

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

**ΣΧΟΛΗ ΕΠΙΣΤΗΜΗΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΠΛΗΡΟΦΟΡΙΚΗΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

ΠΡΟΓΡΑΜΜΑ ΔΙΔΑΚΤΟΡΙΚΩΝ ΣΠΟΥΔΩΝ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

**Ανάπτυξη στοχαστικών αλγορίθμων και μοντέλων
μηχανικής μάθησης για την εξέλιξη πληθυσμών: Εφαρμογή
στο δάκο της Ελιάς**

Ρωμανός Γ. Καλαματιανός-Βούροβ

ΚΕΡΚΥΡΑ

ΟΚΤΩΒΡΙΟΣ 2021

IONIAN UNIVERSITY

**FACULTY OF INFORMATION SCIENCE & INFORMATICS
DEPARTMENT OF INFORMATICS**

PROGRAM OF DOCTORAL STUDIES

PhD THESIS

**Development of stochastic algorithms and machine learning
models for the evolution of populations: Application on the olive
fruit fly**

Romanos G. Kalamatianos-Vourov

CORFU

ΟΚΤΩΒΡΙΟΣ 2021

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Ανάπτυξη στοχαστικών αλγορίθμων και μοντέλων μηχανικής μάθησης για την εξέλιξη
πληθυσμών: Εφαρμογή στο δάκο της Ελιάς

Ρωμανός Γ. Καλαματιανός-Βούροβ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: Μάρκος Αυλωνίτης, Αναπληρωτής Καθηγητής Τμήμα Πληροφορικής,
Ιόνιο Πανεπιστήμιο

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ:

Μάρκος Αυλωνίτης, Αναπληρωτής Καθηγητής Τμήμα Πληροφορικής, Ιόνιο Πανεπιστήμιο

Κάτια-Λήδα Κερμανίδου, Αναπληρώτρια Καθηγήτρια Τμήμα Πληροφορικής, Ιόνιο
Πανεπιστήμιο

Πάυλος Μπουχάγιερ, Επίκουρος Καθηγητής Τμήμα Επιστήμης και Τεχνολογίας
Τροφίμων, Ιόνιο Πανεπιστήμιο

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Μάρκος Αυλωνίτης,

**Αναπληρωτής Καθηγητής Τμήμα
Πληροφορικής, Ιόνιο Πανεπιστήμιο**

**Κάτια-Λήδα Κερμανίδου,
Αναπληρώτρια Καθηγήτρια
Τμήμα Πληροφορικής, Ιόνιο
Πανεπιστήμιο**

Πάυλος Μπουχάγιερ,

**Επίκουρος Καθηγητής Τμήμα Επιστήμης
και Τεχνολογίας Τροφίμων, Ιόνιο
Πανεπιστήμιο**

**Αναστάσιος Καλημέρης,
Αναπληρωτής Καθηγητής
Τμήμα Περιβάλλοντος, Ιόνιο
Πανεπιστήμιο**

Εμμανουήλ Μαραγκουδάκης,

**Καθηγητής Τμήμα Πληροφορικής, Ιόνιο
Πανεπιστήμιο**

**Ιωάννης Καρύδης,
Επίκουρος Καθηγητής
Τμήμα Πληροφορικής, Ιόνιο
Πανεπιστήμιο**

Σπυριδών Σιούτας,

**Καθηγητής Τμήμα Μηχανικών Η/Υ &
Πληροφορικής, Πανεπιστήμιο Πατρών**

Ημερομηνία Εξέτασης: 14 Δεκεμβρίου 2021

PhD THESIS

Development of stochastic algorithms and machine learning models for the evolution of populations:
Application on the olive fruit fly

Romanos G. Kalamatianos-Vourov

SUPERVISOR: **Markos Avlonitis**, Associate Professor Department of Informatics, Ionian University

THREE-MEMBER ADVISORY COMMITTEE:

Markos Avlonitis, Associate Professor Department of Informatics, Ionian University

Katia - Lida Kermanidis, Associate Professor Department of Informatics, Ionian University

Pavlos Bouchagier, Assistant Professor Department of Food Science and Technology, Ionian University

SEVEN-MEMBER EXAMINATION COMMITTEE

Markos Avlonitis,
Associate Professor Department of Informat-
ics, Ionian University

Katia - Lida Kermanidis,
Associate Professor Department
of Informatics, Ionian University

Pavlos Bouchagier,
Assistant Professor Department of Food Sci-
ence and Technology, Ionian University

Anastasios Kalimeris,
Associate Professor Department
of Environment, Ionian Univer-
sity

Emmanouil Maragoudakis,
Professor Department of Informatics, Ionian
University

Ioannis Karydis,
Assistant Professor Department
of Informatics, Ionian University

Spyridon Sioutas,
Professor Department of Computer Engineer-
ing and Informatics, University of Patras

Examination Date: 14 December 2021

ΠΕΡΙΛΗΨΗ

Στην παρούσα διατριβή γίνεται ανάπτυξη στοχαστικών αλγορίθμων και μοντέλων μηχανικής μάθησης για την πρόβλεψη εξέλιξης πληθυσμών και ειδικότερα για την εξέλιξη του πληθυσμού του δάκου της ελιάς. Ο δάκος της ελιάς είναι ένα παρασιτικό έντομο που εξαρτάται από τον καρπό της ελιάς για την διαίωσή του είδους του, με αποτέλεσμα να προκαλεί μεγάλες ζημιές, αν ο πληθυσμός του μείνει ανεξέλεγκτος, στην παραγωγή προϊόντων με βάση τον καρπό της ελιάς, όπως ελαιόλαδο και επιτραπέζιες ελιές. Η παρουσία του εντόμου έχει καταγραφεί σε όλες τις ελαιοπαραγωγικές χώρες και λόγω της εξάρτησης του από τις κλιματικές συνθήκες (θερμοκρασία, υγρασία) ο συνολικός αριθμός γενεών που προκύπτουν ποικίλλει από περιοχή σε περιοχή. Σε αυτό το πλαίσιο, ως προς τους στοχαστικούς αλγορίθμους, αρχικά παρουσιάζεται το λογισμικό προσομοίωσης της χρονικής και χωρικής εξέλιξης του πληθυσμού του δάκου βάσει πραγματικών μετρήσεων στο πεδίο. Στην συνέχεια, περιγράφεται η ανάπτυξη και σύγκριση τριών μοντέλων διασποράς του δάκου με βάση την καρποφορία του ελαιώνα και τέλος, μελετάτε η επιρροή της παρουσίας μικροκλιμάτων στην εξέλιξη του πληθυσμού του δάκου μέσω προσομοίωσης, με χρήση του προαναφερθέντος λογισμικού, και πως, υπό αυτό το πρίσμα, επηρεάζεται η εξέλιξη του πληθυσμού δάκου με την εφαρμογή των συνήθων πρακτικών ελέγχου του πληθυσμού του. Συνεχίζοντας, ως προς την χρήση μεθόδων μηχανικής μάθησης, αρχικά, μελετάται η χρήση των περιβαλλο-ντολογικών συνθηκών για την πρόβλεψη της επόμενης μέτρησης δακοπαγίδας, δεδομένης της γνώσης της προηγούμενης μέτρησης και συγκρίνεται η απόδοση τεσσάρων χαρακτηριστικών διανυσμάτων με χρήση διαφόρων αλγορίθμων μηχανικής μάθησης. Στην συνέχεια, παρουσιάζεται η ανάπτυξη μοντέλου για την αυτόματη αναγνώριση του δάκου, και κατά συνέπεια αυτόματης καταμέτρησης, σε εικόνες του περιεχομένου της παγίδας McPhail που χρησιμοποιείται συνήθως ως μέθοδος ελέγχου και δείκτης έξαρσης του πληθυσμού του δάκου. Τέλος, μελετάται η χρήση μεθόδων μηχανικής μάθησης για την αναγνώριση μικροκλιμάτων, γεγονός που θα βοηθούσε στην πρόβλεψη εστίων προσβολής του ελαιώνα από τον δάκο ανάλογα την εποχή του χρόνου και τα κλιματολογικά χαρακτηριστικά του κάθε μικροκλίματος. Τέλος, εξάγονται τα συμπεράσματα για την απόδοση των προαναφερθέντων πειραμάτων και πιθανά σημεία που θα μπορούσε να εστιάσει τυχόν μελλοντική έρευνα.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Στοχαστική προσομοίωση και μηχανική μάθηση

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: δάκος της ελιάς, μηχανική μάθηση, στοχαστικοί αλγόριθμοι

ABSTRACT

In the present dissertation, stochastic algorithms and machine learning models are developed for the prediction of population evolution and in particular for the evolution of the olive fruit fly population. The olive fruit fly is a parasitic insect that depends on the fruit of the olive tree for the perpetuation of its species, thus causing great damage, if its population remains unchecked, in the production of olive-based products, such as olive oil and table olives. The presence of the insect has been recorded in all olive-producing countries, and due to its dependence on climatic conditions (temperature, humidity) the total number of generations that arise varies from region to region. In this context, as far as stochastic algorithms are concerned, the software for simulating the temporal and spatial evolution of the population of the olive fruit fly based on real measurements in the field is initially presented. Then, the development and comparison of three models of dispersion of the olive fruit fly based on the fruiting of the olive grove is described and finally, the influence of the presence of microclimates, on the evolution of the population of the olive fruit fly, is studied, through simulation, using the aforementioned software, and how, in this context, the evolution of the olive fruit fly population is affected by the application of the usual practices of controlling the population. Continuing, in the use of machine learning methods, initially, the use of environmental conditions to predict the next olive fruit fly trap measurement, given the knowledge of the previous measurement, and a comparison of the performance of four characteristic vectors using various machine learning algorithms is examined. The development of a model for the automatic identification of the olive fruit, and therefore automatic counting, in images of the contents of the McPhail trap, which is commonly used as a method of control and an exacerbation indicator of the population of the olive fruit fly, is presented. Finally, the use of machine learning methods for the identification of microclimates, is examined, which would help to predict outbreaks of infestation of the olive fruit fly in the olive grove depending on the time of year and the climatic characteristics of each microclimate. In conclusion, conclusions are drawn about the performance of the aforementioned experiments and possible directions are pointed out that future research could focus on.

SUBJECT AREA: Stochastic simulation and machine learning

KEYWORDS: olive fruit fly, machine learning, stochastic algorithms

ΠΕΡΙΕΧΟΜΕΝΑ

1	ΕΙΣΑΓΩΓΗ	29
1.1	Μοντέλα στη Βιολογία	29
1.1.1	Σπουδαιότητα Μοντέλων στη Βιολογία	29
1.1.2	Μαθηματικά Μοντέλα στη Βιολογία	29
1.1.2.1	Πληθυσμιακά Μοντέλα Ενός Είδους	30
	Μαλθουσιανό – Εκθετικό Μοντέλο	30
	Εξίσωση Gompertz	31
	Λογιστικό Μοντέλο	31
1.2	Στοχαστική προσομοίωση	33
1.2.1	Σύστημα	33
1.2.1.1	Τύποι συστημάτων	34
1.2.2	Μοντέλο	34
1.2.3	Προσομοίωση	35
1.2.4	Πλεονεκτήματα & Μειονεκτήματα	36
	Πλεονεκτήματα	36
	Μειονεκτήματα	36
1.2.5	Βήματα προσομοίωσης	37
1.2.6	Προσομοίωση εξέλιξης πληθυσμών	37
1.2.6.1	Μοντέλα προσομοίωσης πληθυσμού του δάκου της ελιάς	38
1.3	Μηχανική μάθηση	39
1.3.1	Χαρακτηριστικό διάνυσμα	40
1.3.1.1	Χρησιμότητα	40
1.3.1.2	Παράδειγμα	40
1.3.2	Είδη μάθησης	40
1.3.2.1	Επιβλεπόμενη μάθηση	41
1.3.2.2	Μη-επιβλεπόμενη μάθηση	41
1.3.2.3	Reinforcement learning	42
1.3.3	Το πρόβλημα της ανισορροπίας κλάσης	42

1.3.4	Μέθοδοι αξιολόγησης	43
1.3.4.1	Σύνολα εκπαίδευσης & αξιολόγησης	43
1.3.4.2	Σύνολα εκπαίδευσης, αξιολόγησης & δοκιμής	43
1.3.4.3	k-fold cross validation	43
1.3.5	Μετρικές απόδοσης	43
1.3.5.1	Πίνακας σύγχησης (Confusion matrix)	44
1.3.5.2	Ακρίβεια (Precision)	45
1.3.5.3	Ανάκληση (Recall)	45
1.3.5.4	F1-score	46
1.3.6	Deep Learning	46
1.3.7	Νευρωνικά Δίκτυα	46
1.3.7.1	Νευρώνας	46
1.3.7.2	Συνάψεις	47
1.3.7.3	Είδη νευρωνικών δικτύων	48
	Feed-forward neural network (FFNN)	48
	Recurrent neural network (RNN)	48
	Convolutional neural network (CNN)	48
1.4	Μοντέλο Degree day	49
1.4.1	Μέθοδοι υπολογισμού	50
1.4.1.1	Μέση θερμοκρασία	50
1.4.1.2	Μέθοδος μονού τριγώνου	51
1.4.1.3	Μέθοδος διπλού τριγώνου	52
1.4.1.4	Ημιτονοειδής καμπύλη	52
1.4.1.5	Διπλή ημιτονοειδής καμπύλη	53

2 ΕΛΙΑ ΚΑΙ ΔΑΚΟΣ 55

2.1	Ελαιόδενδρο (Olea europaea)	55
2.1.1	Προϊόντα	55
2.1.2	Οφέλη	55
2.2	Δάκος	57
2.2.1	Δυναμική πληθυσμού	58
2.2.2	Χαρακτηριστικά πληθυσμου	59

2.2.2.1	Αναλογία φύλου	59
2.2.2.2	Ζευγάριωμα	59
2.2.2.3	Διάχυση	59
2.2.3	Επιπτώσεις στην ελαιοκαλλιέργεια	59
2.2.4	Τρόποι αντιμετώπισης	60
2.2.4.1	Παγίδες	60
	McPhail παγίδες	60
	OLIPE (Olivarera los Pedroches) παγίδες	60
	Κίτρινες παγίδες κόλλας	61
	Συσκευή "Attract & kill"	61
	Ηλεκτρονικές/αυτόματες παγίδες	62
2.2.4.2	Ψεκασμοί	64
	Πηλός Καολίνης	64
	Δόλωμα Spinosad	64
2.3	Μικροκλίματα	64
2.3.1	Σχετική έρευνα για την αναγνώριση μικροκλιμάτων	64
2.3.2	Επιρροή μικροκλιμάτων στο δάκο της ελιάς	66
3	ΕΦΑΡΜΟΓΕΣ ΠΡΟΣΟΜΟΙΩΣΗΣ	67
3.1	Λογισμικό προσομοίωσης Δάκου	67
3.1.1	Μονάδες	67
3.1.1.1	Grid module	68
3.1.1.2	Trap module	69
3.1.1.3	Populate module	70
3.1.1.4	Simulation module	70
	Δομή πληθυσμού	71
	Ανάπτυξη	71
	Αναπαραγωγή	72
	Διασπορά	72
	Έξοδος	72
3.1.1.5	Grouping module	72
3.1.1.6	Plot module	73

3.1.1.7	Display module	73
3.1.2	Χαρακτηριστικά	73
3.1.3	Μελλοντικές βελτιώσεις	75
3.2	Ανάπτυξη μοντέλου διασποράς βάση της καρποφορίας	75
3.2.1	Μοντέλα διασποράς	76
3.2.1.1	Μοντέλο 1	76
3.2.1.2	Μοντέλο 2	76
3.2.1.3	Μοντέλο 3	78
3.2.2	Σενάρια προσομοίωσης	79
3.2.3	Αποτελέσματα	80
3.2.3.1	Μοντέλο 1	80
3.2.3.2	Μοντέλο 2	81
3.2.3.3	Μοντέλο 3	83
3.2.4	Αξιολόγηση αποτελεσμάτων	85
	Μοντέλο 1	86
	Μοντέλο 2	86
	Μοντέλο 3	87
3.2.5	Συμπεράσματα	89
3.3	Επίδραση μικροκλιμάτων	90
3.3.1	Περιοχή μελέτης	90
3.3.2	Δεδομένα	91
3.3.3	Σενάρια προσομοίωσης	93
3.3.4	Αποτελέσματα	93
3.3.5	Συμπεράσματα	95

4 ΕΦΑΡΜΟΓΕΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ 99

4.1	Ανάπτυξη μοντέλου για την πρόβλεψη της επόμενης μέτρησης μιας δακοπαγίδας	99
4.1.1	Σχετική Έρευνα	99
4.1.1.1	Αλγόριθμοι μηχανικής μάθησης	99
4.1.1.2	Έρευνα σχετικά με την πρόβλεψη προσβολής από τον δάκο .	102
	Πλεονεκτήματα	102
	Μειονεκτήματα	103

4.1.2	Συλλογή δεδομένων	103
4.1.2.1	Περιβαλλοντολογικά δεδομένα	103
4.1.2.2	Επιλογή χαρακτηριστικών	103
4.1.2.3	Εξαγωγή χαρακτηριστικού διανύσματος	105
4.1.2.4	Το πρόβλημα της ανισορροπίας κλάσης	106
4.1.3	Αξιολόγηση απόδοσης	106
4.1.3.1	Πειραματική διάταξη	107
4.1.3.2	Αποτελέσματα	108
	Μέθοδος 1	108
	Μέθοδος 2	111
	Μέθοδος 3	113
	Μέθοδος 4	115
4.1.3.3	Συμπεράσματα	117
4.2	Ανάπτυξη μοντέλου για την αυτόματη αναγνώριση δάκου σε εικόνες	117
4.2.1	Το toolkit	120
4.2.1.1	Σύνολο δεδομένων	121
4.2.2	Πειραματική αξιολόγηση	123
4.2.2.1	Πειραματική διάταξη	123
4.2.2.2	Αποτελέσματα	124
4.2.3	Συμπεράσματα	128
4.3	Ανάπτυξη μοντέλου για την αγνώριση μικροκλιμάτων	130
4.3.1	Αλγόριθμοι μηχανικής μάθησης	130
4.3.2	Συλλογή δεδομένων	131
4.3.2.1	Περιβαλλοντολογικά δεδομένα	131
4.3.2.2	Προ-επεξεργασία δεδομένων	131
4.3.2.3	Επιλογή χαρακτηριστικών	135
4.3.2.4	Εξαγωγή χαρακτηριστικού διανύσματος	137
4.3.3	Αξιολόγηση απόδοσης	139
4.3.3.1	Πειραματική διάταξη	139
4.3.3.2	Αποτελέσματα	139
4.3.4	Συμπεράσματα	141

5	Επίλογος	143
5.1	Μελλοντική έρευνα	144
	REFERENCES	157

ΛΙΣΤΑ ΕΙΚΟΝΩΝ

1.1	Χρονική εξέλιξη πληθυσμού με βάση το Μαλθουσιανό μοντέλο για διαφορετικές τιμές αύξησης r και αρχικό πληθυσμό 10 άτομα.	31
1.2	Χρονική εξέλιξη πληθυσμού με βάση την εξίσωση Gompertz για $b = 1$, $c = 0,1$ και N_0 τα 10 άτομα.	32
1.3	Χρονική εξέλιξη πληθυσμού με βάση το μοντέλο του Verhulst για διαφορετικές τιμές αύξησης r , $K = 500$ και N_0 τα 10 άτομα.	33
1.4	Σύστημα.	33
1.5	Μεταβολή (α) διακριτής και (β) συνεχούς καταστατικής μεταβλητής στο χρόνο.	34
1.6	Παράδειγμα φυσικού μοντέλου ηλιακού συστήματος.	35
1.7	Βήματα προσομοίωσης συστήματος.	37
1.8	Γενικευμένο μοντέλο μηχανικής μάθησης.	39
1.9	Παράδειγμα της μεθόδου k -fold cross validation για k ίσο με πέντε.	44
1.10	Μοντέλο νευρώνα σε τεχνητό νευρωνικό δίκτυο.	47
1.11	Παράδειγμα Feed-forward neural network	48
1.12	Παράδειγμα Recurrent neural network	49
1.13	Ρυθμός ανάπτυξης σε σχέση με την αύξηση της θερμοκρασίας με άνω και κάτω όριο.	50
1.14	Μέθοδος μονού τριγώνου για τον υπολογισμό των Degree Day.	51
1.15	Μέθοδος ημιτονοειδούς καμπύλης για τον υπολογισμό των Degree Day. Η γκρίζα περιοχή υποδηλώνει την θερμότητα που θα συνεισφέρει στην ανάπτυξη του υπό εξέταση οργανισμού.	52
2.1	Τα προϊόντα που παράγονται από τα ελαιόδεντρα: (α) ελαιόλαδο και (β) επιτραπέζιες ελιές	56
2.2	(α) Θυληκός και (β) αρσενικός δάκος	57
2.3	Καρπός του ελαιόδεντρου	57
2.4	Παγίδα τύπου McPhail εγκατεστημένη σε ελαιώνα	61
2.5	Περιεχόμενο παγίδας McPhail	61
2.6	Παγίδα τύπου OLIPE από Morini33 ^a έχει άδεια βάσει CC BY 3.0	62
3.1	Δείγμα εισόδου επιπέδου ελαιώνων	68

3.2	Δείγμα εισόδου επιπέδου καρποφορίας των ελαιώνων	69
3.3	Δείγμα εισόδου επιπέδου μικροκλιμάτων	70
3.4	Παράδειγμα εξόδου (a) για μηδενική προσβολή του ελαιώνα, (b) για υψηλή προσβολή του ελαιώνα	73
3.5	Σχετικές συχνότητες δάκων σε έναν ελαιώνα χωρίς ελαιόδεντρα μεταξύ 4 – 10 ημερών μετά την αποφυλάκιση. Αποτελέσματα πειραμάτων από τους Fletcher and Karatos [1981]	76
3.6	Σχετικές συχνότητες δάκων σε έναν ελαιώνα με ποσοστό καρποφορίας 30% μεταξύ 4 – 6 ημερών μετά την αποφυλάκιση. Αποτελέσματα πειραμάτων από τους Fletcher and Karatos [1981]	77
3.7	Γράφημα Εξίσωσης 3.3 με d_i ίσο με 60 και T_M ίσο με 35.	77
3.8	Σύγκριση συναρτήσεων μοντέλου (a) 2 και (b) 3.	78
3.9	Πιθανότητα επιλογής κατεύθυνσης σε σχέση με το ποσοστό καρποφορίας. . . .	79
3.10	Σχετική συχνότητα δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Fletcher and Karatos [1981] για 0% παρουσία φρούτων ελιάς για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 1.	80
3.11	Σχετική συχνότητα δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Fletcher and Karatos [1981] για 30% παρουσία φρούτων ελιάς για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 1.	81
3.12	Χωρική κατανομή των δάκων για 0% καρποφορίας στον ελαιώνα για χρονικό διάστημα μιας εβδομάδας, για το μοντέλο διασποράς 2.	82
3.13	Σχετική συχνότητα δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Fletcher and Karatos [1981] για 0% παρουσία φρούτων ελιάς για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 2.	82
3.14	Σχετική συχνότητα δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Fletcher and Karatos [1981] για 30% παρουσία φρούτων ελιάς για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 2.	83
3.15	Χωρική κατανομή των δάκων για 0% καρποφορίας στον ελαιώνα για χρονικό διάστημα μιας εβδομάδας, για το μοντέλο διασποράς 3.	83
3.16	Σχετική συχνότητα δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Fletcher and Karatos [1981] για 0% παρουσία φρούτων ελιάς για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 3.	84
3.17	Χωρική κατανομή των δάκων για 30% καρποφορίας στον ελαιώνα για χρονικό διάστημα μιας εβδομάδας, για το μοντέλο διασποράς 3.	85
3.18	Σχετική συχνότητα δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Fletcher and Karatos [1981] για 30% παρουσία φρούτων ελιάς για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 3.	85

3.19	Σύγκριση δείκτη αξιολόγησης R των τριών μοντέλων διασποράς για τις τρεις εξεταζόμενες περιπτώσεις καρποφορίας.	88
3.20	Σύγκριση δείκτη αξιολόγησης Bias των τριών μοντέλων διασποράς για τις τρεις εξεταζόμενες περιπτώσεις καρποφορίας.	88
3.21	Σύγκριση δείκτη αξιολόγησης RMSE των τριών μοντέλων διασποράς για τις τρεις εξεταζόμενες περιπτώσεις καρποφορίας.	89
3.22	Μέση απόσταση πτήσης δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Rempoulakis and Nestel [2012], για 40% καρποφορίας για χρονικό διάστημα τριών ημερών.	89
3.23	Χάρτης της περιοχής μελέτης στην Κέρκυρα.	90
3.24	Κατακερματισμός περιοχής μελέτης σε μικροκλίματα. Το μπλε χρώμα είναι το μικροκλίμα A, το κόκκινο χρώμα είναι το μικροκλίμα B, το πράσινο χρώμα είναι το μικροκλίμα C και το κίτρινο χρώμα είναι το μικροκλίμα D.	91
3.25	Τοποθεσίες περιβαλλοντικών αισθητήρων στην περιοχή μελέτης	92
3.26	Μέσες ημερήσιες θερμοκρασίες για κάθε μικροκλίμα.	92
3.27	(a) Εξέλιξη πληθυσμού του δάκου σε κάθε μικροκλίμα χωρίς ψεκασμό και κατανομή πληθυσμού (b) στην έναρξη της προσομοίωσης και (c) στο τέλος της προσομοίωσης	94
3.28	(a) Εξέλιξη πληθυσμού του δάκου σε κάθε μικροκλίμα μετά από έναν καθολικό ψεκασμό και κατανομή πληθυσμού (b) στην έναρξη της προσομοίωσης και (c) στο τέλος της προσομοίωσης	95
3.29	Σύγκριση εξέλιξης πληθυσμού για το μικροκλίμα A χωρίς και με ψεκασμό. . .	96
3.30	Σύγκριση εξέλιξης πληθυσμού για το μικροκλίμα B χωρίς και με ψεκασμό. . .	96
3.31	Σύγκριση εξέλιξης πληθυσμού για το μικροκλίμα C χωρίς και με ψεκασμό. . .	97
3.32	Σύγκριση εξέλιξης πληθυσμού για το μικροκλίμα D χωρίς και με ψεκασμό. . .	97
4.1	Σύγκριση Ακρίβειας και Ανάκλησης για πέντε αλγόριθμους ταξινόμησης, για την πειραματική μέθοδο 1	109
4.2	Το δένδρο απόφασης που κατασκευάστηκε από τον αλγόριθμο J48.	109
4.3	Ακρίβεια και Ανάκληση του αλγορίθμου IBk για διαφορετικούς k γείτονες, για την πειραματική μέθοδο 1	110
4.4	Ακρίβεια και Ανάκληση του αλγορίθμου Multilayer Perceptron για διαφορετικό αριθμό κόμβων, για την πειραματική μέθοδο 1	111
4.5	Κατανομή τιμών της κλάσης ταξινόμησης, για την πειραματική μέθοδο 1 . . .	111
4.6	f1 score για όλους τους αλγόριθμους για κάθε τιμή ταξινόμησης, για την πειραματική μέθοδο 1	112

4.7	f1 score για όλους τους αλγόριθμους για κάθε τιμή ταξινόμησης, για την πειραματική μέθοδο 2	112
4.8	Ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο IBk, για την πειραματική μέθοδο 2	113
4.9	Ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο Multilayer Perceptron, για την πειραματική μέθοδο 2	114
4.10	f1 score για όλους τους αλγόριθμους για κάθε τιμή ταξινόμησης, για την πειραματική μέθοδο 3	114
4.11	Ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο IBk, για την πειραματική μέθοδο 3	115
4.12	Ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο Multilayer Perceptron, για την πειραματική μέθοδο 3	116
4.13	f1 score για όλους τους αλγόριθμους για κάθε τιμή ταξινόμησης, για την πειραματική μέθοδο 4	116
4.14	Ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο IBk, για την πειραματική μέθοδο 4	117
4.15	Ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο Multilayer Perceptron, για την πειραματική μέθοδο 4	118
4.16	Σύγκριση Ακρίβειας, ανάκλησης και f1 score μεταξύ των καλύτερων αποτελεσμάτων κάθε μεθόδου	118
4.17	Συλλογή δεδομένων με την εφαρμογή e-Olive.	119
4.18	Διεπαφή σε μορφή ιστοσελίδας του <i>DIRT</i>	120
4.19	Διάγραμμα ροής δημιουργίας και πειραματισμού με το <i>DIRT</i>	121
4.20	Κατανομή των διαστάσεων εικόνων του <i>DIRT</i>	122
4.21	Δείγμα εικόνων από το σύνολο δεδομένων.	122
4.22	Κατανομή εικόνων ανα αριθμό δάκων μέσα σε μία εικόνα.	123
4.23	Δείγμα εικόνων από το σύνολο δεδομένων με ετικετοποίηση	123
4.24	Σύγκριση απόδοσης μοντέλων μετά την εκπαίδευση στο σύνολο δεδομένων εικόνων με δάκους.	125
4.25	Απόδοση μοντέλου 1 για διαφορετικές διάστασεις των εικόνων στο σύνολο δεδομένων.	126
4.26	Απόδοση μοντέλου 1 για διαφορετικές διάστασεις των εικόνων στο σύνολο δεδομένων, μετά την μετατροπή τους στη γκρι χρωματική κλίμακα.	127
4.27	Σύγκριση απόδοσης μεταξύ εικόνων RGB και γκρι κλίμακας για διαφορετικές διαστάσεις των εικόνων.	127

4.28 Σύγκριση ακριβείας ανίχνευσης για σύνολα δεδομένων με βάση τον αριθμό των δάκων.	128
4.29 Τοποθεσίες αισθητήρων	132
4.30 Καταγεγραμμένη θερμοκρασία για τις παραλιακές τοποθεσίες.	133
4.31 Καταγεγραμμένη υγρασία για τις παραλιακές τοποθεσίες.	133
4.32 Καταγεγραμμένη θερμοκρασία για τις τοποθεσίες σε υψώματα.	134
4.33 Καταγεγραμμένη υγρασία για τις τοποθεσίες σε υψώματα.	134
4.34 Καταγεγραμμένη θερμοκρασία για τις τοποθεσίες σε γούβες.	135
4.35 Καταγεγραμμένη υγρασία για τις τοποθεσίες σε γούβες.	135
4.36 Αναλογία κέρδους πληροφορίας των προς εξέταση χαρακτηριστικών	137
4.37 Συσχέτιση των προς εξέταση χαρακτηριστικών με την κλάση ταξινόμησης. . .	138
4.38 Κέρδος πληροφορίας των προς εξέταση χαρακτηριστικών.	138
4.39 Αριθμός συστάδων που χώρισαν τα δεδομένα οι αλγόριθμοι συσταδοποίησης .	140
4.40 Ακρίβεια σωστής συσταδοποίησης για διάφορους αλγόριθμους συσταδοποίησης.	140

ΛΙΣΤΑ ΠΙΝΑΚΩΝ

1.1	Χαρακτηριστικό διάνυσμα οχημάτων	40
1.2	Τιμές χαρακτηριστικών για έξι μοντέλα οχημάτων	41
1.3	Χαρακτηριστικό διάνυσμα παραδείγματος 1.3.1.2 για επιβλεπόμενη μάθηση . .	41
1.4	Πίνακας σύγκρισης για πρόβλημα δυαδικής ταξινόμησης	45
3.1	Δείκτες αξιολόγησης για μοντέλο διασποράς 1.	86
3.2	Δείκτες αξιολόγησης για μοντέλο διασποράς 2.	87
3.3	Δείκτες αξιολόγησης για μοντέλο διασποράς 3.	87
4.1	Hardware configuration.	124
4.2	Στοιχεία εκπαίδευσης μοντέλων.	125
4.3	Σύνολα δεδομένων με βάση τον αριθμό δάκων.	128
4.4	Τοποθεσίες συλλογής δεδομένων	132

1. ΕΙΣΑΓΩΓΗ

Σκοπός αυτού του κεφαλαίου είναι η εισαγωγή στο πεδίο έρευνας της παρούσας διδακτορική διατριβής. Αναλυτικότερα στην ενότητα 1.1 γίνεται μια αναφορά στα μαθηματικά μοντέλα που χρησιμοποιούνται στην βιολογία με ιδιαίτερη έμφαση στα πληθυσμιακά μοντέλα. Στην συνέχεια στην ενότητα 1.2 γίνεται μια εισαγωγή σε βασικές έννοιες σχετικά με την μοντελοποίηση και προσομοίωση και αναφορά εφαρμογών προσομοίωσης στην εξέλιξη πληθυσμών. Στην ενότητα 1.3 γίνεται μια περιγραφή της μηχανικής μάθησης και βασικών εννοιών, Τέλος, στην ενότητα 1.4 παρουσιάζεται το μοντέλο Degree Day καθώς και μερικές μέθοδοι υπολογισμού του.

1.1 Μοντέλα στη Βιολογία

1.1.1 Σπουδαιότητα Μοντέλων στη Βιολογία

Τα μαθηματικά μοντέλα μπορούν να δώσουν άμεσες εκτιμήσεις για πραγματικά βιολογικά προβλήματα, δεδομένων κάποιων υποθέσεων, χρησιμοποιώντας μαθηματικά εργαλεία.

Σκοπός είναι:

- Η μελέτη βιολογικών συστημάτων με χρήση μαθηματικών εργαλείων, και
- Η ανάπτυξη χώρο-χρονικών μαθηματικών μοντέλων για τη μελέτη βιολογικών συστημάτων

1.1.2 Μαθηματικά Μοντέλα στη Βιολογία

Τα μαθηματικά μοντέλα στη βιολογία διακρίνονται, μεταξύ άλλων, ανάλογα με το βιολογικό σύστημα/πρόβλημα το οποίο μελετάνε, σε:

- Πληθυσμιακά μοντέλα: Μοντέλα που περιγράφουν την εξέλιξη του αριθμού (πληθυσμών) βιολογικών ειδών στο χρόνο.
- Μοντέλα Διάχυσης, τα οποία περιγράφουν τη δυναμική εξέλιξη των μεταβλητών σε συγκεκριμένο χώρο.

Ειδικότερα τα πληθυσμιακά μοντέλα χωρίζονται σε [Murray, 2007]:

- Μοντέλα ενός πληθυσμού. Περιγράφουν τη χρονική εξέλιξη του πληθυσμού ενός είδους δίχως την παρουσία κάποιου άλλου είδους που να το επηρεάζει.
- Μοντέλα αλληλεπιδρώντων πληθυσμών. Πρόκειται για μοντέλα που περιγράφουν την χρονική εξέλιξη του πληθυσμού ενός είδους που αλληλεπιδρά με κάποιο άλλο είδος, για παράδειγμα μοντέλα τύπου θηρευτή-θηράματος.

Τα προαναφερθέντα πληθυσμιακά μοντέλα διακρίνονται σε συνεχή και διακριτά.

1.1.2.1 Πληθυσμιακά Μοντέλα Ενός Είδους

Μαλθουσιανό – Εκθετικό Μοντέλο Πρόκειται για το πιο απλό πληθυσμιακό μοντέλο πληθυσμιακής αύξησης που εισήγαγε ο Malthus [2018] το οποίο καταλήγει σε εκθετική αύξηση του πληθυσμού. Το συγκεκριμένο μοντέλο βασίζεται στην υπόθεση ότι ο πληθυσμός ενός είδους έχει απεριόριστες προμήθειες στην διάθεση του.

Έστω $N(t)$ ο πληθυσμός ενός είδους και θεωρώντας ότι δεν υπάρχουν φυσικοί εχθροί και υπάρχει πάντα διαθέσιμο φαγητό, ο ρυθμός μεταβολής του πληθυσμού εξαρτάται από τον ρυθμό γεννήσεων και θανάτων και η γενική μορφή του έχει ως εξής:

$$\frac{dN(t)}{dt} = \text{Births} - \text{Deaths} \pm \text{Migration} \quad (1.1)$$

Κάνοντας τις ακόλουθες υποθέσεις:

- Οι γεννήσεις περιγράφονται με ένα γραμμικό όρο, ως προς τον αριθμό του πληθυσμού N , bN , όπου $b > 0$.
- Οι θάνατοι περιγράφονται με ένα γραμμικό όρο, ως προς τον αριθμό του πληθυσμού N , dN , όπου $0 < d < 1$.
- Ο ρυθμός εισαγωγής ατόμων είναι ίσος με τον ρυθμό εξαγωγής και ως εκ τούτου η μετανάστευση δεν επηρεάζει την εξέλιξη του πληθυσμού N .

Με βάση τις παραπάνω υποθέσεις και ορίζοντας τον ρυθμό αύξησης r , ως $r = b - d$, η εξίσωση 1.1 γράφεται ως εξής:

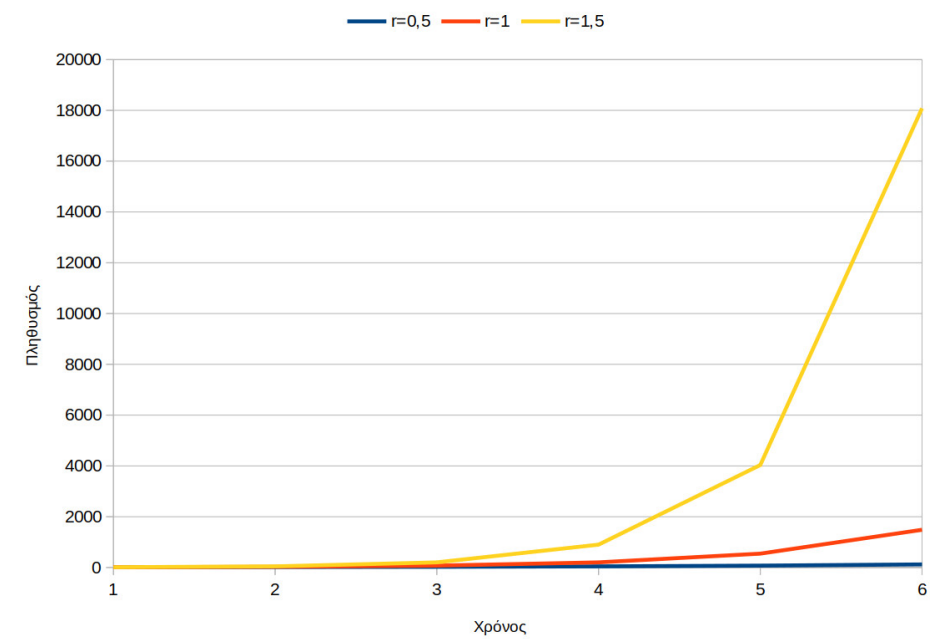
$$\frac{dN(t)}{dt} = (b - d)N(t) = rN(t) \quad (1.2)$$

η λύση της παραπάνω διαφορικής εξίσωσης είναι η ακόλουθη, όπου N_0 ο αρχικός πληθυσμός:

$$N(t) = N_0 e^{rt} \quad (1.3)$$

Η Εικόνα 1.1 παρουσιάζει της γραφική αναπαράσταση της εξίσωσης 1.3 για διάφορες τιμές του ρυθμού αύξησης r και με αρχικό πληθυσμό τα 10 άτομα. Όπως φαίνεται και από τα τρία παραδείγματα η αύξηση του πληθυσμού είναι συνεχής και εκθετική, ενώ όσο ο χρόνος αυξάνεται ο πληθυσμός τείνει προς το άπειρο.

Στην φύση, το παρόν μοντέλο δεν μπορεί να εφαρμοστεί καθώς η απεριόριστη αύξηση του πληθυσμού ενός είδους σε μεγάλα χρονικά διαστήματα δεν μπορεί συμβεί λόγω περιοριστικών παραγόντων στο περιβάλλον που δραστηροποιείται. Οι παράγοντες που περιορίζουν την εξέλιξη του πληθυσμού ενός είδους κατηγοροποιούνται στις εξής δύο κατηγορίες:



Εικόνα 1.1: Χρονική εξέλιξη πληθυσμού με βάση το Μαλθουσιανό μοντέλο για διαφορετικές τιμές αύξησης r και αρχικό πληθυσμό 10 άτομα.

- Πύκνο-εξαρτόμενοι. Πρόκειται για παράγοντες (ασθένειες, ανταγωνισμός κ.α.) που επηρεάζουν τον πληθυσμό όσο η πυκνότητα του μεγαλώνει με την πάροδο του χρόνου [Rockwood, 2015a].
- Πύκνο-ανεξάρτητοι. Πρόκειται για παράγοντες (καιρικές συνθήκες, φυσικές καταστροφές κ.α.) που θα επηρεάσουν τον πληθυσμό ανεξαρτήτως της πυκνότητας του [Rockwood, 2015b].

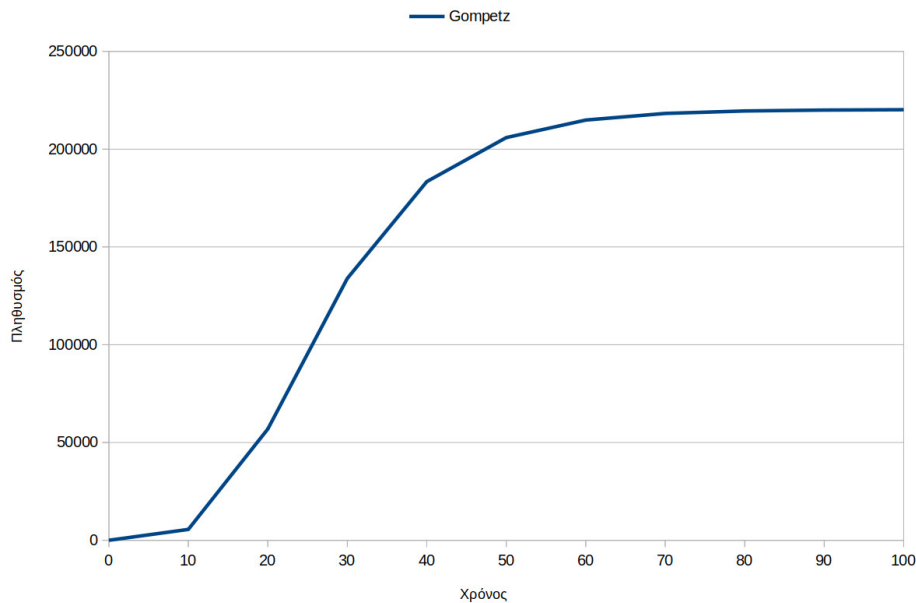
Εξίσωση Gompertz Ο Gompertz [1825] επέκτεινε το Μαλθουσιανό μοντέλο για να περιγράψει τον ρυθμό μεταβολής του πληθυσμού των ανθρώπων υπό πιο ρεαλιστικές συνθήκες. Η υπόθεση που εισήγαγε ο Gompertz είναι ότι η αντίσταση στο θάνατο ενός ατόμου μειώνεται με την πάροδο του χρόνου και το μοντέλο του έχει την εξής μορφή:

$$N(t) = N_0 e^{\frac{b}{c}(1-e^{-ct})} \quad (1.4)$$

όπου N_0 ο αρχικός πληθυσμός και b, c είναι σταθερές.

Η Εικόνα 1.2 παρουσιάζει της γραφική αναπαράσταση της εξίσωσης 1.4 για $b = 1, c = 0,1$ και N_0 τα 10 άτομα. Όπως φαίνεται ο πληθυσμός αρχίζει να αυξάνεται μετά από μικρό χρονικό διάστημα όμως στην συνέχεια δείχνει να σταθεροποιείται.

Λογιστικό Μοντέλο Το Μαλθουσιανό μοντέλο υποθέτει την εξέλιξη του πληθυσμού ενός είδους κάτω από ιδανικές και όχι τόσο ρεαλιστικές συνθήκες. Επιπλέον, δεν είναι τόσο ακρι-



Εικόνα 1.2: Χρονική εξέλιξη πληθυσμού με βάση την εξίσωση Gompertz για $b = 1$, $c = 0,1$ και N_0 τα 10 άτομα.

βές για μεγάλα χρονικά διαστήματα. Για τον λόγο αυτό ο Verhulst [1838] στην προσπάθεια του να βελτιώσει το Μαλθουσιανό μοντέλο, ώστε να μελετήσει την εξέλιξη του πληθυσμού των ανθρώπων, εισήγαγε την ιδιότητα ο ρυθμός μεταβολής του πληθυσμού να μηδενίζεται μετά από μεγάλη πάροδο του χρόνου.

Η εξίσωση που πρότεινε είναι η ακόλουθη:

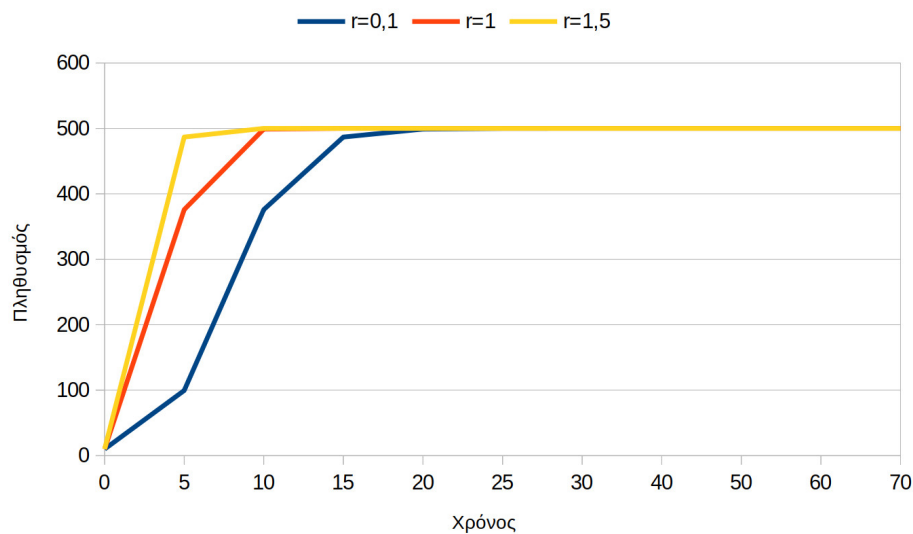
$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \quad (1.5)$$

όπου K είναι η χωρητικότητα του περιβάλλοντος και r ο ρυθμός αύξησης του πληθυσμού.

Η λύση της παραπάνω διαφορικής εξίσωσης είναι η ακόλουθη:

$$N(t) = \frac{N_0 e^{rt}}{1 + \frac{N_0(e^{rt}-1)}{K}} \quad (1.6)$$

Η Εικόνα 1.3 παρουσιάζει της γραφική αναπαράσταση της εξίσωσης 1.6 για διαφορετικές τιμές αύξησης r , $K = 500$ και N_0 τα 10 άτομα. Όπως φαίνεται, όσο μεγαλώνει ο ρυθμός αύξησης r τόσο πιο γρήγορα φτάνει ο πληθυσμός την χωρητικότητα του περιβάλλοντος του.

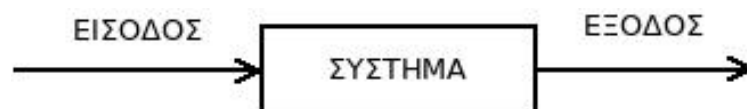


Εικόνα 1.3: Χρονική εξέλιξη πληθυσμού με βάση το μοντέλο του Verhulst για διαφορετικές τιμές αύξησης r , $K = 500$ και N_0 τα 10 άτομα.

1.2 Στοχαστική προσομοίωση

1.2.1 Σύστημα

Ως σύστημα ορίζεται *“μια συλλογή από οντότητες π.χ, άνθρωποι ή μηχανές, που δρουν και αλληλεπιδρούν μαζί προς την επίτευξη κάποιου λογικού τέλους”* [Law, 2015a]. Η επίδραση που δέχεται από το περιβάλλον το σύστημα ορίζεται ως είσοδο ενώ η αντίδραση του συστήματος σε αυτήν την επίδραση ορίζεται ως έξοδος του συστήματος (Εικόνα 1.4). Για παράδειγμα, ως σύστημα μπορεί να οριστεί ένα όχημα το οποίο μία από τις εισόδους που δέχεται από το περιβάλλον είναι η βενζίνη, την οποία χρησιμοποιεί για να επιτευχθεί η κίνηση του οχήματος και μία από τις εξόδους που παράγει είναι οι ρύποι, από την καύση της βενζίνης, που εξέρχονται από την εξάτμιση του οχήματος. Για την κατανόηση της κατάστασης στην οποία βρίσκεται ένα σύστημα γίνεται χρήση ενός συνόλου μεταβλητών, γνωστές και ως καταστατικές μεταβλητές, οι οποίες περιγράφουν το σύστημα σε μια χρονική στιγμή [Law, 2015a].

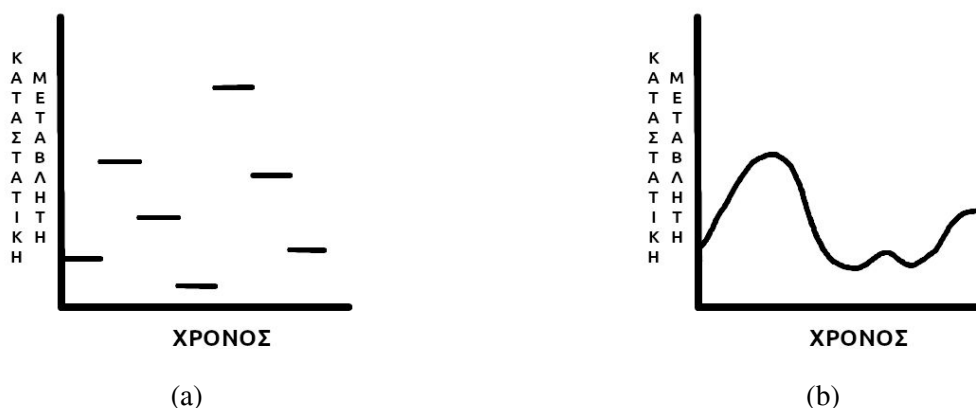


Εικόνα 1.4: Σύστημα.

Τα συστήματα κατηγοριοποιούνται στις ακόλουθες δύο κατηγορίες [Law, 2015a]:

- Διακριτά, όπου οι καταστατικές μεταβλητές μεταβάλλονται απότομα στο χρόνο (Εικόνα 1.5a). Η άφιξη και η αναχώρηση πελατών σε μια ουρά εξυπηρέτησης είναι ένα παράδειγμα καταστατικής μεταβλητής σε ένα διακριτό σύστημα.

- Συνεχή, όπου οι καταστατικές μεταβλητές μεταβάλλονται συνεχώς με την πάροδο του χρόνου (Εικόνα 1.5b). Ένα τέτοιο σύστημα είναι για παράδειγμα η κίνηση ενός οχήματος, όπου η ταχύτητα, η θέση και η επιτάχυνση του αλλάζουν συνεχώς.



Εικόνα 1.5: Μεταβολή (a) διακριτής και (b) συνεχής καταστατικής μεταβλητής στο χρόνο.

Η μελέτη ενός συστήματος, ανάλογα την πολυπλοκότητα του, μπορεί να επιτευχθεί είτε με την παρατήρηση/πειραματισμό του ίδιου του συστήματος είτε με τη δημιουργία ενός μοντέλου αυτού [Law, 2015a].

1.2.1.1 Τύποι συστημάτων

Με βάση την γνώση που υπάρχει για τις διεργασίες που πραγματοποιούνται μεταξύ των οντοτήτων ενός συστήματος, ένα σύστημα μπορεί να χαρακτηριστεί με έναν από τους ακόλουθους τρεις τύπους:

- Μαύρο κουτί. Στον τύπο αυτό η είσοδος και η έξοδος του συστήματος είναι γνωστό, όμως οι διεργασίες εντός του συστήματος είναι άγνωστες.
- Γκρι κουτί. Στον τύπο αυτό η είσοδος και η έξοδος του συστήματος είναι γνωστές, ενώ υπάρχει καποιά πληροφορία για τις εσωτερικές διεργασίες του συστήματος
- Λευκό κουτί. Στο τύπο αυτό η είσοδος και η έξοδος του συστήματος είναι γνωστές και υπάρχει πλήρη γνώση για τις εσωτερικές διεργασίες του συστήματος.

1.2.2 Μοντέλο

Μοντέλο ορίζεται ως μια αναπαράσταση ενός συστήματος. Ένα μοντέλο μπορεί να είναι είτε φυσικό (Εικόνα 1.6) είτε μαθηματικό, για παράδειγμα η εξίσωση 1.7 αναπαριστά τον ρυθμό μείωσης ενός πληθυσμού.

$$\frac{dP}{dt} = kP \quad (1.7)$$

όπου k ο ρυθμός μείωσης.

Η λύση ενός μαθηματικού μοντέλου μπορεί να επιτευχθεί είτε με αναλυτικό τρόπο, με την επίλυση της μαθηματικής εξίσωσης, είτε με την προσομοίωση του εν λόγω μοντέλου. Η τελική επιλογή μεταξύ των δύο προαναφερθέντων τρόπων έγκειται στην πολυπλοκότητα του μοντέλου και στο κόστος επίλυσης του με τον αναλυτικό τρόπο [Law, 2015a].

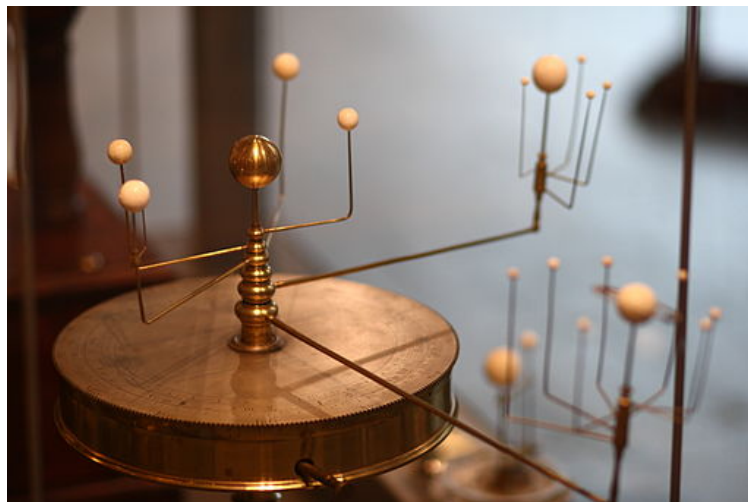
1.2.3 Προσομοίωση

Η προσομοίωση είναι η λειτουργία του μοντέλου υπό χρονική ή/και χωρική κλίμακα με στόχο την ανάλυση της απόδοσης του συστήματος.

Η προσομοίωση ενός μοντέλου μπορεί να κατηγοριοποιηθεί ως εξής [Law, 2015a]:

- Στατική ή δυναμική. Στη δυναμική προσομοίωση μοντέλου, προσομοιώνεται η εξέλιξη του συστήματος στην πάροδο του χρόνου, σε αντίθεση με την στατική προσομοίωση όπου η προσομοίωση του συστήματος γίνεται για μια συγκεκριμένη χρονική στιγμή.
- Ντετερμινιστική ή στοχαστική. Στη ντετερμινιστική προσομοίωση δεν υφίσταται η έννοια της τυχαιότητας και ως εκ τούτου για το ίδιο σύνολο δεδομένων εισόδου η προσομοίωση του μοντέλου θα παράγει το ίδιο αποτέλεσμα όσες φορές και αν επαναληφθεί. Η εκτέλεση μιας μόνο προσομοίωσης είναι αρκετή για αξιολόγηση του αποτελέσματος. Η

⁰a) https://commons.wikimedia.org/wiki/File:Planetarium_in_Putnam_Gallery_2,_2009-11-24.jpg,
b) <https://creativecommons.org/licenses/by-sa/3.0>



Εικόνα 1.6: Παράδειγμα φυσικού μοντέλου ηλιακού συστήματος.
Sage Ross^{1a}, CC BY-SA 3.0^{1b}, μέσω Wikimedia Commons

προσομοίωση της κίνησης των πλανητών είναι ένα παράδειγμα ντετερμινιστικής προσομοίωσης. Στον αντίποδα, στη στοχαστική προσομοίωση υπάρχει κάποια μορφή τυχαιότητας μέσα σε αυτά, με αποτέλεσμα η προσομοίωση του μοντέλου για το ίδιο σύνολο δεδομένων εισόδου θα δώσει διαφορετικό αποτέλεσμα σε κάθε επανάληψη εκτέλεσης της προσομοίωσης. Συνεπώς χρειάζονται αρκετές εκτελέσεις τις προσομοίωσης για να αξιολογηθεί το αποτέλεσμα. Σχετικό παράδειγμα, αν και απλό, είναι η προσομοίωση ρίψης ενός ζαριού, όπου χρειάζεται μεγάλος αριθμός προσομοιώσεων για να διαπιστωθεί πως το αποτέλεσμα ρίψης ενός ζαριού είναι ισοπίθανο.

- Συνεχή ή διακριτή. Ο συγκεκριμένος διαχωρισμός είναι παρόμοιος με τον διαχωρισμό σε συνεχή και διακριτά συστήματα (βλ. υπό-ενότητα 1.2.1).

1.2.4 Πλεονεκτήματα & Μειονεκτήματα

Η χρήση μοντελοποίησης και προσομοίωσης έχει τα ακόλουθα πλεονεκτήματα και μειονεκτήματα.

Πλεονεκτήματα Μερικά από τα πλεονεκτήματα είναι τα εξής [Law, 2015b]:

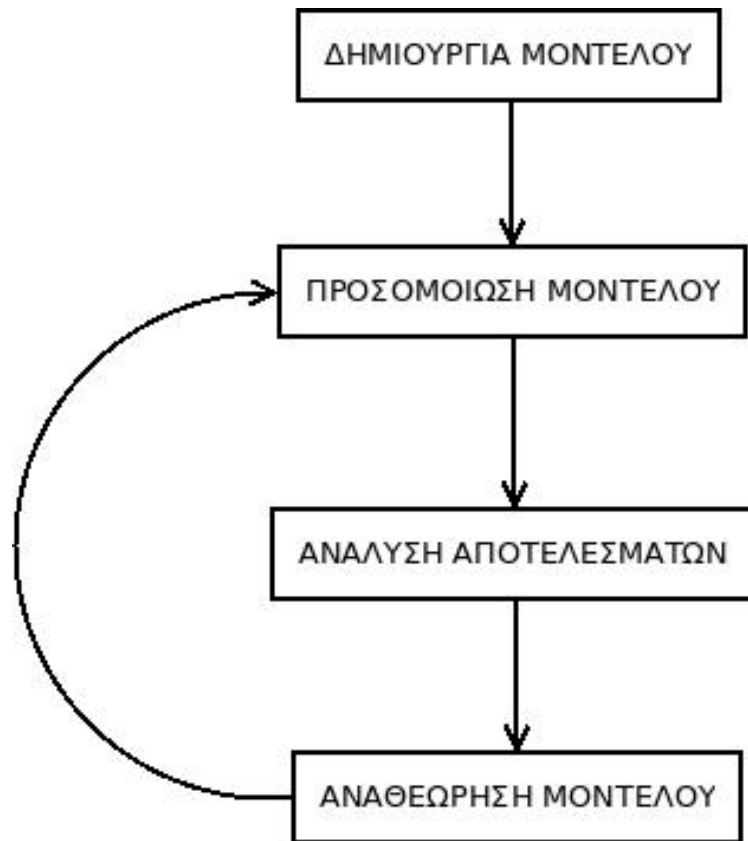
- Ιδανικά για μοντελοποίηση πολύπλοκων συστημάτων. Όταν το μαθηματικό μοντέλο που περιγράφει ένα σύστημα είναι πολύπλοκο με στοχαστικά στοιχεία η χρήση της προσομοίωσης είναι η μόνη όδος.
- Ιδανικά για πειραματισμούς. Μέσω της προσομοίωσης είναι εύκολη η πραγματοποίηση αλλαγών στο σύστημα και η παρατήρηση της επίδρασης που θα έχουν οι αλλαγές αυτές χωρίς να χρειάζεται να πειραχτεί το υπάρχον φυσικό σύστημα.
- Έλεγχος παραμέτρων και συνθηκών. Στην προσομοίωση ενός συστήματος ο ερευνητής έχει τον πλήρη έλεγχο των παραμέτρων/συνθηκών του περιβαλλόντος.
- Δυνατότητα μακροπρόθεσμής μεταβολής του συστήματος. Μέσω της προσομοίωσης μπορεί ο ερευνητής να μελετήσει ένα σύστημα σε βάθος χρόνου.
- Σύγκριση μεταξύ διαφορετικών αρχιτεκτονικών του συστήματος.

Μειονεκτήματα Μερικά από τα μειονεκτήματα είναι τα ακόλουθα [Law, 2015b]:

- Αρκετές φορές η ανάπτυξη ενός μοντέλου προσομοίωσης είναι χρονοβόρα διαδικασία
- Η εκτέλεση μιας στοχαστικής προσομοίωσης παράγει μόνο μια εκτίμηση της κατάστασης ενός συστήματος για ένα συγκεκριμένο σύνολο παραμέτρων εισόδου.

1.2.5 Βήματα προσομοίωσης

Για την πραγματοποίηση προσομοίωσης ενός συστήματος ακολουθούντε τα ακόλουθα βήματα (Εικόνα 1.7):



Εικόνα 1.7: Βήματα προσομοίωσης συστήματος.

1. Δημιουργία μοντέλου. Αρχικά δημιουργείται το μοντέλο που θα αναπαραστήσει το προς εξέταση σύστημα.
2. Προσομοίωση του μοντέλου. Εκτελούνται πειράματα προσομοίωσης με διάφορα σύνολα εισόδου.
3. Ανάλυση αποτελεσμάτων. Τα αποτελέσματα που προκύπτουν από τα πειράματα προσομοίωσης αναλύονται και διαπιστώνεται η εγκυρότητα τους.
4. Αναθεώρηση μοντέλου. Σε περίπτωση μη εγκυρότητας των αποτελεσμάτων προσομοίωσης, το μοντέλο αναθεωρείται και η διαδικασία ξεκινάει ξανά από το δεύτερο βήμα.

1.2.6 Προσομοίωση εξέλιξης πληθυσμών

Παρακάτω αναφέροντε μερικά παραδείγματα από την βιβλιογραφία σχετικά με τη χρήση μοντέλων προσομοίωσης για την εξέλιξη πληθυσμών για διάφορα είδη.

Οι Miller et al. [2015] ανέπτυξαν ένα μοντέλο προσομοίωσης (GrowLS) για την εξέλιξη του πληθυσμού λιονταριών σε μικρές περιοχές διατήρησης με απώτερο σκοπό να προβλέπει το λιγότερο δυνατό πλεόνασμα λιονταριών ύστερα από την εφαρμογή διαφόρων τεχνικών ελέγχου του πληθυσμού τους.

Οι DeGrandi-Hoffman et al. [1989] παρουσίασαν ένα μοντέλο προσομοίωσης (BEEPOP) για την εξέλιξη του πληθυσμού της μέλισσας.

Οι Holm et al. [2008] επιχείρησαν να εξετάσουν, μέσω μοντέλου προσομοίωσης, την επιροή που έχει η ανεξέλεγκτη συγκομιδή του τροπικού φοίνικα *Mauritia flexuosa* στην εξέλιξη του πληθυσμού του είδους και ανέλυσαν τρόπους αειφόρου συγκομιδής.

Ο Chi [1990] ανέπτυξε ένα γενικό μοντέλο προσομοίωσης για την εξέλιξη του πληθυσμού παρασίτων. Σκοπός ήταν μέσω του μοντέλου αυτού να επιλέγονται οι πιο αποδοτικές και οικονομικές μέθοδοι ελέγχου του πληθυσμού του εκάστοτε παρασίτου υπό εξέταση.

Ο Throne [1989] έκανε χρήση ενός μοντέλου προσομοίωσης για την εξέλιξη του πληθυσμού του σκαθαριού σιτήρων ώστε να εξετάσει τη επίδραση βιολογικών μεθόδων ελέγχου του πληθυσμού.

Οι Fand et al. [2015] εξέτασαν την επιροή θερμότερων θερμοκρασιών, για την περιοχή της Ινδίας, στην εξέλιξη του πληθυσμού το σκώληκα κοπής, *Spodoptera litura*, με χρήση ενός μοντέλου προσομοίωσης.

1.2.6.1 Μοντέλα προσομοίωσης πληθυσμού του δάκου της ελιάς

Στην παρούσα υπό-υπό-ενότητα περιγράφονται μοντέλα προσομοίωσης που αφορούν την εξέλιξη του πληθυσμού το δάκου της ελιάς, καθώς σε αυτό το έντομο θα εστιάσει η παρούσα διατριβή.

Οι Comins and Fletcher [1988] πρότειναν ένα μοντέλο που χρησιμοποιεί ένα πιθανοτικό μοντέλο για την εκτίμηση του ρυθμού ανάπτυξης για οποιαδήποτε στιγμή. Ενώ, για να κάνουν τις προβλέψεις φαινολογίας του δάκου χρησιμοποίησαν Monte Carlo προσομοιώσεις. Τέλος, το μοντέλο τους, επίσης, χειρίζεται τη δυναμική των φρούτων, δηλαδή της ελιάς.

Οι Pommois et al. [2006] και οι Bruno et al. [2010] πρότειναν μια προσέγγιση μέσω κυτταρικού αυτόματου, την οποία ονόμασαν FlySim. Το μοντέλο πρόβλεψής τους επικεντρώνεται στη χωρική δυναμική του δάκου μέσα σε έναν ελαιώνα. Η χωρική διάχυση εξαρτάται από τη θερμοκρασία και την παρουσία καρπού ελιάς μέσα σε κάθε κύτταρο. Ενώ η χρονική ανάλυση του μοντέλου τους είναι ωριαία. Δεδομένου ότι χρησιμοποιείται ένα κυτταρικό αυτόματο είναι πιθανό ότι η προσομοίωση μεγάλων χρονικών περιόδων, όπως μήνες ή χρόνια, σε σχέση με το μέγεθος του κυτταρικού αυτόματου θα μπορούσε να επηρεάσει το χρόνο ολοκλήρωσης μιας προσομοίωσης.

Οι Gilioli and Pasquali [2007] εισήγαγε ένα μοντέλο ατομικής βάσης που υπολογίζει αριθμητικά τη δυναμική του πληθυσμού ποικιλόθερων οργανισμών. Η πρόθεσή τους ήταν να παρέχουν μια μέθοδο για την αξιολόγηση των λειτουργιών θνησιμότητας και γονιμότητας κάθε σταδίου ζωής του οργανισμού. Στο έργο τους, παρουσιάστηκε μια περίπτωση χρήσης του μοντέλου

στο δάκο της ελιάς.

Προκειμένου να εξεταστούν οι επιπτώσεις της κλιματικής αλλαγής στη δυναμική του πληθυσμού του δάκου της ελιάς οι Gutierrez et al. [2009] ανέπτυξαν ένα δημογραφικό μοντέλο βασισμένο στις καιρικές συνθήκες που βασίζεται σε φυσιολογικές συνθήκες για να προσομοιώσει την ανάπτυξη, τη φαινολογία και τη δυναμική του πληθυσμού και των δύο ειδών στην Αριζόνα-Καλιφόρνια και την Ιταλία υπό διαφορετικά σενάρια κλιματικής αλλαγής.

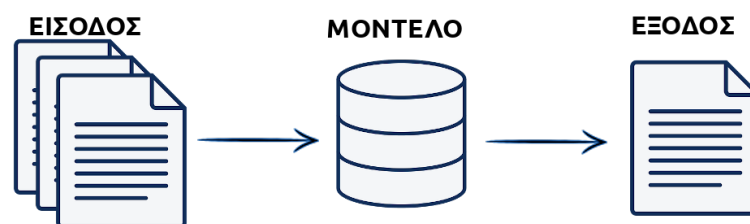
Οι Adeva et al. [2012] και Adeva and Reynolds [2012] χρησιμοποίησαν πεπερασμένες μηχανές κατάστασης για να αναπτύξει έναν προσομοιωτή προσβάσιμο από το διαδίκτυο. Στόχος τους ήταν να προσομοιώσουν τη χωρική και χρονική ανάπτυξη των μυγών φρούτων, του γένους *Bactrocera*, εστίες. Το μοντέλο τους λαμβάνει υπόψη την εποχικότητα του ξενιστή και τη θερμοκρασία για την πτυχή του κλίματος, ενώ η χρονική ανάλυση είναι μια μέρα.

Τέλος, οι Voulgaris et al. [2013] ανέπτυξαν ένα πληροφοριακό σύστημα ικανό να προσομοιώσει την δυναμική του πληθυσμού του δάκου καθώς επίσης την χωρική του διασπορά στο πεδίο, με αποτέλεσμα την πρόβλεψη πληθυσμιακών εξάρσεων.

1.3 Μηχανική μάθηση

Η μηχανική μάθηση θεωρείται ένα υποσύνολο της Τεχνητής Νοημοσύνης. Μέσω της μηχανικής μάθησης γίνεται μια προσπάθεια ένας υπολογιστής/μηχανή να μιμηθεί την ικανότητα του ανθρώπου να μαθαίνει από παλαιότερες εμπειρίες του [Russel and Norvig, 2009a].

Στην Εικόνα 1.8 παρουσιάζεται ένα γενικευμένο μοντέλο μηχανικής μάθησης. Ένα μοντέλο μηχανικής μάθησης δέχεται κάποια δεδομένα ως είσοδο και δίνει μια έξοδο.



Εικόνα 1.8: Γενικευμένο μοντέλο μηχανικής μάθησης.

Για την δημιουργία ενός μοντέλου μηχανικής μάθησης χρειάζεται να γίνει εκπαίδευση σε ένα σύνολο δεδομένων. Στόχος της εκπαίδευσης είναι η δημιουργία ενός μοντέλου που προβλέπει / απαντάει σωστά με μεγάλη ακρίβεια στα δεδομένα που δέχεται ως είσοδο. Για να επιτευχθεί αυτό πραγματοποιούνται τα ακόλουθα βήματα:

1. Συλλογή δεδομένων
2. Προ-επεξεργασία δεδομένων
3. Εκπαίδευση

4. Αξιολόγηση

5. Πρόβλεψη

1.3.1 Χαρακτηριστικό διάνυσμα

Ως χαρακτηριστικό ορίζεται μια μετρήσιμη ιδιότητα ενός φαινομένου / αντικειμένου που έχει παρατηρηθεί [Bishop, 2006]. Η τιμή του χαρακτηριστικού μπορεί να είναι είτε αριθμητική είτε συμβολική. Το σύνολο τέτοιων χαρακτηριστικών ορίζεται ως χαρακτηριστικό διάνυσμα.

1.3.1.1 Χρησιμότητα

Το χαρακτηριστικό διάνυσμα έχει ευρεία χρήση στην μηχανική μάθηση καθώς παρέχει την δυνατότητα να αντιπροσωπεύσει ένα αντικείμενο / φαινόμενο με αριθμητικό τρόπο. Αυτό έχει σαν αποτέλεσμα την εύκολη σύγκριση μεταξύ δύο χαρακτηριστικών διανυσμάτων καθώς υπάρχουν αρκετές τεχνικές / μέθοδοι για να επιτευχθεί αυτό όπως για παράδειγμα η Ευκλείδεια απόσταση.

Στην αναγνώριση εικόνας μέρος ενός χαρακτηριστικού διανύσματος μπορεί να είναι το χρώμα, οι ακμές και πολλά περισσότερα χαρακτηριστικά.

Στην αναγνώριση ομίλιας μέρος ενός χαρακτηριστικού διανύσματος μπορεί να είναι το μήκος του κύματος, επίπεδα θορύβου και άλλα.

1.3.1.2 Παράδειγμα

Έστω ότι θέλαμε μέσω μηχανικής μάθησης να μπορεί να επιτευχθεί η αναγνώριση / ταξινόμηση μοντέλων οχημάτων σε τι είδους όχημα αντιστοιχούν (αυτοκίνητο, ποδήλατο, μηχανή κ.α.). Ένα πιθανό σύνολο χαρακτηριστικών, που θα μπορούσε να χρησιμοποιηθεί ως χαρακτηριστικό διάνυσμα για την περιγραφή οχημάτων (βλ. Πίνακα 1.1), είναι ο συνολικός αριθμός από ρόδες που φέρει το όχημα και αν διαθέτει μηχανή ή όχι.

Table 1.1: Χαρακτηριστικό διάνυσμα οχημάτων

Χαρακτηριστικό	Τύπος τιμής	Παράδειγμα
Συνολικές ρόδες	Αριθμητική	2
Έχει μηχανή	Διακριτή	NAI

Στον Πίνακα 1.2 παρουσιάζονται οι τιμές των χαρακτηριστικών του διανύσματος για έξι μοντέλα οχημάτων.

1.3.2 Είδη μάθησης

Σε αυτήν την υπό-ενότητα θα περιγραφούν τα τρία είδη μηχανικής μάθησης.

Table 1.2: Τιμές χαρακτηριστικών για έξι μοντέλα οχημάτων

Μοντέλο οχήματος	Χαρακτηριστικό διάνυσμα	
	Συνολικές ρόδες	Έχει μηχανή?
Nissan Juke	4	NAI
Toyota Yaris	4	NAI
Kawasaki 750RS ZII Super	2	NAI
Honda CBX400F	2	NAI
Rayleigh Talus 3	2	OXI
Rayleight Tekoa 2	2	OXI

1.3.2.1 Επιβλεπόμενη μάθηση

Στην συγκεκριμένη μέθοδο μάθησης ο αλγόριθμος τροφοδοτείται με δεδομένα εκπαίδευσης (χαρακτηριστικά διανύσματα) και τις αντίστοιχες επιθυμητές εξόδους (κλάση ταξινόμησης) και προσπαθεί να μάθει έναν γενικό κανόνα για να επιτύχει αυτή την αντιστοίχιση [Russel and Norvig, 2009a].

Συνεχίζοντας το παράδειγμα που παρουσιάστηκε στην υπό-υπόενότητα 1.3.1.2 στο χαρακτηριστικό διάνυσμα προστέθηκε το χαρακτηριστικό "Είδος οχήματος" (βλ. Πίνακα 1.3). Με χρήση της επιβλεπόμενης μάθησης ο αλγόριθμος θα εκπαιδευτεί να αντιστοιχεί των συνδυασμό των τιμών των χαρακτηριστικών "Συνολικές ρόδες" και "Έχει μηχανή;" με τις τιμές του χαρακτηριστικού "Είδος οχήματος" που θα είναι και η έξοδος του αλγορίθμου. Λόγω της απλότητας του παραδείγματος η αντιστοίχιση που θα προκύψει είναι ότι για τον συνδυασμό τιμών 4, NAI η έξοδος θα είναι "Αυτοκίνητο", για 2, NAI η έξοδος θα είναι "Μηχανή" και τέλος για 2, OXI θα είναι "Ποδήλατο".

1.3.2.2 Μη-επιβλεπόμενη μάθηση

Σε αντίθεση με την επιβλεπόμενη μάθηση, στην παρούσα μέθοδο ο υπολογιστής τροφοδοτείται με δεδομένα εισόδου χωρίς να καθορίζεται ποια είναι η έξοδος για κάθε ένα από αυτά [Russel and Norvig, 2009a]. Πρακτικά πρέπει ο υπολογιστής να ανακαλύψει μόνος του την δομή των δεδομένων που δέχτηκε ως είσοδο.

Table 1.3: Χαρακτηριστικό διάνυσμα παραδείγματος 1.3.1.2 για επιβλεπόμενη μάθηση

Μοντέλο οχήματος	Χαρακτηριστικό διάνυσμα		
	Συνολικές ρόδες	Έχει μηχανή?	Είδος οχήματος
Nissan Juke	4	NAI	Αυτοκίνητο
Toyota Yaris	4	NAI	Αυτοκίνητο
Kawasaki 750RS ZII Super	2	NAI	Μηχανή
Honda CBX400F	2	NAI	Μηχανή
Rayleigh Talus 3	2	OXI	Ποδήλατο
Rayleight Tekoa 2	2	OXI	Ποδήλατο

1.3.2.3 Reinforcement learning

Σε αυτό το είδος μάθησης ο αλγόριθμος για κάθε πράξη που εκτελεί στο περιβάλλον στο οποίο δραστηριοποιείται λαμβάνει μια ανατροφοδότηση αν ήταν σωστή ή όχι [Russel and Norvig, 2009a]. Η συγκεκριμένη τεχνική θα μπορούσε να παρομοιαστεί με την εκπαίδευση ενός σκυλού. Όπου κατά την διάρκεια της εκπαίδευσης του κάθε φορά που εκτελεί σωστά μια εντολή που του δίνει ο εκπαιδευτής του δέχεται μια λιχουδιά σαν επιβράβευση. Με αυτόν τον τρόπο ο σκύλος μαθαίνει ότι στο άκουσμα της εντολής πρέπει να πραγματοποιεί την συγκεκριμένη ενέργεια για την οποία επιβραβεύτηκε. Παρομοίως ο αλγόριθμος για κάθε ενέργεια ή απόφαση που παίρνει με βάση τα δεδομένα εισόδου κερδίζει κάποια ανταμοιβή την οποία προσπαθεί να μεγιστοποιήσει.

1.3.3 Το πρόβλημα της ανισορροπίας κλάσης

Πολλές φορές υπάρχουν προβλήματα ταξινόμησης όπου καποιά/ές τιμή/ές της κλάσης ταξινόμησης δεν παρατηρείται/ούνται πολύ συχνά σε σχέση με τις υπόλοιπες τιμές της κλάσης. Σε αυτή την περίπτωση η διαφορά στην κατανομή των τιμών μπορεί να είναι αρκετά μεγάλη ώστε να υπάρχει ανισορροπία στην κλάση ταξινόμησης. Εκπαιδεύοντας έναν αλγόριθμο μηχανικής μάθησης με ένα τέτοιο σύνολο δεδομένων θα οδηγήσει ώστε το μοντέλο που θα προκύψει να μπορεί να προβλέπει με μεγάλο ποσοστό επιτυχίας τις τιμές της κλάσης που παρατηρήθηκαν με μεγάλη συχνότητα. Ωστόσο θα δυσκολεύεται να προβλέπει την/τις τιμή/ές ψευοψηφίας καθώς λόγω της έλλειψης σχετικών παραδειγμάτων δεν θα έχει καταφέρει να βρει κάποιο μοτίβο ώστε να μπορεί να συσχετίσει τα δεδομένα εισόδου με την/ις αντίστοιχη/ες τιμή/ές.

Το πρόβλημα της ανισορροπίας κλάσης μπορεί να αντιμετωπιστεί με διάφορους τρόπους. Η τυχαία υπερ-δειγματοληψία (τυχαία αναπαραγωγή των μειονοτικών παραδειγμάτων) και η τυχαία υπο-δειγματοληψία (τυχαία εξάλειψη των πλειοψηφικών παραδειγμάτων) έχουν χρησιμοποιηθεί για την εξισορρόπηση της κατανομής της κλάσης [Japkowicz, 2000]. Σύμφωνα με αρκετούς ερευνητές, η τυχαία υπερ-δειγματοληψία μπορεί να οδηγήσει σε υπερ-προσαρμογή, ενώ η τυχαία υπο-δειγματοληψία συνεπάγεται τον κίνδυνο απομάκρυνσης δυνητικά χρήσιμων δεδομένων, δηλ. δεδομένων που είναι κρίσιμα για τη διαδικασία επαγωγής [Batista et al., 2004].

Επομένως έχουν προταθεί μέθοδοι εστιασμένης αναδειγματοληψίας. Οι εστιασμένες τεχνικές υπερ-δειγματοληψίας ποικίλλουν από την αναπαραγωγή μόνο των παραδειγμάτων μειοψηφίας που εμφανίζονται στην οριακή γραμμή μεταξύ των δύο κατηγοριών [Japkowicz, 2000], στη δημιουργία νέων παραδειγμάτων μειονοτικών κλάσεων με παρεμβολή μεταξύ των υφιστάμενων παραδειγμάτων μειονότητας [Chawla et al., 2002] και προσεγγίσεων που συνδυάζουν υπερ-δειγματοληψία με καθαρισμό δεδομένων (απομάκρυνση παραδειγμάτων και από τις δύο κλάσεις) [Batista et al., 2003].

Οι τεχνικές υπο-δειγματοληψίας ποικίλλουν από την αφαίρεση των αρνητικών παραδειγμάτων που συμμετέχουν σε Tomek links [Tomek, 1976], στην εξεύρεση μιας συνεπούς υποομάδας παραδειγμάτων χρησιμοποιώντας τον κανόνα του πλησιέστερου γείτονα [Hart, 1968] και στον συνδυασμό αυτών των δύο προσεγγίσεων που ονομάζονται μονόπλευρη δειγματοληψία [Ku-

bat et al., 1997]. Ένας άλλος τρόπος αντιμετώπισης του ζητήματος της ανισορροπίας κλάσης είναι η ευαίσθητη στο κόστος μάθηση, δηλαδή η εφαρμογή ταξινομητών που δεν αντιμετωπίζουν όλα τα κόστη εσφαλμένης ταξινόμησης ως ίσα [Domingos, 1999; Provost and Fawcett, 2001].

Τέλος, έχει προταθεί και η τεχνική *metalearning* [Schapire, 2003; Guo and Viktor, 2004] για την αντιμετώπιση μη ισορροπημένων δεδομένων, καθώς αναγκάζει τον μαθητή να επικεντρωθεί σε σκληρά παραδείγματα.

1.3.4 Μέθοδοι αξιολόγησης

Στην μηχανική μάθηση είναι απαραίτητο το μοντέλο πρόβλεψης που έχει εκπαιδευτεί σε ένα σύνολο δεδομένων να αξιολογηθεί η απόδοσή του. Αυτό επιτυγχάνεται με τις ακόλουθες μεθόδους [Shalev-Shwartz and Ben-David, 2014a].

1.3.4.1 Σύνολα εκπαίδευσης & αξιολόγησης

Το σύνολο των δεδομένων χωρίζεται σε δύο υποσύνολα. Το πρώτο υποσύνολο χρησιμοποιείται για την εκπαίδευση του μοντέλου ενώ το δεύτερο υποσύνολο για την αξιολόγηση του.

1.3.4.2 Σύνολα εκπαίδευσης, αξιολόγησης & δοκιμής

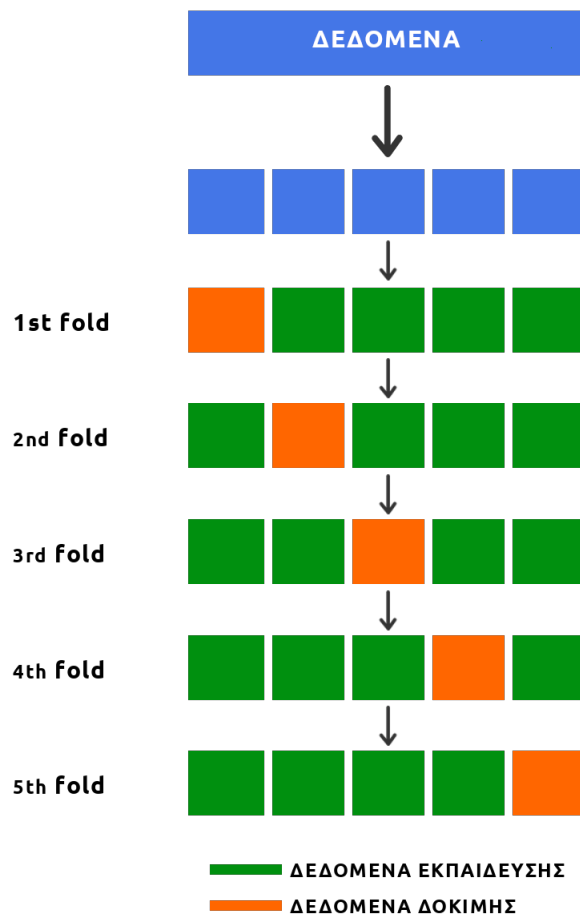
Το σύνολο των δεδομένων χωρίζεται σε τρία υποσύνολα. Το πρώτο υποσύνολο χρησιμοποιείται για την εκπαίδευση του μοντέλου ενώ το δεύτερο υποσύνολο για την αξιολόγηση του. Το τρίτο υποσύνολο χρησιμοποιείται όταν επιλεγεί το καλύτερο μοντέλο, βάση της απόδοσής του στο υποσύνολο αξιολόγησης, για να δοκιμαστεί η απόδοσή του.

1.3.4.3 k-fold cross validation

Με αυτή τη μέθοδο, όπως παρουσιάζεται και στην Εικόνα 1.9, το αρχικό δείγμα χωρίζεται τυχαία σε k υπόδείγματα αφού έχει γίνει μια τυχαία ανακατανομή των στοιχείων του. Ένα από τα k υπόδείγματα διατηρείται ως δεδομένα επικύρωσης για τη δοκιμή του μοντέλου και τα υπόλοιπα $k - 1$ υποδείγματα χρησιμοποιούνται ως δεδομένα εκπαίδευσης. Η διαδικασία διασταυρούμενης επικύρωσης ακολουθώντας επαναλαμβάνεται k φορές, με κάθε ένα από τα k υπόδείγματα να χρησιμοποιείται μόνο μία φορά ως δεδομένα επικύρωσης. Τέλος, τα αποτελέσματα υπολογίζονται κατά μέσον όρο στα k πειράματα που πραγματοποιήθηκαν.

1.3.5 Μετρικές απόδοσης

Οι μετρικές απόδοσης βοηθάνε στην αξιολόγηση της απόδοσης των αλγορίθμων μηχανικής μάθησης και να κατανοήσουμε πόσο καλά τα εκπαιδευμένα μοντέλα αποδίδουν στα δεδομένα



Εικόνα 1.9: Παράδειγμα της μεθόδου k-fold cross validation για k ίσο με πέντε.

εισόδου. Παρακάτω αναφέρονται μερικές μετρικές απόδοσης [Shalev-Shwartz and Ben-David, 2014b].

1.3.5.1 Πίνακας σύγχυσης (Confusion matrix)

Πρόκειται για την ευκολότερη μέθοδο μέτρησης της απόδοσης ενός μοντέλου πάνω σε ένα πρόβλημα ταξινόμησης. Ο πίνακας σύγχυσης, είναι ένας πίνακας δύο ή περισσότερων διαστάσεων, όπου ο αριθμός γραμμών και στηλών είναι ίσος με τον συνολικό αριθμό των τιμών της κλάσης ταξινόμησης. Οι στήλες αντιπροσωπεύουν τις τιμές που προβλέφθηκαν από το μοντέλο και οι γραμμές τις πραγματικές τιμές. Στον Πίνακα 1.4 παρουσιάζεται ένας πίνακας σύγχυσης 2x2 για ένα πρόβλημα δυαδικής ταξινόμησης.

Table 1.4: Πίνακας σύγκρισης για πρόβλημα δυαδικής ταξινόμησης

		ΠΡΟΒΛΕΨΕΙΣ	
		0	1
ΠΑΡΑΤΗΡΗΣΕΙΣ	0	TRUE NEGATIVE	FALSE POSITIVE
	1	FALSE NEGATIVE	TRUE POSITIVE

- True Positives (TP), σε αυτήν την περίπτωση η παρατηρηθείσα τιμή ήταν 1 και το μοντέλο ταξινόμησης προέβλεψε επίσης την τιμή 1.
- True Negatives (TN), σε αυτήν την περίπτωση η παρατηρηθείσα τιμή ήταν 0 και το μοντέλο ταξινόμησης προέβλεψε επίσης την τιμή 0.
- False Positives (FP σε αυτήν την περίπτωση η παρατηρηθείσα τιμή ήταν 1 και το μοντέλο ταξινόμησης προέβλεψε την τιμή 0.)
- False Negatives (FN), σε αυτήν την περίπτωση η παρατηρηθείσα τιμή ήταν 0 και το μοντέλο ταξινόμησης προέβλεψε την τιμή 1.

Στόχος των αλγορίθμων είναι να παράξουν ένα μοντέλο που μεγιστοποιεί τις σωστές προβλέψεις (TP & TN) και ταυτόχρονα ελαχιστοποιεί της λάθος προβλέψεις (FP & FN).

1.3.5.2 Ακρίβεια (Precision)

Η ακρίβεια απαντά στο ερώτημα πόσες από τις θετικές προβλέψεις ήταν όντως θετικές και πόσες ήταν αρνητικές αλλά εσφαλμένα προβλέφθηκαν ως θετικές.

Η τιμή της προκύπτει από τον πίνακα σύγκρισης μέσω του παρακάτω τύπου και κυμαίνεται από μηδέν (0) έως ένα (1).

$$Precision = \frac{\sum TruePositive}{\sum TruePositive + \sum FalsePositive}$$

1.3.5.3 Ανάκληση (Recall)

Η ανάκληση απαντά στο ερώτημα πόσες από τις θετικές τιμές προβλέφθηκαν ορθά ως θετικές και πόσες έπρεπε να προβεφτούν ως θετικές αλλά προβλέφθηκαν το αντίθετο.

Η τιμή της προκύπτει από τον πίνακα σύγκρισης μέσω του παρακάτω τύπου και κυμαίνεται από μηδέν (0) έως ένα (1).

$$Recall = \frac{\sum TruePositive}{\sum TruePositive + \sum FalseNegative}$$

1.3.5.4 F1-score

Η συγκεκριμένη μετρική απόδοσης κάνει χρήση των τιμών ακρίβειας και ανάκλησης και μπορεί να υπολογιστεί από τον παρακάτω τύπο.

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Πρόκειται στην ουσία για έναν μέσο όρο της ακρίβειας και της ανάκλησης ενός ταξινομητή και βοηθάει στην σύγκριση απόδοσης μεταξύ δύο ή περισσότερων ταξινομητών. Όπως προκύπτει και από τον παραπάνω τύπο το F1-score δίνει μεγάλο βάρος στις χαμηλές τιμές καθώς αν μία από τις δύο τιμές είναι χαμηλές τότε και το F1-score θα μειωθεί αντίστοιχα.

1.3.6 Deep Learning

Είναι ένα υποσύνολο της μηχανικής μάθησης με στόχο την μίμηση της δομής του ανθρώπινου εγκεφάλου το οποίο επιτυγχάνεται με την χρήση νευρωνικών δικτύων με πολλά κρυφά επίπεδα. Σε αντίθεση με την μηχανική μάθηση, όπου πρέπει να δημιουργηθεί ένα χαρακτηριστικό διάνυσμα για τα δεδομένα, με την μέθοδο του deep learning το νευρωνικό δίκτυο βρίσκει μόνο του τα χαρακτηριστικά τα οποία θα το βοηθήσουν ώστε να προβλέπει με μεγάλη ακρίβεια. Λόγω της ανεξαρτησίας αυτής, ο όγκος των δεδομένων εκπαίδευσης πρέπει να είναι πολύ μεγάλος.

1.3.7 Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα είναι η μέθοδος αναπαράστασης λειτουργίας του ανθρώπινου εγκεφάλου, όπου τα εγκεφαλικά κύτταρα γνωστά ως νευρώνες δημιουργούν ηλεκτροχημικά δίκτυα μεταξύ τους, στον υπολογιστή [Aggarwal, 2018a].

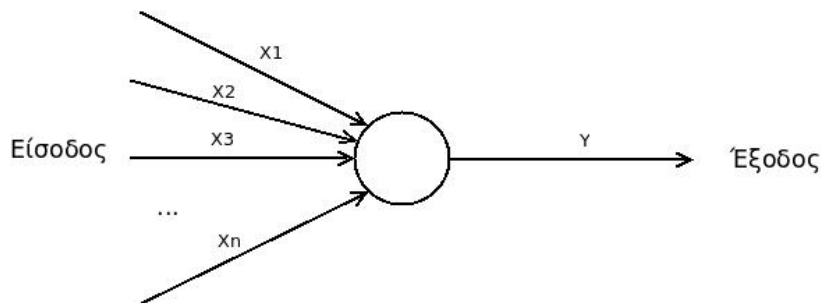
Στις επόμενες υπό-ενότητες θα γίνει αναφορά στα επιμέρους συστατικά από τα οποία αποτελείται ένα νευρωνικό δίκτυο.

1.3.7.1 Νευρώνας

Ένας νευρώνας ή κόμβος (Εικόνα 1.10) είναι η μονάδα που λαμβάνει μια πληροφορία, πραγματοποιεί υπολογισμούς με βάση την πληροφορία που δέχτηκε και προωθεί το αποτέλεσμα [Aggarwal, 2018b].

Ένα τεχνητό νευρωνικό δίκτυο μπορεί να αποτελείται από τα ακόλουθα τρία είδη νευρώνων:

- νευρώνες εισόδου, οι οποίοι λαμβάνουν την πληροφορία από τον εξωτερικό κόσμο
- νευρώνες εξόδου, οι οποίοι παράγουν το αποτέλεσμα από την επεξεργασία που πραγματοποιήθηκε στο εσωτερικό του νευρωνικού δικτύου



Εικόνα 1.10: Μοντέλο νευρώνα σε τεχνητό νευρωνικό δίκτυο.

- κρυφοί νευρώνες, πρόκειται για τους νευρώνες, που βρίσκονται ανάμεσα στους νευρώνες εισόδου και εξόδου, που επεξεργάζονται την πληροφορία που δέχτηκε ως είσοδο το νευρωνικό δίκτυο

Οι νευρώνες σε ένα τεχνητό νευρωνικό δίκτυο είναι σύνηθες να είναι χωρισμένη σε επίπεδα, τα οποία διακρίνονται στα τρία ακόλουθα είδη:

- επίπεδο εισόδου, όπου όπως προδίδει το όνομα τοποθετούνται οι νευρώνες εισόδου
- επίπεδο εξόδου, όπου τοποθετούνται οι νευρώνες εξόδου
- και τα κρυφά επίπεδα, στα οποία τοποθετούνται οι κρυφοί νευρώνες και που γίνεται η ανάλυση της πληροφορίας. Ο αριθμός των κρυφών επιπέδων μπορεί να ποικίλει και εξαρτάται από το πρόβλημα που καλείται να επιλύσει το εκάστοτε νευρωνικό δίκτυο

Στην πιο απλή του μορφή ένα νευρωνικό δίκτυο μπορεί να διαθέτει μόνο ένα επίπεδο εισόδου και εξόδου.

Όλοι οι νευρώνες, εκτός των νευρώνων του επιπέδου εισόδου, το αποτέλεσμα που υπολογίζουν το περνάνε από μία συνάρτηση ενεργοποίησης και το αποτέλεσμα αυτής το προωθούν στο επόμενο επίπεδο νευρώνων. Η προώθηση της πληροφορίας από επίπεδο σε επίπεδο επιτυγχάνεται μέσω των συνάψεων.

1.3.7.2 Συνάψεις

Οι συνάψεις είναι ο τρόπος σύνδεσης μεταξύ ενός νευρώνα με έναν ή περισσότερους νευρώνες που βρίσκονται στο επόμενο επίπεδο. Κάθε σύνναψη έχει ένα βάρος το οποίο επηρεάζει την πληροφορία που μεταδίδεται από τον έναν νευρώνα στον άλλον [Aggarwal, 2018b].

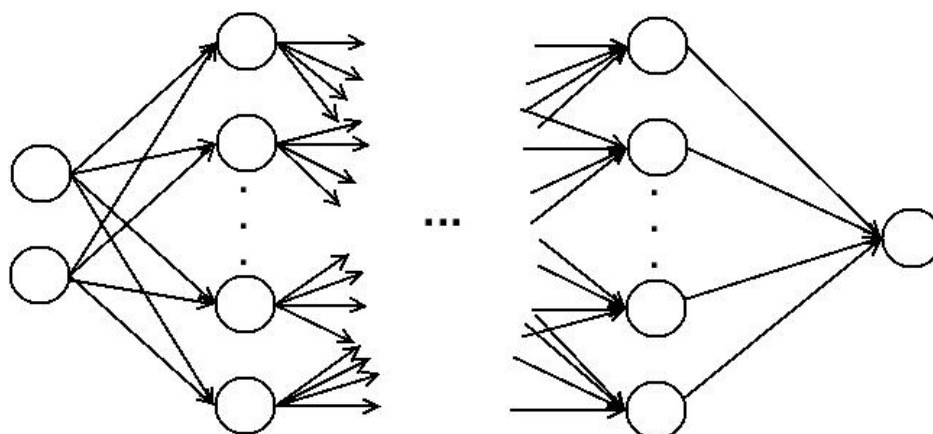
Τα βάρη κάθε σύνναψης στην αρχικοποίηση ενός νευρωνικού δικτύου ορίζονται τυχαία και τροποποιούνται, και κατ' επέκταση βελτιστοποιούνται, κατά την εκπαίδευση του νευρωνικού δικτύου.

1.3.7.3 Είδη νευρωνικών δικτύων

Υπάρχουν πάρα πολλές αρχιτεκτονικές νευρωνικών δικτύων για να αναφερθούν όλες στην παρούσα διατριβή. Για το λόγο αυτό θα γίνει αναφορά στις τρεις πιο διάσημες αρχιτεκτονικές.

Feed-forward neural network (FFNN) Πρόκειται για την πιο απλή μορφή νευρωνικού δικτύου [Rosenblatt, 1958, 1961] (Εικόνα 1.11), με το δίκτυο να τροφοδοτεί την εισερχόμενη πληροφορία από μπρος προς τα πίσω. Το δίκτυο αποτελείται από δύο (είσοδος και έξοδος) ή περισσότερα (είσοδος, κρυφά και έξοδος) επίπεδα, με τα επίπεδα να είναι πλήρως συνδεδεμένα μεταξύ τους. Δηλαδή, κάθε νευρώνας ενός επιπέδου έχει συνάψεις με όλους του νευρώνες του επόμενου επιπέδου.

Λόγο της αρχιτεκτονικής του το δίκτυο δεν έχει μνήμη αφού η πληροφορία μεταφέρεται πάντα προς τα εμπρός, δηλαδή από το ένα επίπεδο στο άλλο. Αυτή η μορφή δικτύου μπορεί να αποδώσει πάρα πολύ καλά στην επιβλεπόμενη μάθηση.

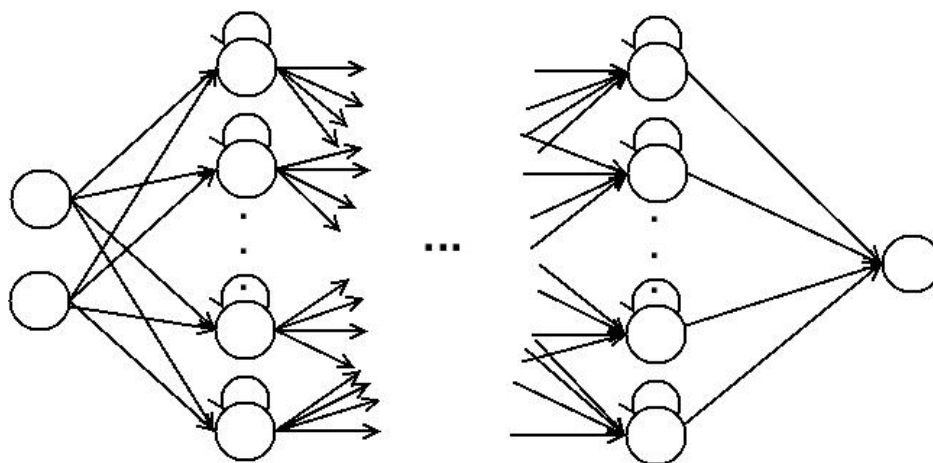


Εικόνα 1.11: Παράδειγμα Feed-forward neural network

Recurrent neural network (RNN) Η συγκεκριμένη αρχιτεκτονική επεκτείνει τα FFNNs προσθέτοντας το χαρακτηριστικό της μνήμης [Elman, 1990] (Εικόνα 1.12). Συγκεκριμένα οι νευρώνες δεν λαμβάνουν πληροφορία μόνο από τους νευρώνες του προηγούμενου επιπέδου αλλά και από τον εαυτό τους από το προηγούμενο πέρασμα. Κατ' επέκταση αυτό επιτρέπει στην συγκεκριμένη αρχιτεκτονική δικτύου να διαχειρίζεται μεταβλητού μήκους δεδομένα εισόδου.

Οι ιδιότητες της συγκεκριμένης αρχιτεκτονικής καθιστούν το νευρωνικό δίκτυο κατάλληλο, μεταξύ άλλων, για αναγνώριση γραφικού χαρακτήρα [Graves et al., 2009] και αναγνώριση ομιλίας [Li and Wu, 2014; Sak et al., 2014].

Convolutional neural network (CNN) Η συγκεκριμένη αρχιτεκτονική, η εμπνεύση της οποίας προήλθε από την κατανόηση της λειτουργίας του οπτικού φλοιού της γάτας [Hubel and Wiesel,



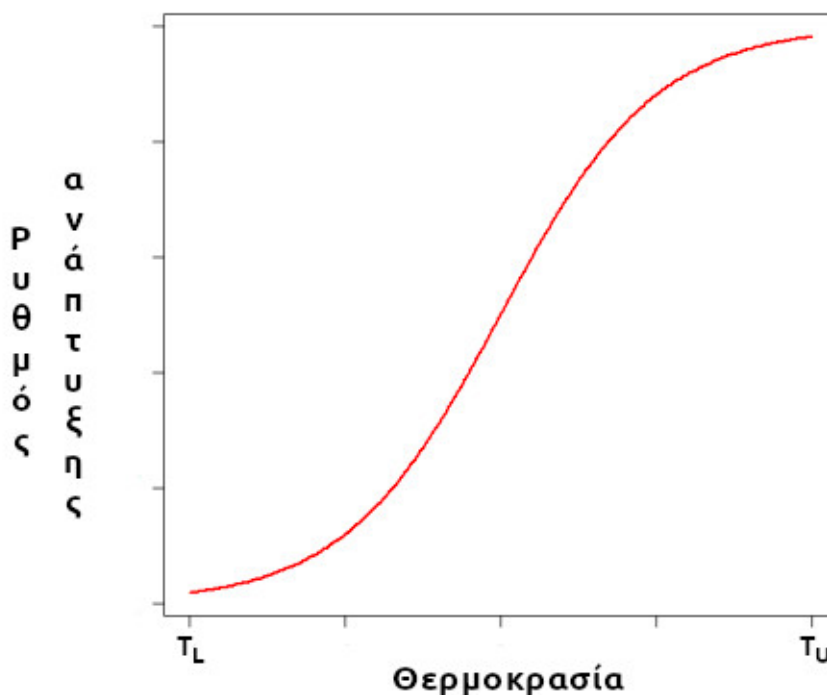
Εικόνα 1.12: Παράδειγμα Recurrent neural network

1959], χρησιμοποιείται κυρίως στην ανάλυση εικόνων [Valueva et al., 2020] καθώς έχει σχεδιαστεί να διαχειρίζεται δομημένους πίνακες δεδομένων όπως οι εικόνες. Έχουν πολλές εφαρμογές στην αναγνώριση και ταξινόμηση εικόνων, διαχείριση φυσικής γλώσσας [Collobert and Weston, 2008] κ.α. Η διαφορά τους με τα κλασσικά νευρωνικά δίκτυα και ως εκ τούτου το πλεονέκτημα τους είναι ότι ένα CNN δεν θα πάρει ως είσοδο όλη την εικόνα. Αν για παράδειγμα έχουμε μια εικόνα 100x100 pixel, ένα κλασσικό νευρωνικό δίκτυο θα χρειαζόταν 10000 νευρώνες εισόδου για να διαβάσει όλα τα pixel. Αντίθετα, ένα CNN θα χρησιμοποιήσει 100 νευρώνες εισόδου που θα σκανάρουν ένα τμήμα 10x10 pixel της εικόνας εισόδου (ξεκινώντας συνήθως από την πάνω αριστερά γωνία). Μόλις περαστούν σαν είσοδο σκανάρεται το επόμενο 10x10 τμήμα της εικόνας ξεκινώντας κατά ένα pixel στα δεξιά.

1.4 Μοντέλο Degree day

Έχει παρατηρηθεί πως η θερμοκρασία του περιβάλλοντος είναι άρρηκτα συνδεδεμένη με την ανάπτυξη φυτών και εντόμων. Πιο συγκεκριμένα, χαμηλές θερμοκρασίες επιβραδύνουν την ανάπτυξη ενός οργανισμού, ενώ όσο η θερμοκρασία αυξάνεται παρατηρείται αύξηση στον ρυθμό ανάπτυξης του οργανισμού μέχρις ότου η θερμοκρασία ξεπεράσει ένα άνω όριο από το οποίο και μετά ο ρυθμός ανάπτυξης μειώνεται ή και παύει (Εικόνα 1.13). Με βάση τα παραπάνω δημιουργήθηκε το μοντέλο degree day (DD) το οποίο υπολογίζει την συνολική θερμότητα που απαιτείται από έναν οργανισμό, μεταξύ ενός άνω (T_U) και κάτω (T_L) ορίου θερμοκρασίας, για να αναπτυχθεί. Ειδικότερα, το μοντέλο degree day υπολογίζει την συνολική θερμότητα που έχει συλλέξει ένα οργανισμός μεταξύ των ορίων ανάπτυξης του μέσα σε μια μέρα [Wilson and Barnett, 1983].

Με τη χρήση αυτού του μοντέλου, μπορεί κανείς να καθορίσει την ποσότητα θερμότητας και κατ' επέκταση τον συνολικό χρόνο που απαιτείται για έναν οργανισμό να μεταβεί από το ένα στάδιο ανάπτυξης στο άλλο. Είναι, επίσης, αξιοσημείωτο πως το μοντέλο Degree day μπορεί να είναι αρωγός στην προσπάθεια ελέγχου των παρασίτων [Pruess, 1983].



Εικόνα 1.13: Ρυθμός ανάπτυξης σε σχέση με την αύξηση της θερμοκρασίας με άνω και κάτω όριο.

1.4.1 Μέθοδοι υπολογισμού

Στην βιβλιογραφία αναφέρονται διάφορες μέθοδοι υπολογισμού των DD με χρήση των ημερήσιων ελάχιστων και μέγιστων θερμοκρασιών. Παρακάτω αναφέρονται μόνο μερικές από αυτές.

1.4.1.1 Μέση θερμοκρασία

Η πιο απλή μέθοδος υπολογισμού κάνει χρήση της μέσης θερμοκρασίας που προκύπτει από την ελάχιστη και μέγιστη ημερήσια θερμοκρασία [Arnold et al., 1960] και υπολογίζεται από την εξίσωση 1.4.1.1.

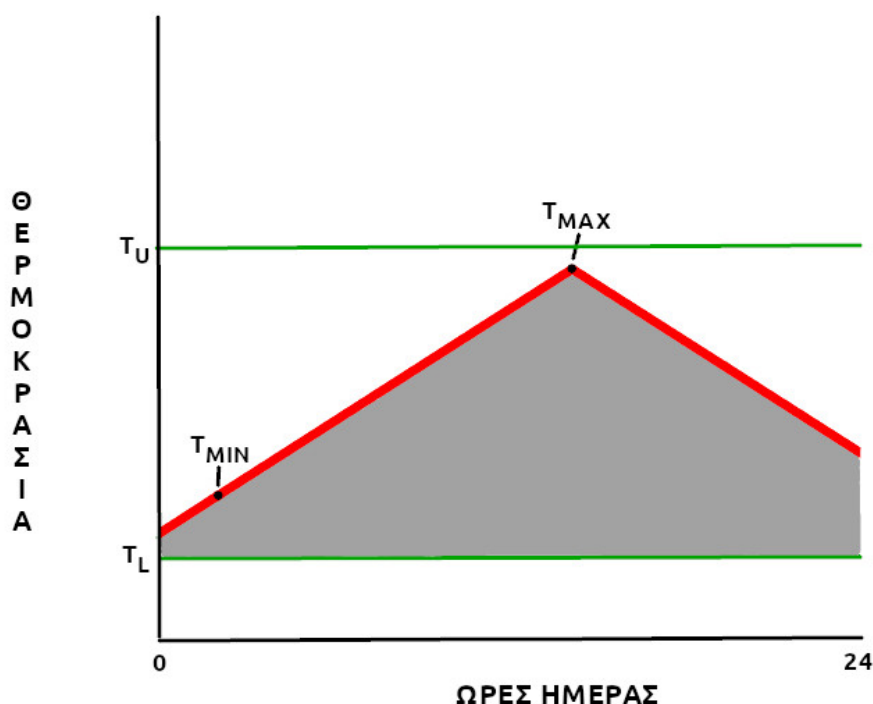
$$DD = \frac{T_{max} + T_{min}}{2} - T_L \quad (1.8)$$

όπου T_{max} η μέγιστη ημερήσια θερμοκρασία, T_{min} η ελάχιστη ημερήσια θερμοκρασία και T_L το κάτω όριο θερμοκρασίας ανάπτυξης.

Η συγκεκριμένη μέθοδος έχει διάφορες παραλλαγές σε ότι αφορά την διαχείριση θερμοκρασιών εκτός των ορίων ανάπτυξης. Για παράδειγμα, οι Bauer et al. [1984] χρησιμοποίησαν την

παραπάνω μέθοδο με την αναπροσαρμογή των μέγιστων και ελάχιστων ημερήσιων θερμοκρασιών, εκτός των ορίων ανάπτυξης, στις αντίστοιχες τιμές των ορίων ανάπτυξης. Δηλαδή, η μέγιστη ημερήσια θερμοκρασία θεωρούνταν ίση με το άνω όριο θερμοκρασίας ανάπτυξης και η ελάχιστη ημερήσια θερμοκρασία θεωρούνταν ίση με το κάτω όριο θερμοκρασίας ανάπτυξης. Μετά την αναπροσαρμογή πραγματοποιούσαν τον υπολογισμό των DD. Από την άλλη μεριά, οι Ottman [2008] αφού υπολόγιζαν την μέση θερμοκρασία αν βρισκόταν εκτός των ορίων ανάπτυξης θεωρούσαν ότι συλλέχθηκαν οι μέγιστοι δυνατοί DD αν ήταν πάνω από το άνω όριο θερμοκρασίας ανάπτυξης ή 0 DD αν η θερμοκρασία ήταν μικρότερη από το κάτω όριο θερμοκρασίας ανάπτυξης. Ειδικά, εκτελούσαν των υπολογισμό τον DD κανονικά.

1.4.1.2 Μέθοδος μονού τριγώνου



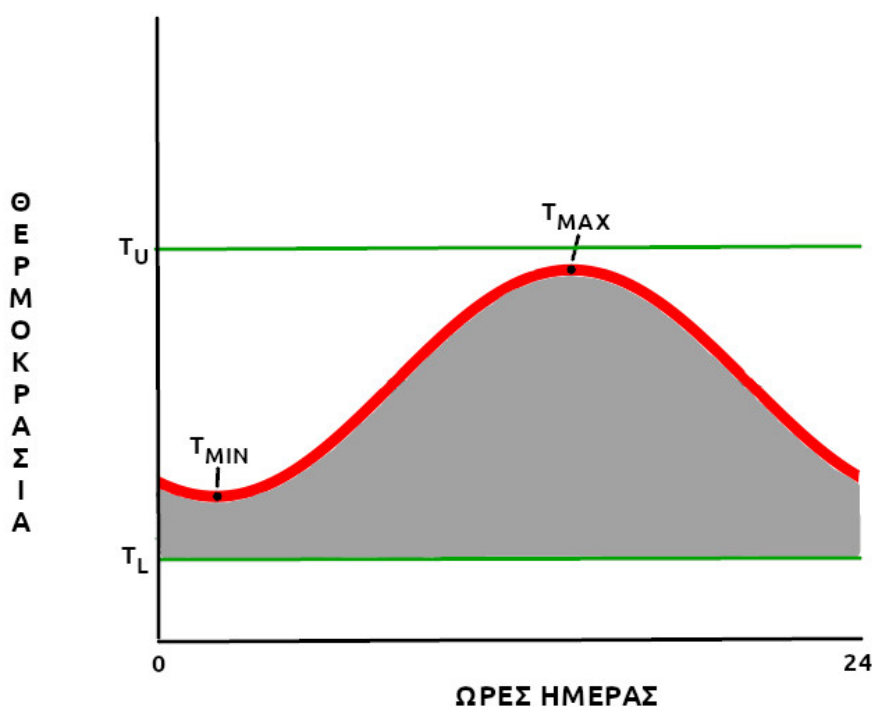
Εικόνα 1.14: Μέθοδος μονού τριγώνου για τον υπολογισμό των Degree Day.

Σε αυτήν την μέθοδο (Εικόνα 1.14) χαράσσεται μια ευθεία γραμμή μεταξύ της ελάχιστης και μέγιστης θερμοκρασίας της ημέρας και αλλή μία από την μέγιστη θερμοκρασία της ημέρας μέχρι την ελάχιστη θερμοκρασία της επόμενης ημέρας, η οποία θεωρείται ότι είναι ίδια με της προηγούμενης. Ο χώρος που οριοθετείται μεταξύ του σχηματισμένου τριγώνου και των άνω και κάτω ορίων θερμοκρασιών ανάπτυξης συνεισφέρει στην ανάπτυξη του οργανισμού [Lindsey and Newman, 1956]

1.4.1.3 Μέθοδος διπλού τριγώνου

Σε αυτήν την μέθοδο χαράσσεται μια ευθεία γραμμή μεταξύ της ελάχιστης και μέγιστης θερμοκρασίας της ημέρας και αλλη μία από την μέγιστη θερμοκρασία της ημέρας μέχρι την πραγματική ελάχιστη θερμοκρασία της επόμενης ημέρας. Επίσης τοποθετείται μια κάθετη ευθεία που περνάει από την μέγιστη θερμοκρασία, δημιουργώντας με αυτόν τον τρόπο δύο τρίγωνα. Το άθροισμα του χώρου που οριοθετείται μεταξύ του πρώτου σχηματισμένου τριγώνου και των άνω και κάτω ορίων θερμοκρασιών ανάπτυξης και του χώρου που οριοθετείται μεταξύ του δεύτερου σχηματισμένου τριγώνου και των άνω και κάτω ορίων θερμοκρασιών ανάπτυξης συνεισφέρει στην ανάπτυξη του οργανισμού [Sevacherian et al., 1977]

1.4.1.4 Ημιτονοειδής καμπύλη



Εικόνα 1.15: Μέθοδος ημιτονοειδούς καμπύλης για τον υπολογισμό των Degree Day. Η γκριζα περιοχή υποδηλώνει την θερμότητα που θα συνεισφέρει στην ανάπτυξη του υπό εξέταση οργανισμού.

Στην συγκεκριμένη μέθοδο (Εικόνα 1.15) χαράσσεται μια ημιτονοειδής καμπύλη με όρια την ελάχιστη (T_{min}) και μέγιστη (T_{max}) ημερήσια θερμοκρασία. Ο χώρος που οριοθετείται μεταξύ της ημιτονοειδούς καμπύλης και των άνω (T_U) και κάτω (T_L) ορίων θερμοκρασιών ανάπτυξης συνεισφέρει στην ανάπτυξη του οργανισμού [Baskerville and Emin, 1969].

1.4.1.5 Διπλή ημιτονοειδής καμπύλη

Πρόκειται για μια παραλλαγή της προαναφερθείσας μεθόδου που προτάθηκε από τους Allen [1976]. Αναλυτικότερα, αρχικά τοποθετείται μια ημιτονοειδής καμπύλη με όρια την ελάχιστη (T_{min1}) και μέγιστη (T_{max}) θερμοκρασία της ημέρας. Στην συνέχεια τοποθετείται μια δεύτερη ημιτονοειδής καμπύλη με όρια την μέγιστη (T_{max}) θερμοκρασία της ημέρας και την ελάχιστη (T_{min2}) θερμοκρασία της επόμενης μέρας. Τα DD υπολογίζονται από το άθροισμα του χώρου που οριοθετείται μεταξύ της πρώτης ημιτονοειδής καμπύλης και των άνω (T_U) και κάτω (T_L) ορίων θερμοκρασιών ανάπτυξης και του χώρου που οριοθετείται μεταξύ της δεύτερης ημιτονοειδής καμπύλης και των άνω (T_U) και κάτω (T_L) ορίων θερμοκρασιών ανάπτυξης.

Ανάπτυξη στοχαστικών αλγορίθμων και μοντέλων μηχανικής μάθησης για την εξέλιξη πληθυσμών: Εφαρμογή στο δάκο της Ελιάς

2. ΕΛΙΑ ΚΑΙ ΔΑΚΟΣ

Σκοπός αυτού του κεφαλαίου είναι η παρουσίαση των οφελών του ελαιόδεντρου στη ζωή του ανθρώπου και τις επιπτώσεις που έχει ο δάκος της ελιάς στην βιομηχανία αυτή. Αναλυτικότερα στην ενότητα 2.1 γίνεται μια περιγραφή σχετικά με τα προϊόντα του ελαιόδεντρου και τα οφέλη που αυτά έχουν στην υγεία του ανθρώπου όταν συμπεριλαμβάνονται στην διατροφή του. Στην συνέχεια στην ενότητα 2.2 γίνεται μια περιγραφή σχετικά με τον δάκο της ελιάς, την δυναμική και τα χαρακτηριστικά του πληθυσμού του, τις οικονομικές επιπτώσεις της δραστηριότητας του και τέλος οι μέθοδοι αντιμετώπισης του. Τέλος, στην ενότητα 2.3 γίνεται μια αναφορά στα μικροκλίματα, στην υπάρχουσα έρευνα σχετικά με την αναγνώριση τους και πως αυτά επηρεάζουν την συμπεριφορά του δάκου μέσα στον ελαιώνα.

2.1 Ελαιόδενδρο (*Olea europaea*)

Τα ελαιόδεντρα βρίσκονται σε μεγάλη αφθονία στη λεκάνη της Μεσογείου με τον αριθμό των δέντρων να φτάνουν τα 800 εκατομμύρια περίπου [Alfonso and Owen, 2002]. Παράλληλα, είναι η κυρίαρχη μόνιμη καλλιέργεια στην ΕΕ όσον αφορά τις κατεχόμενες περιοχές (40% της συνολικής έκτασης μόνιμης καλλιέργειας [Eurostat, 2017]) με περισσότερες από 1500 ποικιλίες [Fogher et al., 2010] μόνο στη Μεσόγειο.

2.1.1 Προϊόντα

Τα κύρια προϊόντα που παράγονται από ένα ελαιόδεντρο είναι:

- το ελαιόλαδο (Εικόνα 2.1a)
- και οι επιτραπέζιες ελιές(Εικόνα 2.1b)

2.1.2 Οφέλη

Προσφέρουν πολλά οφέλη για την υγεία, για παράδειγμα, η κατανάλωση φύλλων ελιάς έχει αντιοξειδωτικές, αντιϋπερτασικές, αντιφλεγμονώδεις, μεταξύ άλλων, ιδιότητες [El and Karakaya, 2009].

Το ελαιόλαδο είναι ένα από τα βασικά συστατικά της διάσημης μεσογειακής διατροφής και έχει τεράστια οφέλη για την υγεία. Σύμφωνα με πολλές μελέτες, το είδος του λίπους που περιλαμβάνεται στο ελαιόλαδο μπορεί να είναι υπεύθυνο για τη μείωση της κακής χοληστερόλης. Οι Mensink and Katan [1989] έδειξαν ότι επιτυγχάνεται σχεδόν μια μείωση 20% σε αριθμούς λιποπρωτεϊνών χαμηλής πυκνότητας. Επιπλέον, το παρθένο ελαιόλαδο προάγει πρόσθετα οφέλη σε αυτά που παρέχονται από άλλα φυτικά έλαια σε παράγοντες κινδύνου για καρδιαγγειακές παθήσεις [Planell et al., 2014]. Επιπλέον, το ελαιόλαδο είναι γνωστό ως ένας



(a)



(b)

Εικόνα 2.1: Τα προϊόντα που παράγονται από τα ελαιόδεντρα: (a) ελαιόλαδο και (b) επιτραπέζιες ελιές

από τους κύριους τρόπους για τη μείωση των θανατηφόρων ασθενειών όπως η αθηροσκλήρωση και το εγκεφαλικό επεισόδιο [Visioli et al., 1995].

Επιπλέον, καθώς μία από τις κύριες κατηγορίες προβλημάτων υγείας σήμερα σχετίζεται με την καρδιά, η έρευνα των De la Torre et al. [2017] δείχνει ότι το ελαιόλαδο ενεργεί σαν ασπίδα για την προστασία των καταναλωτών του. Άλλοι ερευνητές [Escrich et al., 2011; Keys et al., 1986] ανέφεραν ότι η προσθήκη ελαιολάδου στη διατροφή θα μπορούσε να μειώσει τους κινδύνους καρδιακών προβλημάτων.

Ενώ οι συνήθειες πρακτικές μείωσης του λίπους στα καθημερινά διατροφικά σχέδια οδήγησαν στην αποφυγή του ελαιολάδου, το έργο των Esposito et al. [2011] έχει δείξει διαφορετικά, δεδομένου ότι το ελαιόλαδο, και η μεσογειακή διατροφή γενικά, συνδέονται σε μεγάλο βαθμό με την απώλεια βάρους. Επιπλέον, το ελαιόλαδο επηρεάζει την όρεξη και το αίσθημα της πείνας. Μια μελέτη των McManus et al. [2001] συνέκρινε τη μεσογειακή διατροφή με δίαιτες χαμηλών λιπαρών και κατέληξε στο συμπέρασμα ότι η πρώτη έχει καλύτερα αποτελέσματα για την απώλεια βάρους.

Μια δίαιτα με αυξημένη κατανάλωση ελαιολάδου έχει αποδειχθεί ότι σχετίζεται με μειωμένο κίνδυνο διαβήτη [Martínez-González et al., 2008]. Για τους διαβητικούς υπάρχουν επίσης καλά αποτελέσματα από την κατανάλωση ελαιολάδου, καθώς έχει αποδειχθεί ότι βελτιώνει το σάκχαρο στο αίμα και επιτρέπει την απώλεια βάρους σε αντίθεση με τις εναλλακτικές λύσεις χαμηλού λίπους [Lasa et al., 2014].

Σύμφωνα με το Παγκόσμιο Παρατηρητήριο Υγείας¹ η αρτηριακή πίεση του παγκόσμιου πληθυσμού αυξάνεται με την πάροδο των ετών. Η υψηλή αρτηριακή πίεση ευθύνεται για πολλούς θανάσιμους κινδύνους για την υγεία, όπως εγκεφαλικό επεισόδιο και καρδιακές προσβολές. Η κατανάλωση παρθένου ελαιολάδου σύμφωνα με κλινική δοκιμή που διεξήγαγε οι Cândido et al. [2018] έδειξαν βελτίωση στην αρτηριακή πίεση.

¹<https://apps.who.int/gho/data/node.main.A875STANDARD?lang=en>



(a)



(b)

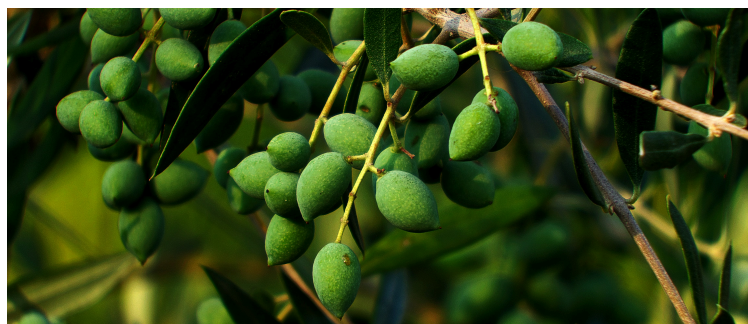
Εικόνα 2.2: (a) Θυληκός και (b) αρσενικός δάκος

Μια άλλη πολύ σημαντική πτυχή για την κατανάλωση ελαιολάδου είναι ότι θα μπορούσε να εμποδίσει την ανάπτυξη ορισμένων μορφών καρκίνου, όπως ο καρκίνος του μαστού [Escrich et al., 2011] και ο καρκίνος του παχέος εντέρου [Storniolo and Moreno, 2016]. Σύμφωνα με πολλά ερευνητικά έργα [Smith-Warner et al., 2001; Trichopoulou et al., 1995; Ayoub et al., 2017], η ανωτερότητα του ελαιολάδου έχει αποδειχθεί σημαντικά υποσχόμενη κατά του καρκίνου του μαστού.

2.2 Δάκος

Ο Δάκος της ελιάς *Bactrocera oleae* (Rossi) (Diptera, Tephritidae) (Εικόνα 2.2) είναι μια μύγα φρούτου που σχετίζεται με το ελαιόδεντρο. Ο καρπός του ελαιόδεντρου (Εικόνα 2.3) είναι ο μόνος τρόπος αναπαραγωγής του εντόμου καθώς εκεί εναποθέτουν τα θυληκά του είδους τα αυγά τους και αποτελεί την μόνη πηγή τροφής των πρώιμων σταδίων του εντόμου κατά την διάρκεια της ανάπτυξης του [Nardi et al., 2005].

Παρότι ο δάκος της ελιάς πιθανώς προέρχεται από την Αφρική [Zohary, 1993], έχει πλέον εξαπλωθεί σε όλη την περιοχή της Μεσογείου και σε μεγάλο ποσοστό χωρών με ελαιοκαλλιέργει-



Εικόνα 2.3: Καρπός του ελαιόδεντρου

σιμες εκτάσεις. Η τρέχουσα κατανομή του επιβλαβούς εντόμου περιλαμβάνει την Νότια και Κεντρική Αφρική [Gaouar and Debouzie, 1995; Munro et al., 1984; Merz, 1992], τη Μεσόγειο [Brnetic et al., 1979], την Ασία [Nouri et al., 2007; Kapoor et al., 1993; Gümusay et al., 1989; Shant et al., 1999; Kandybina et al., 1977], τις ΗΠΑ [Rice et al., 2003] και το Μεξικό [Nardi et al., 2005].

2.2.1 Δυναμική πληθυσμού

Όπως προαναφέρθηκε τα θυληκά του είδους εναποθέτουν τα αυγά τους μέσα στον καρπό του ελαιόδεντρου [Nardi et al., 2005] και συνηθώς προτιμούν να αφήνουν ένα αυγό ανά καρπό [Gutierrez et al., 2009]. Η διαδικασία εναπόθεσης του αυγού στον καρπό της ελιάς διαρκεί μεταξύ 6 με 13 λεπτά [Genç and Nation, 2008]. Στα πειράματα τους οι Genç and Nation [2008] διαπίστωσαν πως ένας θυληκός δάκος μπορεί να εναποθέσει 19 αυγά ανά ημέρα και κατά μέσο όρο 225 αυγά σε όλη τη διάρκεια της ζωής του.

Η προνύμφη που εκκολάπτεται από το αυγό τρέφεται από τον καρπό της ελιάς μέχρι να μπει στο επόμενο στάδιο ανάπτυξης της νύμφης [Daane and Johnson, 2010]. Το στάδιο της νύμφης ανάλογα τις περιβαλλοντικές συνθήκες που επικρατούν το περνάει είτε μέσα στο καρπό της ελιάς είτε μέσα στο έδαφος [Dimou et al., 2003]. Τέλος, μόλις ολοκληρώσει την ανάπτυξη του μέσα στο κέλυφος της νύμφης προκύπτει ο ενήλικος δάκος.

Η ανάπτυξη και η εν γένει δραστηριότητα του δάκου της ελιάς εξαρτάται σε μεγάλο βαθμό από τις περιβαλλοντικές συνθήκες που επικρατούν. Όπως έχει αποδειχθεί η θερμοκρασία και η σχετική υγρασία είναι σημαντικοί περιβαλλοντικοί παράγοντες που επηρεάζουν τη δυναμική του πληθυσμού του δάκου.

Σε ότι αφορά την θερμοκρασία, έχει αποδειχθεί [Tsitsipis and JA, 1977; Tsitsipis, 1980] πως ο ρυθμός ανάπτυξης και των τριών ανώριμων σταδίων αυξάνεται καθώς η περιβαλλοντική θερμοκρασία αυξάνεται και πλησιάζει την ανώτατη θερμοκρασία ανάπτυξης του εκάστοτε σταδίου ανάπτυξης. Η έκθεση σε υψηλές θερμοκρασίες, ακόμη και για μερικές ώρες, επηρεάζει όλα τα στάδια ανάπτυξης όσον αφορά τη μακροζωία, την αναπαραγωγή και την επιβίωση [Pappas et al., 2011]. Επιπλέον, οι υψηλές θερμοκρασίες επηρεάζουν την απόδοση πτήσης του δάκου [Wang et al., 2009], ενώ όταν η θερμοκρασία ξεπεράσει τους 35 °C ο δάκος παύει κάθε δραστηριότητα [Avidon, 1954]. Τέλος, οι Fletcher and Karpatos [1983] διαπίστωσαν πως για θερμοκρασίες μικρότερες από 12 °C δεν αναπτύσσονται οι ωothήκες στους θηλυκούς δάκους.

Ομοίως, χαμηλές τιμές υγρασίας, κάτω του 33%, μειώνει την ωρίμανση των ωοθηκών, καθώς και τη μακροζωία των ενήλικων δάκων. Ενώ, για υψηλότερες τιμές οι προαναφερθείσες παράμετροι αυξάνονται [Broufas et al., 2009].

Συνεπώς βάση των περιβαλλοντικών συνθηκών που επικρατούν ο συνολικός αριθμός γενεών του Δάκου ανά έτος ποικίλλει από περιοχή σε περιοχή. Έχουν αναφερθεί έως και πέντε γενιές στη Νότια Καλιφόρνια, με τους συγγραφείς να δηλώνουν ότι κάτω από ιδανικές συνθήκες θα μπορούσε να φτάσει τις επτά γενιές [Rice et al., 2003]. Εν αντιθέση με την Ιορδανία, για παράδειγμα, που έχουν καταγραφεί έως και τρεις γενιές [Mustafa and Al-Zaghal, 1987].

2.2.2 Χαρακτηριστικά πληθυσμου

2.2.2.1 Αναλογία φύλου

Η αναλογία θηλυκών και αρσενικών δάκων με βάση την βιβλιογραφία [Ant et al., 2012; Moore, 1962; Speranza et al., 2004; Neuenschwander et al., 1979] τείνει κοντά στο 1:1.

2.2.2.2 Ζευγάρισμα

Οι πρόσφατα εκκολαπτόμενοι ενήλικοι δάκοι χρειάζονται τρεις με οχτώ μέρες για να ωριμάσουν σεξουαλικά [Mazomenos, 1984] και κατά συνέπεια να αρχίσουν να ζευγαρώνουν. Οι θηλυκοί δάκοι έχει διαπιστωθεί πως ζευγαρώνουν το πολύ τρεις φορές κατά την διάρκεια της ζωής τους [Zervas, 1982; Zouros and Krimbas, 1970] και πως ένα μεγάλο ποσοστό των θηλυκών δάκων είναι μονογαμικοί [Zouros and Krimbas, 1970]. Από την άλλη πλευρά, τα αρσενικά του είδους είναι πολυγαμικά [Zervas, 1982].

2.2.2.3 Διάχυση

Έχει βρεθεί πως σε διάστημα ενός μήνα μεγάλο ποσοστό του πληθυσμού του δάκου δεν διαχέεται παραπάνω από ένα χιλιόμετρο [Pelekassis et al., 1963]]. Ενώ, όταν προκύπτουν συνθήκες που δυσκολεύουν την αναπαραγωγή του είδους όπως η έλειψη καρπού ή ο υπερπληθυσμός τότε παρατηρήθηκε οι δάκοι να διανύουν μεγάλες αποστάσεις με στόχο την εύρεση ευνοϊκότερων συνθηκών [Fletcher and Kapatos, 1981; Economopoulos et al., 1978; Remund et al., 1976]. Παράταυτα, έχει διαπιστωθεί πως η μέγιστη απόσταση διάχυσης του ενήλικα δάκου κυμαίνεται μεταξύ τεσσάρων με πέντε χιλιομέτρων [Economopoulos et al., 1978; Remund et al., 1976; Pelekassis et al., 1963].

2.2.3 Επιπτώσεις στην ελαιοκαλλιέργεια

Λόγω της εξάρτησης του δάκου, για την διαίωνιση του, από τον καρπό της ελιάς, τόσο για την εναπόθεση των αυγών του όσο και σαν πηγή τροφής για το στάδιο της προνύμφης η ζημιά που προκαλεί στα παραγόμενα προϊόντα του ελαιόδεντρου (βλ. υπό-ενότητα 2.1.1) είναι και ποσοτική και ποιοτική [Neuenschwander and Michelakis, 1978]. Τρία είναι τα κύρια είδη ζημιάς που προκαλεί ο δάκος:

1. Πρώιμη πτώση του ελαιόκαρπου και κατά συνέπεια μείωση της σοδειάς.
2. Μειωμένος πολτός του ελαιόκαρπου λόγω της κατανάλωσης του από την προνύμφη, ως μόνη πηγή τροφής. Μελέτη έχει δείξει πως η κατανάλωση του πολτού κυμαίνεται από 50mg έως 150mg ανάλογα την ποικιλία της ελιάς [Neuenschwander and Michelakis, 1978]. Αυτό έχει σαν άμεσο αποτέλεσμα την μείωση της ποσότητας παραγόμενου ελαιολάδου γύρω στο 20-25% [Silvestri, 1914].

3. Μείωση της ποιότητας του παραγόμενου ελαιολάδου.

Μελέτες έχουν δείξει πως υπό ευνοϊκές συνθήκες και χωρίς εφαρμογή κάποιας μεθόδου αντιμετώπισης η προσβολή του ελαιώνα από τον δάκο μπορεί να ξεπεράσει το 90% [Karatos and Fletcher, 1984; Athar, 2005].

2.2.4 Τρόποι αντιμετώπισης

Για τον έλεγχο του πληθυσμού του δάκου δύο είναι οι κύριες μέθοδοι αντιμετώπισης είτε μέσω ψεκασμών του ελαιώνα είτε με την μαζική τοποθέτηση παγίδων. Ο συνδυασμός των προαναφερθέντων μεθόδων αποτελεί μια συχνή τακτική για την μείωση του πληθυσμού του δάκου σε έναν ελαιώνα.

2.2.4.1 Παγίδες

Οι παγίδες κάνουν συνήθως χρήση κάποιου ελκυστικού και μπορούν να χρησιμοποιηθούν είτε για μείωση του πληθυσμού του δάκου με μαζική τοποθέτηση μέσα στον ελαιώνα ή ως μέθοδος παρακολούθησης του πληθυσμού με σκοπό τον εντοπισμό της έξαρσης του. Τα είδη παγίδων που χρησιμοποιούνται για τον έλεγχο/παρακολούθηση του πληθυσμού του δάκου είναι ένα από τα ακόλουθα:

McPhail παγίδες (Εικόνα 2.4) πρόκειται για μια πλαστική παγίδα αποτελούμενη από τρία διακριτά μέρη

- Διάφανο καπάκι
- Μια ανεστραμμένη χοάνη κίτρινου χρώματος
- Νερό

Ο ενήλικας δάκος, και άλλα έντομα, ελκύονται από το ελκυστικό που είναι τοποθετημένο στο εσωτερικό της παγίδας και εισέρχονται από την τρύπα στην βάση της ανεστραμμένης χοάνης. Τα έντομα προσπαθούν να ξεφύγουν ακολουθώντας το φως που εισέρχεται από το διάφανο καπάκι. Μόλις εξαντληθούν πέφτουν στην βάση της χοάνης και πνίγονται στο νερό που έχει τοποθετηθεί σε αυτή (Εικόνα 2.5). [Vossen et al., 2004; Vossen, 2014]

OLIPE (Olivarera los Pedroches) παγίδες (Εικόνα 2.6) πρόκειται για μια σπιτική παγίδα που αποτελείται από ένα δίλιτρο πλαστικό μπουκάλι με μικρές οπές στο λαιμό του. Το μπουκάλι γεμίζεται με νερό αναμειγμένο με ένα ελκυστικό υλικό το οποίο θα προσελκύσει τον ενήλικα δάκο. [Zalom et al., 2009; Vossen, 2014]



Εικόνα 2.4: Παγίδα τύπου McPhail εγκατεστημένη σε ελαιώνα

Κίτρινες παγίδες κόλλας πρόκειται για μια κολλώδης, από τις δύο πλευρές, επιφάνεια εμποτισμένες με ελκυστικό με τελικό αποτέλεσμα οι δάκοι να κολλάνε στην επιφάνεια και να πεθαίνουν καθώς δεν μπορούν να ξεφύγουν. [Vossen et al., 2004; Vossen, 2014]

Συσκευή "Attract & kill" πρόκειται για ένα κωνοειδές χαρτόνι εμποτισμένο με ελκυστικό και χημικό με τελικό αποτέλεσμα ο ενήλικας δάκος να πεθαίνει μόλις έχει επαφή με το χαρτόνι.



Εικόνα 2.5: Περιεχόμενο παγίδας McPhail

Με αυτή την μέθοδο το χημικό δεν έχει επαφή με το φυτό. [Vossen, 2014]

Ηλεκτρονικές/αυτόματες παγίδες Στις μέρες μας γίνεται μια προσπάθεια με χρήση της τεχνολογίας για την βελτίωση των προαναφερθέντων παγίδων με τελικό στόχο της αυτόματη και απομακρυσμένη μέτρηση των περιεχομένων τους επιτρέποντας με αυτόν τον τρόπο την ταχύτερη αντιμετώπιση της έξαρσης του πληθυσμού του δάκου και όχι μόνο. Πιο συγκεκριμένα:

Οι Tirelli et al. [2011] παρουσίασαν μια συσκευή αυτόματης παρακολούθησης παρασιτικών εντόμων μέσω ενός ασύρματου δικτύου αισθητήρων εικόνας δικτύωσης χαμηλής κατανάλωσης. Εκεί, τοποθετήθηκαν υπαίθριες περιοχές εξέτασης κοντά σε πλάκες με στόχο την καταγραφή των παρασίτων. Στη συνέχεια, οι εικόνες της περιοχής εξέτασης κατά τη δειγματοληψία για παράσιτα συγκρίνονταν με εικόνες αναφοράς της ίδιας περιοχής εξέτασης χωρίς παράσιτα για τον υπολογισμό της διαφοράς τους. Η επεξεργασία εικόνας περιελάμβανε επίσης μείωση θορύβου και μετατροπή σε δυαδική εικόνα, η οποία πραγματοποιούνταν σε κεντρικό διακομιστή. Ο πειραματισμός έδειξε συσχέτιση των παρασίτων που εντοπίζονταν με τις εικόνες αναφοράς.

Μια άλλη έξυπνη παγίδα παρουσιάζεται επίσης από τους Philimis et al. [2013]. Στο έργο αυτό, περιγράφετε ένα σύστημα ασύρματης παρακολούθησης που αποτελείται από αυτοματοποιημένες παγίδες με οπτικές μονάδες και μονάδες ανίχνευσης κίνησης για την συλλογή παρασιτι-



Εικόνα 2.6: Παγίδα τύπου OLIPE από Morini33^a έχει άδεια βάσει CC BY 3.0

^ahttps://es.wikipedia.org/wiki/Trampa_olipe#/media/Archivo:Trampaolipe.jpg

κών εντόμων καθώς και έναν κεντρικό σταθμό που συλλέγει, αποθηκεύει και καθιστά διαθέσιμα επεξεργασμένα αποτελέσματα, όπως τον πληθυσμό των εντόμων. Αυτό το σύστημα επιτρέπει την ανάλυση σε πραγματικό χρόνο εικόνων παρασίτων που οδηγούν σε προσδιορισμός του είδους του (Medfly ή Dacus) καθώς και του φύλου του. Η εργασία είναι μέρος του έργου της Ε.Ε. «e-FlyWatch» [European Commission - Community Research and Development Information Service, 2017], όπου η μεθοδολογία βασίζεται σε χωρικά χαρακτηριστικά που συγκρίνουν το αναγνωρισμένο παράσιτο με πρότυπα σε συνδυασμό με τεχνικές αναγνώρισης βάσει μοτίβου για «μοναδικά χαρακτηριστικά εντόμων, όπως για παράδειγμα την κοιλιά, τα φτερά και ο θώρακας» [European Commission - Community Research and Development Information Service, 2014].

Οι Wang et al. [2016] παρουσίασαν το σχεδιασμό ενός συστήματος αναγνώρισης εικόνας χρησιμοποιώντας υπολογιστή μεθόδους όρασης για μια μεγάλη ποικιλία μύγας φρούτων που επηρεάζουν αρνητικά το διεθνές εμπόριο φρούτων. Το προτεινόμενο σύστημα αντιμετωπίζει 74 είδη που τα περισσότερα αποτελούν παράσιτα της οικογένειας των Tephritidae. Το σύνολο δεδομένων τους περιλαμβάνει τρία μέρη του σώματος, τα πτέρυγια, την κοιλιά και τον θώρακα από τις μύγες των φρούτων με βάση τα οποία πραγματοποιείται η αναγνώριση είτε μεμονωμένα είτε με οποιονδήποτε συνδυασμό. Τα βήματα της διαδικασίας αναγνώρισης περιλαμβάνουν διόρθωση γάμμα για διάφορα θέματα φωτισμού, πολλαπλών προσανατολισμών, λειτουργιών πολλαπλών κλιμάκων με βάση Φίλτρα Gabor και ταξινόμηση k-NN. Ο πειραματισμός έδειξε ένα συνολικό ποσοστό επιτυχίας ταξινόμησης 87% σε επίπεδο ειδών.

Οι Doitsidis et al. [2017] παρουσίασαν μια αυτοματοποιημένη παγίδα McPhail για την παρακολούθηση του πληθυσμού του δάκου. Ο σχεδιασμός της έξυπνης παγίδας επιτρέπει τη λήψη εικόνων από τα περιεχόμενα της παγίδας που στη συνέχεια μεταδίδονται σε διακομιστή και υποβάλλονται σε επεξεργασία προκειμένου να εντοπιστούν οι δάκοι. Η μεθοδολογία για την αναγνώριση των δάκων στις εικόνες βασίστηκε σε διαδικασίες επεξεργασίας εικόνας, όπως αυτόματη διόρθωση φωτεινότητας, ανίχνευση ακμής (CLF), μείωση θορύβου κ.α. Η εκπαίδευση έγινε με τη χρήση 100 χειροκίνητα επισημασμένων εικόνων που οδηγούν σε υπολογισμό μέσο μέγεθος (από την άποψη του ποσοστού της μαύρης περιοχής - μαύρα εικονοστοιχεία) μίας μίας μύγας φρούτων ως ποσοστό της την περιοχή ενδιαφέροντος, δηλαδή την παγίδα, καθιστώντας έτσι απαραίτητη τη χρήση συγκεκριμένης παγίδας ή γνώσης της παγίδας διαστάσεις. Οι εκτεταμένες δοκιμές που έγιναν έδειξαν ότι η ακρίβεια αναγνώρισης έφτασε το 75%.

Οι Potamitis et al. [2017b] πρότειναν την τροποποίηση της τυπικής, χαμηλού κόστους πλαστικής παγίδας μύγας φρούτων (McPhail) με την προσθήκη οπτικοηλεκτρονικών αισθητήρων που παρακολουθούν την είσοδο της παγίδας με στόχο την ανίχνευση και προσδιορισμό του είδους των εισερχόμενων εντόμων, μέσω της οπτικοακουστικής ανάλυσης του φάσματος του χτυπήματος του φτερού τους. Η αναγνώριση γίνεται με βάση τη σύγκριση του πλάτους του χρόνου καταγραφής ενός εισερχόμενου εντόμου στην παγίδα με τις εγγραφές αναφοράς που προέρχονται από πολλές καταγραφές ίδιων εντόμων. Ο εργαστηριακός πειραματισμός έδειξε 0,93 και 0,95 μέση βαθμολογία F1 για όλες τις μύγες φρούτων που δοκιμάστηκαν και για το δάκο της ελιάς, αντίστοιχα.

Οι Shaked et al. [2017] παρουσίασαν το σχέδιο δύο έξυπνων παγίδων για τέσσερα είδη μύγας φρούτων με ποικιλία διαμορφώσεων, οι οποίες χρησιμοποιούν ασύρματη επικοινωνία για να

διαδώσουν τη λήψη εικόνων παγιδευμένων εντόμων σε κολλώδεις επιφάνειες. Το έργο δεν περιείχε μια αυτοματοποιημένη μεθοδολογία αναγνώρισης μύγας φρούτων, ενώ φαίνεται να είναι επέκταση του προηγούμενου έργου των Alorda et al. [2015] (υπό την αιγίδα του ίδιου έργου, "FruitFlyNet") παρουσιάζοντας μια ενεργειακά αποδοτική και χαμηλού κόστους παγίδα για παρακολούθηση του δάκου της ελιάς χρησιμοποιώντας ένα ασύρματο δίκτυο αισθητήρων.

2.2.4.2 Ψεκάσμοι

Πηλός Καολίνης πρόκειται για ένα ορυκτό το οποίο ψεκάζεται στον ελαιώνα και δημιουργεί ένα προστατευτικό φράγμα γύρω από τον καρπό της ελιάς με αποτέλεσμα να δυσκολεύει την εναπόθεση αυγών μέσα σε αυτό. [Vossen, 2014]

Δόλωμα Spinosad πρόκειται για ένα εντομοκτόνο αναμειγμένο με ελκυστικό δόλωμα το οποίο ψεκάζεται σε ένα μικρό τμήμα του δένδρου. Ο ενήλικας δάκος ελκύεται από το δόλωμα, τρέφεται από αυτό και στην συνέχεια πεθαίνει. [Vossen et al., 2004]

2.3 Μικροκλίματα

Όπως προαναφέρθηκε στην υπο-ενότητα 2.2.1 η δυναμική του πληθυσμού του δάκου επηρεάζεται από τους κλιματολογικούς παράγοντες, κυρίως θερμοκρασία και υγρασία, της περιοχής στην οποία δραστηροποιείται. Οι παράμετροι αυτοί επηρεάζονται με τη σειρά τους από μια πληθώρα περιβαλλοντικών / κλιματικών συνθηκών, όπως παραλλαγές της ηλιακής ακτινοβολίας, νεφοκάλυψη, η σκίαση και η εγγύτητα σε θάλασσα / βουνά, μεταξύ άλλων. Για παράδειγμα, η θερμοκρασία είναι γνωστό ότι μειώνεται όσο το υψόμετρο αυξάνεται, γνωστό ως lapse rate, που κυμαίνεται από 5,5 °C, έως 6,5 °C, ανά 1000 μέτρα σε υψόμετρο [Hodkinson, 2005]. Όλοι αυτοί οι παράγοντες συνιστούν στην δημιουργία μικροκλιμάτων. Έχει βρεθεί ότι τα μικροκλίματα παίζουν σημαντικό ρόλο στη φυσιολογία ενός οργανισμού [Holmes and Dingle, 1965; Kearney et al., 2014a], ειδικά για έντομα όπως ο δάκος της ελιάς. Λόγω της σημασίας της παραγωγής ελιάς και της ευρείας παρουσίας της, ένας ελαιώνας μπορεί να ανήκει σε περισσότερα από ένα μικροκλίμα και επομένως η αντιμετώπιση του δάκου απαιτεί αναγνώριση αυτών των μικροκλιμάτων.

2.3.1 Σχετική έρευνα για την αναγνώριση μικροκλιμάτων

Ο Cantlon [1953] πραγματοποίησε μια μελέτη στις πλαγιές μιας κορυφογραμμής με νότιο και βόρειο προσανατολισμό, στο κεντρικό Νιου Τζέρσεϋ, ΗΠΑ, όσον αφορά τις διαφορές σε μια μεγάλη ποικιλία βλάστησης που βασίζεται σε μικροκλίματα, μέσω μετρήσεων από το πεδίο που συλλέχθηκαν σε διάστημα δύο χρόνων. Τα αποτελέσματα που αναφέρονται από τον συγγραφέα δείχνουν σημαντική επίδραση των μικροκλιμάτων στη "δομή και σύνθεση της βλάστησης στις δύο πλαγιές". Επιπλέον, όσον αφορά τις μικροκλιματικές συνθήκες, ο συγγραφέας παρατήρησε ότι μεταξύ των πλαγιών, σε σύγκριση μεταξύ μικροκλιματικών επιπέδων, οι μεγαλύτερες διαφορές στην ατμοσφαιρική υγρασία παρατηρήθηκαν στα χαμηλότερα επίπεδα

(*“στα 5 cm, το χαμηλότερο επίπεδο που παρατηρήθηκε”*) των πλαγιών ενώ παρατηρήθηκε μείωση της διαφοράς στα υψηλότερα επίπεδα (π.χ. *“στα 2 m, το υψηλότερο επίπεδο που παρατηρήθηκε”*).

Οι Suzanne Van Cooten et al. [2002] προσπάθησαν να εντοπίσουν τα μικροκλίματα, σε σχέση με τις βροχοπτώσεις, στη λεκάνη της λίμνης Pontchartrain της Νοτιοανατολικής Λουιζιάνας, ΗΠΑ, αναλύοντας περιβάλλοντολογικά δεδομένα από 17 σταθμούς που είχαν τοποθετηθεί σε ολόκληρη τη λεκάνη από το 1870 έως το 2000. Μέσω της στατιστικής ανάλυσής τους οι συγγραφείς διαπίστωσαν ότι υπήρχε στατιστικά σημαντική διαφορά μεταξύ του μηνιαίου μέσου όρου βροχοπτώσεις στους σταθμούς της βόρειας και της νότιας ακτής.

Η αναγνώριση μικροκλιμάτων στις αστικές περιοχές έχει αποκτήσει μεγάλη προσοχή τα τελευταία χρόνια. Συγκεκριμένα, στο έργο του Shafieiyouni [2014], έγινε μια προσπάθεια εντοπισμού μικροκλιμάτων μέσα στην πόλη Ισφαχάν, Ιράν, χωρίζοντάς την σε πέντε περιοχές με βάση τις επιφάνειές τους, μετρώντας τις περιβάλλοντολογικές παραμέτρους κάθε περιοχής για περίοδο 12 μηνών. Τα αποτελέσματα που προέκυψαν έδειξαν ότι υπάρχουν σημαντικές διαφορές ($p < 0,01$) για όλους τους 12 μήνες του έτους για τη θερμοκρασία του αέρα και τη σχετική υγρασία μεταξύ των σταθμών της πόλεως και των σταθμών αναφοράς με μέγιστη διαφορά στο μέσο όρο μηνιαίας θερμοκρασίας αέρα $6,07\text{ }^{\circ}\text{C}$ και μέγιστη διαφορά στη μέση μηνιαία σχετική υγρασία ίση με 40,73%.

Μια άλλη αστική μελέτη [Dimoudi et al., 2013] ανέλυσε, μέσω μετρήσεων πεδίου, πώς οι μικροκλιματικές συνθήκες επηρεάζουν τις θερμικές συνθήκες των στενών της πόλης των Σερρών, Ελλάδα. Οι συγγραφείς έλαβαν αποτελέσματα που δείχνουν ότι *“η ταχύτητα του ανέμου στο επίπεδο των πεζών (1,8 m) είναι στο $\frac{1}{3}$ με $\frac{1}{4}$ της προαστιακής περιοχής”* καθώς και ότι οι θερμοκρασίες του αέρα στην περιοχή μελέτης είναι περίπου $5,0\text{-}5,5\text{ }^{\circ}\text{C}$ υψηλότερες από ό,τι στην προαστιακή περιοχή, κατά τη διάρκεια του απογεύματος και της νύχτας, ενώ το πρωί, οι θερμοκρασίες του αέρα στην πόλη είναι $7,0\text{ }^{\circ}\text{C}$ χαμηλότερη, καταλήγοντας στο συμπέρασμα ότι η γεωμετρία της πόλης είναι σημαντική στην επίδραση των μικροκλιματικών συνθηκών σε μια αστική περιοχή.

Οι Wong and Jusuf [2010] διερεύνησαν πώς οι συνθήκες του μικροκλίματος διαφέρουν μεταξύ δύο πεζόδρομων στη Σιγκαπούρη, όσον αφορά το πράσινο και τη διανομή κτιρίων. Τα αποτελέσματα που προέκυψαν έδειξαν ότι *“η μέση θερμοκρασία του αέρα μέσα σ’ ένα πεζόδρομο με δέντρα είναι χαμηλότερη κατά περίπου $0,7\text{-}1,18\text{ }^{\circ}\text{C}$ ”* σε σύγκριση με έναν πεζόδρομο χωρίς δέντρα κατά τη διάρκεια της ημέρας, ενώ επίσης ο πεζόδρομος με δέντρα *“διατηρεί επίσης τη δροσιά του περίπου $0,4\text{-}0,58\text{ }^{\circ}\text{C}$ κατά τη διάρκεια της νύχτας”*. Φυσικά, η σχετική υγρασία μετρήθηκε να είναι αυξημένη έως 5% κατά μέσο όρο στο πεζόδρομο με δέντρα σε σύγκριση με το πεζόδρομο χωρίς δέντρα. Έτσι, οι συγγραφείς συμπέραναν ότι η ποσότητα της βλάστησης και της σκιάς από τα γύρω κτίρια είχε επιπτώσεις στο δροσιά των εξεταζόμενων πεζόδρομων.

Οι Stabler et al. [2005] μελέτησαν πώς η φυτική κάλυψη και η χρήση της γης μέσα στον αστικό ιστό σχετίζοντε μεταξύ τους για τη διαμόρφωση μικροκλιμάτων στην πόλη του Φοίνιξ της Αριζόνα, ΗΠΑ. Μέσα από τα ευρήματά τους, κατέληξαν στο συμπέρασμα ότι τα μικροκλίματα είναι μια αλληλεπιδραστική επίδραση της πυκνότητας των φυτών και των αστικών οντοτήτων όπως χώροι στάθμευσης και κτίρια, σε αντίθεση με τη μη υποστηριζόμενη υπόθεση ότι *“η κάλυψη του αστικού δάσους και οι κρυμμένες θερμικές ροές είναι οι βασικοί καθοριστικοί*

παράγοντες του μικροκλίματος στην περιοχή του Φοίνιξ”.

Οι Shahrestani et al. [2015] διερεύνησαν τα χαρακτηριστικά των αστικών μικροκλιμάτων στην πόλη του Λονδίνου, Ηνωμένο Βασίλειο. Τα ευρήματά τους περιλαμβάνουν ότι *”κτίρια σε μια αστική περιοχή λειτουργούν ενάντια στις δικές τους μεμονωμένες μικροκλιματικές μεταβλητές και όχι τα μετεωρολογικά δεδομένα καιρού”* και πως έτσι, η αστική απόδοση και η θερμική και ενεργειακή απόδοση των κτιρίων απαιτούν σημαντική αξιολόγηση. Κατέληξαν στο συμπέρασμα ότι τα αστικά μικροκλίματα είναι αποτέλεσμα της διάταξης και της διαμόρφωσης των κτιρίων, καθώς τα ψηλά κτίρια μπορούν να εμποδίσουν το ηλιακό φως και να μειώσουν τη διαπερατότητα του ανέμου.

Ωστόσο, η έρευνα σχετικά με την πρόβλεψη των μικροκλιμάτων υπήρξε επίσης στο προσκήνιο. Οι Zhang et al. [1997] χρησιμοποίησαν το μοντέλο Energy Balance που πρότειναν οι Avissar and Mahrer [1982] για να προβλέψει το μικροκλίμα μέσα σε ένα θερμοκήπιο επιτυγχάνοντας μέσες τετραγωνικές διαφορές μεταξύ των προβλεπόμενων και των πραγματικών θερμοκρασιών αέρα και φύλλων, σχετική υγρασία και διάρκεια υγρασίας φύλλων στους 1,2 και 1,8 °C, 5,8% και 1,9 h · d⁻¹ αντίστοιχα. Ομοίως, οι Wang and Boulard [2000] προέβλεψαν το μικροκλίμα ενός θερμοκηπίου χρησιμοποιώντας το Gembloux Greenhouse Dynamic μοντέλο επιτυγχάνοντας αποκλίσεις μεταξύ των προβλεπόμενων και πειραματικών παραμέτρων, θερμοκρασία εδάφους, εσωτερική θερμοκρασία αέρα, σχετική υγρασία και διαπνοή συγκομιδής σε 0,5 °C, 0,8 °C, 4,3% και 17,8 W / m², αντίστοιχα. Τέλος, οι Kearney et al. [2014b] πρότειναν ένα μοντέλο μικροκλίματος για την πρόβλεψη μικροκλιματικών συνθηκών σε ηπειρωτική κλίμακα χρησιμοποιώντας δεδομένα για το έδαφος και τον καιρό, επιτυγχάνοντας πρόβλεψη της διακύμανσης όσον αφορά την ωριαία θερμοκρασία εδάφους στο 85% των περιπτώσεων με ακρίβεια 2–3,3 °C.

2.3.2 Επιρροή μικροκλιμάτων στο δάκο της ελιάς

Στη μελέτη τους, οι Kounatidis et al. [2008], μέσω ενός δικτύου 700 δακοπαγίδων, διερεύνησαν πώς οι εστίες πυκνότητας των δάκων αλλάζουν μεταξύ των εποχών σε σχέση με το υψόμετρο. Η ομαδοποίηση των παγιδευμένων δάκων αποδείχθηκε ότι σχετίζεται σημαντικά με το υψόμετρο (που κυμαίνεται από τη στάθμη της θάλασσας έως τα 700 μέτρα πάνω από την στάθμη της θάλασσας) με $p < 0,01$. Οι συγγραφείς διαπίστωσαν ότι κατά το καλοκαίρι οι κλιματολογικές συνθήκες, για υψόμετρο πάνω από 200 μέτρα, ήταν ευνοϊκές για την ανάπτυξη του δάκου και έτσι οδήγησαν στο σχηματισμό εστιών στις δακοπαγίδες, ενώ σε χαμηλότερα υψόμετρα οι μετρήσεις στις δακοπαγίδες ήταν χαμηλές. Από την άλλη πλευρά, κατά τη διάρκεια του φθινοπώρου, οι κλιματολογικές συνθήκες κάτω των 200 μέτρων έγιναν πιο ευνοϊκές για το δάκο, με αποτέλεσμα την δημιουργία εστιών με υψηλές στις δακοπαγίδες και συνεπώς ο αριθμός δάκων στις μετρήσεις των δακοπαγίδων σε υψηλότερα υψόμετρα μειώθηκε.

3. ΕΦΑΡΜΟΓΕΣ ΠΡΟΣΟΜΟΙΩΣΗΣ

Σε αυτό το κεφάλαιο, στην ενότητα 3.1 παρουσιάζεται η αρχιτεκτονική του λογισμικού προσομοίωσης και περιγράφεται κάθε μονάδα από την οποία απαρτίζεται. Στην συνέχεια, στην ενότητα 3.2 παρουσιάζονται τα πειράματα, και η αξιολόγηση των αποτελεσμάτων τους, που διενεργήθηκαν για την επιλογή του μοντέλου διασποράς, που θα κάνει χρήση το λογισμικό προσομοίωσης, με βάση την καρποφορία. Τέλος, στην ενότητα 3.3 παρουσιάζεται η εκτέλεση του λογισμικού προσομοίωσης με στόχο την ανάδειξη της σημαντικότητας της συμπερίληψης των μικροκλιμάτων που υπάρχουν στους ελαιώνες ως παράγοντας στην λήψη απόφασης για την εφαρμογή μέτρων αντιμετώπισης της δακοπροσβολής.

3.1 Λογισμικό προσομοίωσης Δάκου

Το λογισμικό προσομοίωσης που περιγράφεται εδώ, αναπτύχθηκε αρχικά κατά την διάρκεια του προγράμματος "Δάκος"¹ και έχει βελτιωθεί περαιτέρω για να προσομοιώνει καλύτερα την χωροχρονική δυναμική του πληθυσμού του Δάκου. Οι κύριες προσθήκες/αλλαγές που πραγματοποιήθηκαν είναι οι ακόλουθες:

- Προσθήκη δυνατότητας διαχωρισμού προσομοιωμένου ελαιώνα σε μικροκλίματα.
- Προσθήκη δυνατότητας απεικόνισης χρονικής εξέλιξης πληθυσμού για κάθε μικροκλίμα ξεχωριστά.
- Ορισμός ορίων θερμοκρασιών ανάπτυξης για κάθε πρώιμο στάδιο ανάπτυξης του δάκου.
- Εισαγωγή της θερμοκρασίας ως παράγοντα επιροής της διασποράς του δάκου.
- Εισαγωγή του ποσοστού καρποφορίας του ελαιώνα ως παράγοντα επιροής της διασποράς του δάκου.

3.1.1 Μονάδες

Σε αυτήν την ενότητα περιγράφεται κάθε module από το οποίο απαρτίζεται το λογισμικό προσομοίωσης, τα οποία είναι τα ακόλουθα:

- το Grid module,
- το Trap module,
- το Populate module,

¹<http://events.di.ionio.gr/diold/content/research/dakos/>

- το Simulation module,
- το Grouping module,
- το Plot module and
- το Display module.

Τα παραπάνω modules περιγράφονται στις ακόλουθες υπο-ενότητες

3.1.1.1 Grid module

Το Grid module δημιουργεί το πλέγμα το οποίο θα αναπαριστά το πεδίο όπου θα λάβει χώρα η προσομοίωση. Για την δημιουργία του πλέγματος απαιτούνται τρία αρχεία εικόνων, όπου κάθε εικόνα περιγράφει ένα επίπεδο πληροφορία για το πραγματικό πεδίο, να δοθούν ως είσοδο. Κάθε είδος πληροφορίας μεταδίδεται μέσω των RGB τιμών κάθε pixel μέσα στην εικόνα. Να σημειωθεί ότι και οι τρεις εικόνες πρέπει να είναι ίδιων διαστάσεων.

Το πρώτο επίπεδο (Εικόνα 3.1) περιέχει πληροφορία σχετικά με την τοποθεσία των ελαιώνων. Λευκά pixel απότυπώνουν περιοχές χωρίς ελαιώνες (αστικές περιοχές, δρόμους, θάλασσα κλπ.). Από την άλλη μεριά, μαύρα pixel απότυπώνουν περιοχές με ελαιώνες.



Εικόνα 3.1: Δείγμα εισόδου επιπέδου ελαιώνων

Το δεύτερο επίπεδο (Εικόνα 3.2), περιέχει πληροφορία σχετικά με το ποσοστό καρποφορίας των ελαιώνων. Το χρώμα κάθε pixel έχει την ακόλουθη RGB μορφοποίηση $[0, X, 0]$, όπου X παίρνει τιμές μεταξύ 0 και 100 (διαφορετικές αποχρώσεις του πράσινου χρώματος). Μηδενικές τιμές υποδεικνύουν, όπως είναι φυσικό, ότι ο ελαιώνας δεν φέρει καθόλου καρπούς ελιάς.

Σε αντίθεση, τιμές του 100 υποδεικνύουν ότι κάθε ελαιόδενδρο μέσα στον ελαιώνα έχει καρπούς.

Το τελικό επίπεδο πληροφορίας (Εικόνα 3.3) υποδεικνύει τα μικροκλίματα στα οποία είναι χωρισμένο το πεδίο. Τα pixel της εικόνας έχουν την ακόλουθη RGB μορφοποίηση $[X, X, X]$, όπου X μπορεί να έχει τιμές, διαιρούμενες με το 10, μεταξύ 10 και 250. Ως επακόλουθο, δύνεται η δυνατότητα το πεδίο να διακριθεί μέχρι και σε 250 μικροκλίματα. Παραδείγματος χάρη, pixel με RGB τιμές $[10, 10, 10]$ ανήκουν στο μικρόκλιμα Α, ενώ pixel με RGB τιμές $[20, 20, 20]$ ανήκουν στο μικρόκλιμα Β

Το module αναλύει όλες τις εικόνες και δημιουργεί ένα πλέγμα ίδιων διαστάσεων με των εικόνων που δόθηκαν ως είσοδο. Με αυτό το τρόπο επιτυγχάνεται μια προς μια χαρτογράφηση κάθε pixel με κελί του πλέγματος. Κάθε κελί του πλέγματος διατηρεί την ακόλουθη πληροφορία:

- αν αντιπροσωπεύει ελαιώνα ή όχι,
- το ποσοστό καρποφορίας (0% στην περίπτωση που δεν αντιπροσωπεύει ελαιώνα),
- το μικρόκλιμα στο οποίο ανήκει.

3.1.1.2 Trap module

Το Trap module αναγνωρίζει πραγματικές δακοπαγίδες που έχουν τοποθετηθεί στο πεδίο και τοποθετεί ένα εικονικό δίκτυο δακοπαγίδων πάνω από το εικονικό πεδίο που δημιουργήθηκε στην υπο-ενότητα 3.1.1.1.



Εικόνα 3.2: Δείγμα εισόδου επιπέδου καρποφορίας των ελαιώνων



Εικόνα 3.3: Δείγμα εισόδου επιπέδου μικροκλιμάτων

3.1.1.3 Populate module

Αυτό το module είναι υπεύθυνο για τον αρχικό πληθυσμό του που θα περιέχεται μέσα στο εικονικό πεδίο. Για να το επιτύχει αυτό, το module τροφοδοτείται με μετρήσεις δακοπαγίδων από το πραγματικό πεδίο. Κάθε κελί έχει έναν αριθμό ο οποίος υποδεικνύει το συνολικό αριθμό δάκων στην περιοχή.

Έτσι, αυτό το module παράγει σαν έξοδο τον συνολικό αριθμό δακάων σε κάθε κελί του πλέγματος, το οποίο θα χρησιμοποιηθεί στην συνέχεια από το Simulate module ως ο αρχικός πληθυσμός της προσομοίωσης.

3.1.1.4 Simulation module

Σε αυτήν την υπο-ενότητα, παρουσιάζεται ο πυρήνας του λογισμικού προσομοίωσης, το οποίο είναι υπεύθυνο για την χωρο-χρονική προσομοίωση της εξέλιξης του πληθυσμού του Δάκου. Το συγκεκριμένο module απαιτεί τα ακόλουθα για να λειτουργήσει:

- το εικονικό πεδίο που δημιουργήθηκε στην υπο-ενότητα 3.1.1.1,
- περιβαλλοντολογικά δεδομένα, για κάθε μικρόκλιμα. Η τρέχουσα έκδοση του λογισμικού, αναγνωρίζει μόνο θερμοκρασία. Είναι σημαντικό να σημειωθεί ότι το σερ θερμοκρασίας κάθε μικροκλίματος είναι καθολικά το ίδιο για κάθε σημείο που ανήκει στο συγκεκριμένο μικρόκλιμα.
- ψεκασμοί, εάν υπάρχουν.

- τον αρχικό πληθυσμό δάκων (βλ. υπό-ενότητα 3.1.1.3).
- παραμέτρους ανάπτυξης του δάκου. Συγκεκριμένα, ο χρήστης μπορεί να θέσει τιμές σχετικά με το συνολικό αριθμό βαθμο-ημερών που χρειάζεται κάθε στάδιο ανάπτυξης για να ολοκληρωθεί καθώς επίσης τα άνω και κάτω κατώφλια θερμοκρασίας για κάθε στάδιο ανάπτυξης.
- το συνολικά βήματα προσομοίωσης. Να σημειωθεί ότι η χρονική κλίμακα προσομοίωσης είναι ωριαία. Έτσι, ο αριθμός βημάτων αντιπροσωπεύει ώρες

Η διαδικασία της προσομοίωσης περιγράφεται, παρακάτω, από τον Αλγόριθμο 1.

```
for timeStep in TotalTimeSteps, do:
  for fly in OliveFruitFlies, do:
    spray(timeStep, fly)

    if fly.state == alive, then:

      if timeStep is morning and fly.stage == matureAdult, then:
        fly.layEgg()

        if fly.stage == adult and timeStep is dayHour, then:
          fly.drift(grid, temperature)

    fly.age(temperature)
```

Δομή πληθυσμού ο πληθυσμός αποτελείται από τα ακόλουθα στάδια ανάπτυξης:

- αυγό,
- Προνύμφη
- Νύμφη
- Ανώριμος ενήλικας δάκος
- Ενήλικας δάκος

Ανάπτυξη όπως αναφέρθηκε στην υποενότητα 2.2.1 η ανάπτυξη του δάκου επηρεάζεται από τις περιβαλλοντικές συνθήκες της εκάστοτε περιοχής στην οποία δραστηροποιείται. Για αυτό τον λόγο για την προσομοίωση της ανάπτυξης του δάκου χρησιμοποιείται το Degree Day (DD) μοντέλο (Ενότητα 1.4) για την συσχέτιση της θερμοκρασίας του περιβάλλοντος με την ανάπτυξη του. Ο ακόλουθος τύπος χρησιμοποιείται για τον υπολογισμό των ημεροβαθμών.

$$DD(t_i) = (t_i - T_L) * (1 - \frac{1}{1 + e^{-10*(t_i - T_U)}}) \quad (3.1)$$

όπου t_i η θερμοκρασία που επικρατεί στο ι-οστό βήμα της προσομοίωσης, T_L και T_U τα κάτω και άνω κατώφλια θερμοκρασίας ανάπτυξης, αντίστοιχα, για κάθε έντομο. Όστόσο, βάση της βιβλιογραφίας, η χρήση του Degree Hour μοντέλου θεωρείται ότι είναι πιο ακριβής [Gu et al., 2014]. Με βάση αυτό η ακόλουθη συνάρτηση χρησιμοποιείται για τον υπολογισμό των ημεροωρών χρησιμοποιώντας ως βάση την συνάρτηση 3.1

$$DH(t_i) = \frac{DD(t_i)}{24} \quad (3.2)$$

Αναπαραγωγή Δεν γίνεται διάκριση στο φύλο. Όλοι οι ενήλικοι δάκοι μπορούν να γεννήσουν έως και τρία αυγά στη διάρκεια ζωής τους και το πολύ ένα αυγό ανά βήμα προσομοίωσης. Αυτός ο περιορισμός στον αριθμό των συνολικών αυγών ανά δάκο, επιλέχθηκε, έτσι ώστε ο συνολικός πληθυσμός να μην αυξάνεται γρήγορα. Επιπλέον, οι δάκοι μπορούν να γεννήσουν αυγά μόνο τις πρωινές ώρες.

Διασπορά Η διασπορά του δάκου εντός του προσομοιωμένου πεδίου επηρεάζεται από δύο παράγοντες, τη θερμοκρασία και το ποσοστό καρποφορίας. Στην ενότητα 3.2 περιγράφεται η ανάπτυξη και επιλογή του μοντέλου διασποράς του δάκου.

Έξοδος Κατά την διάρκεια της προσομοίωσης δημιουργείται ένα αρχείο καταγραφής όπου καταγράφεται η δραστηριότητα των ενήλικων δάκου σε κάθε βήμα προσομοίωσης.

3.1.1.5 Grouping module

Αυτή η μονάδα είναι υπεύθυνη για την ανάλυση του αρχείου καταγραφής προσομοίωσης, που δημιουργείται από τη μονάδα προσομοίωσης, και για την ομαδοποίηση των δάκων, για κάθε βήμα προσομοίωσης, σε συστάδες $X \times Y$, όπου τα X και Y ορίζονται από τον χρήστη.

Η μονάδα δημιουργεί ένα νέο αρχείο καταγραφής όπου τα δεδομένα αποθηκεύονται στην ακόλουθη φόρμα,

timeStep, blockX, blockY, numberOfOff

όπου το *TimeStep* είναι το χρονικό βήμα της προσομοίωσης, το *blockX* και το *blockY* είναι οι συντεταγμένες του εκάστοτε μπλοκ και *numberOfOff* ο συνολικός αριθμός δάκων μέσα στο τρέχον μπλοκ στο τρέχον χρονικό βήμα.

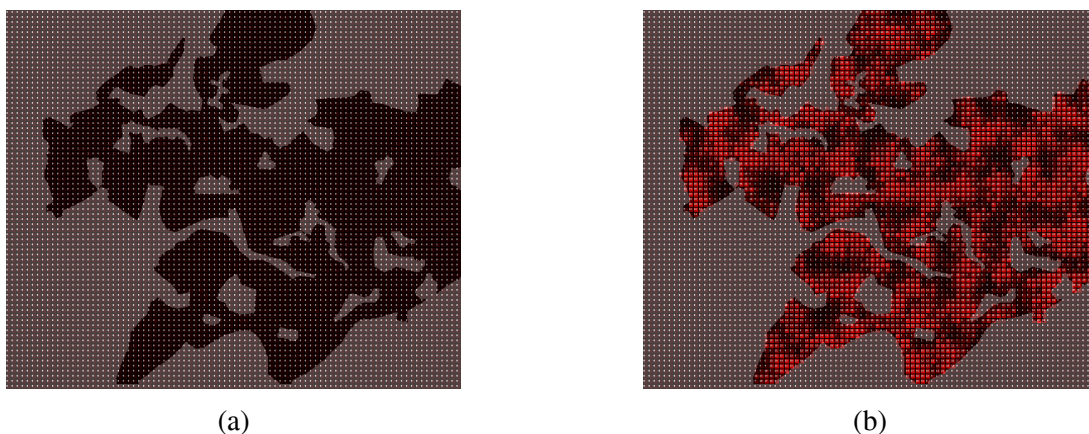
3.1.1.6 Plot module

Η μονάδα Plot δημιουργεί διαγράμματα της χρονικής εξέλιξης του πληθυσμού του δάκου με βάση το αρχείο καταγραφής που δημιουργείται από την μονάδα Grouping. Δημιουργεί μια γραφική παράσταση εξέλιξης του πληθυσμού για ολόκληρο το προσομοιωμένο πεδίο, καθώς και ξεχωριστές γραφικές παραστάσεις για κάθε περιοχή μικροκλίματος στην οποία έχει χωριστεί.

3.1.1.7 Display module

Η μονάδα προβολής είναι υπεύθυνη για τη δημιουργία στιγμιότυπων της χωρικής κατανομής των δάκων για ένα απαιτούμενο χρονικό βήμα. Απαιτούνται τα ακόλουθα για να δημιουργηθεί ένα στιγμιότυπο. Το χρονικό βήμα για να τραβήξει ένα στιγμιότυπο του πεδίου, το αρχείο εικόνας, το οποίο απεικονίζει τη θέση των ελαιώνων και τέλος το αρχείο καταγραφής που δημιουργήθηκε από την μονάδα Ομαδοποίηση.

Ως έξοδος, η ενότητα δημιουργεί ένα αρχείο εικόνας (Εικόνα 3.4) που απεικονίζει τη χωρική κατανομή των δάκων.



Εικόνα 3.4: Παράδειγμα εξόδου (a) για μηδενική προσβολή του ελαιώνα, (b) για υψηλή προσβολή του ελαιώνα

Η μονάδα σχεδιάζει ένα πλέγμα επικάλυψης πάνω στην εικόνα εισόδου όπου κάθε κελί είναι σκιασμένο με βάση τον αριθμό των δάκων που βρίσκονται μέσα σε αυτό. Όσο υψηλότερη είναι η τιμή τόσο πιο κόκκινο γίνεται το κελί, ενώ για μηδενικές τιμές το κελί είναι διαφανές. Η χρωματική σκίαση βοηθά στη διάκριση των ζωνών προσβολής καθώς και του επιπέδου της προσβολής. Επιπλέον, δημιουργείται μια έκδοση κειμένου της εικόνας εξόδου, η οποία περιέχει τον αριθμό των δάκων σε κάθε κελί.

3.1.2 Χαρακτηριστικά

Τα κύρια χαρακτηριστικά του λογισμικού προσομοίωσης είναι τα ακόλουθα:

- Εκτός από τον καθορισμό των συνόρων του ελαιώνα και τη διαθεσιμότητα του καρπού, ο χρήστης μπορεί να προσδιορίσει σε πόσα μικροκλίματα διαιρείται ο ελαιώνας καθώς και την τοποθεσία τους μέσα σε αυτόν. Αυτό είναι ένα σημαντικό χαρακτηριστικό, διότι η τοποθεσία ενός ελαιώνα μπορεί να επηρεάσει τις περιβαλλοντικές παραμέτρους όπως η θερμοκρασία ή η σχετική υγρασία. Για παράδειγμα, ένας ελαιώνας κοντά στη θάλασσα μπορεί να έχει ψυχρότερο κλίμα από έναν ελαιώνα που βρίσκεται στην ενδοχώρα. Φυσικά, το κλίμα μέσα σε έναν ελαιώνα μπορεί να επηρεαστεί από διάφορους παράγοντες όπως η τοπογραφία της περιοχής, η ηλιακή ακτινοβολία κ.λπ.
- Το λογισμικό απαιτεί να τροφοδοτηθεί με μετρήσεις δακοπαγίδων που έχουν πραγματοποιηθεί στο πεδίο. Το λογισμικό χρησιμοποιεί αυτές τις μετρήσεις για να γεμίσει πρώτα τα κελιά παγίδας, που αντιστοιχούν στη φυσική θέση των παγίδων, με τον αντίστοιχο αριθμό δάκων. Στη συνέχεια, γεμίζει όλα τα άλλα κελιά παγίδας χρησιμοποιώντας τεχνικές παρεμβολής με βάση τις μετρήσεις παγίδας που δόθηκαν ως είσοδο. Αυτή η δυνατότητα παρέχει έναν αρχικό πληθυσμό δάκων για την εκτέλεση της προσομοίωσης που βασίζεται σε δεδομένα πεδίου.
- Παρέχεται η δυνατότητα στον χρήστη να παραμετροποιήσει τις παραμέτρους που σχετίζονται με τον κύκλο ζωής του δάκου. Οι εργαστηριακές μελέτες έδειξαν ότι το χαμηλότερο αναπτυξιακό όριο για το στάδιο της προνύμφης κυμαίνεται μεταξύ 10 °C και 12.5 °C, ενώ το ανώτερο όριο κυμαίνεται μεταξύ 30 °C και 32 °C [Tsitsipis, 1980]. Διακυμάνσεις στα αναπτυξιακά κατώφλια παρατηρήθηκαν επίσης στο στάδιο του αυγού με θερμοκρασίες που κυμαίνονται από 7.5 °C και 10 °C και μεταξύ 30 °C και 32 °C, για κατώτατα και ανώτερα όρια αντίστοιχα [Tsitsipis and JA, 1977]. Ως εκ τούτου, οι παράμετροι που επηρεάζουν την ανάπτυξη του δάκου θα πρέπει να τροποποιούνται από τον χρήστη αντί να είναι κωδικοποιημένα μέσα στο λογισμικό.
- Το λογισμικό επιτρέπει επίσης στους χρήστες να προγραμματίζουν εικονικούς ψεκασμούς, που είναι μια από τις πιο χρησιμοποιούμενες μεθόδους για τον έλεγχο των παρασίτων. Η αποτελεσματικότητα των ψεκασμών εξαρτάται από τη διεξαγωγή του ψεκασμού στον σωστό χρόνο και τόπο. Επομένως, η παροχή στον χρήστη της δυνατότητας προγραμματισμού των ψεκασμών κατά τη διάρκεια της προσομοίωσης θα ήταν μεγάλη βοήθεια για τον εντοπισμό της κατάλληλης ώρας και τόπου με το βέλτιστο αποτέλεσμα για τη διεξαγωγή του ψεκασμού ή των ψεκασμών στο χωράφι.
- Το λογισμικό εξάγει τη χρονική εξέλιξη του πληθυσμού του δάκου για ολόκληρο το προσομοιωμένο πεδίο. Καθώς επίσης και χωριστές γραφικές παραστάσεις για την εξέλιξη του πληθυσμού για κάθε μικροκλίμα στο οποίο έχει χωριστεί το πεδίο. Μέσα από αυτά τα διαγράμματα ο χρήστης μπορεί να αναγνωρίσει πότε αναμένεται να υπάρξουν μία ή περισσότερες εστίες έξαρσης του πληθυσμού του δάκου. Εάν ο χρήστης προγραμματίσει ψεκασμούς, μπορεί να παρατηρήσει την επίδραση που θα έχει στον πληθυσμό του δάκου. Επιπλέον, το σύστημα εξάγει τη χωρική κατανομή του δάκου μέσα στον ελαιώνα για ένα ή περισσότερα βήματα προσομοίωσης.

Ωστόσο, προκύπτουν ορισμένα μειονεκτήματα του συστήματος. Η απουσία διεπαφής χρήστη είναι σίγουρα ένα από τα μεγαλύτερα μειονεκτήματα του προτεινόμενου συστήματος. Ανα-

γκάζοντας τον χρήστη να παρέχει την απαραίτητη είσοδο στο σύστημα χειροκίνητα και όχι μέσω κεντρικού τρόπου. Η συλλογή όλων των δεδομένων εισαγωγής και η σωστή τυποποιημένη διαμόρφωσή τους θα μπορούσε να θεωρηθεί ως ένα ακόμα μειονεκτήμα του συστήματος, καθώς μπορεί να είναι χρονοβόρα διαδικασία.

3.1.3 Μελλοντικές βελτιώσεις

Το λογισμικό προσομοίωσης μπορεί να βελτιωθεί περαιτέρω, προσομοιώνοντας την δυναμική του πληθυσμού του δάκου με μεγαλύτερη ακρίβεια, με τις ακόλουθες προσθήκες/αλλαγές:

- Προσθήκη της υγρασίας στους περιβαλλοντικούς παράγοντες που επηρεάζουν την ανάπτυξη και την εν γένει συμπεριφορά του δάκου. Όπως περιγράφηκε στην υπό-ενότητα 2.2.1, η υγρασία είναι ο δεύτερος πιο σημαντικός περιβαλλοντικός παράγοντας, μετά την θερμοκρασία, που επηρεάζει τον δάκο.
- Στην τρέχουσα έκδοση του λογισμικού η θερμοκρασία επηρεάζει την ανάπτυξη των προϊμών σταδίων ανάπτυξης του δάκου και την διασπορά του ενήλικα δάκου στον εικονικό ελαιώνα. Η εισαγωγή της θερμοκρασίας ως παράγοντας που επηρεάζει την μακροζωία, την ανάπτυξη ωοθηκών, για τους θυληκούς δάκους, και την απόδοση της πτήσης του ενήλικα δάκου θα βελτιώσει κατα πολύ την ακρίβεια προσομοίωσης της δυναμικής του πληθυσμού του.
- Μεταβαλλόμενη θερμοκρασία μέσα στο εικονικό πεδίο. Στην τρέχουσα έκδοση η θερμοκρασία είναι ίδια για όλες τις θέσεις είτε για όλο το πεδίο είτε για κάθε μικροκλίμα στο οποίο έχει χωριστεί το πεδίο.

3.2 Ανάπτυξη μοντέλου διασποράς βάση της καρποφορίας

Βάση της βιβλιογραφίας η καρποφορία σε έναν ελαιώνα επιδρά στο πως διασπείρεται ο δάκος της Ελιάς. Πιο συγκεκριμένα οι Fletcher and Karatos [1981] μέσω πειραμάτων που πραγματοποίησαν στο πεδίο έδειξαν πως η απόσταση που θα ταξιδέψει ο δάκος σε έναν ελαιώνα συσχετίζεται με το ποσοστό καρποφορίας. Αναλυτικότερα, εκτέλεσαν δύο πειράματα, στο πρώτο πείραμα ελευθέρωσαν δάκους σε ελαιώνα χωρίς καρπό και διαπίστωσαν πως σε διάστημα μίας εβδομάδας οι δάκοι είχαν διανύσει απόστασης μεγαλύτερη από τα 400 μέτρα κατά μέσο όρο (Εικόνα 3.5). Στο δεύτερο πείραμα, ελευθέρωσαν δάκους σε ελαιώνα με ποσοστό καρποφορίας 30% και διαπίστωσαν πως σε διάστημα μίας εβδομάδας οι δάκοι είχαν διανύσει απόστασης 180 μέτρα κατά μέσο όρο (Εικόνα 3.6).

Βάση αυτών των αποτελεσμάτων επιχειρήθηκε η τροποποίηση της διασποράς του δάκου στο λογισμικό προσομοίωσης με τέτοιο τρόπο ώστε να συνυπολογίζεται και το ποσοστό καρποφορίας μέσα στον ελαιώνα που βρίσκεται.

3.2.1 Μοντέλα διασποράς

Σε αυτήν την υπό-ενότητα περιγράφονται τα τρία μοντέλα διασποράς που εξετάστηκαν για την καλύτερη αναπαραγωγή των δεδομένων που παρήγαγαν στο πεδίο οι Fletcher and Karpatos [1981].

3.2.1.1 Μοντέλο 1

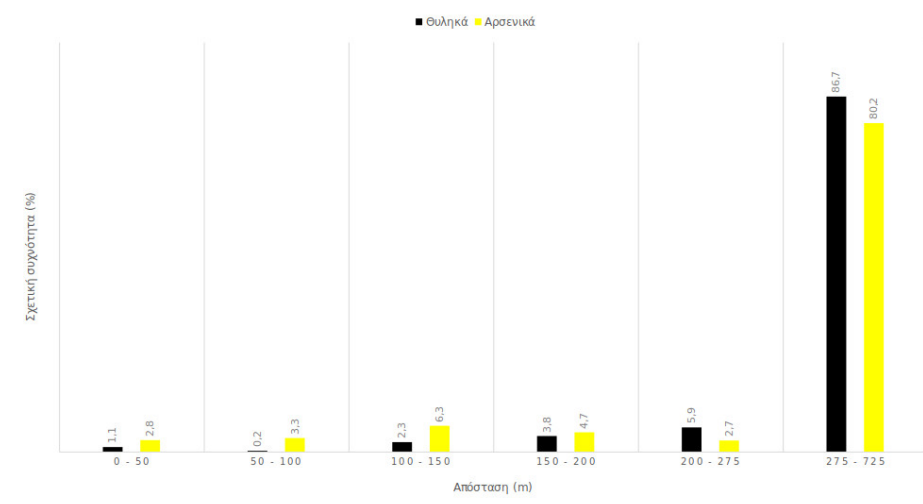
Πρόκειται για το αρχικό μοντέλο διασποράς του λογισμικού προσομοίωσης. Στο συγκεκριμένο μοντέλο οι ενήλικες δάκοι είτε μένουν στάσιμοι είτε κινούνται τυχαία προς κάποια κατεύθυνση βάση του μοντέλου τυχαίου περιπατητή. Η μέγιστη ταχύτητα με την οποία μπορούν να μετακινηθούν σε κάθε βήμα της προσομοίωσης καθορίζεται από την εξίσωση 3.3. Στην Εικόνα 3.7 παρουσιάζεται το γράφημα της εξίσωσης 3.3 με d_i ίσο με 60.

$$Speed = d_i * (1 - (\frac{1}{1 + e^{-10*(t_i - T_M)}})) \quad (3.3)$$

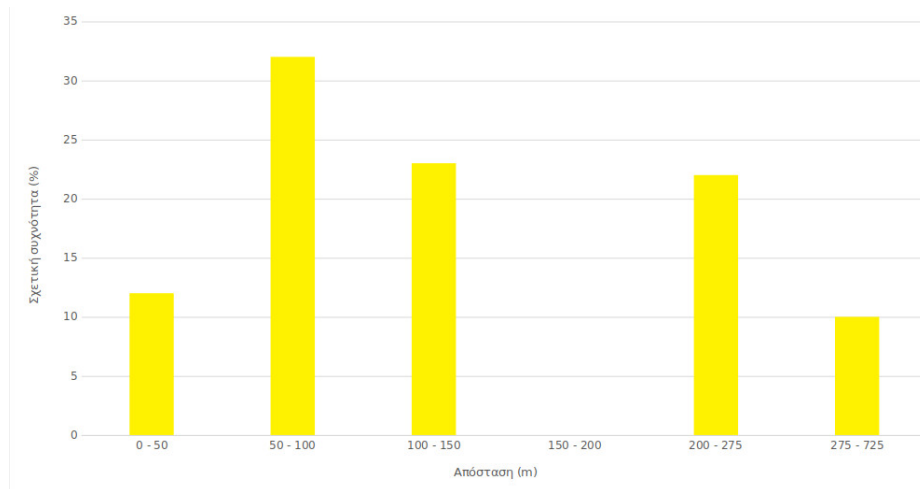
όπου d_i είναι η μέγιστη απόσταση που μπορεί να ταξιδέψει ο ενήλικας δάκος στο i -οστο βήμα προσομοίωσης, t_i είναι η θερμοκρασία του i -οστο βήμα προσομοίωσης και T_M είναι το άνω όριο θερμοκρασίας μετακίνησης.

3.2.1.2 Μοντέλο 2

$$AvgSpeed(f_{x,y}) = \frac{451.8 * e^{-0.04098*f_{x,y}}}{wh} \quad (3.4)$$



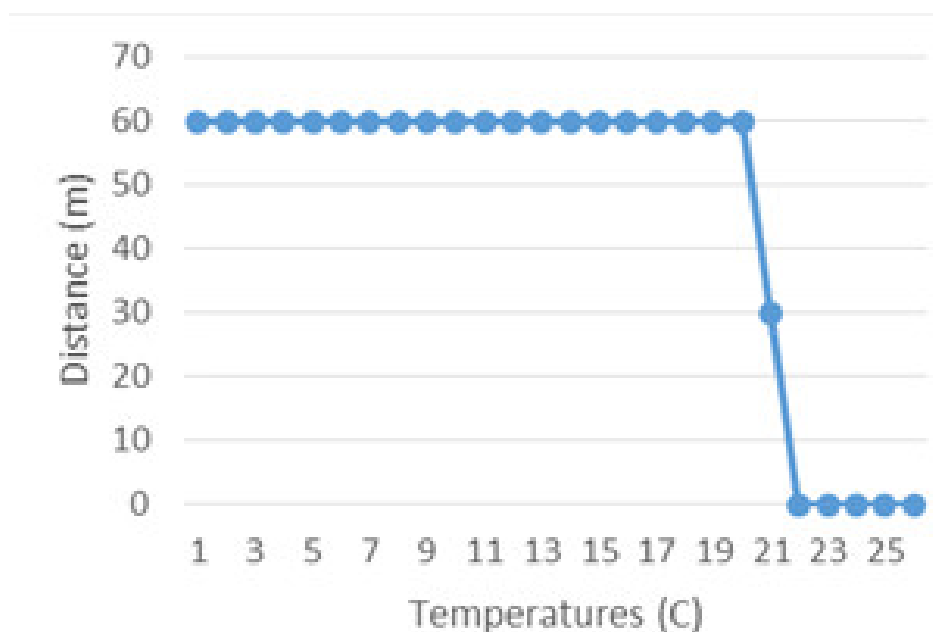
Εικόνα 3.5: Σχετικές συχνότητες δάκων σε έναν ελαιώνα χωρίς ελαιόδεντρα μεταξύ 4 – 10 ημερών μετά την αποφυλάκιση. Αποτελέσματα πειραμάτων από τους Fletcher and Karpatos [1981]



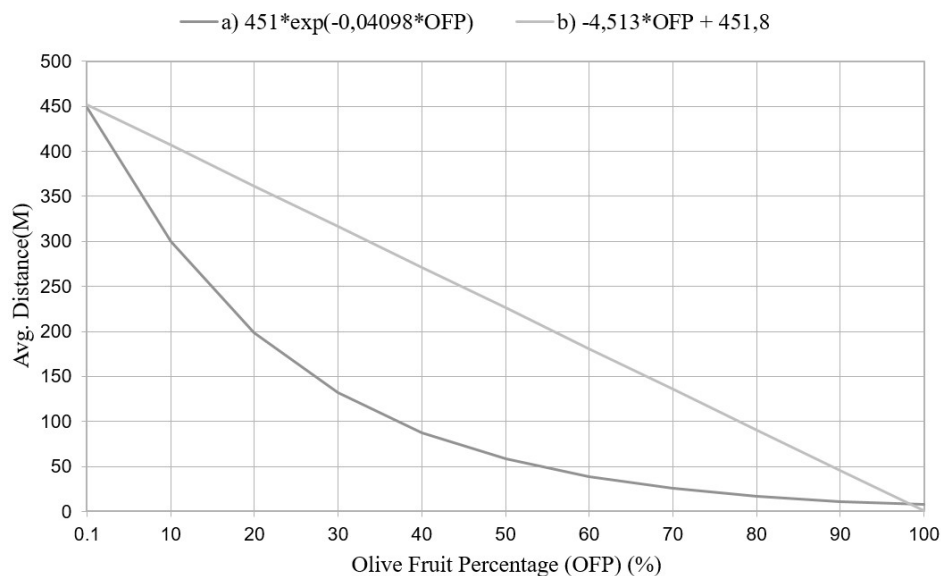
Εικόνα 3.6: Σχετικές συχνότητες δάκων σε έναν ελαιώνα με ποσοστό καρποφορίας 30% μεταξύ 4 – 6 ημερών μετά την αποφυλάκιση. Αποτελέσματα πειραμάτων από τους Fletcher and Karatos [1981]

όπου $f_{x,y}$ είναι το ποσοστό καρποφορίας στις συντεταγμένες (x, y) του πεδίου και wh το σύνολο των ωρών της ημέρας στην τρέχουσα εβδομάδα.

Επιπλέον, τροποποιήθηκε το μοντέλο τυχαίου περιπατητή που χρησιμοποιείται από τους δάκους, στο μοντέλο 1, για να διασκορπιστούν μέσα στο πεδίο με τον ακόλουθο τρόπο. Κάθε δάκος ξεκινά με 50% πιθανότητα να μετακινηθεί αριστερά ή δεξιά και πάνω ή κάτω. Με βάση το ποσοστό καρποφορίας, η πιθανότητα να κινηθεί προς την ίδια κατεύθυνση αυξάνεται, το



Εικόνα 3.7: Γράφημα Εξίσωσης 3.3 με d_i ίσο με 60 και T_M ίσο με 35.



Εικόνα 3.8: Σύγκριση συναρτήσεων μοντέλου (a) 2 και (b) 3.

οποίο σημαίνει ότι μειώνεται η πιθανότητα να κινηθεί προς την αντίθετη κατεύθυνση. αν ο δάκος βρεθεί σε σημείο με διαφορετικό ποσοστό καρποφορίας από την προηγούμενη θέση του τότε η πιθανότητα να μετακινηθεί προς οποιαδήποτε κατεύθυνση επαναφέρεται στο 50% και η διαδικασία επαναλαμβάνεται.

3.2.1.3 Μοντέλο 3

Το τρίτο μοντέλο διασποράς βασίστηκε στο προηγούμενο μοντέλο όπου έγιναν δύο τροποποιήσεις. Αρχικά τροποποιήθηκε η συνάρτηση 3.4 από εκθετική σε γραμμική (Συνάρτηση 3.5), όπως φαίνεται και στην Εικόνα 3.8.

Η συνάρτηση 3.5 υπολογίζει τη μέση ταχύτητα ανά ώρα ενός δάκου με βάση το ποσοστό καρποφορίας της τρέχουσας θέσης του στο πεδίο.

$$AvgSpeed(f_{x,y}) = \frac{-4.513 * f_{x,y} + 451.8}{wh} \quad (3.5)$$

όπου $f_{x,y}$ είναι το ποσοστό καρποφορίας στις συντεταγμένες (x, y) του πεδίου και wh το σύνολο των ωρών της ημέρας στην τρέχουσα εβδομάδα.

Αντί του τροποποιημένου μοντέλου τυχαίου περιπατητή που χρησιμοποιήθηκε στο Μοντέλο 2, χρησιμοποιήθηκε μια εναλλακτική μέθοδος για τη διακίνηση των δάκων μέσα στο πεδίο. Συγκεκριμένα, κάθε ενήλικας δάκος επιλέγει μια τυχαία κατεύθυνση προς την οποία θα κινηθεί. Σε κάθε βήμα προσομοίωσης κάθε δάκος επιλέγει μία από τις ακόλουθες τρεις επιλογές. Είτε παραμένει ακίνητος ή κινείται προς την ίδια κατεύθυνση με το προηγούμενο βήμα, ή κινείται προς την αντίθετη κατεύθυνση. Η Εικόνα 3.9 παρουσιάζει το διάγραμμα της πιθανότητας που έχει κάθε επιλογή σε συνάρτηση του ποσοστού καρποφορίας. Σε συνθήκες μη

ύπαρξης καρπού ο δάκος έχει μεγάλη πιθανότητα να επιλέξει να συνεχίσει στην ίδια κατεύθυνση που ακολουθούσε ενώ όσο το ποσοστό καρποφορίας αυξάνεται και τείνει προς το 100% οι τρεις επιλογές γίνονται ισοπίθανες.

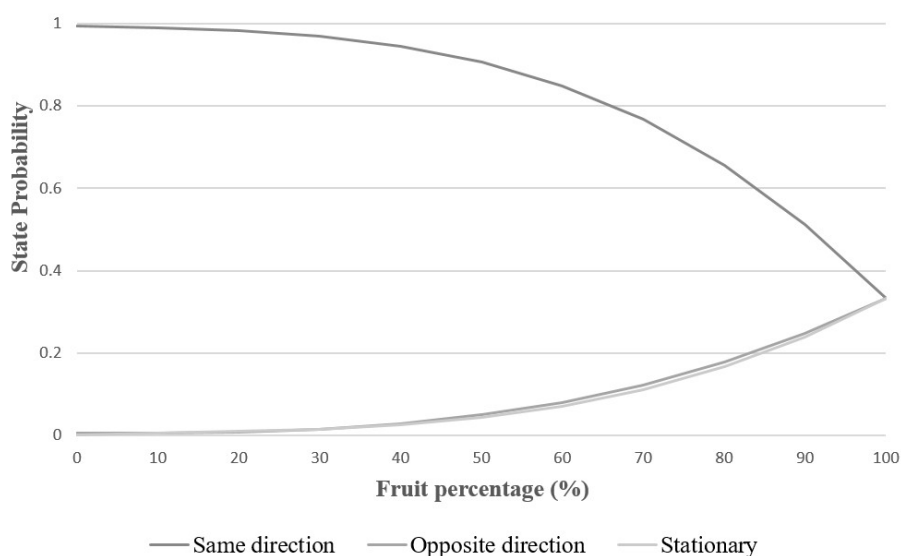
Η τελική ταχύτητα υπολογίζεται από την συνάρτηση 3.6 με βάση τη θερμοκρασία που υπάρχει στην τρέχουσα θέση.

$$FinalSpeed(f_{x,y}, t_i) = AvgSpeed(f_{x,y}) * (1 - \frac{1}{1 + e^{-10*(t_i - MT_U)}}) \quad (3.6)$$

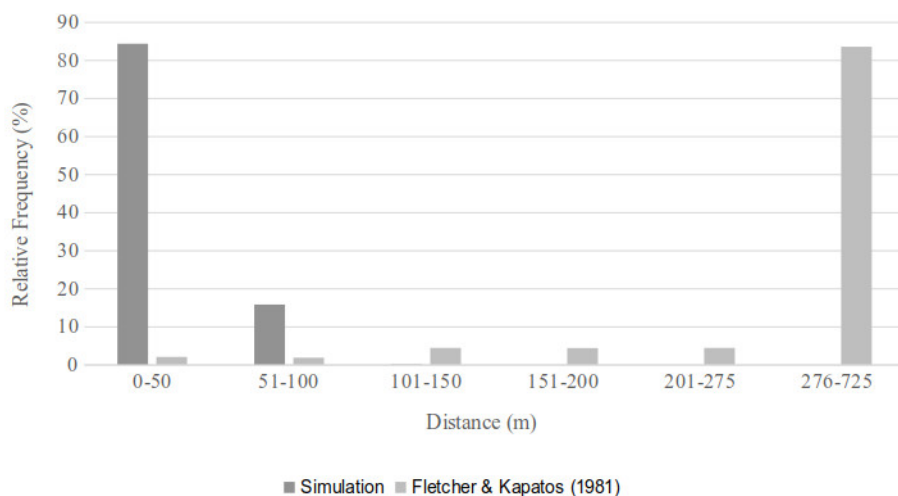
Όπου t_i είναι η θερμοκρασία που υπάρχει στο i -οστό βήμα προσομοίωσης, MT_U είναι το ανώτερο όριο κίνησης, το οποίο έχει οριστεί σε 35 °C. Πέρα από αυτή τη θερμοκρασία, οι δάκοι παραμένουν ακίνητοι [Avidon, 1954].

3.2.2 Σενάρια προσομοίωσης

Εξετάστηκαν τρία σενάρια προσομοίωσης. Σε όλα τα σενάρια προσομοίωσης, 25000 δάκοι τοποθετήθηκαν στο κέντρο ενός ελαιώνα έκτασης 1500m x 1500m. Η χρονική περίοδος προσομοίωσης ορίστηκε σε μία εβδομάδα που αντιστοιχεί σε 168 (ώρες) χρονικά βήματα προσομοίωσης, για τα δύο πρώτα σενάρια. Το τρίτο σενάριο η χρονική περίοδος ορίστηκε σε τρεις μέρες ή 72 βήματα προσομοίωσης. Ο ημερήσιος κύκλος ημέρας/νύχτας ήταν 14 και 10 ώρες αντίστοιχα. Έτσι, η διασπορά ήταν δυνατή μόνο σε 98 (14 ώρες ημέρα x 7 ημέρες) χρονικά βήματα για τα δύο πρώτα σενάρια προσομοίωσης και 42 (14 ώρες ημέρα x 3 ημέρες) για το τελευταίο σενάριο. Δεδομένου ότι, η διασπορά σε σχέση με την παρουσία καρπού ελιάς ήταν το κύριο επίκεντρο των προσομοιώσεων, η θερμοκρασία θεωρήθηκε ότι είναι σταθερή καθόλη την διάρκεια, δηλαδή 27 °C, σε όλα τα στάδια προσομοίωσης και κάτω από τη θερμοκρασία



Εικόνα 3.9: Πιθανότητα επιλογής κατεύθυνσης σε σχέση με το ποσοστό καρποφορίας.



Εικόνα 3.10: Σχετική συχνότητα δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Fletcher and Karpatos [1981] για 0% παρουσία φρούτων ελιάς για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 1.

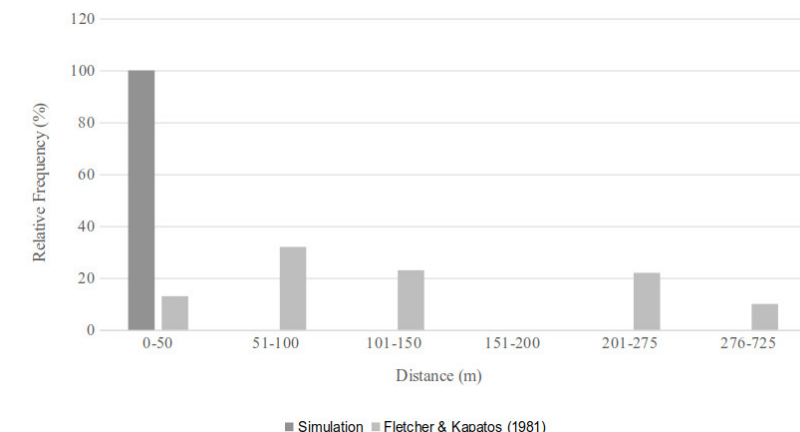
όριο κίνησης. Η μεταβλητή που άλλαξε σε κάθε σενάριο ήταν το ποσοστό των φρούτων ελιάς, το οποίο ορίστηκε σε 0%, 30% και 40% αντίστοιχα. Κάθε σενάριο προσομοίωσης εκτελέστηκε 100 φορές. Η απόδοση των τριών μοντέλων διασποράς συγκρίθηκε, με βάση τα δύο πρώτα σενάρια προσομοίωσης, σε σχέση με τα αποτελέσματα των Fletcher and Karpatos [1981]. Το μοντέλο με την καλύτερη απόδοση χρησιμοποιήθηκε για την εκτέλεση του τρίτου σεναρίου προσομοίωσης και τα αποτελέσματα που προέκυψαν συγκρίθηκαν με τα ευρήματα του πειράματος που εκτέλεσαν στο πεδίο οι Rempoulakis and Nestel [2012], στο οποίο, αν και η διασπορά του δάκου σε σχέση με την καρποφορία του ελαιώνα δεν ήταν το κύριο επίκεντρο της έρευνας τους, προκύπτει ότι για 40% καρποφορία, οι δάκοι ταξίδεψαν κατά μέσο όρο 50 μέτρα τρεις ημέρες μετά την απελευθέρωσή τους στον ελαιώνα.

3.2.3 Αποτελέσματα

Σε αυτήν την υπο-ενότητα παρουσιάζονται τα αποτελέσματα των δύο πρώτων σεναρίων προσομοίωσης για τα μοντέλα διασποράς 1 έως 3.

3.2.3.1 Μοντέλο 1

Η Εικόνα 3.10 εμφανίζει τη σχετική συχνότητα, σε ποσοστό, των δάκων σε καθορισμένα διαστήματα απόστασης μετά το τέλος του πρώτου σεναρίου προσομοίωσης. Τα προσομοιωμένα δεδομένα συγκρίνονται με τα παρατηρούμενα δεδομένα από τους Fletcher and Karpatos [1981] για 0% καρποφορία. Όπως φαίνεται από το διάγραμμα πάνω από το 80% των δάκων έχει διανύσει το πολύ 50 μέτρα από το σημείο εκκίνησης, με το υπόλοιπο 15% να έχει διανύσει 51 έως 100 μέτρα. Για τα υπολοιπούμενα διαστήματα απόστασης δεν υπάρχει παρουσία δάκου. Σε σύγκριση με τα παρατηρούμενα δεδομένα των Fletcher and Karpatos [1981], φαίνεται ότι το



Εικόνα 3.11: Σχετική συχνότητα δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Fletcher and Karpatos [1981] για 30% παρουσία φρούτων ελιάς για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 1.

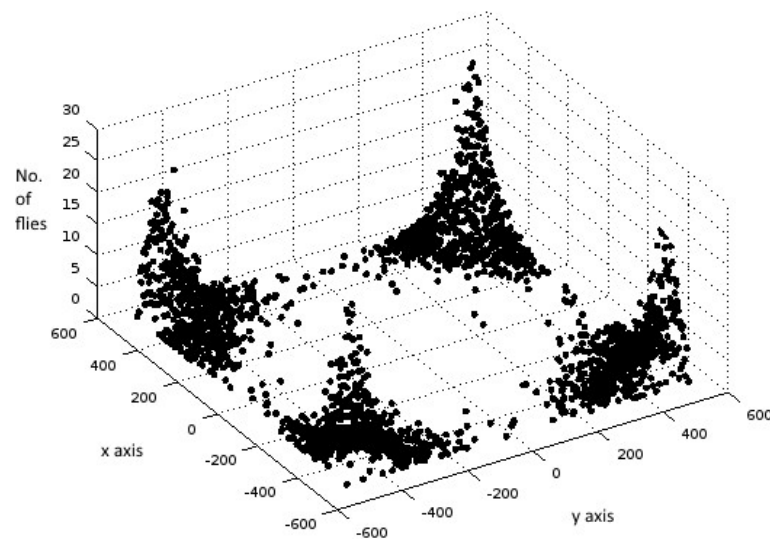
μοντέλο 1, όπως ήταν αναμενόμενο, έχει μεγάλες αποκλίσεις.

Η Εικόνα 3.11 εμφανίζει τη σχετική συχνότητα, σε ποσοστό, των δάκων σε καθορισμένα διαστήματα απόστασης μετά το τέλος του δεύτερου σεναρίου προσομοίωσης. Τα προσομοιωμένα δεδομένα συγκρίνονται με τα παρατηρούμενα δεδομένα από τους Fletcher and Karpatos [1981] για 30% καρποφορία. Όπως φαίνεται από το διάγραμμα όλος ο πληθυσμός των δάκων διένυσε το πολύ 50 μέτρα από το σημείο εκκίνησης, ενώ για τα υπολοιπούμενα διαστήματα απόστασης δεν υπάρχει παρουσία δάκου. Σε σύγκριση με τα παρατηρούμενα δεδομένα των Fletcher and Karpatos [1981], φαίνεται ότι το μοντέλο 1, έχει τεράστια απόκλιση.

3.2.3.2 Μοντέλο 2

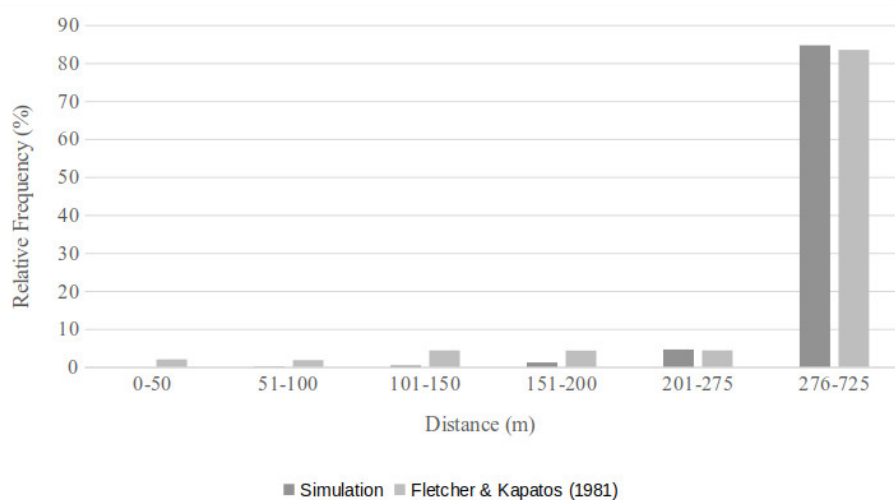
Η Εικόνα 3.12 εμφανίζει την κατανομή της τελικής θέσης των δάκων σε σχέση με την αρχική τους θέση, για το πρώτο σενάριο προσομοίωσης. Οι δάκοι κοντά στη θέση εκκίνησης ήταν πολύ λίγοι και για τους δύο άξονες, ενώ όσο απομακρύνομαστέ από τη θέση εκκίνησης σε κάθε κατεύθυνση η συχνότητα των δάκων αρχίζει να παρουσιάζει αύξηση.

Η Εικόνα 3.13 εμφανίζει τη σχετική συχνότητα, σε ποσοστό, των δάκων σε καθορισμένα διαστήματα απόστασης μετά το τέλος του πρώτου σεναρίου προσομοίωσης. Τα προσομοιωμένα δεδομένα συγκρίνονται με τα παρατηρούμενα δεδομένα από τους Fletcher and Karpatos [1981] για 0% καρποφορία. Όταν δεν υπάρχει καρπός ελιάς στον ελαιώνα, πάνω από το 80% των δάκων έχουν διανύσει απόσταση τουλάχιστον μεγαλύτερη των 275 μέτρων από τη θέση εκκίνησης ενώ για αποστάσεις μεταξύ 150 μέτρων και 275 μέτρων η συχνότητα εμφάνισης δάκων δεν ξεπερνά το 5%. Τέλος για αποστάσεις μικρότερων των 150 μέτρων από το σημείο εκκίνησης η συχνότητα εμφάνισης δάκων είναι σχεδόν μηδενικές. Σε σύγκριση με τα παρατηρούμενα δεδομένα των Fletcher and Karpatos [1981], φαίνεται ότι το μοντέλο 2 αναπαρήγαγε τα παρατηρούμενα δεδομένα αρκετά καλά για τα διαστήματα αποστάσεων 201-275 και 276-725. Για τα υπολοιπούμενα διαστήματα οι διαφορές είναι αρκετά μεγάλες.

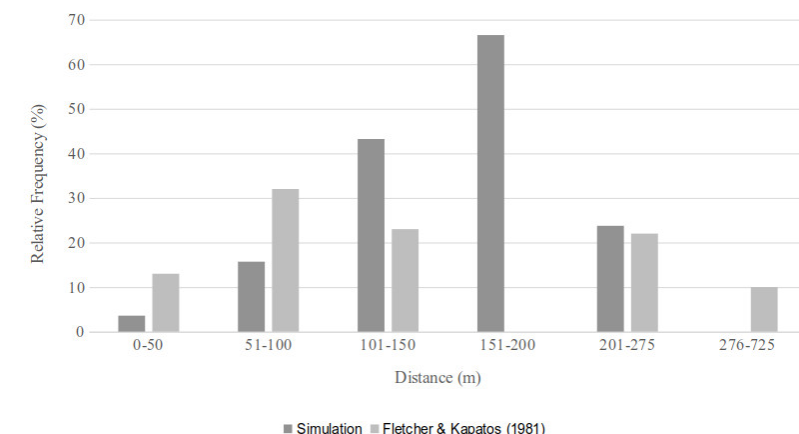


Εικόνα 3.12: Χωρική κατανομή των δάκων για 0% καρποφορίας στον ελαιώνα για χρονικό διάστημα μιας εβδομάδας, για το μοντέλο διασποράς 2.

Η Εικόνα 3.14 εμφανίζει τη σχετική συχνότητα, σε ποσοστό, των δάκων σε καθορισμένα διαστήματα απόστασης μετά το τέλος του δεύτερου σεναρίου προσομοίωσης. Τα προσομοιωμένα δεδομένα συγκρίνονται, και πάλι, με τα παρατηρούμενα δεδομένα από τους Fletcher and Karatos [1981] για 30% καρποφορία. Από το σημείο εκκίνησης μέχρι 200 μέτρα απόσταση από αυτήν η συχνότητα εμφάνισης δάκων αυξάνεται και φτάνει το 66%. Για τα δύο τελευταία διαστήματα απόστασης η σχετική συχνότητα εμφάνισης των δάκων μειώνεται μέχρι το 0%.



Εικόνα 3.13: Σχετική συχνότητα δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Fletcher and Karatos [1981] για 0% παρουσία φρούτων ελιάς για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 2.

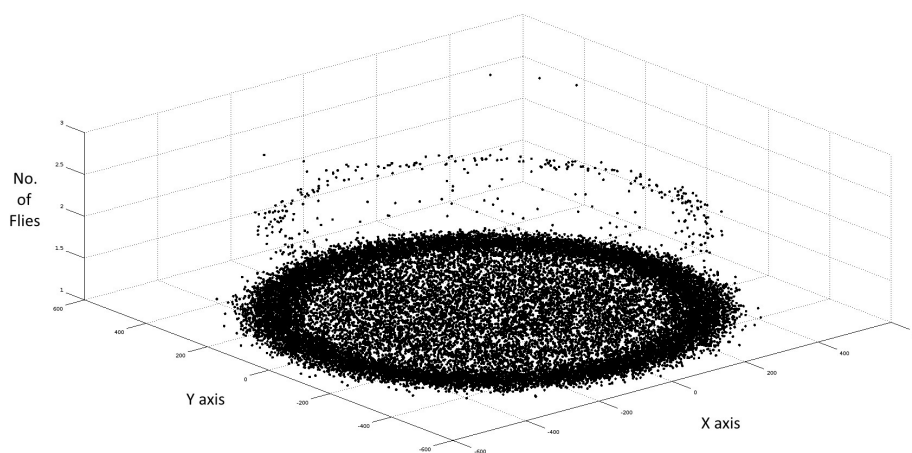


Εικόνα 3.14: Σχετική συχνότητα δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Fletcher and Karpatos [1981] για 30% παρουσία φρούτων ελιάς για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 2.

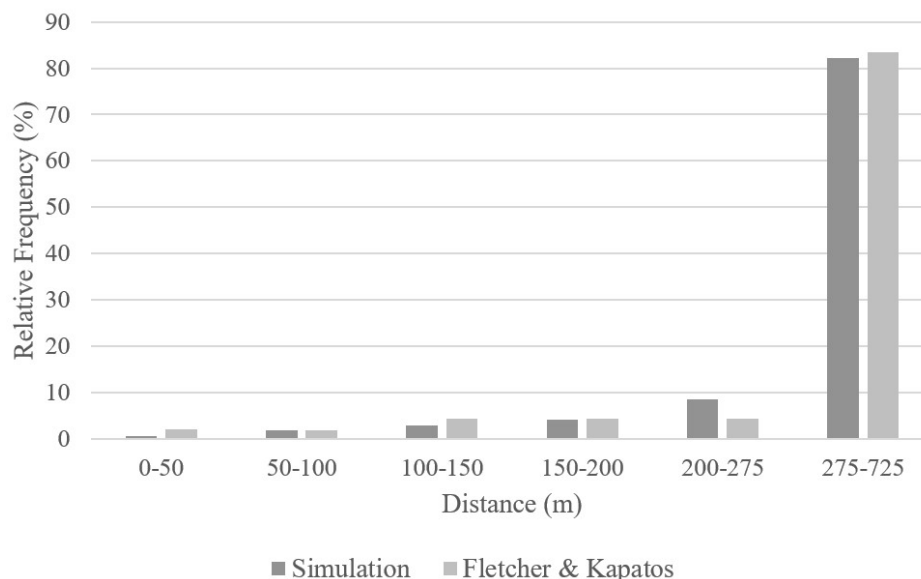
Τα προσομοιωμένα δεδομένα φαίνεται να μην αναπα-ράγουν τη συμπεριφορά των παρατηρούμενων δεδομένων, με μεγάλες διακυμάνσεις μεταξύ των δύο συνόλων δεδομένων σε κάθε διάστημα απόστασης, εκτός από το διάστημα 201-275 όπου οι σχετικές συχνότητες έχουν μικρή απόκλιση.

3.2.3.3 Μοντέλο 3

Η Εικόνα 3.15 απεικονίζει τη χωρική κατανομή των μυγών φρούτων ελιάς για το πρώτο σενάριο προσομοίωσης, όπου το ποσοστό των καρπών της ελιάς στον ελαιώνα ήταν 0%, τα οποία έχουν διασκορπιστεί σε κυκλικό σχήμα. Η κατανομή των δάκων στο εξωτερικό χείλος είναι πυκνότερη από το εσωτερικό χείλος και τον πυρήνα του κυκλικού σχήματος.



Εικόνα 3.15: Χωρική κατανομή των δάκων για 0% καρποφορίας στον ελαιώνα για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 3.



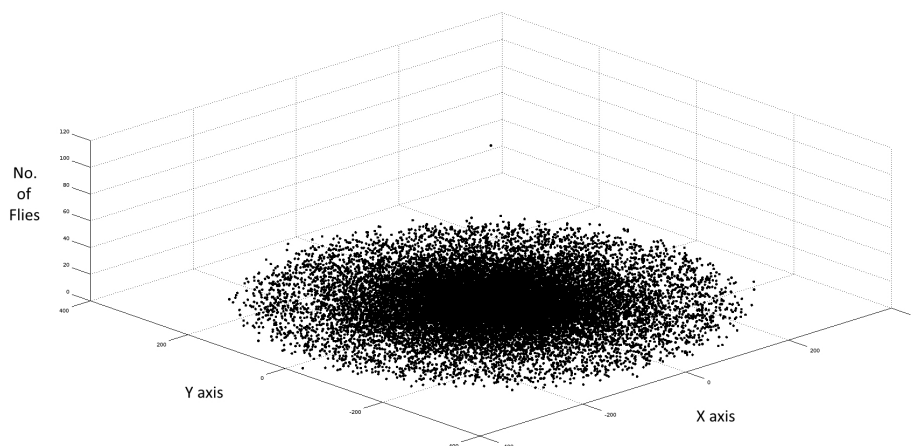
Εικόνα 3.16: Σχετική συχνότητα δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Fletcher and Kapatos [1981] για 0% παρουσία φρούτων ελιάς για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 3.

Η Εικόνα 3.16 εμφανίζει τη σχετική συχνότητα, σε ποσοστό, των δάκων σε καθορισμένα διαστήματα απόστασης μετά το τέλος του πρώτου σεναρίου προσομοίωσης. Τα προσομοιωμένα δεδομένα συγκρίνονται με τα παρατηρούμενα δεδομένα από τους Fletcher and Kapatos [1981] για 0% καρποφορία. Όταν δεν υπάρχει καρπός ελιάς στον ελαιώνα, πάνω από το 80% των δάκων έχουν διανύσει απόσταση τουλάχιστον 275 μέτρων από τη θέση εκκίνησης ενώ για αποστάσεις μικρότερες από 275 μέτρα, παρατηρούνται μικρότερα ποσοστά δάκων.

Σε σύγκριση με τα παρατηρούμενα δεδομένα των Fletcher and Kapatos [1981], φαίνεται ότι το μοντέλο 3 αναπαρήγαγε τα παρατηρούμενα δεδομένα αρκετά καλά. Πράγματι, η παρουσία δάκων γίνεται συχνότερη όσο η απόσταση από τη θέση εκκίνησης αυξάνεται, ενώ η ίδια συμπεριφορά μπορεί να παρατηρηθεί και στα μετρήσιμα δεδομένα.

Η Εικόνα 3.17 απεικονίζει τη χωρική κατανομή των δάκων για το δεύτερο σενάριο προσομοίωσης, όπου το ποσοστό καρποφορίας στον ελαιώνα ήταν 30%. Η χωρική κατανομή των δάκων για την περίπτωση του 30% καρποφορίας είναι το αντίθετο της χωρικής κατανομής για την περίπτωση 0% καρποφορίας. Συγκεκριμένα, η κατανομή είναι πυκνότερη στον πυρήνα και το εσωτερικό χείλος του κυκλικού σχήματος απ' ότι στο εξωτερικό χείλος.

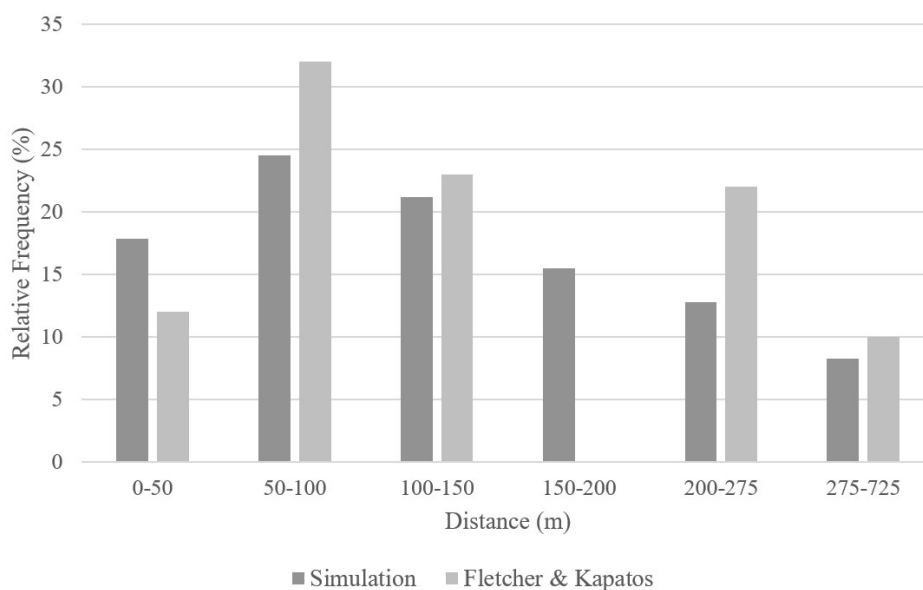
Η Εικόνα 3.18 εμφανίζει τη σχετική συχνότητα, σε ποσοστό, εμφανίζει τη σχετική συχνότητα, σε ποσοστό, των δάκων σε καθορισμένα διαστήματα απόστασης μετά το τέλος του δεύτερου σεναρίου προσομοίωσης. Τα προσομοιωμένα δεδομένα συγκρίνονται, και πάλι, με τα παρατηρούμενα δεδομένα από τους Fletcher and Kapatos [1981] για 30% καρποφορία. Τα προσομοιωμένα δεδομένα αναπαράγουν τη συμπεριφορά των παρατηρούμενων δεδομένων, αν και υπάρχουν μεγάλες διακυμάνσεις μεταξύ των δύο συνόλων δεδομένων σε κάθε διάστημα απόστασης.



Εικόνα 3.17: Χωρική κατανομή των δάκων για 30% καρποφορίας στον ελαιώνα για χρονικό διάστημα μιας εβδομάδας, για το μοντέλο διασποράς 3.

3.2.4 Αξιολόγηση αποτελεσμάτων

Για την αξιολόγηση των μοντέλων διασποράς, πραγματοποιήθηκε σύγκριση μεταξύ των προσομοιωσμένων δεδομένων με τα μετρήσιμα δεδομένα από τους Fletcher and Karpatos [1981], συγκεκριμένα μια σύγκριση μεταξύ των σχετικών συχνοτήτων σε σχέση με καθορισμένα διαστήματα απόστασης. Οι ακόλουθοι δείκτες, οι οποίοι χρησιμοποιήθηκαν επίσης από τους Yang et al. [2012], χρησιμοποιήθηκαν για την εκτίμηση της εγγύτητας των συνόλων δεδομένων: συντελεστής συσχέτισης (R), Bias [Kobayashi and Salam, 2000] (Εξίσωση 3.7) και root mean



Εικόνα 3.18: Σχετική συχνότητα δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Fletcher and Karpatos [1981] για 30% παρουσία φρούτων ελιάς για χρονικό διάστημα μίας εβδομάδας, για το μοντέλο διασποράς 3.

square error (RMSE) (Εξίσωση 3.8). Όταν οι τιμές των δύο τελευταίων δεικτών είναι κοντά στο μηδέν, αυτό υποδεικνύει ότι τα σύνολα δεδομένων βρίσκονται κοντά το ένα στο άλλο. Όλοι οι προαναφερθέντες δείκτες υπολογίστηκαν μέσω ενός script στη γλώσσα προγραμματισμού Python.

$$Bias = \frac{1}{N} * \sum_{i=1}^{N} OBS_i - SIM_i \quad (3.7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (OBS_i - SIM_i)^2}{N}} \quad (3.8)$$

OBS_i και SIM_i είναι η μετρούμενη και προσομοιωμένη τιμή, αντίστοιχα, του i-οστού σημείου δεδομένων από N παρατηρήσεις.

Η αξιολόγηση των μοντέλων διασποράς έγινε για τις περιπτώσεις 0% και 30% καρποφορίας, καθώς μόνο για οι τις δύο αυτές περιπτώσεις υπήρχαν διαθέσιμα μετρήσιμα δεδομένα.

Μοντέλο 1 Ο πίνακας 3.1 παρουσιάζει τις τιμές των δεικτών αξιολόγησης για το μοντέλο διασποράς 1. Συγκεκριμένα, για 0% καρποφορία, η σύγκριση έδωσε τα ακόλουθα αποτελέσματα: Οι τιμές R, Bias και RMSE ήταν -0,26737, 32,08 και 48,278151, αντίστοιχα. Είναι σαφές ότι η εκτιμώμενη σχετική συχνότητα έχει αρνητική συσχέτιση με τις μετρήσιμες τιμές και μεγάλη απόκλιση βάση των δεικτών Bias και RMSE. Ομοίως, για της περίπτωση 30% καρποφορίας υπολογίστηκαν τα ακόλουθα αποτελέσματα: -0,15885, 29 και 40,220227 αντίστοιχα, γεγονός που δείχνει ότι οι εκτιμώμενες σχετικές συχνότητες δεν έχουν συσχέτιση με τις μετρήσιμες τιμές.

Table 3.1: Δείκτες αξιολόγησης για μοντέλο διασποράς 1.

% Καρποφορίας	Δείκτες αξιολόγησης		
	R	Bias	RMSE
0	-0,26737	32,08	48,278151
30	-0,15885	29	40,220227

Μοντέλο 2 Ο πίνακας 3.2 παρουσιάζει τις τιμές των δεικτών αξιολόγησης για το μοντέλο διασποράς 2. Συγκεκριμένα, για 0% καρποφορία, η σύγκριση έδωσε τα ακόλουθα αποτελέσματα: Οι τιμές R, Bias και RMSE ήταν 0,9991, 2,018167 και 2,357282, αντίστοιχα. Είναι σαφές ότι η εκτιμώμενη σχετική συχνότητα συσχετίστηκε καλά με τις μετρήσιμες τιμές. Ωστόσο, για της περίπτωση 30% καρποφορίας υπολογίστηκαν τα ακόλουθα αποτελέσματα: -0,33-473, 20,698367 και 29,695290 αντίστοιχα, γεγονός που δείχνει ότι οι εκτιμώμενες σχετικές συχνότητες έχουν αρνητική συσχέτιση με τις μετρήσιμες τιμές και μεγάλη απόκλιση, βάση των δεικτών Bias και RMSE.

Table 3.2: Δείκτες αξιολόγησης για μοντέλο διασποράς 2.

% Καρποφορίας	Δείκτες αξιολόγησης		
	R	Bias	RMSE
0	0,9991	2,018167	2,357282
30	-0,33473	20,698367	29,695290

Μοντέλο 3 Ο πίνακας 3.3 παρουσιάζει τις τιμές των δεικτών αξιολόγησης για το μοντέλο διασποράς 3. Συγκεκριμένα, για 0% καρποφορία, η σύγκριση έδωσε τα ακόλουθα αποτελέσματα: Οι τιμές R, Bias και RMSE ήταν 0,9980, 1,38167 και 1,931, αντίστοιχα. Είναι σαφές ότι η εκτιμώμενη σχετική συχνότητα συσχετίστηκε καλά με τις μετρήσιμες τιμές. Ωστόσο, για της περίπτωση 30% καρποφορίας υπολογίστηκαν τα ακόλουθα αποτελέσματα: 0,595, 6,771667 και 8,272802 αντίστοιχα, γεγονός που δείχνει ότι οι εκτιμώμενες σχετικές συχνότητες έχουν συσχέτιση με τις μετρήσιμες τιμές.

Table 3.3: Δείκτες αξιολόγησης για μοντέλο διασποράς 3.

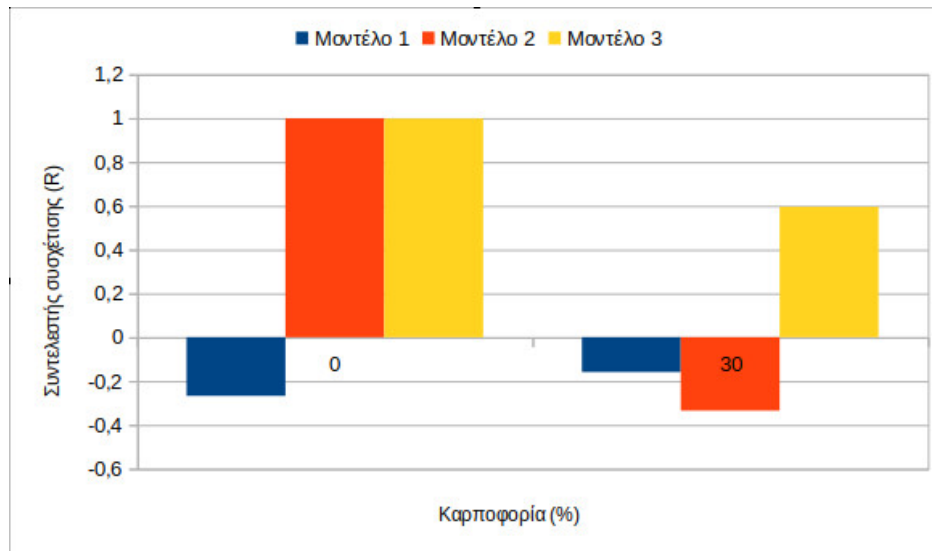
% Καρποφορίας	Δείκτες αξιολόγησης		
	R	Bias	RMSE
0	0,9980	1,38167	1,931
30	0,595	6,771667	8,272802

Η Εικόνα 3.19 παρουσιάζει την σύγκριση των τιμών των τριών μοντέλων διασποράς για τον δείκτη αξιολόγησης R. Από το γράφημα προκύπτει ότι τα μοντέλα 2 και 3 έχουν ισχυρή συσχέτιση με τα παρατηρούμενα δεδομένα για την περίπτωση της 0% καρποφορίας. Το μοντέλο 1 παρουσιάζει αρνητική συσχέτιση και στις δύο εξεταζόμενες περιπτώσεις. Για την περίπτωση 30% καρποφορίας το μοντέλο 3 έχει μια μέση συσχέτιση με τα παρατηρούμενα δεδομένα ενώ το μοντέλο 2 σε αυτήν την περίπτωση παρουσιάζει αρνητική συσχέτιση.

Η Εικόνα 3.20 παρουσιάζει την σύγκριση των τιμών των τριών μοντέλων διασποράς για τον δείκτη αξιολόγησης Bias. Από το γράφημα προκύπτει ότι τα μοντέλα 2 και 3 είναι αρκετά κοντά στα μετρήσιμα δεδομένα για την περίπτωση της 0% καρποφορίας με τιμές μικρότερες του 3. Για την περίπτωση 30% καρποφορίας το μοντέλο 3 έχει την μικρότερη τιμή σε σχέση με τα άλλα δύο μοντέλα. Το μοντέλο 1 και στις δύο εξεταζόμενες περιπτώσεις έχει τις μεγαλύτερες τιμές.

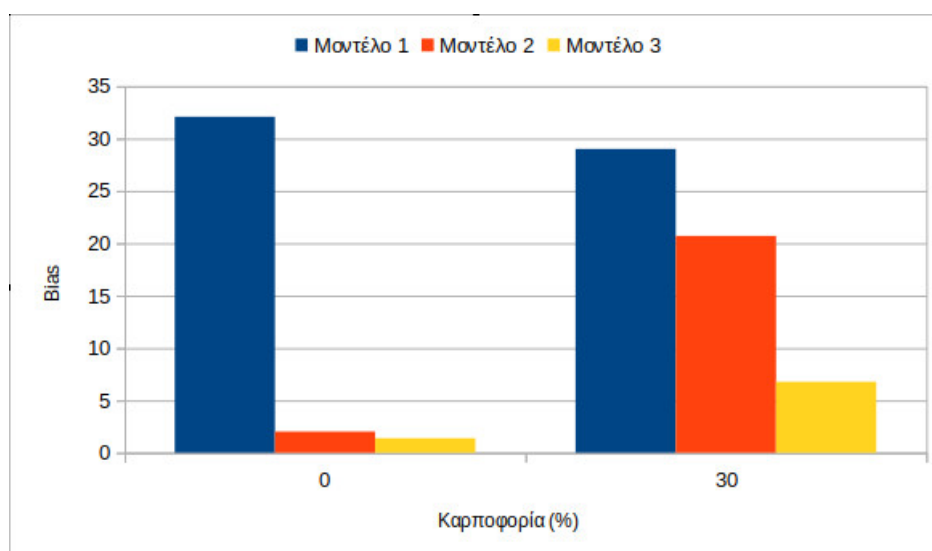
Η Εικόνα 3.21 παρουσιάζει την σύγκριση των τιμών των τριών μοντέλων διασποράς για τον δείκτη αξιολόγησης RMSE. Η τιμές για τον συγκεκριμένο δείκτη έχουν την ίδια συμπεριφορά όπως και για τον δείκτη Bias (Εικόνα 3.20). Ομοίως, τα μοντέλα 2 και 3 είναι αρκετά κοντά στα μετρήσιμα δεδομένα για την περίπτωση της 0% καρποφορίας με τιμές μικρότερες του 3. Για την περίπτωση 30% καρποφορίας το μοντέλο 3 έχει την μικρότερη τιμή σε σχέση με τα άλλα δύο μοντέλα, ενώ το μοντέλο 1 και στις δύο εξεταζόμενες περιπτώσεις έχει τις μεγαλύτερες τιμές.

Με βάση τα παραπάνω γραφήματα (Εικόνες 3.19 - 3.21) το μοντέλο διασποράς 3 απέδωσε καλύτερα στην αναπαραγωγή των παρατηρούμενων δεδομένων.

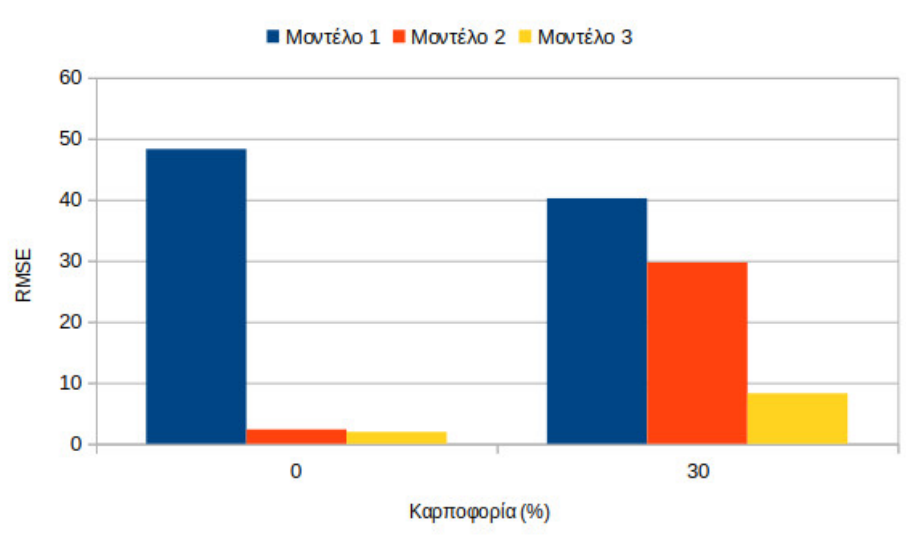


Εικόνα 3.19: Σύγκριση δείκτη αξιολόγησης R των τριών μοντέλων διασποράς για τις τρεις εξεταζόμενες περιπτώσεις καρποφορίας.

Τέλος, το μοντέλο 3 ως καλύτερο μοντέλο διασποράς σε σύγκριση με το μοντέλο 1 και 2 χρησιμοποιήθηκε για την εκτέλεση του τρίτου σεναρίου προσομοίωσης. Η σύγκριση της μέσης απόστασης πτήσης του δάκου από το τρίτο σενάριο προσομοίωσης σε σύγκριση με αυτό που μετρήθηκε από τους Rempoulakis and Nestel [2012], παρουσιάζεται στην Εικόνα 3.22. Όπως φαίνεται το μοντέλο προσομοίωσης διασποράς, με μέση απόσταση 55,11m, έπεσε εκτός κατά περίπου 5m σε σύγκριση με τις μετρήσεις πεδίου των Rempoulakis and Nestel [2012].



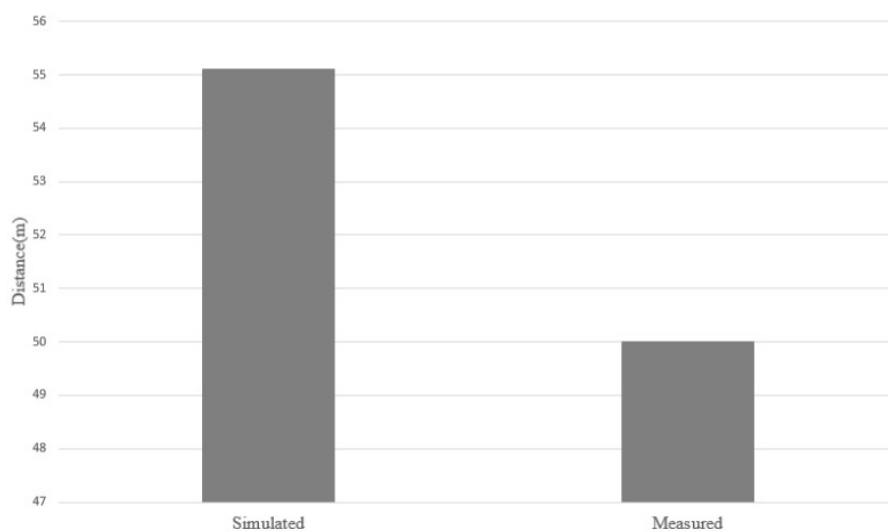
Εικόνα 3.20: Σύγκριση δείκτη αξιολόγησης Bias των τριών μοντέλων διασποράς για τις τρεις εξεταζόμενες περιπτώσεις καρποφορίας.



Εικόνα 3.21: Σύγκριση δείκτη αξιολόγησης RMSE των τριών μοντέλων διασποράς για τις τρεις εξεταζόμενες περιπτώσεις καρποφορίας.

3.2.5 Συμπεράσματα

Τα αποτελέσματα της προσομοίωσης του μοντέλου διασποράς 3 έδειξαν, σε σύγκριση με τα αποτελέσματα των Fletcher and Karatos [1981] ότι για 0% καρποφορία η σχετική συχνότητα των δάκων σε καθορισμένα διαστήματα απόστασης αυξάνεται καθώς κινούμαστε πιο μακριά από το σημείο προέλευσης, το οποίο παρατηρήθηκε και στις μετρήσεις πεδίου. Επιπλέον,



Εικόνα 3.22: Μέση απόσταση πτήσης δάκων σε σύγκριση με τα παρατηρούμενα δεδομένα από τους Rempoulakis and Nestel [2012], για 40% καρποφορίας για χρονικό διάστημα τριών ημερών.

παρατηρήθηκαν μικρές διακυμάνσεις από τα μετρήσιμα δεδομένα. Οι τρεις δείκτες αξιολόγησης αποκάλυψαν ισχυρή εγγύτητα και συσχέτιση μεταξύ των προσομοιωμένων και των παρατηρήσιμων δεδομένων. Για την περίπτωση 30% καρποφορίας η σχετική συχνότητα των δάκων σε καθορισμένα διαστήματα απόστασης έχουν παρόμοια τάση με τις μετρήσεις πεδίου, αν και υπάρχουν μεγάλες διακυμάνσεις μεταξύ τους. Οι δείκτες αξιολόγησης έδειξαν ότι τα προσομοιωμένα με τα μετρήσιμα δεδομένα δεν συσχετίζοντε τόσο ισχυρά όπως και στην περίπτωση της 0% καρποφορίας.

Σε σύγκριση με τα αποτελέσματα των Rempoulakis and Nestel [2012] διαπιστώθηκε ότι υπήρξε απόκλιση 5 μέτρων όσον αφορά τη μέση απόσταση που διανύει ο δάκος μετά από τρεις ημέρες. Ωστόσο, αυτή η διαφορά δεν είναι πολύ μεγάλη, αφού στο χωράφι τα 5 μέτρα θα μπορούσαν να αντιστοιχούν στην απόσταση μεταξύ δύο ελαιόδεντρων.

Εν κατακλείδι το μοντέλο διασποράς 3 αναπαράγει τις μετρήσεις το πεδίου σε ικανοποιητικό βαθμό, ειδικά για την περίπτωση της 0% καρποφορίας.

3.3 Επίδραση μικροκλιμάτων

Στόχος αυτής της ενότητας είναι να διερευνήσει, μέσω προσομοιώσεων, πώς επηρεάζεται η εξέλιξη του πληθυσμού του δάκου από το μικροκλίμα του ελαιώνα στο οποίο δραστηριοποιείται. Καθώς επίσης και την επίδραση που έχουν τα τρέχοντα πρότυπα ψεκασμού στον πληθυσμό του δάκου όταν η ύπαρξη του μικροκλίματος αγνοείται.

3.3.1 Περιοχή μελέτης



Εικόνα 3.23: Χάρτης της περιοχής μελέτης στην Κέρκυρα.

(Δεδομένα χάρτη: Google, TerraMetrics, Data SIO, NOAA, U.S. Navy, NGA, GEBCO, CNES/Airbus)

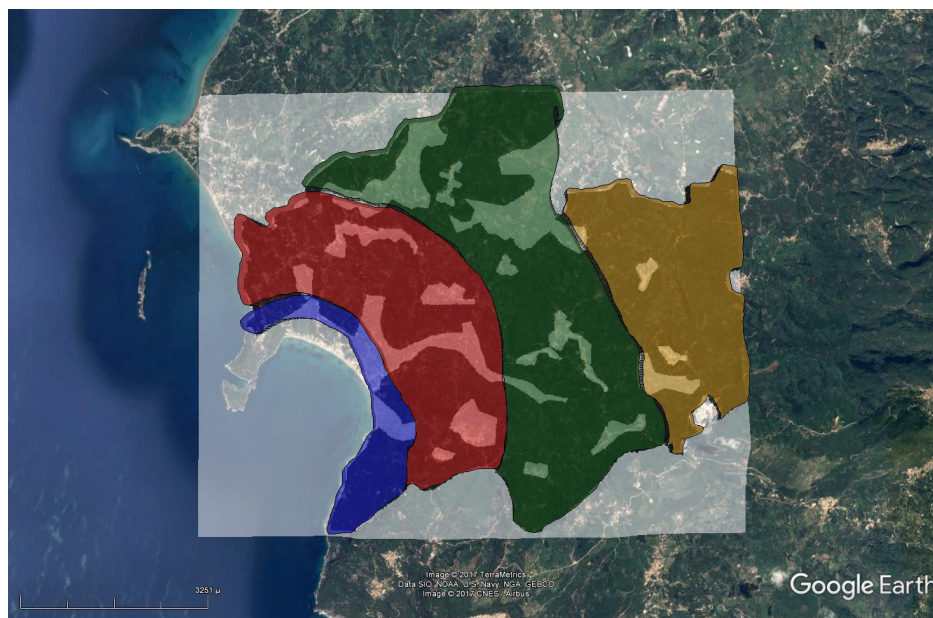
Η περιοχή μελέτης ήταν ο πρώην Δήμος Αγίου Γεωργίου (Εικόνα 3.23) που βρίσκεται στο βορειοδυτικό τμήμα του νησιού της Κέρκυρας. Ο κλιματικός τύπος της περιοχής μελέτης και του νησιού γενικότερα χαρακτηρίζεται από υψηλές θερμοκρασίες και σχεδόν καμία βροχόπτωση κατά τους καλοκαιρινούς μήνες. Η συνολική έκταση είναι 9km x 7,5km.

Αν και η περιοχή της μελέτης μπορεί να αποτελείται από πολλά μικροκλίματα, τα δεδομένα που συλλέξαμε δεν ήταν αρκετά για να διακριθούν τα διαφορετικά μικροκλίματα. Έτσι, για λόγους απλότητας η περιοχή χωρίστηκε αυθαίρετα σε τέσσερα μικροκλίματα όπως απεικονίζονται στην Εικόνα 3.24. Τα προκύπτοντα μικροκλίματα από αριστερά προς τα δεξιά ορίζονται ως μικροκλίμα A, B, C και D.

3.3.2 Δεδομένα

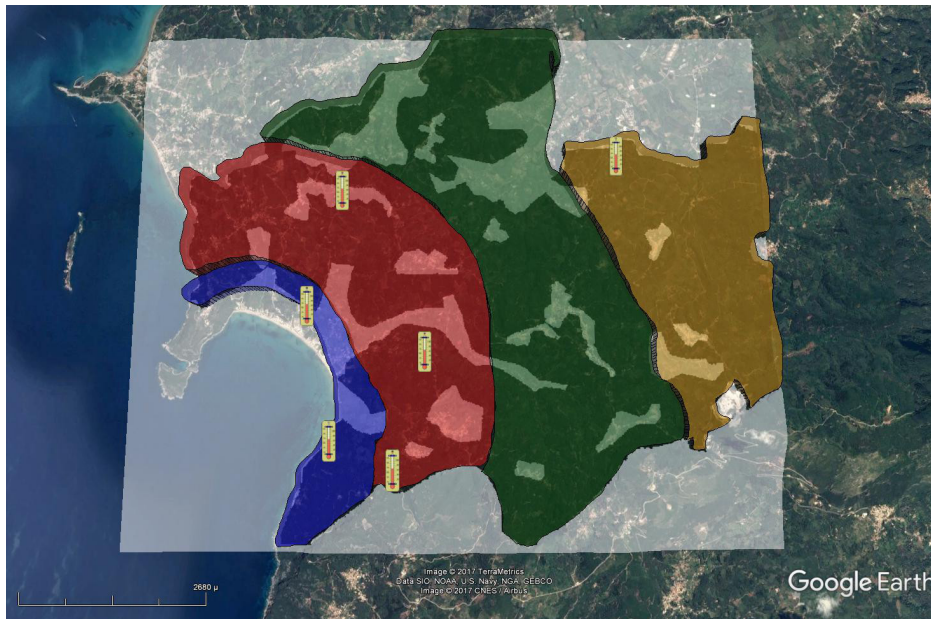
Η Εικόνα 3.25 απεικονίζει τις τοποθεσίες όπου εγκαταστάθηκαν περιβαλλοντικοί αισθητήρες. Σε μικροκλίματα όπου εγκαταστάθηκαν περισσότεροι από ένας αισθητήρες, υπολογίστηκε η μέση τιμή από όλους τους αισθητήρες και χρησιμοποιήθηκε για την αναπαράσταση της θερμοκρασίας για όλες τις περιοχές μέσα στο μικροκλίμα. Για το μικροκλίμα C, όπου δεν εγκαταστάθηκε περιβαλλοντικός αισθητήρας, υπολογίσαμε τη μέση τιμή και από τα δύο γειτονικά μικροκλίμα, και συγκεκριμένα τα μικροκλίματα B και D.

Οι εγκατεστημένοι περιβαλλοντικοί αισθητήρες κατέγραψαν θερμοκρασίες από τις 19 Ιουνίου 2015 μέχρι τις 9 Νοεμβρίου 2015 με διάστημα 15 λεπτών. Για την διεξαγωγή των πειραμάτων χρησιμοποιήθηκαν οι καταγεγραμμένες θερμοκρασίες για τη χρονική περίοδο από 21 Ιουνίου



Εικόνα 3.24: Κατακερματισμός περιοχής μελέτης σε μικροκλίματα. Το μπλε χρώμα είναι το μικροκλίμα A, το κόκκινο χρώμα είναι το μικροκλίμα B, το πράσινο χρώμα είναι το μικροκλίμα C και το κίτρινο χρώμα είναι το μικροκλίμα D.

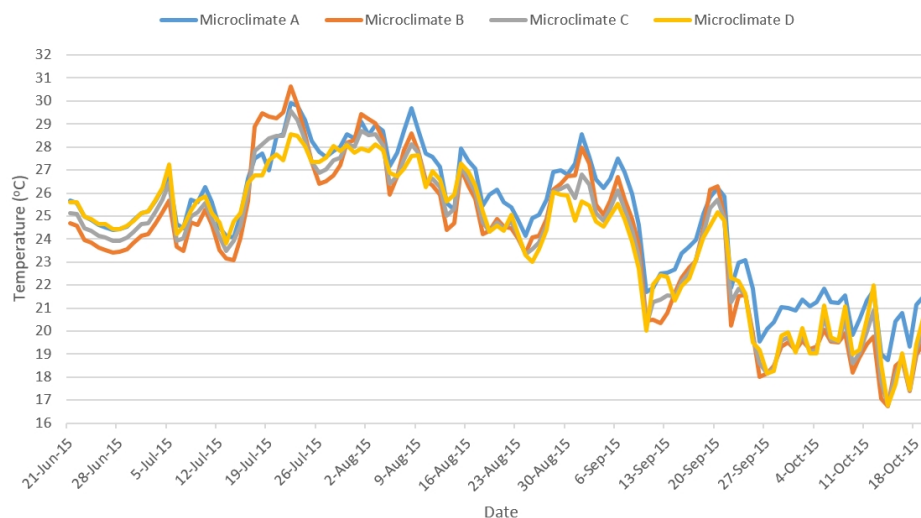
(Δεδομένα χάρτη: Google, TerraMetrics, Data SIO, NOAA, U.S. Navy, NGA, GEBCO, CNES/Airbus)



Εικόνα 3.25: Τοποθεσίες περιβαλλοντικών αισθητήρων στην περιοχή μελέτης (Δεδομένα χάρτη: Google, TerraMetrics, Data SIO, NOAA, U.S. Navy, NGA, GEBCO, CNES/Airbus)

2015 έως 20 Οκτωβρίου 2015. Η Εικόνα 3.26 απεικονίζει τις μέσες ημερήσιες θερμοκρασίες κάθε μικροκλίματος για την επιλεγμένη χρονική περίοδο.

Το μικροκλίμα Α φαίνεται να έχει θερμότερο κλίμα, καθώς και μικροκλίμα Β το οποίο φαίνεται να είναι λίγο πιο δροσερό από το προαναφερθέν μικροκλίμα. Από την άλλη το μικροκλίμα D έχει ψυχρότερο κλίμα από τα υπόλοιπα μικροκλίματα. Ενώ το μικροκλίμα C είναι μεταξύ μικροκλίματος Β και D, δεδομένου ότι οι θερμοκρασίες είναι ο μέσος όρος των προαναφερθέντων μικροκλιμάτων.



Εικόνα 3.26: Μέσες ημερήσιες θερμοκρασίες για κάθε μικροκλίμα.

3.3.3 Σενάρια προσομοίωσης

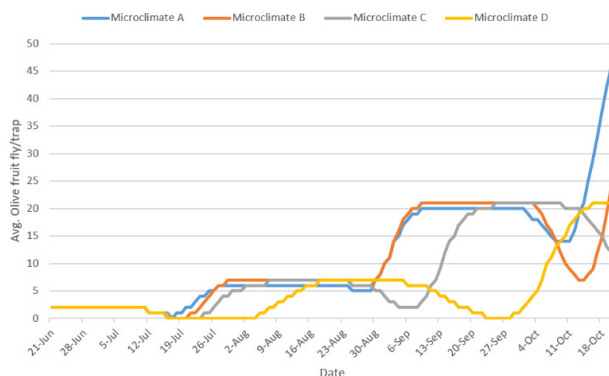
Και για τα δύο παρακάτω σενάρια προσομοίωσης, το προσομοιωμένο πεδίο ήταν ο πρώην δήμος Αγίου Γεωργίου (Εικόνα 3.23) ο οποίος χωρίστηκε σε τέσσερα μικροκλίματα (Εικόνα 3.24). Τα σύνολα θερμοκρασίας για κάθε μικροκλίμα ήταν για το χρονικό διάστημα από 21 Ιουνίου 2015 έως 20 Οκτωβρίου 2015 και το ποσοστό καρποφορίας ορίστηκε σε 75% για όλες τις περιοχές του ελαιώνα εντός του προσομοιωμένου πεδίου. Εκτελέστηκαν τα ακόλουθα δύο σενάρια προσομοίωσης, στο πρώτο σενάριο ο δάκος αναπτύσσεται χωρίς παρεμβολές, δηλαδή ψεκασμό, προκειμένου να παρατηρηθεί η χρονική εξέλιξη του πληθυσμού. Στο δεύτερο σενάριο, πραγματοποιήθηκε ένας ψεκασμός στις 20 Ιουλίου, που είναι γενικά η εποχή κατά την οποία οι τοπικές γεωργικές αρχές ξεκινούν τον πρώτο καθολικό ψεκασμό. Γενικά, οι τοπικές αρχές πραγματοποιούν τρεις ψεκασμούς συνολικά. Ο πρώτος ψεκασμός πραγματοποιείται μεταξύ 10 και 20 Ιουλίου, ο δεύτερος πραγματοποιείται από τις 10 έως τις 15 Αυγούστου και τέλος ο τελευταίος ψεκασμός πραγματοποιείται τις τελευταίες ημέρες του Σεπτεμβρίου έως τις 10 Οκτωβρίου. Ο πρώτος ψεκασμός είναι πάντα καθολικό ψεκασμός, ενώ τα δύο τελευταία είναι τοπικά ψεκασμοί, πράγμα που σημαίνει ότι μόνο συγκεκριμένες περιοχές.

3.3.4 Αποτελέσματα

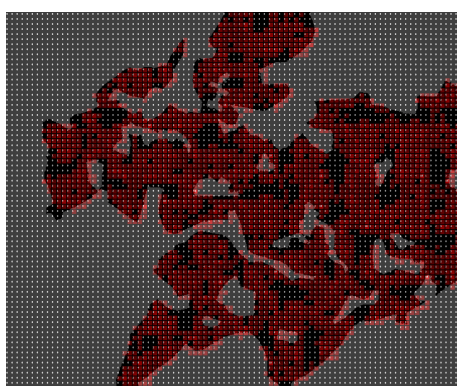
Η Εικόνα 3.27a παρουσιάζει την εξέλιξη του πληθυσμού του δάκου για κάθε μικροκλίμα χωρίς να έχει πραγματοποιηθεί ψεκασμός κατά τη διάρκεια του προσομοιωμένου χρόνου. Είναι προφανές ότι η θερμοκρασία κάθε μικροκλίματος έχει επίδραση στην δυναμική του πληθυσμού του δάκου. Για το μικροκλίμα A και B δύο γενιές είναι ορατές και η 3η γενιά αρχίζει να αναδύεται στο τέλος της προσομοιωμένης περιόδου. Από την άλλη πλευρά, για το μικροκλίμα C και D μόνο δύο γενιές είναι ορατές. Επιπλέον φαίνεται να υπάρχει καθυστέρηση στην ανάπτυξη και την εμφάνιση της πρώτης γενιάς, και κατά συνέπεια για κάθε επόμενη, μεταξύ κάθε μικροκλίματος. Η μικρότερη καθυστέρηση που παρατηρείται είναι μεταξύ των γειτονικών μικροκλίματων A και B που είναι τέσσερις μέρες, ενώ η μεγαλύτερη καθυστέρηση, παρατηρείται μεταξύ των μικροκλίματων A και D όπου η πρώτη γενιά του μικροκλίματος D αναδύεται 19 μέρες μετά την πρώτη γενιά του μικροκλίματος A. Στην Εικόνα 3.27c φαίνεται πως και στα τέσσερα μικροκλίματα στο τέλος της προσομοίωσης όλος ο ελαιώνας που αντιστοιχεί σε κάθε ένα από αυτά υπάρχει έξαρση του πληθυσμού του δάκου.

Η επίδραση ενός μόνο ψεκασμού ένα μήνα μετά την έναρξη της προσομοίωσης, δηλαδή στις 20 Ιουλίου, απεικονίζεται στην Εικόνα 3.28a. Η εξέλιξη του πληθυσμού του δάκου στις περιοχές που αντιστοιχούν στα μικροκλίματα A και B επηρεάζονται σημαντικά. Όμως αυτό δεν ισχύει για την εξέλιξη του πληθυσμού στα μικροκλίματα C και D που μένει ανηπερέαστο από τον ψεκασμό που πραγματοποιήθηκε. Στην Εικόνα 3.28c φαίνεται πως στο τέλος της προσομοίωσης στα μικροκλίματα A και B ο πληθυσμός του δάκου έχει κρατηθεί υπό έλεγχο ενώ υπάρχουν περιοχές (μαύρα κελιά) μέσα στα προαναφερθέντα μικροκλίματα όπου δεν δραστηριοποιείται ο δάκος. Οι περιοχές που αντιστοιχούν στα μικροκλίματα C και D φαίνεται να υπάρχει παρόμοια χωρική κατανομή όπως και στην Εικόνα 3.27c.

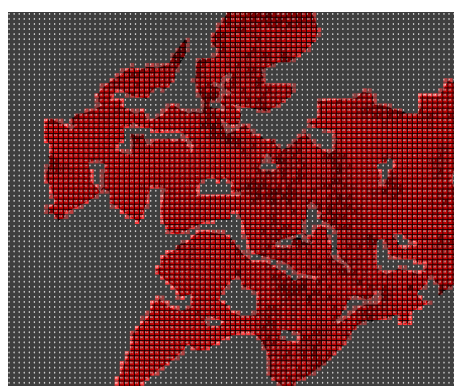
Οι ακόλουθες εικόνες παρουσιάζουν την σύγκριση της εξέλιξης του πληθυσμού του δάκου για τα δύο σενάρια προσομοίωσης για κάθε μικροκλίμα.



(a)



(b)



(c)

Εικόνα 3.27: (a) Εξέλιξη πληθυσμού του δάκου σε κάθε μικροκλίμα χωρίς ψεκασμό και κατανομή πληθυσμού (b) στην έναρξη της προσομοίωσης και (c) στο τέλος της προσομοίωσης

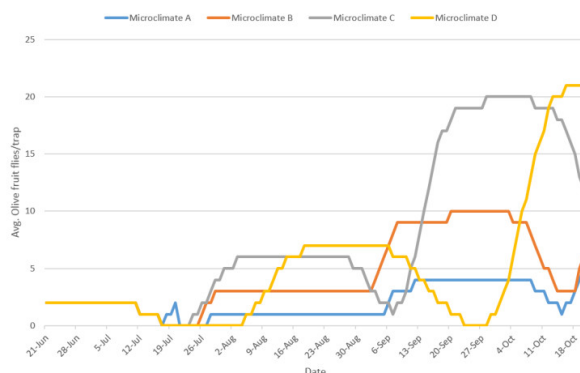
Στην Εικόνα 3.29, παρουσιάζεται το συγκριτικό γράφημα για το μικροκλίμα A. Είναι προφανές ότι ο ψεκασμός είχε καταστροφικές επιπτώσεις για την εξέλιξη του πληθυσμού του δάκου. Είναι σημαντικό να σημειωθεί ότι, το μέγεθος του πληθυσμού μέχρι το τέλος της προσομοίωσης ήταν κάτω από το όριο έξαρσης.

Η σύγκριση για το μικροκλίμα B απεικονίζεται στην Εικόνα 3.30. Ο ψεκασμός που πραγματοποιήθηκε φαίνεται να έχει επηρεάσει σε μεγάλο βαθμό την εξέλιξη του πληθυσμού του δάκου. Αν και, η επίδραση του ψεκασμού δεν είναι τόσο καταστροφική όπως συμβαίνει με το μικροκλίμα A, το μέγεθος του πληθυσμού κάθε επόμενης γενιάς έχει μειωθεί σημαντικά. Το μέγεθος του πληθυσμού για την πρώτη γενιά είναι κάτω από το όριο μόλυνσης και για τη δεύτερη το μέγεθος του πληθυσμού είναι σχεδόν στο μισό σε σύγκριση με τον πληθυσμό του πρώτου σεναρίου προσομοίωσης.

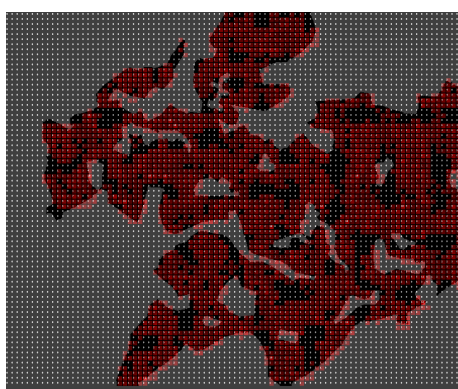
Από την άλλη, το μικροκλίμα C (Εικόνα 3.31) επηρεάζεται ελαφρώς από τον ψεκασμό, δεδομένου ότι, υπάρχει μόνο μια μικρή μείωση στο μέγεθος του πληθυσμού μεταξύ τα δύο σεναρίων προσομοίωσης.

Τέλος, οι περιοχές που ανήκουν στο μικροκλίμα D φαίνεται να είναι ανεπηρέαστες από τον

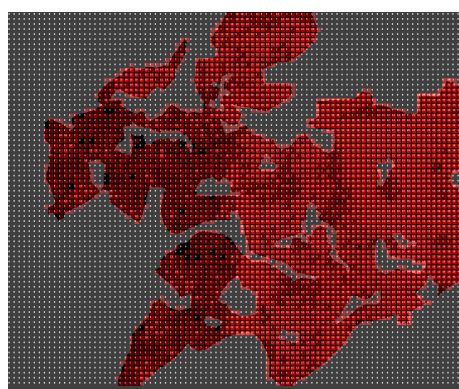
Ανάπτυξη στοχαστικών αλγορίθμων και μοντέλων μηχανικής μάθησης για την εξέλιξη πληθυσμών: Εφαρμογή στο δάκο της Ελιάς



(a)



(b)



(c)

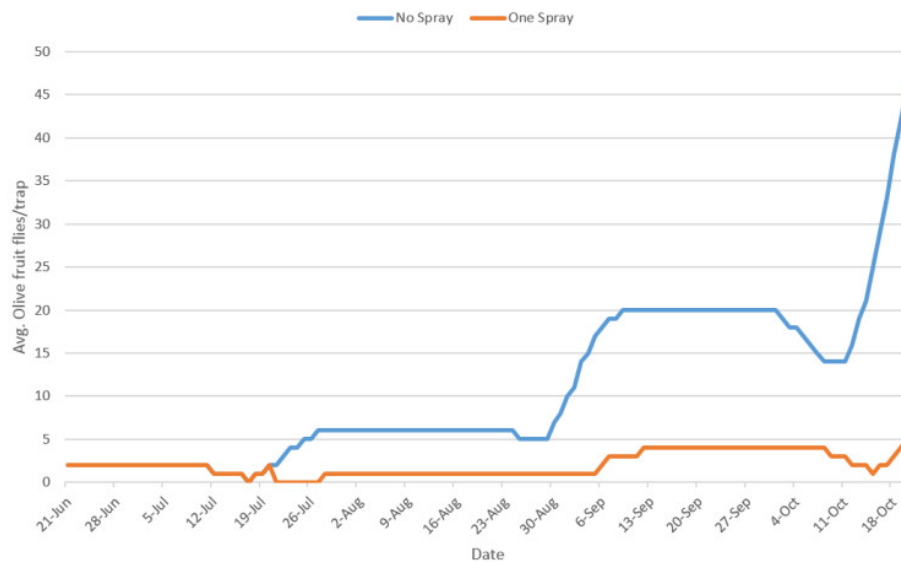
Εικόνα 3.28: (a) Εξέλιξη πληθυσμού του δάκου σε κάθε μικροκλίμα μετά από έναν καθολικό ψεκασμό και κατανομή πληθυσμού (b) στην έναρξη της προσομοίωσης και (c) στο τέλος της προσομοίωσης

ψεκασμό στο δεύτερο σενάριο προσομοίωσης (Εικόνα 3.32), σαν ο ψεκασμός να μην είχε συμβεί.

3.3.5 Συμπεράσματα

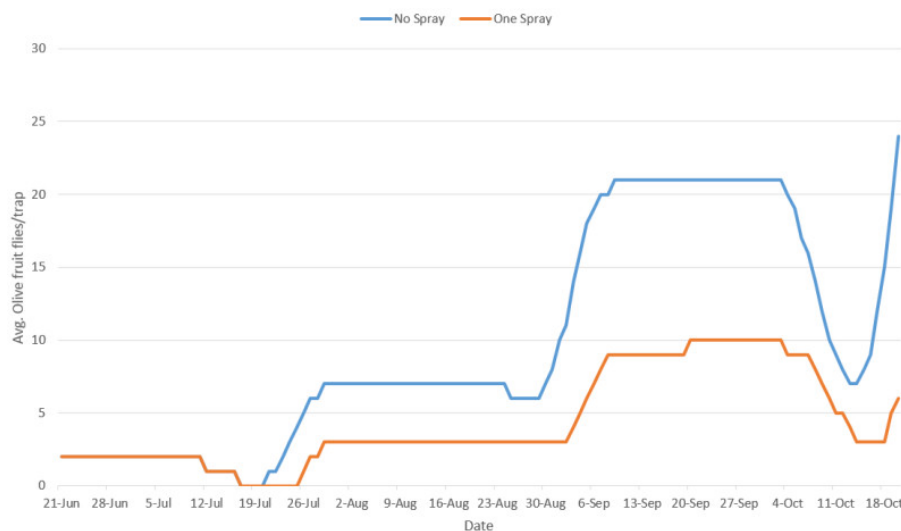
Τα αποτελέσματα των σεναρίων προσομοίωσης έδειξαν ότι λόγω της διαφοράς θερμοκρασίας όσο θερμότερο είναι ένα μικροκλίμα, εντός των ορίων ανάπτυξης του δάκου, τόσο πιο γρήγορα οι ενήλικοι δάκοι αναδύονται. Η χρονική διαφορά της πρώτης γενιάς δάκου μεταξύ του θερμότερου (μικροκλίμα A) και του ψυχρότερου (μικροκλίμα D) μικροκλίματος ήταν σχεδόν ένα μήνας.

Όταν πραγματοποιήθηκε ένας μοναδικός καθολικός ψεκασμός, με βάση τις πρακτικές των τοπικών αρχών στην προσομοίωση. Η εξέλιξη του πληθυσμού στα μικροκλίματα A και B επηρεάζεται σε μεγάλο βαθμό, ενώ στο μικροκλίμα Γ ο πληθυσμός επηρεάζεται πολύ ελαφρά. Τέλος, στο μικροκλίμα D ο ψεκασμός δεν έχει καμία επίδραση στην εξέλιξη του πληθυσμού.

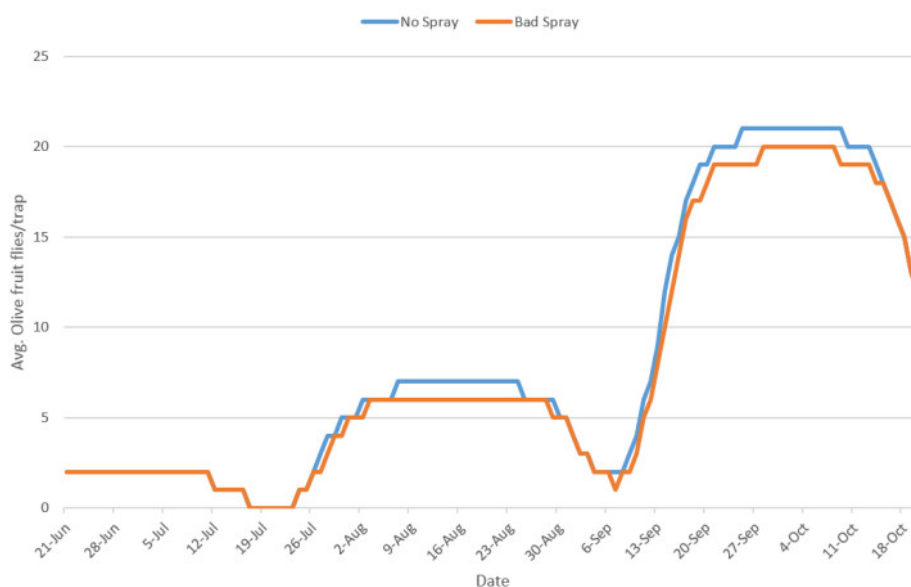


Εικόνα 3.29: Σύγκριση εξέλιξης πληθυσμού για το μικροκλίμα Α χωρίς και με ψεκασμό.

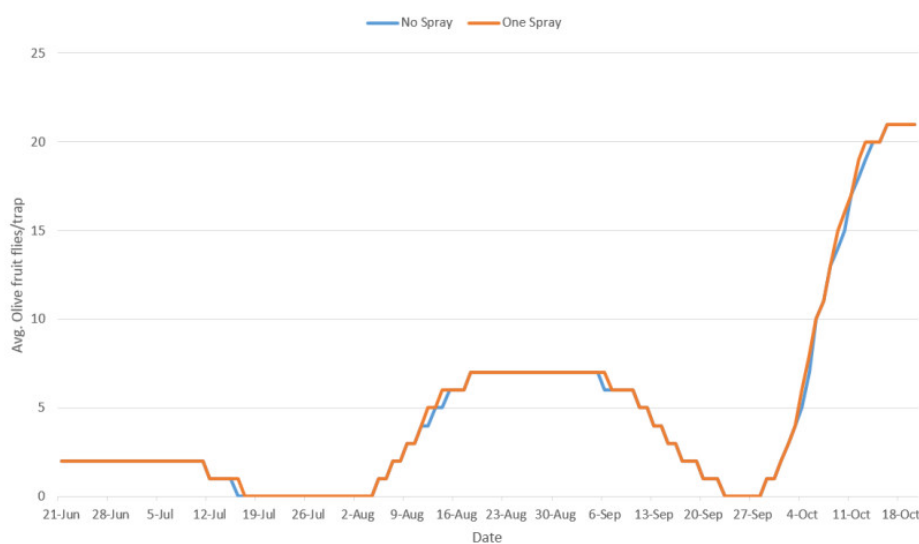
Συμπερασματικά, με βάση τα αποτελέσματα των σεναρίων προσομοίωσης που διενεργήθηκαν, ότι αφορά καθολικές πρακτικές ψεκασμού η ύπαρξη μικροκλίματος θα πρέπει να λαμβάνεται υπόψη. Για τα μικροκλίματα με ψυχρότερο κλίμα οι χημικές ουσίες που χρησιμοποιούνται στον ψεκασμό καθώς και ο χρόνος εφαρμογής του πηγαίνουν χαμένα, καθώς θα έχει μικρή επίδραση στον πληθυσμό του δάκου ή καθόλου.



Εικόνα 3.30: Σύγκριση εξέλιξης πληθυσμού για το μικροκλίμα Β χωρίς και με ψεκασμό.



Εικόνα 3.31: Σύγκριση εξέλιξης πληθυσμού για το μικροκλίμα C χωρίς και με ψεκασμό.



Εικόνα 3.32: Σύγκριση εξέλιξης πληθυσμού για το μικροκλίμα D χωρίς και με ψεκασμό.

Ανάπτυξη στοχαστικών αλγορίθμων και μοντέλων μηχανικής μάθησης για την εξέλιξη πληθυσμών: Εφαρμογή στο δάκο της Ελιάς

4. ΕΦΑΡΜΟΓΕΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ

Στο παρόν κεφάλαιο παρουσιάζονται προτεινόμενες μέθοδοι εφαρμογής μοντέλων μηχανικής μάθησης και τα αποτελέσματά τους με στόχο την καλύτερη αντιμετώπιση της δακοπροσβολής στον ελαιώνα. Αναλυτικότερα στην ενότητα 4.1 χρησιμοποιούνται τεχνικές μηχανικής μάθησης για την πρόβλεψη της επόμενης μέτρησης δακοπαγίδας, δεδομένης της γνώσης της προηγούμενης μέτρησης καθώς επίσης και των περιβαντολογικών συνθηκών. Στην ενότητα 4.2 παρουσιάζεται η ανάπτυξη μοντέλου για την αυτόματη αναγνώριση του δάκου, και κατά συνέπεια αυτόματη καταμέτρηση, σε εικόνες του περιεχομένου της παγίδας McPhail που χρησιμοποιείται συνήθως ως μέθοδος ελέγχου του πληθυσμού. Τέλος, στην ενότητα 4.3 γίνεται χρήση μεθόδων μηχανικής μάθησης για την αναγνώριση μικροκλιμάτων, γεγονός που θα βοηθούσε στην πρόβλεψη εστίων προσβολής του ελαιώνα από τον δάκο ανάλογα την εποχή του χρόνου και τα κλιματολογικά χαρακτηριστικά του κάθε μικροκλίματος.

4.1 Ανάπτυξη μοντέλου για την πρόβλεψη της επόμενης μέτρησης μιας δακοπαγίδας

Η επίδραση των εντόμων που προκαλούν βλάβες στις καλλιέργειες, έχει μεγάλη σημασία. Τα παράσιτα μολύνουν και τρέφονται από τους καρπούς και τους σπόρους των γεωργικών προϊόντων, μειώνοντας έτσι σημαντικά την αξία τους. Αυτό με τη σειρά του οδηγεί σε απώλεια τόσο των πρώτων υλών όσο και των επενδυμένων κεφαλαίων [Murray et al., 2013; Haniotakis, 2005]. Ο δάκος της ελιάς δεν αποτελεί εξαίρεση καθώς η εξάρτηση του από τον καρπό της ελιάς για την αναπαραγωγή του μπορεί να προκαλέσει μεγάλες ποιοτικές και οικονομικές ζημιές.

Η υπάρχουσα έρευνα για το δάκο της ελιάς έχει εστιάσει κυρίως σε πτυχές όπως η βιολογία, η οικολογία, η διαχείριση και ο αντίκτυπος στην παραγωγή ελιάς. Από την άλλη πλευρά, η πρόβλεψη προσβολής, με χρήση μεθόδων μηχανικής μάθησης, δεν έχει λάβει ακόμη σημαντική προσοχή, παρά την ευρεία διαθεσιμότητα δεδομένων από τους ελαιώνες, καθώς και τα θετικά αποτελέσματα μιας τέτοιας πρόβλεψης που θα μπορούσε να επιφέρει στην αντιμετώπιση των προσβολών του δάκου και κατά συνέπεια βελτιώνοντας την παραγωγή του καρπού της ελιάς.

4.1.1 Σχετική Έρευνα

Αυτή η ενότητα περιγράφει τις απαραίτητες βασικές πληροφορίες σχετικά με τις μεθόδους μηχανικής μάθησης καθώς και με την υπάρχουσα έρευνα σχετικά με την πρόβλεψη προσβολής από τον δάκο της ελιάς.

4.1.1.1 Αλγόριθμοι μηχανικής μάθησης

Υπάρχουν διάφοροι αλγόριθμοι ταξινόμησης που είναι κατάλληλοι για τους πειραματικούς σκοπούς. Καθώς οι προσεγγίσεις ταξινόμησης στην πρόβλεψη προσβολής για τον δάκο της

ελιάς είναι εξαιρετικά περιορισμένες σε αριθμό στην βιβλιογραφία, η επιλογή των αλγορίθμων έχει βασιστεί σε μεγάλο βαθμό σε διερευνητικά κριτήρια με στόχο την κάλυψη διαφόρων οικογενειών μάθησης συμπεριλαμβανομένης της μετα-μάθησης. Για τα πειράματα που ακολουθούν χρησιμοποιήθηκαν οι παρακάτω αλγόριθμοι μηχανικής μάθησης:

- J48 [Witten and Frank, 2005] είναι ένας αλγόριθμος δένδρου απόφασης και είναι μια έκδοση του C4.5, ενός παλιότερου αλγορίθμου που αναπτύχθηκε από τον J. Ross Quinlan [1996]. Ο αλγόριθμος C4.5 δημιουργεί ένα δέντρο αποφάσεων με βάση το κέρδος πληροφορίας των χαρακτηριστικών στα εκπαιδευτικά δεδομένα. Πιο συγκεκριμένα, σε κάθε επανάληψη αναγνωρίζεται το χαρακτηριστικό του οποίου οι τιμές διακρίνουν σαφέστερα τα παραδείγματα εκπαίδευσης σύμφωνα με την ετικέτα ταξινόμησης. Ο αλγόριθμος τερματίζει την εκτέλεση του όταν δεν υπάρχουν άλλα χαρακτηριστικά για εξερεύνηση ή όταν όλα τα παραδείγματα εκπαίδευσης έχουν διαχωριστεί σύμφωνα με την ετικέτα ταξινόμησης τους. Επιπροσθέτως, ο αλγόριθμος J48 δύο μεθοδολογίες κλαδέματος δέντρων: Η πρώτη είναι γνωστή ως *subtree Replication* και αντικαθιστά έναν κόμβο σε ένα δέντρο αποφάσεων με το αντίστοιχο φύλλο, εάν το συγκεκριμένο υπό-δέντρο δεν βοηθά στην ακρίβεια της ταξινόμησης. Αυτή η διαδικασία κλαδέματος ξεκινά από τα φύλλα του πλήρως διαμορφωμένου δέντρου και κινείται από κάτω προς τα πάνω προς τη ρίζα του δέντρου. Η δεύτερη μεθοδολογία είναι γνωστή ως *subtree raise* στην οποία ένας κόμβος μπορεί να αντικαταστήσει άλλους κόμβους ενώ μετακινείται προς τη ρίζα. Αυτός ο τύπος κλαδέματος τις περισσότερες φορές έχει ασήμαντη επίδραση στα μοντέλα δέντρων αποφάσεων
- Sequential Minimal Optimization ή SMO [Platt, 1998], είναι ένας βελτιωμένος αλγόριθμος για την εκπαίδευση μηχανών υποστήριξης διανυσμάτων (support vector machines). Ο αλγόριθμος SMO τεμαχίζει ένα μεγάλο πρόβλημα βελτιστοποίησης τετραγωνικού προγραμματισμού μετατρέποντας το σε μικρότερα υπό-προβλήματα. Τα υπό-προβλήματα επιλύονται γρήγορα με μια αναλυτική προσέγγιση όπου ο αλγόριθμος SMO αποφεύγει να χρησιμοποιήσει επιπλέον χρόνο για την βελτιστοποίηση αριθμητικού τετραγωνικού προγραμματισμού σαν εσωτερικό βρόγχο. Αυτή η τεχνική επιτρέπει στον αλγόριθμο SMO να μειώσει σημαντικά το χρόνο υπολογισμού.
- Naive Bayes [Witten and Frank, 2005], είναι ένας πιθανός ταξινομητής που βασίζεται στην υπόθεση ανεξαρτησίας υπό όρους [Russel and Norvig, 2009b], η οποία υποθέτει ότι η εμφάνιση ενός συγκεκριμένου χαρακτηριστικού δεδομένης της τιμής ταξινόμησης δεν σχετίζεται με την εμφάνιση οποιουδήποτε άλλης χαρακτηριστικού σε ένα σύνολο δεδομένων. Αν και δεν ισχύει στην πραγματικότητα, αυτή η υπόθεση έχει αποδειχθεί ότι ανταπεξέρχεται σε διάφορα προβλήματα ταξινόμησης. Ένα επιπρόσθετο πλεονέκτημα αυτού του αλγορίθμου είναι η ανάγκη για ένα μικρό όγκο εκπαιδευτικών δεδομένων για τον προσδιορισμό των παραμέτρων που απαιτούνται για την ταξινόμηση. Λόγω της υπόθεσης των ανεξάρτητων μεταβλητών, δεν χρειάζεται να εκτιμηθεί ολόκληρος ο πίνακας συνδιακύμανσης αλλά μόνο οι διαφοροποιήσεις των μεταβλητών για κάθε κλάση.
- Random Forest [Breiman, 2001], είναι ένας αλγόριθμος ταξινόμησης μετα-μάθησης που εκτελείται επαναληπτικά. Σε κάθε επανάληψη, ένα δέντρο αποφάσεων δημιουργείται

από ένα τυχαία επιλεγμένο υποσύνολο των χαρακτηριστικών. Ο αριθμός των επαναλήψεων είναι προκαθορισμένος, ενώ το τελικό σφάλμα ταξινόμησης είναι το μέσο σφάλμα όλων των επαναλήψεων.

- AdaBoost [Freund and Schapire, 1996], είναι ένας άλλος αλγόριθμος μετα-μάθησης, ο οποίος αλλάζει επαναληπτικά τα βάρη των παραδειγμάτων, με βάση το εάν αυτά ταξινομήθηκαν σωστά (ή όχι) σε προηγούμενη επανάληψη. Με αυτόν τον τρόπο, ο αλγόριθμος αναγκάζεται να επικεντρωθεί σε περιπτώσεις που είναι δύσκολο να ταξινομηθούν. Η τελική ταξινόμηση προέρχεται από τη στάθμιση των μοντέλων που δημιουργούνται μετά από κάθε επανάληψη.
- Ibk [Witten and Frank, 2005], είναι μια υλοποίηση του αλγορίθμου k κοντινότερων γειτόνων. Χρησιμοποιώντας την Ευκλείδεια απόσταση ως μετρικό δείκτη, προσδιορίζει τα k παραδείγματα εκπαίδευσης που είναι πλησιέστερα σε ένα συγκεκριμένο δοκιμαστικό παράδειγμα και, μέσω της πλειοψηφίας, επιλέγει την κλάση που θα το ταξινομήσει. Η τιμή k μπορεί να προσδιοριστεί ρητά ή να εκτιμηθεί βέλτιστα χρησιμοποιώντας διασταυρούμενη επικύρωση
- Multilayer Perceptron, είναι ένα τεχνητό νευρωνικό δίκτυο [Rosenblatt, 1961], όπου κάθε ένα από τα πολλαπλά στρώματα των κόμβων είναι πλήρως συνδεδεμένο με το επόμενο.

Πολύ συχνά οι αλγόριθμοι μηχανικής μάθησης αντιμετωπίζουν το πρόβλημα της υπέρ-προσαρμογής (overfitting), δηλαδή αποδίδουν βέλτιστα όταν αξιολογούν δεδομένα εκπαίδευσης, αλλά φτωχά όταν αξιολογούν νέα αθέατα δεδομένα. Συνήθως, όσο πιο περίπλοκο είναι το μοντέλο που παράγεται από έναν αλγόριθμο μάθησης, τόσο πιο επιρρεπής είναι στην υπέρ-προσαρμογή. Στα πειράματά μας αντιμετωπίσαμε το πρόβλημα της υπέρ-προσαρμογής με τους ακόλουθους τρόπους:

- Εφαρμόζοντας μέτα-κλάδεμα, δηλαδή την αύξηση των υπόδενδρων, των δημιουργημένων δένδρων από τους ταξινομήτες που βασίζονται σε δένδρα (J48 και RandomForest)
- Χρησιμοποιώντας μια πολυωνυμική συνάρτηση χαμηλής πολυπλοκότητας στον πυρήνα της διανυσματικής μηχανής υποστήριξης, δηλαδή ένα πολυώνυμο πρώτου βαθμού.
- Επιλέγοντας μια χαμηλής πολυπλοκότητας δομή, που περιλαμβάνει μόνο ένα κρυφό στρώμα και εξετάζοντας την επίδοση του με διάφορους αριθμούς κόμβων, ξεκινώντας από την χαμηλή πολυπλοκότητα ενός κόμβου μέχρι τους 15 κόμβους.
- Διερευνώντας άλλες δομές νευρωνικών δικτύων χαμηλής πολυπλοκότητας με διαφορετικό αριθμό εισόδων και κρυφών στρωμάτων και κατά συνέπεια ποικιλία στην πολυπλοκότητα

Όσο αφορά το πρόβλημα των υπερβολικών τιμών, ανάμεσα στα χαρακτηριστικά του προτεινόμενου χαρακτηριστικού διανύσματος μόνο ένα χαρακτηριστικό είχε αριθμητικές τιμές. Μετά από ανάλυση των αριθμητικών χαρακτηριστικών δεν βρέθηκαν τιμές με μια ασυνήθιστη/έντονα αποκλίνουσα τιμή. Επομένως ο χώρος του χαρακτηριστικού μας διανύσματος δεν περιέχει υπερβολικές τιμές.

4.1.1.2 Έρευνα σχετικά με την πρόβλεψη προσβολής από τον δάκο

Όπως περιγράφηκε στην ενότητα 1.2.6.1 έχουν υπάρξει αρκετές περιπτώσεις πρόβλεψης δακοπροσβολής μέσω μοντελοποίησης και προσομοίωσης.

Οι del Sagrado and del Águila [2007] πρότειναν την πρόβλεψη προσβολής από τον δάκο χρησιμοποιώντας πληροφορίες για την υγεία του δένδρου της ελιάς καθώς επίσης μετρήσεις από δακοπαγίδες. Η έννοια της υγείας της σοδειάς, όπως περιγράφεται, περιλαμβάνει μετρήσεις που αφορούν και τα τρία στάδια που συνδέονται με την καλλιέργεια: παρακολούθηση και της καλλιέργειας. Η έρευνά τους χρησιμοποίησε ταξινομικά δέντρα και Bayesian δίκτυα για τον προσδιορισμό των προγνωστικών που υποδεικνύουν απαιτήσεις φυτοϋγειονομικής επεξεργασίας σε σχέση με την προσβολή από τον δάκο. Τα αποτελέσματα που προέκυψαν δείχνουν ότι τα ταξινομικά δέντρα αποδίδουν ευνοϊκά σε σχέση με τα Bayesian δίκτυα σε ότι αφορά τους όρους απλότητας, επιτυχίας και ευαισθησίας, αν και με φτωχή απόδοση ως προς την εξειδίκευση. Επιπλέον, οι del Sagrado κ.ά. προσπάθησαν να μειώσουν την πολυπλοκότητα με τη χρήση ενός υποσυνόλου των διαθέσιμων μεταβλητών με αβέβαια αποτελέσματα.

Η προσέγγιση που προτείνεται εδώ διαφοροποιείται από την προαναφερθείσα [del Sagrado and del Águila, 2007] προτείνοντας ένα διάνυσμα χαρακτηριστικών το οποίο βασίζεται στις μετρήσεις των δακοπαγίδων καθώς και σε περιβαλλοντικούς παράγοντες, όπως η θερμοκρασία, αντί της υγείας της ελιάς. Η προσέγγιση έχει τα ακόλουθα πλεονεκτήματα και μειονεκτήματα σε σχέση με το έργο που επιτελέστηκε από τους del Sagrado and del Águila [2007]:

Πλεονεκτήματα

- Η χρήση του Degree day (DD) μοντέλου ως χαρακτηριστικό. Το DD είναι ένα μοντέλο συσχετισμού μεταξύ της θερμοκρασίας και την ανάπτυξης, και γενικότερα δραστηριότητας, του πληθυσμού ενός οργανισμού [Gonçalves and Torres, 2011]. Το μοντέλο αυτό υπολογίζει την θερμότητα που συσσωρεύεται από έναν οργανισμό κατά τη διάρκεια της ημέρας, μεταξύ κατώτατα και ανώτατα όρια. Χρησιμοποιώντας αυτό το μοντέλο, για παράδειγμα, η ποσότητα της θερμότητας και κατ'επέκταση ο συνολικός χρόνος που απαιτείται για να μετασχηματιστεί ένας οργανισμός από το ένα αναπτυξιακό στάδιο στο άλλο.
- Το προτεινόμενο διάνυσμα χαρακτηριστικών θα μπορούσε να εφαρμοστεί για την πρόβλεψη προσβολής άλλων παρασίτων πλην του δάκου της ελιάς. Δεδομένου ότι, όπως περιγράφηκε παραπάνω, το χαρακτηριστικό DD συσχετίζει την ανάπτυξη ενός οργανισμού με τη θερμοκρασία του περιβάλλοντος. Ως εκ τούτου, η DD χρησιμοποιείται στη γεωργία για να προβλέψει ορισμένα γεγονότα, όπως η πρόβλεψη φυτικών σταδίων [Miller et al., 2001] ή η πρόβλεψη ανάπτυξης εντόμων και εν συνεχεία τη δραστηριότητα επιβλαβών οργανισμών [Herms, 2004].
- Ο υπολογισμός του DD είναι μια τυποποιημένη μεθοδολογία και διάφορες μέθοδοι μπορούν να χρησιμοποιηθούν [Gonçalves and Torres, 2011; Brown, 1989]. Επιπλέον, πρόκειται για ποσοτικό χαρακτηριστικό και η διαδικασία υπολογισμού μπορεί να γίνει με τη χρήση ενός αλγορίθμου. Αντίθετα, στο έργο των del Sagrado and del Águila [2007]

απαιτείται η χρήση εμπειρογνομώνων για την αξιολόγηση της υγείας των καλλιεργειών με βάση το δείγμα που συλλέγεται. Στην περίπτωση που υπάρχουν πολλοί εμπειρογνώμονες, υπάρχει η πιθανότητα να υπάρξει διαφορά απόψεων.

Μειονεκτήματα

- Ο προσδιορισμός της ημερομηνίας έναρξης της συσσώρευσης DD μπορεί να είναι μια δυσκίνητη εργασία. Στην περίπτωση του, η περίοδος έναρξης είναι όταν τα πρώτα αυγά τοποθετούνται μέσα στα ελαιόδεντρα. Ως εκ τούτου, απαιτείται έντονη παρακολούθηση του ελαιώνα για να καθοριστεί πότε έχουν τοποθετηθεί τα πρώτα αυγά. Για παράδειγμα, στην Κέρκυρα, υπό ευνοϊκές περιβαλλοντικές συνθήκες, δάκος της ελιάς ξεκινάει την ωοτοκία από τα τέλη Ιουνίου έως τις αρχές Ιουλίου. Ωστόσο, υπήρξαν περιπτώσεις όπου, εξαιτίας του ζεστού καιρού, η ωοτοκία άρχισε στα τέλη Αυγούστου. Επομένως, ο καθορισμός της ημερομηνίας έναρξης της συσσώρευσης μπορεί να κυμανθεί από μερικές ημέρες έως μήνες.

4.1.2 Συλλογή δεδομένων

4.1.2.1 Περιβαλλοντολογικά δεδομένα

Τα δεδομένα που χρησιμοποιήθηκαν στα πειράματα που πραγματοποιήθηκαν συλλέχθηκαν από περιβαλλοντικούς αισθητήρες και δακωπαγίδες που τοποθετήθηκαν σε 16 θέσεις στη βορειοδυτική πλευρά του νησιού της Κέρκυρας. Οι μετρήσεις των δακωπαγίδων έδειξαν τον συνολικό αριθμό των δάκων που πιάστηκαν στην παγίδα. Κάθε μέτρηση πραγματοποιήθηκε σε όλες τις θέσεις των παγίδων κάθε πέντε ημέρες για την περίοδο από τις 10 Ιουνίου 2015 έως τις 29 Σεπτεμβρίου 2015 και από τις 8 Ιουλίου 2016 έως την 1η Οκτωβρίου 2016. Οι αισθητήρες σε κάθε θέση κατέγραφαν τιμές θερμοκρασίας κάθε 15 λεπτά.

4.1.2.2 Επιλογή χαρακτηριστικών

Για την πραγματοποίηση πειραμάτων ταξινόμησης, τα προαναφερθέντα περιβαλλοντικά δεδομένα μετατράπηκαν στο ακόλουθο σύνολο αριθμητικών χαρακτηριστικών (για να αντιπροσωπεύσουν τις μετρήσεις ως διανύσματα χαρακτηριστικών εκμάθησης):

1. Μέση θερμοκρασία των τελευταίων πέντε ημερών πριν την επόμενη μέτρηση
2. Μέση μέγιστη θερμοκρασία των τελευταίων πέντε ημερών πριν την επόμενη μέτρηση
3. Μέση ελάχιστη θερμοκρασία των τελευταίων πέντε ημερών πριν την επόμενη μέτρηση
4. Ημέρα 1 Μέση θερμοκρασία
5. Ημέρα 1 Μέγιστη θερμοκρασία

6. Ημέρα 1 Ελάχιστη θερμοκρασία
7. Ημέρα 2 Μέση θερμοκρασία
8. Ημέρα 2 Μέγιστη θερμοκρασία
9. Ημέρα 2 Ελάχιστη θερμοκρασία
10. Ημέρα 3 Μέση θερμοκρασία
11. Ημέρα 3 Μέγιστη θερμοκρασία
12. Ημέρα 3 Ελάχιστη θερμοκρασία
13. Ημέρα 4 Μέση θερμοκρασία
14. Ημέρα 4 Μέγιστη θερμοκρασία
15. Ημέρα 4 Ελάχιστη θερμοκρασία
16. Ημέρα 5 Μέση θερμοκρασία
17. Ημέρα 5 Μέγιστη θερμοκρασία
18. Ημέρα 5 Ελάχιστη θερμοκρασία
19. Τελευταία μέτρηση δακοπαγίδας
20. Degree Days, που θα έχουν συλλεχθεί μέχρι την επόμενη μέτρηση
21. Επόμενη μέτρηση δακοπαγίδας

Διάφορες μέθοδοι μπορούν να χρησιμοποιηθούν για τον υπολογισμό του χαρακτηριστικού Degree Days (DD) όπως οι μέθοδοι max-min, η μέθοδος "saw-tooth" και η διπλή ημιτονοειδής καμπύλη, όπως περιγράφεται από τους Wilson and Barnett [1983]. Ωστόσο, οι πιο συνηθισμένες είναι η μονη ημιτονοειδής καμπύλη και μέθοδοι μέσης θερμοκρασίας [Brown, 1989]. Για τον υπολογισμό της DD σε αυτό το έργο, χρησιμοποιήθηκε η ακόλουθη μέθοδος υπολογισμού που παρουσιάστηκε αρχικά στο [27], όπως φαίνεται στην Εξίσωση 4.1

$$DD = (t - T_L) * \left(1 - \frac{1}{1 + e^{-10*(t-T_U)}}\right) \quad (4.1)$$

όπου t η μέση θερμοκρασία της ημέρας T_L και T_U είναι το κάτω και άνω, αντίστοιχα, κατώφλι θερμοκρασία ανάπτυξης του δάκου της ελιάς.

Σύμφωνα με τους συγγραφείς Croveti et al. [1982] και Gonçalves and Torres [2011], για να εκκολαφθούν τα αυγά του δάκου της ελιάς να απαιτούνται 49,77 DD, κατά μέσο όρο, ενώ για να φτάσει στο στάδιο του ενήλικα, ο δάκος της ελιάς απαιτεί 379,02 DD, κατά μέσο όρο, μετρώντας από την ημέρα εναπόθεσης του αυγού στους καρπούς της ελιάς. Η διαίσθηση για τη χρήση αυτού του σύνθετου χαρακτηριστικού είναι ότι, έχοντας DD μέσα στο διάνυσμα

χαρακτηριστικών, κάνοντας παραδοχή ως προς την ημερομηνία εναπόθεσης των αυγών πριν από την μέτρηση της δακοπαγίδας. Αυτό, όπως επίσης υποστηρίζεται στη βιβλιογραφία, αναμένεται να υποστηρίξει την πρόβλεψη έξαρσης.

Τέλος, το χαρακτηριστικό "Επόμενη μέτρηση δακοπαγίδας", χρησιμοποιήθηκε ως κλάση ταξινόμησης.

4.1.2.3 Εξαγωγή χαρακτηριστικού διανύσματος

Η διαδικασία εξαγωγής των χαρακτηριστικών για το χαρακτηριστικό διάνυσμα από τα δεδομένα αισθητήρα αυτοματοποιήθηκε με τη χρήση ενός script. Το script γράφτηκε στη προγραμματιστική γλώσσα Python και εξήγαγε όλα τα διανύσματα σε αρχείο CSV (Comma Separated Values) για περαιτέρω επεξεργασία. Τέλος, οι μετρήσεις των δακοπαγίδων προστέθηκαν χειροκίνητα σε κάθε αντίστοιχο παράδειγμα διανυσμάτων.

Τα χαρακτηριστικά που σχετίζονται με την θερμοκρασία, αρχικά αριθμητικά, διακρίνονται στις ακόλουθες τρεις διακριτές κλάσεις:

- <15, για θερμοκρασίες μικρότερες από 15 °C
- 15 to 32, για θερμοκρασίες μεταξύ 15 °C και 32 °C
- >32, για θερμοκρασίες μεγαλύτερες από 32 °C

Η διακριτοποίηση των τιμών θερμοκρασίας βασίστηκε στην περιοχή θερμοκρασιών (μεταξύ 15 °C και 32 °C [Vossen, 2014]), στην οποία ο δάκος είναι ενεργός. Σε περιπτώσεις όπου η θερμοκρασία του περιβάλλοντος είναι μικρότερη από το χαμηλότερό ή μεγαλύτερη από το ανώτερο όριο, τότε ο δάκος είναι ακίνητος λόγω κρύου ή ζέστης. Συνεπώς, υποθέτουμε ότι εκτός της βέλτιστης περιοχής θερμοκρασιών για τον δάκο της ελιάς, οι παγίδες δεν θα συλλαμβάνουν τυχόν μύγες.

Τα χαρακτηριστικά σχετικά με την μέτρηση της δακοπαγίδας έχουν επίσης διακριτοποιηθεί στις ακόλουθες κλάσεις:

- 0 to 4, κανένας ή μέχρι τέσσερεις δάκοι μέσα στην παγίδα,
- 5 to 6, πέντε με έξι δάκοι μέσα στην παγίδα
- ≥ 7 , μεγαλύτερες ή ίσες με επτά δάκους μέσα στην παγίδα

Η ποσοτικοποίηση των μετρήσεων των παγίδων είναι τριμερής με βάση τη θεωρητική ανάλυση, δηλ. το όριο προσβολής εξαρτάται με την εποχή που πραγματοποιούνται οι μετρήσεις. Συγκεκριμένα, τους καλοκαιρινούς μήνες το κατώφλι προσβολής ορίζεται σε επτά δάκους ανά παγίδα την εβδομάδα. Από την άλλη, από τον Σεπτέμβριο και μετά, το κατώφλι προσβολής μειώνεται σε πέντε δάκους ανά παγίδα την εβδομάδα, λόγω των ψυχρότερων καιρικών συνθηκών [8]. Επομένως, παρόλο που η τελευταία τιμή θα έδειχνε πάντα την προσβολή, η δεύτερη τιμή του θα εξαρτιόταν από την εποχή.

4.1.2.4 Το πρόβλημα της ανισορροπίας κλάσης

Η ανάλυση των μετρήσεων των παγίδων έδειξε χαμηλή συχνότητα εμφάνισης για την τιμή "5 to 6" σε συσσωρευμένα δεδομένα δύο ετών.

Αυτό οδηγεί σε υπερ-εκπροσώπηση της πρώτης και τελευταίας τιμής στα δεδομένα, σε σύγκριση με την δεύτερη τιμή. Η πρόβλεψη περιπτώσεων μειονοτικής κλάσης υποφέρει, λόγω της αραιότητας της.

Στην παρούσα έρευνα έχουν χρησιμοποιηθεί διάφορα μέσα για την αντιμετώπιση του προβλήματος της ανισορροπίας:

- Συνθετική υπερ-δειγματοληψία εφαρμόζεται στο σύνολο των δεδομένων χρησιμοποιώντας την τεχνική SMOTE [Chawla et al., 2002]. Η υπέρ-δειγματοληψία επιλέγεται έναντι της υπο-δειγματοληψίας λόγω του περιορισμένου αριθμού διαθέσιμων παραδειγμάτων. Η SMOTE (Synthetic Minority Over-sampling TEchnique) δημιουργεί νέα συν-παραδείγματα από ένα δείγμα τάξης μειοψηφίας, εξετάζοντας τους πλησιέστερους γείτονές του και συνδυάζοντας γραμμικά το δείγμα με κάθε γείτονα. Το SMOTE έχει χρησιμοποιηθεί εκτεταμένα για την ανίχνευση της διείσδυσης του δικτύου [Cieslak et al., 2006], για την ανίχνευση ορίων φράσης στην ομιλία [Liu et al., 2006], για την πρόβλεψη διανομής ειδών [Johnson et al., 2012], για την ανίχνευση του καρκίνου του μαστού [Fallahi and Jafari, 2011], καθώς και σε εφαρμογές βιοπληροφορικής [Batuwita and Palade, 2009; Xiao et al., 2011; MacIsaac et al., 2006; Wang et al., 2006; Doyle et al., 2011].
- Η μετα-μάθηση εφαρμόζεται για να «αναγκάσει» τον ταξινομητή να προσέξει τα λανθασμένα ταξινομημένα παραδείγματα. Επίσης πειραματιστήκαμε με την μέθοδο της ενίσχυσης καθώς και τον αλγόριθμο Random Forest.
- Οι αρχικές τρεις κλάσεις των μετρήσεων δακοπαγίδας μειώνονται δεδομένου του προαναφερθέντος διπλού κατωφλίου στην προσβολή με βάση τη θερμοκρασία περιβάλλοντος, και οι δύο κλάσεις "5 to 6" και " ≥ 7 " εξυπηρετούν ποιοτικά τον ίδιο σκοπό. Έτσι, οι κλάσεις "5 to 6" και " ≥ 7 " ενσωματώνονται στην υπερκλάση " ≥ 5 ", ενώ η ένδειξη μόλυνσης ενσωματώνεται στην ίδια κλάση.

4.1.3 Αξιολόγηση απόδοσης

Για την υποστήριξη της αποτελεσματικότητας του προτεινόμενου διανύσματος χαρακτηριστικών και των εξετασθέντων αλγορίθμων μηχανικής μάθησης, αυτή η ενότητα παρουσιάζει έναν αριθμό πειραμάτων που έχουν εκτελεστεί. Περιγράφεται επίσης μια συνοπτική περιγραφή της πλατφόρμας πειραμάτων και των δεδομένων, ακολουθούμενη από μια ανάλυση απόδοσης.

4.1.3.1 Πειραματική διάταξη

Δεδομένου ότι η προτεινόμενη προσέγγιση μηχανικής μάθησης, για το συγκεκριμένο πρόβλημα που περιγράφεται εδώ, γίνεται για πρώτη φορά, η επιλογή της αρχιτεκτονικής των προτεινόμενων μοντέλων ήταν καθαρά διερευνητική. Επομένως, μια βέλτιστη αρχιτεκτονική δεν μπορεί να γίνει γνωστή εκ των προτέρων. Τα προτεινόμενα μοντέλα επιλέχθηκαν με βάση την ικανότητά τους να αντιμετωπίσουν ικανοποιητικά το πρόβλημα της υπέρ-προσαρμογής, το είδος των χαρακτηριστικών και το μικρό αριθμό εκπαιδευτικών παραδειγμάτων που είχαμε στη διάθεσή μας.

Η πλατφόρμα μηχανικής μάθησης WEKA¹ χρησιμοποιήθηκε για τη διεξαγωγή των πειραμάτων ταξινόμησης. Στον αλγόριθμο Ibk, ο αριθμός των πλησιέστερων γειτόνων κυμαινόταν από 1 έως 33 γείτονες, ενώ επιλέχθηκαν μόνο μονές τιμές, προκειμένου να αποφευχθούν οι ισοπαλίες και η πλειοψηφία να είναι εφικτή. Για τον αλγόριθμο Multilayer Perceptron πραγματοποιήθηκαν πειράματα με ένα κρυφό στρώμα με τον αριθμό των κόμβων να κυμαίνονται από 1 έως 10.

Σε ορισμένες περιπτώσεις, χρησιμοποιήθηκε η τεχνική SMOTE [Chawla et al., 2002] προκειμένου να επαναληφθεί η δειγματοληψία στο σετ δεδομένων όταν περιελάμβανε μια κλάση μειοψηφίας. Κατά συνέπεια, εξετάζονται οι ακόλουθοι τέσσερις μέθοδοι:

Μέθοδος 1 Χρήση των ακόλουθων χαρακτηριστικών ως διάνυσμα χαρακτηριστικών με δεδομένα εκπαίδευσης τα δεδομένα του 2015. Στην διάθεση μας είχαμε 184 εκπαιδευτικά παραδείγματα, λόγω του μικρού μεγέθους των δεδομένων εκπαίδευσης, δεν ήταν δυνατή η υποβολή δοκιμαστικού σετ για την επικύρωση των αποτελεσμάτων. Ως εκ τούτου, χρησιμοποιήθηκε η μέθοδος k-fold cross validation (βλ. Υπό-ενότητα 1.3.4.3) με k να έχει οριστεί ίσο με δέκα.

1. Μέση θερμοκρασία των τελευταίων πέντε ημερών πριν την επόμενη μέτρηση
2. Μέση μέγιστη θερμοκρασία των τελευταίων πέντε ημερών πριν την επόμενη μέτρηση
3. Μέση ελάχιστη θερμοκρασία των τελευταίων πέντε ημερών πριν την επόμενη μέτρηση
4. Ημέρα 1 Μέση θερμοκρασία
5. Ημέρα 1 Μέγιστη θερμοκρασία
6. Ημέρα 1 Ελάχιστη θερμοκρασία
7. Ημέρα 2 Μέση θερμοκρασία
8. Ημέρα 2 Μέγιστη θερμοκρασία
9. Ημέρα 2 Ελάχιστη θερμοκρασία
10. Ημέρα 3 Μέση θερμοκρασία
11. Ημέρα 3 Μέγιστη θερμοκρασία

¹<http://www.cs.waikato.ac.nz/ml/weka/>

12. Ημέρα 3 Ελάχιστη θερμοκρασία
13. Ημέρα 4 Μέση θερμοκρασία
14. Ημέρα 4 Μέγιστη θερμοκρασία
15. Ημέρα 4 Ελάχιστη θερμοκρασία
16. Ημέρα 5 Μέση θερμοκρασία
17. Ημέρα 5 Μέγιστη θερμοκρασία
18. Ημέρα 5 Ελάχιστη θερμοκρασία
19. Τελευταία μέτρηση δακοπαγίδας
20. Επόμενη μέτρηση δακοπαγίδας

Μέθοδος 2 Χρησιμοποιήθηκε το διάνυσμα χαρακτηριστικών της Μεθόδου 1, με εφαρμογή της τεχνικής SMOTE στα δεδομένα του 2015 (89 παραδείγματα) για εκπαίδευση ενώ τα δεδομένα του 2016 (112 παραδείγματα) χρησιμοποιήθηκαν για αξιολόγηση. Η μείωση των παραδειγμάτων εκπαίδευσης για τα δεδομένα του 2015 σε σχέση με την Μέθοδο 1 οφείλεται ότι λήφθηκαν υπόψη μόνο μετρήσεις από αρχές Ιουλίου και έπειτα.

Μέθοδος 3 Χρήση του ακόλουθου διανύσματος χαρακτηριστικών, με εφαρμογή της τεχνικής SMOTE στα δεδομένα του 2015 (89 παραδείγματα) για εκπαίδευση ενώ τα δεδομένα του 2016 (112 παραδείγματα) χρησιμοποιήθηκαν για αξιολόγηση.

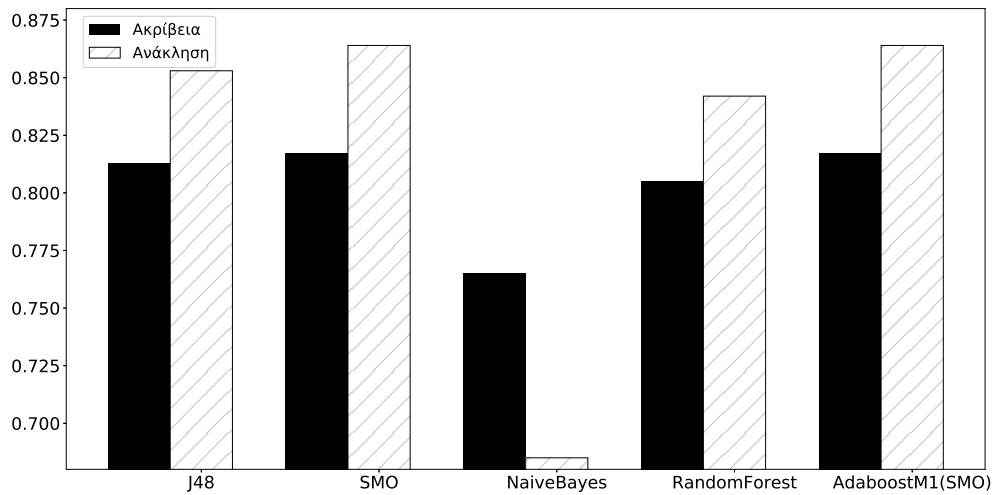
1. Τελευταία μέτρηση δακοπαγίδας
2. Degree Days, που θα έχουν συλλεχθεί μέχρι την επόμενη μέτρηση
3. Επόμενη μέτρηση δακοπαγίδας

Μέθοδος 4 Το διάνυσμα χαρακτηριστικών που χρησιμοποιήθηκε στη Μέθοδο 3, με μείωση των τιμών, των χαρακτηριστικών σχετικά με τις μετρήσεις των δακοπαγίδων, από τρεις τιμές σε δύο, τόσο για τα δεδομένα του 2015 (89 παραδείγματα) όσο και για του 2016 (112 παραδείγματα). Και τα δύο σετ δεδομένων δόθηκαν για εκπαίδευση των αλγορίθμων ενώ για την αξιολόγηση χρησιμοποιήθηκε η μέθοδος του 10-fold cross validation.

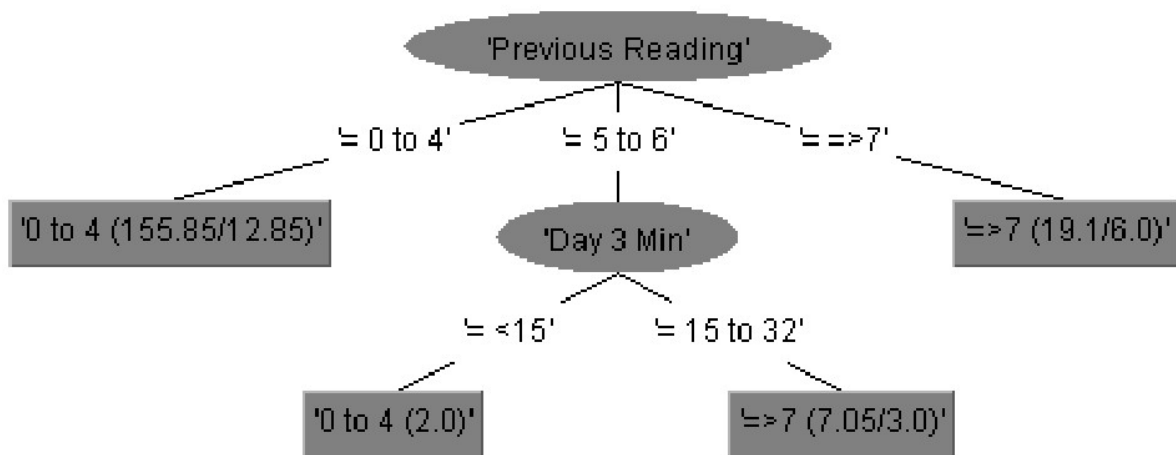
4.1.3.2 Αποτελέσματα

Μέθοδος 1 Στην Εικόνα 4.1 εμφανίζονται τα αποτελέσματα ταξινόμησης των πέντε προαναφερόμενων αλγορίθμων μηχανικής μάθησης. Ο αλγόριθμος SMO παράγει τα καλύτερα αποτελέσματα σε ακρίβεια και ανάκληση, με τον αλγόριθμο J48 να βρίσκεται αρκετά κοντά. Ο AdaBoostM1 παράγει τα ίδια αποτελέσματα χρησιμοποιώντας τον αλγόριθμο SMO ως βασικό εκπαιδευτή. Από την άλλη πλευρά, ο NaiveBayes παράγει τα χειρότερα αποτελέσματα σε σύγκριση με τους άλλους αλγορίθμους, με σημαντική μείωση στην ακρίβεια και μια αρκετά αξιοσημείωτη χαμηλή ανάκληση.

Η Εικόνα 4.2 παρουσιάζει το κλαδεμένο δέντρο που κατασκευάστηκε από τον αλγόριθμο J48. Είναι σαφές ότι ο αλγόριθμος θεωρεί το χαρακτηριστικό "Τελευταία μέτρηση δακοπαγίδας"



Εικόνα 4.1: Σύγκριση Ακρίβειας και Ανάκλησης για πέντε αλγόριθμους ταξινόμησης, για την πειραματική μέθοδο 1



Εικόνα 4.2: Το δένδρο απόφασης που κατασκευάστηκε από τον αλγόριθμο J48.

ως το πιο σημαντικό χαρακτηριστικό για ταξινόμηση. Με την ελάχιστη θερμοκρασία της ημέρας 3 να χρησιμοποιείται όταν η προηγούμενη τιμή έχει τιμή από "5 to 6".

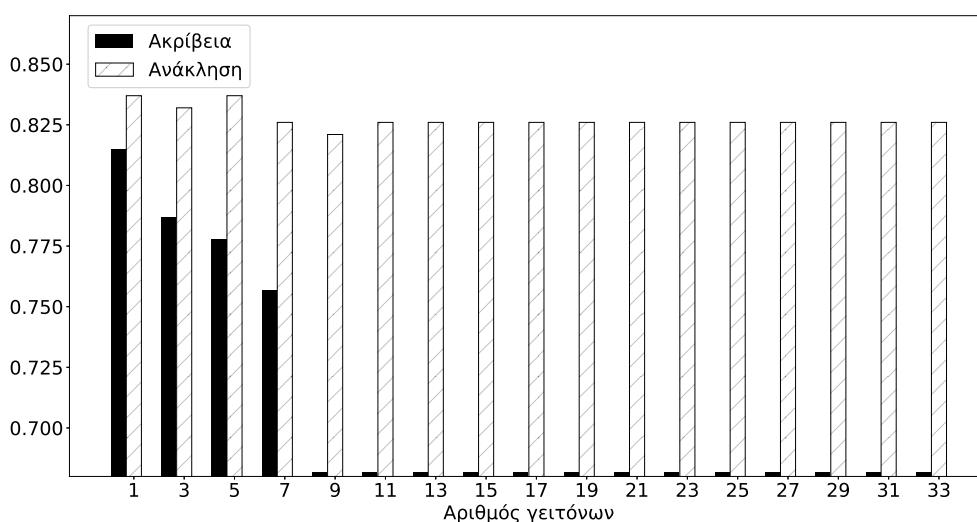
Το επόμενο πείραμα χρησιμοποιεί τον αλγόριθμο IBk, όπως φαίνεται στην Εικόνα 4.3, με μέγιστη ανάκληση για έναν και πέντε πλησιέστερους γείτονες. Το καλύτερο αποτέλεσμα τόσο σε ανάκληση όσο και σε ακρίβεια επιτυγχάνετε όταν χρησιμοποιείτε έναν πλησιέστερο γείτονα. Από τρεις πλησιέστερους γείτονες και μετά η ακρίβεια του αλγόριθμου μειώνεται, με σημαντική πτώση μετά από εννέα πλησιέστερους γείτονες. Μετά από έντεκα πλησιέστερους γείτονες, η ακρίβεια και η ανάκληση παραμένουν σταθερές.

Τέλος, ο αλγόριθμος Multilayer Perceptron, που φαίνεται στην Εικόνα 4.4, παρουσιάζει μεγάλη διαφορά μεταξύ των τιμών ανάκλησης και ακρίβειας για οποιοδήποτε αριθμό κόμβων στο κρυφό στρώμα. Το καλύτερο αποτέλεσμα τόσο σε ακρίβεια όσο και σε ανάκληση προκύπτει για τρεις κόμβους στο κρυφό στρώμα.

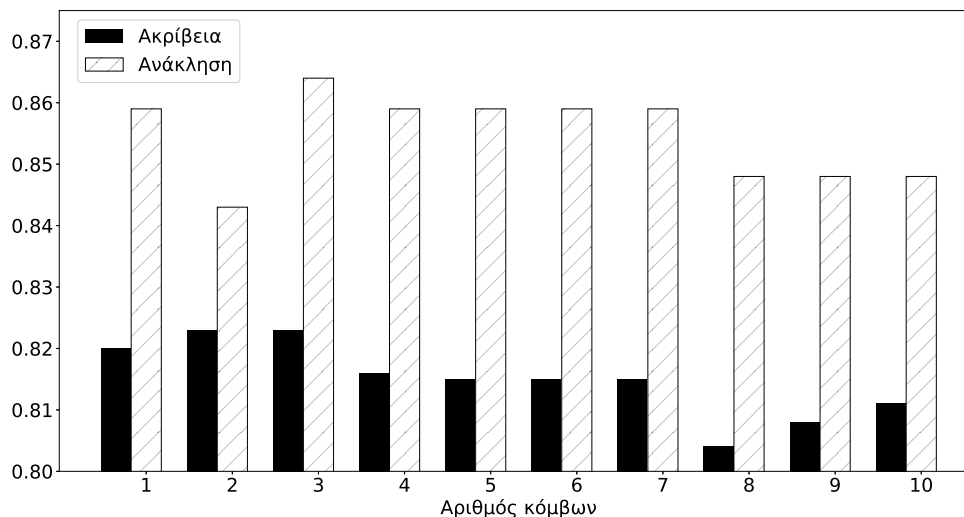
Συγκρίνοντας όλους τους αλγόριθμους μηχανικής μάθησης που χρησιμοποιήθηκαν στα πειράματα, οι καλύτερες επιδόσεις επιτεύχθηκαν από τον αλγόριθμο SMO και τον αλγόριθμο AdaBoostM1, χρησιμοποιώντας τον αλγόριθμο SMO ως βασικό εκπαιδευτή.

Είναι σημαντικό να σημειωθεί ότι οι τιμές του χαρακτηριστικού ταξινόμησης δεν είναι ισορροπημένες. Η Εικόνα 4.5 απεικονίζει την κατανομή των τριών τιμών μεταξύ των 184 περιπτώσεων. Πάνω από το 80% των περιπτώσεων έχουν τιμή "0 to 4", ενώ η τιμή "5 to 6" είναι η τιμή που εμφανίζεται λιγότερες φορές, με εννέα περιπτώσεις. Αυτό το γεγονός εξηγεί γιατί ο αλγόριθμος J48, όπως φαίνεται από τον κατασκευασμένο δέντρο στην Εικόνα, δεν μπορεί να ταξινομήσει κανένα στιγμιότυπο στην τιμή "5 to 6".

Στην Εικόνα 4.6 εμφανίζονται τα αποτελέσματα ταξινόμησης όλων των προαναφερθέντων αλγορίθμων μηχανικής μάθησης για κάθε τιμή ταξινόμησης. Στην περίπτωση αυτή, το χαρακτηριστικό διάνυσμα αποτελείται από 20 χαρακτηριστικά, 18 από τα οποία σχετίζονται με τη θερμοκρασία ενώ τα δύο τελευταία είναι σχετικά με τις μετρήσεις δακόπαγίδας. Η τιμή "5 to 6" περιλαμβάνει πολύ λίγα δεδομένα για να παραχθούν χρήσιμα αποτελέσματα. Τα αποτελέσματα των αλγορίθμων J48, SMO, AdaBoostM1, IBk και Multilayer perceptron είναι συγκρίσιμα για την τιμή "0 to 4", ενώ για την τιμή " ≥ 7 " ο αλγόριθμος SMO παράγει τα καλύτερα αποτελέσματα. Από την άλλη πλευρά, ο αλγόριθμος Naïve Bayes παράγει τα χειρότερα αποτελέσματα σε σύγκριση με τους άλλους αλγορίθμους, με μια σημαντική μείωση σε όλες τις τιμές.

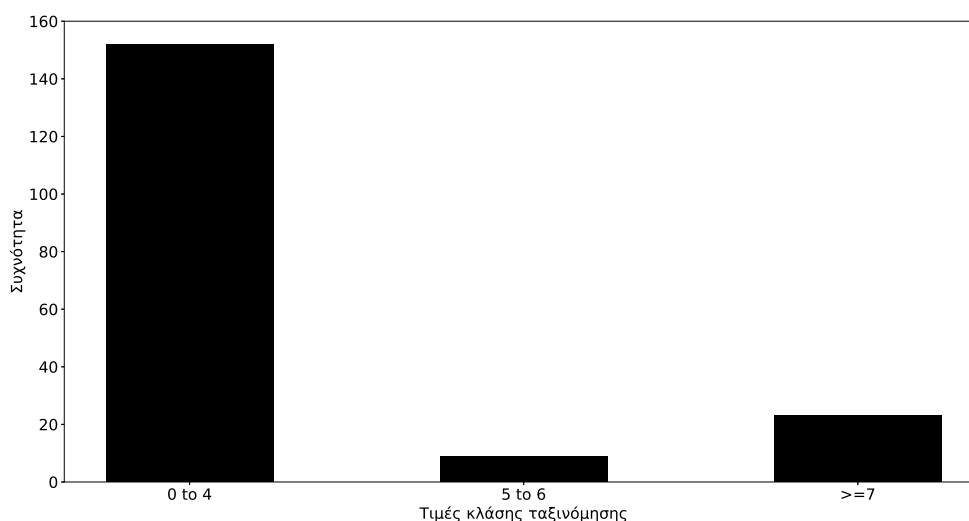


Εικόνα 4.3: Ακρίβεια και Ανάκληση του αλγορίθμου IBk για διαφορετικούς k γείτονες, για την πειραματική μέθοδο 1

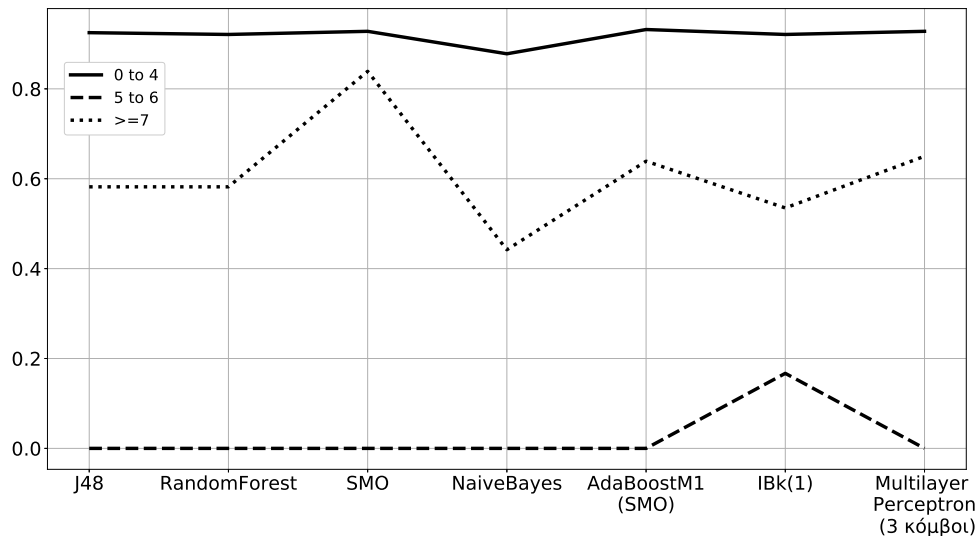


Εικόνα 4.4: Ακρίβεια και Ανάκληση του αλγορίθμου Multilayer Perceptron για διαφορετικό αριθμό κόμβων, για την πειραματική μέθοδο 1

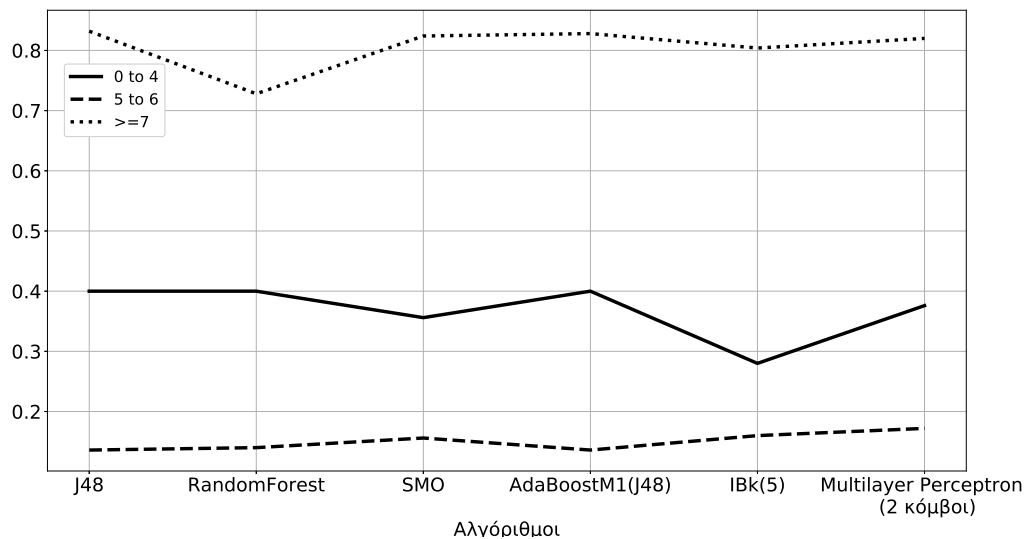
Μέθοδος 2 Στο επόμενο σύνολο πειραμάτων, το διάνυσμα χαρακτηριστικών της Μεθόδου 1 χρησιμοποιήθηκε με εφαρμογή της τεχνικής SMOTE στα δεδομένα του 2015 για εκπαίδευση ενώ τα αρχικά δεδομένα του 2016 χρησιμοποιήθηκαν ως αξιολόγηση. Στόχος μας σε αυτό το πείραμα είναι να εξεταστεί εάν, με την εξισορρόπηση του εκπαιδευτικού σετ του 2015 και αξιολογώντας το με τα νέα δεδομένα που συλλέχθηκαν από το 2016, θα παραχθούν καλύτερα αποτελέσματα. Η Εικόνα 4.7 παρουσιάζει το f1 score που όλοι οι αλγόριθμοι ταξινόμησης



Εικόνα 4.5: Κατανομή τιμών της κλάσης ταξινόμησης, για την πειραματική μέθοδο 1



Εικόνα 4.6: f1 score για όλους τους αλγόριθμους για κάθε τιμή ταξινόμησης, για την πειραματική μέθοδο 1



Εικόνα 4.7: f1 score για όλους τους αλγόριθμους για κάθε τιμή ταξινόμησης, για την πειραματική μέθοδο 2

έχουν επιτύχει για κάθε μία τιμή ταξινόμησης. Οι αλγόριθμοι J48, SMO, AdaBoostM1, IBk και Multilayer perceptron παράγουν συγκρίσιμες επιδόσεις για την τιμή "≥ 7", ενώ ο RandomForest παρουσιάζει σημαντικά χαμηλότερες επιδόσεις.

Στην τιμή "0 to 4", οι αλγόριθμοι J48, RandomForest και AdaBoostM1 παράγουν συγκρίσιμες επιδόσεις με τον αλγόριθμο SMO να είναι λίγο χαμηλότερα και τους αλγορίθμους IBk και

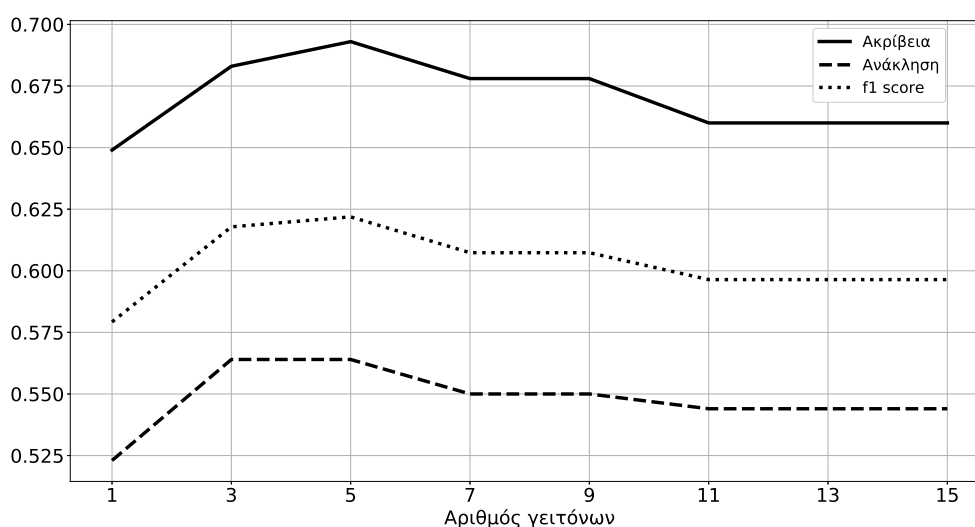
Multilayer perceptron να έχουν σημαντικά χαμηλότερες επιδόσεις. Τα αποτελέσματα της τιμής "5 to 6" είναι σαφώς χαμηλότερα από τις άλλες δύο τιμές με τους αλγορίθμους SMO, IBk και Multilayer perceptron να παρουσιάζουν τα καλύτερα αποτελέσματα.

Ομοίως, η Εικόνα 4.8 παρουσιάζει την ακρίβεια, ανάκληση και f1 score που έχει επιτύχει ο αλγόριθμος IBk. Το καλύτερο f1 score επιτυγχάνετε για πέντε πλησιέστερους γείτονες. Τέλος, η Εικόνα 4.9 παρουσιάζει την ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο Multilayer perceptron, με μία κρυφή στρώση. Το καλύτερο f1 score επιτεύχθηκε για 2 κόμβους.

Μέθοδος 3 Ακολουθώντας το παράδειγμα της προηγούμενης πειραματικής μεθόδου, εφαρμόστηκε η τεχνική SMOTE στα δεδομένα του 2015 ενώ τα δεδομένα του 2016 χρησιμοποιήθηκαν για αξιολόγηση. Το διάγραμμα χαρακτηριστικών στην παρούσα πειραματική μέθοδο μειώθηκε σε τρία χαρακτηριστικά, αναφορικά

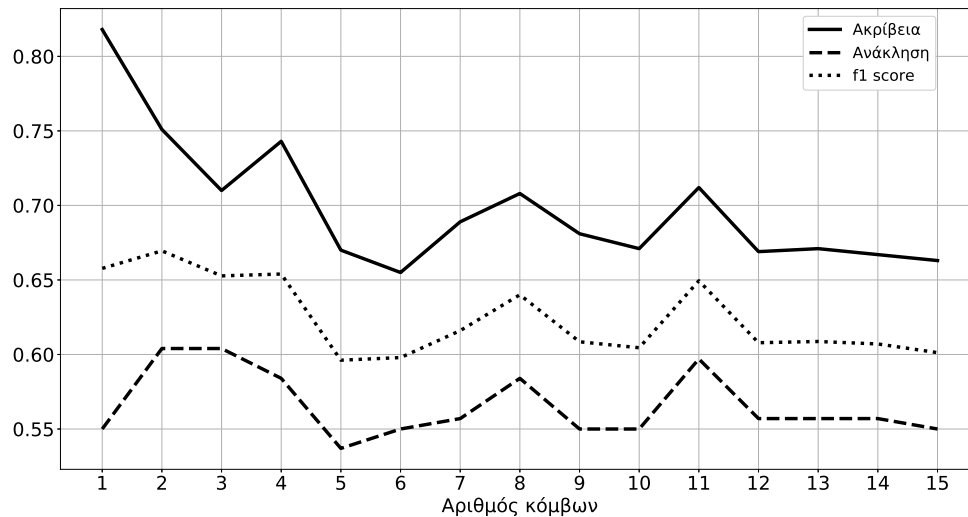
1. Τελευταία μέτρηση δακοπαγίδας
2. Degree Days, που θα έχουν συλλεχθεί μέχρι την επόμενη μέτρηση
3. Επόμενη μέτρηση δακοπαγίδας

εφόσον το χαρακτηριστικό Degree Days προκύπτει από την θερμοκρασία, αποφασίστηκε να αφαιρεθούν όλα τα χαρακτηριστικά που είναι σχετικά με την θερμοκρασία. Έτσι, σε αυτό το πείραμα, εξετάζεται αν εφαρμόζοντας την διαδικασία της Μεθόδου 2 στο νέο διάγραμμα χαρακτηριστικών θα παράξει καλύτερα αποτελέσματα.

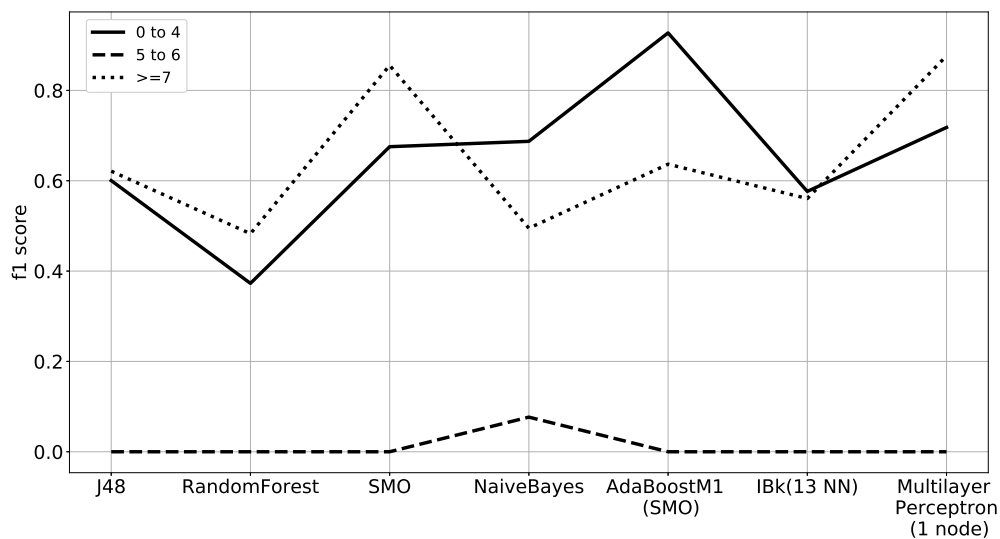


Εικόνα 4.8: Ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο IBk, για την πειραματική μέθοδο 2

Ανάπτυξη στοχαστικών αλγορίθμων και μοντέλων μηχανικής μάθησης για την εξέλιξη πληθυσμών: Εφαρμογή στο δάκο της Ελιάς



Εικόνα 4.9: Ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο Multilayer Perceptron, για την πειραματική μέθοδο 2



Εικόνα 4.10: f1 score για όλους τους αλγόριθμους για κάθε τιμή ταξινόμησης, για την πειραματική μέθοδο 3

Η Εικόνα 4.10 παρουσιάζει το f1 score των αλγορίθμων J48, RandomForest, SMO, Naïve Bayes, AdaBoostM1(SMO), IbK και Multilayer Perceptron. Για την τιμή " ≥ 7 " οι αλγόριθμοι SMO και Multilayer Perceptron έχουν την καλύτερη απόδοση, με τους αλγόριθμους J48, AdaBoostM1(SMO) και IBk να ακολουθούν με χαμηλότερα αλλά συγκρίσιμα αποτελέσματα. Ενώ, οι αλγόριθμοι RandomForest και Naïve Bayes έχουν την χαμηλότερη απόδοση. Την καλύτερη απόδοση για την τιμή "0 to 4" πετυχαίνει ο αλγόριθμος AdaBoostM1(SMO), με τους αλγόριθμους J48, SMO,

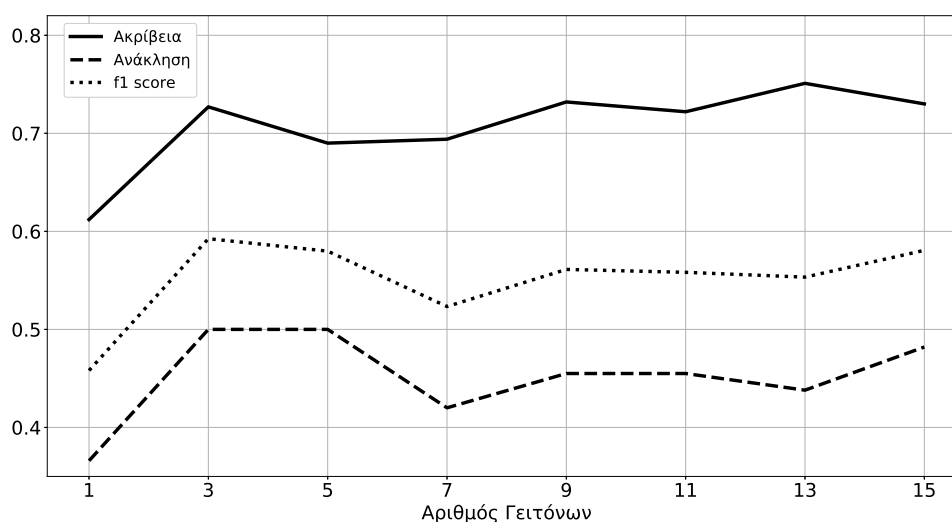
Naïve Bayes, IBk και Multilayer Perceptron να ακολουθούν, ενώ ο αλγόριθμος RandomForest έχει την χειρότερη απόδοση. Στον αντίποδα, για την τιμή "5 to 6" υπήρχαν λίγα δεδομένα διαθέσιμα για να παραχθούν χρήσιμα αποτελέσματα, παρά την ελάχιστην αυξημένη απόδοση του αλγορίθμου Naïve Bayes.

Η Εικόνα 4.11 παρουσιάζει την ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο IBk. Το καλύτερο f1 score επιτυγχάνεται με τρεις πλησιέστερους γείτονες. Τέλος, η Εικόνα 4.12 παρουσιάζει την ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο Multilayer Perceptron με ένα κρύφο επίπεδο νευρώνων. Το καλύτερο f1 score επιτυγχάνεται με ένα κόμβο στο κρυφό επίπεδο

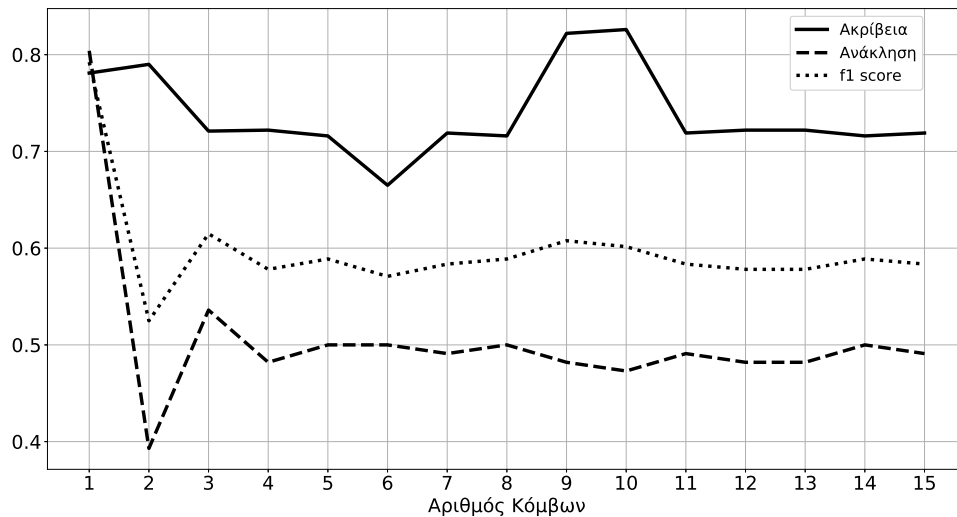
Μέθοδος 4 Σε αυτήν την πειραματική μέθοδο χρησιμοποιήθηκε το διάνυσμα χαρακτηριστικών της Μεθόδου 3, με μείωση των τιμών, των χαρακτηριστικών σχετικά με τις μετρήσεις των δακοπαγίδων, από τρεις τιμές σε δύο στα δεδομένα του 2015 και 2016. Σε αυτό το πείραμα γίνεται προσπάθεια να ισορροπηθεί η κλάση ταξινόμησης συγχωνεύοντας δύο τιμές σε μία. Η Εικόνα 4.13 παρουσιάζει το f1 score των αλγορίθμων J48, RandomForest, SMO, Naïve Bayes, IBk και Multilayer perceptron που επιτυγχάνεται για κάθε τιμή της κλάσης ταξινόμησης. Παρόλο που δεν ήταν σημαντικά καλύτερη, ο αλγόριθμος IBk είχε την καλύτερη απόδοση από τους υπόλοιπους αλγορίθμους με τον αλγόριθμο Multilayer perceptron να είναι πολύ κοντά και στις δύο τιμές, "0 έως 4" και "≥ 5".

Παρομοίως, οι αλγόριθμοι J48, SMO και Naïve Bayes είχαν ελαφρώς χαμηλότερη απόδοση, αντίθετα με τον αλγόριθμο RandomForest να έχει την χειρότερη απόδοση και στις δύο τιμές.

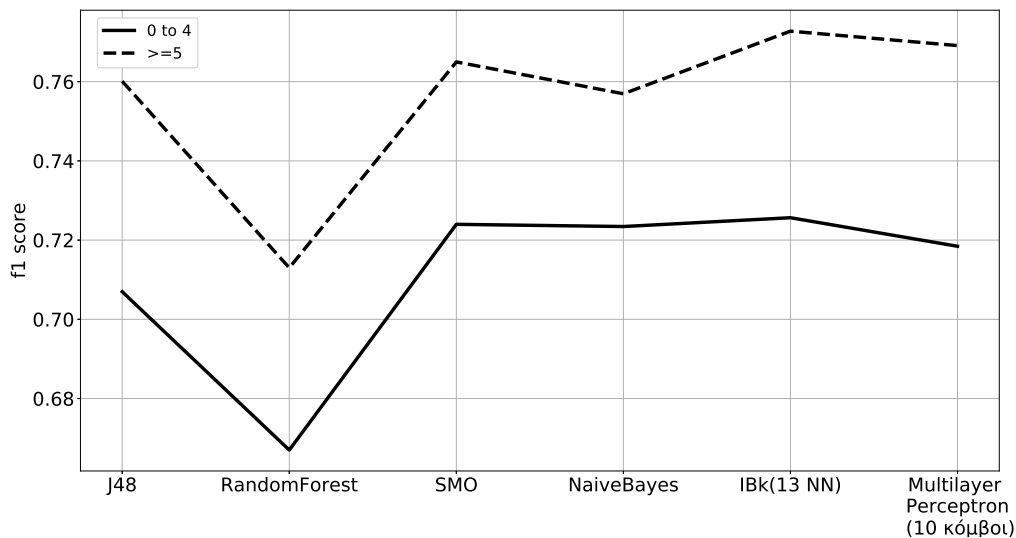
Η Εικόνα 4.16 απεικονίζει τα συλλογικά αποτελέσματα. Οι υψηλότερες επιδόσεις της Μεθόδου 1 είναι παραπλανητικές, καθώς η συγκεκριμένη μέθοδος έχει ένα εξαιρετικά ανισόρροπο



Εικόνα 4.11: Ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο IBk, για την πειραματική μέθοδο 3



Εικόνα 4.12: Ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο Multilayer Perceptron, για την πειραματική μέθοδο 3



Εικόνα 4.13: f1 score για όλους τους αλγόριθμους για κάθε τιμή ταξινόμησης, για την πειραματική μέθοδο 4

σύνολο δεδομένων και οι αναφερόμενες μετρήσεις αποτελούν τον μέσο όρο τιμών σε όλες τις τιμές της κλάσης ταξινόμησης. Οι Μέθοδοι 2, 3 και 4 αντιμετωπίζουν ένα ισορροπημένο σύνολο δεδομένων, παρόλο που οι μέσες τιμές απόδοσης για όλες τις τιμές της κλάσης ταξινόμησης είναι χαμηλότερες. Μεταξύ των τριών προαναφερόμενων μεθόδων η Μέθοδος 4 αποδίδει καλύτερα και στις τρεις μετρικές απόδοσης.

4.1.3.3 Συμπεράσματα

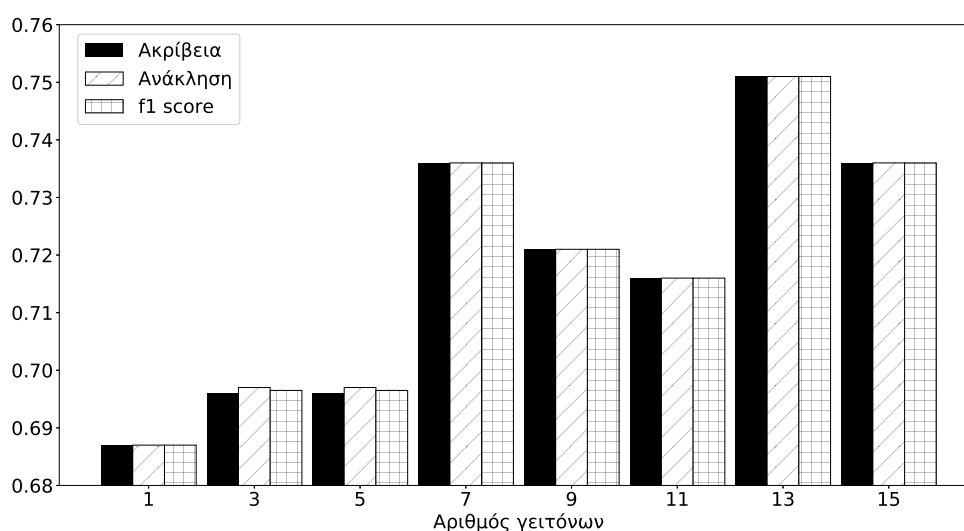
Τα αποτελέσματα που παράχθηκαν από τα πειράματα που πραγματοποιήθηκαν είναι πολύ υποσχόμενα δείχνοντας παράλληλα την ανωτερότητα της μεθόδου 4 με την εισαγωγή του χαρακτηριστικού DD το οποίο είναι άρηκτα συνδεδεμένο με τον βιολογικό κύκλο του δάκου.

Αυτό μπορεί να εξηγηθεί βάση της εξίσωσης 4.1. Η παρουσία των τιμών άνω και κάτω κατωφλιού θερμοκρασίας ανάπτυξης, T_L , T_U μπορεί να οδηγήσει σε μια δραματική αλλαγή στα στάδια ανάπτυξης του δάκου, ακόμη και αν οι διαφορές στις θερμοκρασίες είναι μικρές. Ως αποτέλεσμα, ένας αλγόριθμος μηχανικής μάθησης που βασίζεται μόνο σε απόλυτες τιμές της θερμοκρασίας δεν θα μπορεί να προβλέψει σωστά την έξαρση του δάκου στον ελαιώνα.

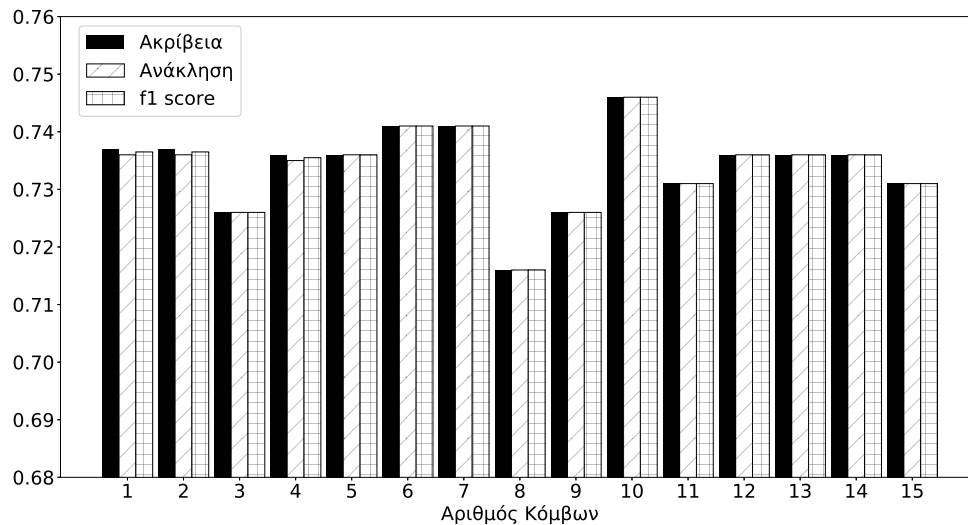
Σε μελλοντική έρευνα θα μπορούσε να λάφθούν υπόψη περισσότερες περιβαλλοντικές παράμετροι όπως η σχετική υγρασία, η ποσότητα φωτός που εκτίθενται οι δάκοι και χαρακτηριστικά διάχυσης. Τέλος, η παροχή περισσότερων εκπαιδευτικών περιπτώσεων στους αλγόριθμους μηχανικής μάθησης θα βοηθούσε στην παραγωγή καλύτερων αποτελεσμάτων.

4.2 Ανάπτυξη μοντέλου για την αυτόματη αναγνώριση δάκου σε εικόνες

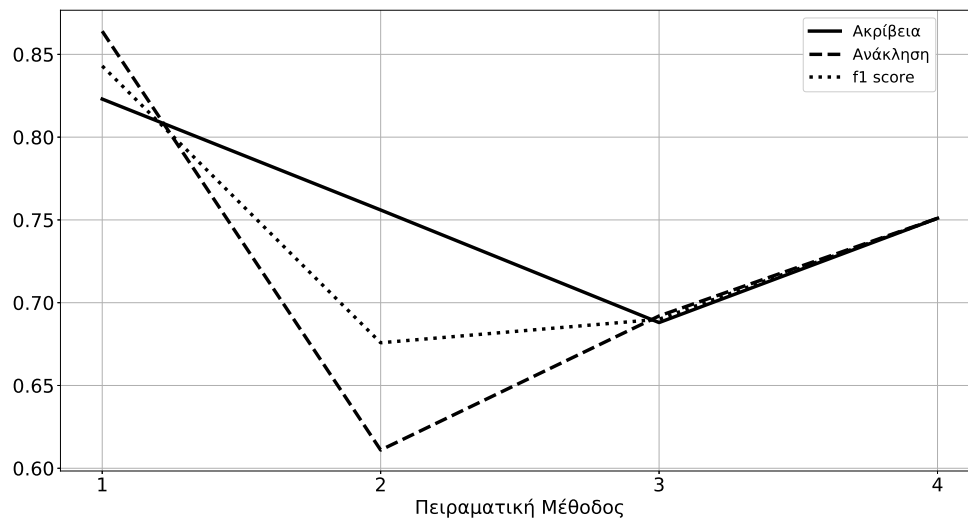
Όπως αναφέρθηκε στην ενότητα 2.2.4 μια από τις μεθόδους ελέγχου του πληθυσμού του δάκου είναι ο ψεκασμός του ελαιώνα. Για να πραγματοποιηθεί όμως ο ψεκασμός του ελαιώνα πρέπει να διαπιστωθεί ότι υπάρχει προσβολή αυτού από τον δάκο της ελιάς. Ο έλεγχος προσβολής ενός ελαιώνα γίνεται κυρίως με χειροκίνητες μεθόδους που περιλαμβάνουν παγίδες, ενώ μία από τις βασικές προϋποθέσεις για την επαλήθευση έξαρσης του πληθυσμού έγκειται στη μέτρηση των δάκων που συλλέγονται σε μια παγίδα για ένα χρονικό διάστημα [Shaked et al.,



Εικόνα 4.14: Ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο IBk, για την πειραματική μέθοδο 4



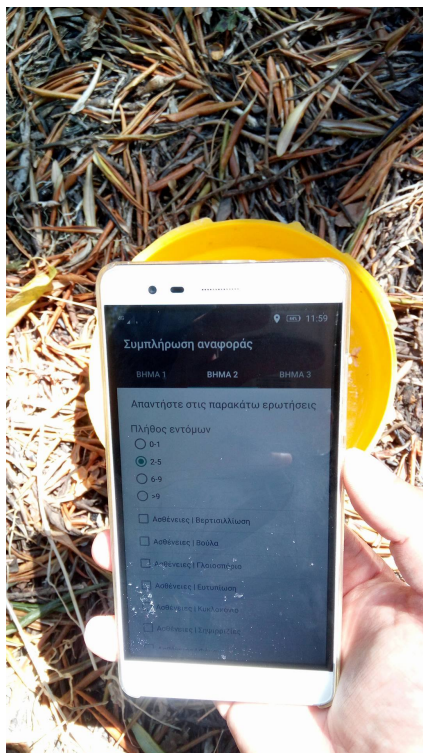
Εικόνα 4.15: Ακρίβεια, ανάκληση και f1 score για τον αλγόριθμο Multilayer Perceptron, για την πειραματική μέθοδο 4



Εικόνα 4.16: Σύγκριση Ακρίβειας, ανάκλησης και f1 score μεταξύ των καλύτερων αποτελεσμάτων κάθε μεθόδου

2017]. Αυτή η διαδικασία απαιτεί συχνούς και χρονοβόρους μη αυτόματους ελέγχους.

Για τον λόγο αυτό έχουν αναπτυχθεί προηγμένες παγίδες, ή έξυπνες παγίδες, οι οποίες διαθέτουν μια κάμερα που φωτογραφίζει τους δάκους που συλλέγονται από την παγίδα και στη συνέχεια εξετάζονται από τα ενδιαφερόμενα μέρη [Doitsidis et al., 2017; Alorda et al., 2015; Wang et al., 2016; Philimis et al., 2013; Tirelli et al., 2011]. Με βάση τις εικόνες των δάκων



Εικόνα 4.17: Συλλογή δεδομένων με την εφαρμογή e-Olive.

που συλλέγονται στις παγίδες, τα ενδιαφερόμενα μέρη όπως οι αγρότες, οι εντομολόγοι, οι γεωπόνοι κ.λπ. μπορούν να αναγνωρίσουν και να μετρήσουν τους συλλεγόμενους δάκους, να προσαρμόσουν τη συχνότητα της εξέτασης της παγίδας, ενώ ταυτόχρονα να ελαχιστοποιήσουν τον χρόνο εξέτασης και τη σχετική δυσκολία.

Παρά τις προαναφερθείσες προόδους των έξυπνων παγίδων, για να μπορέσουμε να εξαγάγουμε γνώσεις από τα προαναφερθέντα δεδομένα έξυπνης παγίδας, οι υπάρχουσες μεθοδολογίες πρέπει να είναι σε θέση να συγκρίνουν τα αποτελέσματα. Έτσι, η χρήση ενός κοινού συνόλου δεδομένων παρατήρησης παγίδας είναι απαραίτητη για τον έλεγχο της αποτελεσματικότητας και της αποτελεσματικότητας των μεθόδων, παρέχοντας παράλληλα αναφορά για σύγκριση νέων και υπαρχουσών μεθόδων για να δείξει πρόοδο, όπως συμβαίνει με τα περισσότερα επιστημονικά σύνολα δεδομένων [Uhlir, 2015; Mayernik et al., 2016]. Από ό, τι γνωρίζουμε, οι υπάρχουσες ερευνητικές εργασίες στην αυτοματοποιημένη αναγνώριση του δάκου δεν λειτουργούν στα ίδια δεδομένα και έτσι τα αποτελέσματα που παρουσιάζονται σε αυτά δεν είναι εύκολα συγκρίσιμα.

Η συλλογή δεδομένων από τις δακόπαγίδες είναι μια μακρά διαδικασία που απαιτεί πολλές τοποθεσίες ελαιώνων, συχνή φυσική προσοχή στις παγίδες, δευτερεύουσες εντομολογικές γνώσεις, κατάλληλο υλικό, ενώ είναι δυνατό μόνο για τη χρονολογική περίοδο που οι δάκοι είναι ενεργοί [Potamitis et al., 2017a]. Όλοι αυτοί οι παράγοντες καθιστούν τη συλλογή δεδομένων δακόπαγίδων μάλλον δύσκολη. Έτσι, η αξιολόγηση των νέων μεθόδων παρεμποδίζεται από την έλλειψη εύκολα προσβάσιμων δεδομένων για τη δοκιμή των μεθόδων.

Ανάπτυξη στοχαστικών αλγορίθμων και μοντέλων μηχανικής μάθησης για την εξέλιξη πληθυσμών: Εφαρμογή στο δάκο της Ελιάς

Για την αντιμετώπιση των προαναφερθεισών απαιτήσεων, δημιουργήθηκε το Dacus Image Recognition Toolkit (*DIRT*), μια συλλογή από διαθέσιμα στο κοινό δεδομένα, δείγματα κώδικα προγραμματισμού και διαδικτυακές υπηρεσίες που επικεντρώνονται στην υποστήριξη της έρευνας με στόχο τη διαχείριση της ελιάς.

4.2.1 To toolkit

Τα επιμέρους στοιχεία από τα οποία αποτελείται το toolkit είναι τα ακόλουθα:

DIRT online detector

Please provide a picture of the contents of a McPhail trap and our neural network will try to detect all olive fruit flies for you.

Upload a picture (max. 2 Megabytes) of a McPhail trap using the form below.

Choose File api_test.jpg

Detect

Detected 28 olive fruit flies

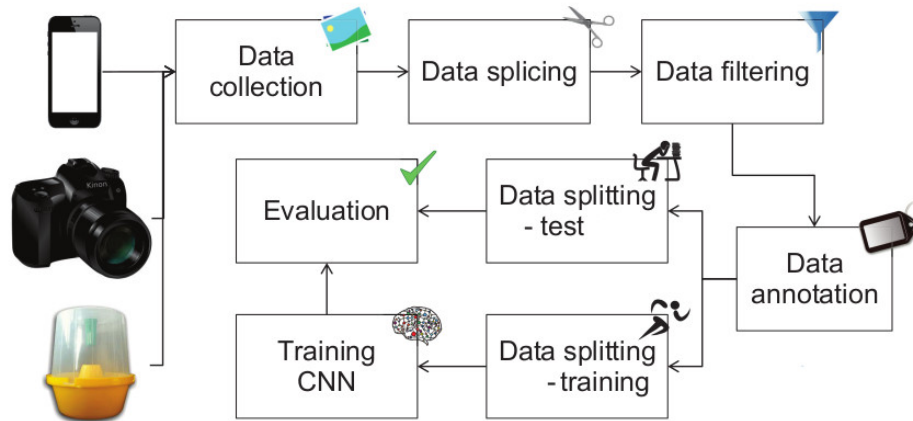


Εικόνα 4.18: Διεπαφή σε μορφή ιστοσελίδας του *DIRT*

- Σύνολο δεδομένων.
- Δείγματα κώδικα, που θα επιτρέψουν στους ενδιαφερόμενους ερευνητές να επιταχύνουν τη χρήση των περιεχομένων του *DIRT* καθώς και να τους καθοδηγήσουν σε κάποιο στοιχειώδη πειραματισμό.
- API. Το *DIRT* περιλαμβάνει επίσης ένα δημόσιο API που απαντά σε ερωτήματα για αναγνώριση δάκου σε εικόνες που παρέχονται από τον χρήστη. Με αυτόν τον τρόπο, η πολυπλοκότητα της αναγνώρισης δάκου/ων στις εικόνες μετριάζεται από τους χρήστες που χρειάζεται μόνο να διαχειριστούν την αλληλεπίδραση με το API του *DIRT*, το οποίο αντιμετωπίζεται επίσης από μια απλοποιημένη διεπαφή σε μορφή ιστοσελίδας² (Εικόνα 4.18).

²<https://olive.di.ionio.gr/DIRT/>

4.2.1.1 Σύνολο δεδομένων



Εικόνα 4.19: Διάγραμμα ροής δημιουργίας και πειραματισμού με το *DIRT*

Τα δεδομένα του *DIRT* αποτελούνται από εικόνες, η πλειονότητα των οποίων απεικονίζει δάκους μέσα σε παγίδες McPhail, που συλλέχθηκαν από το 2015 έως το 2017 σε διάφορες τοποθεσίες της Κέρκυρας. Οι εικόνες αποκτήθηκαν κυρίως μέσω της smartphone εφαρμογής e-Olive³ (Εικόνα 4.17), η οποία επιτρέπει στους χρήστες να υποβάλλουν αναφορές σχετικά με τις μετρήσεις του δάκου και να ανεβάζουν εικόνες που τραβήχτηκαν από την παγίδα στο πεδίο. Καθώς η συλλογή εικόνων έχει πραγματοποιηθεί χρησιμοποιώντας μια ποικιλία υλικού (smartphone και tablet που εκτελούν την εφαρμογή e-Olive, φωτογραφικές μηχανές διαθέσιμες στο πεδίο κατά την επιθεώρηση παγίδας, κ.λπ.) οι εικόνες του συνόλου δεδομένων δεν είναι τυποποιημένες. Η Εικόνα 4.20 δείχνει την κατανομή των εικόνων όσον αφορά τις διαστάσεις τους.

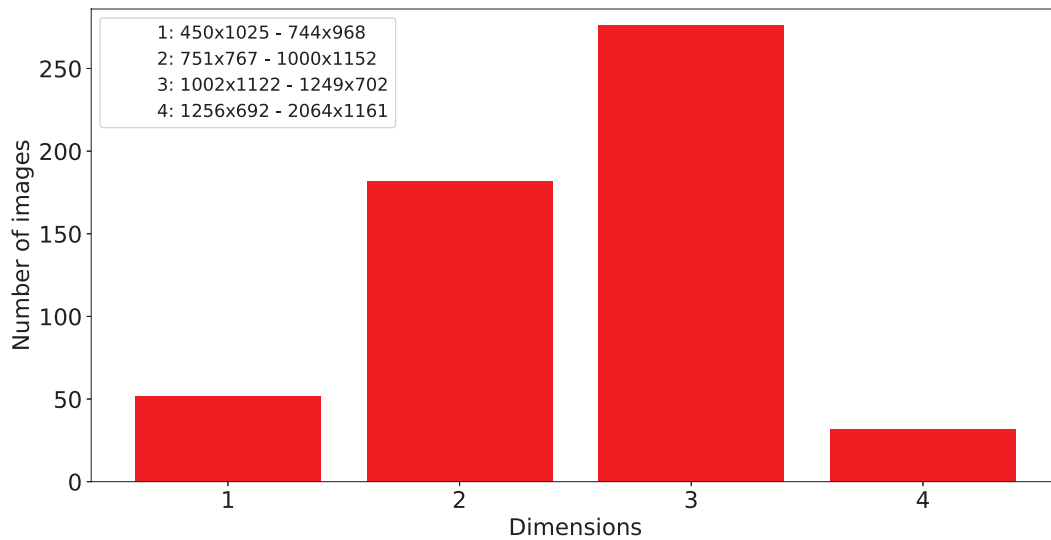
Το αρχικό σύνολο δεδομένων αποτελείται από 336 εικόνες, αλλά μετά την απόρριψη εικόνων που είτε ήταν πολύ θολές για να ξεχωρίσουν οι δάκοι από άλλα έντομα είτε δεν υπήρχαν δάκοι, το μέγεθος του συνόλου δεδομένων μειώθηκε σε 202 εικόνες. Επιπλέον, δεδομένου ότι η εκπαίδευση σε αυτό το σύνολο δεδομένων για 50.000 βήματα απαιτούσε ανώτερο από το συνήθως διαθέσιμο υλικό, λόγω των απαιτήσεων μνήμης που σχετίζονται με το μεγάλο μέγεθος των διαθέσιμων φωτογραφιών, αποφασίσαμε να κόψουμε κάθε εικόνα σε τέσσερα ή περισσότερα μέρη. Έτσι, μετά την απόρριψη τμημάτων εικόνας που δεν απεικόνιζαν κανένα δάκο, το τελικό σύνολο δεδομένων περιλαμβάνει 542 εικόνες, ένα δείγμα των οποίων φαίνεται στο σχήμα 4.21. Από αυτές τις εικόνες, 486 επιλέχθηκαν τυχαία για εκπαίδευση, ενώ οι υπόλοιπες, επίσης επιλέχθηκαν τυχαία, 56 εικόνες χρησιμοποιήθηκαν για αξιολόγηση στα πειράματά μας. Η Εικόνα 4.22 παρουσιάζει το ιστόγραμμα των χειροκίνητα σχολιασμένων δάκων σε εικόνες, πριν και μετά τη διαδικασία τεμαχισμού των εικόνων, όπως προαναφέρθηκε.

Η προσθήκη ετικέτας στο σύνολο των δεδομένων μας έγινε με το εργαλείο ετικετοποίησης LabelImg⁴ που επιτρέπει στο χρήστη να σχεδιάζει χειροκίνητα κουτιά γύρω από αντικείμενα σε εικόνες και να συσχετίζει κάθε πλαίσιο με μια ετικέτα, όπως φαίνεται στην Εικόνα 4.23. Είναι

³<https://play.google.com/store/apps/details?id=gr.cwa.eolive&hl=el&gl=US>

⁴<https://github.com/tzutalin/labelImg>

Ανάπτυξη στοχαστικών αλγορίθμων και μοντέλων μηχανικής μάθησης για την εξέλιξη πληθυσμών: Εφαρμογή στο δάκο της Ελιάς

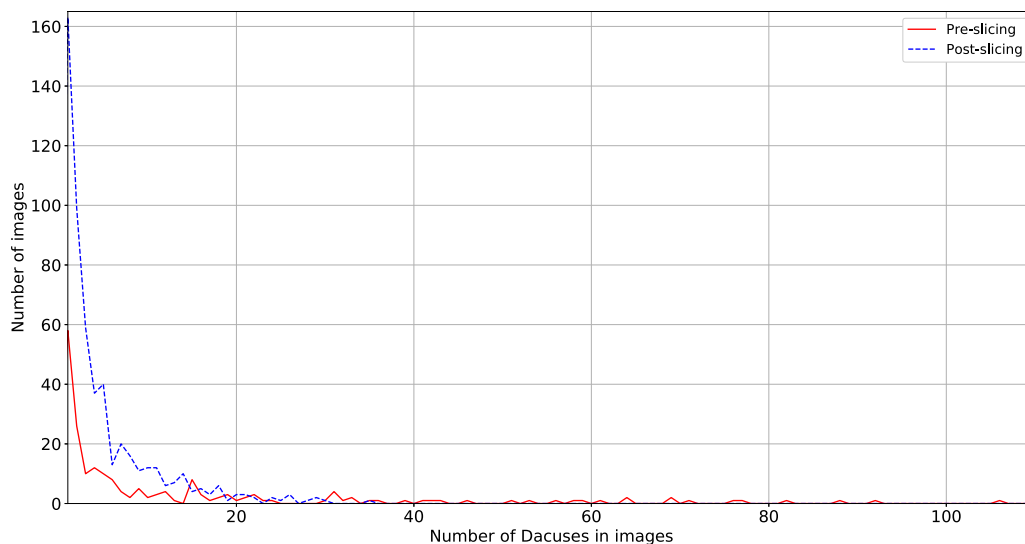


Εικόνα 4.20: Κατανομή των διαστάσεων εικόνων του *DIRT*



Εικόνα 4.21: Δείγμα εικόνων από το σύνολο δεδομένων.

σημαντικό να σημειωθεί, ότι για εικόνες που απεικονίζουν παγίδες McPhail, ετικετοποιήσαμε ως επί το πλείστον δάκους που επιπλέουν στο υγρό διάλυμα εκτός από τους διακριτούς βυθισμένους δάκους, καθώς όταν βυθίζονται, οι δάκοι τείνουν να δημιουργούν συστάδες που κάνουν την εργασία ταυτοποίησης πολύ δύσκολη. Δεσμευμένοι με αυτόν τον περιορισμό ετικετοποιήσαμε σε σύνολο 542 εικόνες 2672 δάκους.



Εικόνα 4.22: Κατανομή εικόνων ανα αριθμό δάκων μέσα σε μία εικόνα.



Εικόνα 4.23: Δείγμα εικόνων από το σύνολο δεδομένων με ετικετοποίηση

4.2.2 Πειραματική αξιολόγηση

4.2.2.1 Πειραματική διάταξη

Για τα πειράματά μας επιλέξαμε το Tensorflow Object Detection API ⁵ που είναι ένα πλαίσιο ανοιχτού κώδικα που βασίζεται στο Tensorflow ⁶, ένα πλαίσιο μηχανικής μάθησης ανοιχτού κώδικα. Το API ανίχνευσης αντικειμένων Tensorflow παρέχει αρκετά προ-εκπαιδευμένα μοντέλα για χρήση σε πειράματα. Τα μοντέλα ανίχνευσης που παρέχονται ήταν προ-εκπαιδευ-

⁵https://github.com/tensorflow/models/tree/master/research/object_detection

⁶<https://www.tensorflow.org/>

μένα στα COCO⁷, KITTI⁸ και Open Images⁹ σύνολα δεδομένων. Όλες οι εκπαιδεύσεις εκτελέστηκαν για 100.000 βήματα, με μέγεθος παρτίδας ένα. Έτσι, η εκπαίδευση διεξήχθη για 184 εποχές. Η διαμόρφωση υλικού όπου πραγματοποιήθηκαν όλα τα πειράματα φαίνεται στον Πίνακα 4.1.

Table 4.1: Hardware configuration.

CPU	Intel Core i7 920 @ 2.67 GHz
GPU	NVIDIA TITAN Xp 11 GB
RAM	16 GB

4.2.2.2 Αποτελέσματα

Για να επαληθευτεί η χρησιμότητα του προτεινόμενου συνόλου δεδομένων, διεξήχθη μια ποικιλία πειραμάτων, που αφορούσαν κυρίως την ικανότητα αυτόματου εντοπισμού των δάκων στις εικόνες χρησιμοποιώντας ως χωρική αλήθεια επισημασμένες εικόνες με δάκους.

Πρώτον, πραγματοποιήθηκε εκπαίδευση στα ακόλουθα προ-εκπαιδευμένα μοντέλα και αυτό με την καλύτερη απόδοση επιλέχθηκε για τη διεξαγωγή περαιτέρω πειραμάτων.

- faster_rcnn_inception_v2_coco
- faster_rcnn_resnet50_coco
- faster_rcnn_resnet50_lowproposals_coco
- rfcn_resnet101_coco
- faster_rcnn_resnet101_coco
- faster_rcnn_inception_resnet_v2_atrous_coco

Ο πίνακας 4.2 παρουσιάζει τη συνολική απώλεια και τον συνολικό χρόνο ολοκλήρωσης της εκπαίδευσης στα καθορισμένα βήματα για καθένα από τα προαναφερθέντα μοντέλα, μετά την εκπαίδευση στο σύνολο δεδομένων *DIRT*.

Η Εικόνα 4.24 παρουσιάζει την απόδοση όλων των μοντέλων που έχουν εκπαιδευτεί στο σύνολο δεδομένων μας. Το μοντέλο που παρουσίασε τη χειρότερη απόδοση ήταν το μοντέλο 4 με mean Average Precision (mAP) 57,42%. Τα μοντέλα ανίχνευσης 3, 5 και 6 απέδωσαν σχετικά καλά, κυμαινόμενα μεταξύ 68% και 80%. Ωστόσο, τα μοντέλα 1 και 2 ξεπέρασαν όλα τα υπόλοιπα με τιμή χάρτη 91,52% και 90,03% αντίστοιχα. Αν και η διαφορά στην απόδοση είναι μικρή (1,49 %) μεταξύ των δύο, επιλέξαμε το μοντέλο 1 για τα υπόλοιπα πειράματά μας, καθώς ο χρόνος προπόνησής του είναι σχεδόν τέσσερις φορές πιο γρήγορος από το μοντέλο 2 (βλ. Πίνακα 4.2).

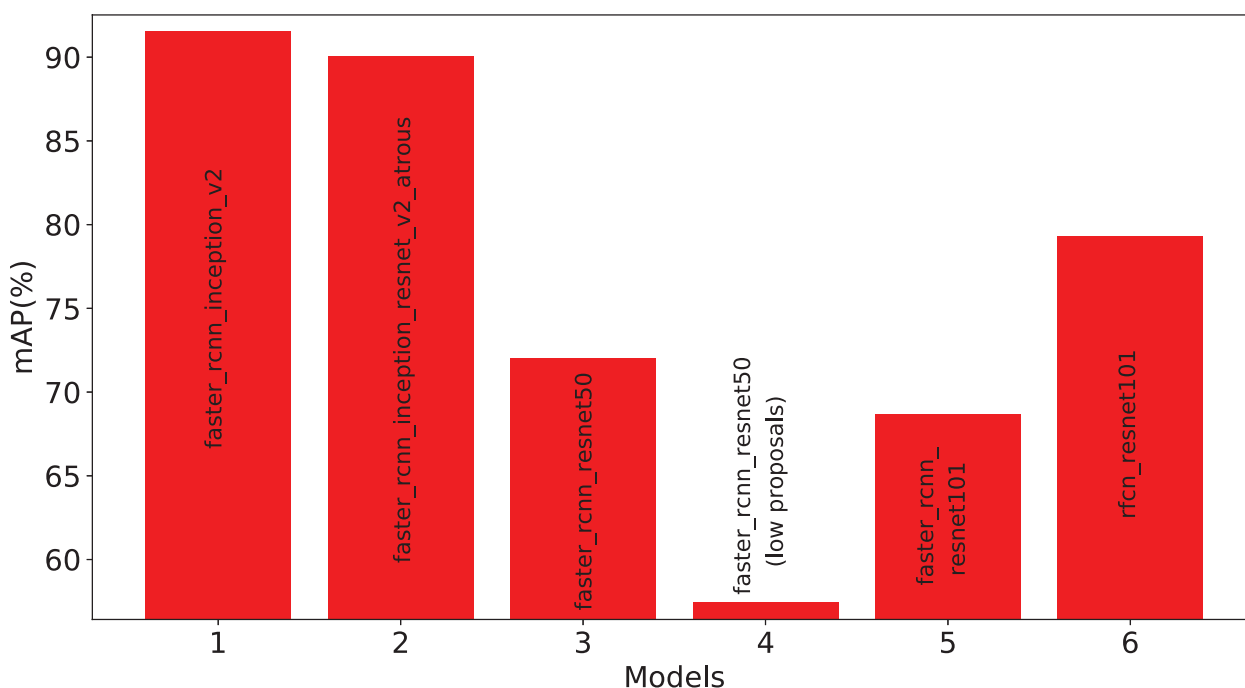
⁷<http://cocodataset.org/>

⁸<http://www.cvlibs.net/datasets/kitti/>

⁹<https://storage.googleapis.com/openimages/web/index.html>

Table 4.2: Στοιχεία εκπαίδευσης μοντέλων.

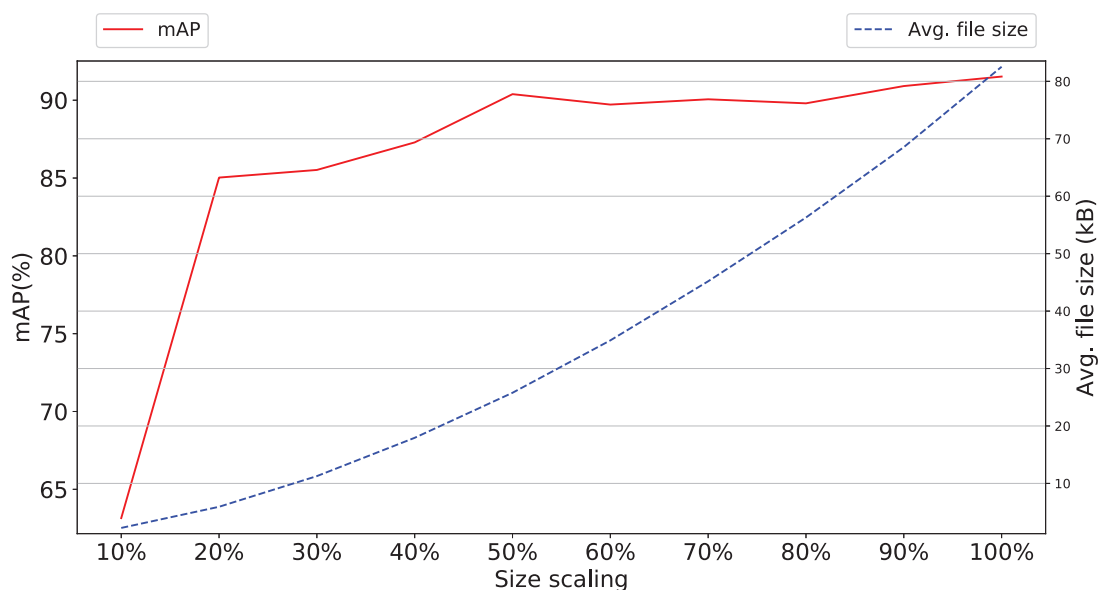
Index	Μοντέλο	Ολική απώλεια	Χρόνος Εκπαίδευσης
1	faster_rcnn_inception_v2	0.04886	5 h 35 m 51 s
2	faster_rcnn_inception_resnet_v2_atrous	0.1883	23 h 21 m 39 s
3	faster_rcnn_resnet50	0.1029	8 h 35 m 53 s
4	faster_rcnn_resnet50_low_proposals	0.5773	8 h 29 m 54 s
5	faster_rcnn_resnet101	0.1608	12 h 57 m 49 s
6	rfcn_resnet101	0.1349	13 h 45 m 51 s



Εικόνα 4.24: Σύγκριση απόδοσης μοντέλων μετά την εκπαίδευση στο σύνολο δεδομένων εικόνων με δάκους.

Αφού επιλέξαμε το μοντέλο με την καλύτερη απόδοση, διερευνήσαμε τον τρόπο με τον οποίο επηρεάζεται η απόδοση από την λεπτομέρεια της εικόνας, εκπαιδεύοντας έτσι σε εικόνες μικρότερων διαστάσεων από το αρχικό σύνολο δεδομένων. Αναλυτικά, εκπαιδεύσαμε το μοντέλο μας από το 10% μέχρι το 100% των αρχικών διαστάσεων με βήμα αύξησης 10%. Επιπλέον, το ίδιο πείραμα επαναλήφθηκε, μόνο αυτή τη φορά όλες οι εικόνες μετατράπηκαν στη γκριζα χρωματική κλίμακα για να επαληθευτεί η επίδραση του χρώματος στα αποτελέσματα που ελήφθησαν.

Η Εικόνα 4.25 δείχνει πώς αλλάζει η απόδοση του μοντέλου ανίχνευσης 1 όταν εκπαιδεύεται σε διαφορετικές διαστάσεις των εικόνων από το αρχικό σύνολο δεδομένων του δάκου. Για το 10% των αρχικών διαστάσεων, το μοντέλο έχει αρκετά χαμηλή απόδοση, με mAP 63,15%, σχετικά με τα επόμενα μεγέθη διαστάσεων. Για 20% και άνω, η ακρίβεια ανίχνευσης κυμαίνεται μεταξύ 85% και 91%. Είναι ενδιαφέρον, αρκετά, οι τιμές του mAP για διαστάσεις μεταξύ



Εικόνα 4.25: Απόδοση μοντέλου 1 για διαφορετικές διαστάσεις των εικόνων στο σύνολο δεδομένων.

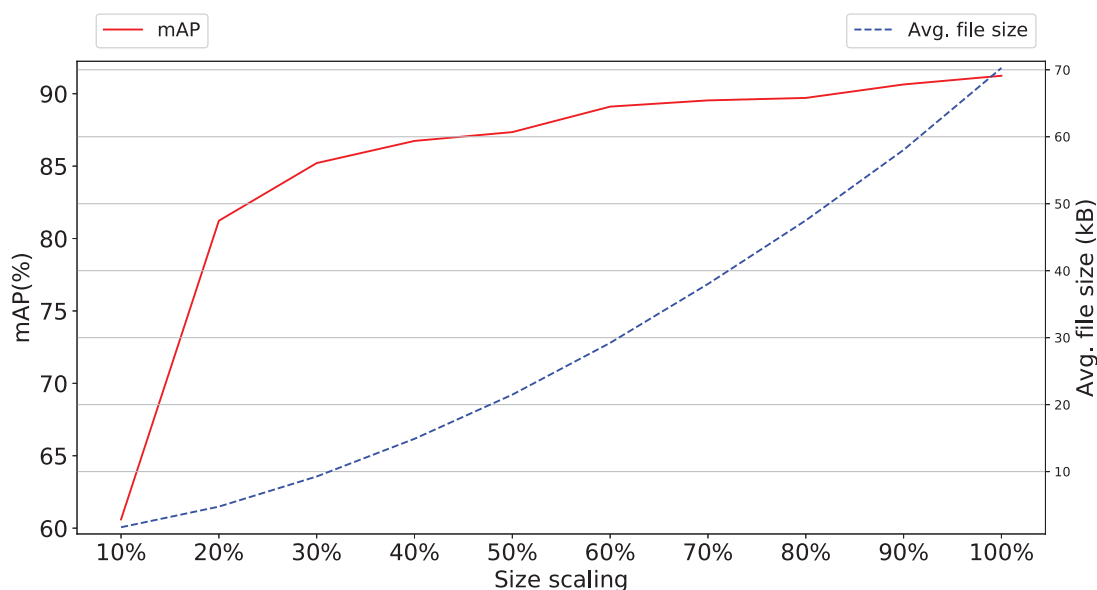
50% και 100% είναι σχεδόν σταθερές με μικρές διακυμάνσεις, ενώ η διαφορά μεταξύ των εικόνων πλήρους μεγέθους και των εικόνων που αλλάζουν μέγεθος στο μισό είναι 1,13%.

Ομοίως, η Εικόνα 4.26 παρουσιάζει την αλλαγή στην απόδοση της ανίχνευσης για το μοντέλο 1 για διαφορετικές διαστάσεις των εικόνων αλλά με μετατροπή στη γκρι χρωματική κλίμακα. Για άλλη μια φορά, για εικόνες μεγέθους 10% η ακρίβεια ανίχνευσης είναι χαμηλή (60,6%) σε σύγκριση με τα υπόλοιπα μεγέθη. Ενώ, μεταξύ 20% και 100% κλιμάκωση, ο δείκτης mAP κυμαίνεται μεταξύ 81% και 91%. Τέλος, μετά από κλιμάκωση 60%, η ακρίβεια ανίχνευσης είναι αρκετά σταθερή με μικρές διακυμάνσεις.

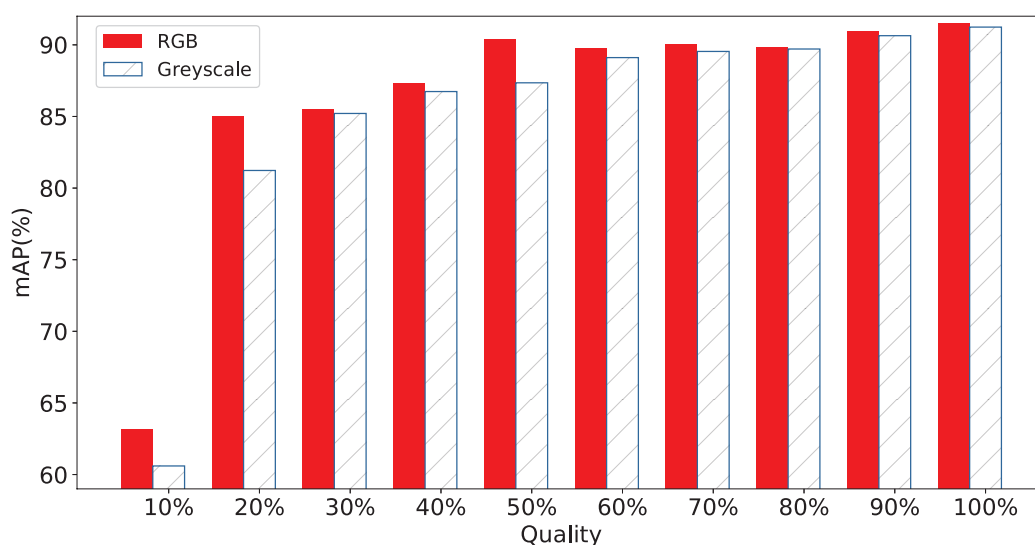
Στην Εικόνα 4.27, συγκρίνεται η ακρίβεια ανίχνευσης μεταξύ των συνόλων δεδομένων γκριζας κλίμακας και χρώματος (RGB) από τα δύο προηγούμενα πειράματα. Και τα δύο έχουν παρόμοια τάση στην αύξηση της ακρίβειας καθώς πλησιάζει το αρχικό μέγεθος των εικόνων. Ωστόσο, η μέση διαφορά στην ακρίβεια ανίχνευσης μεταξύ κλίμακας 10% και 50% είναι 2,05% υπέρ του συνόλου δεδομένων RGB. Από την άλλη πλευρά, από 60% και μετά, η ακρίβεια μεταξύ των συνόλων δεδομένων RGB και γκριζας κλίμακας φτάνει περίπου την ίδια με μια μέση διαφορά 0,354 % υπέρ του συνόλου δεδομένων RGB.

Στο επόμενο πείραμα, διερευνήθηκε πώς επηρεάζεται η απόδοση του επιλεγμένου μοντέλου σε σχέση με τον συνολικό αριθμό δάκων σε μια εικόνα. Επομένως, δημιουργήθηκαν τέσσερα νέα δευτερεύοντα σύνολα δεδομένων (βλ. Πίνακα 4.3) με βάση το αρχικό σύνολο δεδομένων, τόσο για τα εκπαιδευτικά όσο και για τα σετ δοκιμών. Συγκεκριμένα, δοκιμάσαμε την απόδοση σε εικόνες που περιείχαν 3, 7, 10 και 14 δάκους για να επαληθεύσουμε την ικανότητα του προτεινόμενου μοντέλου να διατηρήσει υψηλή απόδοση ανεξάρτητα από την πυκνότητα των δάκων που αναγνωρίζονται σε μια εικόνα. Οι τιμές των δάκων επιλέχθηκαν με βάση τη διαθεσιμότητα.

Τέλος, η απόδοση της προτεινόμενης μεθοδολογίας πειραματίστηκε σε σχέση με τον πραγματικό αριθμό των δάκων (δηλαδή, με μη αυτόματη καταμέτρηση) καθώς και τον συνολικό αριθμό δάκων σε κάθε εικόνα. Η Εικόνα 4.28 παρουσιάζει την ακρίβεια ανίχνευσης για τέσσερις ομάδες εικόνων, όπου κάθε ομάδα περιέχει αποκλειστικά εικόνες με τον ίδιο αριθμό δάκων. Σε όλους τους δοκιμασμένους αριθμούς δάκων (3, 7, 10, 14) η ακρίβεια ανίχνευσης δεν πέφτει



Εικόνα 4.26: Απόδοση μοντέλου 1 για διαφορετικές διάστασεις των εικόνων στο σύνολο δεδομένων, μετά την μετατροπή τους στη γκρι χρωματική κλίμακα.



Εικόνα 4.27: Σύγκριση απόδοσης μεταξύ εικόνων RGB και γκρι κλίμακας για διαφορετικές διαστάσεις των εικόνων.

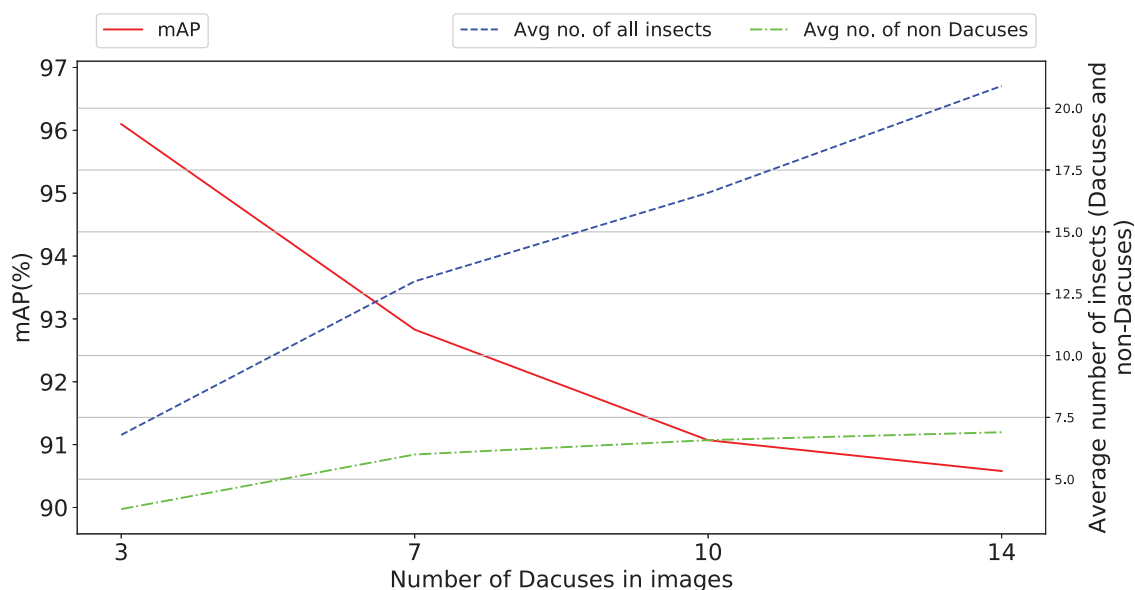
Table 4.3: Σύνολα δεδομένων με βάση τον αριθμό δάκων.

Αριθμός δάκων	Αριθμός Εικόνων Εκπαίδευσης	Αριθμός Εικόνων Αξιολόγησης
3	483	59
7	522	20
10	530	12
14	532	10

κάτω από 89%. Η περιορισμένη διακύμανση που υπάρχει (η ανίχνευση τριών δάκων παράγει την υψηλότερη τιμή mAP 96,1% με τη χαμηλότερη απόδοση ανίχνευσης 89,5% για δέκα δάκους) αποδίδεται στο μέγεθος του διαθέσιμου περιεχομένου εικόνας κάθε εξεταζόμενης ομάδας.

4.2.3 Συμπεράσματα

Τα πειραματικά αποτελέσματα στο μέγεθος των εικόνων που λαμβάνονται από την έξυπνη δακοπαγίδα, όπως φαίνεται στην Εικόνα 4.25, έδειξαν ότι οι υψηλής λεπτομέρειας φωτογραφίες πράγματι επηρεάζουν την απόδοση της προτεινόμενης μεθοδολογίας, αλλά η αναλογία της αύξησης της απόδοσης πέφτει απότομα μετά την απομάρκυνση του 80% των αρχικών πληροφοριών, ενώ η διαφορά μεταξύ της απομάρκυνσης 50–10% πληροφορίας είναι περίπου 1% και ως εκ τούτου σχεδόν αμελητέα, για ορισμένες εφαρμογές. Κατά συνέπεια, η ευρεία διαθεσιμότητα των φωτογραφικών μηχανών υψηλής ανάλυσης, αν και έχει αποδειχθεί ότι



Εικόνα 4.28: Σύγκριση ακριβείας ανίχνευσης για σύνολα δεδομένων με βάση τον αριθμό των δάκων.

αυξάνει την αποτελεσματικότητα του εντοπισμού των δάκων, οδηγεί σε μια ανταλλαγή μεταξύ οριακά υψηλότερης απόδοσης και αυξημένου όγκου δεδομένων που ενδεχομένως πρέπει να αποθηκευτούν τοπικά σε περιορισμένα μέσα αποθήκευσης ή να μεταφερθούν μέσω ασύρματων συνδέσεων (π.χ. GSM modem) ή σε ad-hoc δίκτυα, με αποτέλεσμα να επηρεάζουν το φορτίο του δικτύου.

Ομοίως με το προηγούμενο επιχείρημα, η πληροφορία του χρώματος των εικόνων που λαμβάνονται από την έξυπνη δακοπαγίδα, όπως φαίνεται στις Εικόνες 4.26 και 4.27, φαίνεται να είναι δευτερεύουσας σημασία όπως, τόσο στις εικόνες μικρότερων διαστάσεων γκρι κλίμακας όσο και σε εικόνες αρχικών διαστάσεων μετά τη μετατροπή τους σε γκρι κλίμακα, η επίδραση του χρώματος RGB στην απόδοση είναι σχεδόν αμελητέα. Αυτό υποστηρίζει περαιτέρω το προηγούμενο επιχείρημα σχετικά με τον μειωμένο ρόλο των εικόνων υψηλής ανάλυσης ακόμη και όταν αυτές είναι γκριζας κλίμακας και συνεπώς απαιτούν περίπου το ένα τρίτο των RGB ισοδύναμων εικόνων. Έτσι, η επιλογή μεταξύ φωτογραφικών μηχανών RGB ή γκριζας κλίμακας επανέρχεται στην προαναφερθείσα αντιστάθμιση μεταξύ της μικρής αύξησης της ακρίβειας αναγνώρισης του δάκου έναντι του όγκου δεδομένων.

Η ικανότητα της προτεινόμενης μεθόδου να διατηρεί υψηλή απόδοση ανεξάρτητα από τον αριθμό των δάκων που βρίσκονται μέσα στην παγίδα είναι πολύ σημαντικό ώστε να μπορεί να αντιμετωπίσει μια ποικιλία σεναρίων σε σχέση με τον σχεδιασμό των παγίδων, των γεωγραφικών και καιρικών χαρακτηριστικών του ελαιώνα, τις ποικιλίες ελιών κλπ. Η διαφορά σε απόδοση που φαίνεται στην Εικόνα 4.28 είναι περίπου 6,5% και συνεπώς απαιτεί περαιτέρω εξέταση, ακόμα κι αν για τις συγκεκριμένες παγίδες που χρησιμοποιούνται, όλες οι τιμές του αριθμού των δάκων που είναι ίσες ή μεγαλύτερες από επτά θεωρούνται ομοίως ότι αποτελούν σημαντική ένδειξη προσβολής του ελαιώνα. Συνολικά, μια γενική τάση είναι εμφανής για τις παραμέτρους που δοκιμάστηκαν: ο αριθμός των εντόμων (και κατ'επέκταση των δάκων) επηρεάζει αντίστροφα την ακρίβεια ανίχνευσης, ένα αποτέλεσμα που αποδίδεται στον αναλογικά υψηλότερο αριθμό εντόμων στην παγίδα όταν αυξάνεται ο αριθμός των δάκων μέσα σε αυτήν.

Τα αποτελέσματα έδειξαν ότι είναι δυνατόν να επιτευχθεί η αναγνώριση δάκου σε εικόνες με ακρίβεια απόδοσης (mAP) 91,52% με βάση δημόσια διαθέσιμα προ-εκπαιδευμένα μοντέλα και περαιτέρω εκπαιδευμένα σε δάκους. Εξάλλου, τα πειραματικά αποτελέσματα έδειξαν μια αντιστάθμιση τόσο στις διαστάσεις της εικόνας όσο και στην πληροφορία χρώματος (και οι δύο παράμετροι αυξάνουν τις λεπτομέρειες μιας εικόνας, το μέγεθος αρχείου και την πολυπλοκότητα των προσεγγίσεων) και ακρίβεια απόδοσης (mAP) που μπορεί να χρησιμοποιηθεί επιλεκτικά για την καλύτερη αντιμετώπιση των αναγκών κάθε σεναρίου χρήσης.

Μελλοντικά η αύξηση του συνόλου δεδομένων του *DIRT* θα βοηθούσε πιθανώς στην αύξηση της ακρίβειας εντοπισμού δάκων σε μια εικόνα. Επιπλέον, η επέκταση του μοντέλου αναγνώρισης με την δυνατότητα να αναγνωρίζει το φύλο των δάκων είναι μια κατεύθυνση μεγάλης σημασίας. Τέλος, το φαινόμενο που παρατηρήθηκε σε ορισμένες εικόνες με βυθισμένους δάκους ή συστάδες δάκων που ήταν δύσκολο να διακριθούν, θα μπορούσε να συμπεριληφθεί στην επέκταση του μοντέλου ώστε να αντιμετωπίζει τέτοιες καταστάσεις.

4.3 Ανάπτυξη μοντέλου για την αγνώριση μικροκλίματων

Όπως προαναφέρθηκε στην υπό-ενότητα 2.3 τα μικροκλίματα επηρεάζουν την φυσιολογία ενός οργανισμού και όπως απέδειξαν οι Kounatidis et al. [2008] οι εστίες πυκνότητας δακοπροβολής αλλάζουν αναλογώς την εποχή με βάση το υψόμετρο. Η υπάρχουσα έρευνα έχει δείξει ισχυρή εξάρτηση τόσο από τη θερμοκρασία όσο και από τη σχετική υγρασία της δυναμικής του πληθυσμού του δάκου, αλλά δεν έχει επικεντρωθεί στις τοπικές περιβάντολογικές/κλιματικές συνθήκες (μικροκλίματα) που σχετίζονται με τον κύκλο ζωής του παρασίτου. Είναι λοιπόν προφανές ότι η αναγνώριση των μικροκλίματος που σχετίζεται με τον δάκο ελιάς είναι μια πολύ σημαντική παράμετρος στην προσπάθεια αντιμετώπισης του. Οι επιπτώσεις της υψηλής ακρίβειας μοντελοποίησης του κύκλου ζωής του δάκου (με βάση τα προαναφερθέντα μικροκλίματα) είναι εκτεταμένες: μια συστηματική προσομοίωση του κύκλου ζωής του δάκου θα οδηγήσει σε ακριβή πρόγνωση της προσβολής ενός ελαιώνα, επιτρέποντας έτσι την προληπτική αντιμετώπιση του που θα ελαχιστοποιήσει τόσο την περιττή εφαρμογή διαχείρισης επιβλαβών οργανισμών, ειδικά όταν γίνεται με χημικούς αναστολείς, καθώς και την ένταση της εφαρμογής στο απαιτούμενο επίπεδο. Αυτό με τη σειρά του θα επιτρέψει στους αγρότες να κάνουν την καλύτερη στρατηγική επιλογή, με στόχο την βελτιωμένη ποιότητα της καλλιέργειας και την ελαχιστοποίηση του κόστους και του χρόνου εφαρμογής των μεθόδων αντιμετώπισης του δάκου.

Για τους παραπάνω λόγους χρησιμοποιήθηκε ένας πλούτος μετρήσεων περιβαντολογικών αισθητήρων και μέθοδοι μηχανικής μάθησης προκειμένου να γίνει προσδιορισμός τοπικών περιβαλλοντικών/κλιματικών συνθηκών που επηρεάζουν τον κύκλο ζωής του δάκου.

4.3.1 Αλγόριθμοι μηχανικής μάθησης

Υπάρχουν διάφοροι αλγόριθμοι ομαδοποίησης που είναι κατάλληλοι για τους πειραματικούς σκοπούς. Πιο συγκεκριμένα, χρησιμοποιήθηκαν οι παρακάτω αλγόριθμοι μηχανικής μάθησης για ομαδοποίηση δεδομένων:

- Canopy [McCallum et al., 2000] είναι ένας αλγόριθμος προ-ομαδοποίησης μη επιβλεπόμενης μάθησης. Πραγματοποιεί ομαδοποίηση των δεδομένων σε δύο στάδια. Στο πρώτο στάδιο ομαδοποιεί γρήγορα τα δεδομένα σε υποσύνολα τα οποία μπορεί να επικαλύπτονται, ενώ στο δεύτερο στάδιο πραγματοποιείται μια πιο αυστηρή ομαδοποίηση των δεδομένων στα υπάρχοντα υποσύνολα. Το πλεονέκτημα αυτού του αλγορίθμου είναι ότι το σύνολο των δεδομένων εκπαίδευσης που πρέπει να συγκριθούν μεταξύ τους σε κάθε βήμα μειώνεται.
- Cobweb [Fisher, 1987a,b] είναι ένας αλγόριθμος ιεραρχικής ομαδοποίησης που οργανώνει τις παρατηρήσεις με την χρήση ενός δένδρου ταξινόμησης.
- EM [Dempster et al., 1977] είναι ένας πιθανοτικός αλγόριθμος ομαδοποίησης καθώς αναθέτει σε κάθε παρατήρηση μια κατανομή πιθανοτήτων η οποία υποδηλώνει την πιθανότητα να ανήκει σε κάθε μία από τις εξεταζόμενες ομάδες.

- FarthestFirst [Hochbaum and Shmoys, 1985] είναι ένας αλγόριθμος ομαδοποίησης ο οποίος επιλέγει αυθαίρετα τα πρώτα σημεία, ενώ κάθε διαδοχικό σημείο είναι όσο το δυνατόν πιο μακριά από το σύνολο των προηγούμενων επιλεγμένων σημείων.
- Filtered clusterer [Hall et al., 2011] είναι ένας αυθαίρετος ομαδοποιητής δεδομένων που έχει περάσει από ένα αυθαίρετο φίλτρο η δομή του οποίου βασίζεται αποκλειστικά στα δεδομένα εκπαίδευσης.
- HierarchicalClusterer [Hall et al., 2011] είναι μια ανάλυση συστάδων που στοχεύει στη δημιουργία μιας ιεραρχίας συστάδων χρησιμοποιώντας τη συσσωματική προσέγγιση.
- MakeDensityClusterer [Hall et al., 2011] είναι ένας αλγόριθμος μετά-ομαδοποίησης που περικλείει αλγόριθμους ομαδοποίησης που στοχεύει να παράγει μια κατανομή πιθανότητας και πυκνότητας.
- K-means [Forgy, 1965] είναι ένας από τους απλούστερους αλγόριθμους μάθησης χωρίς επίβλεψη που στοχεύουν στο διαχωρισμό των παρατηρήσεων σε k ομάδες στις οποίες κάθε παρατήρηση ανήκει στην ομάδα με τον πλησιέστερο μέσο όρο.

4.3.2 Συλλογή δεδομένων

4.3.2.1 Περιβαλλοντολογικά δεδομένα

Τα δεδομένα που χρησιμοποιήθηκαν στα πειράματα που πραγματοποιήθηκαν συλλέχθηκαν από περιβαλλοντικούς αισθητήρες που τοποθετήθηκαν σε εννέα θέσεις στη βορειοδυτική πλευρά του νησιού της Κέρκυρας, από τις 10 Σεπτεμβρίου 2018 έως τις 10 Οκτωβρίου 2018. Οι αισθητήρες σε κάθε θέση κατέγραφαν τιμές θερμοκρασίας και υγρασίας κάθε 15 λεπτά. Η επιλογή των θέσεων έγινε με βάση τα χαρακτηριστικά των τοποθεσιών. Συγκεκριμένα, επιλέχθηκαν τρία σημεία από τα παράλια του νησιού, τρία σημεία από κορυφές υψωμάτων και, τέλος, τρία σημεία που περικλείονται από υψώματα/λόφους και έχουν βορειοανατολικό προσανατολισμό, τις οποίες θα αποκαλούμε γούβες από εδώ και πέρα. Οι τοποθεσίες που επιλέχθηκαν αναφέρονται στο Πίνακα 4.4 και οι θέσεις τους στον χάρτη εμφανίζονται στην Εικόνα 4.29.

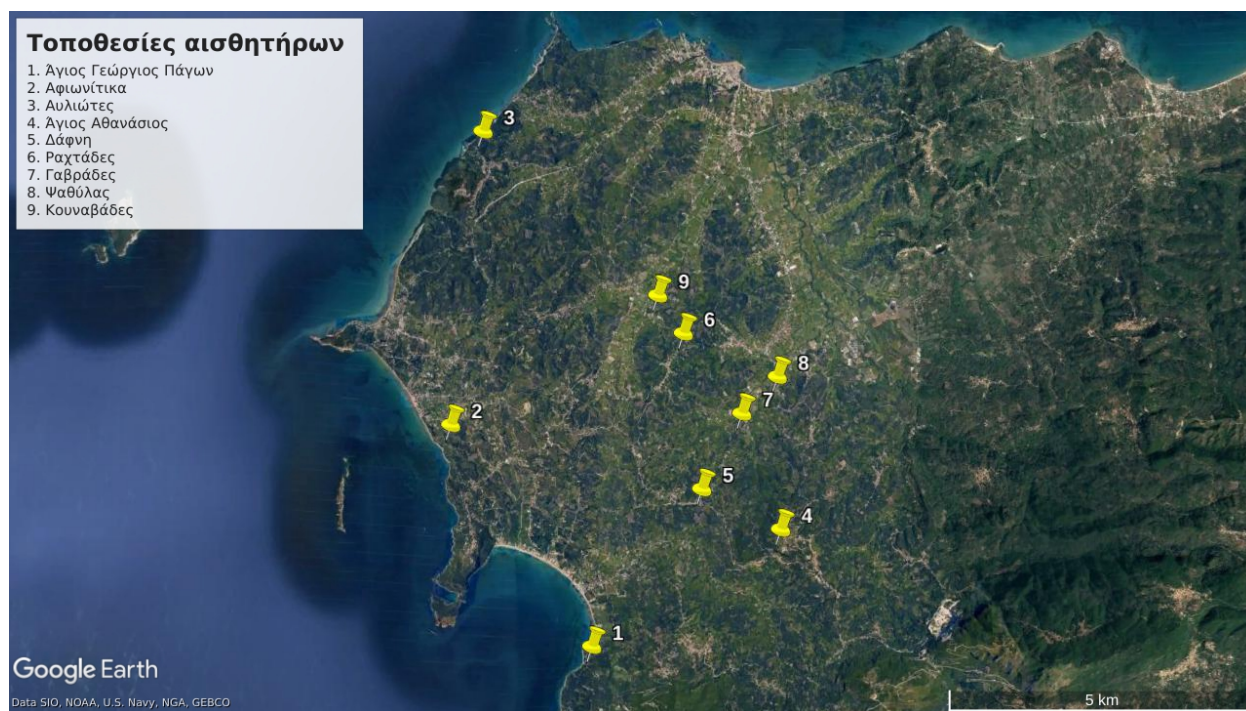
4.3.2.2 Προ-επεξεργασία δεδομένων

Όπως προαναφέρθηκε η συλλογή δεδομένων πραγματοποιήθηκε από τις 10 Σεπτεμβρίου 2018 έως και τις 10 Οκτωβρίου 2018. Όμως, προτού χρησιμοποιηθούν τα δεδομένα για την διεξαγωγή πειραμάτων πραγματοποιήθηκε μια προ-επεξεργασία αυτών. Αρχικά εντοπίστηκε το κοινό διάστημα όπου όλοι οι αισθητήρες διέθεταν μετρήσεις, καθώς υπήρχε το ενδεχόμενο κάποιος αισθητήρας να μην καταγράψει δεδομένα για τεχνικούς λόγους ή να διακόψει την λειτουργία του λόγω της απώλειας ενέργειας. Το κοινό χρονικό διάστημα είναι μεταξύ 24 Σεπτεμβρίου 2018 και 8 Οκτωβρίου 2018. Τέλος, μέσα σε αυτό το διάστημα αφαιρέθηκαν οι μετρήσεις ημερομηνιών που υπήρχε έλλειψη δεδομένων. Από αυτό το φιλτράρισμα αφαιρέθηκαν οι ακόλουθες ημερομηνίες:

Ανάπτυξη στοχαστικών αλγορίθμων και μοντέλων μηχανικής μάθησης για την εξέλιξη πληθυσμών: Εφαρμογή στο δάκο της Ελιάς

Table 4.4: Τοποθεσίες συλλογής δεδομένων

#	Τίτλος τοποθεσίας	Υψόμετρο (m)	Γεωγραφικό μήκος	Γεωγραφικό πλάτος
Παραλία				
1	Άγιος Γεώργιος Πάγων	144	39,728882	19,702466
2	Αφιωνίτικα	24	39,737251	19,654854
3	Αυλιώτες	106	39,78027306	19,65994629
Υψώματα				
4	Άγιος Αθανάσιος	212	39,7243062	19,7172308
5	Δάφνη	144	39,72888173	19,7024659
6	Ραχτάδες	135	39,75143104	19,69852521
Γούβες				
7	Γαβράδες	58	39,73986665	19,71007243
8	Ψαθύλας	39	39,74539178	19,71654322
9	Κουναβάδες	39	39,75688385	19,69359849

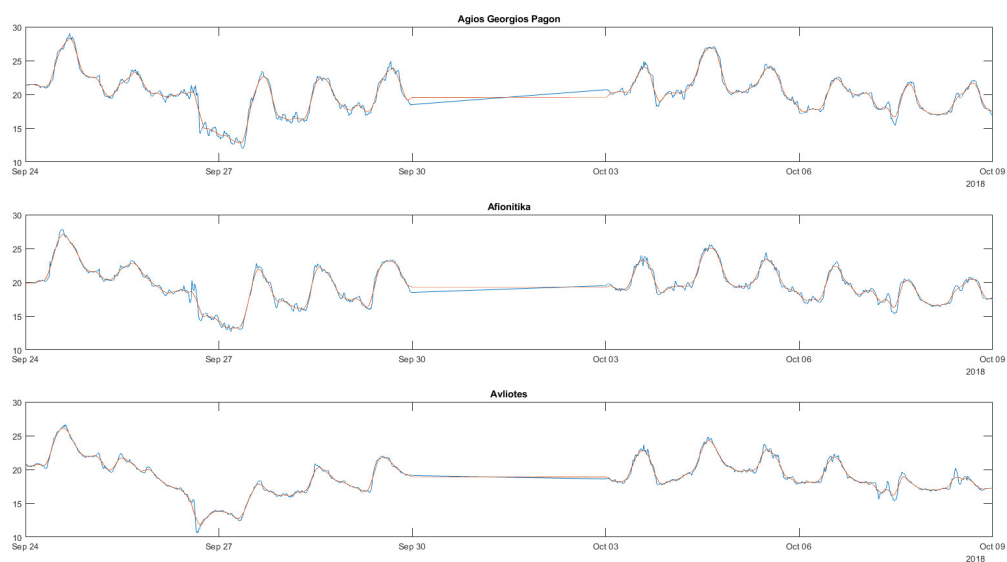


Εικόνα 4.29: Τοποθεσίες αισθητήρων

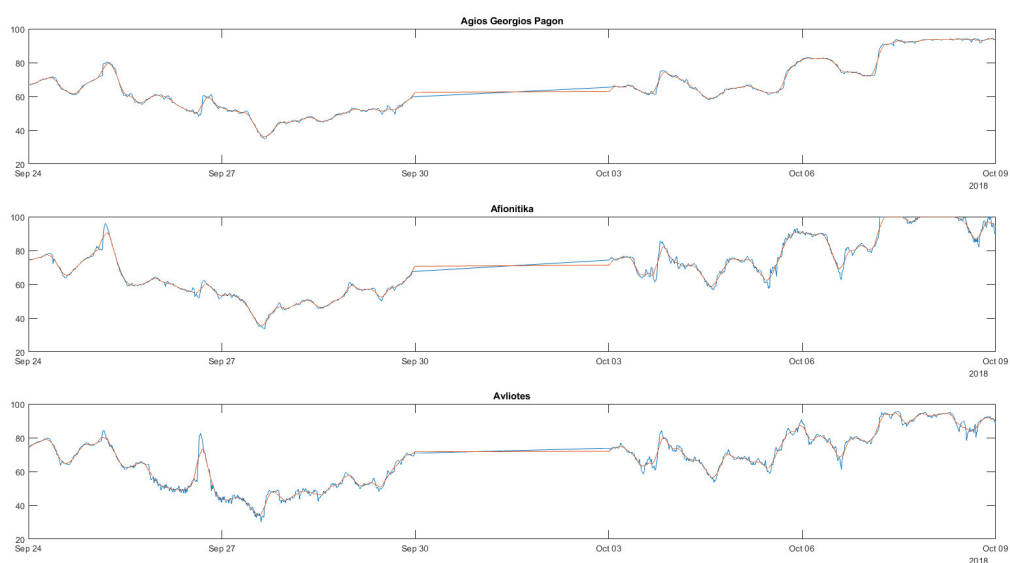
- 30 Σεπτεμβρίου 2018
- 1 Οκτωβρίου 2018
- 2 Οκτωβρίου 2018

Στις Εικόνες 4.30 έως 4.35 παρουσιάζονται τα δεδομένα θερμοκρασίας και υγρασίας που συλλέχθηκαν στο κοινό χρονικό διάστημα που αναφέρθηκε παραπάνω. Από τις παρακάτω

εικόνες φαίνεται πως οι τρεις τοποθεσίες για κάθε είδος προτεινόμενου μικροκλίματος παρουσιάζουν το ίδιο μοτίβο αυξομειώσης μεταξύ τους και στις δύο περιβαλλοντολογικές παραμέτρους.

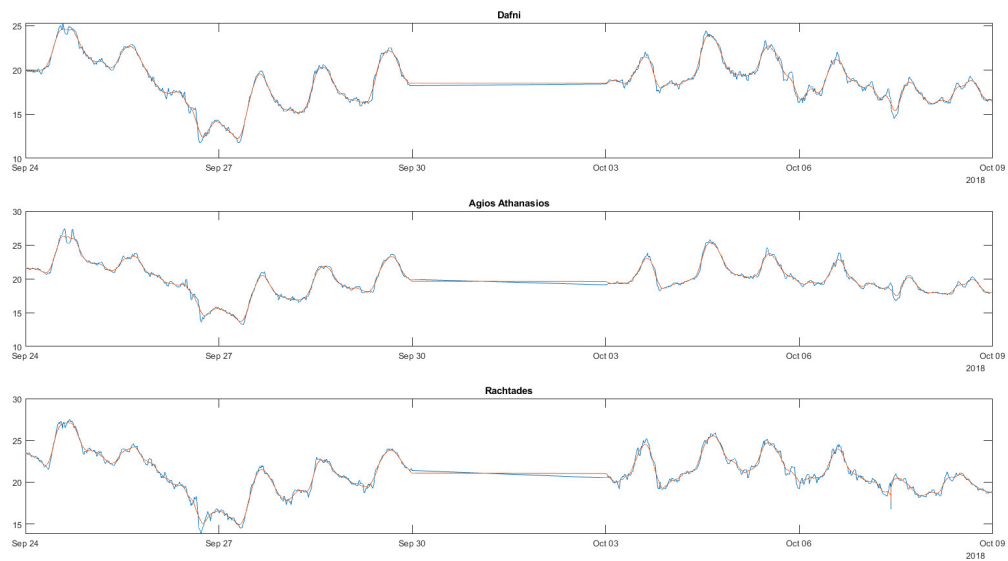


Εικόνα 4.30: Καταγεγραμμένη θερμοκρασία για τις παραλιακές τοποθεσίες.

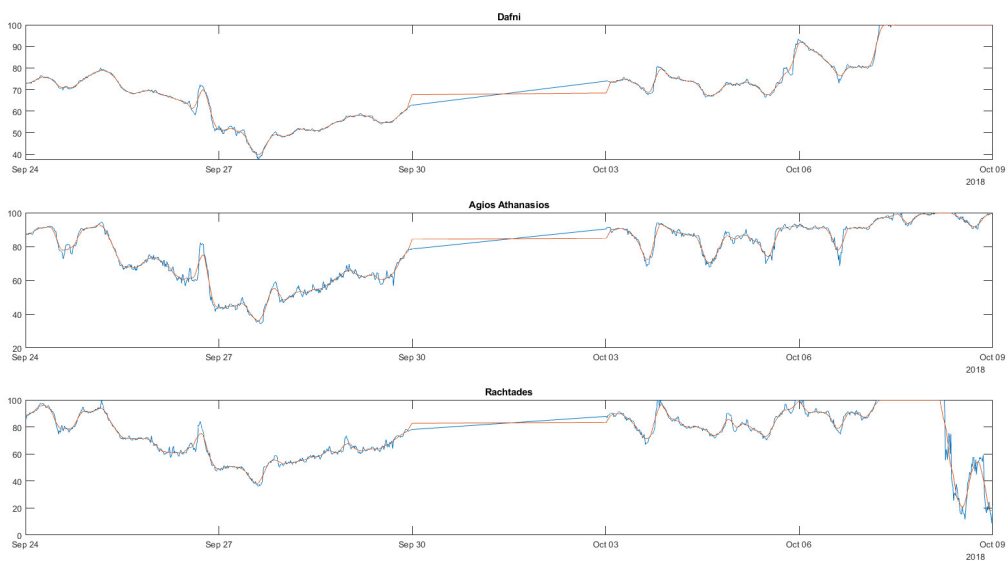


Εικόνα 4.31: Καταγεγραμμένη υγρασία για τις παραλιακές τοποθεσίες.

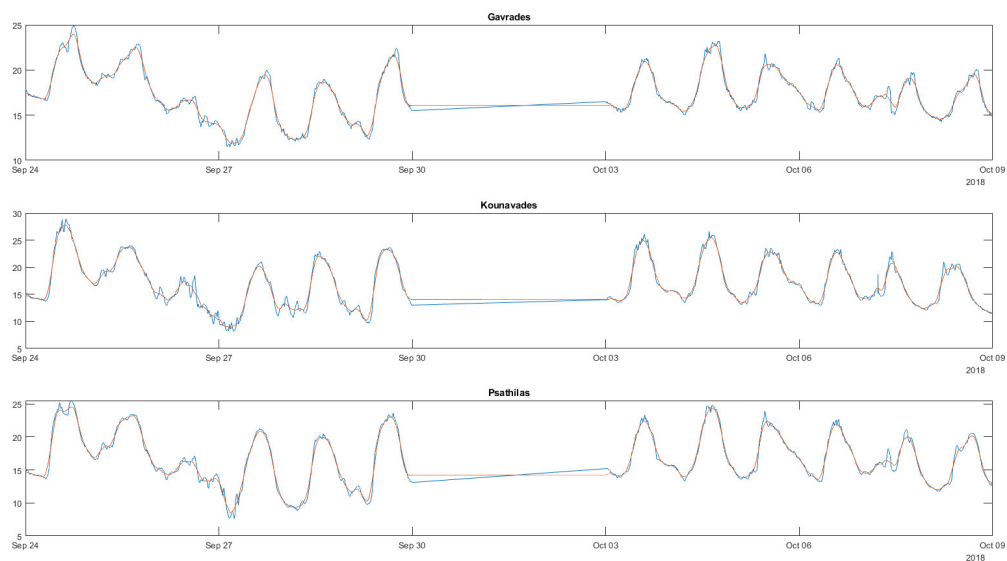
Ανάπτυξη στοχαστικών αλγορίθμων και μοντέλων μηχανικής μάθησης για την εξέλιξη πληθυσμών: Εφαρμογή στο δάκο της Ελιάς



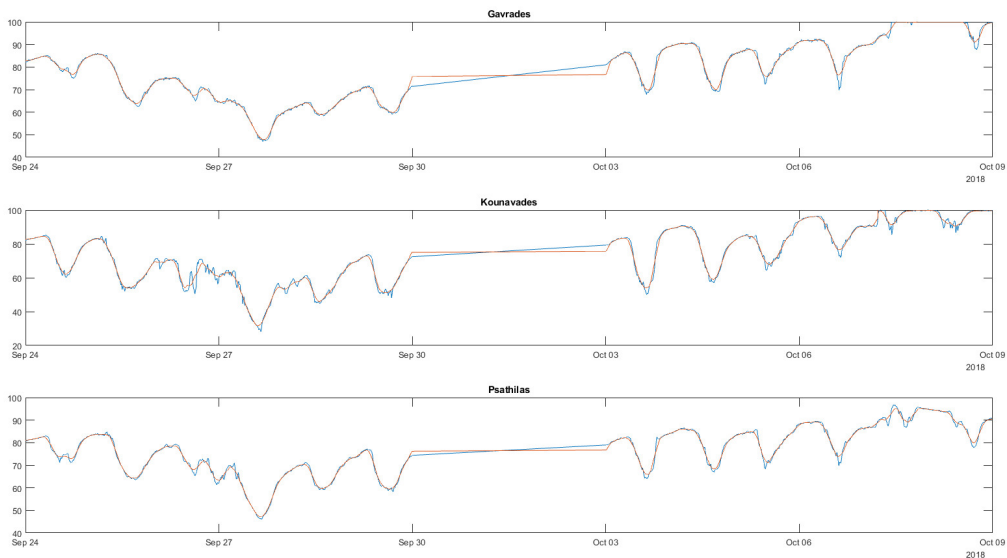
Εικόνα 4.32: Καταγεγραμμένη θερμοκρασία για τις τοποθεσίες σε υψώματα.



Εικόνα 4.33: Καταγεγραμμένη υγρασία για τις τοποθεσίες σε υψώματα.



Εικόνα 4.34: Καταγεγραμμένη θερμοκρασία για τις τοποθεσίες σε γούβες.



Εικόνα 4.35: Καταγεγραμμένη υγρασία για τις τοποθεσίες σε γούβες.

4.3.2.3 Επιλογή χαρακτηριστικών

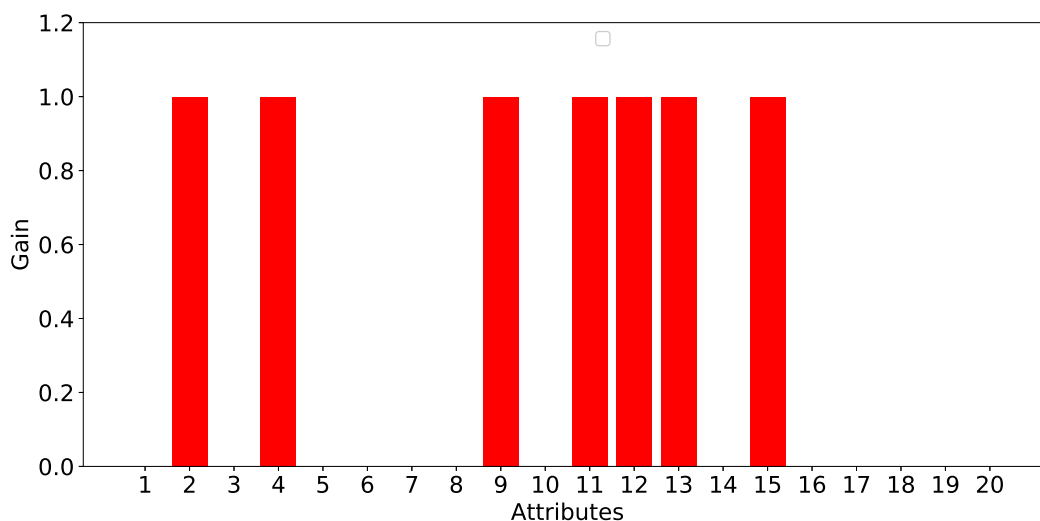
Για την πραγματοποίηση πειραμάτων ομαδοποίησης, τα περιβαλλοντικά δεδομένα που συλλέχθηκαν μετατράπηκαν στο ακόλουθο σύνολο αριθμητικών χαρακτηριστικών:

- Θερμοκρασία.

1. Ελάχιστη θερμοκρασία της χρονοσειράς.
 2. Μέση θερμοκρασία της χρονοσειράς.
 3. Μέγιστη θερμοκρασία της χρονοσειράς.
 4. Τυπική απόκλιση θερμοκρασίας της χρονοσειράς.
 5. Απόλυτη διαφορά μέγιστης και ελάχιστης θερμοκρασίας.
 6. Μέσος ρυθμός αύξησης από την ελάχιστη ημερήσια θερμοκρασία στην μέγιστη ημερήσια θερμοκρασία.
 7. Μέσος ρυθμός μείωσης από την μέγιστη ημερήσια θερμοκρασία στην τοπικά ελάχιστη ημερήσια θερμοκρασία.
 8. Μέση μέγιστη ημερήσια θερμοκρασία.
 9. Μέση ελάχιστη ημερήσια θερμοκρασία.
 10. Μέση απόλυτη διαφορά μέγιστης και ελάχιστης ημερήσιας θερμοκρασίας.
 11. Μέσος βαθμός ομοιότητας ημερήσιων χρονοσειρών θερμοκρασίας με το μικροκλίμα Α
 12. Μέσος βαθμός ομοιότητας ημερήσιων χρονοσειρών θερμοκρασίας με το μικροκλίμα Β
 13. Μέσος βαθμός ομοιότητας ημερήσιων χρονοσειρών θερμοκρασίας με το μικροκλίμα Γ
- Υγρασία
14. Ελάχιστη υγρασία της χρονοσειράς.
 15. Μέση υγρασία της χρονοσειράς.
 16. Μέγιστη υγρασία της χρονοσειράς.
 17. Τυπική απόκλιση υγρασίας της χρονοσειράς.
 18. Απόλυτη διαφοράς μέγιστης και ελάχιστης υγρασίας.
 19. Μέση ελάχιστη ημερήσια υγρασία.
 20. Μέση μέγιστη ημερήσια υγρασία.

Το τελευταίο χαρακτηριστικό που προστίθεται, το οποίο θα είναι η κλάση ομαδοποίησης, και το οποίο είναι κατηγορηματικό είναι το “Μικροκλίμα” το οποίο έχει τις ακόλουθες τρεις τιμές:

- Ύψωμα
- Γούβα
- Παραλία



Εικόνα 4.36: Αναλογία κέρδους πληροφορίας των προς εξέταση χαρακτηριστικών

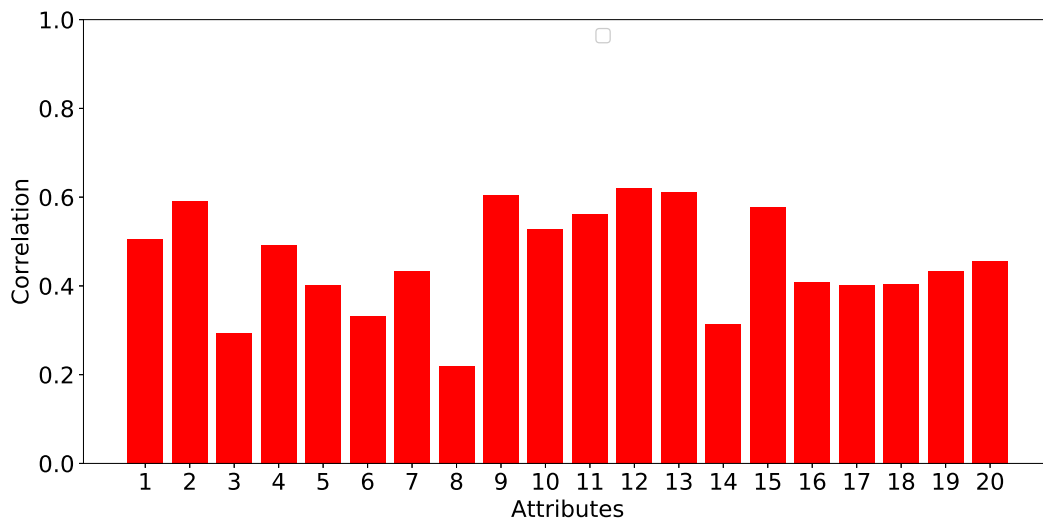
4.3.2.4 Εξαγωγή χαρακτηριστικού διανύσματος

Από τα 20 χαρακτηριστικά που αναφέρθηκαν παραπάνω η τελική επιλογή των χαρακτηριστικών για την δημιουργία του χαρακτηριστικού διανύσματος έγινε με βάση την αξιολόγηση κάθε χαρακτηριστικού ως προς το κέρδος πληροφορίας (Information gain), την συσχέτιση (Correlation) και την αναλογία κέρδους της πληροφορίας σε σχέση με την ουσιαστική πληροφορία (Information gain ratio) ως προς την κλάση Μικροκλίμα.

Παρατηρώντας το διάγραμμα της Εικόνας 4.36 βλέπουμε ότι επτά από τα 20 αρχικά χαρακτηριστικά έχουν αναλογία κέρδους 1 ενώ τα υπόλοιπα 0. Πιο συγκεκριμένα πρόκειται για τα εξής χαρακτηριστικά:

- Μέση θερμοκρασία της χρονοσειράς.
- Τυπική απόκλιση θερμοκρασίας της χρονοσειράς.
- Μέση ελάχιστη ημερήσια θερμοκρασία.
- Μέσος βαθμός ομοιότητας ημερήσιων χρονοσειρών θερμοκρασίας με το μικροκλίμα Α
- Μέσος βαθμός ομοιότητας ημερήσιων χρονοσειρών θερμοκρασίας με το μικροκλίμα Β
- Μέσος βαθμός ομοιότητας ημερήσιων χρονοσειρών θερμοκρασίας με το μικροκλίμα Γ
- Μέση υγρασία της χρονοσειράς.

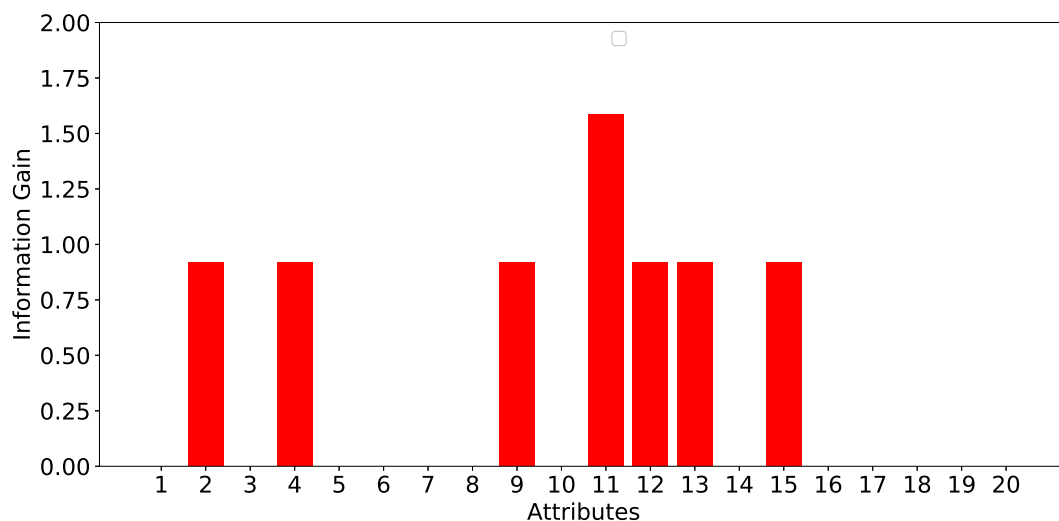
Το παρακάτω διάγραμμα (Εικόνα 4.37) παρουσιάζει την αξία κάθε χαρακτηριστικού μέσω του συντελεστή συσχέτισης τους με την κλάση ομαδοποίησης “Μικροκλίμα”. Όπως προκύπτει



Εικόνα 4.37: Συσχέτιση των προς εξέταση χαρακτηριστικών με την κλάση ταξινόμησης.

από το διάγραμμα τα επτά χαρακτηριστικά που εμφάνισαν την μεγαλύτερη αναλογία κέρδους (βλ. Εικόνα 4.36) συν το χαρακτηριστικό 10 (“Μέση απόλυτη διαφορά μέγιστης και ελάχιστης ημερήσιας θερμοκρασίας”) έχουν την μεγαλύτερη συσχέτιση (>0.5) σε σχέση με τα υπόλοιπα χαρακτηριστικά.

Τέλος, το διάγραμμα της Εικόνας 4.38 παρουσιάζει το κέρδος πληροφορίας των υποψήφιων χαρακτηριστικών για εισαγωγή στο χαρακτηριστικό διάνυσμα. Όπως προκύπτει από το διάγραμμα τα ίδια επτά χαρακτηριστικά με το διάγραμμα τις Εικόνας 4.36 εμφανίζουν το μεγαλύτερο κέρδος πληροφορίας. Συγκεκριμένα το χαρακτηριστικό 11 δίνει το μεγαλύτερο κέρδος



Εικόνα 4.38: Κέρδος πληροφορίας των προς εξέταση χαρακτηριστικών.

πληροφορίας (1.585), ενώ τα υπολειπόμενα έξι χαρακτηριστικά (2, 4, 9, 12, 13, 15) δίνουν το ίδιο κέρδος πληροφορίας, 0.918. Τα εναπομείναντα χαρακτηριστικά δεν προσδίδουν κανένα κέρδος πληροφορίας.

Από την παραπάνω ανάλυση τα χαρακτηριστικά που επιλέγονται για την σύνθεση του χαρακτηριστικού διανύσματος είναι τα ακόλουθα:

- Μέσος βαθμός ομοιότητας ημερήσιων χρονοσειρών θερμοκρασίας με το μικροκλίμα Α
- Μέσος βαθμός ομοιότητας ημερήσιων χρονοσειρών θερμοκρασίας με το μικροκλίμα Β
- Μέσος βαθμός ομοιότητας ημερήσιων χρονοσειρών θερμοκρασίας με το μικροκλίμα Γ
- Μέση ελάχιστη ημερήσια θερμοκρασία.
- Μέση υγρασία της χρονοσειράς.
- Μέση θερμοκρασία της χρονοσειράς.
- Τυπική απόκλιση θερμοκρασίας της χρονοσειράς.

4.3.3 Αξιολόγηση απόδοσης

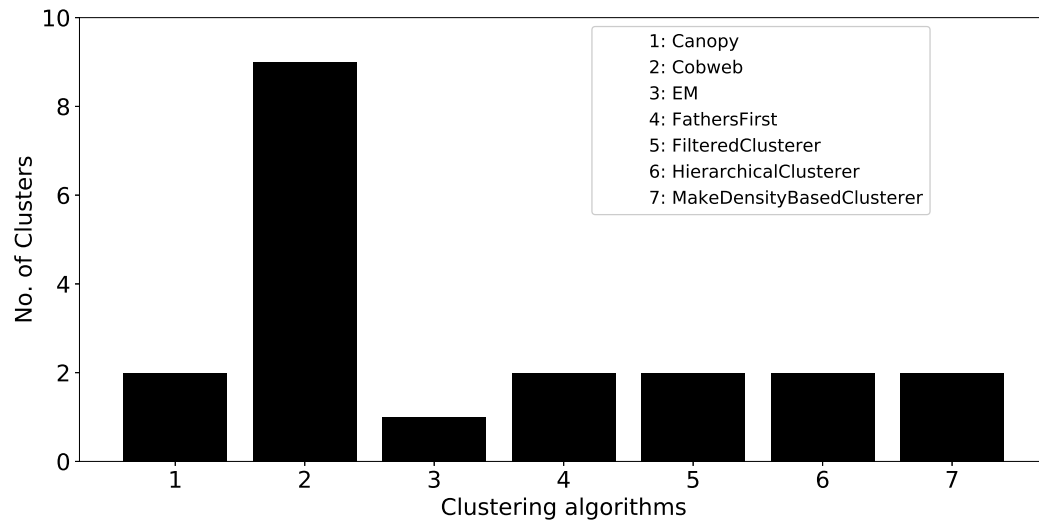
Για την υποστήριξη της αποτελεσματικότητας του προτεινόμενου διανύσματος χαρακτηριστικών και των εξετασθέντων αλγορίθμων μηχανικής μάθησης, αυτή η ενότητα παρουσιάζει έναν αριθμό πειραμάτων που έχουν εκτελεστεί. Περιγράφεται επίσης μια συνοπτική περιγραφή της πλατφόρμας πειραμάτων και των δεδομένων, ακολουθούμενη από μια ανάλυση απόδοσης.

4.3.3.1 Πειραματική διάταξη

Η πλατφόρμα μηχανικής μάθησης WEKA χρησιμοποιήθηκε για τη διεξαγωγή των πειραμάτων ομαδοποίησης. Για τον αλγόριθμο K-Means ο αριθμός συστάδων κυμαινόταν από 2 έως 3 και εξετάστηκαν οι συναρτήσεις απόστασης Manhattan και Euclidean. Στην διάθεση μας είχαμε 9 εκπαιδευτικά παραδείγματα. Παρότι ο αριθμός παραδειγμάτων είναι μικρός για να βγουν ασφαλή συμπεράσματα, επαρκούν ωστόσο για μια πρώτη.

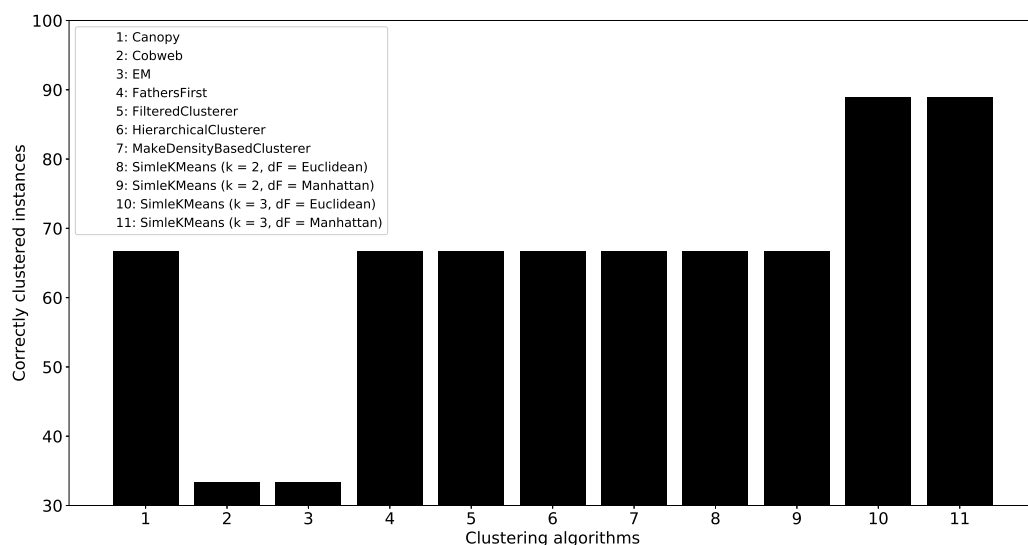
4.3.3.2 Αποτελέσματα

Η Εικόνα 4.39 παρουσιάζει των αριθμό συστάδων που χωρίστηκαν τα δεδομένα μετά την εφαρμογή των αλγορίθμων συσταδοποίησης (Canopy, Cobweb, EM, FarthersFirst, FilteredClusterer, HierarchicalClusterer, MakeDensityBasedClusterer). Η πλειοψηφία των αλγορίθμων χώρισε τις ελεγχόμενες τοποθεσίες σε δύο συστάδες, ενώ ο αλγόριθμος EM ομαδοποίησε και τις εννέα τοποθεσίες σε μία συστάδα. Τέλος, ο αλγόριθμος Cobweb ανέθεσε κάθε τοποθεσία σε δικιά της συστάδα.



Εικόνα 4.39: Αριθμός συστάδων που χώρισαν τα δεδομένα οι αλγόριθμοι συσταδοποίησης

Τέλος, η Εικόνα 4.40 παρουσιάζει την ακρίβεια σωστής συσταδοποίησης των παραδειγμάτων από τους αλγόριθμους συσταδοποίησης που χρησιμοποιήθηκαν στα πειράματά μας. Την χειρότερη απόδοση είχαν οι αλγόριθμοι EM και Cobweb με ακρίβεια 33%. Σε αντίθεση με τον αλγόριθμο K-means, που είχε την καλύτερη απόδοση (89%), με $k=3$ και για τις δύο εξεταζόμενες συναρτήσεις απόστασης. Οι εναπομείναντες αλγόριθμοι είχαν παρόμοια απόδοση, η οποία ήταν 66%.



Εικόνα 4.40: Ακρίβεια σωστής συσταδοποίησης για διάφορους αλγόριθμους συσταδοποίησης.

4.3.4 Συμπεράσματα

Με βάση το χαρακτηριστικό διάνυσμα που χρησιμοποιήθηκε και τα πειράματα που πραγματοποιήθηκαν καταλήγουμε στο συμπέρασμα ότι οι τοποθεσίες που επιλέχθηκαν μπορούν να διαχωριστούν σε δύο μικροκλίματα. Συγκεκριμένα, στα μικροκλίματα “Γουβα” και “Παράλια – Υψώματα” που είναι μια συγχώνευση των περιοχών που άνηκαν στα μικροκλίματα “Παράλια” και “Υψώματα” με βάση την αρχική μας υπόθεση. Πρέπει όμως να επισημανθεί ότι το παραπάνω συμπέρασμα προκύπτει λόγω του μικρού αριθμού τοποθεσιών για τις οποίες έγινε ανάλυση. Οπότε με την εισαγωγή περισσότερων τοποθεσιών με παρόμοια χαρακτηριστικά θα μπορούσε να επιβεβαιωθεί η αρχική μας υπόθεση για την ύπαρξη τριών μικροκλιμάτων.

Ανάπτυξη στοχαστικών αλγορίθμων και μοντέλων μηχανικής μάθησης για την εξέλιξη πληθυσμών: Εφαρμογή στο δάκο της Ελιάς

5. ΕΠΙΛΟΓΟΣ

Η παραπάνω διατριβή πραγματοποιήθηκε στο πλαίσιο των διδακτορικών σπουδών μου στο Τμήμα Πληροφορικής του Ιονίου Πανεπιστημίου και πραγματεύεται την ανάπτυξη στοχαστικών αλγορίθμων και την ανάπτυξη μοντέλων μηχανικής μάθησης σε σχέση με την εξέλιξη πληθυσμού του δάκου της ελιάς. Ο δάκος είναι μια μύγα φρούτου που παρασιτεί στον καρπό του ελαιόδεντρου, χρησιμοποιώντας τον ως το μέσο ανάπτυξης των πρώιμων σταδίων του. Λόγω, αυτής της εξάρτησης του εντόμου με την ελιά η μη αντιμετώπιση του μπορεί να προκαλέσει ανυπολογιστες ζημιές, τόσο ποσότητες όσο και ποιοτικές, στα προϊόντα που παράγονται από το ελαιόδεντρο. Η δυνατότητα πρόβλεψης έξαρσης της δακοπροσβολής σε έναν ελαιώνα θα επέτρεπε στην ταχεία και αποτελεσματικότερη αντιμετώπιση του παρασίτου.

Στα πλαίσια αυτής της προσπάθειας, πραγματοποιήθηκε βελτιστοποίηση υπάρχοντος λογισμικού προσομοίωσης της χωροχρονικής εξέλιξης του πληθυσμού του δάκου της ελιάς. Οι κυριότερες βελτιστοποιήσεις που είσηχθησαν στο λογισμικό ήταν, αρχικά, η δημιουργία ενός στοχαστικού μοντέλου διασποράς του δάκου με βάση το επίπεδο καρποφορίας του ελαιώνα στον οποίο δραστηροποιείται. Επίσης, στο μοντέλο διασποράς προστέθηκε η επιροή της θερμοκρασίας του περιβάλλοντος, με ανασταλτικό χαρακτήρα. Τέλος, προστέθηκε η δυνατότητα κατακερματισμού του πεδίου προσομοίωσης σε μικροκλίματα.

Στα πειράματα που εκτελέστηκαν για την επιλογή του μοντέλου διασποράς τα αποτελέσματα του μοντέλου που επικράτησε έδειξαν ότι ήταν ικανό να αναπαραστήσει σε ικανοποιητικό βαθμό δεδομένα διασποράς του δάκου, σε σχέση με την καρποφορία του ελαιώνα, που είχαν καταγραφεί στο πεδίο. Επίσης, με την εισαγωγή του χαρακτηριστικού διαχωρισμού του πεδίου σε μικροκλίματα εξετάστηκε, μέσω προσομοιώσεων, πρώτον η επιροή που έχουν στην εξέλιξη του πληθυσμού του δάκου και δεύτερον αν θα πρέπει να λαμβάνονται υπόψη στην λήψη αποφάσης για την εφαρμογή μέτρων αντιμετώπισης της δακοπροσβολής. Τα πειράματα έδειξαν ότι αναλόγα τα χαρακτηριστικά του μικροκλίματος η ανάπτυξη των πρώιμων σταδίων του δάκου θα είναι ταχύτερη ή πιο αργή το οποίο κατα συνέπεια επηρεάζει το πότε θα υπάρξει έξαρση της δακοπροσβολής στον ελαιώνα. Τέλος, το συμπέρασμα που προέκυψε σε σχέση με την συσχέτιση των μικροκλιμάτων με την εφαρμογή μέτρων αντιμετώπισης, είναι ότι η αγνόηση τους είναι πιθανό να οδηγήσει σε άσκοπη χρήση χημικών και άρα περιττά έξοδα για τον ελαιοκαλλιεργητή.

Το ζήτημα προσεγγίστηκε και από την πλευρά της τεχνητής νοημοσύνης, με την ανάπτυξη διάφορων μοντέλων μηχανικής μάθησης, με κάθε ένα από αυτά να προσπαθούν να λύσουν ένα διαφορετικό πρόβλημα. Αρχικά, δημιουργήθηκαν και αξιολογήθηκαν τέσσερα μοντέλα μηχανικής μάθησης με στόχο την πρόβλεψη της επόμενης μέτρησης μια δακοπαγίδας. Στη συνέχεια, λόγω της ύπαρξης αρκετών "έξυπνων"/ηλεκτρονικών δακοπαγίδων επιχειρήθηκε η δημιουργία ενός μοντέλου, και ενός συνόλου εργαλειών γενικότερα, αυτόματης αναγνώρισης δάκων σε εικόνες. Τέλος, δημιουργήθηκε ένα μοντέλο με στόχο την επιτυχή αναγνώριση και ομαδοποίηση μικροκλιμάτων.

Από τα πειράματα που εκτελέστηκαν για τα παραπάνω μοντέλα τα αποτελέσματα ήταν αρκετά υποσχόμενα, με το μοντέλο για την αυτόματη αναγνώριση δάκων σε εικόνες να παράγει πολύ καλά αποτελέσματα.

5.1 Μελλοντική έρευνα

Φυσικά όλα τα παραπάνω μπορούν να βελτιωθούν περαιτέρω. Σε ότι αφορά το λογισμικό προσομοίωσης μερικές προσθήκες/αλλαγές που θα συνεισφέρουν σε αυτό τον σκοπό είναι οι ακόλουθες:

- Προσθήκη της υγρασίας στους περιβαντολογικούς παράγοντες που επηρεάζουν την ανάπτυξη και την εν γένει συμπεριφορά του δάκου.
- Εισαγωγή της θερμοκρασίας ως παράγοντας που επηρεάζει την μακροζωία, την ανάπτυξη ωοθηκών, για τους θυληκούς δάκους, και την απόδοση της πτήσης του ενήλικα δάκου.
- Μεταβαλλόμενη θερμοκρασία μέσα στο εικονικό πεδίο. Στην τρέχουσα έκδοση η θερμοκρασία είναι ίδια για όλες τις θέσεις είτε για όλο το πεδίο είτε για κάθε μικροκλίμα στο οποίο έχει χωριστεί το πεδίο.

Για τα μοντέλα μηχανικής μάθησης, γενικότερα, η συλλογή περισσότερων παραδειγμάτων εκπαίδευσης θα βοηθήσει στην βελτιώση της ακρίβειας πρόβλεψης. Ειδικότερα, για το μοντέλο πρόβλεψης επόμενης μέτρησης δακοπαγίδας η ενσωμάτωση της υγρασίας είτε στο χαρακτηριστικό διάνυσμα είτε ως παράγοντας που επηρεάζει την θερμοκρασία και ως εκ τούτου το χαρακτηριστικό DD θα έχει ενδιαφέρον να εξεταστεί μελλοντικά. Για το μοντέλο αναγνώρισης δάκων σε εικόνες το επόμενο στάδιο έρευνας θα μπορούσε να εξετάσει την αναγνώριση του φύλου των παγιδευμένων δάκων. Τέλος, για το μοντέλο αναγνώρισης μικροκλιμάτων θα μπορούσε να εξεταστεί η εισαγωγή καινούργιων χαρακτηριστικών στο χαρακτηριστικό διάνυσμα για την επίτευξη μεγαλύτερης ακρίβειας.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Adeva, J. G., Botha, J., and Reynolds, M. (2012). A simulation modelling approach to forecast establishment and spread of bactrocera fruit flies. *Ecological Modelling*, 227:93–108.
- Adeva, J. G. and Reynolds, M. (2012). Web-based simulation of fruit fly to support biosecurity decision-making. *Ecological informatics*, 9:19–36.
- Aggarwal, C. C. (2018a). *Neural Networks and Deep Learning*, chapter 1. An Introduction to Neural Networks, pages 1–4. Springer.
- Aggarwal, C. C. (2018b). *Neural Networks and Deep Learning*, chapter 2. The Basic Architecture of Neural Networks, pages 4–21. Springer.
- Alfonso, M. and Owen, J. (2002). Alternative methods for controlling the olive fly, bactrocera oleae, involving semiochemicals. use of pheromones and other semi chemicals in integrated production. *IOBC wprs Bulletin*, 25.
- Allen, J. C. (1976). A modified sine wave method for calculating degree days. *Environmental Entomology*, 5(3):388–396.
- Alorda, B., Valdes, F., Mas, B., Leza, M., Almenar, L., Feliu, J., Ruiz, M., and Miranda, M. (2015). Design of an energy efficient and low cost trap for olive fly monitoring using a zigbee based wireless sensor network. In *European Conference Precision Agriculture*.
- Ant, T., Koukidou, M., Rempoulakis, P., Gong, H.-F., Economopoulos, A., Vontas, J., and Alphey, L. (2012). Control of the olive fruit fly using genetics-enhanced sterile insect technique. *BMC biology*, 10(1):1–8.
- Arnold, C. Y. et al. (1960). Maximum-minimum temperatures as a basis for computing heat units. In *Proceedings. American Society for Horticultural Science*, volume 76, pages 682–92.
- Athar, M. (2005). Infestation of olive fruit fly, bactrocera oleae, in california and taxonomy of its host trees. *Agriculturae Conspectus Scientificus*, 70(4):135–138.
- Avidov, Z. (1954). Further investigations on the ecology of the olive fly (dacus oleae gmel.) in israel. *Ktavim*, 4:39–50.
- Avissar, R. and Mahrer, Y. (1982). Verification study of a numerical greenhouse microclimate model. *Transactions of the ASAE*, 25(6):1711–1720.
- Ayoub, N. M., Siddique, A. B., Ebrahim, H. Y., Mohyeldin, M. M., and El Sayed, K. A. (2017). The olive oil phenolic (-)-oleocanthal modulates estrogen receptor expression in luminal breast cancer in vitro and in vivo and synergizes with tamoxifen treatment. *European journal of pharmacology*, 810:100–111.

- Baskerville, G. L. and Emin, P. (1969). Rapid estimation of heat accumulation from maximum and minimum temperatures. *Ecology*, 50(3):514–517.
- Batista, G., Bazzan, A. L., and Monard, M. C. (2003). Balancing training data for automated annotation of keywords: a case study. In *Proceedings of the Second Brazilian Workshop on Bioinformatics*, pages 20–28.
- Batista, G., Prati, R. C., and Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM Sigkdd Explorations Newsletter*, 6(1):20–29.
- Batuwita, R. and Palade, V. (2009). micropred: effective classification of pre-mirnas for human mirna gene prediction. *Bioinformatics*, 25(8):989–995.
- Bauer, A., Fanning, C., Enz, J. W., and Eberlein, C. (1984). Use of growing-degree days to determine spring wheat growth stages. *Extension bulletin-North Dakota State University of Agriculture and Applied Science, Cooperative Extension Service (USA)*.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Brnetic, D. et al. (1979). Control of the olive fly (*dacus oleae* gmel.) with regard to protection of the environment and oil production. *Poljoprivreda i Sumarstvo*, 25(1):3–16.
- Broufas, G., Pappas, M., and Koveos, D. (2009). Effect of relative humidity on longevity, ovarian maturation, and egg production in the olive fruit fly (diptera: Tephritidae). *Annals of the Entomological Society of America*, 102(1):70–75.
- Brown, P. W. (1989). Heat units. *Bull.*, 8915.
- Bruno, V., Baldacchini, V., and Di Gregorio, S. (2010). Temperature effects on olive fruit fly infestation in the flysim cellular automata model. In *Natural Computing*, pages 125–132. Springer.
- Cândido, F. G., Valente, F. X., da Silva, L. E., Coelho, O. G. L., Peluzio, M. d. C. G., and Alfenas, R. d. C. G. (2018). Consumption of extra virgin olive oil improves body composition and blood pressure in women with excess body fat: a randomized, double-blinded, placebo-controlled clinical trial. *European journal of nutrition*, 57(7):2445–2455.
- Cantlon, J. E. (1953). Vegetation and microclimates on north and south slopes of cusetunk mountain, new jersey. *Ecological Monographs*, 23(3):241–270.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.
- Chi, H. (1990). Timing of control based on the stage structure of pest populations: a simulation approach. *Journal of Economic entomology*, 83(4):1143–1150.
- Cieslak, D. A., Chawla, N. V., and Striegel, A. (2006). Combating imbalance in network intrusion datasets. In *GrC*, pages 732–737.

- Collobert, R. and Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167.
- Comins, H. N. and Fletcher, B. S. (1988). Simulation of fruit fly population dynamics, with particular reference to the olive fruit fly, *dacus oleae*. *Ecological modelling*, 40(3):213–231.
- Crovetti, A., Quaglia, F., Loi, G., Rossi, E., Malfatti, P., Chesi, F., Conti, B., Belcari, A., Raspi, A., and Paparatti, B. (1982). Influenza di temperatura e umidità sullo sviluppo degli stadi preimaginali di *dacus oleae* (gmelin). *Frustr Entom*, 18:133–166.
- Daane, K. M. and Johnson, M. W. (2010). Olive fruit fly: managing an ancient pest in modern times. *Annual review of entomology*, 55:151–169.
- De la Torre, R., Corella, D., Castañer, O., Martínez-González, M. A., Salas-Salvador, J., Vila, J., Estruch, R., Sorli, J. V., Arós, F., Fiol, M., et al. (2017). Protective effect of homovanillyl alcohol on cardiovascular disease and total mortality: Virgin olive oil, wine, and catechol-methylthion. *The American journal of clinical nutrition*, 105(6):1297–1304.
- DeGrandi-Hoffman, G., Roth, S. A., Loper, G., and Erickson Jr, E. (1989). Beepop: A honeybee population dynamics simulation model. *Ecological modelling*, 45(2):133–150.
- del Sagrado, J. and del Águila, I. M. (2007). Olive fly infestation prediction using machine learning techniques. In *Conference of the Spanish Association for Artificial Intelligence*, pages 229–238. Springer.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.
- Dimou, I., Koutsikopoulos, C., Economopoulos, A., and Lykakis, J. (2003). Depth of pupation of the wild olive fruit fly, *bactrocera (dacus) oleae* (gmel.)(dipt., tephritidae), as affected by soil abiotic factors. *Journal of applied entomology*, 127(1):12–17.
- Dimoudi, A., Kantzioura, A., Zoras, S., Pallas, C., and Kosmopoulos, P. (2013). Investigation of urban microclimate parameters in an urban center. *Energy and Buildings*, 64:1–9.
- Doitsidis, L., Fouskitakis, G. N., Varikou, K. N., Rigakis, I. I., Chatzichristofis, S. A., Papafilippaki, A. K., and Birouraki, A. E. (2017). Remote monitoring of the *bactrocera oleae* (gmelin) (diptera: Tephritidae) population using an automated mcphail trap. *Computers and Electronics in Agriculture*, 137:69 – 78.
- Domingos, P. (1999). Metacost: A general method for making classifiers cost-sensitive. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 155–164. ACM.
- Doyle, S., Monaco, J., Feldman, M., Tomaszewski, J., and Madabhushi, A. (2011). An active learning based classification strategy for the minority class problem: application to histopathology annotation. *BMC bioinformatics*, 12(1):424.

- Economopoulos, A., Haniotakis, G., Mathioudis, J., Missis, N., and Kinigakis, P. (1978). Long-distance flight of wild and artificially-reared *dacus oleae* (gmelin)(diptera, tephritidae). *Zeitschrift für Angewandte Entomologie*, 87(1-4):101–108.
- El, S. N. and Karakaya, S. (2009). Olive tree (*olea europaea*) leaves: potential beneficial effects on human health. *Nutrition reviews*, 67(11):632–638.
- Elman, J. L. (1990). Finding structure in time. *Cognitive science*, 14(2):179–211.
- Escrich, E., Moral, R., and Solanas, M. (2011). Olive oil, an essential component of the mediterranean diet, and breast cancer. *Public health nutrition*, 14(12A):2323–2332.
- Esposito, K., Kastorini, C.-M., Panagiotakos, D. B., and Giugliano, D. (2011). Mediterranean diet and weight loss: meta-analysis of randomized controlled trials. *Metabolic syndrome and related disorders*, 9(1):1–12.
- European Commission - Community Research and Development Information Service (2014). E-FLYWATCH Report Summary. Τελευταία πρόσβαση Ιούνιος 2021, https://cordis.europa.eu/result/rcn/141151_en.html.
- European Commission - Community Research and Development Information Service (2017). Development of an innovative automated and wireless trap with warning and monitoring modules for integrated management of the mediterranean (*ceratitis capitata*) & olive (*dacus oleae*) fruit flies. Τελευταία πρόσβαση Ιούνιος 2021, https://cordis.europa.eu/project/rcn/96182_en.html.
- Eurostat (2017). Agri-environmental indicator - cropping patterns. Accessed October 2018, https://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_cropping_patterns.
- Fallahi, A. and Jafari, S. (2011). An expert system for detection of breast cancer using data preprocessing and bayesian network. *Int J Adv Sci Technol*, 34:65–70.
- Fand, B. B., Sul, N. T., Bal, S. K., and Minhas, P. S. (2015). Temperature impacts the development and survival of common cutworm (*spodoptera litura*): Simulation and visualization of potential population growth in india under warmer temperatures through life cycle modelling and spatial mapping. *PLOS ONE*, 10(4):1–25.
- Fisher, D. H. (1987a). Improving inference through conceptual clustering. In *AAAI*, volume 87, pages 461–465.
- Fisher, D. H. (1987b). Knowledge acquisition via incremental conceptual clustering. *Machine learning*, 2(2):139–172.
- Fletcher, B. and Kapatatos, E. (1981). Dispersal of the olive fly, *dacus oleae*, during the summer period on corfu. *Entomologia Experimentalis et Applicata*, 29(1):1–8.

- Fletcher, B. and Kapatos, E. (1983). The influence of temperature, diet and olive fruits on the maturation rates of female olive flies at different times of the year. *Entomologia experimentalis et applicata*, 33(3):244–252.
- Fogher, C., Busconi, M., Sebastiani, L., and Bracci, T. (2010). Chapter 2 - olive genomics. In Preedy, V. R. and Watson, R. R., editors, *Olives and Olive Oil in Health and Disease Prevention*, pages 17 – 24. Academic Press, San Diego.
- Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *biometrics*, 21:768–769.
- Freund, Y. and Schapire, R. E. (1996). Experiments with a new boosting algorithm. In *International Conference on Machine Learning*, volume 96, pages 148–156.
- Gaouar, N. and Debouzie, D. (1995). Within-tree vertical pattern in *bactrocera oleae* gmel.(dipt, tephritidae) infestations and optimization of insecticide applications. *Journal of Applied Entomology*, 119(1-5):251–254.
- Genç, H. and Nation, J. L. (2008). Maintaining *bactrocera oleae* (gmelin.)(diptera: Tephritidae) colony on its natural host in the laboratory. *Journal of pest science*, 81(3):167.
- Gilioli, G. and Pasquali, S. (2007). Use of individual-based models for population parameters estimation. *ecological modelling*, 200(1):109–118.
- Gompertz, B. (1825). Xxiv. on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. frs &c. *Philosophical transactions of the Royal Society of London*, (115):513–583.
- Gonçalves, M. F. and Torres, L. M. (2011). The use of the cumulative degree-days to predict olive fly, *bactrocera oleae* (rossi), activity in traditional olive groves from the northeast of portugal. *Journal of Pest Science*, 84(2):187–197.
- Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., and Schmidhuber, J. (2009). A novel connectionist system for unconstrained handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):855–868.
- Gu, S., Sabuwala, A., and Gohil, H. (2014). Comparison of growing degree hours based on hourly average temperatures with growing degree days based on daily minimum and maximum or average temperatures to interpret heat summation. In *HORTSCIENCE*, volume 49, pages S160–S160. AMER SOC HORTICULTURAL SCIENCE 113 S WEST ST, STE 200, ALEXANDRIA, VA 22314
- Gümusay, B., ÖOzilbey, U., Ertem, G., and Oktar, A. (1989). Studies on the susceptibility of some important table and oil olive cultivars of aegean region to olive fly (*dacus oleae* gmel.) in turkey. In *International Symposium on Olive Growing* 286, pages 359–362.
- Guo, H. and Viktor, H. L. (2004). Learning from imbalanced data sets with boosting and data generation: the databoost-im approach. *ACM SIGKDD Explorations Newsletter*, 6(1):30–39.

- Gutierrez, A. P., Ponti, L., and Cossu, Q. (2009). Effects of climate warming on olive and olive fly (*bactrocera oleae* (gmelin)) in california and italy. *Climatic Change*, 95(1-2):195–217.
- Hall, M., Witten, I., and Frank, E. (2011). *Data Mining: practical machine learning tools and techniques*.
- Haniotakis, G. E. (2005). Olive pest control: present status and prospects. *IOBC wprs Bulletin*, 28(9):1.
- Hart, P. (1968). The condensed nearest neighbor rule (corresp.). *IEEE Transactions on Information Theory*, 14(3):515–516.
- Hermes, D. A. (2004). Using degree-days and plant phenology to predict pest activity. *IPM (integrated pest management) of midwest landscapes*, pages 49–59.
- Hochbaum and Shmoys (1985). A best possible heuristic for the k-center problem. *Mathematics of Operations Research*, 10(2):180–184.
- Hodkinson, I. D. (2005). Terrestrial insects along elevation gradients: species and community responses to altitude. *Biological reviews*, 80(3):489–513.
- Holm, J. A., Miller, C. J., and Cropper Jr, W. P. (2008). Population dynamics of the dioecious amazonian palm *mauritia flexuosa*: simulation analysis of sustainable harvesting. *Biotropica*, 40(5):550–558.
- Holmes, R. and Dingle, A. N. (1965). The relationship between the macro-and microclimate. *Agricultural Meteorology*, 2(2):127–133.
- Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574–591.
- Japkowicz, N. (2000). The class imbalance problem: Significance and strategies. In *Proc. of the Int'l Conf. on Artificial Intelligence*. Citeseer.
- Johnson, R. A., Chawla, N. V., and Hellmann, J. J. (2012). Species distribution modeling and prediction: A class imbalance problem. In *Intelligent Data Understanding (CIDU), 2012 Conference on*, pages 9–16. IEEE.
- Kandybina, M. et al. (1977). The larvae of fruit-flies (diptera, tephritidae). keys to the fauna of the ussr no. 114. *The larvae of fruit-flies (Diptera, Tephritidae). Keys to the fauna of the USSR no. 114*.
- Kapatos, E. and Fletcher, B. (1984). The phenology of the olive fly, *dacus oleae* (gmel.)(diptera, tephritidae), in corfu. *Zeitschrift für Angewandte Entomologie*, 97(1-5):360–370.
- Kapoor, V. C. et al. (1993). *Indian fruit flies:(Insecta: Diptera: Tephritidae)*. International Science Publisher.
- Kearney, M. R., Isaac, A. P., and Porter, W. P. (2014a). microclim: Global estimates of hourly microclimate based on long-term monthly climate averages. *Scientific data*, 1(1):1–9.

- Kearney, M. R., Shamakhy, A., Tingley, R., Karoly, D. J., Hoffmann, A. A., Briggs, P. R., and Porter, W. P. (2014b). Microclimate modelling at macro scales: a test of a general microclimate model integrated with gridded continental-scale soil and weather data. *Methods in Ecology and Evolution*, 5(3):273–286.
- Keys, A., Mienotti, A., Karvonen, M. J., Aravanis, C., Blackburn, H., Buzina, R., Djordjevic, B., Dontas, A., Fidanza, F., Keys, M. H., et al. (1986). The diet and 15-year death rate in the seven countries study. *American journal of epidemiology*, 124(6):903–915.
- Kobayashi, K. and Salam, M. U. (2000). Comparing simulated and measured values using mean squared deviation and its components. *Agronomy Journal*, 92(2):345–352.
- Kounatidis, I., Papadopoulos, N., Mavragani-Tsipidou, P., Cohen, Y., Tertivanidis, K., Nomikou, M., and Nestel, D. (2008). Effect of elevation on spatio-temporal patterns of olive fly (*bactrocera oleae*) populations in northern greece. *Journal of Applied Entomology*, 132(9-10):722–733.
- Kubat, M., Matwin, S., et al. (1997). Addressing the curse of imbalanced training sets: one-sided selection. In *ICML*, volume 97, pages 179–186. Nashville, USA.
- Lasa, A., Miranda, J., Bulló, M., Casas, R., Salas-Salvadó, J., Larretxi, I., Estruch, R., Ruiz-Gutiérrez, V., and Portillo, M. P. (2014). Comparative effect of two mediterranean diets versus a low-fat diet on glycaemic control in individuals with type 2 diabetes. *European Journal of Clinical Nutrition*, 68(7):767–772.
- Law, A. M. (2015a). *Simulation Modeling and Analysis*, chapter Basic Simulation Modeling, pages 3–6. McGraw-Hill Education.
- Law, A. M. (2015b). *Simulation Modeling and Analysis*, chapter Basic Simulation Modeling, pages 70–72. McGraw-Hill Education.
- Li, X. and Wu, X. (2014). Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition. *CoRR*, abs/1410.4281.
- Lindsey, A. A. and Newman, J. E. (1956). Use of official wather data in spring time: temperature analysis of an indiana phenological record. *Ecology*, 37(4):812–823.
- Liu, Y., Chawla, N. V., Harper, M. P., Shriberg, E., and Stolcke, A. (2006). A study in machine learning from imbalanced data for sentence boundary detection in speech. *Computer Speech & Language*, 20(4):468–494.
- MacIsaac, K. D., Gordon, D. B., Nekludova, L., Odom, D. T., Schreiber, J., Gifford, D. K., Young, R. A., and Fraenkel, E. (2006). A hypothesis-based approach for identifying the binding specificity of regulatory proteins from chromatin immunoprecipitation data. *Bioinformatics*, 22(4):423–429.
- Malthus, T. R. (2018). *Essay on the Principle of Population*. Yale University Press.
- Martínez-González, M. Á., De la Fuente-Arrillaga, C., Nunez-Cordoba, J., Basterra-Gortari, F., Beunza, J. J., Vazquez, Z., Benito, S., Tortosa, A., and Bes-Rastrollo, M. (2008). Adherence to mediterranean diet and risk of developing diabetes: prospective cohort study. *Bmj*, 336(7657):1348–1351.

- Mayernik, M. S., Phillips, J., and Nienhouse, E. (2016). Linking publications and data: Challenges, trends, and opportunities. *D-Lib Magazine*, 22(5/6):11.
- Mazomenos, B. E. (1984). Effect of age and mating on pheromone production in the female olive fruit fly, *dacus oleae* (gmel.). *Journal of insect physiology*, 30(10):765–769.
- McCallum, A., Nigam, K., and Ungar, L. H. (2000). Efficient clustering of high-dimensional data sets with application to reference matching. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '00, page 169–178, New York, NY, USA. Association for Computing Machinery.
- McManus, K., Antinoro, L., and Sacks, F. (2001). A randomized controlled trial of a moderate-fat, low-energy diet compared with a low fat, low-energy diet for weight loss in overweight adults. *International journal of obesity*, 25(10):1503–1511.
- Mensink, R. P. and Katan, M. B. (1989). Effect of a diet enriched with monounsaturated or polyunsaturated fatty acids on levels of low-density and high-density lipoprotein cholesterol in healthy women and men. *New England Journal of Medicine*, 321(7):436–441.
- Merz, B. (1992). The fruit flies of the canary islands (diptera: Tephritidae). *Insect Systematics & Evolution*, 23(2):215–231.
- Miller, P., Lanier, W., and Brandt, S. (2001). Using growing degree days to predict plant stages. *Ag/Extension Communications Coordinator, Communications Services, Montana State University-Bozeman, Bozeman, MO*, 59717(406):994–2721.
- Miller, S. M., Funston, P. J., and Tambling, C. J. (2015). Growls: Lion (*panthera leo*) population growth simulation for small reserve management planning. *African Journal of Wildlife Research*, 45(2):169–177.
- Moore, I. (1962). Further investigations on the artificial breeding of the olive fly—*dacus oleae* gmel.—under aseptic conditions. *Entomophaga*, 7(1):53–57.
- Munro, H. K. et al. (1984). A taxonomic treatise on the dacidae (tephritoidea, diptera) of africa. *A taxonomic treatise on the Dacidae (Tephritoidea, Diptera) of Africa.*, (61).
- Murray, D. A., Clarke, M. B., and Ronning, D. A. (2013). Estimating invertebrate pest losses in six major australian grain crops. *Australian Journal of Entomology*, 52(3):227–241.
- Murray, J. (2007). *Mathematical Biology I. An Introduction*. Springer, 3rd edition.
- Mustafa, T. M. and Al-Zaghal, K. (1987). Frequency of *dacus oleae* (gmelin) immature stages and their parasites in seven olive varieties in jordan. *International Journal of Tropical Insect Science*, 8(2):165–169.
- Nardi, F., Carapelli, A., Dallai, R., Roderick, G. K., and Frati, F. (2005). Population structure and colonization history of the olive fly, *bactrocera oleae* (diptera, tephritidae). *Molecular Ecology*, 14(9):2729–2738.

- Neuenschwander, P. and Michelakis, S. (1978). The infestation of *dacus oleae* (gmel.)(diptera, tephritidae) at harvest time and its influence on yield and quality of olive oil in crete. *Zeitschrift für angewandte Entomologie*, 86(1-4):420–433.
- Neuenschwander, P., Michelakis, S., et al. (1979). Mcphail trap captures of *dacus oleae* (gmel.)(diptera, tephritidae) in comparison to the fly density and population composition as assessed by sondage technique in crete, greece. *Mitteilungen der Schweizerischen Entomologischen Gesellschaft*, 52(4):343–357.
- Nouri, H. et al. (2007). Integrated olive pest management in iran. *Integrated olive pest management in Iran*, 30(9).
- Ottman, M. (2008). Predicting wheat growth using the csm-cropsim-ceres-wheat crop model.
- Pappas, M., Broufas, G., Koufali, N., Pieri, P., and Koveos, D. (2011). Effect of heat stress on survival and reproduction of the olive fruit fly *bactocera* (*dacus*) *oleae*. *Journal of Applied Entomology*, 135(5):359–366.
- Pelekassis, C. E., Mourikos, P., and Bantzios, D. (1963). Preliminary studies of the field movement of the olive fruit fly (*dacus oleae* gmel.) by labelling a natural population with radioactive phosphorus (p 32). In *Radiation and Radioisotopes Applied to Insects of Agricultural Importance. Proceedings of the Symposium on the Use and Application of Radioisotopes and Radiation in the Control of Plant and Animal Insect Pests*.
- Philimis, P., Psimolophitis, E., Hadjiyiannis, S., Giusti, A., Perello, J., Serrat, A., and Avila, P. (2013). A centralised remote data collection system using automated traps for managing and controlling the population of the mediterranean (*ceratitis capitata*) and olive (*dacus oleae*) fruit flies. In *International Conference on Remote Sensing and Geoinformation of the Environment*, volume 8795, pages 8795 – 8795 – 8, Paphos, Cyprus.
- Planell, M. I. C., de la Torre, R., and Fitó, M. (2014). Scientific evidence of the benefits of virgin olive oil for human health. *Medicina balear*, 29(2):39–46.
- Platt, J. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines. Technical report, Microsoft Research. technical report msr-tr-98-14.
- Pommois, P., Brunetti, P., Bruno, V., Mazzei, A., Baldacchini, V., and Di Gregorio, S. (2006). Flysim: a cellular automata model of *bactrocera oleae* (olive fruit fly) infestation and first simulations. In *International Conference on Cellular Automata*, pages 311–320. Springer.
- Potamitis, I., Eliopoulos, P., and Rigakis, I. (2017a). Automated remote insect surveillance at a global scale and the internet of things. *Robotics*, 6(3):19.
- Potamitis, I., Rigakis, I., and Tatlas, N.-A. (2017b). Automated surveillance of fruit flies. *Sensors*, 17(1):110.
- Provost, F. and Fawcett, T. (2001). Robust classification for imprecise environments. *Machine learning*, 42(3):203–231.

- Pruess, K. P. (1983). Day-degree methods for pest management. *Environmental Entomology*, 12(3):613–619.
- Quinlan, J. R. (1996). Bagging, boosting, and c4.5. In *AAAI/IAAI, Vol. 1*, pages 725–730.
- Rempoulakis, P. and Nestel, D. (2012). Dispersal ability of marked, irradiated olive fruit flies [*bactrocera oleae* (rossi)(diptera: Tephritidae)] in arid regions. *Journal of applied entomology*, 136(3):171–180.
- Remund, U., Boller, E., Economopoulos, A., and Tsitsipis, J. (1976). Flight performance of *dacus oleae* reared on olives and artificial diet. *Zeitschrift für Angewandte Entomologie*, 82(1-4):330–339.
- Rice, R., Phillips, P., Stewart-Leslie, J., Sibbett, G., et al. (2003). Olive fruit fly populations measured in central and southern california. *California Agriculture*, 57(4):122–127.
- Rockwood, L. L. (2015a). *Introduction to population ecology*, chapter 2. Density dependent growth and intraspecific competition, pages 33–36. Wiley Blackwell, 2nd edition.
- Rockwood, L. L. (2015b). *Introduction to population ecology*, chapter 1. Density independent growth, pages 12–12. Wiley Blackwell, 2nd edition.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.
- Rosenblatt, F. (1961). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Spartan Books.
- Russel, S. and Norvig, P. (2009a). *Artificial intelligence: a modern approach*, chapter Learning from examples, pages 693–697. Pearson Education Limited London.
- Russel, S. and Norvig, P. (2009b). *Artificial intelligence: a modern approach*, chapter Learning Probabilistic Models, pages 802–829. Pearson Education Limited London.
- Sak, H., Senior, A. W., and Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling.
- Schapire, R. E. (2003). The boosting approach to machine learning: An overview. In *Nonlinear estimation and classification*, pages 149–171. Springer.
- Sevacherian, V., Stern, V., and Mueller, A. (1977). Heat accumulation for timing lygus control measures in a safflower-cotton complex. *Journal of Economic Entomology*, 70(4):399–402.
- Shafieiyou, E. (2014). Identification of micro-climates of isfahan city and its effect on air temperature, relative air humidity and reference crop evapotranspiration. In *Proceedings of the 3rd ScienceOne International Conference on Environmental Sciences, Dubai, United Arab Emirates*, pages 21–23.
- Shahrestani, M., Yao, R., Luo, Z., Turkbeyler, E., and Davies, H. (2015). A field study of urban microclimates in london. *Renewable energy*, 73:3–9.

- Shaked, B., Amore, A., Ioannou, C., Valdés, F., Alorda, B., Papanastasiou, S., Goldshtein, E., Shenderey, C., Leza, M., Pontikakos, C., Perdakis, D., Tsiligridis, T., Tabilio, M. R., Sciarretta, A., Barceló, C., Athanassiou, C., Miranda, M. A., Alchanatis, V., Papadopoulos, N., and Nestel, D. (2017). Electronic traps for detection and population monitoring of adult fruit flies (diptera: Tephritidae). *Journal of Applied Entomology*, 142(1-2):43–51.
- Shalev-Shwartz, S. and Ben-David, S. (2014a). *Understanding Machine Learning From Theory to Algorithms*, chapter 11. Model Selection and Validation, pages 116–120. Cambridge University Press.
- Shalev-Shwartz, S. and Ben-David, S. (2014b). *Understanding Machine Learning From Theory to Algorithms*, chapter 17. Multiclass, Ranking and Complex Performance Measures, pages 206–207. Cambridge University Press.
- Shant, P. et al. (1999). Olive (*olea europaea* wall.) cultivation in jammu and kashmir. *Scientific Horticulture*, 6:71–78.
- Silvestri, F. (1914). *Report of an expedition to Africa in search of the natural enemies of fruit flies (Trypanidae): With descriptions, observations and biological notes*. Number 3. Board of Agriculture and Forestry, Division of Entomology.
- Smith-Warner, S. A., Spiegelman, D., Yaun, S.-S., Adami, H.-O., Beeson, W. L., Van Den Brandt, P. A., Folsom, A. R., Fraser, G. E., Freudenheim, J. L., Goldbohm, R. A., et al. (2001). Intake of fruits and vegetables and risk of breast cancer: a pooled analysis of cohort studies. *Jama*, 285(6):769–776.
- Speranza, S., Bellocchi, G., and Pucci, C. (2004). Ipm trials on attract-and-kill mixtures against the olive fly *bactrocera oleae* (diptera tephritidae). *Bulletin of Insectology*, 57(2):111–115.
- Stabler, L. B., Martin, C. A., and Brazel, A. J. (2005). Microclimates in a desert city were related to land use and vegetation index. *Urban forestry & urban greening*, 3(3-4):137–147.
- Storniolo, C. and Moreno, J. (2016). Effect of extra virgin olive oil components on the arachidonic acid cascade, colorectal cancer and colon cancer cell proliferation. *Grasas y aceites*, 67(4):159.
- Suzanne Van Cooten, N., NWSFO New Orleans-Baton Rouge, S., DD, B., DJA, M., and DG, C. (2002). Identification of precipitation microclimates and rainfall trends across the lake pontchartrain basin of southeast louisiana.
- Throne, J. E. (1989). Effects of Noncatastrophic Control Technologies that Alter Life History Parameters on Insect Population Growth: A Simulation Study . *Environmental Entomology*, 18(6):1050–1055.
- Tirelli, P., Borghese, N. A., Pedersini, F., Galassi, G., and Oberti, R. (2011). Automatic monitoring of pest insects traps by zigbee-based wireless networking of image sensors. In *IEEE International Instrumentation and Measurement Technology Conference*, pages 1–5.
- Tomek, I. (1976). Two modifications of cnn. *IEEE Transactions on Systems, Man and Communications*, 6:769–772.

- Trichopoulou, A., Katsouyanni, K., Stuver, S., Tzala, L., Gnardellis, C., Rimm, E., and Trichopoulos, D. (1995). Consumption of olive oil and specific food groups in relation to breast cancer risk in greece. *JNCI: Journal of the National Cancer Institute*, 87(2):110–116.
- Tsitsipis, J. and JA, T. (1977). Effect of constant temperatures on the eggs of the olive fruit fly, *dacus oleae* (diptera, tephritidae).
- Tsitsipis, J. A. (1980). Effect of constant temperatures on larval and pupal development of olive fruit flies reared on artificial diet. *Environmental Entomology*, 9(6):764–768.
- Uhlir, P. (2015). The value of open data sharing. Accessed October 2018, https://www.earthobservations.org/documents/dsp/20151130_the_value_of_open_data_sharing.pdf.
- Valueva, M. V., Nagornov, N., Lyakhov, P. A., Valuev, G. V., and Chervyakov, N. I. (2020). Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Mathematics and Computers in Simulation*, 177:232–243.
- Verhulst, P.-F. (1838). Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.*, 10:113–126.
- Visioli, F., Bellomo, G., Montedoro, G., and Galli, C. (1995). Low density lipoprotein oxidation is inhibited in vitro by olive oil constituents. *Atherosclerosis*, 117(1):25–32.
- Vossen, P. (2014). Monitoring and control of olive fruit fly (olf) for oil production in california. Technical report, University of California Cooperative Extension. <http://cesonoma.ucanr.edu/files/203835.pdf>.
- Vossen, P., Varel, L., and Alexandra, D. (2004). Olive fruit fly. Technical report, University of California Cooperative Extension. <http://fruitsandnuts.ucdavis.edu/files/74133.pdf>.
- Voulgaris, S., Stefanidakis, M., Floros, A., and Avlonitis, M. (2013). Stochastic modeling and simulation of olive fruit fly outbreaks. *Procedia Technology*, 8:580–586.
- Wang, J., Xu, M., Wang, H., and Zhang, J. (2006). Classification of imbalanced data by using the smote algorithm and locally linear embedding. In *2006 8th international Conference on Signal Processing*, volume 3. IEEE.
- Wang, J.-n., Chen, X.-l., Hou, X.-w., Zhou, L.-b., Zhu, C.-D., and Ji, L.-q. (2016). Construction, implementation and testing of an image identification system using computer vision methods for fruit flies with economic importance (diptera: Tephritidae). *Pest Management Science*, 73(7):1511–1528.
- Wang, S. and Boulard, T. (2000). Predicting the microclimate in a naturally ventilated plastic house in a mediterranean climate. *Journal of agricultural engineering research*, 75(1):27–38.
- Wang, X.-G., Johnson, M. W., Daane, K. M., and Opp, S. (2009). Combined effects of heat stress and food supply on flight performance of olive fruit fly (diptera: Tephritidae). *Annals of the Entomological Society of America*, 102(4):727–734.

- Wilson, L. and Barnett, W. (1983). Degree-days: an aid in crop and pest management. *California Agriculture*, 37(1):4–7.
- Witten, I. H. and Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Wong, N. H. and Jusuf, S. K. (2010). Study on the microclimate condition along a green pedestrian canyon in singapore. *Architectural science review*, 53(2):196–212.
- Xiao, J., Tang, X., Li, Y., Fang, Z., Ma, D., He, Y., and Li, M. (2011). Identification of microrna precursors based on random forest with network-level representation method of stem-loop structure. *BMC bioinformatics*, 12(1):1.
- Yang, J., Zhao, J., Yu, H., Wang, Y., Wang, R., and Tang, L. (2012). Mathematical study of the effects of temperature and humidity on the morphological development of *pleurotus eryngii* fruit body. In *International Conference on Computer and Computing Technologies in Agriculture*, pages 312–323. Springer.
- Zalom, F. G., Steenwyk, R. A. V., Burrack, H. J., and Johnson, M. W. (2009). Olive fruit fly. integrated pest management for home gardeners and landscape professionals. Technical report, University of California Agriculture and Natural Resources. <http://ipm.ucanr.edu/PDF/PESTNOTES/pnolivefruitfly.pdf>.
- Zervas, G. (1982). Reproductive physiology of *dacus oleae* (gmelin) diptera tephritidae. comparison of wild and laboratory reared flies. *Geoponika (Greece)*.
- Zhang, Y., Mahrer, Y., and Margolin, M. (1997). Predicting the microclimate inside a greenhouse: an application of a one-dimensional numerical model in an unheated greenhouse. *Agricultural and forest meteorology*, 86(3-4):291–297.
- Zohary, D. (1993). The wild genetic resources of the cultivated olive. In *II International Symposium on Olive Growing 356*, pages 62–65.
- Zouros, E. and Krimbas, C. (1970). Frequency of female digamy in a natural population of the olive fruit fly *dacus oleae* as found by using enzyme polymorphism. *Entomologia experimentalis et applicata*, 13(1):1–9.