γεγονότος που υπάρχει, πχ. υγιές/ασθενές. Είναι ένα μαθηματικό μοντέλο της στατιστικής που εκτιμά την πιθανότητα πραγματοποίησης ενός συμβάντος δοθέντων κάποιων προηγούμενων δεδομένων. Η λογιστική παλινδρόμηση λειτουργεί με δυαδικά δεδομένα, όπου είτε πραγματοποιείται το συμβάν (1) είτε δεν πραγματοποιείται (0).

Επιστημονικές Μελέτες

Σε ένα επιστημονικό άρθρο του 2006 χρησιμοποιήθηκαν δεδομένα ασθενών για τον εντοπισμό αυτών που διατρέχουν υψηλό κίνδυνο νοσηλείας μελλοντικά. Στα δεδομένα συμπεριλήφθηκαν όλοι οι ασθενείς που είχαν εισηχθεί επειγόντως στο νοσοκομείο National Health Service της Αγγλίας μεταξύ 1ης Απριλίου 2000 και 31 Μαρτίου 2001. Το μοντέλο πρόβλεψης αναπτύχθηκε χρησιμοποιώντας πολλαπλή λογιστική παλινδρόμηση και παρήγαγε ένα χαρακτηριστικό αποτέλεσμα καμπύλης λειτουργίας δέκτη (ROC) 0,72 [23].

Σε άλλο επιστημονικό άρθρο του 2012 αναλύθηκαν δεδομένα από το Get With The Guidelines-Heart Failure (GWTG-HF) του American Heart Association μεταξύ 2009 και 2012 για να περιγραφούν οι τάσεις και η πιθανή μείωση των ποσοστών επανανοσηλείας σε περίοδο 30 ημερών για ασθενείς με καρδιακή ανεπάρκεια. Χρησιμοποιήθηκε ένα μοντέλο γραμμικής παλινδρόμησης για τον προσδιορισμό παραγόντων σε επίπεδο νοσοκομείου που σχετίζονται με μείωση της πιθανότητας επανανοσηλείας σε εύρος 30 ημερών μεταξύ του 2009 και 2012[24].

2.1.2 Δέντρα Απόφασης

Το δέντρο απόφασης είναι το πιο ισχυρό και δημοφιλές εργαλείο ταξινόμησης και πρόβλεψης. Είναι ένα διάγραμμα ροής σαν δομή δέντρου, όπου κάθε εσωτερικός κόμβος δηλώνει την εξέταση της τιμής ενός χαρακτηριστικού, κάθε κλάδος αντιπροσωπεύει το αποτέλεσμα αυτής της εξέτασης και κάθε κόμβος φύλλων κρατά μια ετικέτα κλάσης. Ένα δέντρο μπορεί να «μάθει» χωρίζοντας το σύνολο δεδομένων σε υποσύνολα με βάση την τιμή ενός χαρακτηριστικού. Αυτή η διαδικασία επαναλαμβάνεται σε κάθε παράγωγο υποσύνολο με έναν αναδρομικό τρόπο που ονομάζεται αναδρομική διαμέριση. Τα δέντρα αποφάσεων μπορούν να χειριστούν δεδομένα μεγάλων διαστάσεων και σε γενικές γραμμές πετυχαίνουν καλή ακρίβεια για αυτό έχουν χρησιμοποιηθεί εκτενώς στην πρόβλεψη κινδύνου νοσηλείας και επανανοσηλείας.

Επιστημονικές Μελέτες

Στο επιστημονικό άρθρο του Ofir Ben-Assulia και Rema Padman εξετάζεται ο εντοπισμός μιας μελλοντικής επανανοσηλείας ασθενών με μεγάλο αριθμό επανειλημμένων νοσηλείων. Για αυτή την ταξινόμηση των ασθενών συγκρίθηκαν πέντε μέθοδοι: Logistic Regression, Boosted Decision Trees (BDTs), Μηχανές διανυσμάτων υποστήριξης (Support Vector Machine - SVM), Bayes Point Machine (BPM) και Two-Class Neural Network (TCNN) και επιχειρείται η πρόβλεψη της τελευταίας νοσηλείας κάθε ασθενούς. Το δέντρο αποφάσεων παρουσιάζει οριακά καλύτερη AUC (περιοχή κάτω από την καμπύλη ROC), με τιμή 0.92, σε σχέση με τη λογιστική παλινδρόμηση (0.912) και το BPM (0.913), ακολουθούμενο από το TCNN (0.878) και το SVM (0.846)[25].

Η μελέτη του Eun Whan Lee στοχεύει στον προσδιορισμό των παραγόντων κινδύνου που προβλέπουν μια επικείμενη νοσηλεία συγκρίνοντας τρία μοντέλα και επιλέγοντας το πιο επιτυχημένο. Σε

αυτή τη μελέτη συμμετείχαν 11.951 εσωτερικοί ασθενείς μεταξύ Ιανουαρίου και Δεκεμβρίου 2009, προκειμένου να προβλεφθεί ο κίνδυνος επανανοσηλείας εντός 28 ημερών μετά την έξοδο από το νοσοκομείο. Τα μοντέλα πρόβλεψης που κατασκευάστηκαν ήταν: λογιστική παλινδρόμηση, δέντρο αποφάσεων και νευρωνικό δίκτυο. Στη συνέχεια τα μοντέλα συγκρίθηκαν και αξιολογήθηκαν με βάση το ποσοστό εσφαλμένης ταξινόμησης, το ασυμπτωτικό τυπικό σφάλμα ρίζας, το διάγραμμα ανύψωσης και τη χαρακτηριστική καμπύλη ROC. Από τη σύγκριση του ποσοστού εσφαλμένης ταξινόμησης και του ασυμπτωτικού τυπικού σφάλματος ρίζας για την αξιολόγηση της προγνωστικής ισχύος των τριών μοντέλων, το ασυμπτωτικό τυπικό σφάλμα ρίζας των αποτελεσμάτων ήταν 0,385 για παλινδρόμηση, 0,373 για το δέντρο αποφάσεων και 0,384 για το νευρωνικό δίκτυο. Έτσι, το δέντρο αποφάσεων έδειξε την υψηλότερη προγνωστική ισχύ όσον αφορά στο ασυμπτωτικό τυπικό σφάλμα ρίζας. Το ποσοστό εσφαλμένης ταξινόμησης ήταν 0,214 για την παλινδρόμηση, 0,180 για το δέντρο αποφάσεων και 0,214 για το νευρωνικό δίκτυο. Έτσι, το δέντρο αποφάσεων έδειξε επίσης την υψηλότερη προγνωστική ισχύ σύμφωνα με το ποσοστό εσφαλμένης ταξινόμησης. Χρησιμοποιήθηκε επίσης το διάγραμμα ανύψωσης και η καμπύλη ROC, τα οποία χρησιμοποιούνται ευρέως για την αξιολόγηση της προγνωστικής ισχύος ενός μοντέλου, και από τα αποτελέσματα, τόσο το διάγραμμα ανύψωσης όσο και η καμπύλη ROC έδειξαν ότι το δέντρο αποφάσεων έχει την ισχυρότερη προγνωστική ικανότητα. Κατ' επέκταση, το δέντρο αποφάσεων επιλέχθηκε για να προβλέψει τον κίνδυνο επανανοσηλείας των ασθενών[26].

2.1.3 Μηχανές Διανυσμάτων Υποστήριξης

Ο στόχος του αλγορίθμου των μηχανών διανυσμάτων υποστήριξης (Support Vector Machines - SVM) είναι η εύρεση ενός υπερεπιπέδου σε έναν n -διάστατο χώρο, όπου n ο αριθμός των χαρακτηριστικών, που διαχωρίζει επιτυχημένα τα δείγματα των δύο κλάσεων. Ωστόσο υπάρχουν πολλά πιθανά υπερεπίπεδα που θα μπορούσαν να επιλεγούν. Στόχος είναι να βρεθεί αυτό που απέχει τη μέγιστη απόσταση από τα σημεία-δείγματα και των δύο κλάσεων. Η μεγιστοποίηση αυτή εξυπηρετεί στο ότι τα μελλοντικά σημεία-δείγματα θα μπορούν να ταξινομηθούν με μεγαλύτερη πιθανότητα επιτυχίας.

Επιστημονικές Μελέτες

Εφαρμογή του μοντέλου SVM συναντάμε στην παρακάτω μελέτη για την πρόβλεψη του κινδύνου επανανοσηλείας σε ασθενείς με ΣΔ.

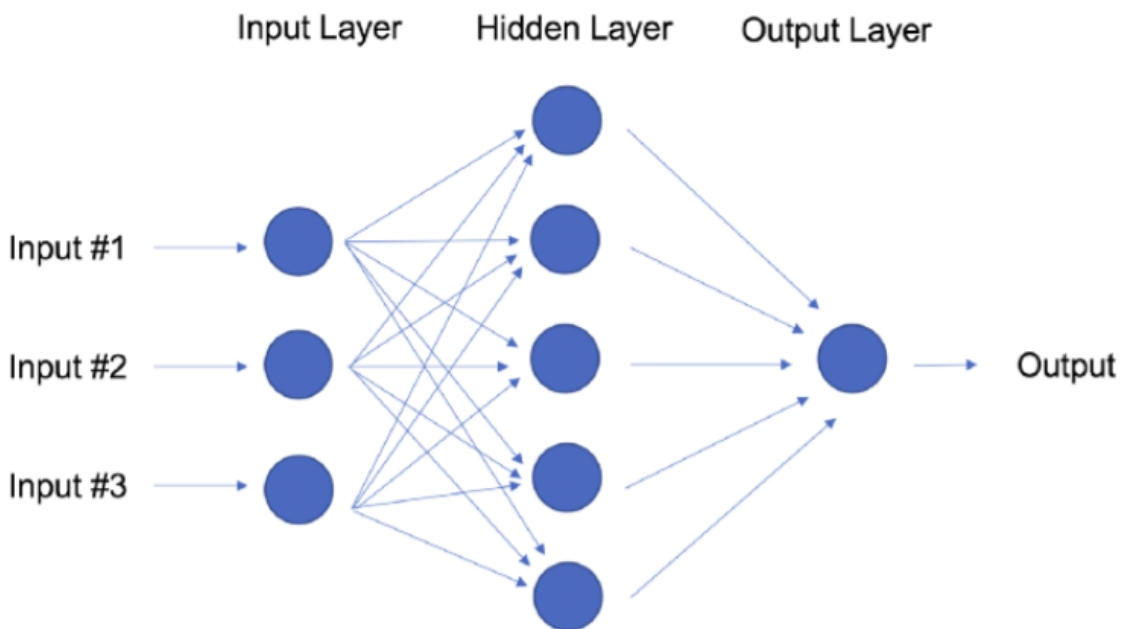
Αυτή η μελέτη παρουσιάζει μια νέα μέθοδο που συνδυάζει το SVM και έναν γενετικό αλγόριθμο για τη δημιουργία του μοντέλου πρόβλεψης κινδύνου επανανοσηλείας. Αυτό το μοντέλο στοχεύει στην παροχή υποστήριξης στις αποφάσεις των κλινικών ιατρών για τη διαχείριση ασθενών με ΣΔ. Τα πειράματα πραγματοποιήθηκαν σε ένα σύνολο 8756 ιατρικών περιπτώσεων με 50 διαφορετικά χαρακτηριστικά σχετικά με την επανανοσηλεία του ασθενή με ΣΔ. Στη συνέχεια εφαρμόστηκε 5-fold cross-validation για τη σύγκριση της απόδοσης της προτεινόμενης μεθόδου με άλλες μεθόδους (LACE score, logistic regression, naïve bayes, decision tree and feed forward neural networks). Τα πειραματικά αποτελέσματα δείχνουν ότι η προτεινόμενη μέθοδος που βασίζεται σε SVM επιτυγχάνει accuracy 81,02%, sensitivity 82,89%, specificity 79,23%, και ξεπερνά άλλους δημοφιλείς αλγόριθμους για τον εντοπισμό ασθενών με ΣΔ που μπορεί να επανανοσηλευτούν[27].

Μία άλλη μελέτη προτείνει ένα μοντέλο ταξινόμησης για την πρόβλεψη του κινδύνου επανανοσηλείας ασθενών χρησιμοποιώντας particle swarm optimization (PSO) και SVMs. Ένα μεγάλο ποσοστό των επανανοσηλείων προέρχεται από τη χαμηλή ποιότητα περίθαλψης κατά τη διάρκεια της παραμονής των ασθενών σε νοσοκομεία. Για την αποτελεσματική πρόληψη της επανανοσηλείας, είναι σημαντικό για τα νοσοκομεία να εντοπίζουν ασθενείς με υψηλό κίνδυνο επανανοσηλείας. Σε αυτήν την έρευνα, προτείνονται τα SVMs με διαφορετικές λειτουργίες πυρήνα για την ταξινόμηση των ασθενών σε κατηγορίες υψηλού και χαμηλού κινδύνου με βάση τα χαρακτηριστικά του ασθενούς. Επιπλέον, το PSO υλοποιείται για την αναζήτηση της βέλτιστης ρύθμισης των παραμέτρων λειτουργίας του πυρήνα του SVM. Το προτεινόμενο μοντέλο έχει εφαρμοστεί για την ταξινόμηση ασθενών με καρδιακή ανεπάρκεια. Τα αποτελέσματα δείχνουν ότι η ικανότητα ταξινόμησης αυτού του προτεινόμενου μο-

ντέλου PSO-SVM ξεπερνά τα άλλα όσον αφορά στην ακρίβεια, την ευαισθησία και το F-measure. Κατά συνέπεια, τα νοσοκομεία μπορούν να χρησιμοποιούν τους πόρους τους πιο αποτελεσματικά για να ενισχύουν τις παρεμβάσεις σε ασθενείς υψηλού κινδύνου και να μειώσουν το συνολικό ποσοστό επανανοσηλείας[28].

2.1.4 Τεχνητά Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα, ή αλλιώς νευρωνικά δίκτυα, είναι υπολογιστικά συστήματα που δημιουργήθηκαν ύστερα από μίμηση των βιολογικών νευρικών δικτύων που εμφανίζονται στους ζωικούς εγκεφάλους. Ένα τεχνητό νευρωνικό δίκτυο βασίζεται σε ένα σύνολο συνδεδεμένων κόμβων που ονομάζονται τεχνητοί νευρώνες, οι οποίοι μοντελοποιούν κατά κάποιο τρόπο τους νευρώνες ενός βιολογικού εγκεφάλου. Κάθε σύνδεση, όπως οι συνάψεις σε έναν βιολογικό εγκέφαλο, μπορεί να μεταδώσει ένα σήμα από ένα νευρώνα σε άλλους. Ένας τεχνητός νευρώνας που λαμβάνει σήμα στη συνέχεια το επεξεργάζεται και μπορεί να σηματοδοτήσει άλλους νευρώνες που συνδέονται με αυτόν. Το "σήμα" σε μια σύνδεση είναι ένας πραγματικός αριθμός και η έξοδος κάθε νευρώνα υπολογίζεται από κάποια μη γραμμική συνάρτηση του αθροίσματος των εισόδων του. Οι συνδέσεις ονομάζονται ακμές. Οι νευρώνες και οι ακμές συνδέονται συνήθως με ένα βάρος που προσαρμόζεται καθώς προχωρά η μάθηση. Το βάρος αυξάνει ή μειώνει την ισχύ του σήματος σε μια σύνδεση. Επίσης οι νευρώνες μπορεί να έχουν ένα κατώφλι έτσι ώστε ένα σήμα να αποστέλλεται μόνο εάν το συνολικό σήμα ξεπερνά την τιμή του κατωφλίου. Οι νευρώνες, εκτός των άλλων, συνήθως οργανώνονται σε επίπεδα. Διαφορετικά επίπεδα ενδέχεται να εκτελούν διαφορετικούς μετασχηματισμούς στις εισόδους τους. Τα σήματα μετακινούνται από το πρώτο επίπεδο (το επίπεδο εισόδου), στο τελευταίο επίπεδο (το επίπεδο εξόδου), πιθανώς αφού περάσουν πολλές φορές από τα ενδιάμεσα επίπεδα[29].

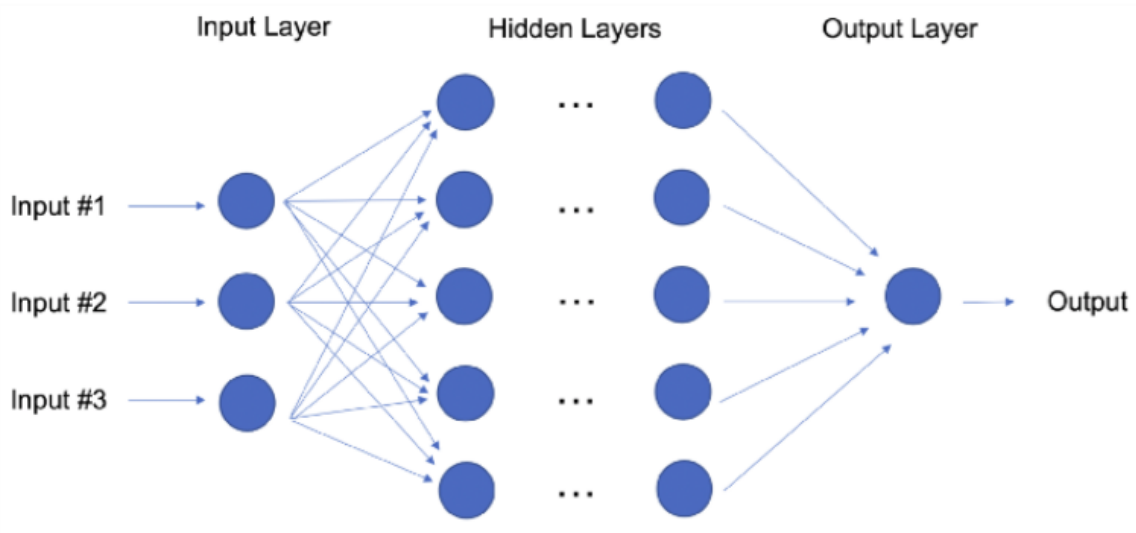


Σχήμα 11: Νευρωνικό δίκτυο με ένα κρυμμένο επίπεδο

Ο στόχος των νευρωνικών δικτύων είναι να συλλάβει μη γραμμικά μοτίβα στα δεδομένα προσθέτοντας αυτά τα επίπεδα παραμέτρων στο μοντέλο. Στο σχήμα 11, το απλό αυτό νευρωνικό δίκτυο έχει τρεις εισόδους, ένα κρυφό επίπεδο με πέντε παραμέτρους και ένα επίπεδο εξόδου. Στην πραγματικότητα, η δομή των νευρωνικών δικτύων είναι αρκετά ευέλικτη.

Βαθιά Μάθηση

Ο όρος "βαθιά μάθηση" αναφέρεται σε πολυεπίπεδες αρχιτεκτονικές τεχνητών νευρωνικών δικτύων με πολλά κρυμμένα επίπεδα (βλέπε σχήμα 12). Η ιδέα είναι συγκρίσιμη με τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου. Όσο αυξάνεται το βάθος των επιπέδων, τόσο περισσότερο πυκνά σε πληροφορία είναι τα χαρακτηριστικά που εξάγονται και καταγράφονται. Η βαθιά μάθηση είναι επίσης ισχυρή στην πρόβλεψη, αλλά ο υπολογισμός είναι συχνά βαρύς λόγω των πολύπλοκων δομών των επιπέδων. Ευτυχώς, η πρόσφατη ανάπτυξη της ικανότητας ανάλυσης μεγάλων όγκων δεδομένων διευκόλυνε την εφαρμογή αυτών των εξαιρετικά ικανών στην πρόβλεψη αλγορίθμων. Για την καλύτερη απόδοση, οι τεχνικές βαθιάς μάθησης απαιτούν πολλά δεδομένα και πολύ υπολογιστική ισχύ, καθώς η μέθοδος αυτορυθμίζει πολλές παραμέτρους μέσα σε τεράστιες αρχιτεκτονικές. Συγκεκριμένα, οι τεχνικές βαθιάς μάθησης είναι εξαιρετικά επιτυχημένες στους τομείς της ταξινόμησης και αναγνώρισης εικόνας, κειμένου, ήχου και βίντεο[22].



Σχήμα 12: Βαθιά μάθηση: Νευρωνικό δίκτυο με πολλά κρυμμένα επίπεδα

Αναδρομικά Νευρωνικά Δίκτυα

Τα αναδρομικά νευρωνικά δίκτυα (RNN) είναι σχεδιασμένα για την ανάλυση ακολουθιακών δεδομένων μέσω των κρυφών δομών τους. Σε ορισμένες εφαρμογές όπως η επεξεργασία κειμένου, η αναγνώριση ομιλίας και οι ακολουθίες DNA, η εκάστοτε έξοδος εξαρτάται από υπολογισμούς σε προηγούμενα στάδια. Δεδομένου ότι τα RNN ασχολούνται με ακολουθιακά δεδομένα, είναι κατάλληλα για τον τομέα της μηχανικής μάθησης στην υγεία, όπου διατίθενται τεράστιες ποσότητες ακολουθιακών δεδομένων προς επεξεργασία[30]. Γενικά, τα RNN τροφοδοτούνται με δεδομένα εισόδου σε μορφή ακολουθιών που περιέχουν αρκετές αλληλεξαρτήσεις. Ως εκ τούτου, διαθέτουν μια σημαντική δομή για τη διατήρηση των πληροφοριών προηγούμενων βημάτων. Η έξοδος που παράγεται στο χρόνο $t1$ επηρεάζει την παράμετρο που εισάγεται στο χρόνο $t1 + 1$. Με αυτό τον τρόπο, τα RNN διατηρούν δύο είδη εισόδου για να παράγουν την έξοδο: την παροντική τιμή εισόδου και την τιμή του πρόσφατου παρελθόντος. Λόγω ορισμένων αδυναμιών του RNN μοντέλου έχουν δημιουργηθεί διάφορες βελτιωμένες παραλλαγές του: το αναδρομικό νευρωνικό δίκτυο μακράς βραχείας μνήμης (Long Short-Term Memory - LSTM) και το νευρωνικό δίκτυο αναδρομικών πυλών (Gated Recurrent Units - GRU).

Επιστημονικές Μελέτες

Τα νευρωνικά δίκτυα έχουν χρησιμοποιηθεί ιδιαιτέρως στην υλοποίηση μοντέλων πρόβλεψης του κινδύνου νοσηλείας και επανανοσηλείας όπως φαίνεται από τις επιστημονικές μελέτες που παρατίθενται παρακάτω.

Στη συγκεκριμένη μελέτη χρησιμοποιήθηκαν δεδομένα από περισσότερες από 300.000 νοσηλείες στην Καλιφόρνια σύμφωνα με το σύστημα ηλεκτρονικών αρχείων υγείας του Sutter Health. Δημιουργήθηκε και αξιολογήθηκε ένα νευρωνικό δίκτυο με βάση τη βιβλιοθήκη TensorFlow της Google. Μέσω της σύγκρισης με άλλα παραδοσιακά και μη παραδοσιακά μοντέλα, αποδείχθηκε ότι τα νευρωνικά δίκτυα είναι εξαιρετικοί υποψήφιοι για να συλλάβουν την πολυπλοκότητα και την αλληλεξάρτηση διαφόρων παραγόντων που καταγράφονται στο ηλεκτρονικό αρχείο υγείας. Το LACE, ως το τρέχον πρότυπο σύγκρισης, παρουσίασε τιμή 0,20 στην μετρική πιστότητας (precision) όσον αφορά στον εντοπισμό ασθενών υψηλού κινδύνου για επανανοσηλεία. Αντίθετα, το νευρωνικό δίκτυο κατάφερε τιμή πιστότητας 0,24, το οποίο είναι μια βελτίωση 20% σε σχέση με το LACE[31].

Σε μια διαφορετική μελέτη, χρησιμοποιήθηκαν πραγματικά δεδομένα από περισσότερους από 7500 ασθενείς με συμφορητική καρδιακή ανεπάρκεια που νοσηλεύτηκαν μεταξύ 2012 και 2016 στη Σουηδία. Δημιουργήθηκε και αξιολογήθηκε ένα μοντέλο βαθιάς μάθησης για την πρόβλεψη μη προγραμματισμένης επανανοσηλείας εντός 30 ημερών. Παρουσιάστηκε μια διαμόρφωση ενός αναδρομικού νευρωνικού δικτύου μακράς βραχείας μνήμης (LSTM) χρησιμοποιώντας εξειδικευμένα χαρακτηριστικά και ενσωματώνοντας έννοιες κλινικής ιατρικής. Αξιολογήθηκε η συμβολή κάθε χαρακτηριστικού στην ακρίβεια της πρόβλεψης με τις μετρικές ROC-AUC και F1-μέτρο. Αποδεικνύεται ότι το μοντέλο επιτυγχάνει υψηλή ικανότητα διάκρισης με AUC: 0,77 και F1: 0,51, ξεπερνώντας τα απλούστερα μοντέλα σε τουλάχιστον δύο μετρικές αξιολόγησης[32].

Σε μία τρίτη μελέτη οι ερευνητές βασίζονται σε διαφορετικούς αλγόριθμους μηχανικής μάθησης (απλή λογιστική παλινδρόμηση, δέντρα αποφάσεων, Naive Bayes, Random Forest και μοντέλο βαθιάς μάθησης) προβλέπουν την πιθανότητα επανανοσηλείας σε περίοδο 30 ημερών σε δύο χρονικές στιγμές: κατά την εισαγωγή στο νοσοκομείο και κατά την έξοδο από το νοσοκομείο, χρησιμοποιώντας δεδομένα 1.633.099 νοσηλείων στο Κεμπέκ του Καναδά μεταξύ του 1995 και 2012. Η απόδοση των προβλέψεων μετράται σύμφωνα με τη μετρική AUC. Τα αποτελέσματα έδειξαν ότι η βαθιά μάθηση παρήγαγε εξαιρετική πρόβλεψη επανανοσηλείας σε ολόκληρη την επαρχία ενώ ο αλγόριθμος Random Forest έφτασε σε πολύ παρόμοια επίπεδα. Η AUC για αυτούς τους δύο αλγόριθμους έφτασε πάνω από το 78% στην εκτίμηση κατά την εισαγωγή στο νοσοκομείο και πάνω από το 87% κατά την έξοδο από το νοσοκομείο. Οι τιμές της AUC για όλους τους αλγόριθμους που εξετάστηκαν φαίνονται στον πίνακα 1[33].

Σε αυτό το σημείο αξίζει να αναφερθεί και μία μελέτη που προτείνει ένα συμβουλευτικό σύστημα υποστήριξης βασισμένο στα νευρωνικά δίκτυα σχετικά με το καθεστώς ινσουλίνης και την προσαρμογή της δόσης σε ασθενείς με διαβήτη τύπου 1. Το σύστημα αποτελείται από δύο τροφοδοτικά νευρωνικά δίκτυα (feed-forward neural networks), εκπαιδευμένα με τον αλγόριθμο οπίσθιας διάδοσης (backpropagation) και προσαρμοζόμενο ρυθμό μάθησης. Η είσοδος του μοντέλου απαρτίζεται από τα επίπεδα γλυκόζης του ασθενούς, την ποσότητα πρόσληψης ινσουλίνης και τα παρατηρούμενα συμπτώματα υπογλυκαιμίας σε σύντομο χρονικό διάστημα. Η έξοδος του πρώτου νευρωνικού δικτύου παρέχει το καθεστώς ινσουλίνης, το οποίο εφαρμόζεται ως είσοδος στο δεύτερο νευρωνικό δίκτυο για την εκτίμηση των κατάλληλων δόσεων ινσουλίνης για μικρό χρονικό διάστημα. Η ικανότητα του συστήματος να προτείνει το καθεστώς ινσουλίνης είναι εξαιρετική, ενώ η απόδοσή του στην προσαρμογή των δόσεων ινσουλίνης για έναν συγκεκριμένο ασθενή εξαρτάται σε μεγάλο βαθμό από το σύνολο δεδομένων που χρησιμοποιήθηκε κατά τη διάρκεια της διαδικασίας προπόνησης. Παρά τους περιορισμούς των προσεγγίσεων που βασίζονται σε υπολογιστή, αυτή η μελέτη δείχνει ότι τα τεχνητά νευρικά δίκτυα μπορούν να βοηθήσουν τους ασθενείς με διαβήτη στην προσαρμογή της ινσουλίνης [34].

Μια ακόμη σχετική μελέτη στοχεύει στην ανάπτυξη και αξιολόγηση ενός εξατομικευμένου συστήματος παροχής συμβουλών για έγχυση ινσουλίνης (Insulin Infusion Advisory System - IIAS), ικανό να παρέχει εκτιμήσεις σε πραγματικό χρόνο για τον κατάλληλο ρυθμό έγχυσης ινσουλίνης για

Πίνακας 1: AUC ανά αλγόριθμο και μοντέλο

Algorithm	Logistic regression	Decision Tree	Naïve Bayes	Random Forest	Deep Learning
Admission	0.7121	0.5017	0.7463	0.7814	0.7877
Discharge	0.7889	0.7264	0.8155	0.8706	0.8776

ασθενείς με ΣΔΤ1 που χρησιμοποιούν συνεχή έλεγχο γλυκόζης και αντλίες ινσουλίνης. Το σύστημα βασίζεται σε έναν μη γραμμικό ελεγκτή πρόβλεψης μοντέλου (Nonlinear Model-Predictive Controller - NMPC) που χρησιμοποιεί ένα εξατομικευμένο μοντέλο μεταβολισμού γλυκόζης-ινσουλίνης, που αποτελείται από δύο διαμεριστικά μοντέλα και ένα αναδρομικό νευρωνικό δίκτυο. Το μοντέλο λαμβάνει ως είσοδο τις πληροφορίες του ασθενούς σχετικά με την πρόσληψη γεύματος, τις μετρήσεις γλυκόζης και τα ποσοστά έγχυσης ινσουλίνης και παρέχει προβλέψεις γλυκόζης. Οι προβλέψεις τροφοδοτούνται στο NMPC, προκειμένου το τελευταίο να εκτιμήσει τους βέλτιστους ρυθμούς έγχυσης ινσουλίνης. Το IIAS έχει αξιολογηθεί χρησιμοποιώντας ένα κατάλληλο περιβάλλον προσομοίωσης και ήταν σε θέση να χειριστεί διάφορα προφίλ γεύματος, συνθήκες νηστείας, μεταβλητότητα μεταξύ των ασθενών, διακύμανση στις φυσιολογικές παραμέτρους εντός της ημέρας και σφάλματα στις εκτιμήσεις της ποσότητας γεύματος[35].

2.2 Τεχνικές Διαχείρισης Μη-ισορροπημένων Δεδομένων

Τα μη-ισορροπημένα σύνολα δεδομένων είναι μια ειδική περίπτωση για προβλήματα ταξινόμησης όπου η κατανομή των δεδομένων δεν είναι ομοιόμορφη μεταξύ των κλάσεων. Συνήθως, πρόκειται για προβλήματα ταξινόμησης δύο κλάσεων: την πλειοψηφική (αρνητική) τάξη και την μειοψηφική (θετική) τάξη. Μη-ισορροπημένα σύνολα δεδομένων μπορεί να βρεθούν σε πολλούς διαφορετικούς τομείς, όπως η ιατρική διάγνωση, το φιλτράρισμα ανεπιθύμητων μηνυμάτων και ο εντοπισμός οικονομικής απάτης.

Οι περισσότεροι αλγόριθμοι μηχανικής μάθησης λειτουργούν καλύτερα όταν ο αριθμός των δειγμάτων σε κάθε κλάση είναι περίπου ίσος. Αυτό συμβαίνει επειδή οι περισσότεροι αλγόριθμοι έχουν σχεδιαστεί με σκοπό τη μεγιστοποίηση της ακρίβειας και την ελαχιστοποίηση του σφάλματος. Ωστόσο σε περιπτώσεις μη-ισορροπημένων δεδομένων, η ακρίβεια δεν είναι η καλύτερη μετρική που μπορεί να χρησιμοποιηθεί για την αξιολόγηση του μοντέλου, καθώς μπορεί να είναι πολύ παραπλανητική. Οι μετρικές που μπορούν να παρέχουν καλύτερη εικόνα περιλαμβάνουν:

- **Confusion Matrix:** ένας πίνακας που μετράει τον αριθμό των σωστών προβλέψεων και τα είδη των λανθασμένων προβλέψεων.
- **Πιστότητα (precision):** ο αριθμός των σωστών θετικών προβλέψεων διαιρεμένος με τον αριθμό όλων των θετικών προβλέψεων. Η μετρική αυτή ονομάζεται επίσης positive predictive value και είναι ένα μέτρο της ακρίβειας του ταξινομητή. Η χαμηλή πιστότητα υποδηλώνει μεγάλο αριθμό διαψευσμένων θετικών προβλέψεων.
- **Ευαισθησία (recall ή sensitivity ή true positive rate):** ο αριθμός των σωστών θετικών προβλέψεων διαιρεμένος με τον αριθμό των δεδομένων που ανήκουν στη θετική κλάση. Η χαμηλή ευαισθησία υποδηλώνει μεγάλο αριθμό διαψευσμένων αρνητικών προβλέψεων.
- **F1-Score:** ο σταθμισμένος μέσος όρος του precision και του recall.

Μερικές συνήθεις μέθοδοι για την αντιμετώπιση των μη-ισορροπημένων δεδομένων είναι οι παρακάτω:

- **Αλλαγή αλγορίθμου μηχανικής μάθησης**

Ενώ σε κάθε πρόβλημα μηχανικής μάθησης, είναι καλός κανόνας να εξετάζεται μια ποικιλία αλγορίθμων, όσον αφορά στα μη-ισορροπημένα σύνολα δεδομένων, αυτό μπορεί να αποβεί ιδιαίτερα ωφέλιμο. Τα δέντρα αποφάσεων συχνά παρουσιάζουν καλή απόδοση σε μη-ισορροπημένα δεδομένα.

- **Υπερδειγματοληψία στην κλάση μειοψηφίας**

Η υπερδειγματοληψία μπορεί να οριστεί ως η προσθήκη περισσότερων αντιγράφων προερχόμενων από τα δεδομένα της κλάσης μειοψηφίας. Αυτή η τεχνική μπορεί να είναι μια καλή επιλογή όταν δεν υπάρχουν πολλά δεδομένα για εργασία.

- **Υποδειγματοληψία στην κλάση πλειοψηφίας**

Η υποδειγματοληψία μπορεί να οριστεί ως η αφαίρεση ορισμένων δεδομένων της πλειοψηφικής τάξης. Η υποδειγματοληψία μπορεί να είναι μια καλή επιλογή όταν υπάρχει υπερπληθώρα δεδομένων προς εργασία. Ωστόσο ένα μειονέκτημα της μεθόδου είναι ότι αφαιρούνται πληροφορίες που μπορεί να είναι πολύτιμες. Αυτό θα μπορούσε να οδηγήσει σε underfitting και φτωχά αποτελέσματα στο test set.

- **Δημιουργία τεχνητών δεδομένων**

Μια τεχνική παρόμοια με τη δειγματοληψία είναι η δημιουργία τεχνητών δεδομένων. Αυτή επιτυγχάνεται πρακτικά με την τεχνική SMOTE ή Synthetic Minority Oversampling. Η SMOTE χρησιμοποιεί έναν αλγόριθμο πλησιέστερων γειτόνων για τη δημιουργία νέων, τεχνητών δεδομένων που μπορούν να χρησιμοποιηθούν για την αποτελεσματικότερη εκπαίδευση του μοντέλου.

- **Συνδυασμός υπερδειγματοληψίας και υποδειγματοληψίας**

Η υπερδειγματοληψία μπορεί να οδηγήσει σε overfitting, το οποίο μπορεί να αποφευχθεί με την εφαρμογή μεθόδων "καθαρισμού" μέσω υπο-δειγματοληψίας.

- **Μέθοδοι Συλλογικής Μάθησης (Ensemble Learning methods)**

Οι μέθοδοι υποδειγματοληψίας συνεπάγονται ότι ορισμένα δείγματα της κλάσης-πλειοψηφίας χάνονται κατά τη διαδικασία εξισορρόπησης. Οι συνδυαστικές μέθοδοι προσφέρουν μια εναλλακτική λύση με σκοπό τη χρήση των περισσότερων δειγμάτων: δημιουργία ισορροπημένων συνόλων δεδομένων και χρήση αυτών για την εκπαίδευση οποιουδήποτε ταξινομητή[36][37].

Κεφάλαιο 3

Μηχανική μάθηση και ερμηνευσιμότητα - Η χρήση της για την ανάπτυξη ερμηνεύσιμων μοντέλων στον χώρο της υγείας

Η ταχεία και ακριβής αναγνώριση των νοσοκομειακών ασθενών που διατρέχουν υψηλό κίνδυνο για επανανοσηλεία, ασθένεια, παρατεταμένη διάρκεια παραμονής στο νοσοκομείο, θνητότητα κλπ., έχει τη δυνατότητα να βελτιώσει την ποιότητα της περίθαλψης και να μειώσει τις δυσχέρειες και το κόστος. Ωστόσο, οι περισσότερες μελέτες για την πρόβλεψη κινδύνου νοσηλείας έχουν παράγει μη-ερμηνεύσιμα μαύρα κουτιά, γεγονός που εμποδίζει την αποτελεσματική τους χρήση από τα συστήματα υποστήριξης νοσοκομειακών αποφάσεων. Συνπώς η δημιουργία ερμηνεύσιμων μοντέλων μηχανικής μάθησης στο πλαίσιο της υγειονομικής περίθαλψης και ειδικά για το έργο της πρόβλεψης του κινδύνου επανανοσηλείας αποτελεί άμεση ανάγκη[38].

3.1 Ταξινόμηση των μεθόδων ερμηνευσιμότητας

Οι μέθοδοι ερμηνείας της μηχανικής μάθησης μπορούν να ταξινομηθούν σύμφωνα με διάφορα κριτήρια που παρουσιάζονται παρακάτω [39]:

- Εγγενής και εκ των υστέρων ερμηνεία

Αυτό το κριτήριο διακρίνει εάν η ερμηνευσιμότητα επιτυγχάνεται περιορίζοντας την πολυπλοκότητα του μοντέλου μηχανικής μάθησης (εγγενής) ή εφαρμόζοντας μεθόδους που αναλύουν το μοντέλο μετά την εκπαίδευση (εκ των υστέρων). Η εγγενής ερμηνευσιμότητα αναφέρεται σε μοντέλα μηχανικής εκμάθησης που θεωρούνται ερμηνεύσιμα λόγω της απλής δομής τους, όπως δέντρα βραχείας απόφασης ή αραιά γραμμικά μοντέλα. Η εκ των υστέρων ερμηνεία αναφέρεται στην εφαρμογή μεθόδων ερμηνευσιμότητας μετά την εκπαίδευση του μοντέλου. Οι μέθοδοι εκ των υστέρων ερμηνείας μπορούν επίσης να εφαρμοστούν σε εγγενώς ερμηνεύσιμα μοντέλα.

- Αποτέλεσμα της μεθόδου ερμηνευσιμότητας

Οι διάφορες μέθοδοι ερμηνευσιμότητας μπορούν να διαφοροποιηθούν ανάλογα με τα αποτελέσματά τους.

- Στατιστικά περιγραφής χαρακτηριστικών: Πολλές μέθοδοι ερμηνείας παρέχουν συνοπτικά στατιστικά στοιχεία για κάθε χαρακτηριστικό. Ορισμένες μέθοδοι επιστρέφουν έναν μεμονωμένο αριθμό ανά χαρακτηριστικό, όπως η βαρύτητα του χαρακτηριστικού ή ένα πιο περίπλοκο αποτέλεσμα, όπως τη βαρύτητα της αλληλεπίδρασης δύο χαρακτηριστικών, η οποία αποτελείται από έναν αριθμό για κάθε ζεύγος χαρακτηριστικών.
- Οπτικοποίηση περιγραφής χαρακτηριστικών: Τα περισσότερα στατιστικά στοιχεία ανά χαρακτηριστικό μπορούν επίσης να απεικονιστούν. Ορισμένες περιγραφές χαρακτηριστικών έχουν ουσιαστικά νόημα μόνο εάν οπτικοποιηθούν και ένας πίνακας στατιστικών στοιχείων θα ήταν λανθασμένη επιλογή.

- Εσωτερικές παράμετροι μοντέλου (π.χ. βάρη): Η ερμηνεία των εγγενώς ερμηνεύσιμων μοντέλων εμπίπτει σε αυτήν την κατηγορία. Παραδείγματα είναι τα βάρη σε γραμμικά μοντέλα ή η ερμηνεία μέσω της δομής ενός δέντρου (τα χαρακτηριστικά και τα κατώτατα όρια που χρησιμοποιούνται για τα διαχωριστικά στον αλγόριθμο) των δέντρων απόφασης. Οι γραμμές είναι θολές όσον αφορά στην κατηγοριοποίηση των αποτελεσμάτων της μεθόδου ερμηνευσιμότητας μεταξύ εσωτερικών παραμέτρων των μοντέλων και των στατιστικών περιγραφής χαρακτηριστικών. Παραδείγματος χάρη σε γραμμικά μοντέλα, τα βάρη είναι ταυτόχρονα εσωτερικά στοιχεία του μοντέλου και στατιστικά στοιχεία που αφορούν τα χαρακτηριστικά.
 - Σημείο δεδομένων: Αυτή η κατηγορία περιλαμβάνει όλες τις μεθόδους που επιστρέφουν σημεία δεδομένων (υπάρχουν ήδη ή έχουν δημιουργηθεί πρόσφατα) για να κάνουν ένα μοντέλο ερμηνεύσιμο. Μια μέθοδος ονομάζεται ερμηνεία με αντιπαράδειγμα και περιγράφει τη μικρότερη αλλαγή στις τιμές των χαρακτηριστικών που αλλάζει την πρόβλεψη σε μια διαφορετική, προκαθορισμένη έξοδο. Για να είναι χρήσιμες, οι μέθοδοι ερμηνείας που εξάγουν νέα σημεία δεδομένων απαιτούν τα ίδια τα σημεία να είναι ερμηνεύσιμα. Αυτό λειτουργεί καλά για δεδομένα εικόνων και κειμένων, αλλά είναι λιγότερο χρήσιμο για δεδομένα πίνακα. .
 - Εγγενώς ερμηνεύσιμα μοντέλα: Μία λύση για την ερμηνεία μοντέλων μαύρου κουτιού είναι η προσέγγιση τους (είτε ολικά είτε τοπικά) με ένα ερμηνεύσιμο μοντέλο. Το ίδιο το ερμηνεύσιμο μοντέλο ερμηνεύεται εξετάζοντας τις εσωτερικές παραμέτρους του μοντέλου ή τα στατιστικά στοιχεία περιγραφής χαρακτηριστικών.
- Ερμηνεία ανεξαρτήτως μοντέλου ή για συγκεκριμένο μοντέλο

Τα εργαλεία ερμηνείας για συγκεκριμένο μοντέλο περιορίζονται σε συγκεκριμένες κατηγορίες μοντέλων. Η ερμηνεία των βαρών παλινδρόμησης σε ένα γραμμικό μοντέλο είναι μια ερμηνεία συγκεκριμένη για αυτό το μοντέλο, αφού - εξ ορισμού - η ερμηνεία των εγγενώς ερμηνεύσιμων μοντέλων είναι πάντα συγκεκριμένη για το μοντέλο. Τα εργαλεία ερμηνείας ανεξαρτήτως μοντέλου μπορούν να χρησιμοποιηθούν σε οποιοδήποτε μοντέλο μηχανικής μάθησης και εφαρμόζονται μετά την εκπαίδευση του μοντέλου. Αυτές οι αγνωστικές μέθοδοι λειτουργούν συνήθως αναλύοντας ζεύγη εισόδου και εξόδου. Εξ ορισμού, αυτές οι μέθοδοι δεν μπορούν να έχουν πρόσβαση σε εσωτερικές παραμέτρους του μοντέλου όπως βάρη ή δομικές πληροφορίες.
 - Ολική ή τοπική ερμηνεία

Η μέθοδος ερμηνείας μπορεί να εξηγήσει μια μεμονωμένη πρόβλεψη ή ολόκληρη τη συμπεριφορά του μοντέλου και για αυτό διαφοροποιείται σε τοπική και ολική ερμηνεία αντίστοιχα.

3.2 Μέθοδοι ερμηνευσιμότητας

3.2.1 Σημαντικότητα των Χαρακτηριστικών

Η μέθοδος εύρεσης της σημαντικότητας των χαρακτηριστικών μετρά το μέγεθος της αύξησης του σφάλματος πρόβλεψης του μοντέλου μετά την τροποποίηση των τιμών ενός χαρακτηριστικού. Μετράται ουσιαστικά η σημασία ενός χαρακτηριστικού υπολογίζοντας την αύξηση του σφάλματος πρόβλεψης του μοντέλου μετά την τροποποίηση της τιμής του. Ένα χαρακτηριστικό είναι "σημαντικό" αν η τροποποίηση των τιμών του αυξάνει το σφάλμα του μοντέλου, επειδή σε αυτήν την περίπτωση το μοντέλο βασίστηκε σε αυτό για να καθορίσει την πρόβλεψη. Ένα χαρακτηριστικό είναι "ασημαντο" εάν τροποποιώντας τις τιμές του, το σφάλμα του μοντέλου παραμένει αμετάβλητο, επειδή στην περίπτωση αυτή το μοντέλο αγνόησε το χαρακτηριστικό αυτό για την πρόβλεψη. Η μέτρηση της σημαντικότητας ενός χαρακτηριστικού μέσω τροποποιήσεων εισήχθη από τον Breiman (2001) εφαρμοσμένη πάνω στο μοντέλο δάσους τυχαίας απόφασης (Random Forest). Με βάση αυτήν την

ιδέα, οι Fisher, Rudin και Dominici (2018) πρότειναν μια ανεξαρτήτως-μοντέλου έκδοση της μεθόδου σημαντικότητας χαρακτηριστικών και την ονόμασαν "μοντέλο εμπιστοσύνης" [39].

Συγκεκριμένα μία μελέτη του τμήματος Υπολογιστικής Βιολογίας και Εφαρμοσμένων Αλγορίθμων του Ινστιτούτου Max Planck εισάγει μια μέθοδο που βασίζεται σε επαναλαμβανόμενες τροποποιήσεις του διανύσματος αποτελεσμάτων για την εκτίμηση της κατανομής της μετρούμενης σημαντικότητας για κάθε χαρακτηριστικό. Η τιμή P της παρατηρούμενης σημαντικότητας παρέχει ένα διορθωμένο μέτρο της σημαντικότητας των χαρακτηριστικών. Εφαρμόζοντας τη μέθοδο σε προσομοιωμένα δεδομένα αποδεικνύεται ότι (1) τα μη-σημαντικά χαρακτηριστικά (όσον αφορά στην προβλεπτική τους συμβολή) δεν λαμβάνουν σημαντικές τιμές P, (2) τα σημαντικά χαρακτηριστικά μπορούν να διακριθούν και να διαφοροποιηθούν επιτυχώς σε σχέση με τα μη-σημαντικά και (3) οι τιμές P που υπολογίστηκαν με τη μέθοδο "σημαντικότητας χαρακτηριστικών με τροποποιήσεις" (permutation importance - PIMP) είναι καθοριστικές για την εύρεση της σημαντικότητας των μεταβλητών και, ως εκ τούτου, βελτιώνουν την ερμηνευσιμότητα του μοντέλου. Επιπλέον, το PIMP χρησιμοποιήθηκε για τη διόρθωση μετρήσεων σημαντικότητας χαρακτηριστικών σε μοντέλο δάσους τυχαίας απόφασης πάνω σε δύο υπαρκτά προβλήματα [40].

3.2.2 Τοπική μέθοδος ερμηνευσιμότητας ανεξαρτήτως μοντέλου (LIME)

Η μέθοδος LIME (Local interpretable model-agnostic explanations) χρησιμοποιεί τοπικά υποκατάστατα μοντέλα τα οποία είναι ερμηνεύσιμα και αξιοποιούνται για να εξηγήσουν μεμονωμένες προβλέψεις μοντέλων μηχανικής μάθησης. Τα υποκατάστατα μοντέλα εκπαιδεύονται για να προσεγγίσουν τις προβλέψεις του υποκείμενου μοντέλου μαύρου κουτιού. Αντί να εκπαιδεύσει ένα συνολικό υποκατάστατο μοντέλο, το LIME εστιάζει στην εκπαίδευση τοπικών υποκατάστατων μοντέλων για να εξηγήσει μεμονωμένες προβλέψεις.

Η ιδέα είναι αρκετά διαισθητική. Αρχικά, αγνοεί τα δεδομένα εκπαίδευσης και χρησιμοποιεί μόνο το μοντέλο του μαύρου κουτιού στο οποίο μπορούν να εισαχθούν σημεία δεδομένων και να ληφθούν οι προβλέψεις του μοντέλου. Ο στόχος είναι να κατανοηθεί γιατί το μοντέλο μηχανικής μάθησης έκανε μια συγκεκριμένη πρόβλεψη. Το LIME δημιουργεί ένα νέο σύνολο δεδομένων που αποτελείται από διαταραγμένα δείγματα και τις αντίστοιχες προβλέψεις του μοντέλου μαύρου κουτιού για αυτά. Σε αυτό το νέο σύνολο δεδομένων, το LIME εκπαιδεύει στη συνέχεια ένα ερμηνεύσιμο μοντέλο, το οποίο σταθμίζεται από την εγγύτητα των νέων δειγμάτων στο δείγμα ενδιαφέροντος. Το ερμηνεύσιμο μοντέλο μπορεί να είναι για παράδειγμα ο "λιγότερο απόλυτος χειριστής συρρίκνωσης και επιλογής" (least absolute shrinkage and selection operator - LASSO) ή ένα δέντρο αποφάσεων. Το εκπαιδευμένο ερμηνεύσιμο μοντέλο πρέπει να αποτελεί μια καλή προσέγγιση των προβλέψεων του μοντέλου μηχανικής μάθησης τοπικά, αλλά όχι απαραίτητα μια καλή ολική προσέγγιση [39].

Η ερμηνευσιμότητα των μοντέλων μηχανικής μάθησης είναι κρίσιμη για τις προσπάθειες του τομέα της ιατρικής ακριβείας. Η ιατρική ακριβείας έχει μεγάλο μέλλον στην υγειονομική περίθαλψη καθώς προσαρμόζει την ιατρική περίθαλψη στη μοναδική κατάσταση της ασθένειας ενός ατόμου. Ωστόσο, τα πολύ αποδοτικά μοντέλα ως προς την προγνωστική τους ικανότητα είναι γενικά περίπλοκα και είναι δύσκολο να ερμηνευθούν. Στην παρακάτω μελέτη λοιπόν χρησιμοποιήθηκε ο αλγόριθμος LIME, παρουσιάζοντας πώς πολύπλοκα μοντέλα όπως το Random Forest μπορούν να γίνουν ερμηνεύσιμα. Χρησιμοποιώντας το σύνολο δεδομένων της έξυπνης παρακολούθησης πολλαπλών παραμέτρων της μονάδας εντατικής φροντίδας, προβλέφθηκε επιτυχώς η θνησιμότητα στην μονάδα εντατικής θεραπείας με ακρίβεια 80% και επιτεύχθηκε επίσης η ερμηνεία της σχετικής επίδρασης των χαρακτηριστικών στην πρόβλεψη σε ατομικό επίπεδο [41].

Σε διαφορετική μελέτη στόχος είναι η ανάπτυξη ενός συστήματος υποστήριξης αποφάσεων ακτινολογίας για την ταξινόμηση των μαγνητικών τομογραφιών. Προς αυτή την κατεύθυνση αναπτύχθηκαν διαφορετικά προγνωστικά μοντέλα μηχανικής μάθησης: γραμμικής Παλινδρόμησης, μηχανές διανυσμάτων υποστήριξης, μοντέλο δάσους τυχαίας απόφασης και XGBoost, για να αυτοματοποιηθεί η διαδικασία ταξινόμησης και να αντληθούν πληροφορίες που είναι χρήσιμες για τη μελλοντική λήψη κλινικών αποφάσεων και τη βέλτιστη αξιοποίηση των πόρων. Στην συνέχεια πραγματοποιήθηκε περαιτέρω διερεύνηση της ερμηνείας των αλγορίθμων μαύρου κουτιού χρησιμοποιώντας το μοντέλο

LIME για την άντληση επιπλέον πληροφοριών που αφορούσαν σε αποφάσεις που εξήχθηκαν χρησιμοποιώντας ένα μοντέλο XGBoost, με την ερμηνεία να αφορά μεμονωμένους ασθενείς. Ο αλγόριθμος LIME είναι ένα σημαντικό πρώτο βήμα προς την ανάπτυξη ενός ολοκληρωμένου συστήματος υποστήριξης αποφάσεων σε επίπεδο ασθενούς για την ταξινόμηση των μαγνητικών τομογραφιών [42].

Η επόμενη μελέτη ασχολείται με τα γλοιώματα τα οποία είναι οι πιο συνήθεις πρωτογενείς κακοήθεις όγκοι του εγκεφάλου σε ενήλικες. Η ακριβής βαθμολόγηση της κατάστασης του ασθενούς είναι ζωτικής σημασίας καθώς οι θεραπευτικές στρατηγικές είναι συχνά διαφορετικές για διαφορετικούς βαθμούς και μπορεί να επηρεάσουν την προβλεπόμενη κατάσταση των ασθενών. Η μελέτη στοχεύει στην παροχή μιας αυτοματοποιημένης πλατφόρμας βαθμολόγησης γλοιώματος βασισμένη στα μοντέλα ηχανικής μάθησης. Η πλατφόρμα ταξινομεί τα γλοιώματα σε βαθμούς II, III και IV. Επιπλέον, ερμηνεύονται ποσοτικά και αναδεικνύονται οι σημαντικές παραμέτρους που συνέβαλαν στην βαθμολόγηση χρησιμοποιώντας τον αλγόριθμο LIME. Η ποσοτική ανάλυση και εξήγηση μπορεί να βοηθήσει τους κλινικούς ιατρούς να κατανοήσουν καλύτερα την ασθένεια και, κατά συνέπεια, να επιλέξουν βέλτιστες θεραπείες για τη βελτίωση των κλινικών αποτελεσμάτων [43].

3.2.3 Μέθοδος ερμηνευσιμότητας Άγκυρες (Anchors)

Η μέθοδος Anchors εξηγεί μεμονωμένες προβλέψεις οποιουδήποτε μοντέλου ταξινόμησης μαύρου κουτιού βρίσκοντας έναν κανόνα απόφασης που "αγκυρώνει" την πρόβλεψη επαρκώς. Ένας κανόνας αγκυρώνει μια πρόβλεψη με βάση τις τιμές των χαρακτηριστικών των παρατηρήσεων. Έτσι εάν οι τιμές των χαρακτηριστικών μιας παρατήρησης ικανοποιούν όλες τις προϋποθέσεις του κανόνα, τότε υπάρχει πολύ μεγάλη πιθανότητα η πρόβλεψη να ταιριάζει με την πρόβλεψη του μοντέλου. Σε αυτή την περίπτωση οι αλλαγές σε άλλες τιμές χαρακτηριστικών δεν επηρεάζουν την πρόβλεψη. Οι Ribeiro, Singh και Guestrin πρότειναν τον αλγόριθμο το 2018 - οι ίδιοι ερευνητές που εισήγαγαν και τον αλγόριθμο LIME.

Όπως και ο προκάτοχός της, η προσέγγιση αγκυρών αναπτύσσει μια στρατηγική βασισμένη σε διαταραχές-τροποποιήσεις του δείγματος προς ερμηνεία για τη δημιουργία τοπικών εξηγήσεων για τις προβλέψεις των μοντέλων μηχανικής μάθησης μαύρου κουτιού. Ωστόσο, αντί των υποκατάστατων μοντέλων που χρησιμοποιούνται από το LIME, οι εξηγήσεις που προκύπτουν εκφράζονται ως εύχρηστοι κανόνες IF-THEN, που ονομάζονται άγκυρες. Κάτι τέτοιο επιτρέπει στην προσέγγιση να αγνοεί τη δομή του μαύρου κουτιού και τις εσωτερικές του παράμετρους, έτσι ώστε αυτές να μπορούν να παραμείνουν ανεπιτήρητες και αμετάβλητες. Έτσι, ο αλγόριθμος είναι ανεξάρτητος μοντέλου και κατ' επέκταση μπορεί να εφαρμοστεί σε οποιαδήποτε κατηγορία μοντέλου [39].

3.2.4 Ερμηνευσιμότητα μέσω των τιμών Shapley (SHAP)

Ο στόχος του SHAP είναι να εξηγήσει την πρόβλεψη ενός στιγμιότυπου υπολογίζοντας τη συμβολή κάθε χαρακτηριστικού στην πρόβλεψη. Η μέθοδος ερμηνευσιμότητας SHAP υπολογίζει τις τιμές Shapley από τη θεωρία παιγνίων συνασπισμού. Οι τιμές των χαρακτηριστικών ενός στιγμιότυπου ενεργούν ως παίκτες σε έναν συνασπισμό. Οι τιμές Shapley μας λένε πώς να κατανεύουμε δίκαια την "αμοιβή" (payout) (στην περίπτωση μας την τιμή της πρόβλεψης) μεταξύ των χαρακτηριστικών. Ένας παίκτης μπορεί να θεωρηθεί ένα μεμονωμένο χαρακτηριστικό του στιγμιότυπου, παραδείγματος χάρι στα δεδομένα πίνακα. Εναλλακτικά μπορεί να θεωρηθεί και μια ομάδα χαρακτηριστικών. Για παράδειγμα, για να ερμηνευθεί μια εικόνα, τα pixel μπορούν να ομαδοποιηθούν σε super-pixel(παίκτες) και η πρόβλεψη να κατανέμεται μεταξύ τους [39].

Παρακάτω παρατίθενται επιστημονικές μελέτες στον τομέα της υγείας στις οποίες η μέθοδος SHAP αναδείχθηκε:

Αυτή η μελέτη στοχεύει στην ανάπτυξη ενός ερμηνεύσιμου και γενικευμένου μοντέλου πρόβλεψης πρόωρης θνησιμότητας σε ηλικιωμένους ασθενείς με σύνδρομο πολλαπλής δυσλειτουργίας οργάνων (ΣΠΔΟ), οι οποίοι διατρέχουν υψηλό κίνδυνο θανάτου. Η απόδοση των σημερινών συστημάτων βαθμολογίας που αξιολογούν τη σοβαρότητα του ΣΠΔΟ ανά ασθενή και την πιθανότητα θνησιμότητας του παραμένει μη ικανοποιητική. Έτσι, χρησιμοποιήθηκε το μοντέλο XGBoost σε συνδιασμό

με τη μέθοδο SHAP για τη διεξαγωγή πρώιμων και ερμηνεύσιμων προβλέψεων ως προς το αποτέλεσμα της νοσηλείας των ασθενών. Το ερμηνεύσιμο μοντέλο με τη βέλτιστη απόδοση που αναπτύχθηκε επικυρώθηκε σε ξεχωριστά σύνολα δεδομένων και οι επιδόσεις του μοντέλου στην πρόβλεψη της θνησιμότητας εντός του νοσοκομείου όπως επικυρώθηκε από τα τρία σύνολα δεδομένων ήταν: Πρώτο σύνολο: AUC 0,885, ευαισθησία 0,834 και ειδικότητα 0,705 Δεύτερο σύνολο: AUC 0,849, ευαισθησία 0,763 και ειδικότητα 0,784 Τρίτο σύνολο: AUC 0,838, ευαισθησία 0,888 και ειδικότητα 0,669. Οι συγκρίσεις της AUC μεταξύ αυτού του μοντέλου και των βασικών μοντέλων υπέδειξαν ανώτερες επιδόσεις αυτού του μοντέλου. Το ερμηνεύσιμο μοντέλο μηχανικής μάθησης που αναπτύχθηκε σε αυτήν τη μελέτη ήταν ανθεκτικό και γενικευμένο. Αυτό το μοντέλο ξεπέρασε τα βασικά μοντέλα και αρκετές κλινικές βαθμολογίες για πρόωρη πρόβλεψη θνησιμότητας σε ηλικιωμένους ασθενείς στη Μονάδα Εντατικής Θεραπείας. Η ερμηνευτική φύση αυτού του μοντέλου παρείχε στους κλινικούς ιατρούς μια αποτελεσματική κατάταξη των χαρακτηριστικών κινδύνου θνησιμότητας [44].

Ο στόχος της παρακάτω μελέτης είναι να αναπτύξει και να αξιολογήσει ένα ερμηνεύσιμο εξατομικευμένο μοντέλο πρόβλεψης κινδύνου για θανατηφόρα ή μη θανατηφόρα συχνότητα εμφάνισης καρδιαγγειακής νόσου σε άτομα με σακχαρώδη διαβήτη τύπου 2 (ΣΔΤ2). Η καρδιαγγειακή νόσος είναι μια σημαντική αιτία αναπηρίας και θανάτου μεταξύ των ατόμων με σακχαρώδη διαβήτη (ΣΔ). Οι διεθνείς κλινικές οδηγίες για τη διαχείριση του ΣΔΤ2 βασίζονται στην πρωτογενή και δευτερογενή πρόληψη και ευνοούν την αξιολόγηση των παραγόντων κινδύνου που σχετίζονται με την καρδιαγγειακή νόσο προς την κατάλληλη έναρξη της θεραπείας. Η ενσωμάτωση μεθόδων ερμηνευσιμότητας στα μοντέλα πρόβλεψης μπορεί να ενισχύσει την κατανόηση του ανθρώπου, να μεγιστοποιήσει τη διαφάνεια και να εμπλουτίσει την εμπιστοσύνη στην υιοθέτηση των μοντέλων στην κλινική πρακτική. Στην παρούσα μελέτη αναπτύσσεται μια ερμηνεύσιμη προσέγγιση βασισμένη στο μοντέλο του XGBoost σε συνδιασμό με το Tree SHAP για τον υπολογισμό του 5ετούς κινδύνου εμφάνισης καρδιαγγειακής νόσου και τη δημιουργία μεμονωμένων επεξηγήσεων σχετικά με τις αποφάσεις του μοντέλου. Τα αποτελέσματα που προέκυψαν (AUC = 71,13%) υποδηλώνουν τις δυνατότητες της προτεινόμενης προσέγγισης παρέχοντας παράλληλα κλινικά σημαντικές πληροφορίες σχετικά με τη διαδικασία λήψης αποφάσεων του μοντέλου [45].

Κεφάλαιο 4

Ανάπτυξη ερμηνεύσιμου μοντέλου εκτίμησης του κινδύνου νοσηλείας και επανανοσηλείας σε άτομα με Σακχαρώδη Διαβήτη Τύπου 1

4.1 Σύνολο Δεδομένων

Τα δεδομένα προέρχονται από τη βάση "SWEET" και παραχωρήθηκαν από το Νοσοκομείο Παιδων "Η Αγία Σοφία". Σε αυτά περιλαμβάνονται 127 ασθενείς με ΣΔΤ1 και διατίθενται μετρήσεις των

Πίνακας 2: Χαρακτηριστικά

	Πρώτος Χρόνος		Δεύτερος Χρόνος	
	Μη-νοσηλεία N ₀₀ =106	Νοσηλεία N ₀₁ =21	Μη-νοσηλεία N ₁₀ =121	Νοσηλεία N ₁₁ =6
Ηλικία (χρόνια)	10.42 ± 3.65	9.77 ± 4.33	13.63 ± 3.99	11.24 ± 5.78
Διάρκεια διαβήτη (χρόνια)	3.17 ± 3.55	3.75 ± 2.72	3.94 ± 3.50	4.57 ± 2.91
Γλυκοζυλιωμένη αιμοσφαιρίνη (HbA _{1c}) (%)	7.13 ± 0.93	7.84 ± 0.98	7.23 ± 0.94	9.84 ± 2.04
Ύψος (Z-score)	0.35 ± 1.08	0.21 ± 0.86	0.32 ± 1.06	0.03 ± 0.77
Βάρος (Z-score)	0.56 ± 0.98	0.55 ± 1.02	0.54 ± 0.97	0.57 ± 1.05
Δείκτης μάζας σώματος (Z-score)	0.58 ± 0.92	0.61 ± 1.04	0.56 ± 0.89	0.78 ± 0.92
Συστολική αρτηριακή πίεση (Z-score)	0.29 ± 0.77	0.28 ± 0.97	0.21 ± 0.86	1.05 ± 1.44
Διαστολική αρτηριακή πίεση (Z-score)	0.24 ± 0.74	0.34 ± 1.02	0.19 ± 0.89	1.27 ± 0.71
Συνολική ημερήσια δόση ινσουλίνης (U/kg)	0.70 ± 0.28	0.79 ± 0.32	0.74 ± 0.28	0.97 ± 0.20
Προγευματική δόση ινσουλίνης/ Συνολική ημερήσια δόση ινσουλίνης	0.61 ± 0.14	0.65 ± 0.13	0.62 ± 0.12	0.62 ± 0.11
Δόση ινσουλίνης μακράς δράσης/ Συνολική ημερήσια δόση ινσουλίνης	0.39 ± 0.14	0.35 ± 0.13	0.38 ± 0.12	0.38 ± 0.11
Φύλο				
Αρσενικό	54 (51%)	13 (62%)	64 (53%)	3 (50%)
Θηλυκό	52 (49%)	8 (38%)	57 (47%)	3 (50%)

συγκεκριμένων χαρακτηριστικών τους από την πρώτη και τη δεύτερη επίσκεψη με απόσταση μεταξύ τους το πολύ δύο ετών. Το 16,5% των ασθενών (21 άτομα) νοσηλεύτηκαν εκτάκτως στο νοσοκομείο λόγω διαβητικής κετοξέωσης, ενώ το 4,72% (6 άτομα) επανανοσηλεύτηκαν. Για την ολοκληρωμένη και ακριβή διαμόρφωση των δεδομένων εισόδου του μοντέλου λήφθηκαν υπόψη κλινικές και σωματομετρικές μετρήσεις, εργαστηριακά δεδομένα και μετρήσεις σχετικές με τη θεραπεία του κάθε ασθενούς.

4.2 Εννοιολογικό Πλαίσιο

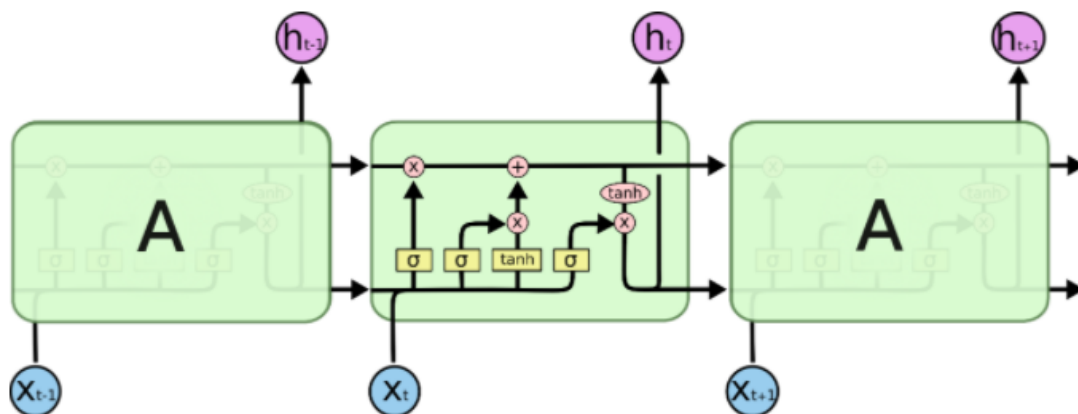
Η ανάπτυξη του μοντέλου βασίστηκε στα Νευρωνικά Δίκτυα Μακράς Βραχείας Μνήμης (Long Short-Term Memory Neural Networks- LSTM), τα οποία έχουν διακριθεί για την αποτελεσματικότητά τους στον τομέα της υγείας. Για τη διαχείριση της μη ισορροπημένης φύσης των δεδομένων, υιοθετήθηκε προσέγγιση συλλογικής μάθησης, σύμφωνα με την οποία κατασκευάστηκαν δύο μικρότερα υποσύνολα του αρχικού συνόλου δεδομένων και εκπαιδεύτηκαν δύο πρωταρχικά μοντέλα LSTM, των οποίων οι αποφάσεις συνδυάστηκαν με στόχο την εκτίμηση της τελικής πιθανότητας. Το συνδυαστικό μοντέλο εξήγαγε την πιθανότητα νοσηλείας εντός ενός έτους μετά την πρώτη επίσκεψη (χρονική στιγμή 1) και την πιθανότητα επανανοσηλείας εντός ενός έτους μετά τη δεύτερη επίσκεψη (χρονική στιγμή 2). Στη συνέχεια, για να ερμηνευθούν οι αποφάσεις του συνδυαστικού μοντέλου εφαρμόστηκε η μέθοδος ερμηνευσιμότητας LIME.

4.3 Μοντέλο βασισμένο στα Νευρωνικά Δίκτυα Μακράς Βραχείας Μνήμης

4.3.1 Ανάλυση μοντέλου

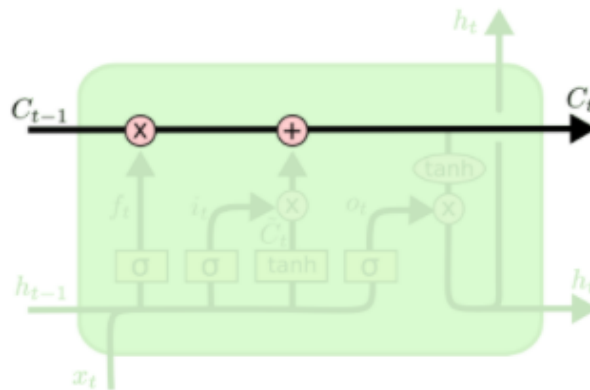
Τα νευρωνικά δίκτυα LSTM αποτελούν ένα ειδικό είδος αναδρομικών νευρωνικών δικτύων, ικανό να εκπαιδευτεί και να αναγνωρίζει μακροπρόθεσμες εξαρτήσεις. Εισηχθησαν από τους Hochreiter Schmidhuber (1997) και βελτιώθηκαν και διαδόθηκαν από πολλούς ανθρώπους που εργάστηκαν πάνω σε αυτά. Λειτουργούν εξαιρετικά καλά σε μια μεγάλη ποικιλία προβλημάτων και χρησιμοποιούνται πλέον ευρέως.

Τα LSTM έχουν σχεδιαστεί ρητά ώστε η απομνημόνευση πληροφοριών για μεγάλα χρονικά διαστήματα να είναι ουσιαστικά η προεπιλεγμένη συμπεριφορά τους. Όλα τα αναδρομικά νευρωνικά δίκτυα έχουν τη μορφή μιας αλυσίδας επαναλαμβανόμενων μονάδων νευρωνικού δικτύου. Η επαναλαμβανόμενη μονάδα των LSTM διαθέτει τέσσερα επίπεδα, που αλληλεπιδρούν με έναν πολύ ειδικό τρόπο όπως φαίνεται στο σχήμα 13.



Σχήμα 13: Η επαναλαμβανόμενη μονάδα σε ένα LSTM περιλαμβάνει τέσσερα επίπεδα

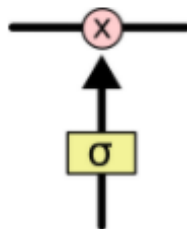
Το κλειδί για τα LSTM είναι η κατάσταση μνήμης, η οριζόντια γραμμή που διατρέχει το πάνω μέρος του διαγράμματος.



Σχήμα 14: Κατάσταση Μνήμης

Η κατάσταση μνήμης είναι σαν μια μεταφορική ταινία. Τρέχει κατά μήκος ολόκληρης της αλυσίδας, με μόνο μερικές μικρές γραμμικές τροποποιήσεις. Είναι πολύ εύκολο για τις πληροφορίες να ρέουν αμετάβλητα μέσω της κατάστασης μνήμης. Το LSTM έχει τη δυνατότητα να αφαιρεί ή να προσθέτει πληροφορίες στην κατάσταση μνήμης, μέσω προσεκτικών ρυθμίσεων που εκτελούνται από δομές που ονομάζονται πύλες.

Οι πύλες είναι ένας τρόπος για να διατηρούνται επιλεκτικά πληροφορίες. Αποτελούνται από ένα στρώμα νευρωνικού δικτύου (σύστημα βαρών) με συνάρτηση ενεργοποίησης τη σιγμοειδή (στρώμα σιγμοειδούς).

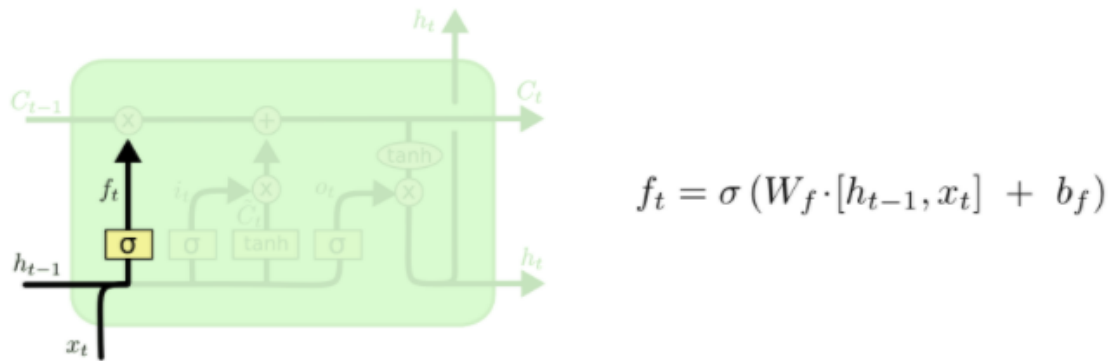


Σχήμα 15: Πύλη

Το στρώμα σιγμοειδούς εξάγει αριθμούς μεταξύ μηδέν και ένα, περιγράφοντας σε τι ποσοστό θα επιτραπεί στην τιμή της κάθε συνιστώσας να "περάσει". Η τιμή μηδέν σημαίνει ότι η συνιστώσα εξαλείφεται, ενώ η τιμή ένα σημαίνει ότι η συγκεκριμένη συνιστώσα "περνάει" αναλλοίωτη. Ένα LSTM χρησιμοποιεί τρεις από αυτές τις πύλες, για τον έλεγχο της κατάστασης μνήμης.

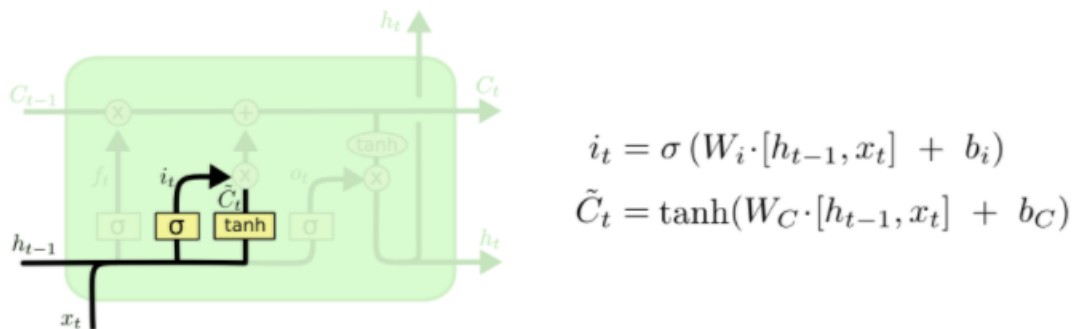
Το πρώτο βήμα για το LSTM είναι να αποφασίσει ποιες πληροφορίες πρόκειται να εξασθενήσουν από την κατάσταση μνήμης. Αυτή η απόφαση λαμβάνεται από ένα σιγμοειδές στρώμα νευρωνικού δικτύου που ονομάζεται «στρώμα πύλης αδιαφορίας» (forget gate layer). Πληροφορίες που είναι καθοριστικές για τις τιμές μελλοντικών προβλέψεων που όμως έφτασε η στιγμή που χρησιμοποιήθηκαν και ολοκληρώθηκε η εν λόγω πρόβλεψη, δεν είναι πλέον απαραίτητες οπότε διαγράφονται μέσω αυτής της πύλης. Η συγκεκριμένη πύλη υλοποιείται ως εξής: έχει ως είσοδο την ενοποίηση του διανύσματος εισόδου x_t και του διανύσματος κρυφής κατάστασης h_{t-1} το οποίο λαμβάνεται από το προηγούμενο χρονικό βήμα (timestep). Αφού ενοποιηθούν αυτά τα διανύσματα, το συνολικό διάνυσμα $[h_{t-1}, x_t]$ περνάει από το στρώμα νευρωνικού δικτύου της πύλης όπου πολλαπλασιάζεται με έναν πίνακα βαρών και εξάγεται ένα νέο διάνυσμα διαστάσεων ίσων με του C_{t-1} . Ύστερα κάθε στοιχείο αυτού του διανύσματος περνάει από τη σιγμοειδή συνάρτηση και έτσι όλα του τα στοιχεία μετατρέπονται σε

αριθμούς μεταξύ 0 και 1. Αυτό το τελικό διάνυσμα πολλαπλασιάζεται με το διάνυσμα C_{t-1} (στοιχείο με στοιχείο) με αποτέλεσμα να εξαλείφονται τα πλέον άχρηστα δεδομένα μνήμης από το C_{t-1} .



Σχήμα 16: Forget Gate

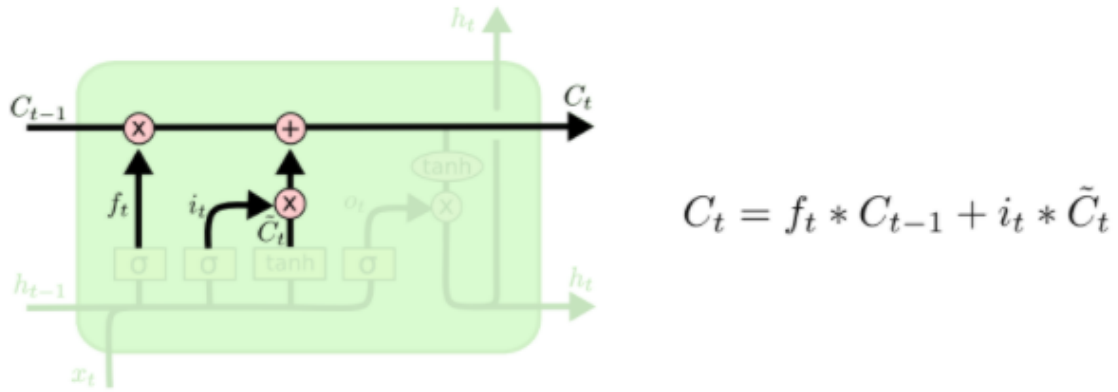
Το επόμενο βήμα είναι να αποφασιστεί ποιες νέες πληροφορίες πρόκειται να αποθηκευτούν στην κατάσταση μνήμης. Αυτό για να επιτευχθεί απαιτεί δύο μέρη. Πρώτον, ένα στρώμα νευρωνικού δικτύου υπερβολικής εφαπτομέτης (tanh) που δημιουργεί ένα διάνυσμα με τις νέες υποψήφιες τιμές, \tilde{C} , που θα μπορούσαν να προστεθούν στην παρούσα κατάσταση μνήμης και που αποτελούν ένα σύνολο επιλογών για την πρόβλεψη του LSTM για το εκάστοτε timestep. Αξίζει να τονιστεί ότι το συγκεκριμένο διάνυσμα \tilde{C} δεν περιέχει μόνο τις δυνατές επιλογές πρόβλεψης αλλά περιλαμβάνει και τις επιλογές που πρέπει να αποφευχθούν θέτοντας τους αρνητικό πρόσημο. Γι' αυτό σε αυτό το στρώμα είναι απαραίτητο να χρησιμοποιηθεί η tanh και όχι η σιγμοειδής. Το δεύτερο μέρος είναι ένα σιγμοειδές στρώμα που ονομάζεται "στρώμα πύλης εισόδου" (input gate layer) που φιλτράρει τις νέες τιμές \tilde{C} που παρήχθησαν ώστε να μην "περνάνε" όλες οι επιλογές για την τιμή της πρόβλεψης, ακόμη και οι λιγότερο εύστοχες, που λόγω πλήθους θα οδηγήσουν σε αποπροσανατολισμό του μοντέλου.



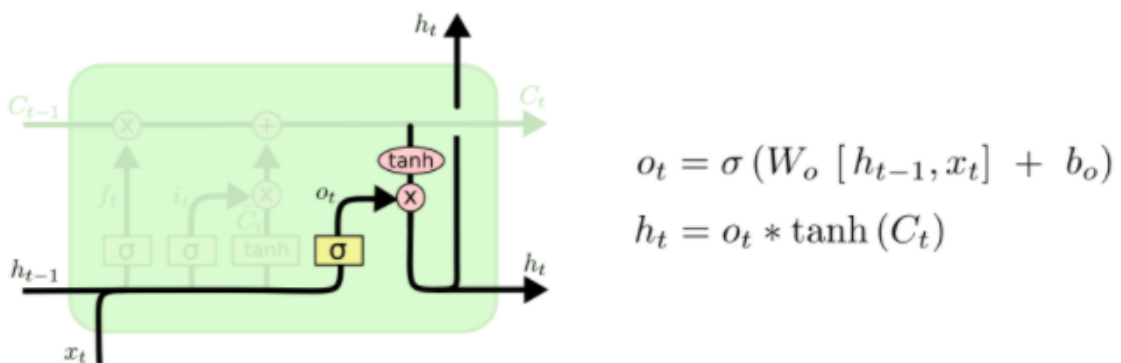
Σχήμα 17: Input Gate and tanh layer

Αφού στα προηγούμενα βήματα το forget gate και το input gate έχουν κάνει την κατάλληλη προετοιμασία, μπορεί πλέον να ενημερωθεί η παλιά κατάσταση μνήμης, C_{t-1} , στη νέα κατάσταση μνήμης C_t . Αρχικά, η παλιά κατάσταση έχει πολλαπλασιαστεί με f_t , αποδυναμώνοντας τα σημεία που αποφασίστηκε να ξεχαστούν. Στη συνέχεια προστίθεται το $i_t \times \tilde{C}_t$ που περιέχει τις φιλτραρισμένες νέες επιλογές πρόβλεψης. Σημειώνεται ότι μέσω της πρόσθεσης μπορούν επίσης να εξαλειφθούν προϋπάρχουσες μνήμες του C_{t-1} , καθώς εάν εμφανίζονται με θετικό πρόσημο μέσα στο διάνυσμα C_{t-1} και με αρνητικό πρόσημο στο διάνυσμα $i_t \times \tilde{C}_t$ τότε μέσω της πρόσθεσης θα αλληλοαναιρεθούν και δεν θα περάσει αυτή η πληροφορία μνήμης. Κατά τον παραπάνω τρόπο λοιπόν ενημερώνεται το διάνυσμα κατάστασης μνήμης C_t .

Τέλος, πρέπει να αποφασίσουμε τι θα εξάγουμε από τη συγκεκριμένη μονάδα ως πρόβλεψη στο συγκεκριμένο timestep. Αυτή η έξοδος θα βασίζεται στην κατάσταση μνήμης που έχουμε, αλλά θα



Σχήμα 18: Ενημέρωση κατάστασης μνήμης



Σχήμα 19: Ενημέρωση κρυφού επιπέδου

είναι μια φιλτραρισμένη εκδοχή της καθώς δεν θέλουμε όλες οι μνήμες που διατηρούμε από το παρελθόν (μέσω της κατάστασης μνήμης) να εξάγονται ως προβλεψεις στα επιμέρους timesteps. Έτσι αφού το διάνυσμα κατάστασης μνήμης C_t εφαρμοστεί σε μια \tanh (για να περιοριστούν οι τιμές του μεταξύ -1 και 1) πολλαπλασιάζεται με την έξοδο ενός σιγμοειδούς στρώματος νευρωνικού δικτύου που ονομάζεται "στρώμα πύλης εξόδου" (output gate layer). Αυτό το στρώμα λειτουργεί ως φίλτρο της κατάστασης μνήμης επιλέγοντας ποιές τιμές της θα εξαχθούν. Το αποτέλεσμα αυτών των πράξεων είναι το νέο διάνυσμα κρυμμένης κατάστασης h_t που αποτελεί και την τιμή της πρόβλεψης για το συγκεκριμένο timestep. [46] [47].

4.3.2 Υπερπαράμετροι

Οι υπερπαράμετροι που χρησιμοποιήθηκαν για τα πρωταρχικά μοντέλα LSTM φαίνονται στον πίνακα 3.

4.4 Εκπαίδευση πρωταρχικών μοντέλων

Για την αξιολογήση του μοντέλου χρησιμοποιήθηκε η μέθοδος της διασταυρωμένης επικύρωσης δέκα πτυχών (10 fold cross-validation). Σύμφωνα με την τεχνική αυτή, τα δεδομένα χωρίζονται σε 10 ισομερή υποσύνολα. Ύστερα κάθε υποσύνολο χρησιμοποιείται διαδοχικά για έλεγχο (test set) και τα υπόλοιπα αξιοποιούνται για την εκπαίδευση του μοντέλου (training set). Έτσι, μία αξιόπιστη εκτίμηση για την ακρίβεια του μοντέλου προκύπτει από τον μέσο όρο των δέκα εκτιμήσεων.

Στην παρούσα διπλωματική εργασία, τα επιμέρους υποσύνολα για την εφαρμογή της διασταυ-

Πίνακας 3: Υπερπαράμετροι του LSTM

Υπερπαράμετρος	Σημασία
units	θετικός ακέραιος που αντιπροσωπεύει το μέγεθος του διανύσματος κρυφής κατάστασης και κατ' επέκταση τις διαστάσεις της εξόδου
return_sequences	λογική μεταβλητή που εκφράζει το αν θα επιστραφεί μόνο η τελευταία έξοδος του τελευταίου timestep ή μια ακολουθία εξόδων που θα αντιπροσωπεύει την έξοδο σε κάθε timestep. Default: False

ρωμένης επικύρωσης κατασκευάστηκαν χωρίς την χρήση κάποιας έτοιμης συνάρτησης διαχωρισμού του αρχικού συνόλου. Καθώς ο αριθμός των υποσυνόλων ήταν δέκα, το 10% των δεδομένων σε κάθε επανάληψη χρησιμοποιήθηκε για έλεγχο και το 90% για την εκπαίδευση του μοντέλου.

Τα δείγματα διαχωρίζονταν στις εξής κατηγορίες:

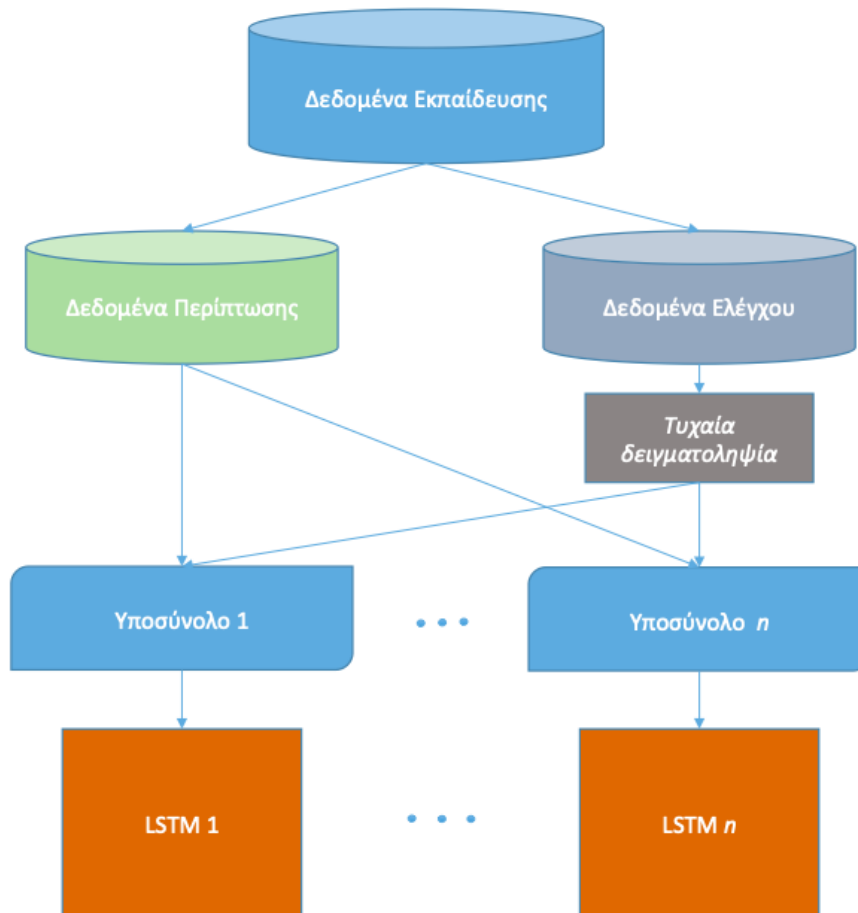
- Περίπτωση 1: Ο ασθενής νοσηλεύτηκε στην πρώτη χρονική στιγμή αλλά όχι στη δεύτερη (15 δείγματα)
- Περίπτωση 2: Ο ασθενής νοσηλεύτηκε και στις δύο χρονικές στιγμές (6 δείγματα)
- Περίπτωση 3: Ο ασθενής δεν νοσηλεύτηκε σε καμία εκ των δυο χρονικών στιγμών (106 δείγματα)

Τελικά τα πρώτα 4 υποσύνολα περιείχαν 2 δείγματα ασθενών της Περίπτωσης 1 και 11 δείγματα της Περίπτωσης 3, ενώ τα υπόλοιπα 6 υποσύνολα περιέχουν ένα δείγμα της Περίπτωσης 2, ένα δείγμα της Περίπτωσης 1 και 10 δείγματα της Περίπτωσης 3. Συνεπώς διατηρήθηκε μια σχετική ομοιογένεια στις κατανομές των επιμέρους υποσυνόλων.

4.4.1 Συνδυαστικό μοντέλο

Όπως έχει αναφερθεί και παραπάνω, το σύνολο δεδομένων είναι μη ισορροπημένο. Το γεγονός αυτό θα μπορούσε να αποτελέσει εμπόδιο στην επίτευξη υψηλής επίδοσης, καθώς γενικά οι περισσότεροι αλγόριθμοι μηχανικής μάθησης έχουν σχεδιαστεί θεωρώντας ίσοποση κατανομή της κάθε κλάσης στα δεδομένα. Λαμβάνοντας υπόψη τα παραπάνω, στο πλαίσιο της ανάπτυξης του μοντέλου αξιοποιήθηκαν συνδυαστικές τεχνικές διαχείρισης μη-ισορροπημένων συνόλων δεδομένων με βάση την υποδειγματοληψία. Σκοπός της τεχνικής είναι η δημιουργία υποσυνόλων δεδομένων με ίσο ποσοστό εμφάνισης δειγμάτων νοσηλευμένων και μη-νοσηλευμένων ασθενών. Κάθε τέτοιο υποσύνολο απαρτίζεται από όλο το πλήθος των νοσηλευμένων ασθενών (είτε νοσηλεύτηκαν μόνο την πρώτη χρονική στιγμή είτε και τις δύο) και ενός μέρους των δειγμάτων από μη-νοσηλευμένους ασθενείς, διαφορετικό σε κάθε υποσύνολο, όπως φαίνεται στο σχήμα 20. Το ποσοστό νοσηλευμένων και μη-νοσηλευμένων στο κάθε υποσύνολο επιλέχθηκε να είναι 29% και 71% αντιστοίχως καθώς αυτή η αναλογία απέδωσε τα καλύτερα αποτελέσματα. Εξετάστηκε ωστόσο και η αναλογία 50%-50% και 64% και 36%.

Επομένως προέκυψαν 2 ισορροπημένα υποσύνολα με 66 δείγματα το καθένα (19 νοσηλευμένοι και 47 μη-νοσηλευμένοι).



Σχήμα 20: Στάδιο εκπαίδευσης

Πίνακας 4: Τιμές Υπερπαραμέτρων που ελέγχθηκαν

Υπερπαραμέτρος	Τιμές
units	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
εποχές	10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200
αριθμός συστάδας	1, 2

Τα ισορροπημένα υποσύνολα αξιοποιήθηκαν για την εκπαίδευση δύο ανεξάρτητων μεταξύ τους LSTM μοντέλων. Κατά την εκπαίδευση των μοντέλων πραγματοποιήθηκε διερεύνηση των παραμέτρων των LSTM, προκειμένου να βρεθεί ο αποδοτικότερος συνδυασμός παραμέτρων. Συγκεκριμένα διερευνήθηκαν οι παράμετροι: αριθμός των κόμβων (units), αριθμός εποχών και αριθμός συστάδας (batch size). Το κριτήριο που χρησιμοποιήθηκε για την αξιολόγηση της απόδοσης του μοντέλου για τις διαφορετικές τιμές παραμέτρων ήταν η Area Under the ROC Curve (AUC). Οι τιμές που διερευνήθηκαν φαίνονται στον πίνακα 4.

Τα μοντέλα LSTM εκπαιδεύτηκαν με τις ίδιες τιμές παραμέτρων. Στον πίνακα 5, φαίνονται οι επιλεγμένες τιμές των παραμέτρων όπως διαμορφώθηκαν σύμφωνα με την βέλτιστη απόδοση των μοντέλων με βάση το κριτήριο AUC.

Συνοψίζοντας, για τις τελικές προβλέψεις ακολουθήθηκαν τα εξής βήματα:

Πίνακας 5: Επιλογή τελικών υπερπαραμέτρων

Υπερπαραμέτρος	Τελική Τιμή
units	30
εποχές	200
αριθμός συστάδας	1

- Το αρχικό σύνολο δεδομένων χωρίστηκε σε 10 υποσύνολα τα οποία διατήρησαν τις αναλογίες των κλάσεων του αρχικού συνόλου. Κάθε υποσύνολο τέθηκε με τη σειρά του ως σύνολο ελέγχου ενώ τα εναπομείναντα ενοποιήθηκαν σε ένα σύνολο εκπαίδευσης.
- Το σύνολο εκπαίδευσης δεδομένου ότι δεν ήταν ισορροπημένο (καθώς διατήρησε τις αρχικές αναλογίες) διαχωρίστηκε σε 2 ισορροπημένα σύνολα, όπως εξηγήθηκε παραπάνω.
- Κάθε ισορροπημένο υποσύνολο, τροφοδοτήθηκε σε ένα εκ των 2 μοντέλων LSTM (με κοινές υπερπαραμέτρους) ώστε να επιτευχθεί η εκπαίδευση δύο μοντέλων LSTM με ισορροπημένο σύνολο δεδομένων εισόδου.
- Αφού ολοκληρώθηκε η εκπαίδευση των πρωταρχικών μοντέλων, οι προβλέψεις τους για τον κίνδυνο νοσηλείας (στιγμιότυπο 1) και επανανοσηλείας (στιγμιότυπο 2) συνδυάστηκαν με χρήση του απλού μέσου όρου, όπως φαίνεται στο σχήμα 21. Εξετάστηκε και το συνδυαστικό σχήμα του σταθμισμένου μέσου όρου καθώς και το σχήμα ψηφοφορίας που όμως δεν απέδωσαν ικανοποιητικά αποτελέσματα.
- Στη συνέχεια αξιολογήθηκε η επίδοση του συνδυαστικού μοντέλου στο σύνολο ελέγχου μέσω πέντε διαφορετικών κριτηρίων (Ακρίβεια (accuracy), Ευαισθησία (sensitivity), Ειδικότητα (specificity), Area Under the ROC Curve (AUC) και Brier Score).
- Η διαδικασία εφαρμόστηκε για κάθε ένα από τα δέκα διαφορετικά σύνολα ελέγχου.

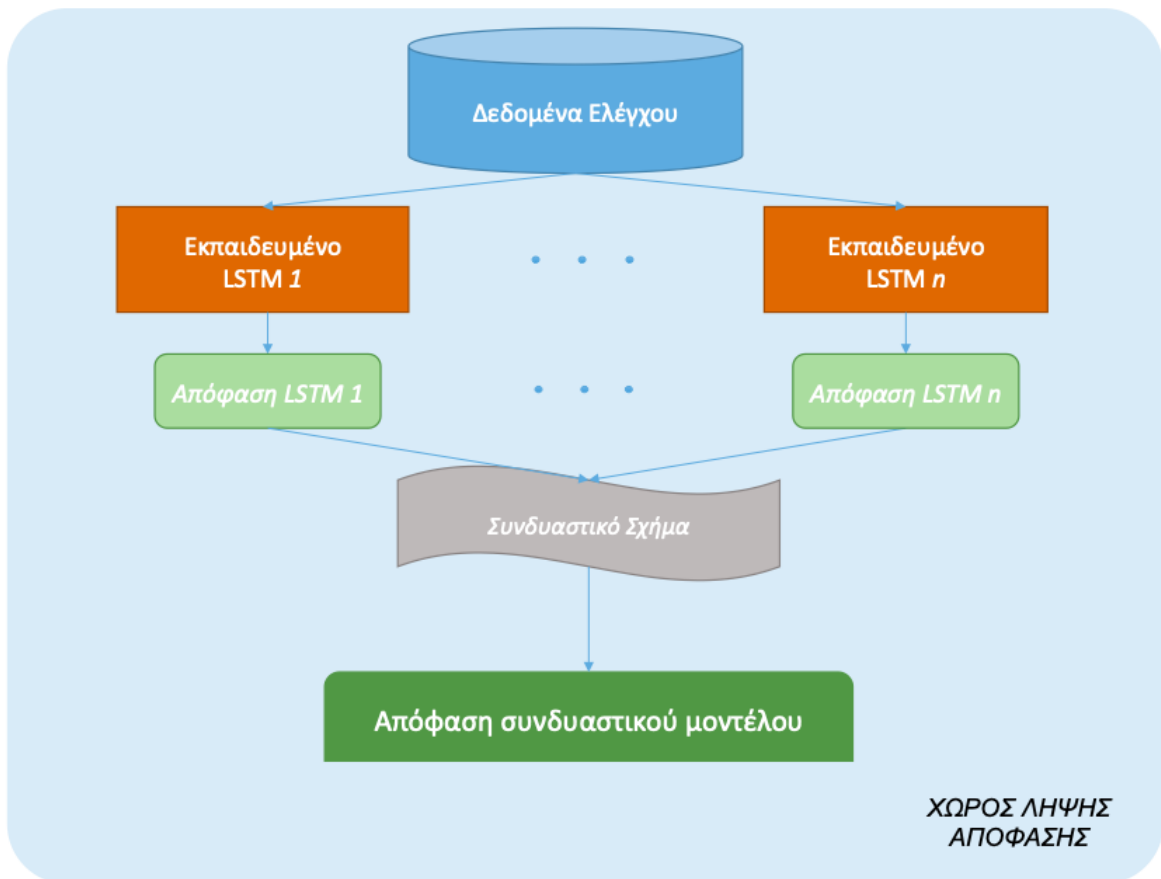
4.5 Ερμηνευσιμότητα κατά LIME

4.5.1 Ανάλυση μηχανισμών μεθόδου

Πίσω από τη λειτουργία του LIME βρίσκεται η υπόθεση ότι κάθε περίπλοκο μοντέλο είναι γραμμικό σε τοπική κλίμακα. Έτσι, η μέθοδος επιχειρεί να προσαρμόσει ένα απλό μοντέλο γύρω από ένα μόνο δείγμα που θα μιμείται τη συμπεριφορά του σύνθετο-γενικού μοντέλου σε αυτή την περιοχή. Το απλό μοντέλο μπορεί στη συνέχεια να χρησιμοποιηθεί για να εξηγήσει τις προβλέψεις του πιο περίπλοκου μοντέλου τοπικά.

Ο αλγόριθμος που ακολουθεί η μέθοδος LIME για την επίτευξη αυτού του στόχου περιλαμβάνει τα εξής βήματα:

1. Λαμβάνει το δείγμα που επιθυμούμε να εξηγηθεί και παράγει νέα, τεχνητά "διαταραγμένα" δείγματα με βάση το σύνολο δεδομένων εκπαίδευσης.
2. Προβλέπει πιθανότητες για όλα τα διαταραγμένα δείγματα χρησιμοποιώντας το σύνθετο-γενικό μοντέλο.
3. Υπολογίζει την απόσταση του δείγματος ενδιαφέροντος από όλα τα διαταραγμένα δείγματα.
4. Μετατρέπει αυτή την απόσταση σε βαθμό ομοιότητας.



Σχήμα 21: Λήψη απόφασης

5. Επιλέγει m χαρακτηριστικά των διαταραγμένων δεδομένων που περιγράφουν καλύτερα το αποτέλεσμα του σύνθετου-γενικού μοντέλου .
6. Βρίσκει το κατάλληλο απλό μοντέλο που αντιπροσωπεύει ακριβέστερα τις προβλέψεις του σύνθετου-γενικού μοντέλου αξιοποιώντας τους βαθμούς ομοιότητας των διαταραγμένων δεδομένων και τα m χαρακτηριστικά τους.
7. Εξάγει τα βάρη χαρακτηριστικών από το απλό μοντέλο και τα χρησιμοποιεί ως εξηγήσεις για την πρόβλεψη του δείγματος ενδιαφέροντος δηλαδή για την τοπική συμπεριφορά του σύνθετου-γενικού μοντέλου.

Από τα παραπάνω είναι ξεκάθαρο ότι υπάρχει μεγάλο περιθώριο βελτιστοποίησης της εξήγησης. Το πιο σημαντικό μεταξύ των επιλογών που επηρεάζουν την ποιότητα της εξήγησης είναι ο τρόπος δημιουργίας των διαταραγμένων δεδομένων, ο τρόπος υπολογισμού του βαθμού ομοιότητας, ο αριθμός των χαρακτηριστικών που επιλέγονται όπως και ο τρόπος που συμβαίνει αυτό και, τέλος, ποιο μοντέλο χρησιμοποιείται ως απλό μοντέλο. Ορισμένες από αυτές τις επιλογές είναι προκαθορισμένες από τον κώδικα του LIME, ενώ άλλες μπορούν να καθοριστούν από τον χρήστη. Παρακάτω αναλύονται όλες αυτές οι επιλογές.

Δημιουργώντας τα διαταραγμένα δείγματα

Όσον αφορά στη διατάραξη ενός δείγματος, η μέθοδος LIME διαχωρίζει την προσέγγιση της ανάλογα με τον τύπο των δεδομένων εισόδου. Επί του παρόντος υποστηρίζονται δύο τύποι εισόδου: δεδομένα τύπου πίνακα και δεδομένα κειμένου.

- Δεδομένα πίνακα

Όταν πρόκειται για δεδομένα πίνακα, η διατάραξη του δείγματος ενδιαφέροντος εξαρτάται από το σύνολο εκπαίδευσης. Κατά τη δημιουργία του επεξηγητή που εκπροσωπεί τη μέθοδο LIME εξάγονται τα στατιστικά στοιχεία των τιμών που λαμβάνει κάθε χαρακτηριστικό στο σύνολο εκπαίδευσης. Έτσι, παίρνοντας δείγματα από μια κανονική κατανομή (0,1) και ύστερα κάνοντας την αντίστροφη διαδικασία του κεντραρίσματος γύρω από τη μέση τιμή (mean-centering) και της κλιμάκωσης (scaling) χρησιμοποιώντας τις μέσες τιμές και τις τυπικές αποκλίσεις που βρέθηκαν 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 σημαντικό, καθώς τελικά αυτός ο βαθμός είναι το μόνο πράγμα που εξασφαλίζει την τοπικότητα γύρω από το δείγμα ενδιαφέροντος.

- **Δεδομένα κειμένου**

Όταν το αποτέλεσμα προβλέψεων κειμένου είναι αυτό που πρέπει να εξηγηθεί, τα διαταραγμένα δείγματα δημιουργούνται αφαιρώντας τυχαία λέξεις από το αρχικό δείγμα.

Υπολογίζοντας βαθμούς ομοιότητας

Ακριβώς όπως τα διαταραγμένα δείγματα δημιουργούνται με διαφορετικό τρόπο αναλόγως με τον τύπο των δεδομένων εισόδου, έτσι και οι βαθμοί ομοιότητας υπολογίζονται με διαφορετικούς τρόπους για τον ίδιο λόγο. Για δεδομένα κειμένου χρησιμοποιείται ως μέτρο ομοιότητας το συνημίτονο, το οποίο προτιμάται γενικότερα στην ανάλυση κειμένου καθώς μετρά αποτελεσματικά τη διαφορά γωνίας μεταξύ δύο διανυσμάτων χαρακτηριστικών. Για δεδομένα πίνακα απαιτείται περισσότερη σκέψη και η βέλτιστη λύση θα εξαρτηθεί από το είδος των δεδομένων εισόδου. Κατ' αρχάς, τα κατηγορικά χαρακτηριστικά θα κωδικοποιηθούν με βάση το αν έχουν ίδια τιμή με αυτή του δείγματος ενδιαφέροντος. Για τα συνεχή χαρακτηριστικά, η απόσταση από το δείγμα ενδιαφέροντος υπολογίζεται με βάση την επιλεγμένη από το χρήστη μετρική απόστασης (προεπιλογή: ευκλείδια) και μετατρέπεται σε βαθμό ομοιότητας χρησιμοποιώντας έναν εκθετικό πυρήνα εξομάλυνσης ενός, καθορισμένου από το χρήστη, πλάτους (προεπιλογή: 0,75 φορές την τετραγωνική ρίζα του πλήθους των χαρακτηριστικών). Ένα μικρό πλάτος πυρήνα σημαίνει ότι ένα δείγμα πρέπει να είναι πολύ κοντά στο δείγμα ενδιαφέροντος για να επηρεάσει το τοπικό μοντέλο, ενώ ένα μεγαλύτερο πλάτος πυρήνα σημαίνει ότι τα δείγματα που βρίσκονται πιο μακριά επηρεάζουν επίσης το μοντέλο.

Επιλέγοντας τα m χαρακτηριστικά

Η μέθοδος LIME έχει τη δυνατότητα να εφαρμόσει μια σειρά διαφορετικών προσεγγίσεων για την επιλογή χαρακτηριστικών από τις οποίες ο χρήστης είναι ελεύθερος να επιλέξει. Πρώτα όμως, πρέπει να επιλεγεί ο αριθμός των χαρακτηριστικών. Ο αριθμός αυτός (m) πρέπει να εξισορροπεί την πολυπλοκότητα του μοντέλου και την απλότητα της εξήγησης. Έτσι, ο χρήστης πρέπει να επιλέξει συνεπώς. Όσον αφορά στην επιλογή των m χαρακτηριστικών, η μέθοδος LIME υποστηρίζει τους ακόλουθους αλγόριθμους:

- Χωρίς επιλογή: Χρήση όλων των χαρακτηριστικών για εξήγηση.
- Επιλογή προς τα εμπρός: Τα χαρακτηριστικά προστίθενται ένα-ένα ανάλογα με τη βελτίωση που προσφέρουν στην προσέγγιση του σύνθετου γενικού μοντέλου από το απλό.
- Επιλογή προς τα πίσω: Αυτό σημαίνει ότι ξεκινώντας από το πλήρες μοντέλο (που περιέχει όλα τα χαρακτηριστικά) εξετάζεται ένα-ένα κάθε χαρακτηριστικό ως προς το εάν θα φέρει μεγαλύτερη βελτίωση η αφαίρεση του, έως ότου επιτευχθεί ένα μοντέλο με m χαρακτηριστικά.
- Υψηλότερα βάρη: Επιλέγονται τα χαρακτηριστικά m με το υψηλότερο απόλυτο βάρος στην παλινδρόμηση κορυφογραμμών (ridge regression) που έχει προκύψει ύστερα από προσαρμογή στα αποτελέσματα-προβλέψεις του σύνθετου-γενικού μοντέλου.

- LASSO: Επιλέγονται τα m χαρακτηριστικά που είναι λιγότερο επιρρεπή σε συρρίκνωση βάσει της διαδρομής κανονικοποίησης μιας LASSO regression εφαρμοσμένης στα αποτελέσματα του σύνθετου-γενικού μοντέλου. Αυτή η μέθοδος αποτελεί προεπιλογή της μεθόδου LIME για την επιλογή των m χαρακτηριστικών. Συγκεκριμένα ο αλγόριθμος περιλαμβάνει τα εξής βήματα:

- Δίνεται ένα σύνολο χαρακτηριστικών x_1, x_2, \dots, x_p για κάθε δείγμα εισόδου και μια πρόβλεψη y για κάθε δείγμα με αυτά τα χαρακτηριστικά.
- Το LASSO ταιριάζει ένα γραμμικό μοντέλο με συντελεστές b_1, \dots, b_p για όλα τα δείγματα εισόδου και για κάθε ένα υπολογίζει το $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$.
- Ελαχιστοποιεί το

$$\sum (y - \hat{y})^2 \quad (4.1)$$

απαιτώντας

$$\sum |b_j| \leq s \quad (4.2)$$

Το πρώτο άθροισμα εφαρμόζεται σε όλα τα δείγματα του συνόλου δεδομένων, ενώ το s είναι μια παράμετρος συντονισμού. Όταν το s είναι αρκετά μεγάλο, ο περιορισμός δεν έχει καμία επίδραση και η λύση είναι απλώς η συνήθης παλινδρόμηση πολλαπλών γραμμικών ελαχίστων τετραγώνων του y . Ωστόσο, για μικρότερες τιμές s ($s \geq 0$) οι λύσεις είναι συρρικνωμένες παραλλαγές των εκτιμήσεων των ελαχίστων τετραγώνων. Συχνά, ορισμένοι από τους συντελεστές b_j είναι μηδέν. Η επιλογή του s είναι σαν να επιλέγεται ο αριθμός των χαρακτηριστικών που θα χρησιμοποιηθούν σε ένα μοντέλο παλινδρόμησης.

- Αυτόματη: Χρησιμοποιεί επιλογή προς τα εμπρός εάν $m \leq 6$, αλλιώς υψηλότερα βάρη.

Προσαρμογή ενός απλού μοντέλου στα διαταραγμένα και μειωμένων χαρακτηριστικών δεδομένα

Μόλις δημιουργηθούν τα διαταραγμένα δεδομένα, υπολογιστούν οι βαθμοί ομοιότητας και επιλεχθούν τα m χαρακτηριστικά, έφτασε η στιγμή να γίνει η προσαρμογή του απλού μοντέλου. Εάν το σύνθετο μοντέλο εξάγει πιθανότητες, τότε το απλό μοντέλο θα προβλέψει άμεσα την έξοδο του σύνθετου μοντέλου. Εάν το σύνθετο μοντέλο είναι ταξινομητής, τότε το απλό μοντέλο θα προβλέψει την πιθανότητα της επιλεγμένης κλάσης.

Η μόνη απαίτηση για το απλό μοντέλο είναι να είναι εύκολο να εξαχθεί εξήγηση από αυτό και να μπορεί να λειτουργήσει με σταθμισμένη είσοδο καθώς αυτή σταθμίζεται με βάση τον βαθμό ομοιότητας που έχει υπολογιστεί ώστε να διασφαλίζεται ότι τα σφάλματα στα δείγματα που είναι πιο κοντά στο δείγμα ενδιαφέροντος έχουν μεγαλύτερη βαρύτητα από τα υπόλοιπα. Η μέθοδος LIME υποστηρίζει πολλούς τύπους απλών μοντέλων, ωστόσο σαν προεπιλογή χρησιμοποιεί ένα ridge regression μοντέλο.

Τέλος, μαθηματικά, η μέθοδος LIME μπορεί να εκφραστεί ως εξής:

$$\text{explanation}(x) = \arg \min_{g \in G} L(f, g, \pi_x) + \Omega(g) \quad (4.3)$$

Όπου το μοντέλο εξήγησης για το δείγμα x είναι το μοντέλο g (π.χ. μοντέλο γραμμικής παλινδρόμησης) που ελαχιστοποιεί την απώλεια L (π.χ. μέσο τετράγωνο σφάλμα), η οποία μετρά πόσο κοντά είναι η εξήγηση στην πρόβλεψη f του σύνθετου-γενικού μοντέλου (π.χ. LSTM), ενώ η πολυπλοκότητα του μοντέλου, $\Omega(g)$, διατηρείται χαμηλή. Το G είναι η οικογένεια πιθανών εξηγήσεων, για παράδειγμα όλα τα πιθανά μοντέλα γραμμικής παλινδρόμησης. Το μέτρο εγγύτητας

$$\pi_x = \exp(-D(x, z)^2 / \sigma^2) \quad (4.4)$$

με $D(x, z)$ την συνάρτηση απόστασης μεταξύ του διαταραγμένου δείγματος και του δείγματος ενδιαφέροντος, σ να έχει προεπιλεγμένη τιμή 0,75 φορές την τετραγωνική ρίζα του πλήθους των χαρακτηριστικών, καθορίζει πόσο μεγάλη είναι η γειτονιά που θεωρούμε γύρω από το δείγμα x .

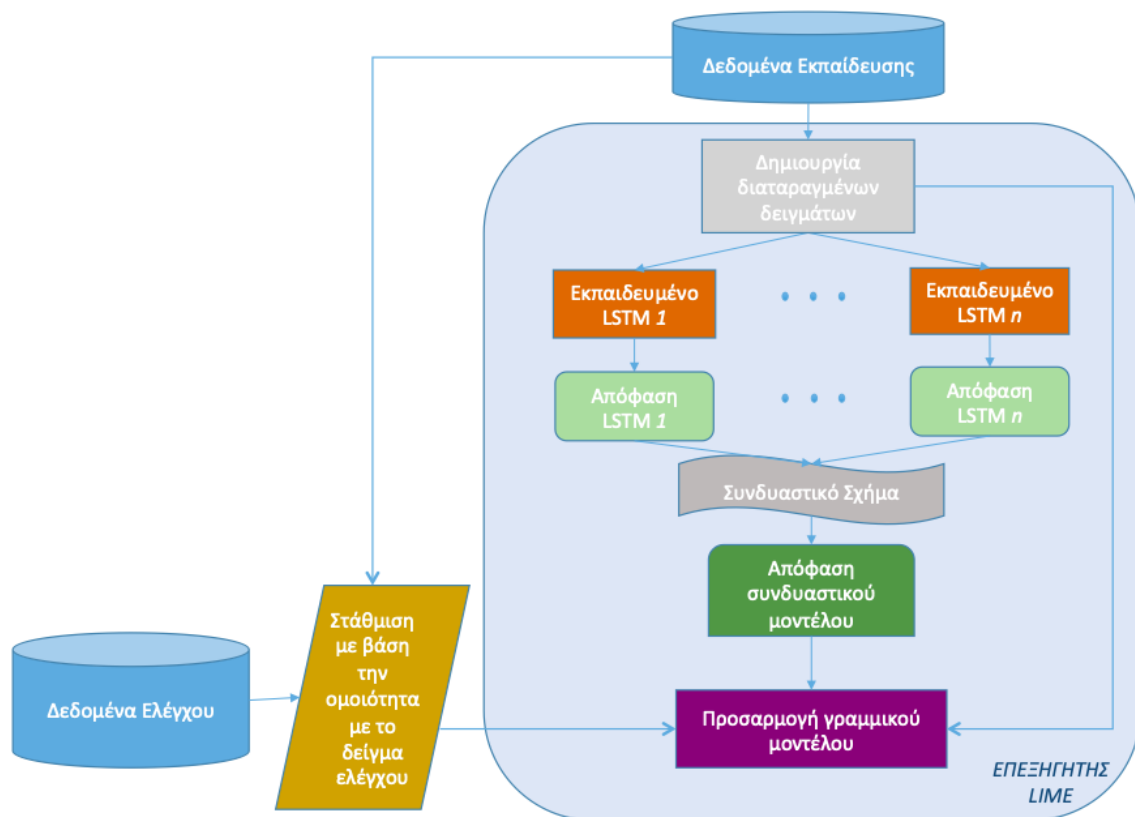
Στην πράξη, η μέθοδος LIME βελτιστοποιεί μόνο το μέρος απώλειας, ενώ ο χρήστης πρέπει να καθορίσει την πολυπλοκότητα μέσω της επιλογής του πλήθους m των χαρακτηριστικών [39] [48] [49] [50].

Συγκεκριμένα, η απώλεια ορίζεται από τον τύπο:

$$L(f, g, \pi_x) = \sum_{z, z' \in Z} \pi_x(z) (f(z) - g(z'))^2 \quad (4.5)$$

4.5.2 LIME και LSTM

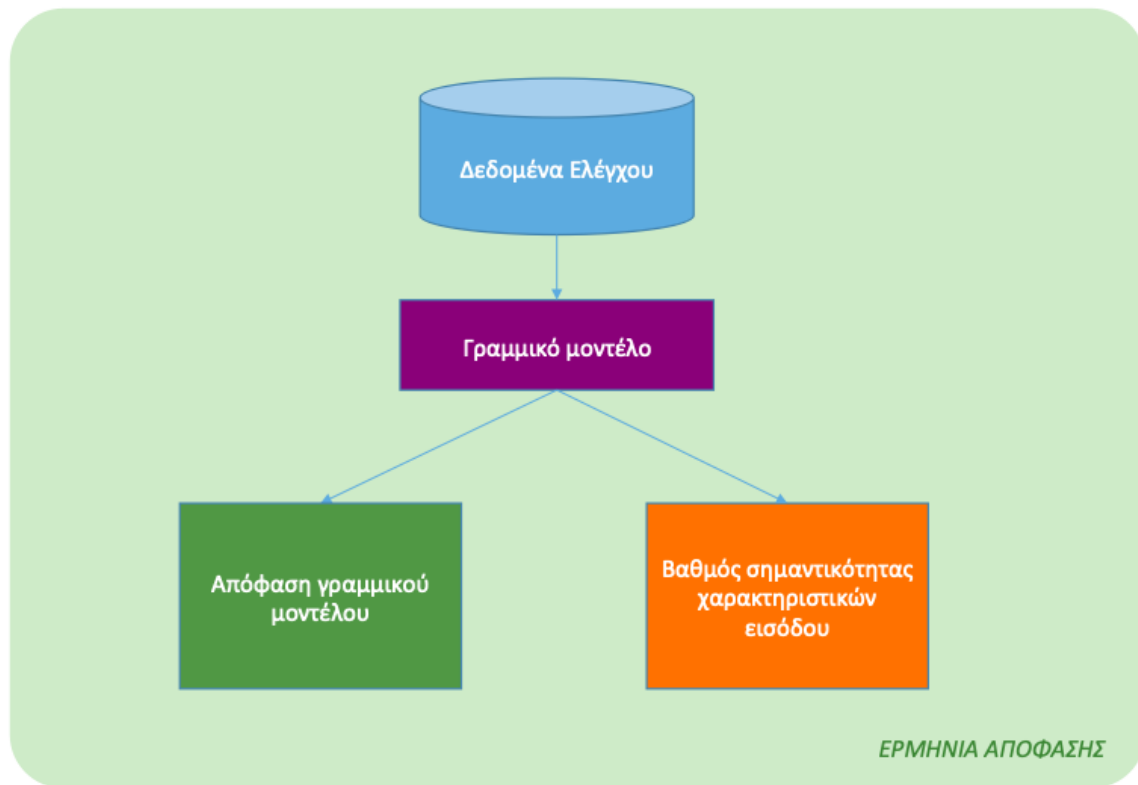
Μετά την αναλυτική παρουσίαση των μηχανισμών της μεθόδου LIME απομένει να παρουσιαστεί ο τρόπος που εφαρμόστηκε η συγκεκριμένη μέθοδος στο συνδιαστικό μοντέλο LSTM που περιγράφηκε παραπάνω. Το σχήμα 23 σε συνδυασμό με το σχήμα 24 δίνει μια οπτικοποίηση της βασικής δομής του επεξηγητή LIME πάνω στο συνδιαστικό μοντέλο LSTM έως την τελική εξαγωγή της ερμηνείας για κάποιο δείγμα.



Σχήμα 22: Εσωτερική δομή επεξηγητή LIME

Είναι σημαντικό να σημειωθεί ότι το LIME δεν παρέχει τη δυνατότητα ερμηνείας πολλαπλών προβλέψεων για ένα μόνο δείγμα όπως συμβαίνει στην δική μας περίπτωση όπου κάθε δείγμα διαθέτει δύο λίστες χαρακτηριστικών (μία για κάθε timestep) και προβλέπονται δύο τιμές πιθανότητας, μία για κάθε timestep. Χρησιμοποιώντας την εκδοχή του LIME που αφορά σε αναδρομικά νευρωνικά δίκτυα, παρέχεται η δυνατότητα να δίνονται μεν δείγματα με πολλαπλές λίστες χαρακτηριστικών (πολλά timesteps) αλλά από κάθε τέτοιο δείγμα, το μοντέλο προς ερμηνεία να παράγει μόνο μία τελική τιμή πρόβλεψης, αυτή του τελικού timestep. Συνεπώς η περίπτωση μας, όπου για κάθε timestep παράγεται και μία πρόβλεψη, δεν ήταν στα πλαίσια των περιπτώσεων που χειρίζεται κατευθείαν ο έτοιμος κώδικας του LIME.

Η τροποποίηση που έγινε ώστε να εφαρμοστεί η μέθοδος LIME στο συνδιαστικό μοντέλο LSTM που είχε παραχθεί βασίστηκε στα εξής βήματα:



Σχήμα 23: Τελική ερμηνεία

- Κατά την δημιουργία του ερμηνευτή LIME απαιτείται ως είσοδος το σύνολο εκπαίδευσης του μοντέλου. Εδώ δόθηκε ένα ελαφρώς τροποποιημένο, ως προς τις διαστάσεις, σύνολο. Αντί για σύνολο τριών διαστάσεων (αριθμός δειγμάτων x αριθμός timesteps x αριθμός χαρακτηριστικών) δόθηκε ένα σύνολο δύο διαστάσεων (αριθμός δειγμάτων x νέος αριθμός χαρακτηριστικών) όπου στην ουσία οι δύο λίστες χαρακτηριστικών (μία για κάθε timestep) συνενώθηκαν σε μία λίστα με μέγεθος το διπλάσιο του αριθμού των χαρακτηριστικών.
- Στη συνέχεια δόθηκε το δείγμα προς ερμηνεία και πάλι στην μορφή όπου τα χαρακτηριστικά των δύο timesteps ήταν συνενωμένα.
- Ο ερμηνευτής LIME παράγαγε τα διαταραγμένα δεδομένα λαμβάνοντας υπόψιν τις διαστάσεις του συγκεκριμένου δείγματος προς ερμηνεία και έτσι τα νέα δεδομένα που προέκυψαν είχαν την ίδια μορφή. Αυτά τα δεδομένα τροφοδοτούνται σε μία συνάρτηση που πρέπει να εξάγει για κάθε δεδομένο δύο πιθανότητες (μία για κάθε κλάση) και αυτή η συνάρτηση δίνεται ως είσοδος στο LIME. Πρέπει να τονιστεί ότι αυτές οι πιθανότητες αφορούν στο ένα από τα δύο timesteps κάθε φορά καθώς δεν γίνεται να εξαχθούν τέσσερις πιθανότητες (δύο για κάθε timestep). Συνεπώς αυτή η συνάρτηση παίρνει και ως είσοδο τον αριθμό του timestep που μας ενδιαφέρει να ερμηνεύσουμε την πρόβλεψη του. Αν θέλουμε να κάνουμε ερμηνεία και για τα δύο timesteps πρέπει να τρέξουμε δύο φορές την εντολή ερμηνείας δίνοντας κάθε φορά διαφορετική τιμή εισόδου στην συνάρτηση (μηδέν για το πρώτο timestep ή ένα για το δεύτερο). Αυτή η συνάρτηση λοιπόν κατασκευάστηκε εκ του μηδενός και υπενθυμίζεται ότι λαμβάνει ως είσοδο τα νέα δεδομένα και το timestep και πρέπει να δώσει έξοδο στην μορφή που εξηγήθηκε παραπάνω. Συνεπώς ακολουθεί τα εξής βήματα:
 - Τα νέα δεδομένα επαναφέρονται στην αρχική μορφή των δύο ξεχωριστών λιστών (μία για κάθε timestep) ώστε να τροφοδοτηθούν σε κάθε μοντέλο από τα δύο LSTM το οποίο φυσικά απαιτεί αυτή την μορφή στην είσοδο του.

- Έχοντας τέσσερις προβλέψεις για κάθε διαταραγμένο δείγμα, δύο για κάθε timestep, κρατάμε το σύνολο των δύο προβλέψεων που αφορούν στο timestep που μας ενδιαφέρει και στη συνέχεια λαμβάνεται ο απλός μέσος όρος αυτών και κατ' επέκταση εξάγεται μία τελική τιμή πιθανότητας για κάθε διαταραγμένο δείγμα (στο συγκεκριμένο timestep). Αυτή η τιμή αφορά στην πιθανότητα το δείγμα (για αυτό το timestep) να ανήκει στην κλάση "νοσηλείας".
- Για κάθε δείγμα λοιπόν παράγεται μία λίστα δύο πιθανοτήτων, μία που αφορά στην πιθανότητα "νοσηλείας" και μία στην πιθανότητα "μη-νοσηλείας" ($1 - p[\text{"νοσηλείας"}]$). Έτσι τελικά η συνάρτηση δίνει ως έξοδο έναν πίνακα διαστάσεων (αριθμός διαταραγμένων δεδομένων) $\times 2$
- Αυτές οι τελικές τιμές πιθανότητας θεωρούνται ως η πρόβλεψη-πιθανότητα $f(z)$, για διατάραξη z , του συνολικού σύνθετου-γενικού μοντέλου. Στη συνέχεια, ο αλγόριθμος του LIME εξακολουθεί όπως προβλέπεται από τα βήματα που περιγράφηκαν αρχικά.

Κεφάλαιο 5

Αποτελέσματα και Μελλοντική Εργασία

Στο παρόν κεφάλαιο παρουσιάζονται αναλυτικά η αποδοτικότητα του σύνθετου μοντέλου LSTM χρησιμοποιώντας τα εξής κριτήρια: ακρίβεια, ευαισθησία, ειδικότητα, επιφάνεια κάτω από την καμπύλη ROC (AUC) και brier score. Επιπλέον παρατίθενται τα αποτελέσματα της ερμηνείας αυτού του μοντέλου, για τα χαρακτηριστικά που χρησιμοποιήθηκαν στην πρόβλεψη, μέσω διαγραμμάτων για την καλύτερη οπτικοποίηση των αποτελεσμάτων.

Για την αξιολόγηση της απόδοσης και της ικανότητας γενίκευσης του συνδυαστικού μοντέλου εφαρμόστηκε διασταυρωμένη επικύρωση 10 πτυχών. Τα κριτήρια που αξιολογήθηκαν στα δεδομένα εκπαίδευσης σε κάθε επανάληψη του αλγορίθμου διασταυρωμένης επικύρωσης ορίζονται ως εξής:

- Ακρίβεια (accuracy): εκφράζει το ποσοστό επιτυχίας του μοντέλου στην ταξινόμηση των δειγμάτων στις σωστές κατηγορίες επί του συνόλου δεδομένων.

$$\text{Accuracy} = (\text{TP} + \text{TN})/\text{total}$$

- Ευαισθησία (sensitivity): εκφράζει την ικανότητα του ταξινομητή να διακρίνει τις πραγματικά θετικές προβλέψεις.

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN})$$

- Ειδικότητα (specificity): εκφράζει την ικανότητα του ταξινομητή να διακρίνει τις πραγματικά αρνητικές προβλέψεις.

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP})$$

- Area Under the ROC Curve: Η καμπύλη AUC - ROC είναι μια μετρική της απόδοσης για ένα πρόβλημα ταξινόμησης σε διάφορες ρυθμίσεις κατωφλίου. Το ROC είναι μια καμπύλη πιθανότητας και το AUC αντιπροσωπεύει βαθμό ή μέτρο διαχωρισμού. Δείχνει δηλαδή σε τι βαθμό το μοντέλο μπορεί να διαχωρίσει τις κλάσεις. Όσο υψηλότερη είναι η AUC, τόσο καλύτερα το μοντέλο προβλέπει τα μηδενικά ως μηδενικά και τους άσσους ως άσσους. Αναλογικά, όσο υψηλότερη είναι η AUC, τόσο καλύτερα το μοντέλο διακρίνει μεταξύ ασθενών που θα νοσηλευτούν και που δεν θα νοσηλευτούν. Η καμπύλη ROC έχει στον άξονα y τα TPR(ποσοστό των πραγματικά θετικών τιμών) και στον άξονα x το FPR(ποσοστό των λανθασμένα θετικών τιμών).
- Brier Score: είναι μέτρο αξιολόγησης των μοντέλων, όσον αφορά στην ακρίβεια των προβλέψεων πιθανότητας, για το οποίο υπολογίζεται η μέση τετραγωνική απόκλιση των εκτιμώμενων πιθανοτήτων κάποιων γεγονότων και του αποτελέσματός τους.

Οι συντομογραφίες TP, TN, FP, FN και total ορίζονται ως εξής:

- True Positive (TP): Το σύνολο των δειγμάτων για τα οποία η πρόβλεψη είναι θετική και η πραγματική τιμή είναι επίσης θετική.
- True Negative (TN): Το σύνολο των δειγμάτων για τα οποία η πρόβλεψη είναι αρνητική και η πραγματική τιμή είναι επίσης αρνητική.

Πίνακας 6: Κριτήρια αξιολόγησης με χρήση απλού μέσου όρου

Κριτήριο Αξιολόγησης	Μέση τιμή ± Τυπική απόκλιση για συνδυαστικό μοντέλο LSTM
Accuracy	72,96% ± 9,57%
Sensitivity	70% ± 22,11%
Specificity	73,29% ± 10,12%
AUC	81,26% ± 14,20%
Brier Score	9,03% ± 2,73%

Πίνακας 7: Κριτήρια αξιολόγησης με χρήση σταθμισμένου μέσου όρου

Κριτήριο Αξιολόγησης	Μέση τιμή ± Τυπική απόκλιση για συνδυαστικό μοντέλο LSTM
Accuracy	86,13% ± 6,71%
Sensitivity	51,67% ± 25,22%
Specificity	90,25% ± 7,67%
AUC	80,25% ± 10,05%
Brier Score	7,93% ± 1,56%

Πίνακας 8: Κριτήρια αξιολόγησης με χρήση σχήματος ψηφοφορίας

Κριτήριο Αξιολόγησης	Μέση τιμή ± Τυπική απόκλιση για συνδυαστικό μοντέλο LSTM
Accuracy	88,53% ± 6,29%
Sensitivity	41,67% ± 25,00%
Specificity	94,20% ± 6,54%

- False Positive (FP): Το σύνολο των δειγμάτων για τα οποία η πρόβλεψη είναι θετική και η πραγματική τιμή είναι αρνητική.
- False Negative (FN): Το σύνολο των δειγμάτων για τα οποία η πρόβλεψη είναι αρνητική και η πραγματική τιμή είναι θετική.
- total: Το πλήθος των συνολικών δειγμάτων.

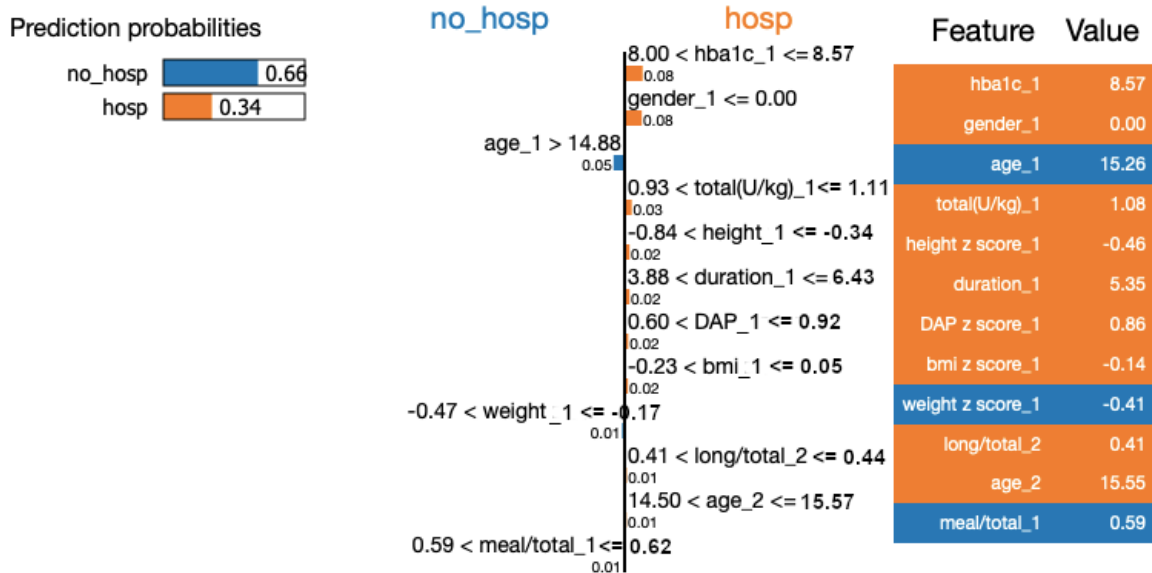
Όσον αφορά στο συνδυαστικό μοντέλο που περιγράφηκε στα προηγούμενα κεφάλαια, οι τιμές των μετρικών για το συνδυαστικό σχήμα του απλού μέσου όρου είναι αυτές που παρουσιάζονται στον πίνακα 6.

Ο πίνακας 7 παρουσιάζει τις τιμές των μετρικών για το συνδυαστικό σχήμα του σταθμισμένου μέσου όρου ενώ ο πίνακας 8 για το συνδυαστικό σχήμα ψηφοφορίας.

Για την κατηγοριοποίηση κάθε δείγματος στην θετική ή την αρνητική κλάση χρησιμοποιήθηκε το κατώφλι 0.25. Δηλαδή για τιμές πιθανότητας πάνω από 0.25, το δείγμα κατατάχθηκε στην θετική κλάση (νοσηλεία), διαφορετικά κατατάχθηκε στην αρνητική (μη-νοσηλεία).

5.1 Ατομικά Διαγράμματα LIME

Η μέθοδος ερμηνευσιμότητας του LIME παρέχει την δυνατότητα οπτικοποίησης της εξήγησης που αφορά σε συγκεκριμένο ασθενή. Παρακάτω θα παρουσιαστούν τα ατομικά διαγράμματα για τις εξής περιπτώσεις ασθενών:

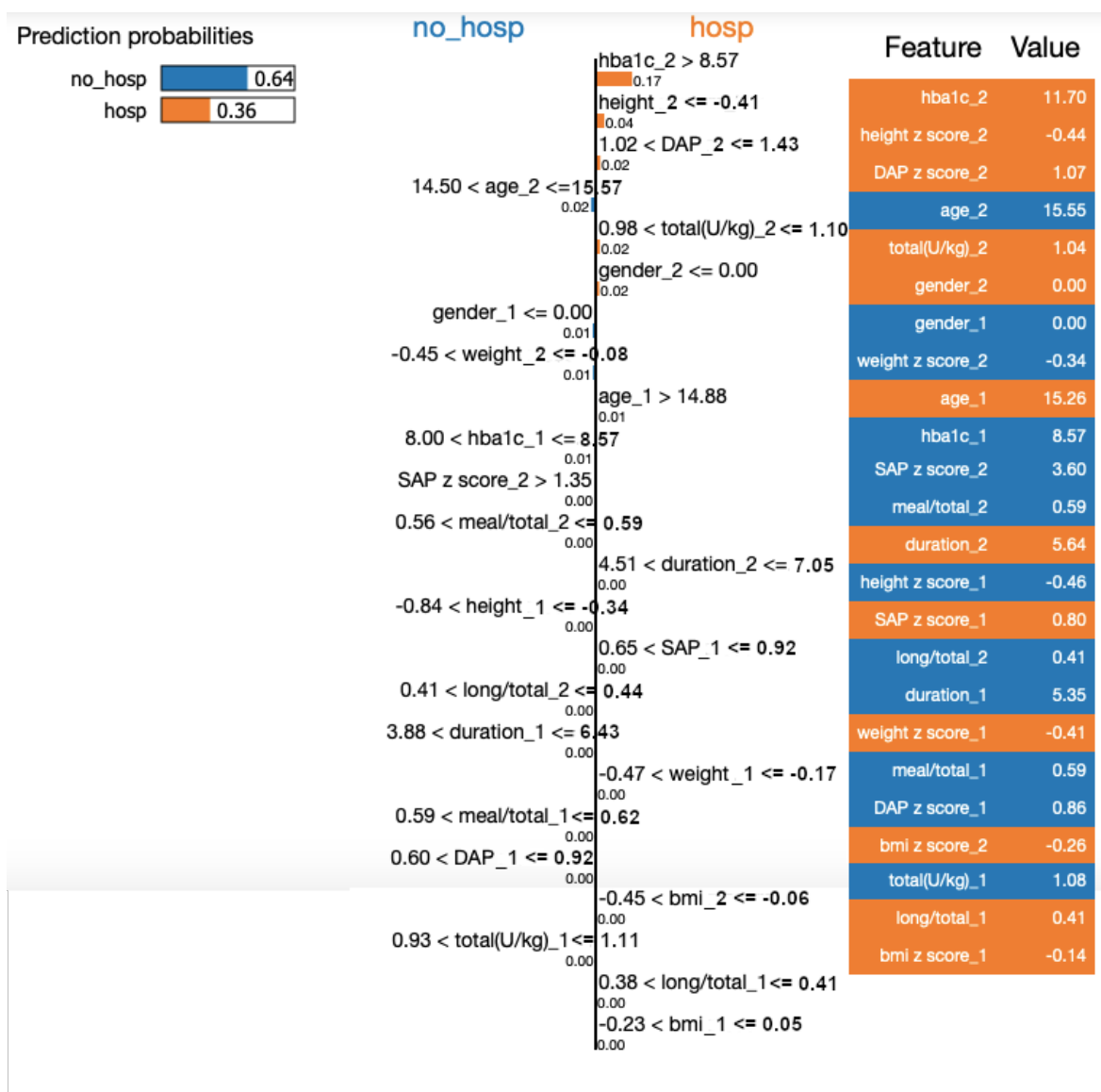


Σχήμα 24: Ασθενής 1 - Ερμηνεία απόφασης για τον κίνδυνο νοσηλείας (χρονική στιγμή 1)

1. Περίπτωση ασθενούς με νοσηλεία και επανανοσηλεία

Όσον αφορά στην πρώτη χρονική στιγμή, η πιθανότητα που εξήγαγε το συνδυαστικό μοντέλο LSTM είναι 0.41, τιμή που προσεγγίστηκε σε καλό βαθμό από το γραμμικό μοντέλο του LIME, το οποίο έδωσε πιθανότητα 0.34. Το κυριότερο χαρακτηριστικό που συνέβαλε πιο καθοριστικά στην εξαγωγή της προηγούμενης τιμής φαίνεται να είναι η υψηλή τιμή της γλυκοζυλιωμένης αιμοσφαιρίνης με συντελεστή βαρύτητας 8%, πράγμα που ενισχύει την αξιοπιστία της ερμηνευτικής μεθόδου καθώς σύμφωνα με τη βιβλιογραφία, η τιμή της γλυκοζυλιωμένης αποτελεί όντως έναν καθοριστικό παράγοντα όσον αφορά στον κίνδυνο νοσηλείας από διαβητική κετοξέωση. Επιπλέον παράγοντες που επέδρασαν ήταν το αρσενικό φύλο με 8%, η συνολική ημερήσια δόση ινσουλίνης με 3%, το ύψος με 2%, η χρονική διάρκεια του διαβήτη με 2% και ο δείκτης μάζας σώματος του ασθενούς με, επίσης, 2%. Από τους παραπάνω παράγοντες είναι λογικό να επέδρασε, σύμφωνα με τη βιβλιογραφία, η ημερήσια δόση ινσουλίνης, η διάρκεια του διαβήτη, η πίεση και ο δείκτης μάζας σώματος του ασθενούς. Στους λιγότερο σημαντικούς παράγοντες, φαίνεται να συμπεριλαμβάνονται και κάποιες τιμές χαρακτηριστικών της δεύτερης χρονικής στιγμής, ωστόσο δεν πρέπει να ληφθούν υπόψιν καθώς συμπεριλήφθηκαν στην είσοδο του LIME με σκοπό την επίλυση τεχνικών προβλημάτων του LIME και επιπλέον είναι φανερό ότι και η ίδια η μέθοδος του LIME αναγνωρίζει την χαμηλή σημαντικότητά τους.

Όσον αφορά στην δεύτερη χρονική στιγμή, η πιθανότητα που εξήγαγε το συνδυαστικό μοντέλο LSTM είναι 0.76, τιμή που δεν προσεγγίζεται επαρκώς από το γραμμικό μοντέλο του LIME, το οποίο έδωσε πιθανότητα 0.36. Ωστόσο αποδεικνύεται εξαιρετικά ικανό να αναγνωρίσει τα καθοριστικότερα χαρακτηριστικά στην περίπτωση της επανανοσηλείας με πρώτο και κύριο την ιδιαίτερα υψηλή τιμή της γλυκοζυλιωμένης αιμοσφαιρίνης στη δεύτερη χρονική στιγμή με συντελεστή βαρύτητας 17% και στη συνέχεια το ύψος με 4%, την υψηλή διαστολική αρτηριακή πίεση με 2% και την ημερήσια δόση ινσουλίνης με επίσης 2%. Τα χαρακτηριστικά της πρώτης χρονικής στιγμής λήφθηκαν επίσης υπόψιν αλλά δεν φαίνεται να επέδρασαν ιδιαίτερα στην τελική τιμή πιθανότητας.

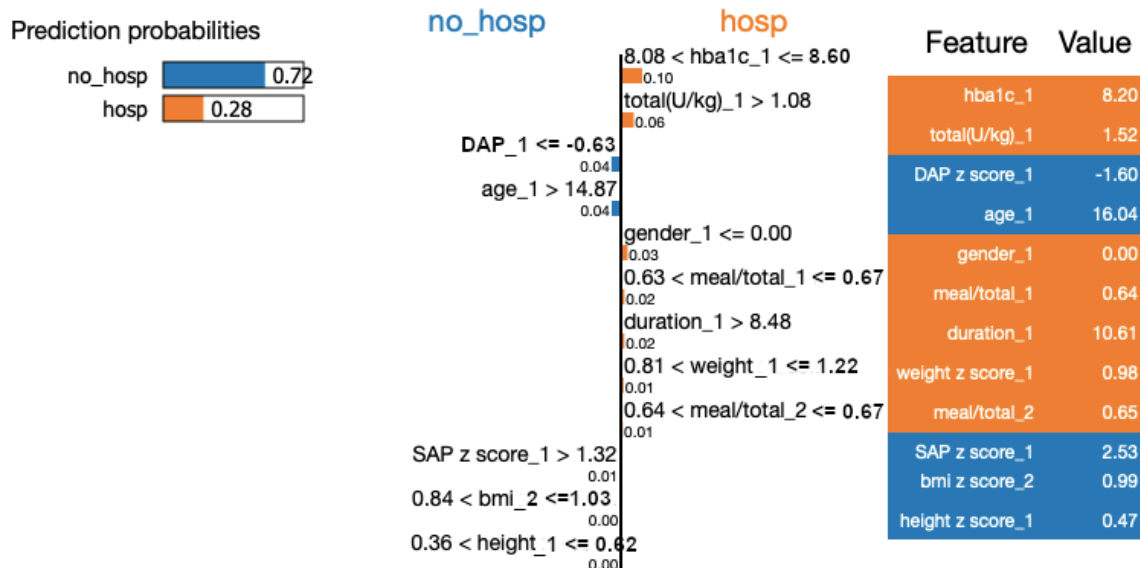


Σχήμα 25: Ασθενής 1 - Ερμηνεία απόφασης για τον κίνδυνο επανανοσηλείας (χρονική στιγμή 2)

2. Περίπτωση ασθενούς με νοσηλεία αλλά όχι επανανοσηλεία

Όσον αφορά στην πρώτη χρονική στιγμή, η πιθανότητα που εξήγαγε το συνδυαστικό μοντέλο LSTM είναι 0.42, τιμή που δεν προσεγγίστηκε ικανοποιητικά από το γραμμικό μοντέλο του LIME, το οποίο έδωσε πιθανότητα 0.28. Τα χαρακτηριστικά που επέδρασαν σε μεγαλύτερο βαθμό στην διαμόρφωση της τελικής τιμής πιθανότητας είναι η υψηλή τιμή της γλυκοζυλιωμένης αιμοσφαιρίνης με συντελεστή βαρύτητας 10%, στη συνέχεια η τιμή της ημερήσιας δόσης ινσουλίνης με μικρότερο συντελεστή 6%, το αρσενικό φύλο με 3%, η τιμή της γευματικής ινσουλίνης προς την συνολική με 2% και η μεγάλη χρονική διάρκεια του διαβήτη με 2%. Η χαμηλή διαστολική πίεση και η ηλικία του ασθενούς φαίνεται να επηρεάζουν προς την αντίθετη κατεύθυνση (της μή-νοσηλείας) πράγμα που συνάδει με τη βιβλιογραφία.

Όσον αφορά στην δεύτερη χρονική στιγμή, η πιθανότητα που εξήγαγε το συνδυαστικό μοντέλο LSTM είναι 0.15, τιμή που προσεγγίζεται επαρκώς από το γραμμικό μοντέλο του LIME, το οποίο έδωσε πιθανότητα 0.12. Και πάλι φαίνεται να κυριαρχούν ως καθοριστικότερα, τα χαρακτηριστικά της δεύτερης χρονικής στιγμής, όπως είναι και το επιθυμητό. Η υψηλή τιμή της γλυκοζυλιωμένης έχει συντελεστή βαρύτητας 6% και συνάδει περισσότερο με το ενδεχόμενο της νοσηλείας. Ωστόσο η χαμηλή διαστολική πίεση και η ικανή ηλικία φαίνεται να συμφωνούν



Σχήμα 26: Ασθενής 2 - Ερμηνεία απόφασης για τον κίνδυνο νοσηλείας (χρονική στιγμή 1)

με τη μη-νοσηλεία και να λαμβάνουν συντελεστές βαρύτητας 4% και 3%, αντίστοιχα. Στη συνέχεια φαίνεται να παίζει κάποιο ρόλο η τιμή της ημερήσιας δόσης ινσουλίνης και ο δείκτης μάζας σώματος, της πρώτης χρονικής στιγμής.

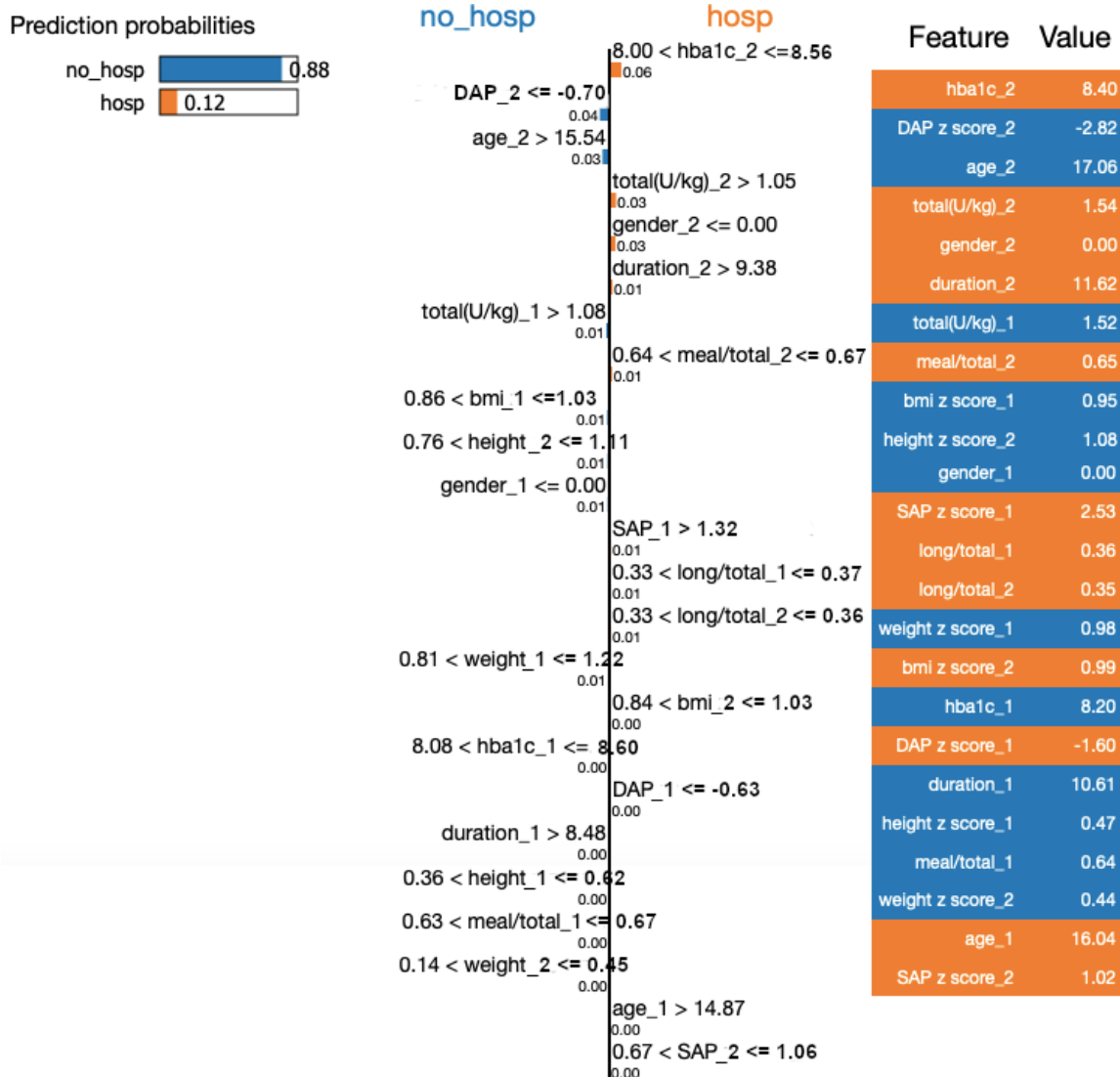
3. Περίπτωση ασθενούς χωρίς νοσηλεία

Όσον αφορά στην πρώτη χρονική στιγμή, η πιθανότητα που εξήγαγε το συνδυαστικό μοντέλο LSTM είναι ίδια με την τιμή που δίνει το γραμμικό μοντέλο του LIME, δηλαδή 0.2. Τα χαρακτηριστικά που επέδρασαν σε μεγαλύτερο βαθμό στην διαμόρφωση της τελικής τιμής πιθανότητας είναι η φυσιολογική τιμή της γλυκοζυλιωμένης αιμοσφαιρίνης με συντελεστή βαρύτητας 13%, στη συνέχεια το θηλυκό φύλο με 12%, το φυσιολογικό βάρος με 3% και η τιμή της ημερήσιας δόσης ινσουλίνης με συντελεστή 2%. Η μικρή χρονική διάρκεια που έχει εμφανιστεί η ασθένεια συνάδει περισσότερο με το ενδεχόμενο νοσηλείας όπως είναι το αναμενόμενο καθώς η έλλειψη εμπειρίας στην διαχείριση της ασθένειας είναι αυτή που συντελεί στην εμφάνιση επεισοδίου διαβητικής κετοξέωσης και κατ' επέκταση νοσηλείας.

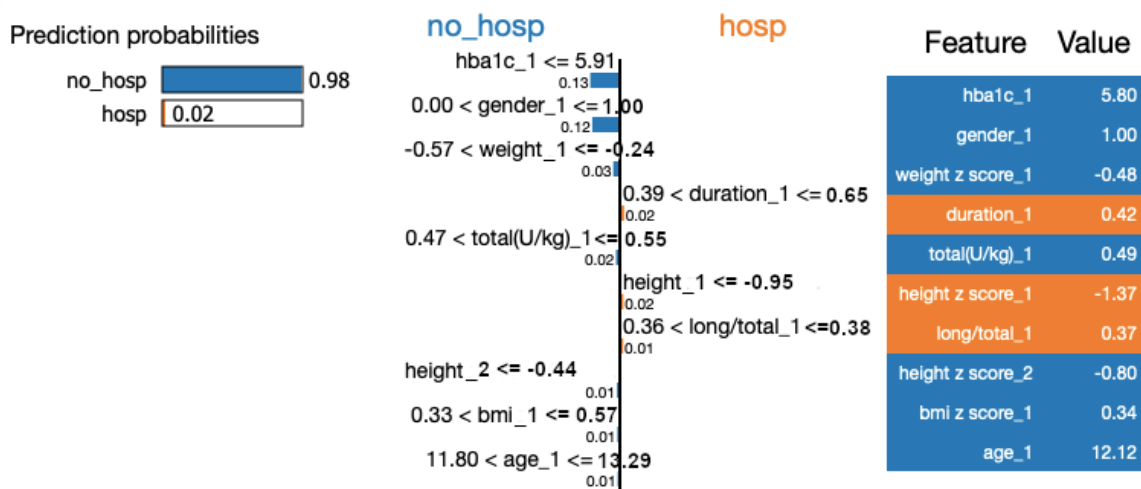
Όσον αφορά στην δεύτερη χρονική στιγμή, η πιθανότητα που εξήγαγε το συνδυαστικό μοντέλο LSTM είναι 0.009, τιμή που απέχει σε μικρό βαθμό από αυτή του γραμμικού μοντέλου του LIME, το οποίο έδωσε πιθανότητα 0.04. Και πάλι φαίνεται να κυριαρχούν ως σημαντικότερα, τα χαρακτηριστικά της δεύτερης χρονικής στιγμής, όπως είναι και το επιθυμητό. Παρατηρείται η χαμηλή τιμή της γλυκοζυλιωμένης έχει συντελεστή βαρύτητας 13%, το θηλυκό γένος με 11%, το χαμηλό βάρος με 5% και η τιμή της γευματικής ινσουλίνης προς την συνολική με 2%. Το χαμηλό ύψος φαίνεται να οδηγεί στην νοσηλεία, ενώ το θηλυκό γένος και η χαμηλή γλυκοζυλιωμένη της πρώτης χρονικής στιγμής φαίνεται να παίζουν επίσης έναν μικρό ρόλο προς την κατεύθυνση της νοσηλείας. Αυτό έρχεται σε αντίθεση με τη βιβλιογραφία και με την σημαντικότητα που δόθηκε στο γένος στη δεύτερη χρονική στιγμή.

Σε γενικές γραμμές, η μέθοδος του LIME φαίνεται να υποδηλώνει ότι το συνδυαστικό μοντέλο LSTM εξήγαγε αξιόπιστες πιθανότητες βασισμένες στις τιμές των κυριότερων, σύμφωνα με τη βιβλιογραφία, χαρακτηριστικών.

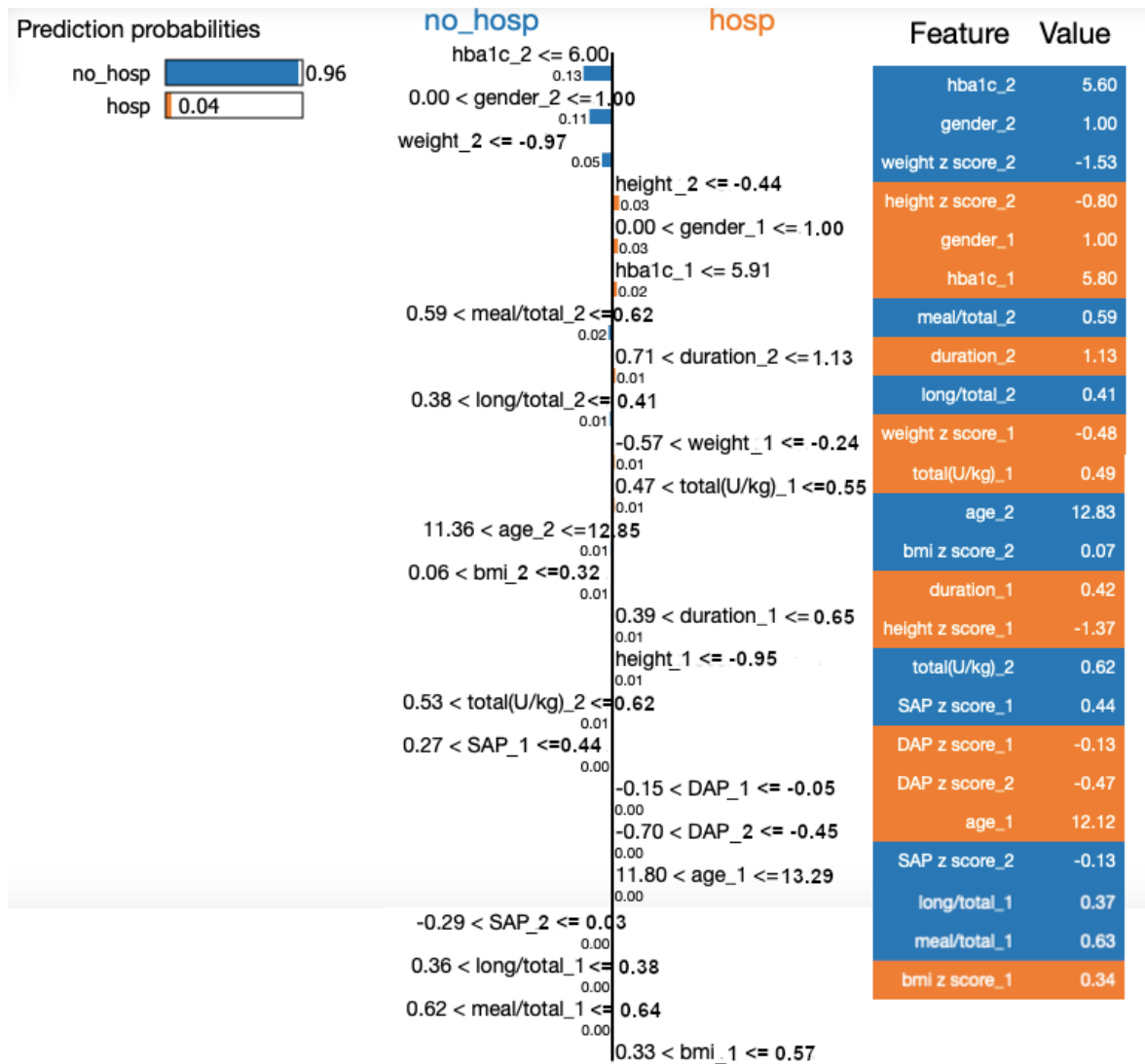
Στα διαγράμματα, για κάθε ασθενή, παρουσιάζονται οι ερμηνείες των δύο προβλέψεων (μία για την πρώτη και μία για τη δεύτερη χρονική στιγμή). Στο πρώτο μέρος του διαγράμματος παρουσιάζονται οι πιθανότητες που υπολόγισε το γραμμικό μοντέλο του lime. Στο τρίτο μέρος φαίνονται οι τιμές των χαρακτηριστικών για τον συγκεκριμένο ασθενή για τον οποίο εξάγεται η ερμηνεία της πιθανότητας που έδωσε το μοντέλο για την περίπτωση του. Στο μεσαίο και κυριότερο μέρος κατατάσσονται τα



Σχήμα 27: Ασθενής 2 - Ερμηνεία απόφασης για τον κίνδυνο επανανοσηλείας (χρονική στιγμή 2)



Σχήμα 28: Ασθενής 3 - Ερμηνεία απόφασης για τον κίνδυνο νοσηλείας (χρονική στιγμή 1)

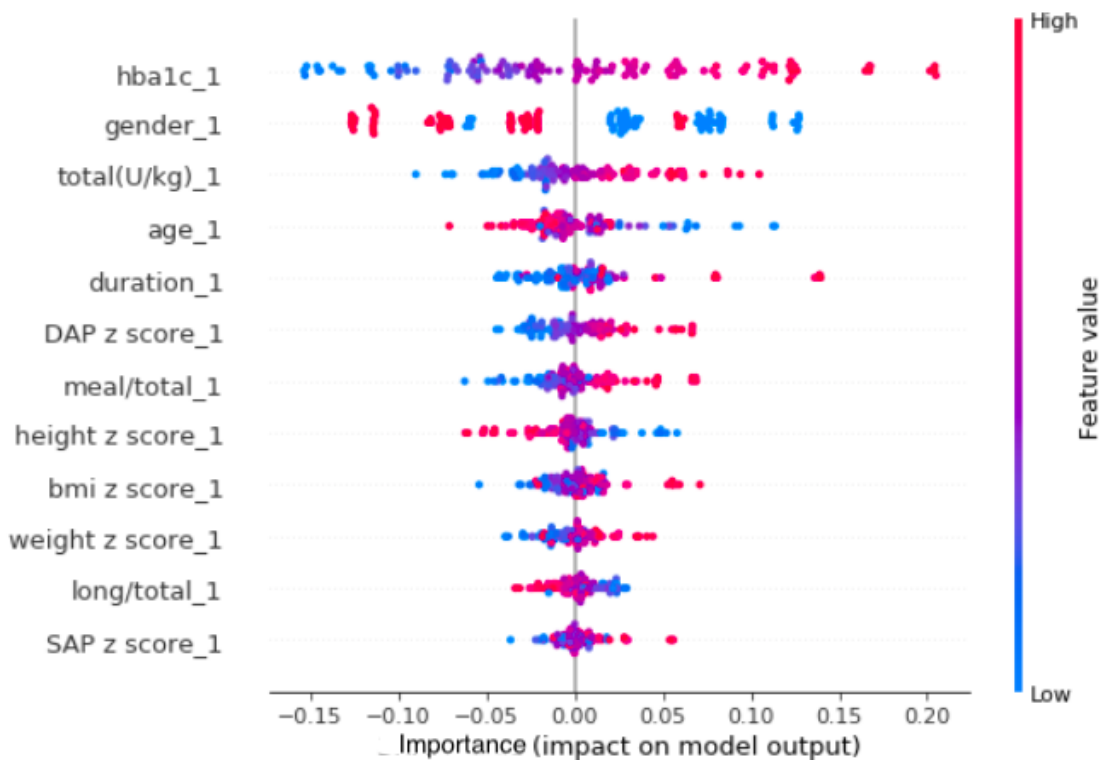


Σχήμα 29: Ασθενής 3 - Ερμηνεία απόφασης για τον κίνδυνο επανανοσηλείας (χρονική στιγμή 2)

χαρακτηριστικά με σειρά φθίνουσας σημαντικότητας. Για κάθε χαρακτηριστικό συμπεριλαμβάνεται μια οριοθέτηση ανάμεσα σε κάποιες τιμές οι οποίες αποτελούν την αιτία που η συγκεκριμένη τιμή του χαρακτηριστικού καθόρισε την τελική απόφαση. Ακριβώς κάτω από το όνομα κάθε χαρακτηριστικού βρίσκεται ο συντελεστής βαρύτητας που του δόθηκε. Έτσι μπορούμε να διακρίνουμε τα σημαντικότερα χαρακτηριστικά αλλά και να κατανοήσουμε ότι αφού η τιμή τους βρέθηκε ανάμεσα σε κάποια συγκεκριμένα όρια, αυτό είχε ως αποτέλεσμα να επηρεάζουν σε έναν συγκεκριμένο βαθμό την τελική πιθανότητα.

5.2 Συνολικά διαγράμματα σημαντικότητας χαρακτηριστικών

Στα διαγράμματα του σχήματος 27 και 28, στον κατακόρυφο άξονα είναι τοποθετημένα όλα τα χαρακτηριστικά σε φθίνουσα σειρά σημαντικότητας και στον οριζόντιο άξονα είναι η τιμή του συντελεστή βαρύτητας του συγκεκριμένου χαρακτηριστικού για τον εκάστοτε ασθενή. Κάθε κουκίδα αντιπροσωπεύει ένα δείγμα από το σύνολο δεδομένων, το χρώμα της δείχνει πόσο χαμηλή ή υψηλή τιμή έχει το χαρακτηριστικό στο οποίο την ευθεία είναι τοποθετημένη, ενώ η θέση του στον οριζόντιο άξονα υποδηλώνει την συνεισφορά του χαρακτηριστικού αυτού στο ενδεχόμενο νοσηλείας. Οι αρνητικές τιμές του οριζόντιου άξονα υποδηλώνουν θετική συνεισφορά στο ενδεχόμενο μη-νοσηλείας ή,



Σχήμα 30: Διάγραμμα σημαντικότητας χαρακτηριστικών για την εκτίμηση του κινδύνου νοσηλείας σε επίπεδο πληθυσμού

αλλιώς, αρνητική συνεισφορά στο ενδεχόμενο νοσηλείας.

Ακολουθούν δύο διαγράμματα: ένα για κάθε χρονική στιγμή. Αυτό που αφορά στην πρώτη χρονική στιγμή περιλαμβάνει 12 χαρακτηριστικά ενώ αυτό της δεύτερης χρονικής στιγμής περιλαμβάνει 24 καθώς λαμβάνει υπόψιν τα χαρακτηριστικά της πρώτης και της δεύτερης χρονικής στιγμής με τον αντίστοιχο δείκτη 1 ή 2, αντίστοιχα.

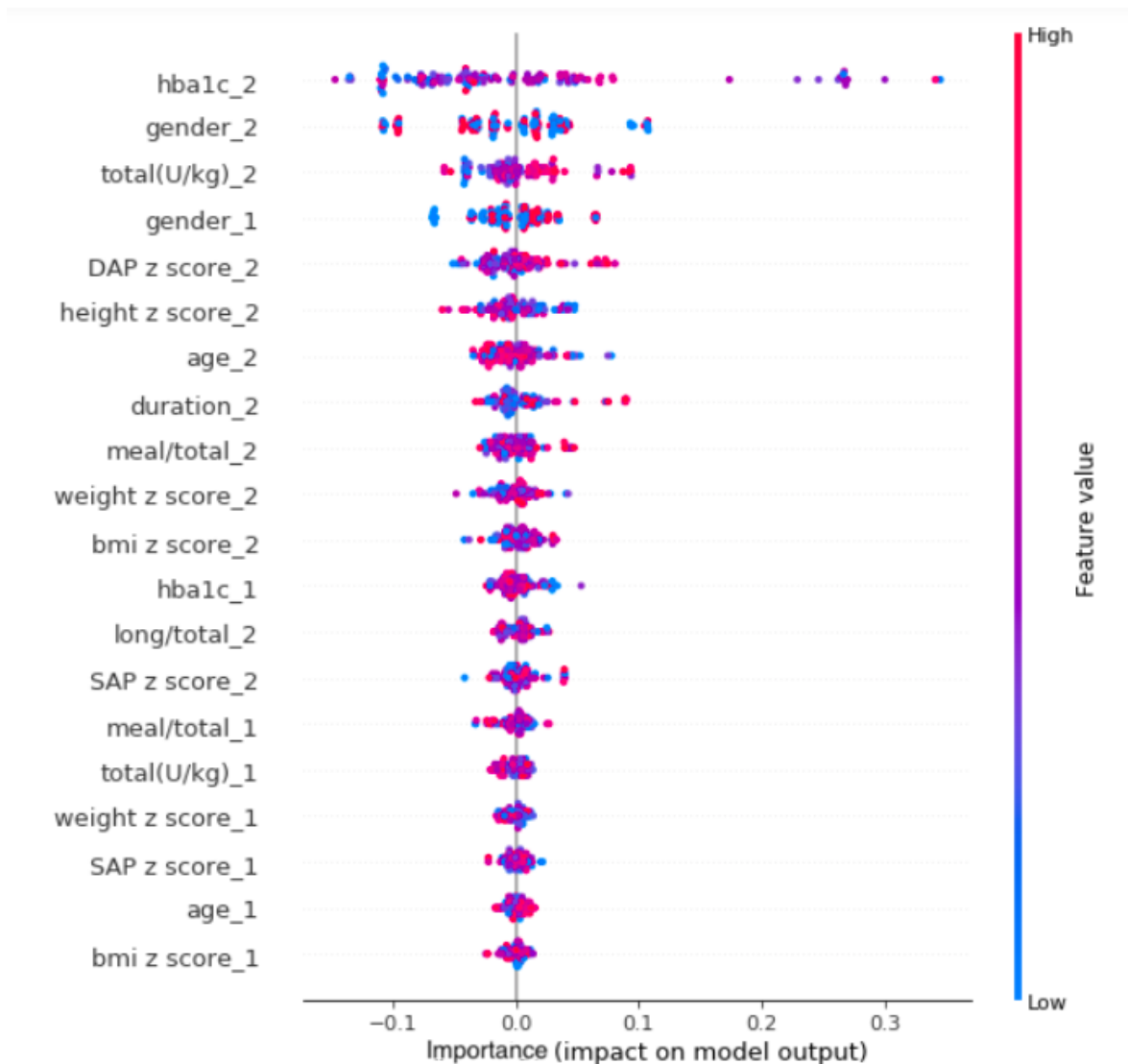
Από τα σχήματα γίνεται αντιληπτό ότι το καθοριστικότερο χαρακτηριστικό είναι η γλυκοζυλιωμένη αιμοσφαιρίνη, η οποία όταν λαμβάνει μεγάλες τιμές, δέχεται ταυτόχρονα και υψηλή θετική τιμή σημαντικότητας, που συνεπάγεται νοσηλεία, ενώ όταν λαμβάνει μικρές τιμές δέχεται υψηλή αρνητική τιμή σημαντικότητας που συνεπάγεται συνεισφορά στην μη-νοσηλεία. Τα παραπάνω συνάδουν πλήρως με τα γνωστά από την βιβλιογραφία.

Όσον αφορά στο φύλο που φαίνεται να είναι το δεύτερο σημαντικότερο χαρακτηριστικό, παρατηρείται στο πρώτο διάγραμμα ότι το ανδρικό φύλο οδηγείται κυρίως σε νοσηλεία ενώ το θηλυκό όχι. Από το δεύτερο διάγραμμα δεν φαίνεται να μπορεί να εξαχθεί κάποιο συμπέρασμα που να αφορά στο φύλο.

Στη συνέχεια παρατηρείται ότι η συνολική ημερήσια ποσότητα ινσουλίνης, όταν είναι υψηλή, οδηγεί σε νοσηλεία ενώ όταν είναι χαμηλή σε μη-νοσηλεία. Αυτό δεν συνάδει με όσα γνωρίζουμε για τον ΣΔ καθώς η διαβητική κετοξέωση αποτελεί μιά επιπλοκή, που οφείλεται στην έλλειψη της ινσουλίνης και όχι στο περίσσειμα αυτής.

Όσον αφορά στην ηλικία, φαίνεται κυρίως από το πρώτο διάγραμμα ότι οι νεότερες ηλικιακές ομάδες τείνουν να νοσηλεύονται περισσότερο σε σχέση με τις υψηλότερες ηλικιακά. Αυτό είναι λογικό καθώς τα παιδιά μεγαλύτερης ηλικίας αντιλαμβάνονται καλύτερα τους τρόπους διαχείρισης της ασθένειας.

Όσον αφορά στην χρονική διάρκεια που έχει εμφανιστεί ο ΣΔ θα αναμενόταν οι υψηλότερες τιμές να συνδέονται με μη-νοσηλεία και οι χαμηλότερες με νοσηλεία λόγω έλλειψης εμπειρίας στην διαχείριση της ινσουλίνης και συνολικά της ασθένειας. Ωστόσο, σύμφωνα με το πρώτο διάγραμμα,



Σχήμα 31: Διάγραμμα σημαντικότητας χαρακτηριστικών για την εκτίμηση του κινδύνου επανανοσηλείας σε επίπεδο πληθυσμού

οι ασθενείς με μικρότερη διάρκεια φαίνεται να μην νοσηλεύονται και αντίθετα ορισμένοι ασθενείς με υψηλές τιμές χρονικής διάρκειας να νοσηλεύονται.

Οι υψηλές τιμές διαστολικής πίεσης στον ασθενή συνδέονται συχνά με επιπλοκές λόγω του ΣΔ καθώς η συσσώρευση γλυκόζης στην κυκλοφορία του αίματος μπορεί να προκαλέσει βλάβες στα αιμοφόρα αγγεία και τα νεφρά, όργανα που παίζουν βασικό ρόλο στη διατήρηση της υγιούς αρτηριακής πίεσης. Τα διαγράμματα επιβεβαιώνουν ότι οι υψηλές τιμές πίεσης λαμβάνουν και υψηλές τιμές σημαντικότητας προς την κατεύθυνση της νοσηλείας.

Η τιμή της ινσουλίνης γέυματος προς την συνολική ημερήσια ποσότητα ινσουλίνης αναμένεται να συνδέεται με νοσηλεία όταν είναι χαμηλή ενώ με μη-νοσηλεία όταν είναι υψηλότερη. Αυτό, ωστόσο, δεν επαληθεύεται από τα διαγράμματα.

Το ύψος φαίνεται να παίζει σημαντικό ρόλο σε περιπτώσεις νοσηλείας όταν είναι χαμηλό, ενώ αντίθετα σχετίζεται με μη-νοσηλεία όταν είναι υψηλό. Αυτό πιθανότατα δικαιολογείται από το γεγονός ότι το ύψος σχετίζεται με τον δείκτη μάζας σώματος και την παχυσαρκία που αποτελεί καθοριστικά αρνητικό παράγοντα στην υγεία των ασθενών με ΣΔ.

Ο παράγοντας του δείκτη μάζας σώματος, σύμφωνα με το πρώτο διάγραμμα, φαίνεται να επιδρά όπως αναμενόταν καθώς όταν έχει υψηλή τιμή, σχετίζεται με νοσηλεία και λαμβάνει υψηλότερες

τιμές σημαντικότητας ενώ όταν είναι χαμηλός αποτελεί ένδειξη μη-νοσηλείας. Το ίδιο συμβαίνει και με το βάρος.

Όσον αφορά στον παράγοντα της δόσης ινσουλίνης μακράς δράσης προς την συνολική ημερήσια δόση ινσουλίνης, παρατηρείται ότι όταν αυτή η τιμή είναι χαμηλή, το χαρακτηριστικό αυτό λαμβάνει θετικές τιμές σημαντικότητας που συνδέονται με νοσηλεία, ενώ όταν είναι υψηλή συμβαίνει το αντίθετο. Το δεύτερο διάγραμμα δεν μας δίνει κάποιο σαφές συμπέρασμα όσον αφορά σε αυτό το χαρακτηριστικό.

Τέλος η συστολική αρτηριακή πίεση ανταποκρίνεται στα αναμενόμενα δεδομένης της βιβλιογραφίας καθώς όταν λαμβάνει υψηλές τιμές συνδέεται με νοσηλεία ενώ όταν λαμβάνει χαμηλές συνδέεται κυρίως με μη-νοσηλεία.

5.3 Συμπεράσματα-Μελλοντική έρευνα

Στο πλαίσιο της παρούσας διπλωματικής αναπτύχθηκε ένα συνδυαστικό μοντέλο πρόβλεψης βασισμένο στο αναδρομικό μοντέλο μακράς βραχείας μνήμης, οι αποφάσεις του οποίου ερμηνεύτηκαν μέσω της μεθόδου ερμηνευσιμότητας LIME. Τα αποτελέσματα έδειξαν ότι το μοντέλο έχει ικανοποιητική διακριτική ικανότητα και είναι ικανό να παράγει ακριβείς προβλέψεις, όπως φαίνεται από τα κριτήρια.

Ωστόσο, το γεγονός ότι ο ΣΔ, όπως αποδείχθηκε και από τις στατιστικές μελέτες, αφορά ένα όλο και μεγαλύτερο μέρος του πληθυσμού, σε συνδυασμό με την αυξανόμενη εξοικείωση των ιατρών με την τεχνολογία και την εμπιστοσύνη στα μοντέλα πρόβλεψης, καθιστά επιτακτική την ανάγκη περαιτέρω εξέλιξης και βελτιστοποίησης των μοντέλων πρόβλεψης που σχετίζονται με τον ΣΔ. Συγκεκριμένα για την παρούσα εργασία οι επιστημονικές επεκτάσεις που θα μπορούσαν να πραγματοποιηθούν είναι οι ακόλουθες:

- Να εξετασθούν οι ερμηνείες των προβλέψεων του μοντέλου από επαγγελματίες υγείας προκειμένου να αξιολογηθεί ο τρόπος λήψης αποφάσεων του μοντέλου.
- Να χρησιμοποιηθεί ένα νέο σύνολο δεδομένων ως είσοδος στο συνδυαστικό μοντέλο LSTM ώστε να αξιολογηθεί η απόδοσή του και να διορθωθούν οι ελλείψεις του. Επιπλέον ένα πλουσιότερο σύνολο δεδομένων με περισσότερα δείγματα αλλά και περισσότερες χρονικές στιγμές θα βοηθούσε το συγκεκριμένο μοντέλο να αναδείξει τις ικανότητές του και να εντοπίσει περισσότερες συσχετίσεις μεταξύ των χαρακτηριστικών και του κινδύνου νοσηλείας και επανανοσηλείας.
- Να εφαρμοστεί το ίδιο σύνολο δεδομένων εισόδου και σε διαφορετικά μοντέλα ώστε να συγκριθεί η απόδοση του σύνθετου μοντέλου LSTM που δημιουργήθηκε με άλλα μοντέλα μηχανικής μάθησης. Καθώς πρόκειται για σειριακά δεδομένα θα ήταν χρήσιμο να διαπιστωθεί εάν το LSTM αποδίδει καλύτερα στις περιπτώσεις τέτοιων δεδομένων.
- Τέλος, να εφαρμοστούν και άλλες μέθοδοι ερμηνευσιμότητας, όπως η μέθοδος SHAP, προκειμένου να συγκριθούν τα αποτελέσματα, ώστε να αξιολογηθεί η απόδοση της τεχνικής του LIME αλλά και να πραγματοποιηθεί μία εμβάθυνση στην εσωτερική λειτουργία του σύνθετου μοντέλου LSTM.

Βιβλιογραφία

- [1] The National Institute of Diabetes and Digestive and Kidney Diseases, *What is Diabetes?* 2016. [Online]. Available: <https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes>.
- [2] Wikipedia contributors, *Wikipedia, The Free Encyclopedia*, 2020. [Online]. Available: <https://en.wikipedia.org/wiki/Diabetes>.
- [3] International Diabetes Federation, *What is diabetes*, 2020. [Online]. Available: <https://www.idf.org/aboutdiabetes/what-is-diabetes.html>.
- [4] Stephen L. Aronoff, Kathy Berkowitz, Barb Shreiner and Laura Want, “Glucose Metabolism and Regulation: Beyond Insulin and Glucagon,” *Diabetes Spectrum*, vol. 17, no. 3, pp. 183–190, 2004.
- [5] Γ.Π. ΚΥΡΙΑΚΙΔΗΣ, Δ.Φ. ΦΛΩΡΟΣ, Γ.Χ. ΣΥΜΕΩΝΙΔΗΣ, *ΑΜΥΛΙΝΗ, ΠΡΑΜΑΙΝΤΙΑΗ ΣΑΚΧΑΡΩΔΗΣ ΔΙΑΒΗΤΗΣ*, 2013. [Online]. Available: <https://kkeram1441.wordpress.com/2013/07/17/%CE%B1%CE%BC%CF%85%CE%BB%CE%AF%CE%BD%CE%B7-%CF%83%CF%85%CF%83%CF%87%CE%AD%CF%84%CE%B9%CF%83%CE%B7-%CE%BC%CE%B5-%CF%83%CE%B1%CE%BA%CF%87%CE%B1%CF%81%CF%8E%CE%B4%CE%B7-%CE%B4%CE%B9%CE%B1%CE%B2%CE%AE/#more-615>.
- [6] Mayo Clinic Staff, *Diabetes symptoms: When diabetes symptoms are a concern*, 2019. [Online]. Available: <https://www.mayoclinic.org/diseases-conditions/diabetes/in-depth/diabetes-symptoms/art-20044248>.
- [7] —, *Diabetes Diagnosis*, 2020. [Online]. Available: <https://www.mayoclinic.org/diseases-conditions/diabetes/diagnosis-treatment/drc-20371451>.
- [8] International Diabetes Federation, *IDF DIABETES ATLAS*, 2019. [Online]. Available: <https://diabetesatlas.org/en/>.
- [9] British Diabetic Association, *Complications of diabetes*, 2020. [Online]. Available: <https://www.diabetes.org.uk/guide-to-diabetes/complications>.
- [10] American Diabetes Association, *Complications*, 2020. [Online]. Available: <https://www.diabetes.org/diabetes/complications>.
- [11] S. G. Mougiakakou, C. S. Bartsocas, E. Bozas, N. Chaniotakis, D. Iliopoulou, I. Kouris, S. Pavlopoulos, A. Prountzou, M. Skevofilakas, A. Tsoukalis, K. Varotsis, A. Vazeou, K. Zarkogianni, and K. S. Nikita, “SMARTDIAB: 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*, vol. 14, no. 3, pp. 622–633, 2010. doi: 10.1109/TITB.2009.2039711.
- [12] M. Skevofilakas, K. Zarkogianni, B. Karamanos, and K. Nikita, “A hybrid Decision Support System for the Risk Assessment of retinopathy development as a long term complication of Type 1 Diabetes Mellitus,” *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2010, pp. 6713–6, Aug. 2010. doi: 10.1109/IEMBS.2010.5626245.
- [13] K. Dalakleidi, K. Zarkogianni, K. Vassilis, A. Thanopoulou, and K. Nikita, “A Hybrid Genetic Algorithm for the Selection of Critical Features for Cardiovascular Complications Prognosis in Type 2 Diabetes Patients,” Nov. 2013. doi: 10.1109/BIBE.2013.6701620.

- [14] G. Fico, L. Hernandez, J. Cancela, A. Dagliati, L. Sacchi, A. Martinez-Millana, J. Posada, L. Manero, J. Verdu, A. Facchinetti, *et al.*, “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*, vol. 19, p. 163, 2019.
- [15] K. Zarkogianni, E. Litsa, A. Vazeou, and K. S. Nikita, “Personalized glucose-insulin metabolism model based on self-organizing maps for patients with Type 1 Diabetes Mellitus,” in *13th IEEE International Conference on BioInformatics and BioEngineering*, 2013, pp. 1–4. doi: 10.1109/BIBE.2013.6701604.
- [16] K. Zarkogianni, K. Mitsis, M. Arredondo, G. Fico, A. Fioravanti, and K. Nikita, “Neuro-fuzzy based glucose prediction model for patients with Type 1 diabetes mellitus,” Jun. 2014, pp. 252–255, isbn: 978-1-4799-2131-7. doi: 10.1109/BHI.2014.6864351.
- [17] K. Zarkogianni, K. Mitsis, E. Litsa, M. T. Arredondo, G. Fico, A. Fioravanti, and K. S. Nikita, “Comparative assessment of glucose prediction models for Patients with Type 1 Diabetes Mellitus applying sensors for glucose and physical activity monitoring,” *Medical and Biological Engineering and Computing*, vol. 53, 2015.
- [18] K. Zarkogianni, E. Litsa, K. Mitsis, P.-Y. Wu, C. Kaddi, C. Cheng, M. Wang, and K. Nikita, “A Review of Emerging Technologies for the Management of Diabetes Mellitus,” *IEEE transactions on bio-medical engineering*, vol. 62, Aug. 2015. doi: 10.1109/TBME.2015.2470521.
- [19] K. Dalakleidi, K. Zarkogianni, A. Thanopoulou, and K. Nikita, “Comparative assessment of statistical and machine learning techniques towards estimating the risk of developing type 2 diabetes and cardiovascular complications,” *Expert Systems*, 2017.
- [20] K. Zarkogianni, M. Athanasiou, A. C. Thanopoulou, and K. S. Nikita, “Comparison of machine learning approaches towards assessing the risk of developing Cardiovascular disease as a long-term diabetes complication,” *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1–1, 2018. doi: 10.1109/JBHI.2017.2765639.
- [21] K. Zarkogianni and K. S. Nikita, “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, 2015.
- [22] Jorge Castañón, *10 Machine Learning Methods that Every Data Scientist Should Know*, 2019. [Online]. Available: <https://towardsdatascience.com/10-machine-learning-methods-that-every-data-scientist-should-know-3cc96e0e0000>.
- [23] Alex Bottle and Paul Aylin and Azeem Majeed, “Identifying Patients at High Risk of Emergency Hospital Admissions: A Logistic Regression Analysis,” *Journal of the Royal Society of Medicine*, vol. 99, no. 8, pp. 406–414, 2006. [Online]. Available: <https://doi.org/10.1177/014107680609900818>.
- [24] Bergethon, Kristin E and Ju, Christine and DeVore, Adam D and Hardy, N Chantelle and Fonarow, Gregg C and Yancy, Clyde W and Heidenreich, Paul A and Bhatt, Deepak L and Peterson, Eric D and Hernandez, Adrian F, “Trends in 30-Day Readmission Rates for Patients Hospitalized With Heart Failure: Findings From the Get With The Guidelines-Heart Failure Registry,” *Circulation. Heart failure*, vol. 9, no. 6, 2016. [Online]. Available: <https://europepmc.org/articles/PMC4928632>.
- [25] O. Ben-Assuli and R. Padman, “Analysing repeated hospital readmissions using data mining techniques,” *Health Systems*, vol. 7, no. 2, pp. 120–134, 2018. [Online]. Available: <https://doi.org/10.1080/20476965.2017.1390635>.
- [26] L. E. Whan, “Selecting the Best Prediction Model for Readmission,” *J Prev Med Public Health*, vol. 45, no. 4, pp. 259–266, 2012. [Online]. Available: <http://www.jpmp.org/journal/view.php?number=1296>.

- [27] S. Cui, D. Wang, Y. Wang, P.-W. Yu, and Y. Jin, "An improved support vector machine-based diabetic readmission prediction," *Computer Methods and Programs in Biomedicine*, vol. 166, pp. 123–135, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0169260718308083>.
- [28] J. Zhang, S. W. Yoon, M. Khasawneh, K. Srihari, and S. Poranki, "Hospital readmission prediction using swarm intelligence-based support vector machines," *IIE Annual Conference and Expo 2013*, pp. 1522–1531, 2013.
- [29] Wikipedia contributors, *Wikipedia, The Free Encyclopedia*, 2020. [Online]. Available: https://en.wikipedia.org/wiki/Artificial_neural_network.
- [30] S. Mougiakakou, A. Proutzou, K. Zarkogianni, C. Bartsocas, K. Nikita, and A. Gerasimidi-Vazeou, "Prediction of glucose profile in children with type 1 diabetes mellitus using continuous glucose monitors and insulin pumps," *Hormone Research*, vol. 70, pp. 22–23, Jan. 2008.
- [31] M. Jamei, A. Nisnevich, E. Wetchler, S. Sudat, and E. Liu, "Predicting all-cause risk of 30-day hospital readmission using artificial neural networks," *PLOS ONE*, vol. 12, no. 7, pp. 1–14, Jul. 2017. [Online]. Available: <https://doi.org/10.1371/journal.pone.0181173>.
- [32] A. Ashfaq, A. Sant'Anna, M. Lingman, and S. Nowaczyk, "Readmission prediction using deep learning on electronic health records," *Journal of Biomedical Informatics*, vol. 97, p. 103 256, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1532046419301753>.
- [33] D. Échevin, Q. Li, and M.-A. Morin, "Hospital Readmission is Highly Predictable from Deep Learning," *Chaire de recherche Industrielle Alliance sur les enjeux économiques des changements démographiques*, Tech. Rep. 1705, 2017. [Online]. Available: <https://ideas.repec.org/p/lvl/criacr/1705.html>.
- [34] M. S. G. and N. K. S., "A Neural Network Approach for Insulin Regime and Dose Adjustment in Type 1 Diabetes," *Diabetes Technology & Therapeutics*, vol. 2, no. 3, pp. 381–389, 2000. doi: 10.1089/15209150050194251. eprint: <https://doi.org/10.1089/15209150050194251>. [Online]. Available: <https://doi.org/10.1089/15209150050194251>.
- [35] K. Zarkogianni, A. Vazeou, S. Mougiakakou, A. Proutzou, and K. Nikita, "An Insulin Infusion Advisory System Based on Autotuning Nonlinear Model-Predictive Control," *IEEE transactions on bio-medical engineering*, vol. 58, pp. 2467–77, May 2011. doi: 10.1109/TBME.2011.2157823.
- [36] G. Lemaitre, F. Nogueira, and C. K. Aridas, "Imbalanced-Learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 559–563, Jan. 2017, issn: 1532-4435.
- [37] Tara Boyle, *Dealing with Imbalanced Data*, 2019. [Online]. Available: <https://towardsdatascience.com/methods-for-dealing-with-imbalanced-data-5b761be45a18>.
- [38] Jialiang Jiang, *Interpretable Learning for Hospital Readmission Prediction from Healthcare Data*, 2020. [Online]. Available: <https://search.proquest.com/openview/29350f9131c77e941f08cca554f79160/1?cbl=18750&diss=y&loginDisplay=true&pq-origsite=gscholar#>.
- [39] C. Molnar, *Interpretable Machine Learning, A Guide for Making Black Box Models Explainable*. 2019, <https://christophm.github.io/interpretable-ml-book/>.
- [40] A. Altmann, L. Tološi, O. Sander, and T. Lengauer, "Permutation importance: a corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, Apr. 2010, issn: 1367-4803. doi: 10.1093/bioinformatics/btq134. eprint: <https://academic.oup.com/bioinformatics/article-pdf/26/10/1340/16892402/btq134.pdf>. [Online]. Available: <https://doi.org/10.1093/bioinformatics/btq134>.

- [41] G. J. Katuwal and R. Chen, *Machine Learning Model Interpretability for Precision Medicine*, 2016. arXiv: 1610.09045 [q-bio.QM].
- [42] A. Y. Zhang, S. S. W. Lam, N. Liu, Y. Pang, L. L. Chan, and P. H. Tang, “Development of a Radiology Decision Support System for the Classification of MRI Brain Scans,” in *2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT)*, 2018, pp. 107–115. doi: 10.1109/BDCAT.2018.00021.
- [43] X. Wang, D. Wang, Z. Yao, B. Xin, B. Wang, C. Lan, Y. Qin, S. Xu, D. He, and Y. Liu, “Machine Learning Models for Multiparametric Glioma Grading With Quantitative Result Interpretations,” *Frontiers in Neuroscience*, vol. 12, p. 1046, 2019, issn: 1662-453X. doi: 10.3389/fnins.2018.01046. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnins.2018.01046>.
- [44] X. Liu, P. Hu, Z. Mao, P.-C. Kuo, P. Li, C. Liu, J. Hu, D. Li, D. Cao, R. G. Mark, L. A. Celi, Z. Zhang, and F. Zhou, *Interpretable Machine Learning Model for Early Prediction of Mortality in Elderly Patients with Multiple Organ Dysfunction Syndrome (MODS): a Multicenter Retrospective Study and Cross Validation*, 2020. arXiv: 2001.10977 [physics.med-ph].
- [45] M. Athanasiou, K. Sfrintzeri, K. Zarkogianni, A. C. Thanopoulou, and K. S. Nikita, *An explainable XGBoost-based approach towards assessing the risk of cardiovascular disease in patients with Type 2 Diabetes Mellitus*, 2020. arXiv: 2009.06629 [physics.med-ph].
- [46] Christopher Olah, *Understanding LSTM Networks*, 2015. [Online]. Available: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- [47] Brandon Rohrer, *Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM)*, 2017. [Online]. Available: <https://www.youtube.com/watch?v=WCUNPb-5EYI&t=134s>.
- [48] Thomas Lin Pedersen Michaël Benesty, *Understanding lime*, 2019. [Online]. Available: https://cran.r-project.org/web/packages/lime/vignettes/Understanding_lime.html.
- [49] Abhishek Sharma, *Decrypting your Machine Learning model using LIME*, 2018. [Online]. Available: <https://towardsdatascience.com/decrypting-your-machine-learning-model-using-lime-5adc035109b5>.
- [50] Sharayu Rane, *LIME: Explaining predictions of machine learning models*, 2019. [Online]. Available: <https://towardsdatascience.com/lime-explaining-predictions-of-machine-learning-models-1-2-1802d56addf9>.