UNIVERSITY OF CRETE
DEPARTMENT OF CHEMISTRY

POSTGRADUATE PROGRAM IN CHEMISTRY

M.Sc. Thesis

LASER-INDUCED BREAKDOWN SPECTROSCOPY (LIBS) COMBINED WITH MACHINE LEARNING MODELS AND NEURAL NETWORKS ENABLES SCREENING AND CLASSIFICATION OF HARD TISSUES SUCH AS BONES AND TEETH, WITH POTENTIAL APPLICATIONS IN ARCHAEOLOGICAL SCIENCE.

LAMPROS SPANOS

SUPERVISOR: DEMETRIOS ANGLOS

HERAKLION 2021

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΧΗΜΕΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΗ ΧΗΜΕΙΑ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ

ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΚΑΙ ΤΑΞΙΝΟΜΗΣΗ ΟΣΤΙΚΩΝ ΚΑΙ ΟΔΟΝΤΙΚΩΝ ΙΣΤΩΝ ΜΕΣΩ ΣΤΟΙΧΕΙΑΚΗΣ ΑΝΑΛΥΣΗΣ ΜΕ ΦΑΣΜΑΤΟΣΚΟΠΙΑ ΠΛΑΣΜΑΤΟΣ ΛΕΙΖΕΡ, ΣΕ ΣΥΝΔΥΑΣΜΟ ΜΕ ΜΕΘΟΔΟΥΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ. ΠΙΘΑΝΕΣ ΕΦΑΡΜΟΓΕΣ ΣΤΗΝ ΑΡΧΑΙΟΛΟΓΙΚΗ ΕΡΕΥΝΑ.

ΛΑΜΠΡΟΣ ΣΠΑΝΟΣ

ΕΠΙΒΛΕΠΩΝ: ΔΗΜΗΤΡΙΟΣ ΑΓΓΛΟΣ

ΗΡΑΚΛΕΙΟ 2021

“If life seems jolly rotten

There's something you've forgotten

And that's to laugh and smile and dance and sing

When you're feeling in the dumps

Don't be silly chumps

Just purse your lips and whistle, that's the thing”

Monty Python

THESIS COMMITTEE MEMBERS (alphabetical)

Demetrios Anglos

Professor, Department of Chemistry, University of Crete (Supervisor)

Pavlos Pavlidis

Principal Researcher, Institute of Computer Science, Foundation for Research and Technology - Hellas

Apostolos Spiros

Associate Professor, Department of Chemistry, University of Crete

ΜΕΛΗ ΤΡΙΜΕΛΟΥΣ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ (αλφαβητικά)

Δημήτριος Άγγλος

Καθηγητής Τμήματος Χημείας Πανεπιστημίου Κρήτης (Επιβλέπων)

Παύλος Παυλίδης

Κύριος Ερευνητής Ινστιτούτου Πληροφορικής Ιδρύματος Τεχνολογίας και Έρευνας

Απόστολος Σπύρος

Αναπληρωτής Καθηγητής Τμήματος Χημείας Πανεπιστημίου Κρήτης

Ευχαριστίες

Με την ολοκλήρωση αυτής της εργασίας, ολοκληρώνεται πέραν των σπουδών μου και ένα μεγάλο κομμάτι της ζωής μου. Η διαδρομή μου τόσο στα προπτυχιακά αλλά και στα μεταπτυχιακά φοιτητικά χρόνια μου χάρισε εμπειρίες και φίλους που θα με καθορίζουν, ως προσωπικότητα και ως άνθρωπο, και θα με συντροφεύουν καθ' όλη την διάρκεια της ζωής μου. Ακόμη, δημιούργησε την μεγάλη μου αγάπη για την πόλη του Ηρακλείου, που παρ' ότι δεν είναι αξιοθαύμαστη εκ πρώτης όψεως, είναι σίγουρα μοναδική και όπως όλα τα αληθινά όμορφα πράγματα έχει την δική της ιδιαίτερη και μοναδική ομορφιά.

Η αλληλεπίδρασή, αυτά τα δυόμισι χρόνια, με διάφορους ανθρώπους καθόρισε σε μεγάλο βαθμό τόσο την πορεία προς την ολοκλήρωση της παρούσας εργασίας και κατ' επέκταση των σπουδών μου, όσο και στην διαμόρφωση της προσωπικότητάς μου. Θα ήταν, λοιπόν, παράληψη να μην εκφράσω τις ευχαριστίες μου στους ανθρώπους που συνέβαλαν σε αυτό, ο καθένας με τον δικό του μοναδικό τρόπο.

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα της εργασίας μου, Καθηγητή Δημήτριο Άγγλο, για την δυνατότητα που μου έδωσε να εξερευνήσω ένα πεδίο μακριά από τα μέχρι τότε ενδιαφέροντα μου, καθώς και για τις πολύτιμες συμβουλές που μου παρείχε καθ' όλη την διάρκεια των σπουδών μου. Επίσης, θα ήθελα να ευχαριστήσω το Ίδρυμα Τεχνολογίας κι Έρευνα και συγκεκριμένα το Ινστιτούτο Δομής και Λέιζερ για την υλικοτεχνική υποδομή που παρείχε καθώς και την χρηματοδότηση της ερευνητικής μου δουλειάς. Ακόμη, θα ήθελα να ευχαριστήσω τα υπόλοιπα μέλη της τριμελούς επιτροπής εξέτασης της εργασίας μου, τους Αναπληρωτή Καθηγητή Απόστολο Σπύρο και Κύριο Ερευνητή Παύλο Παυλίδη, που δέχθηκαν να εξετάσουν την δουλειά που εκπόνησα. Τον τελευταίο εκ των δύο θα ήθελα να τον ευχαριστήσω ιδιαιτέρως για την συνολική παρουσία του καθ' όλη την εκπόνηση της εργασίας, τις πολύτιμες συμβουλές του και την καθοδήγηση που μου παρείχε.

Θα ήθελα, επίσης, να σταθώ ιδιαίτερα στην παρουσία τριών ατόμων που με βοήθησαν σε μεγάλο βαθμό στην εκπόνηση της εργασίας και χωρίς την παρουσία τους θα ήταν πολύ διαφορετική η πορεία της δουλειάς μου. Πρόκειται για τους Δρ. Παναγιώτη Σιώζο, Δρ. Άγγελο Φιλιππίδη και Άγγελο Κοροπούλη, που με την καθοδήγηση, τις συμβουλές αλλά και κάθε μέσω το οποίο διέθεταν, με βοήθησαν και με δίδαξαν.

Προχωρώντας, θα ήθελα να σταθώ σε όλους τους ανθρώπους που με την παρουσία τους βοήθησαν και έκαναν πραγματικά ευχάριστο και απολαυστικό το εργασιακό περιβάλλον του εργαστηρίου. Ευχαριστώ τους Ειρήνη Μαλεγιαννάκη, Αντωνία Ιντζέ, Μαριλένα Κωνσταντίνου, Δημήτρη Χατζηγιάννη, Κώστα Χατζηγιαννάκη, Δρ. Γιώργο Τσερεβελάκη και Δρ. Όλγα Κοκκινάκη, τόσο για την βοήθεια που μου παρείχαν όσο και για την ευχάριστη και φιλική ατμόσφαιρα που μου χάρισαν. Ακόμη, δεν θα μπορούσα να παραλείψω τα μέλη του εργαστηρίου «Φωτονικής με Εφαρμογές στην Αγροδιατροφή και το Περιβάλλον», Γιώργο Σταυρακάκη, Μάνο Ορφανάκη και Νίκο Φραγκούλη για την συμβολή τους στο ευχάριστο εργασιακό περιβάλλον καθώς και

τις πολύτιμες συμβουλές που μου παρείχαν. Ενώ, δεν θα έπρεπε να ξεχάσω τον Δρ. Σπυρίδων Χαυλή για την αμέριστη βοήθειά του και τις συμβουλές και διευκρινήσεις που μου έδωσε όταν όλα δείχνανε να μην βγάζουν νόημα.

Η ιδιαιτερότητα της παρούσας εργασίας εντοπίζεται στο γεγονός πως για την πραγματοποίησή της ήταν απαραίτητη η συλλογή σκελετικών δειγμάτων. Πάνω στο κομμάτι αυτό, θα ήθελα να ευχαριστήσω τους Μάνο Τσιχλή, Κέλλυ Τσιχλή, Μήνα Λοΐζου και Σούλα Κουλουτμπάνη για την ανιδιοτελή παραχώρηση πλήθους δειγμάτων προς εξέταση, διότι χωρίς την συμβολή τους η παρούσα εργασία δεν θα είχε πραγματοποιηθεί. Επίσης, θα ήθελα να ευχαριστήσω τις Αθανασία Παπουτσή και Ιωάννα Πανδή για τα δείγματα που μου προσέφεραν, παρ' ότι αυτά δεν χρησιμοποιήθηκαν στο πλαίσιο της παρούσας εργασίας.

Κλείνοντας, θα ήθελα να αναφερθώ και σε ξεχωριστούς ανθρώπους που με την στήριξη τους και την καλή τους συντροφιά βοήθησαν και βοηθούν όχι μόνο στην παρούσα εργασία, αλλά και συνολικά στο να διαμορφωθώ και να γαλουχηθώ σαν άνθρωπος και να είναι απολαυστικότερη, χαρούμενη και όμορφη η καθημερινότητά μου. Αρχικά, θα ήθελα να ευχαριστήσω την οικογένειά μου, τους Σάββα Σπανό, Κέλλυ Μαρούτση, Δήμητρα Μαρούτση και Κωνσταντίνο Σπανό, για την αμέριστη αγάπη και βοήθεια που μου έχουν προσφέρει όλα αυτά τα χρόνια. Επίσης, ξεχωριστή αναφορά οφείλω να κάνω στην Μαρία Συσκάκη, που με το εκρηκτικό της ταπεραμέντο φροντίζει να μην είναι βαρετή καμία στιγμή της ημέρας μου. Τέλος, δεν θα μπορούσα να παραλείψω τους φίλους μου, Νίκο, Ντιντσέρ, Μιχαλάκη, Παναγιώτη, Μαρία, Γιώργο και όχι μόνο, που βρίσκονταν και βρίσκονται πάντα εκεί για να με βοηθούν και να με στηρίζουν, ο καθένας με τον δικό του ξεχωριστό τρόπο.

CURRICULUM VITAE

PERSONAL INFORMATION

NAME	Lampros Spanos
DATE OF BIRTH	20/04/1995
PLACE OF BIRTH	Athens, Greece
NATIONALITY	Greek
STUDIES	Chemistry
FAMILY STATUS	Unmarried
HOME ADDRESS	Pambotidos 72, Glyfada, Athens, Greece, 16561
CURRENT ADDRESS	Ved Andebakken 6 st th 2000, Frederiksberg, Denmark
PHONE NUMBER	+30 2109626497, +30 6936680335

EDUCATION

September 2013 – February 2018	**University of Crete** Department of Chemistry B.Sc. in Chemistry GPA: 7.36/10 (Very Good) Thesis Title: "*Synthesis of protein-perfluorinated polymer bioconjugates.*"
October 2018 – April 2021	**University of Crete** Department of Chemistry M.Sc. in Chemistry (Advanced Materials, Physical Chemistry) Thesis Title: "*Laser-induced breakdown spectroscopy (LIBS) combined with machine learning models and neural networks enables screening and classification of hard tissues such as bones and teeth, with potential applications in archaeological science.*"

WORKING EXPERIENCE

November 2017 - January 2018	Internship, Science Teaching at High School Level (Agios Myronas High School, Heraklion, Crete)
April 2018 - July 2018	Erasmus+ Internship Prof. G. Mantovani group, Molecular Therapeutics and Formulation Division, School of Pharmacy, Faculty of Science, University of Nottingham, UK
May 2021 - July 2021	Erasmus+ Internship Prof. N. Hatzakis group, Nano-Science Center, Department of Chemistry, University of Copenhagen, Denmark

MEMBERSHIPS & ACTIVITIES

2018	Association of Greek Chemists
May 2019	Member of the organizing committee of the "21st Postgraduate Students Conference on Chemistry" (Heraklion)

CONFERENCES

June 2018	20th Postgraduate Students Conference on Chemistry (Heraklion) Poster Presentation: Grafanaki E., Andrikopoulos N., Kalfaoglou S., Liarou E., Mandriotis P., Papageorgiou M., Skordalidis P., Spanos L., Dramountanis N., Theodorou A., Charalampidis G., Anastasaki A., Coutsolelos A., Haddleton D., Velonia K.. *"New Families of Functional Giant-Amphiphiles"*
September 2020	11th International Conference on Laser-Induced Breakdown Spectroscopy (Kyoto/ Web Conference) Poster Presentation: L. Spanos, P. Siozos, A. Koropoulis, N. Hausmann, M. Holst, P. Pavlidis, D. Anglos *"Laser Induced Breakdown Spectroscopy (LIBS) combined with machine learning methods or neural networks enables screening and classification of archaeological hard tissue remains"*

RESEARCH EXPERIENCE

March 2017 - February 2018	B. Sc. Research Thesis: *"Synthesis of protein-perfluorinated polymer bioconjugates"*, Laboratory of Synthetic Biomaterials, Department of Material Science and Technology, University of Crete
April 2018 - July 2018	Erasmus Internship Research: Project 1: *"Synthetic Glycopolymers"* Project 2: *"Novel route to RAFT chain-transfer agents starting from functional maleimides"* Prof. G. Mantovani group, Molecular Therapeutics and Formulation Division, School of Pharmacy, Faculty of Science, University of Nottingham, UK
December 2018 - April 2021	M. Sc. Thesis Research: "*Laser-induced breakdown spectroscopy (LIBS) combined with machine learning models and neural networks enables screening and classification of hard tissues such as bones and teeth, with potential applications in archaeological science.*", Laboratory of Applied Spectroscopy, Institute of Electronic Structure and Laser, Foundation for Research and Technology-Hellas

LANGUAGES

Languages	Greek (Native), English (B2)
Programming Languages	R, Python

ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ

ΑΤΟΜΙΚΑ ΣΤΟΙΧΕΙΑ

ΟΝΟΜΑΤΕΠΩΝΥΜΟ	Λάμπρος Σπανός
ΗΜΕΡΟΜΗΝΙΑ ΓΕΝΝΗΣΕΩΣ	20/05/1995
ΕΙΔΙΚΟΤΗΤΑ	Χημικός
ΟΙΚΟΓΕΝΕΙΑΚΗ ΚΑΤΑΣΤΑΣΗ	Άγαμος
ΔΙΕΥΘΥΝΣΗ ΚΑΤΟΙΚΙΑΣ	Παμβώτιδος 72, Άνω Γλυφάδα, Αττική, τ.κ.16561
ΠΡΟΣΦΑΤΗ ΔΙΕΥΘΥΝΣΗ	Ved Andebakken 6 st th 2000, Frederiksberg, Denmark
ΤΗΛΕΦΩΝΟ ΚΑΤΟΙΚΙΑΣ	2109626497-6936680335

ΣΠΟΥΔΕΣ

Σεπτέμβριος 2013 - Φεβρουάριος 2018	**Πανεπιστήμιο Κρήτης** Τμήμα Χημείας Βαθμός Πτυχίου: Λίαν Καλώς (7,36) Διπλωματική Εργασία: "*Synthesis of protein-perfluorinated polymer bioconjugates.*"
Οκτώβριος 2018 - Απρίλιος 2021	**Πανεπιστήμιο Κρήτης** Τμήμα Χημείας Μεταπτυχιακό Δίπλωμα Ειδίκευσης στη Χημεία (Κατεύθυνση: Προηγμένα Υλικά-Φυσικοχημεία) Διπλωματική Εργασία: *«Χαρακτηρισμός και ταξινόμηση οστικών και οδοντικών ιστών μέσω στοιχειακής ανάλυσης με φασματοσκοπία πλάσματος λέιζερ, σε συνδυασμό με μεθόδους μηχανικής μάθησης και νευρωνικά δίκτυα. Πιθανές εφαρμογές στην αρχαιολογική έρευνα.»*

ΕΡΓΑΣΙΑΚΗ ΕΜΠΕΙΡΙΑ

Νοέμβριος 2017 - Ιανουάριος 2018	Πρακτική Άσκηση Διδακτικής Θετικών και Τεχνολογικών Επιστημών (Γενικό Λύκειο Αγίου Μύρωνα, Ηράκλειο Κρήτης)
Απρίλιος 2018 - Ιούλιος 2018	Πρακτική Άσκηση Erasmus+ Prof. G. Mantovani group, Molecular Therapeutics and Formulation Division, School of Pharmacy, Faculty of Science, University of Nottingham, UK
Μάιος 2021 - Ιούλιος 2021	Πρακτική Άσκηση Erasmus+ Prof. N. Hatzakis group, Nano-Science Center, Department of Chemistry, University of Copenhagen, Denmark

ΕΠΙΣΤΗΜΟΝΙΚΕΣ ΣΥΝΔΡΟΜΕΣ & ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ

2018	Ένωση Ελλήνων Χημικών
Μάιος 2019	Μέλος της επιτροπής οργάνωσης του 21ο Συνεδρίου Μεταπτυχιακών Φοιτητών Χημείας (Ηράκλειο)

CONFERENCES

Ιούνιος 2018	20^{ο} Συνέδριο Μεταπτυχιακών Φοιτητών Χημείας (Ηράκλειο) Αναρτημένης Ανακοίνωσης: Grafanaki E., Andrikopoulos N., Kalfaoglou S., Liarou E., Mandriotis P., Papageorgiou M., Skordalidis P., Spanos L., Dramountanis N., Theodorou A., Charalampidis G., Anastasaki A., Coutsolelos A., Haddleton D., Velonia K.. *"New Families of Functional Giant-Amphiphiles"*
Σεπτέμβριος 2020	11th International Conference on Laser-Induced Breakdown Spectroscopy (Kyoto/ Web Conference) Αναρτημένης Ανακοίνωσης: L. Spanos, P. Siozos, A. Koropoulis, N. Hausmann, M. Holst, P. Pavlidis, D. Anglos *"Laser Induced Breakdown Spectroscopy (LIBS) combined with machine learning methods or neural networks enables screening and classification of archaeological hard tissue remains"*

ΕΡΕΥΝΗΤΙΚΗ ΔΡΑΣΤΗΡΙΟΤΗΤΑ

Μάρτιος 2017 - Φεβρουάριος 2018	Ερευνητική Πτυχιακή Εργασία: *"Synthesis of protein-perfluorinated polymer bioconjugates"*, Εργαστήριο Συνθετικών Βιοϋλικών, Τμήμα Επιστήμης και Τεχνολογίας Υλικών, Πανεπιστήμιο Κρήτης
Απρίλιος 2018 - Ιούλιος 2018	Ερευνητική Πρακτική Άσκηση Erasmus+: Έργο 1: *"Synthetic Glycopolymers"* Έργο 2: *"Novel route to RAFT chain-transfer agents starting from functional maleimides"* Prof. G. Mantovani group, Molecular Therapeutics and Formulation Division, School of Pharmacy, Faculty of Science, University of Nottingham, UK
Δεκέμβριος 2018 - Απρίλος 2021	Ερευνητική Διπλωματική Εργασία: *«Χαρακτηρισμός και ταξινόμηση οστικών και οδοντικών ιστών μέσω στοιχειακής ανάλυσης με φασματοσκοπία πλάσματος λέιζερ, σε συνδυασμό με μεθόδους μηχανικής μάθησης και νευρωνικά δίκτυα. Πιθανές εφαρμογές στην αρχαιολογική έρευνα.»*, Εργαστήριο Εφηρμοσμένης Φασματοσκοπίας, Ινστιτούτο Ηλεκτρονικής Δομής και Λέιζερ, Ίδρυμα Τεχνολογίας κι Έρευνας

ΛΟΙΠΕΣ ΓΝΩΣΕΙΣ

Ομιλούμενες Γλώσσες	Ελληνικά (Μητρική Γλώσσα), Αγγλικά (B2)
Γλώσσες Προγραμματισμού	R, Python

Abstract

The excavation of mass graves and sites of accidents or natural disasters, which reveal numbers of hard tissue remains originating from multiple individuals, is usual in archaeology and forensic science. The discrimination of the individuals and the classification of their remains is useful for revealing the identity, as well as information about life and death of those individuals. However, poor preservation makes the task of discrimination/classification extremely difficult and time consuming, using conventional methods based on morphological characteristics or DNA analysis. Thus, the need for simple, direct and cost-effective analysis of hard tissue remains, with minimal damage to the artifacts, has emerged.

In the current work, Laser Induce Breakdown Spectroscopy (LIBS), combined with Machine Learning algorithms and a simple Artificial Neural Network, were employed for the discrimination and classification of hard tissue remains. Several bone fragments and teeth were studied, using a LIBS microscopy setup (micro-LIBS) for data collection, while Machine Learning algorithms and a Neural Network were used for data analysis.

Micro-LIBS is a micro-destructive, fast and transferable method, with high spatial resolution (around 50μm/spot) that enables analysis of the surface or the cross-section of samples, with little or no sample preparation, providing massive amounts of data in little time. Thus, it is a suitable technique to be combined with machine learning algorithms for the analysis of the collected data.

Hydroxyapatite ($Ca_5(PO_4)_3OH$) is the main component of both bones and teeth, while proteinaceous materials (mainly collagen) and water, in different ratios, complete the hard tissue matrix. Magnesium (Mg), Strontium (Sr) and Barium (Ba) can replace Calcium (Ca) in metabolic processes and thus can replace it in hydroxyapatite's crystal. Spectral emission lines from biogenic elements in the remains are observed across the spectral range used (200 - 660 nm). Hence, the data collected provide significant information to the algorithms employed.

Machine Learning and Neural Networks enable computers to learn from experience following a similar process with several living organisms. This process is based on pattern recognition on given data, improving future decisions and giving computers the ability to learn without being explicitly programmed. This pattern recognition on LIBS data is the aim of this work. The comparison of four different methods (k-Nearest Neighbors, Random Forest, Support Vector Machine, Artificial Neural Network) with gradual complexity, after parameter tuning and feature selection, provided the best behaved model to achieve the requested task.

Artificial Neural Network had significantly better results compared to the rest of the models used, while the selection of specific spectral areas corresponding in spectral lines from biogenic elements increased the resulting classification accuracy. The

achieved classification varied from decent to excellent, giving a good classification accuracy regardless of the used data.

Concluding, the present work is an attempt for development of a fast, accurate and easily accessible and applicable methodology for the discrimination and classification of hard tissue remains, based on the analysis of LIBS data using machine learning models.

Key words: Laser Induce Breakdown Spectroscopy (LIBS), micro-LIBS, Machine Learning, Neural Networks, Hard Tissue Remains, Archaeology, Forensic Science

Περίληψη

Διαχρονικά η αρχαιολογική και εγκληματολογική έρευνα, έρχονται αντιμέτωπες με την ανασκαφή και αποκάλυψη μαζικών ταφικών χώρων που περιέχουν πλήθος οστικών και οδοντικών υπολειμμάτων μεγάλου αριθμού υποκειμένων. Για τη βέλτιστη δυνατή μελέτη των ευρημάτων αυτών και την εξαγωγή συμπερασμάτων τόσο για την ζωή όσο και για τα αίτια θανάτου των υποκειμένων είναι αναγκαίος ο διαχωρισμός και η κατηγοριοποίηση των υπολειμμάτων σε υποκείμενα, για την ανακατασκευή τμημάτων ή και ολόκληρου του σκελετού. Η σύγχρονη οστεολογία και οστεοαρχαιολογία δίνουν συχνά λύση, χρησιμοποιώντας μεθόδους βασιζόμενες σε μορφολογικά χαρακτηριστικά των οστών. Η ταχύτητα και η ακρίβεια των μεθόδων αυτών είναι άμεσα εξαρτώμενες από την κατάσταση διατήρησης των ευρημάτων με συνέπεια η διαδικασία χαρακτηρισμού να είναι, σε αρκετές περιπτώσεις, χρονοβόρα και το αποτέλεσμα ανακριβές. Για τον λόγο αυτό, η ανάπτυξη εναλλακτικών, εύκολα εφαρμόσιμων μεθόδων που αυξάνουν τόσο την ταχύτητα όσο και την ακρίβεια του διαχωρισμού, απλοποιώντας τις παραπάνω διαδικασίες, θα ήταν δυνατό να λειτουργήσει ευεργετικά τόσο στην αρχαιολογική όσο και στην εγκληματολογική επιστημονική έρευνα.

Στην παρούσα εργασία, μελετήθηκε η εφαρμογή της φασματοσκοπίας πλάσματος επαγόμενου από λέιζερ (Laser Induced Breakdown Spectroscopy, LIBS), σε συνδυασμό με την επεξεργασία των δεδομένων με μεθόδους μηχανικής μάθησης (Machine Learning) και με τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks) με σκοπό το διαχωρισμό και την κατηγοριοποίηση σε υποκείμενα, οστικών και οδοντικών υπολειμμάτων. Μελετήθηκαν ανθρώπινα και ζωικά δείγματα, με χρήση διάταξης μικροσκοπίας LIBS (micro-LIBS) για τη συλλογή των φασματικών δεδομένων, ενώ η επεξεργασία και ανάλυση αυτών πραγματοποιήθηκε μέσω των μεθόδων μηχανικής μάθησης καθώς και με τη χρήση ενός απλού τεχνητού νευρωνικού δικτύου.

Με τη διάταξη micro-LIBS πραγματοποιείται ανάλυση σε τομές σκελετικών υπολειμμάτων, διαστάσεων μερικών cm^2, με υψηλή χωρική ανάλυση (περίπου 50 μm/σημείο) ενώ επιτυγχάνεται ταχεία συλλογή μεγάλου αριθμού φασματικών δεδομένων, απαραίτητη προϋπόθεση για τη βέλτιστη εφαρμογή των αλγορίθμων που χρησιμοποιήθηκαν για την ανάλυση των δεδομένων. Επίσης η τεχνική LIBS είναι μικροκαταστρεπτική, διεξάγεται απευθείας στην επιφάνεια του προς ανάλυση αντικειμένου και δεν απαιτεί προεπεξεργασία του δείγματος, γεγονός που την καθιστά κατάλληλη για τη στοιχειακή ανάλυση ευαίσθητων δειγμάτων και αντικειμένων πολιτιστικής κληρονομίας.

Τα οστά και τα δόντια των οργανισμών, έχουν ως κύριο συστατικό τον υδροξυαπατίτη ($Ca_5(PO_4)_3OH$), ενώ περιέχουν και μικρά ποσοστά οργανικών ενώσεων, όπως το κολλαγόνο, καθώς και νερό. Παράλληλα, στοιχεία όπως το μαγνήσιο (Mg), το στρόντιο (Sr) και το βάριο (Ba) ενσωματώνονται στο πλέγμα ως προσμίξεις, αντικαθιστώντας το ασβέστιο (Ca). Στη φασματική περιοχή στην οποία

πραγματοποιήθηκαν οι μετρήσεις LIBS (200 -660 nm) παρατηρούνται κορυφές εκπομπής όλων των προαναφερθέντων στοιχείων, γεγονός που συνεπάγεται την ύπαρξη επαρκούς πληροφορίας για τη στοιχειακή σύσταση των μελετώμενων ιστών. Αυτό καθιστά τα δεδομένα που συλλέγονται κατάλληλα για την άντληση αυξημένης πληροφορίας από τους αλγορίθμους που χρησιμοποιήθηκαν για την ανάλυση.

Η μηχανική μάθηση και τα τεχνητά νευρωνικά δίκτυα, βασίζονται στη χρήση αλγορίθμων και στατιστικών μοντέλων για την εκπαίδευση/εκμάθηση ενός υπολογιστή βάσει παραδειγμάτων και την εύρεση μοτίβων, χωρίς ανθρώπινη παρέμβαση, με τρόπο εκμάθησης παρόμοιο με αυτόν του ανθρώπινου εγκεφάλου καθώς και άλλων ζώντων οργανισμών. Η αναζήτηση ανάλογων μοτίβων στα καταγραφόμενα φάσματα σκελετικών υπολειμμάτων, που μπορούν να βοηθήσουν σε ικανοποιητική κατηγοριοποίησή τους, αποτελεί και τον βασικό σκοπό της εργασίας. Οι μέθοδοι που διερευνήθηκαν (k-Nearest Neighbors, Random Forest, Support Vector Machine, Artificial Neural Network) διαθέτουν κλιμακούμενη πολυπλοκότητα και απαιτούν διαφορετική υπολογιστική ισχύ. Πραγματοποιήθηκε συγκριτική ανάλυση των αποτελεσμάτων μεταξύ των τεσσάρων μεθόδων με σκοπό την εύρεση των βέλτιστων παραμέτρων με τις οποίες μπορεί να επιτευχθεί η ανάπτυξη ενός αξιόπιστου μοντέλου διαχωρισμού και κατηγοριοποίησης σκελετικών υπολειμμάτων. Ταυτόχρονα ερευνήθηκε ο περιορισμός των χρησιμοποιούμενων φασματικών περιοχών για την εκ νέου βελτιστοποίηση καθώς και την επιτάχυνση της ανάλυσης των δεδομένων.

Από τα αποτελέσματα της ανάλυσης προέκυψε η υπεροχή του νευρωνικού δικτύου, συγκριτικά με τους υπόλοιπους αλγορίθμους, καθώς και η επιλογή συγκεκριμένων φασματικών περιοχών για την επίτευξη του βέλτιστου διαχωρισμού και κατηγοριοποίησής των δεδομένων. Με την χρήση των προαναφερθέντων, η ακρίβεια κατηγοριοποίησης των τεσσάρων διαφορετικών συνόλων δεδομένων που εξετάστηκαν κυμάνθηκε από ικανοποιητική έως εξαιρετική, επιτυγχάνοντας ένα τουλάχιστον επαρκές ποσοστό διαχωρισμού ανεξαρτήτως αυτών.

Συνοψίζοντας, στο πλαίσιο της παρούσας εργασίας μελετήθηκε η ανάπτυξη μιας ταχείας, προσβάσιμης και με ικανοποιητική ακρίβεια μεθοδολογίας διαχωρισμού και ταξινόμησης οστικών και οδοντικών υπολειμμάτων, βασιζόμενη στην επεξεργασία φασματικών δεδομένων εκπομπής πλάσματος επαγόμενου από λέιζερ, με τη χρήση μεθόδων μηχανικής μάθησης και νευρωνικά δίκτυα, με πιθανή εφαρμογή σε αρχαιολογικές μελέτες.

Λέξεις κλειδιά: Φασματοσκοπίας Πλάσματος Επαγόμενου από Λέιζερ*, Μηχανική Μάθηση, Νευρωνικά Δίκτια, Σκελετικά Υπολείμματα,* Οστικά *Υπολείμματα*, Οδοντικά *Υπολείμματα, Αρχαιολογία, Εγκληματολογία*

Table of Contents

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
CCD	Charge-Coupled Device
KNN	K Nearest Neighbors
LA-ICP-MS	Laser Ablation Inductive Coupled Plasma Mass Spectroscopy
LIBS	Laser Induced Break Down Spectroscopy
ML	Machine Learning
PCA	Principal Component analysis
RF	Random Forest
SD	Standard Deviation
SVM	Support Vector Machine
SVMP	Polynomial Support Vector Machine
SVMR	Radial Support Vector Machine

1 Hard Tissue Remains

1.1 Excavations and Hard Tissues in Archaeology

Archaeology and forensic science are both well-established disciplines, focusing, among others, on the scientific study of remains of past human lives and activities and the investigation of crime scenes, respectively. (Darvill, 2019; Daniel, 2019; Giannelli, 2006) Their common ground is found in forensic archaeology and anthropology that investigate archaeological and forensic excavations. (Hunter, 2009) Hard tissue remains (e. g. bones, teeth) from animals and/or humans constitute important and often key findings in many excavation sites. Considering the large number of excavations going on all around the globe every year, studying and analyzing biological remains can provide useful information. (Villagran et al., 2009; Samek et al., 2001; Darvill, 2019) For example, excavated hard tissue remains may be useful for revealing the identity, as well as the way of life and death of individuals they belong to and as a result they are important not only in the scientific and humanitarian context but also they may have legal implications. (Moncayo et al., 2014; Hunter et al., 2001, Pietrusewsky, 2007, Kumar & Sharma, 2018) Thus, the discrimination among individuals and the identification of the remains, is very important for the progress of scientific research, especially in forensic archaeological studies, which face the task of excavating sites of accidents, natural disasters and mass graves, with remains by multiple individuals (Figure 1.1). (Moncayo et al., 2014; Crossland, 2000, Samek et al., 2001; Lindley, 1977)

Figure 1.1: *Hard tissue remains of multiple individuals in Igreja de São João Evangelista - Igreja dos Lóios, Evora, Portugal.*

In addition to the large amounts of hard tissue remains that can be found in an excavation, it is also their condition that can vary greatly due to a number of reasons, which adds to the complexity of the problem. Individuals can be found either as complete skeletons, or as dismembered assemblages or even fragmented and poorly preserved as a consequence of several factors, for example, scavenging animals,

scattering and burial because of agricultural activity or downslides, or disturbance by local foot traffic. In particular, in the case of mass graves, victims may have been intentionally incinerated or dismembered by use of explosives to prevent identification. (Haglund, Connor & Scott, 2001; Hunter et al., 2001)

In order to achieve identification, discrimination between individuals and classification of the fragmented hard tissues, is necessary and several methods are used, relying on multidisciplinary expertise and various techniques of physical and chemical analysis. (Moncayo et al., 2014; Hunter et al., 1994) Some of the most frequently used techniques are forensic pathology, forensic odontology and DNA analysis, while micromorphology of the tissues can be very useful, too. (Moncayo et al., 2014; Villagran et al., 2009, Samek et al., 2001) Considering that health and dental records may not exist and DNA availability might be limited, especially for archaeological researches, those methods are unable to provide clear information. (Moncayo et al., 2014) Thus, the need for simple, direct and cost-effective analysis of hard tissue remains, preferably with non- or minimally invasive means, is obvious. In recent years, several projects based on the analysis of hard tissue remains using laser sampling techniques have been reported. For example, LA-ICP-MS and LIBS combined with chemometrics and machine learning algorithms, are the most commonly non-DNA analytical techniques used for supplementing forensic archaeological studies. (Moncayo et al., 2014; Kumar & Sharma, 2018; Rehse, Salimnia & Miziolek, 2012; Siozos et al., 2021, Samek et al., 2001, Castro et al., 2010)

1.2 Structural and Chemical Characteristics of Hard Tissues

Structural characteristics of bones and teeth are quite different. Both of them consist of several types of tissues that differ from one another, either in morphology and/or in chemical composition. Almost all types of bones consist of two different tissue types (Figure 1.2). On the external layers, a lamellar tissue forms a smooth surface, which is called cortical bone and it is compact and bulky, giving the bone its shape. On the inner part, a spongy or cancellous tissue type, called trabecular bone, forms a fine network of thin bone beams. This spongy and flimsy tissue fills the inner parts of the bone, allowing blood and bone marrow to flow while it also reduces the total mass of the bone. The surface of trabecular bone is significantly larger and its regeneration is much faster, than that of the cortical bone. (Castro et al., 2010; Lynnerup & Klaus, 2019) As a result, trabecular bone is more sensitive to diagenetic changes and burial contamination, hence it is less valuable especially in archaeological research which deals with samples that are long buried even for centuries. (Castro et al., 2010)

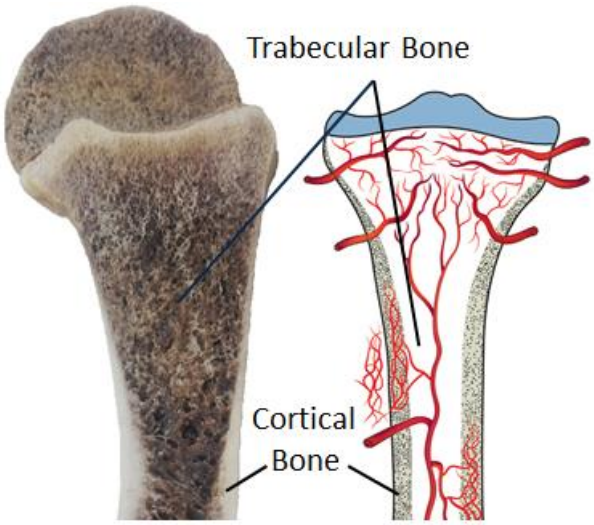

Figure 1.2: *The outside surface in bones is covered by a compact and bulky tissue, the cortical bone, which gives the shape and strength to the bones. The inner part of each bone consists of a spongy fine network of thin bone beams that allow blood to flow and minimizes bone mass. This tissue form is called trabecular bone. (Lynnerup & Klaus, 2019)*

Teeth consist of two different sections, the crown and the root, and four distinct tissues, enamel, dentine, cementum and plum (Figure 1.3). The crown of a tooth is the exposed part of it, outside of the dental arch, while the root is inside the dental arch. Plum is the only non-calcified tissue of a tooth, occupying the plum cavity and contains cells, nerves fibers, blood and lymph vessels. The plum cavity is encircled by a calcified tissue which is called dentine, forming the main body of the tooth. Surrounding dentine, the two different areas of the tooth, crown and root, differ in their composition. The outer layers of the crown are mainly formed by enamel, the hardest calcified tissue in the human body with high concentration of mineral matter, protecting the inner layers of the tooth. On the other hand, the outer layers of the root are formed by cementum, a hard tissue very similar to dentine. (Türp & Alt, 1998; Castro et al., 2010) Because of their excess, dentine and enamel are the calcified tissues that are used for analytical purposes. The regeneration dentine is subject to makes it useful for probing more recent exposure in environmental changes. At the same time, its less mineralized and softer structure makes it vulnerable in degradation and burial contamination. (Castro et al., 2010) Enamel is the tissue most resistant to diagenesis, compared with the rest of the body's hard tissues, reflecting more accurately the elemental composition for the studied organism. (Klepinger, 1984, Castro et al., 2010) Thus, in archaeological studies enamel is the preferable tissue for determining elemental and isotopic composition, because of its hardness and its reduced probability for being affected by diagenesis and burial contamination, followed by the cortical bone tissue. (Castro et al., 2010)

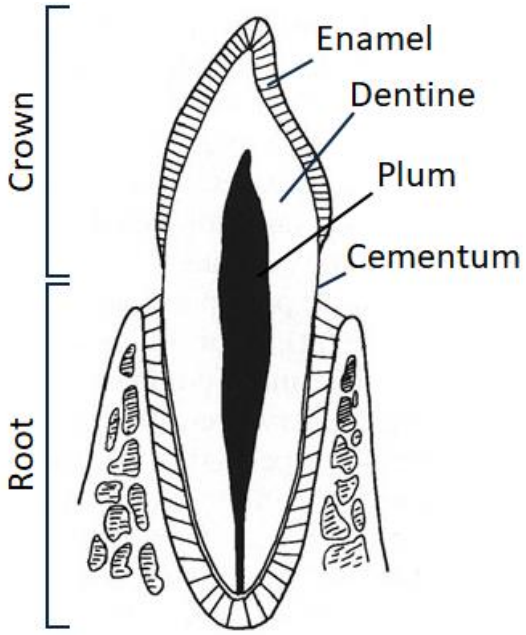

Figure 1.3: *The main body of the tooth is formed by dentine. Inside dentine is the plum cavity, where the non-calcified tissue of the tooth, nerves and organic matter, is well protected. The outer surface of it is separated in two different parts. In the crown, the hardest tissue of the body, enamel, forms protective layers on the surface of the tooth, giving to it its remarkable strength. In the root, a softer and very similar with dentine tissue, cementum, covers the outer layer of it. (Türp & Alt, 1998)*

The chemical characteristics of bones and teeth are very similar. The inorganic component in both hard tissues is mainly hydroxyapatite ($Ca_5(PO_4)_3OH$), a calcium phosphate biomineral, which is more frequently encountered in its hexagonal crystal structure form. (Castro et al., 2010; Ma & Liu, 2009) Other elements that can be found in those tissues in minor and trace concentration levels, include Magnesium (Mg), Strontium (Sr), Sodium (Na), Barium (Ba), Iron (Fe), Zinc (Zn), Copper (Cu), Manganese (Mn), Lead (Pb) and many more. (Castro et al., 2010; Kasem et al., 2014). Magnesium and strontium are above and below calcium in group 2 of the periodic table, sharing many of its chemical properties and having the ability to replace it in the course of the metabolic processes occurring in living organisms and thus to replace it in the hydroxyapatite crystal. (Klepinger, 1984) The majority of magnesium and strontium in animal bodies are located in the hard tissues. In particular, 99% of strontium and 60-65% of magnesium are located in the animals' skeleton. The higher levels of Sr and Mg in plants result in differences in the values of concentration ratios such as Sr/Ca and Mg/Ca between herbivores and carnivores. Furthermore, differences can be observed either in omnivorous organisms depending on the availability of meat and vegetation in their diet, or in herbivores depending on the vegetation that is preferred for consumption. Thus, studying the Sr/Ca and Mg/Ca ratios can facilitate the reconstruction of dietary habits of several organisms. (Klepinger, 1984; Kasem et al., 2014) Similarly with the inorganic phase, the organic phase of both bones and teeth is very similar, mainly consists of type I collagen. The combination of those two materials

with the presence of water, forms a composite with remarkable mechanical properties. (Samek, 2001; Currey, 2008)

In particular, the composition of bones varies according to their type. Hydroxyapatite constitutes 50% to 60% of the bone, collagen 20% and water 15% to 20%. Other components of the bone matrix include carbonates, phosphates and proteins amounting to approximately 7% of the total bone mass. Teeth differ significantly as regards the percentage of hydroxyapatite they contain. (Samek, 2001) As already mentioned, enamel is the hardest calcified tissue in the human body composed of approximately 95% of hydroxyapatite, 4% of water and just 1% of organic matter. Dentine and cementum are much softer than enamel having a significantly lower fraction of hydroxyapatite. Dentine consists of approximately 70% of hydroxyapatite and 20% of organic matter, mostly collagen fibers. The remaining 10% of dentine is water. (Castro et al., 2010; Samek, 2001; Türp & Alt, 1998) Cementum is similar to dentine, with very small changes differentiating those tissues. Cementum's main component is hydroxyapatite that constitutes 61% of it, while the rest is formed by 27% of organic matter and 12% of water. (Türp & Alt, 1998)

1.3 Samples and Sample Sets

In the present work, two main categories of samples were used. The first one was archaeological bone fragments and the other one was modern teeth.

1.3.1 Archaeological Bone Fragments

This sample set has already been described and analyzed in previous works (Siozos et al., 2021). The bone fragments originate from archaeological excavations of burials at the Cross Street Unitarian Chapel in Manchester (United Kingdom) and the burial period is considered to have been between the 18th and 19th century. All skeletons were at least 50% complete and their state of preservation ranged from "good" to "extremely poor". The sample set consisted of seven bone fragments originating from five different human individuals. Two of them were assigned two bone fragments each, while the rest just one. (Table 1.1) Each sample had been washed with water to remove dirt and impurities and air dried following excavation.

Table 1.1: *Archaeological bone fragments. Individuals and samples*

Individual	Number of Samples	Name of Samples	Figure 1.4
Ind_1	2	SK 3.25 Femur SK 3.25 Scapula	a b
Ind_2	2	SK 3.50 Clavicle SK 3.50 Femur	c d
Ind_3	1	SK 4.36 Ilium	e
Ind_4	1	SK 4.37 Tibia	f
Ind_5	1	SK 4.45 Humerus	g

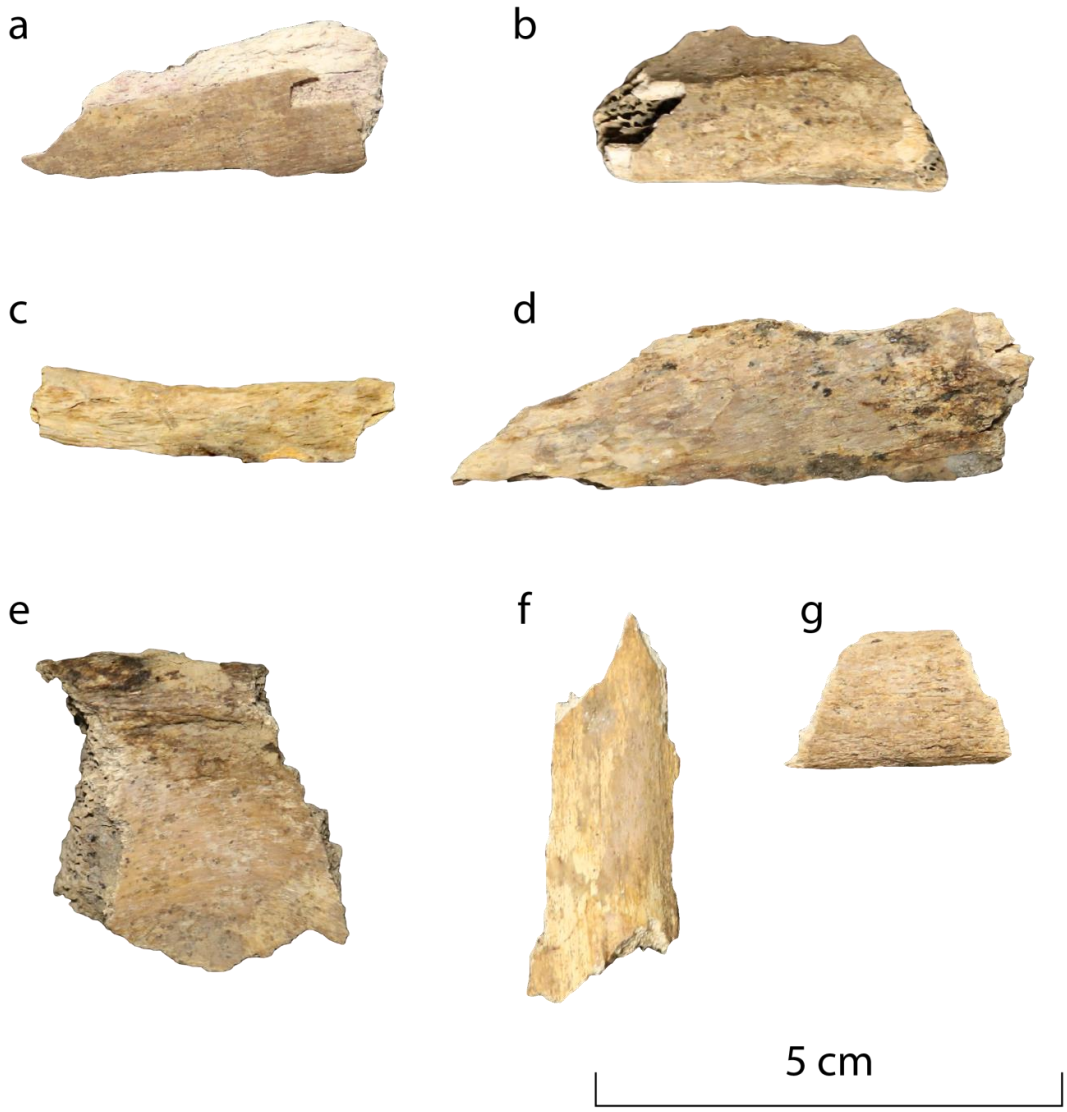

Figure 1.4: *a) SK 3.25 Femur, b) SK 3.25 Scapula, c) SK 3.50 Clavicle, d) SK 3.50 Femur, e) SK 4.36 Ilium, f) SK 4.37 Tibia, g) SK 4.45 Humerus. (Siozos et al., 2021)*

1.3.2 Modern Teeth Samples

Two different teeth sample sets were collected and tested for the purposes of this work. A set of adult and baby human teeth composed of 20 samples coming from 16 different individuals and a set of domestic sheep teeth composed of 11 samples from 5 individuals. Each individual in the human set was represented by at least one tooth, two of them were assigned two teeth each, while for one individual three different teeth samples were available (Table 1.2). The majority of samples were collected from dentists in Attica and Crete (Greece) and the rest were donated directly by their owners. Most of the teeth had been previously soaked in an aqueous bleach solution in order to be cleaned from any remains of soft tissue and blood. After cleaning, each sample was washed with water and air dried. The second sample set consisted of teeth coming from 5 different domestic sheep individuals. Only one individual had just one sample, while the rest had two or three samples each. (Table 1.3) All samples had been

collected from already cooked animals and they had been washed with water and air dried, prior to analysis.

All samples from both domestic sheep and human sets were cut with a low speed saw (Buehler Isomet Low Speed Saw) and the cross sections were evened to sleek using sandpapers, for the improvement of the experimental process. Measurements for each sample were taken in the enamel's cross section, avoiding any contamination and impurities of the outer surface.

Table 1.2: *Human modern teeth. Individuals and samples.*

Individual	Age	Number of Samples	Name of Samples	Figure 1.5
Ds_1	Baby	2	Ds_1_tooth_1 Ds_1_tooth_2	a b
Dx_1	Adult	1	Dx_1_tooth_1	c
Gr_1	Adult	2	Gr_1_tooth_1 Gr_1_tooth_2	d e
Kg_1	Adult	1	Kg_1_tooth_1	f
Ko_1	Baby	1	Ko_1_tooth_1	g
Ls_1	Baby	3	Ls_1_tooth_3 Ls_1_tooth_4 Ls_1_tooth_5	h i j
Mn_1	Adult	1	Mn_1_tooth_1	k
Mn_2	Adult	1	Mn_2_tooth_1	l
Mn_3	Adult	1	Mn_3_tooth_1	m
Mn_4	Adult	1	Mn_4_tooth_1	n
Mn_5	Adult	1	Mn_5_tooth_1	o
Mn_6	Adult	1	Mn_6_tooth_1	p
Mn_7	Adult	1	Mn_7_tooth_1	q
Mn_8	Adult	1	Mn_8_tooth_1	r
Sl_1	Adult	1	Sl_1_tooth_1	s
Sl_2	Adult	1	Sl_2_tooth_1	t

Table 1.3: *Domestic sheep cooked teeth. Individuals and samples.*

Individual	Number of Samples	Name of Samples	Figure 1.6
Lmp_1	2	Lmp_1_tooth_2_uj_1 Lmp_1_tooth_3_uj_1	a
Lmp_2	1	Lmp_2_tooth_3_lj_1	b
Lmp_3	2	Lmp_3_tooth_2_lj_2 Lmp_3_tooth_3_lj_2	c
Lmp_4	3	Lmp_4_tooth_2_lj_2 Lmp_4_tooth_2_lj_1 Lmp_4_tooth_3_lj_1	d e f
Lmp_5	3	Lmp_5_tooth_2_lj_1 Lmp_5_tooth_3_lj_1 Lmp_5_tooth_4_lj_1	g

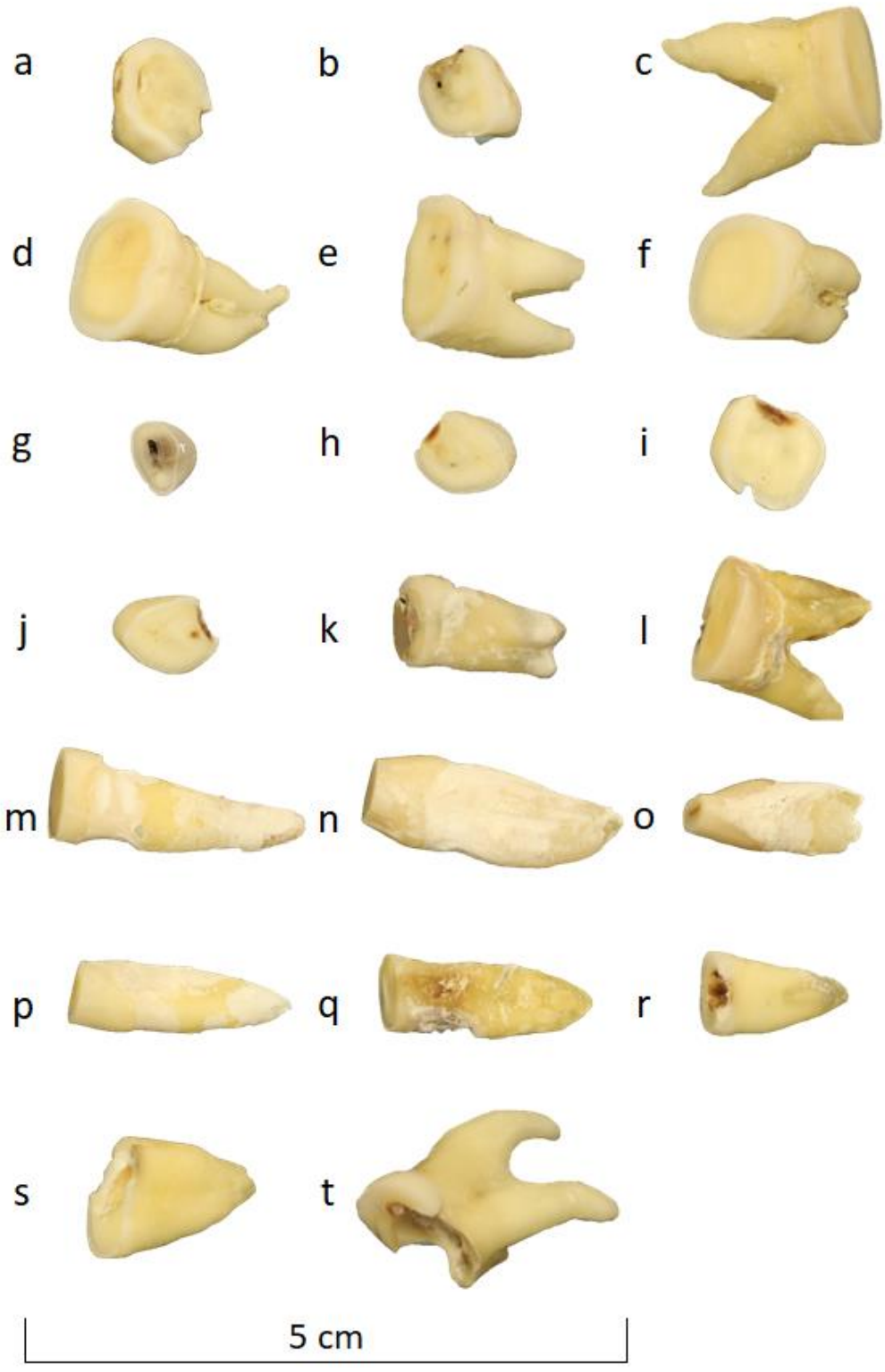

Figure 1.5: *Modern Human teeth samples. a) Ds_1_tooth_1, b) Ds_1_tooth_2, c) Dx_1_tooth_1, d) Gr_1_tooth_1, e) Gr_1_tooth_2, f) Kg_1_tooth_1, g) Ko_1_tooth_1, h) Ls_1_tooth_3, i) Ls_1_tooth_4, j) Ls_1_tooth_5, k) Mn_1_tooth_1, l) Mn_2_tooth_1, m) Mn_3_tooth_1, n) Mn_4_tooth_1, o) Mn_5_tooth_1, p) Mn_6_tooth_1, q) Mn_7_tooth_1, r) Mn_8_tooth_1, s) Sl_1_tooth_1, t) Sl_2_tooth_1.*

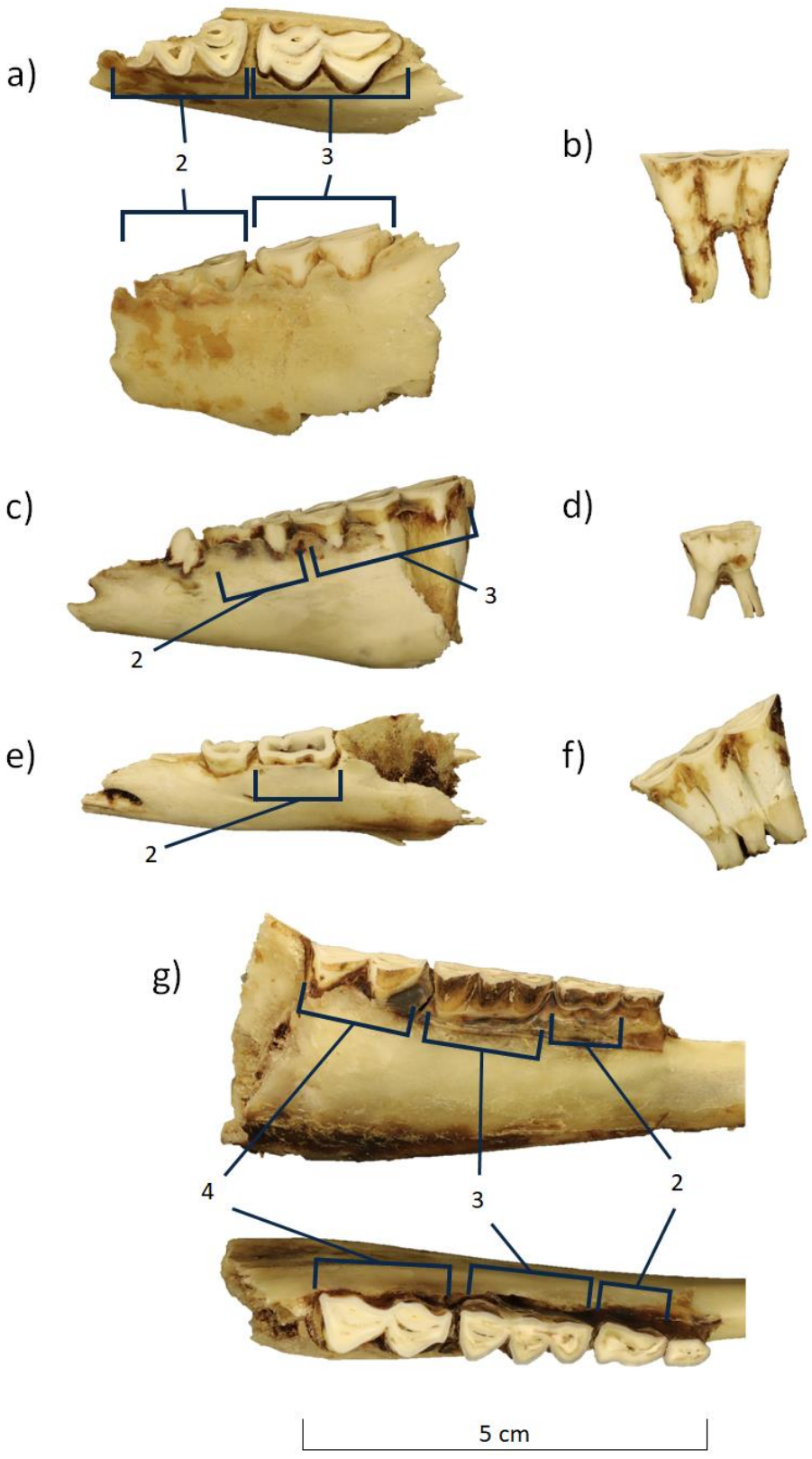

Figure 1.6: *Domestic Sheep teeth samples. a) Lmp_1_tooth_2_uj_1 and Lmp_1_tooth_3_uj_1, b) Lmp_2_tooth_3_lj_1, c) Lmp_3_tooth_2_lj_2 and Lmp_3_tooth_3_lj_2, d) Lmp_4_tooth_2_lj_2, e) Lmp_4_tooth_2_lj_1, f) Lmp_4_tooth_3_lj_1, g) Lmp_5_tooth_2_lj_1, Lmp_5_tooth_3_lj_1 and Lmp_5_tooth_2_lj_1.*

2 Laser Induced Breakdown Spectroscopy

2.1 LIBS History and Development

Plasma generation and its use for spectroscopic purposes was a field of interest at least since the late 1900s but the origins of Laser Induced Breakdown (or Plasma) Spectroscopy lie several decades later, in the 1960s, following the construction of the first pulsed laser. (Cremers, Radziemski & Loree, 1984; Cremers & Radziemski, 2013) The development of the first pulsed ruby laser in 1960 led to the systematic observation of laser-induced plasma during the following years and the development of further instrumentation led to the systematic use of Laser Induced Plasma Spectroscopy for obtaining qualitative and quantitative information on a variety of samples and materials. The strong interest of Los Alamos Laboratory led to extensive research and development of the technique during the 1980s, increasing the use and the applications of it. (Cremers & Radziemski, 2013) The following two decades faced a dramatic increase, year by year, on the research around LIBS and extensive increase in its applications with a wide range of them extending from Industry, to Cultural Heritage, Archaeology and Forensic science, to Medical science and Space exploration. (Cremers & Radziemski, 2013; Cremers, Multari & Knight, 2016)

2.2 LIBS Fundamentals

Laser Induced Breakdown Spectroscopy is an analytical technique based on the principles of atomic emission spectroscopy. In particular, this method is based on the formation of plasma, within the focus of a laser pulse directly onto the surface of a solid material. Plasma formation is accompanied by a bright flash of light that provides substantial information about the tested material (Cremers & Radziemski, 2013; Cremers, Multari & Knight, 2016) A short-duration (5 ns - 20 ns) and low-energy laser pulse, typically tens to hundreds of millijoules, is focused on the surface of a material with the use of a focusing lens, leading to material breakdown and plasma formation. (Cremers & Radziemski, 2013) The created plasma, which can be seen as a highly ionized gas is a local assembly of three different entities. Atoms, ions and free electrons coexist in the plasma creating an electrically neutral assemblage. (Cremers & Radziemski, 2013) The light emitted from the plasma depends on the unique spectral emission of atoms and ions of the elements in the ablated material, providing fingerprint spectra of the material's elemental components. (Cremers, Multari & Knight, 2016) LIBS has several advantages over other analytical techniques. A wide range of samples can be analyzed with it, plasma creation can be induced on the surface or within (in transparent materials) any type of sample, gases, liquids or solids, with little or no sample preparation prior to the analysis. Additionally, the technique is able for in situ analysis, since only optical access to the sample is required, providing simultaneous multielement detection and sustaining good sensitivity for many elements, too. (Cremers, Multari & Knight, 2016)

2.3 Plasma Formation

During the interaction of light and matter, atoms can typically absorb a single photon of radiation for each transition, with energy either equal to the transition's energy or equal to (or higher than) the ionization energy. The first consideration of a different behavior was made in the early 1930s, when a two-photon transition was considered possible. The invention of the first laser in the early 1960s led, among others, to the observation of the multiphoton excitation and multiphoton ionization phenomenon that was introduced in the second half of the same decade. (Mainfray & Manus, 1991; Agostini et al., 1968) Thus, it became clear that electrons in atoms are able to absorb more than one photon through nonlinear processes.

$$M + nhv \rightarrow M^{+} + e^{-}$$

This phenomenon is widely observed during the interaction of matter with high power density radiation (>1 GW/cm^2), for example a laser pulse. (Mainfray & Manus, 1991; Agostini et al., 1968; Cremers & Radziemski, 2013) After the production of the first few free electrons, through the multiphoton ionization process, their energy and velocity are increased by the inverse-Bremsstralung effect (free-free absorption), during their interaction with the electric field of the laser pulse. (Cremers & Radziemski, 2013; Bloembergen, 1997; Geltman, 1973)

$$e^{-} + hv \leftrightarrow e^{-*}$$

This leads to the production of free electrons capable of ionizing neutral atoms by collisions.

$$M + e^{-*} \rightarrow M^{+} + 2e^{-}$$

In a dense material, the increasing number of electrons, ions and collisions, leads to further free electron multiplication. This phenomenon is called avalanche or cascade ionization and creates a growing number of free electrons and ions in the surface of the material. (Cremers & Radziemski, 2013; Bloembergen, 1997) These two ionization approaches balance each other. In standard conditions of pressure and for radiance values that typically used by LIBS, avalanche ionization dominates the electron multiplication, but for higher radiance values there is multiphoton ionization that prevails the ionization of neutral atoms, leading to plasma formation. (Cremers & Radziemski, 2013)

In solids the arrival of a nanosecond laser pulse is followed, as was mentioned above, by the excitation of the free electrons in the lattice and their acceleration. The excited free electrons increase the collision rate, transferring energy to the lattice phonons. This leads to a significant increase of the local temperature that, first, gives rise to melting and then evaporation of the material. After evaporation of the material, the incoming laser pulse leads to bond breaking and ionization along with free electrons generation resulting in the formation of a weakly ionized plasma in the surface of the material (Figure 2.1). (Cremers & Radziemski, 2013; Chaudhary, Rizvi & Ali, 2016)

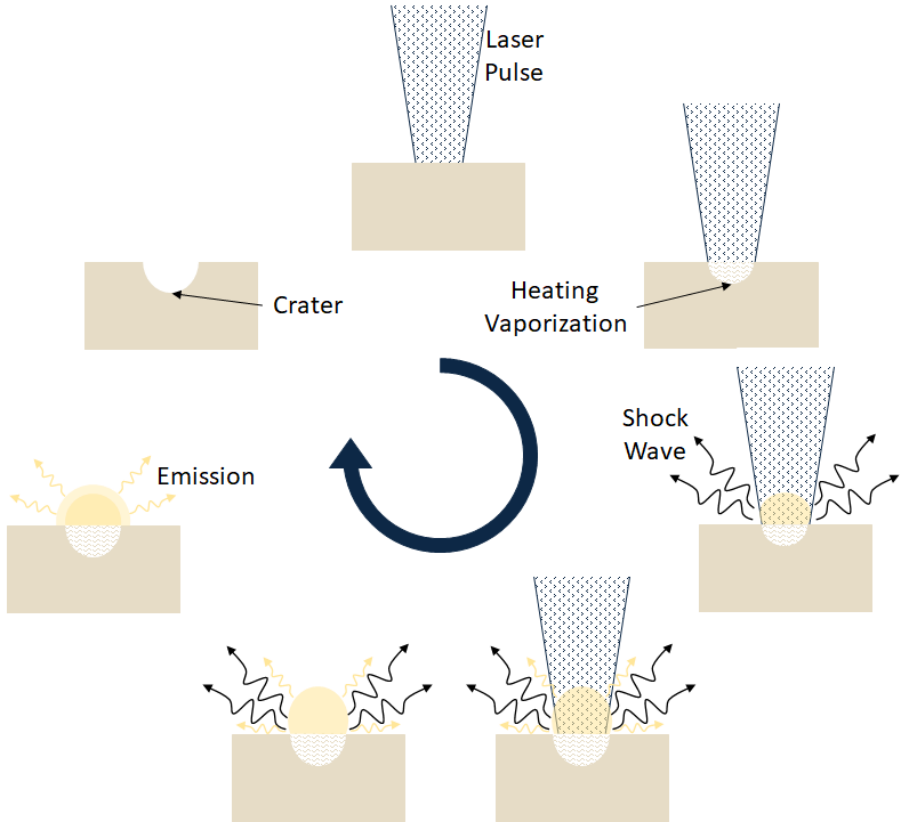

Figure 2.1: *Illustration of the main events occurring in interaction of a nanosecond laser pulse with a solid surface. The arrival of the pulse, the creation, expansion and cooling of the plasma and the formation of a crater in the surface of the material, can be seen. (Cremers, Multari & Knight, 2016)*

2.4 Plasma Expansion and Cooling

Ionization degree, temperature and electron density are the main parameters for the characterization of a plasma. LIBS plasma falls typically in the range of weakly ionized plasmas with the ratio of the electrons over the other species being below 10%. At the same time, plasma temperature on LIBS varies between 6000 and 10000 K. (Cremers & Radziemski, 2013; Cremers, Multari & Knight, 2016)

Following generation of the first few free electrons the interaction of them with the electric field of the ongoing laser pulse accelerates them, leading to the avalanche ionization phenomenon and the creation of plasma. Within the duration of the laser pulse, the plasma created continues to expand because of the energy supplied by the light electric field, reaching its full expansion at the end of the laser pulse. At this point, plasma is in its highest energy state with extensive ionization of its species and the highest temperature. Following interruption of the supplied energy, the plasma gradually starts cooling via loss of energy in the form of emitted radiation. An illustration of plasma generation, expansion and cooling can be seen in Figure 2.2, as a proportion of the optical signal intensity through time. At the first stages of plasma cooling, free electrons decelerate by the Bremsstralung effect producing continuum radiation (Figure 2.3). During the deceleration of the free electrons, collisions with the existing ions lead to formation of neutral atoms and the decrease of the free electrons

in the plasma. At that point, the emitted radiation from the transition on the existing ions, dominates the light produced by the plasma. As energy decreases and the collisions between free electrons and ions continue, the ion population decreases and so does the intensity of the corresponding ion emission spectral lines. Then radiation coming from transition between energy levels of neutral atoms dominates the range of the emitted light. In the final stages of the plasma, the energy reduction is extensive that collisions between atoms result in the formation of small molecules. (Cremers & Radziemski, 2013)

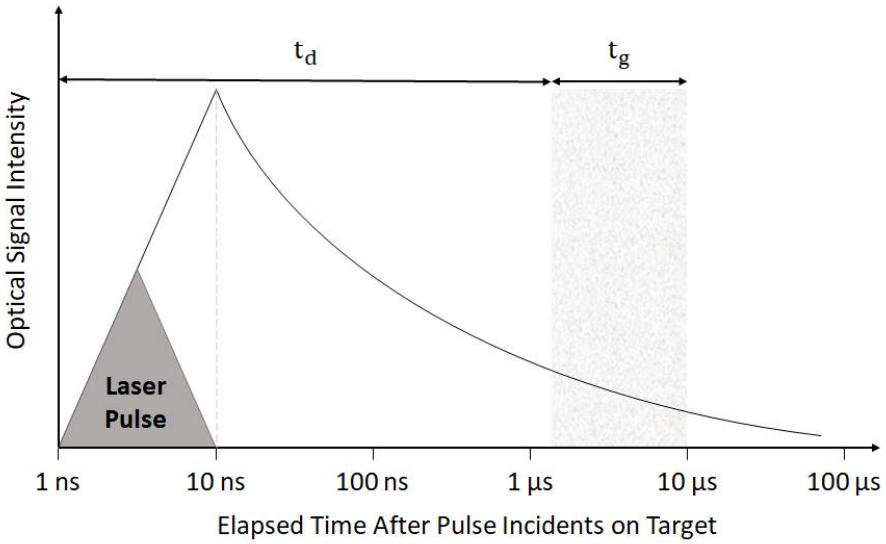

Figure 2.2: *Optical signal intensity of the plasma through its lifetime. The arrival of the laser pulse generates the creation of a plasma, which reaches the maximum of its expansion at the end of the laser pulse. A cooling process starts soon after termination of the pulse and interruption of the supplied energy, during which the plasma emits radiation. t_d represents the delay time for the beginning of the collection of the emitted radiation, while t_g represents the gate time, the time period that the spectrometers collects the emitted radiation. Time delay can be hundreds of nanoseconds, to avoid the continuum radiation of the existing free electrons, while gate time can be several microseconds to maximize the collected radiation from the plasma. (Cremers & Radziemski, 2013)*

As a result, the first stages of plasma cooling are dominated by the existence of a continuum radiation in the form of background signal, masking any atomic and ionic spectral lines and preventing the extraction of useful information concerning spectrochemical analysis. The continuum background decays with time along with the reduction of the free electron density, revealing the ionic and atomic spectral lines in the form of sharp emission peaks. The gradual cooling of the plasma leads to the creation of small molecules during the final stages of plasma's existence that can be observed in the form of vibronic bands in the LIBS spectrum. Considering that the continuum radiation on the first stages of the plasma cooling cannot provide any important information, the collection of the emitted radiation starts hundreds of nanoseconds after the arrival of the laser pulse. The time period that the emitted radiation is not collected is called "Delay Time" and can vary according to the needs of the measurements and the used instrumentation. The time period that the emitted

light is collected from the spectrometer is called “Gate Time” (Figure 2.2). (Cremers & Radziemski, 2013)

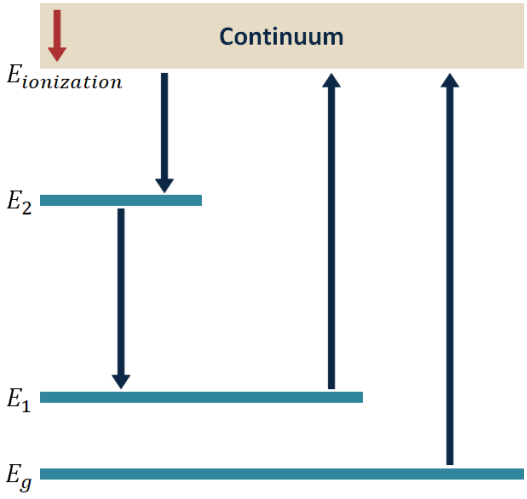

Figure 2.3: *Typical transitions in an atom or an ion. On the right side the ionization from the ground and an excited state can be seen. On the left side, the transitions of the free electrons are illustrated by the red arrow, while with blue arrows are illustrated the transition for bond creation and for the decay in lower excitation states. (Cremers & Radziemski, 2013)*

2.5 Line Shape and Broadening

The shape of spectral lines in LIBS plasma is a well-studied topic, even from the very first appearance of the technique in the early 1960s. (Gornushkin et al., 1999) This line broadening can be caused by two major mechanisms. Doppler broadening and pressure broadening. The first one is caused by the Doppler effect due to velocity distribution of the emitters, while pressure broadening comes from the interactions of the emitter with the surrounding atomic and molecular species. (Gigosos, 2014) The second mechanism can be further categorized in resonance pressure broadening, that results from interactions between identical atoms in the form of a dipole-dipole interaction, and Stark broadening which occurs with the interaction between charged species in the plasma. (Gornushkin et al., 1999; Gigosos, 2014) Due to the high electron density of the plasma, Stark broadening dominates over the rest of the broadenings in LIBS, giving a typical Lorentzian profile in the observed spectral lines. (Harilal et al., 1997; Cremers & Radziemski, 2013)

2.6 Matrix Effect

In many analytical techniques, including LIBS, the elemental composition of the sample can affect the observed elemental signal. This effect is known as matrix effect. In particular, matrix effect results in the modification of the produced signal of elemental components with constant concentration after changes in the concentration of one or more components forming the sample matrix. There are two different categories of matrix effects based on the physical and chemical properties of the samples. Physical matrix effect depends on the physical properties of the samples that affect the conversion of the ablated mass into plasma. Such properties are heat of vaporization, thermal conductivity, absorption coefficient and water content of

samples. On the other hand, chemical matrix effect occurs when the presence of one element affects the emission characteristics of another. (Cremers, Multari & Knight, 2016; Takahashi & Thornton, 2017)

2.7 Sample Damage

The formation of plasma on the surface of samples indicates the degradation of the material in the focal point of the laser pulse. The amount of degradation depends on the energy and time width of the laser pulse. Shorter duration indicates lower damage. When a nanosecond laser pulse interacts with a solid sample the dominant mechanism for plasma ignition is thermal vaporization. This leads to a significant and unsymmetrical crater reaming, due to the meltdown of the material (Figure 2.4). The damage by a nanosecond laser in hard tissues, can be seen in Figure 2.5. By contrast, a laser that produces pulses with a shorter time, for example a femtosecond laser, concentrates energy in much less time. In these conditions the interaction between light and matter differs significantly. Instead of thermal vaporization, the main bond-breaking mechanism is non-thermal. Because of the amount of energy in a shorter period of time, the material cannot transfer energy to the lattice or melt and thus evaporates instantly. During the interaction of laser pulse with the sample, extreme multiphoton ionization takes place resulting in the required conditions for plasma formation through Coulomb's explosion, leading to a more precise drilling and more well-shaped craters (Figure 2.4). (Cremers & Radziemski, 2013; Chaudhary, Rizvi & Ali, 2016)

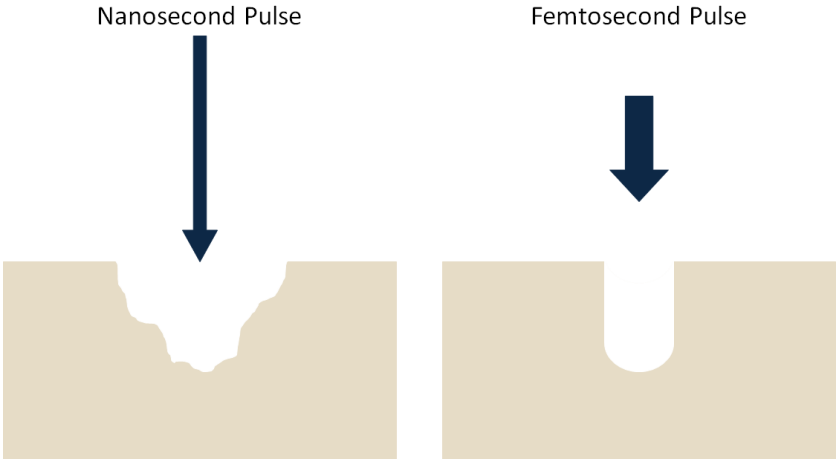

Figure 2.4: *Illustration of crater creation for two different types of laser pulses. On the left, an unsymmetrical crater is created after the interaction of the material with a nanosecond laser pulse, while on the right side, a well-shaped crater is created after the interaction of the material with a femtosecond laser pulse. (Cremers & Radziemski, 2013)*

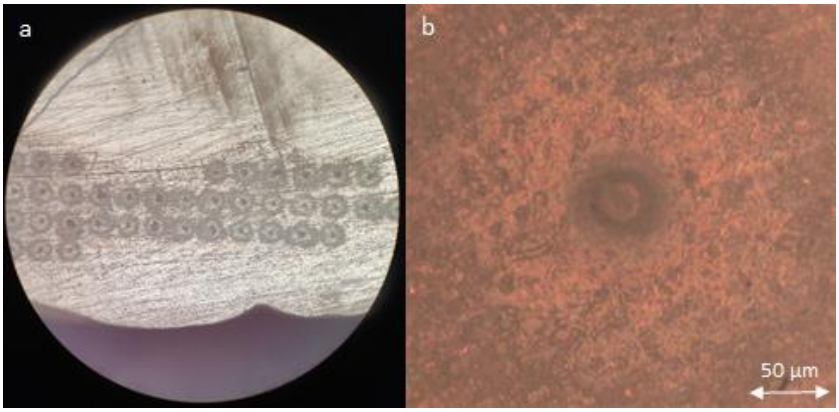

Figure 2.5: *Craters produced by a Nd:YAG laser after 10 consecutive pulses of 6 mJ each, on the same area, as they appear under an optical microscope. a) Several craters after mapping in the enamel of a tooth. b) One of the previously shown craters in magnification. The diameter of the crater is around 50 µm.*

2.8 Laser Fundamentals

To understand better the properties and applications of LIBS it would be necessary to briefly overview the theory of lasers. The term laser originates from the combination of the first letter in each word of the phrase "Light Amplification by the Stimulated Emission of Radiation". This explains the basic idea of a laser, which is the amplification of the emitted radiation of a medium, through the phenomenon of stimulated emission. Each laser consists of the same parts. A resonant cavity contains a lasing material, which allocates the appropriate energy levels to achieve the population inversion between an upper and a lower level of a transition, during a sufficiently strong pumping excitation by an external source. The external source that can provide the appropriate excitation of the lasing medium can be a flash lamp, a diode stack or even another laser. When the population inversion is achieved, photons produced by spontaneous emission between the previously mentioned upper and lower energy levels, can initiate an avalanche of emitted photons with equal energy through stimulated emission. This leads to a rapid amplification of the emitted radiation (Figure 2.6). Most common laser systems are three and four-level laser systems (Figure 2.7). In both cases, pumping excitation leads to the transition of the system (atom, ion, molecule) from a lower (normally the ground state) to a higher energy level, via which the upper level of the lasing transition is populated. A long lifetime of this level favors population inversion which eventually leads to a fast and rapid decay via stimulated emission producing laser radiation. A typical example of a four-level-system laser is the Nd:YAG laser. (Cremers & Radziemski, 2013; Hecht, 2019; Thyagarajan, & Ghatak, 2010)

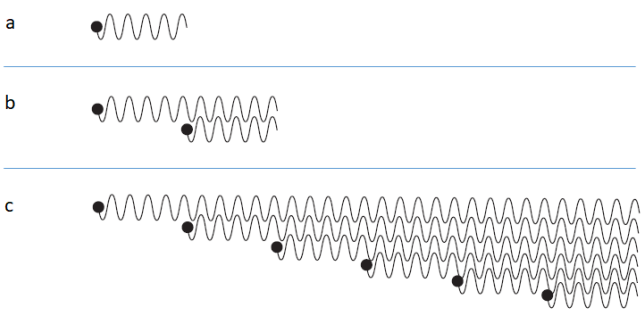

Figure 2.6: *a) Initial spontaneous emission. b) Stimulated emission of another photon shearing the same frequency. c) Stimulated emission of several photons with the same frequency after the creation of the first one. (Hecht, 2019)*

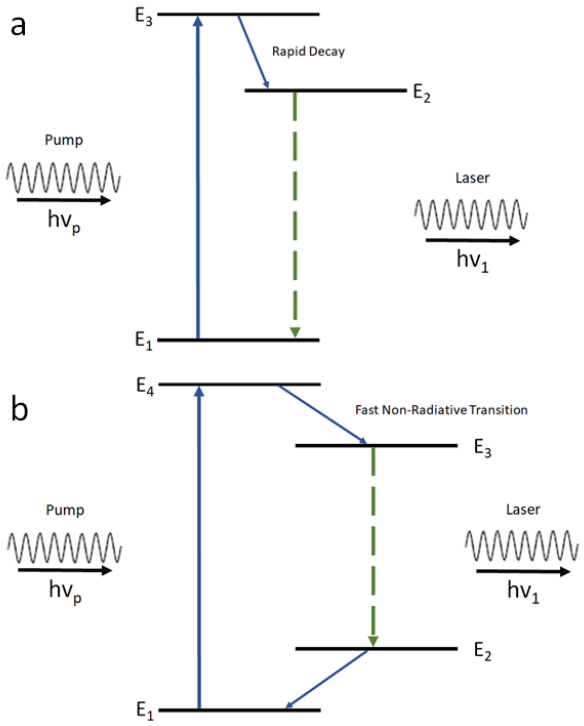

Figure 2.7: *a) Three-level laser system. The pumping energy leads to the excitation of an electron from the E_1 energy state to the E_3 energy state followed by a rapid decay from E_3 to E_2. The transition between E_2 and E_1 has low probability, increasing the population of E_2 and resulting in the population inversion between E_2 and E_1. Through stimulated emission the population of E_2 rapidly decays to E_1, producing the lasing radiation. b) Four-level laser system. As in the three level system, the pumping is followed by a rapid decay in a lower energy state. Population inversion takes place in E_3 energy level and through the decay to E_2 energy level by stimulated emission, the creation of lasing radiation can be achieved. A rapid decay to the lowest level follows the lasing transition. (Thyagarajan, & Ghatak, 2010)*

The majority of LIBS measurements are performed by using Neodymium-doped Yttrium Aluminum Garnet laser or Nd:YAG laser. In a Nd:YAG laser the lasing material is made by an Yttrium Aluminum Garnet doped by Nd^{3+} ions. A flashlamp irradiates the crystal, producing excitation on the lasing material by the absorption of a small percentage of the pumping light from the Nd^{3+} ions (Figure 2.8). Due to the electronic energy levels of the Nd^{3+} ions the intended population inversion can be achieved,

leading in a highly populated upper level of the lasing atomic transition between $^4F_{3/2}$ and $^4I_{11/2}$ energy states (Figure 2.9). To achieve the high powers needed for the ablation of a material, it is necessary to use an electro-optic Q switched shutter to create a short and high-power pulse. The Q switch is placed in the cavity to prevent photons, with the lasing frequency, to complete the path through the cavity and induce random stimulated emission from the lasing medium (relaxation oscillations), increasing the population inversion. At a certain time, the Q switch becomes transparent allowing photons to complete their path and induce stimulated emission, resulting in a short and high-power pulse. (Cremers & Radziemski, 2013; Stafe, Marcu & Puscas, 2014)

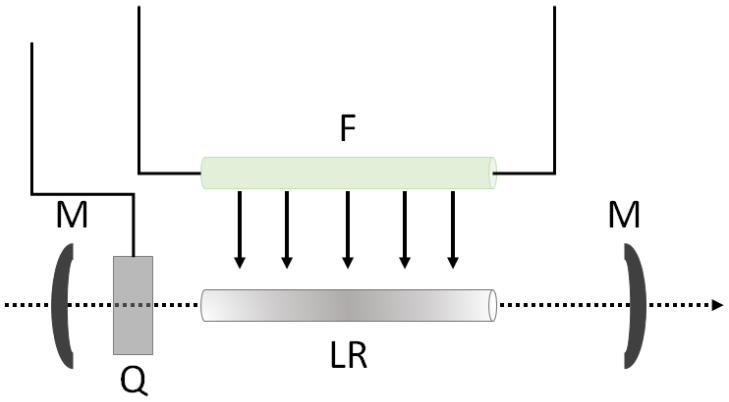

Figure 2.8: *Q switch laser cavity and flashlamp schematic. In a Nd:YAG laser, a mirror cavity contains the laser rod made by an Y and Al garnet doped with Nd^{3+} ions. Flashlamp light leads to the excitation of Nd^{3+} ions resulting in the population inversion of the lasing material. An 'opaque' switch prevents stimulated emission until a certain point that becomes transparent allowing a rapid decay and an instant emission of radiation (Cremers & Radziemski, 2013).*

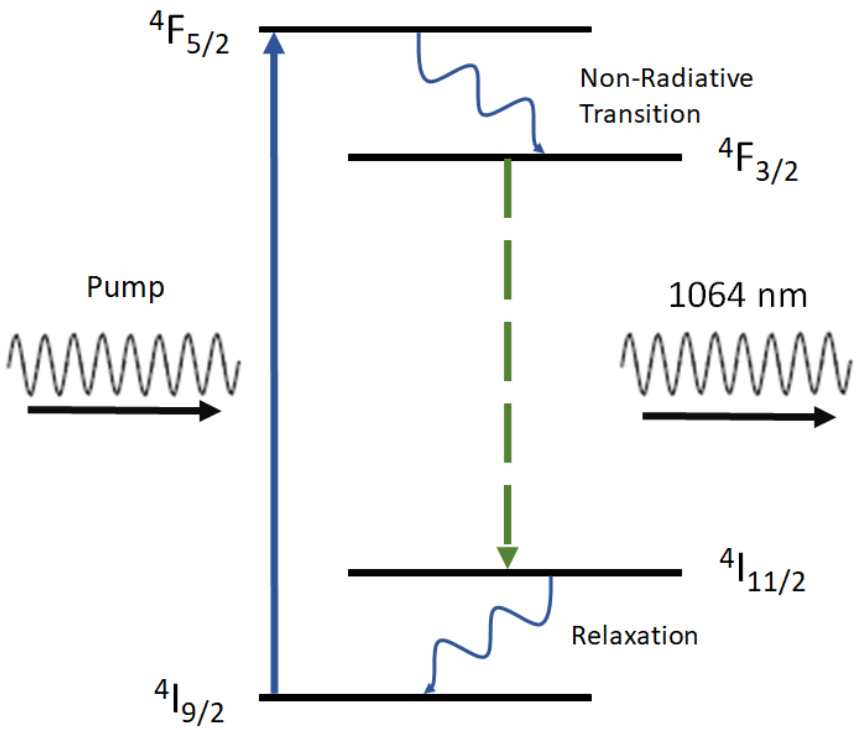

Figure 2.9: *The Nd:YAG laser is a typical example of a four level laser system. A flashlamp pumps the electrons to $^4F_{5/2}$ or a higher energy level. After a non-radiative transition from the high energy level to the $^4F_{3/2}$ energy state the population inversion is achieved. Stimulated emission through the transition to $^4I_{11/2}$ energy level produces the lasing radiation, followed by relaxation to lower energy levels. (Stafe, Marcu & Puscas, 2014)*

2.9 Experimental Setup

Data collection was performed by the use of a customized LIBS microspectrometer, which has been previously described in several publications. (Hausmann et al., 2017, 2019; Siozos, Philippidis & Anglos, 2017) The LIBS system was combined with a XYZ transition stage that could be either computer-controlled or manually controlled by the user and enabled scanning of the sample in a linear fashion or a 2D mapping. An infrared (1064) Q-switched Nd:YAG laser (Spectron Laser Systems), with a 10 ns pulse duration, was focused directly onto the sample surface using an objective lens (10x magnification, 28 mm focal length, LMH-10× −1064, Thorlabs) with infrared anti-reflection coating. Additionally, a clear image of the sampling area was formed, using another lens (f = +120 mm), on the sensor of a camera (2736 × 2192 pixels) at a magnification of 4.2 : 1, providing an overview of the sampling area. (Figure 2.10)

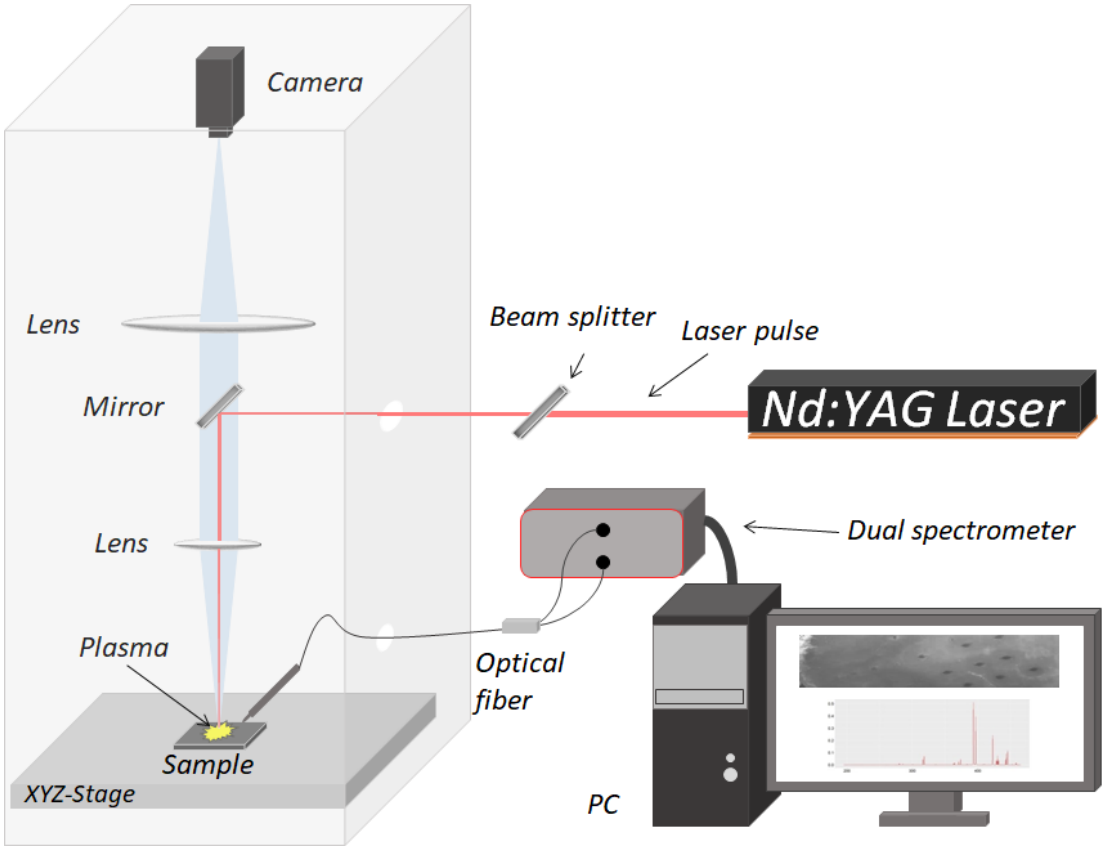

Figure 2.10: *Schematic of the LIBS microspectrometer. A Nd:YAG laser is used as a radiation source and the produced beam is focused on the surface of the sample using an infrared anti-reflection coating and an objective lens. Each pulse results in the creation of a plasma plume and the emitted light is collected by an optical fiber and transmitted into a Czerny Turner spectrometer unit. The LIBS system is combined with a XYZ transition stage, allowing the movement of the samples in space depending on the needs for each measurement. Sample movement and sampling area can be observed by a camera at a magnification of 4.2 : 1.*

A luminous plume of ionized material was created on the focal point of the laser beam onto the surface of the sample. The light emitted by the plasma plume was collected by a quartz fiber and transmitted into a Czerny Turner spectrometer unit (Avaspec-2048-2-USB2, Avantes), that records emission spectra across a wavelength range extending from 200 to 466 nm, with resolution of 0.2 nm. The delay time applied on the CCD was $\tau_D = 1.28\ \mu s$ and the integrational time was $\tau_G = 1.05\ ms$.

2.10 Experimental Procedure

The experimental procedure that was followed, was slightly different for each different dataset.

2.10.1 Archaeological Bone Fragments

The size and the texture of the archaeological bone fragments, combined with their cultural importance, prevent a potential smooth cut. The samples were placed without any preparation on the XYZ stage. The laser was focused directly onto the surface of the sample with an energy of 10 mJ/ pulse. Fifty five to sixty (55-60) points were analyzed on the surface of each sample, moving the stage manually to each point and correcting, if needed, samples position in agreement with the focal point of the

laser beam. At each point on the sample, the first 5 pulses were used to remove superficial dirt and dust, revealing a clean surface, while the next 5 produced an averaged LIBS spectrum per each sampling point. (Siozos et al., 2021)

2.10.2 Modern Domestic Sheep Teeth

Considering that the samples of this dataset had not had any historical or medical importance they were sectioned and evened, creating a sleek cross section in which the measurements could be performed easily and efficiently. The samples were placed one at a time on the XYZ stage with the cross section surface parallel to the stage surface and the laser was focused directly to the surface of the sample with an energy of 6 mJ/ pulse. Twenty five (25) points were analyzed in the enamel surface of each sample moving the stage manually to each point. In each point, the first 5 pulses were used to remove superficial dirt and dust, as previously mentioned, while the next 5 produced an averaged LIBS spectrum per each sampling point.

2.10.3 Modern Human Teeth

As mentioned above, considering the low historical and medical importance, each sample of this set was cut and evened, creating a sleek cross. The samples were placed one at a time on the XYZ stage with the cross section surface parallel to the stage surface. The laser was focused directly to the surface of the sample with an energy of 6 mJ/ pulse. More than sixty (60) preselected points were analyzed in the enamel surface of each sample using the auto-controlled mode for the stage movement. In each point the first 5 pulses were used to remove superficial dirt and dust, revealing a clean surface, while the next 5 produced an averaged LIBS spectrum per each sampling point.

2.11 Data Sets/ Data Form

After the formation of the plasma plume, the emitted light was collected by a quartz fiber and transmitted into a Czerny Turner spectrometer unit (Avaspec-2048-2-USB2, Avantes).The spectrometer was recording emission spectra from 195,117 to 465.855 nm, with resolution of 0.2 nm and the CCD detector comprised 2048 pixels, resulting to a sequence of 2048 intensity values matched with specific wavelength values. Every point on each sample was represented by a spectrum, originating from the accumulation of five single-shot spectra. As a result, data for each sample set are stored in a matrix of 2048 columns and as many rows as the points in each sample. Data matrices of samples are combined to create the final data matrices for the three different datasets, one for the Modern Human Teeth Samples, one for the Domestic Sheep Teeth samples and one for the Archaeological Human Bone Fragments. (Siozos et al., 2021) Additionally, an extra dataset consisted of both Modern Human Teeth data and Domestic Sheep Teeth data was created, with the combination of the already existing data from the previously created datasets.

2.12 Elemental Analysis and Spectra

LIBS can provide information on the elemental composition of the samples. Considering that the hard tissue remains mainly consist of hydroxyapatite, the

resulting spectra would contain emission lines mainly from calcium (Ca), along with minor emissions from other biogenic elements and impurities.

2.12.1 Modern Domestic Sheep Teeth

The spectrum for the domestic sheep teeth samples mainly consist of calcium spectral lines all across the spectral range investigated. However, spectral lines for other elements can be seen in the spectra. At $\lambda < 300$ nm, four different spectral lines coming from phosphorus emission can be seen, while several spectral lines coming from magnesium are spread across the spectrum. Additionally, two emission lines of Strontium can be seen at $\lambda > 400$ nm and a tiny spectral line coming from sodium (the well-known yellow doublet) appears, too. The resulted average spectrum for the Domestic Sheep samples can be seen in Figure 2.11.

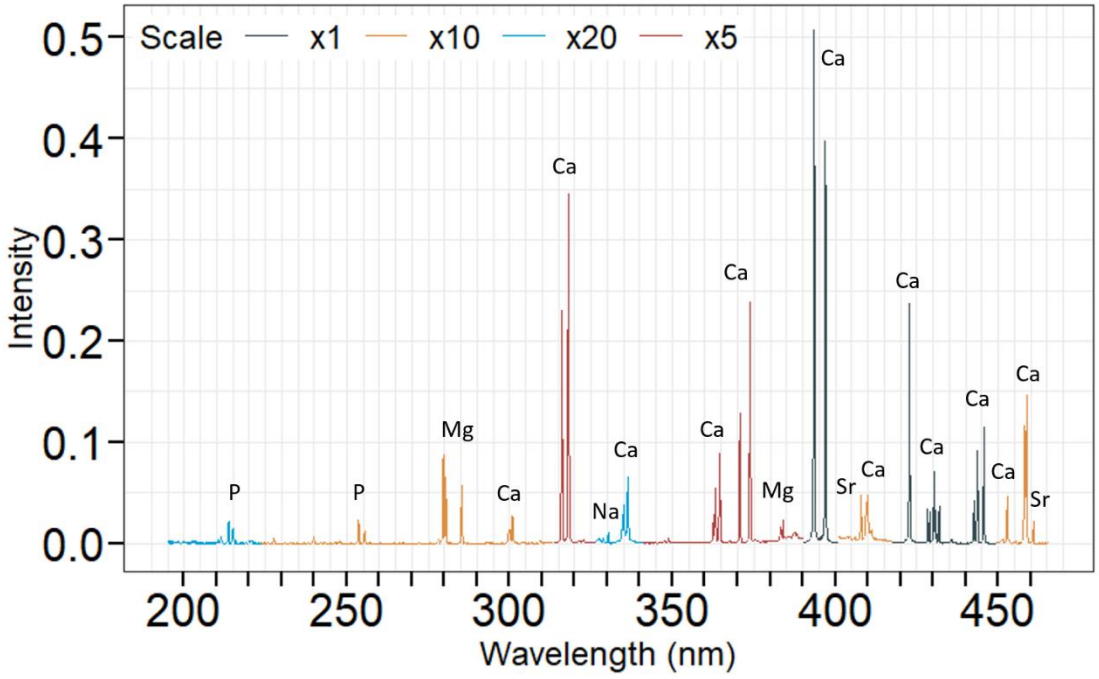

Figure 2.11: *Average spectrum using all the retrieved data from the Modern Domestic Sheep dataset. Several spectral lines can be seen, coming from the emission of Ca and P, which are the main elements on the hydroxyapatite, as well as Mg, Sr and Na that can replace Ca in hydroxyapatite crystal.*

2.12.2 Modern Human Teeth

The resulted spectra from the human teeth samples are quite similar to those of the domestic sheep teeth. In Figure 2.12, the average spectrum from all the human teeth samples and the elemental composition of them, can be seen. Comparing it with the spectrum for the domestic sheep teeth it is easily observed that are almost identical.

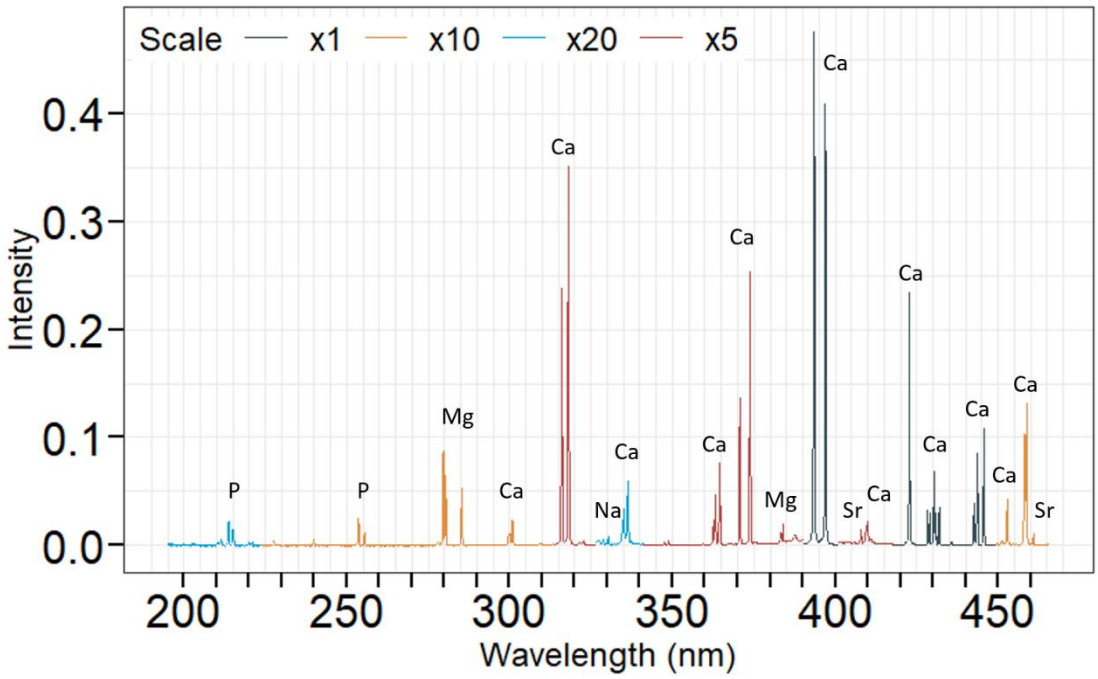

Figure 2.12: *Average spectrum using all the retrieved data from the Modern Human dataset. Similarly to the Domestic Sheep spectrum, spectral lines coming from the emission of Ca, P, Mg, Sr and Na can be seen. The spectrum is very similar to the one in Figure 2.11.*

2.12.3 Archaeological Human Bone Fragments

In contrast to the similarities observed in the spectra of the previously mentioned modern samples, the spectrum for the archaeological bone fragments is quite different. This spectrum is richer, consisting of spectral lines from several elements. Besides the spectral lines coming from Ca, Mg, P, Sr and Na, the spectra of the archaeological samples contained also spectral lines from iron (Fe), magnesium (Mn), aluminum (Al), copper (Cu) and barium (Ba), resulting in a richer spectrum as can be seen in Figure 2.13. The existence of these elements is attributed to potential burial contamination from the surrounding area that were excavated. Further analysis for the spectral characteristics of those samples has been presented in previous works. (Siozos et al., 2021)

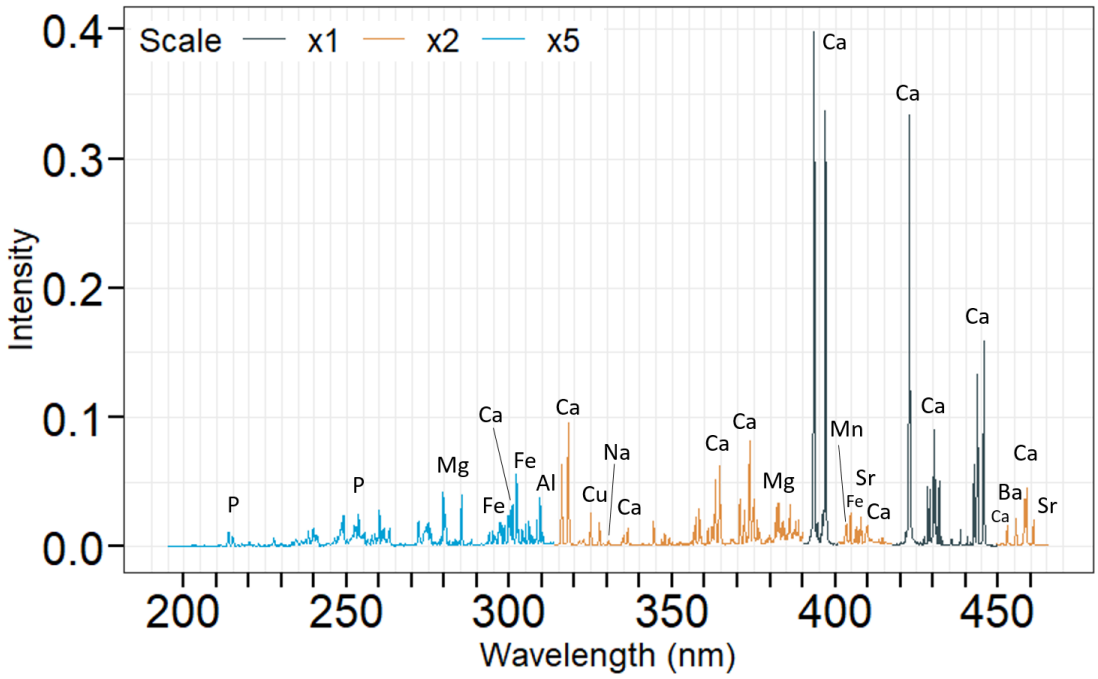

Figure 2.13: *Average spectrum using all the retrieved data from the Archaeological dataset. Several spectral lines can be seen, coming from the emission of both biogenic and non-biogenic elements of the hard tissues. Ca and P, which are the main elements on the hydroxyapatite, as well as Mg, Sr and Na that can replace Ca in hydroxyapatite crystal, can be seen in the spectrum, similarly with the spectra from the other datasets. On the contrary, in the current spectrum several new spectral lines can be observed, coming from the emission of non-biogenic element as Fe, Al, Cu and Mn.*

3 Machine Learning and Neural Networks

3.1 Machine Learning Fundamentals

The branch of artificial intelligence in which the main goal is the recognition of hardly visible patterns by an algorithm, was the inspiration behind machine learning. Machine learning is defined as the field of study that gives computers the ability to achieve specific tasks without being explicitly programmed, based on given examples. This process starts by the observation of hidden patterns in data and the attempt to improve future decisions. Hence, the main goal for computers is to learn or be trained in a way similar with the one that living organisms learn from experience. Machine learning is divided in three main branches that differentiate on the learning process and provided data's nature. In supervised learning the computer is fed with labeled input data and the main goal is the learning of general rules that map input, to the desirable output. In unsupervised learning no labeling for the input data is given to the algorithm, leaving it to discover itself hidden patterns. The last of the three categories is called reinforcement learning and is strongly linked with AI and game theory. In this approach the algorithm interacts dynamically with the environment by making discussions and discovering errors without given instructions whether it has come close to its goal or not. A further categorization of supervised learning can provide two different categories, classification and regression. Those categories differ on the nature of the desirable output. On classification the output variable takes discrete values, in the form of class labels, identifying a group membership, while on regression the output variables are continues. (Koropoulis, Alachiotis & Pavlidis, 2020; Alpaydin, 2010; Venables & Ripley, 2010)

In current work, different open-access supervised machine learning algorithms were used for classification tasks. In particular, the algorithms that were used are K Nearest Neighbors, Support Vector Machine using a polynomial and a radial kernel, Random Forest and a relatively simple Artificial Neural Network. Principal Components Analysis were used only for visualization of the data in 2D. All the algorithms were developed and performed in R programming language using a variety of already developed packages. (Wickham, 2011; Wickham, 2016; Meyer et al., 2015; Liaw & Wiener, 2002)

3.2 Principal Components Analysis

Principal Components Analysis is a commonly used unsupervised machine learning method based on coordination transformation, preserving as much of the relevant information as possible in the newly formed coordination system. The method firstly introduced by Pearson, back in 1901 and by Hotelling in 1933 (Pearson, 1901; Hotelling, 1933), as a method for dimensional reduction by the variance on the data and found its use among several applications in a variety of fields (Pearson, 1901; Jolliffe, 2002). PCA is the linear transformation of the original coordinate system to a newly formed, with the new set of variables being called principal components and maximize the variance among data. This result can be achieved by fitting the best fitted lines on the data that reduce the sum of squared distances of the data points

(Pearson, 1901; Jolliffe, 2002). The newly formed variables can be sorted from first to last following the decrease in variance of the data, creating the Principal Components. The use of principal components is widely spread considering that can be used for dimensional reduction and data visualization, lossy data compression, as well as feature selection and feature extraction. (Jolliffe, 2002; Bishop, 2006)

Hotelling's approach differs from that of Pearson's (Hotelling, 1933; Jolliffe, 2002; Bishop, 2006). PCA seeks for desirable linear combinations of the n dimensions of the initial dataset x to maximize the variance (Hotelling, 1933; Jolliffe, 2002; Jolliffe & Cadima, 2016). This linear combination can be given by the equation:

$$\sum_{i=1}^{n} u_i x_i = xu$$

Where u is a vector of constants u_1, …, u_n. For any of these linear combinations the variance can be computed by the equation:

$$var(xu) = u'Au$$

Where A is the covariance matrix of the whole dataset and $'$ denotes transpose.

For the computation of the covariance matrix, the use of two different equations is necessary. To compute the covariance between two variables, the equation:

$$cov(x_1, x_2) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

is used. However, the computation of the variance of each variable is necessary, too. This is achieved using the following equation:

$$var(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

With respect to the previous two equations the covariance matrix $(n\ x\ n)$ can be computed.

	x_1	x_2	…	x_n
x_1	$var(x_1)$	$cov(x_1, x_2)$	…	$cov(x_1, x_n)$
x_2	$cov(x_1, x_2)$	$var(x_2)$	…	$cov(x_2, x_n)$
…	…	…	…	…
x_n	$cov(x_1, x_n)$	$cov(x_2, x_n)$	…	$var(x_n)$

Continuing, identifying the linear combination that maximizes the variance for xu ($var(xu)$), it is equivalent to obtain a vector which maximizes $u'Au$. For achieving the maximization without turn on infinite, it is necessary to impose a normalization constraint in which $u'u = 1$, or that the sum of squares of elements for u equals 1. To maximize $u'Au$, considering $u'u = 1$, the standard approach is the use of Lagrange Multipliers:

$$u'Au - \lambda(u'u - 1)$$

where λ is a Lagrange Multiplier. Differentiation with respect to u gives:

$$Au - \lambda u = 0$$

or

$$(A - \lambda I_n)u = 0$$

where I_n is the $(n\ x\ n)$ identity matrix. As it is easily observed, λ is the eigenvalue of A and u is any corresponding eigenvector. Solving the equation results in $n\ x\ \lambda$ eigenvalues $(\lambda_1, \lambda_2, \dots, \lambda_n)$, corresponding in $n\ x\ u$ eigenvectors $(u_1, u_2, \dots, u_n)$. Sorting the eigenvalues from greatest to least, the corresponding eigenvectors can be sorted, too, with respect to the variance reduction. (Jolliffe, 2002; Jolliffe & Cadima, 2016)

Finally, after sorting the eigenvectors, they are used to transform the original data into new data

$$x\ u = x'$$

with a new coordination system and increased variance in the first few axis.

Considering the simplicity and the low computational time that requires, PCA is an easily applicable machine learning technique capable to resolve the dimensionality problems of high dimensional data, like spectra. For those reasons is one of the first techniques that are used for the classification of samples using spectral data and the first that was used in the current study searching for possible clusters among data and later for their visualization.

3.3 K Nearest Neighbors

K nearest neighbors or KNN was firstly introduced by Fix and Hodges in 1951 as a nonparametric discrimination analysis (Fix & Hodges, 1989) and established as a pattern recognition classifier by Cover and Hart in 1967 (Cover & Hart, 1967). This algorithm is a rather simple, yet widely used machine learning method with its basic principle based on finding the nearest neighbors of a data point in a dataset. Then, the data point can be classified by the majority of its nearest neighbors' class labels (Figure 3.1). Because of its simplicity the algorithm is used in a variety of fields not only for classification tasks, but also for feature selection, pattern recognition and clustering. (Xia et al., 2015)

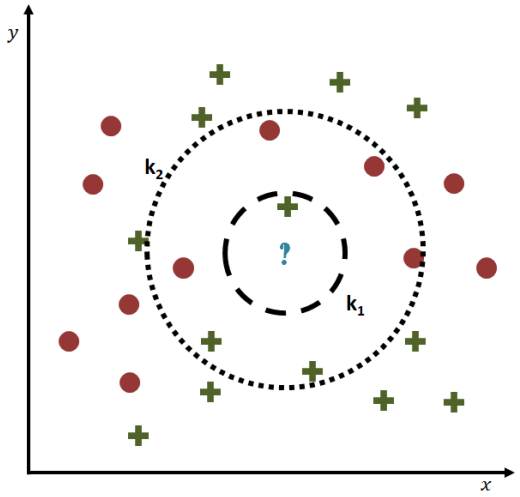

Figure 3.1: *K nearest neighbors representation in a two dimensional space, using two different K values. The red circles and the green crosses represent data points of two different class labels, while the blue question mark represents the unknown data point, which needs to be classified.*

The main idea behind this machine learning method is the calculation of the distances between an unknown sample and the training data points. Using majority voting of the class labels, for the K closest training data points, the unknown sample is assigned in one of the existing classes. The most commonly used methods for distance calculation are Euclidean distance, Manhattan distance, Chebychev distance and Minkowski distance. (Mulak & Talhar, 2015; Ooi, Ng & Lim, 2013) For the current work Euclidean distance was used. Assuming that arbitrary data correspond in a N-dimensional space R^N and an unknown sample is represented by the feature vector $U = (x_{1u}, x_{2u}, \dots, x_{Nu})$, where x_{ku} donates the value of the kth dimension of the unknown sample. The Euclidian distance between U and a known sample $A = (x_{1a}, x_{2a}, \dots, x_{Na})$ is given by the equation:

$$d(U, A) = \sqrt{\sum_{k=1}^{N} (x_{ku} - x_{ka})^2}$$

The number of distances that will be calculated each time, equals the number of the training data points. The closest K training data points are identified as the K nearest neighbors and their class labels will attribute to the assignment of the unknown sample, in the major class (Figure 3.2). (Sarkar & Leong, 2000; Mitchell, 1997)

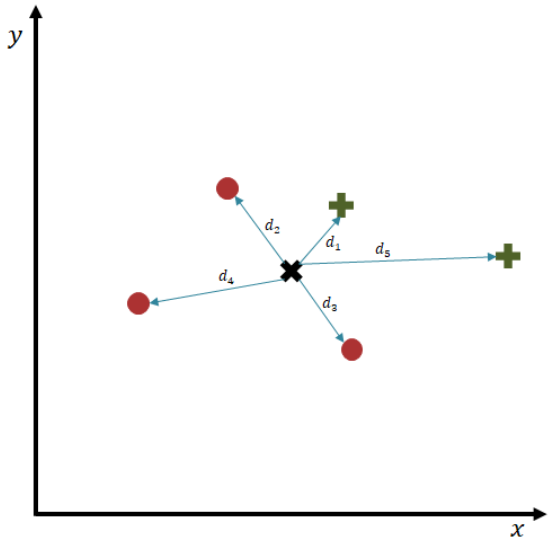

Figure 3.2: *Representation of K nearest neighbors algorithm in a two dimensional space. The red circles and the green crosses represent data points of two different class labels. The algorithm calculates all the distances between the unlabeled sample (black x) and the training data points. Following the calculations of the distances, the algorithms sorts them from least to greatest ($d_1 < d_2 = d_3 < d_4 < d_5$). Finally, the model classifies through majority voting the unknown data point. The number of nearest neighbors that are taken into consideration is very important for the result, since it can drastically change the class label of the unknown samples. Considering $K = 1$, the unknown samples would be assigned as green cross, however changing K value in $K = 3$, results in a different assignment of the unknown sample, as a red circle.*

As it is easily observed, KNN is not strictly a learning classifier but rather a memory-based classifier. (Koropoulis, Alachiotis & Pavlidis, 2020) Its simplicity makes it a popular choice for use, among machine learning methods, but its low accuracy in complex datasets limits its usefulness.

3.4 Support Vector Machine

Support vector machine is a machine learning algorithm, firstly introduced back in 1992 and 1995, based on the use of the best fitted hyperplane on the closest data points between the different classes for the classification of unknown data. (Boser, Guyon & Vapnik, 1992; Cortez & Vapnik, 1995) This algorithm uses the marginal data points from each class as support vectors for the desirable hyperplane that separates the samples. (Figure 3.3) The distance of those marginal data points from the hyperplane is called margin and the main goal of the algorithm is to maximize it, to increase the probability for sufficient classification of unknown samples in the correct class. (Koropoulis, Alachiotis & Pavlidis, 2020)

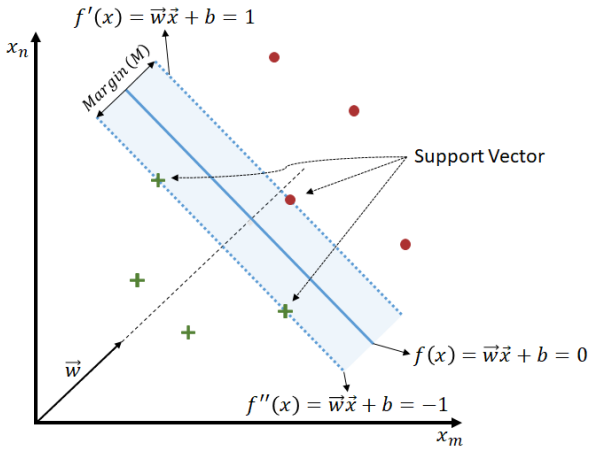

Figure 3.3: *Best fitting hyperplane, based on the support vectors of the different classes. (Boser, Guyon & Vapnik, 1992; Cortez & Vapnik, 1995)*

The hyperplane that separates the classes is defined by the equation:

$$f(x) = \vec{w} \cdot \vec{x} + b = 0$$

where $\vec{w}$ is the perpendicular vector to the hyperplane, b is a bias and $\vec{x} = (x_1, \dots, x_N)$ is the random N dimensional vector on the hyperplane. The hyperplanes on which the support vectors lay, are given by the equations:

$$f'(x) = \vec{w} \cdot \vec{x} + b = 1$$

$$f''(x) = \vec{w} \cdot \vec{x} + b = -1$$

Therefore, the areas outside of the gutter that is defined by the support vectors are given by:

$$\vec{w} \cdot \vec{x} + b \geq 1, \quad \vec{w} \cdot \vec{x} + b \leq -1$$

Considering two support vectors and calculating there difference, it can easily be shown that maximizing the margin equals to maximize the value of $\frac{2}{\|w\|}$. (Figure 3.4) In other words, to maximize the margin equals to minimize the $\|w\|$. (Boser, Guyon, & Vapnik, 1992; Cortez, Vapnik, 1995)

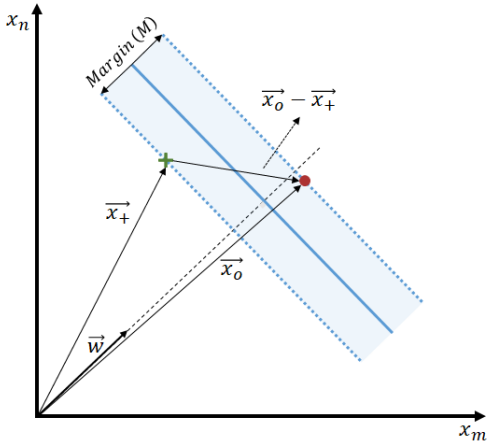

Figure 3.4: *Distance between the marginal values.*

The minimization of $\|w\|$ should be performed honoring the constraint:

$$y(\vec{w} \cdot \vec{x} + b) \geq 1$$

where $y = -1$ for $\vec{w} \cdot \vec{x} + b \leq -1$ and $y = 1$ for $\vec{w} \cdot \vec{x} + b \geq 1$. (Boser, Guyon, & Vapnik, 1992; Cortez, Vapnik, 1995)

When two classes are not linearly separable, the use of kernel functions that transform the data into a more convenient/separable form is needed. (Figure 3.5)

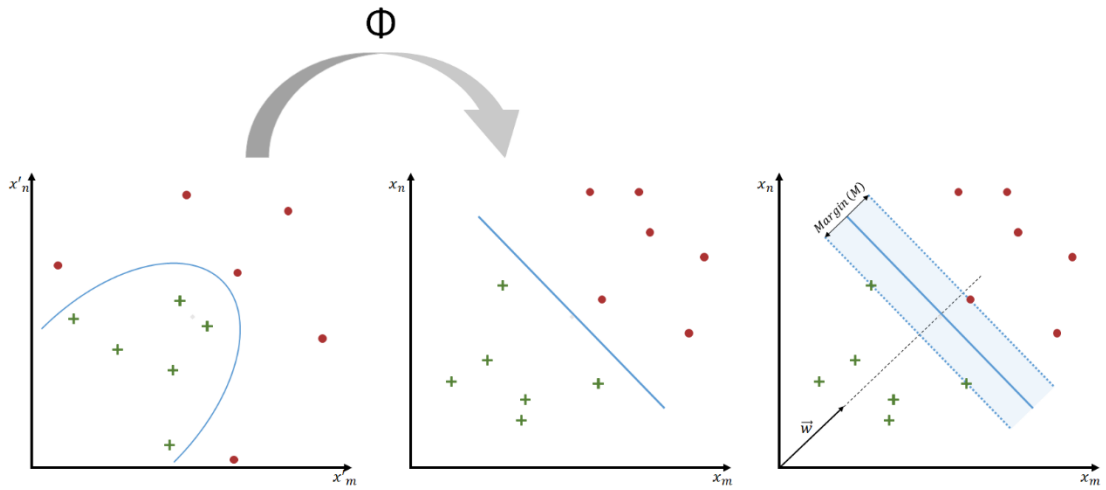

Figure 3.5: *Data transformation in a linearly separable form. This transformation is achieved by the use of a kernel functions.*

Popular kernel functions are:

$$K_r(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$$

$$K_p(x_i, x_j) = (x_i \cdot x_j + 1)^q$$

K_r is the Radial kernel function with $\sigma > 0$ as a free parameter determining the width of the Gaussian function and K_p is the Polynomial kernel function with q as the polynomial degree. (Boser, Guyon & Vapnik, 1992; Müller, et al, 1997; Vapnik, 1998; Cristianini, & Ricci, 2008)

The example that has been previously mentioned is called hard margin SVM and it is applicable when the used data are separable. For non-separable data it is necessary to penalize errors during training, resulting in a different form of SVM called Soft Margin. In this case a loss function is applied to the algorithm in order to penalize wrong classification during the training process. Most commonly used loss function in SVM is Hinge Loss. Using this function, the maximization of the margin is achieved by minimizing:

$$\left[\frac{1}{N}\sum_i^N max(0, 1 - y_i(\vec{w}_i \cdot \vec{x}_i + b))\right] + \lambda\|w\|^2$$

where the first part refers to the summation of the penalty values for each mistakenly classified data point, while the second part refers to the margin value, with λ being a free value to weight the impact of $\|w\|$ to the final result. (Vapnik, Guyon & Hastie, 1995)

As it is observed, SVM is a machine learning method that can be easily applied in various situations. It can be used for regression problems, as well as binary and multiclass classification problems, while using the kernel trick can be applied in non-linearly separable data, too. (Bishop, 2006; Boser, Guyon & Vapnik, 1992) Furthermore, it is able to solve the "curse of dimensionality" by having the ability to work with high dimensional datasets. (Koropoulis, Alachiotis & Pavlidis, 2020) Those characteristics make this classifier capable of solving a variety of classification or regression problems, making it a powerful tool with many applications.

3.5 Random Forest

According to Leo Breiman, who introduced the idea in 2001, a Random Forest is defined as a classifier consisting of a large ensemble of tree-structured classifiers $\{h(x, \Theta_k), k = 1, \dots, n\}$, where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x. (Breiman, 2001b) This classifier has been used in a variety of fields and it is widely spread due to its accuracy and its ability to highly perform for both small and large, high-dimensional data sets. (Biau & Scornet, 2016) The building block of the random forest, decisional tree, is a conceptually simple classifier that can be used for classification, as well as regression and is based on space separation in regions using specific features or a linear combination of them. Each tree consists of decision nodes which split the data and the space in separate areas, resulting in a tree structure model (Figure 3.6). As a result, while growing a tree the algorithm needs to decide on the splitting variables, splitting points and also the topology of it. To find best partitioning, is necessary to evaluate the best splitting criteria, scanning through all the possible inputs. This evaluation is performed by measuring the node impurity for each possible leaf and split. In classification trees, node impurity can be measured in three different ways, by Misclassification error, Gini index and Cross-entropy (or deviance). (Hastie, Tibshirani & Friedman, 2009)

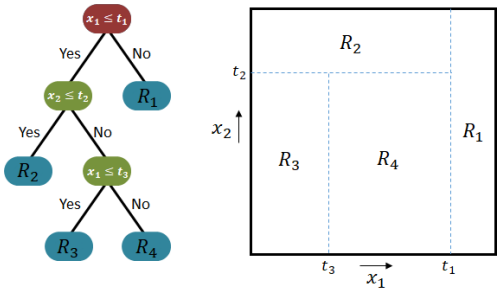

***Figure 3.6**: Decisional tree representation and space partition. On the left side, is the representation of a classification tree using binary splitting. The root of the tree is colored red, while the nodes are colored green. Leafs or terminal nodes are colored blue. On the right side,*

the representation of the space partition by the tree, can be seen. (Hastie, Tibshirani & Friedman, 2009)

In an arbitrary node m, representing a region R_m with N_m observations, the proportion of the observations of a class k is given by:

$$p_{mk} = \frac{1}{N_m} \sum_{i}^{N_m} I(y_i = k)$$

The impurity of that node could be measure by the previously mentioned ways as follows:

$$Misclassification\ error \quad \rightarrow \quad Q = \frac{1}{N_m} \sum_{i}^{N_m} I(y_i \neq k) = 1 - p_{mk}$$

$$Gini\ index \quad \rightarrow \quad Q = \sum_{k \neq k'} p_{mk} p_{mk'} = \sum_{k}^{K} p_{mk}(1 - p_{mk})$$

$$Cross - entropy \quad \rightarrow \quad Q = -\sum_{k}^{K} p_{mk} \log p_{mk}$$

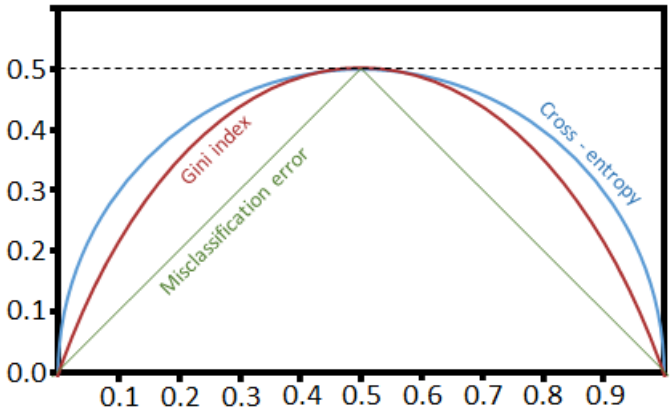

Figure 3.7: *Node impurity in a two-class classification model, as a function of the proportion of one of the two classes, measured by the three different values (Misclassification error, Gini index and Cross – entropy). Cross - entropy is scaled to pass through the point (0.5, 0.5). (Hastie, Tibshirani & Friedman, 2009)*

For a two-class classification problem, the values of the three measures for the node impurity are $1 - \max(p, 1-p)$, $2p(1-p)$, $-p \log p - (1-p)\log(1-p)$, respectively, considering p as the proportion for one of the two classes. A typical representation of the values can be seen in (Figure 3.7). (Hastie, Tibshirani & Friedman, 2009)

Cross – entropy and Gini index are the most popular, out of the three, values that are used in random forest's building blocks. (Koropoulis, Alachiotis & Pavlidis, 2020; "Random forests - classification description", 2021; Breiman, 1996b)

The impurity values of the descendants are weighted averaged for the calculation of the overall impurity for each split and the final value is compared with the impurity value of the parent node. Thus,

$$Q_{split} = \sum_{i}^{I} n_i Q_i$$

where Q_{split} is the final impurity value of the split, while Q_i and n_i are the impurity value and the sample size of each descendant, respectively. The split takes place only if the averaged impurity value is lower than the parent node impurity $(Q_{parent} > Q_{split})$, or until reaching the minimum growth of the tree, that has been previously set. When the stopping criteria is met, unsplit nodes are called "terminal nodes". Following this splitting process, the ending leafs or terminal nodes are, eventually, used for classification. Each of the terminal nodes is assigned to a specific class, by computing the most frequently appeared class on its representative space (Figure 3.8). (Culter, A., Culter, D.R. & Stevens, 2012) The discussion above is focused on the method named CART (Classification & Regression Trees) for the growth of decisional trees.

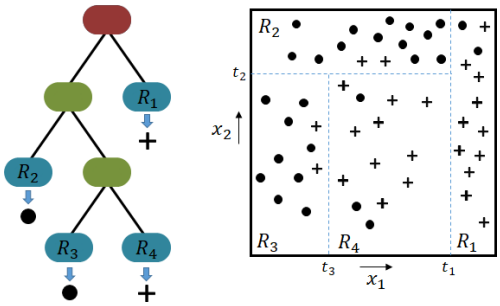

Figure 3.8: *Representation of a classification tree and the space partition that it provides. Each terminal node is assigned to the most frequently appeared class in its corresponding space section. (Hastie, Tibshirani & Friedman, 2009)*

Despite their advantages, decisional trees suffer from low accuracy because of their dependence from the data used to create them. (Hastie, Tibshirani & Friedman, 2009) Random forest bypasses this problem by creating a large number of independent tree-structured classifiers, which are taken into consideration for the final result of the classification or regression task. For the creation of this large amount of trees by a single dataset, a procedure called "bootstrap aggregating", or in one word "bagging", is used. Using this approach, several bootstrapped datasets are generated by the original dataset and are used for creating different trees. For further randomization on each tree, the best split in each node is chosen by a randomly selected subset of the initial features. This parameter is called m_{try} and typically is the $\sqrt{p}$, for classification and $p/3$, for regression trees, where p is the total number of features. The ensemble of the created trees predict unknown data by aggregating the final prediction of each decisional tree (Figure 3.9). Thus, in a classification task the prediction is achieved by majority voting, while in a regression task is achieved by averaging the predicted values. (Liaw, & Wiener, 2002; Breiman, 1996a)

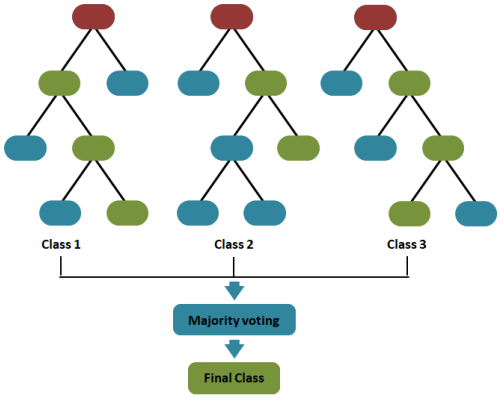

Figure 3.9: *Random forest sematic representation consisting of three tree-structured classifiers, voting equally for the final class. Nodes colored green are the representation of the path that the model follows in each tree for a hypothetical unknown sample. The roots of the trees are colored red, while the nodes that are out of use are colored blue. (Liaw & Wiener, 2002; Breiman, 1996a)*

Random forest is considered as one of the most useful and powerful machine learning methods reaching excellent performances in classification tasks, comparable to Support Vector Machine. (Díaz-Uriarte & Alvarez de Andrés, 2006)

3.6 Artificial Neural Networks

In the 1940s, McCulloch and Pitts introduced the idea of an algorithm that could complete computational tasks, inspired by biological neuros. (McCulloch & Pitts, 1943; Bishop, 2006; Hassabis et al., 2017) Since then, the research on the development of artificial neural network models has been very wide, leading to their use in a variety of fields and applications, like image and speech recognition, predictions on the activity of potential drug molecules, speech transcription into text and relevant results selection on search. (LeCun, Bengio & Hinton, 2015; Ma et al., 2015) Despite their recent success, neural networks suffered from lack of accuracy over other machine learning methods, through the years. The recent success can be explained by the increased data availability and the high computational power of modern computers. This plethora of data and computational power has outgrown the limits of traditional machine learning algorithms (Figure 3.10), while neural networks can handle that amount of information easier. (Aggarwal, 2018)

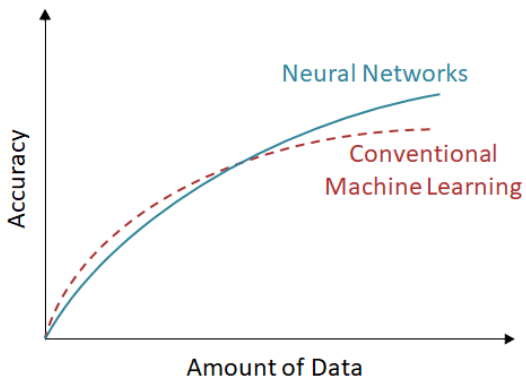

Figure 3.10: *Illustrative comparison between the accuracy of a traditional machine learning algorithm and the accuracy of a deep neural network. Increasing the computational power and*

consequently the amount of data that can be handled, neural network algorithms have become more attractive than conventional machine learning methods. (Aggarwal, 2018)

An artificial neural network consists of several layers of neurons. Each neuron is connected with all the neurons of the previous and the next layer, forming a network. Each path is a sequence of computational operations depending on the input values. The first layer of neurons is called input layer and represents the layer that feeds the model with the data, while the last layer of neurons is called output layer and provides the final results of the model. In between those extremum levels, lays a number of different layers of neurons, which are called hidden layers (Figure 3.11). The number of those layers combined with the number of nodes on each one affect the complexity of the model and determine the depth and the width of it. (Rumelhart, Hinton & Williams, 1986; LeCun, Bengio & Hinton, 2015; Aggarwal, 2018) For the purpose of this work, models with only one hidden layer were used for classification tasks and for that reason, only single-layer neural networks will be discussed. Neural network models with multiple hidden layers appertain to deep learning, following almost the same principals, but will be out of interest for the current work.

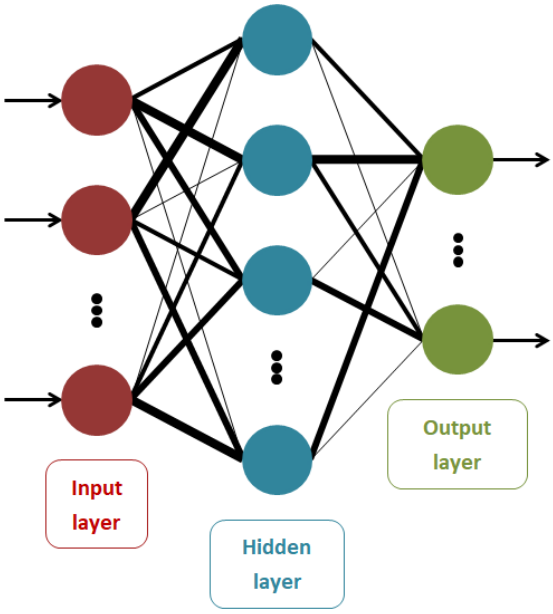

Figure 3.11: *Illustration of a single-layer neural network. Disks represent neurons on each layer, while the black lines represent the connections between the neurons. The different width of the lines represent the differences in the weight values. Input layer is colored red and is the only layer that neurons do not represent a computational operation. The hidden layer is colored blue and the output layer green. (Bishop, 1994; LeCun, Bengio & Hinton, 2015; Aggarwal, 2018)*

The nodes of a network are the representation of a computational operation, the transformation of the input value by an activation function, which provides the output value of the node. Sigmoid and tanh functions, were the most commonly used activation functions throughout the evolution of the neural networks (Figure 3.12).

$$\Phi(u) = \frac{1}{1+e^{-u}} \qquad (Sigmoid\ function)$$

$$\Phi(u) = \frac{e^{2u}-1}{e^{2u}+1} \qquad (Tanh\ function)$$

However, in recent years the ReLU ($\Phi(u) = \max\{u, 0\}$) and the hard tanh functions ($max\{min[u, 1], -1\}$) have heavily replaced the previously used functions on the modern neural network models (Figure 3.12). (LeCun, Bengio & Hinton, 2015; Krizhevsky, Sutskever & Hinton, 2017; Aggarwal, 2018)

$$\Phi(u) = \max\{u, 0\} \qquad (ReLu\ function)$$

$$\Phi(u) = max\{min[u, 1], -1\} \qquad (Hard\ \tanh\ function)$$

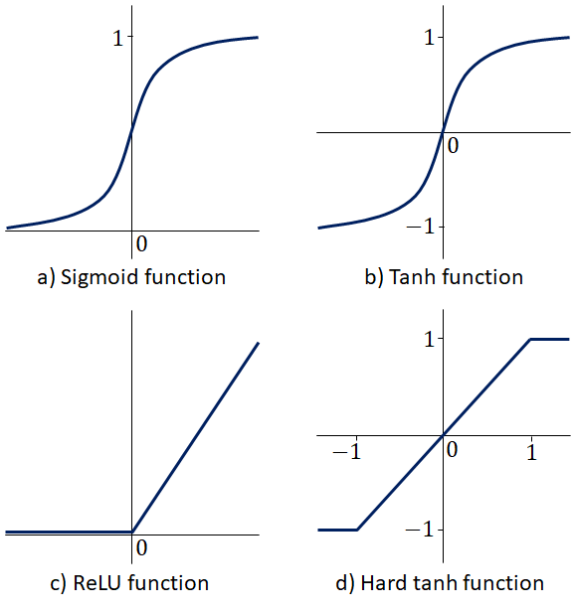

Figure 3.12: *Activation functions. (Aggarwal, 2018)*

The input value of the activation function in each node is provided by an additive contribution of each connected node's output. In other words, the input to a node is the weighted sum of the outputs from each of the connected nodes from the previous layer, added in a bias or offset term. (Bishop, 1994; Aggarwal, 2018)

$$\left(\sum_{i=1}^{n} x_i^{(L)} \cdot w_{ij}^{(L)}\right) + b_j^{(L)}$$

where $i = 1, \dots, n$ is the number of the node on the previous layer, $j = 1, \dots, k$ is the number of the node in the current layer, $L = 1, \dots, r$ is the number of the current layer, $x_i^{(L)}$ represents the initial input of the model or the output of a previous node, $w_{ij}^{(L)}$ is the weighted value for each computational path and $b_j^{(L)}$ is the offset term. In Figure 3.13, a mathematical and a graphical representation between the connections of a node with its previous layer can be seen. (Bishop, 1994; Aggarwal, 2018)

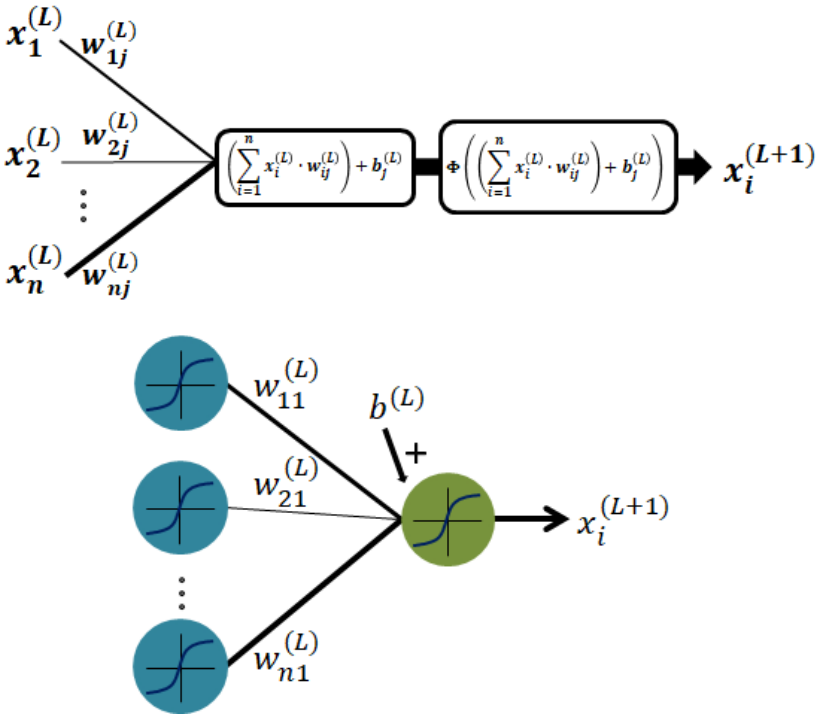

Figure 3.13: *Illustration of the connection between a node and the previous layer of nodes. In the upper half of the image the sequence of computational operations that take place during the connection of a node with the previous layer of nodes, can be seen. In the lower half of the image, a schematic representation of this procedure is illustrated. The differences in the width of the connecting lines represent the different weight values. Inside of each node a typical representation of an activation function can be seen. (Bishop, 1994; Aggarwal, 2018)*

During the training process of an artificial neural network, the aim is to find the proper set of weights and biases ensuring that the predicted result will be sufficiently close to the desired output, for any input vector. (Rumelhart, Hinton & Williams, 1986) The evaluation of the model's result is given by a loss function (L_t) that evaluates the agreement between the given and the expected value. Choosing the most suitable loss function in each application is crucial for the correct defining of the outputs. For regression tasks, squared loss and hinge loss are the most commonly used loss functions.

$$L_t = \sum_{j=1}^{J} (y_j - d_j)^2 \qquad (Squared\ loss)$$

$$L_t = \sum_{j=i}^{J} max\{0, 1 - y_j \cdot d_j\} \qquad (Hinge\ loss)$$

where $j = 1, \dots, J$ is the number of training vectors, y_j is the desirable output vector and d_j is the predicted output vector. However, for classification tasks in which softmax output is probabilistic, cross-entropy loss is preferred.

$$L_t = -\sum_{j=1}^{J} y_j \log p(d_j) \qquad (Cross - entropy)$$

where $j = 1, \dots, J$ is the number of training vectors, y_j is the desirable output vector, d_j is the predicted output vector and $p(d_j)$ is predicted output vector of probabilities. (Rumelhart, Hinton & Williams, 1986; Bishop, 2006; Janocha & Czarnecki, 2017;

Aggarwal, 2018) For the optimization of a neural network it is necessary to minimize the value of L_t ($L_t \to 0$). The minimization of L_t (Figure 3.14) is performed by gradient descent algorithm, which requires the calculation of the partial derivatives of the loss value for all weight and bias variables in the network (∇L_t). This optimization can be performed simultaneously for all the variables. A small sift is calculated for each different variable by the value of the ∇L_t, for the current variable, weighted by a learning rate. The extraction of that sifts gives the new values for the variables:

$$\theta_{new} = \theta - \eta \cdot \nabla L_t(\theta)$$

where θ is a random weight or bias value and η is the learning rate. (Rumelhart, Hinton & Williams, 1986; Ketkar, 2017; Jurafsky & Martin, 2009)

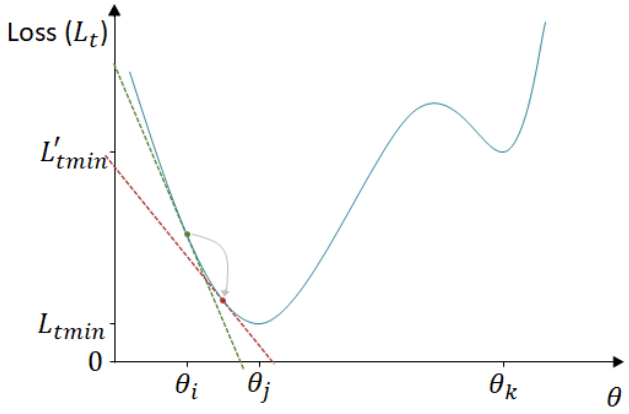

Figure 3.14: *Illustration of the Loss function L_t for a parameter θ. The algorithm performs changes in the value of θ, for each iteration, to move the value of the Loss function closer to the total minimum (L_{tmin}). The changes are based on the value of the slope for tangent line in each point. After the calculation of the derivative for a point (green), a new value for θ with reduced derivative (red) is calculated using gradient descent. L_{tmin} represents the total minimum of the Loss function and L'_{tmin} represents a local minimum of it. (Bishop, 1994; Jurafsky & Martin, 2009)*

Consequently, calculating the loss value for every training vector, starting with random weight and threshold values and using the results for the correction of each variant, can gradually decrease the error and increase the accuracy of the model. This method is called Error Backpropagation and it is the main learning process of the neural networks. (Rumelhart, Hinton & Williams, 1986) Summing up, the model evaluates the results for the training vectors using random variables and then correcting those variables reevaluates the results. This process continues until reaching a minimum value for the Loss function or reaching the maximum value of iterations. (Rumelhart, Hinton & Williams, 1986; Bishop, 2006) Using gradient descent approach, the model has to perform a great amount of computations, especially for large datasets, leading to significantly high training time and demanding memory. Another, common problem is the potential assumption of local minima as the total minimum of the loss function. To overcome those problems, most of neural network models use stochastic gradient descent. In this approach, a randomly selected subset of the training vectors is used for each iteration of the Backpropagation method, significantly reducing the impact of the training data on the Loss value. (Ketkar, 2017)

Artificial neural networks are currently the edge of machine learning research, finding use in many different fields, from scientific tasks, to business and everyday life technology. The evolution of computer science and computer hardware, combined with the availability of data and information have renewed the interest in that branch of artificial intelligence and have turned neural networks into the potentially most powerful tool for data analysis, in the present and possibly in the near future ,too. (LeCun, Bengio & Hinton, 2015; Ma et al., 2015)

3.7 Overfitting and Stratified Nested Cross-Validation

It has been widely noted that supervised machine learning algorithms suffer from an inadequacy to generalize from observed data to unseen data, because of overfitting. A machine learning model, which is trained using a finite number of data, can perform perfectly on training data but poorly on unknown samples, which are used as testing data. The cause of this phenomenon is the dependence of the optimization process of the model from the data that have been used for training. The noise of the training data can be considered as an existing pattern by the model and act as a basis of predictions, leading to biased models that are unable to correctly perform unknown data (Figure 3.15). (Ying, 2019; Lever, Krzywinski & Altman, 2016)

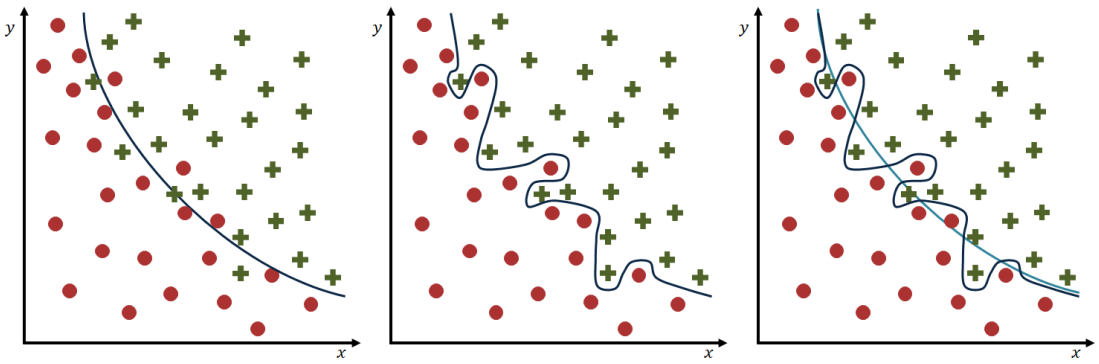

Figure 3.15: *Illustration of the fitting line of an arbitrary model in arbitrary training data. On the left side, the representation of the best fitted line that separates the training data sufficiently, while retaining a high probability to separate unknown data, can be seen. On the middle, is the representation of an overfitted line which perfectly separates the training data, while on the right side, the comparison between the two lines can be seen.*

Considering that real life data always occur many biases, to secure an unbiased result of a machine learning algorithm, avoiding overfitting, it is necessary to remove the impact that the training and testing data have on the final algorithm. The most commonly used method to achieve that is k-fold cross validation and it is based on developing several models each one trained and tested by different subsets from the original data. (Lever, Krzywinski & Altman, 2016; Hastie, Tibshirani & Friedman, 2009) To achieve that a partition of the data in a number of subsets m is performed. Those subsets are kept one at a time as validation sets $(y_i, j \in m)$, while the training of the model is performed using the rest of the data $(y - y_i, j \in m)$. The accuracy of each model is estimated each time by the remaining data and the final accuracy is estimated by averaging the accuracies of all the developed models (Figure 3.16). (Hastie, Tibshirani & Friedman, 2009)

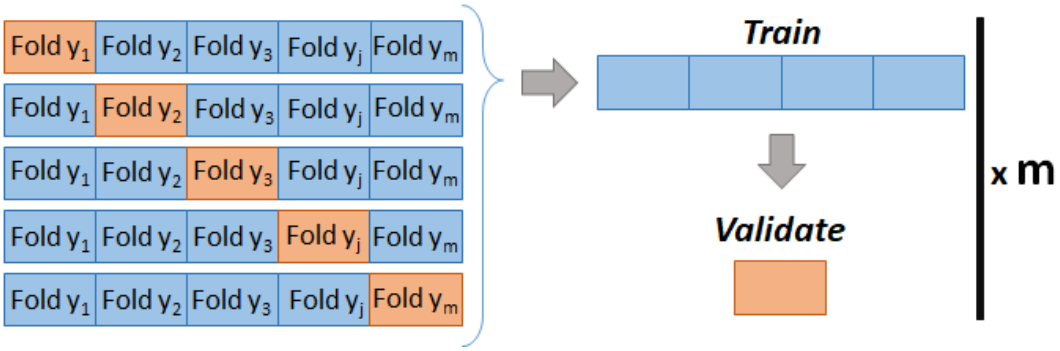

Figure 3.16: *Schematic representation of cross-validation process.*

Another use of the k fold cross-validation technique is during the parameters selection process that is necessary for the optimal performance of the algorithms. The optimal values of the parameters for each algorithm can be found by performing the cross-validation method using different values in each iteration and then selecting those created the highest performing model for use in the final trained model. For more robust outcomes the process can be repeated for several times. (Tsamardinos, Rakhashani & Lagani, 2014)

However, it is corroborated that cross-validation overestimates performance when it is used during parameter selection. (Tsamardinos, Rakhashani & Lagani, 2014) To overcome this problem, a method called nested cross-validation is recommended. In this method a new outer loop is introduced, working similarly with the cross-validation method. The previously mentioned parameter selection remains as an inner loop in the process. As a result, using this method the original data set is split in n different subsets. Each subset is kept as a testing set, while the rest of the data are used as the dataset for training and validation in the inner cross-validation process. After the selection of the optimal values for the parameters, the final model is tested using the data that have been kept previously, as a test set. The training-testing process of the algorithm is performed n times, each time using a different subset $(x_i, i \in n)$ as testing set and the rest of the subset as training set $(x - x_i, i \in n)$, resulting in n different outcomes (Figure 3.17). (Tsamardinos, Rakhashani & Lagani, 2014) The estimation of the final model's results is achieved by averaging the n different outcomes.

The training process and the result of a model can be affected by the partition of the initial dataset in subsets. To avoid the effects of biased partitioning, on the training model, an approach called stratification is recommended. Stratification forces the created subsets to have the same distribution of samples from each category with the initial data set. Using this technique is secured that each fold has a representative number of samples from each different class. (Tsamardinos, Rakhashani & Lagani, 2014)

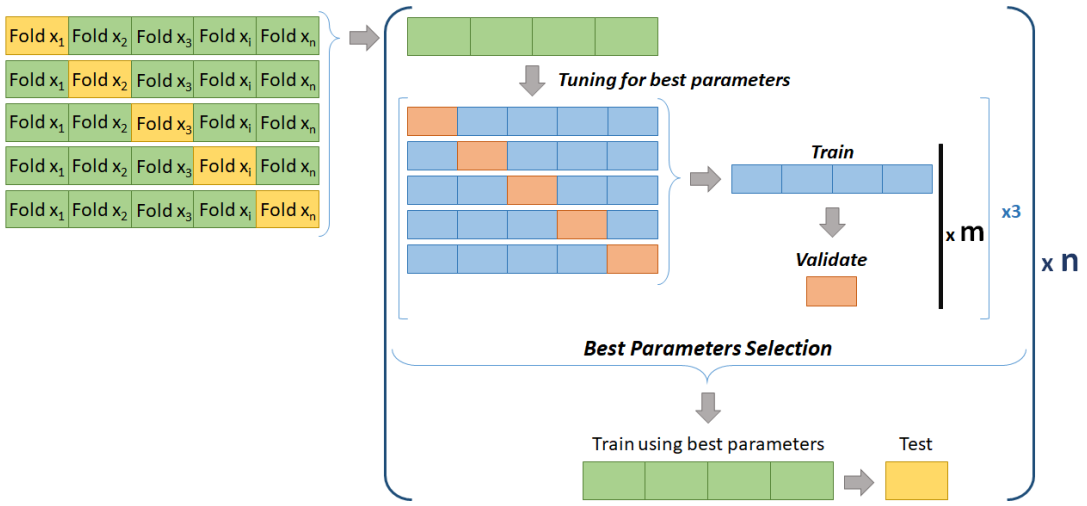

Figure 3.17: *Schematic representation of nested cross-validation process.*

3.8 Feature Selection and Extraction

As it has been discussed already in section 2.11, each data point consists of a spectrum and each spectrum consists of a sequence of 2048 intensity values. That number of dimensions is quite large, even for a machine learning problem. Considering the enormous computational time is needed for the processing of that amount of information and the fact that the majority of those values coming from spectral background, providing not useful information, it is imperative to reduce the number of dimensions for the optimal performance of the algorithms. (Guyon & Elisseeff, 2003) For that reason, several approaches for feature selection and feature extraction were employed.

Firstly, a feature selection function was developed, based on the feature variance of each class. The main idea was to isolate the features with the lowest in-class variance and the greatest among-classes variance. In order to achieve this selection, a new function was developed in which the ratio between the product of the in-class variances and the among-classes variance was calculated and sorted from the least to the greatest.

$$F_k = \frac{var_1 \cdot var_2 \cdot \cdots \cdot var_n}{var_T^n}$$

where var_i is the variance of the i^{th} class for a specific feature (pixel) k, var_T is the variance among different classes for a previously mentioned feature (pixel) and n is the number of the different classes. The K features (pixels) with the lowest values given by the previously mentioned function were considered important, as features, for classification using the machine learning algorithms, with K being an arbitrary number defined by the analyst.

A different methodology for feature selection was followed as well, based on a previous work (Siozos et al., 2021), targeting specific areas of the spectrum with characteristic emission spectral lines chosen directly from the analyst. Trimming the spectrum by targeting areas with emission lines coming from biogenic elements of the

sample, results in a significantly decreased number of features, followed by a desirable increase on the models accuracy and reducing the impact of potential contamination on the samples.

The last feature selection method used was based on the spectral lines intensities. Similarly with the previously mentioned method, specific spectral lines were targeted, but in this particular case only the pixels with the highest intensity were considered important as features. A function that isolates the pixel with the highest intensity for each given peak was developed.

In a different approach for solving the high dimensionality problem, instead of reducing the number of features by targeting only a small number of them, it is possible to extract newly formed features using the already existing. By this approach the information from the parental features is incorporated in the newly formed. (Guyon & Elisseeff, 2003) Based on this idea, a new function that calculates the integral for several specific spectral lines was developed. The calculated integrals for specific biogenic spectral lines were used as new features for feeding the models, significantly reducing the computational time.

3.9 Metrics

The resulting output of machine learning models can be evaluated by a variety of different metrics. The use of each different metric depend on the question that needs to be answered, as well as the nature of the problem itself. Some commonly used metrics are Accuracy, Area Under Curve (AUC), Choen's Kappa, Sensitivity, Specificity, F1 score, Mean Absolute Error, Mean Squared Error and many more. In the current work, Accuracy, alongside Sensitivity and Choen's Kappa were considered the most suitable metrics for the evaluation of this particular classification problem.

3.9.1 Accuracy

Accuracy is the most commonly used metric for the evaluation of many statistical techniques and machine learning methods. It is defined as the closeness of agreement between a test result and the accepted reference value, while it is calculated by the sum of the correctly predicted values divided by the number of total predictions. (Miller & Miller, 2014; Veropoulos, Campbell & Cristianini, 1999) Accuracy is given by the equation:

$$Accuracy = \frac{\sum(Correctly\ Predicted\ Valeus)}{Total\ Predicted\ Valeus}$$

3.9.2 Sensitivity

This metric calculates the proportion between the correctly predicted values in each group of samples and the total number of values in the same group. Alternatively, it is a way to compare the correctly predicted values with the values that had to be predicted correctly for the different categories of samples, if the rater was performing perfectly. (Lee et al., 2001; Veropoulos, Campbell & Cristianini, 1999)

Table 3.1: *Representation of the predicted values of the rater considering the reference values. True Positive represents the correctly predicted A values, while True Negative the correctly predicted B values. . On the other hand, False Positive and False Negative represent the incorrect A and B values, respectively (Lee et al., 2001; Veropoulos, Campbell & Cristianini, 1999).*

		Reference	
		A	B
Rater	A	True Positive	False Positive
	B	False Negative	True Negative

Sensitivity is given by the equation:

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

3.9.3 Choen's Kappa

Kappa or Choen's Kappa was introduced by Jacob Choen back in 1960 and it refers to the proportion of agreement between two raters, after chance agreement has been removed from consideration (Cohen, 1960). In this particular work, Kappa refers to the agreement of each rater with the reference, after the subtraction of chance agreement.

Table 3.2: *Representation of a matrix of probabilities for each different case of prediction by the rater. $P_{A,A}$ is the probability of agreement between rater and reference for A. Similarly, $P_{B,B}$, $P_{A,B}$ and $P_{B,A}$ are the probability of agreement between the rater and the reference for B and the probability of disagreement for B and A, respectively. The addition values are the overall probabilities of the rater to predict A and B and the overall probability of the reference to be A and B (Cohen, 1960).*

		Reference		Rater
		A	B	A,B
Rater	A	$P_{A,A}$	$P_{B,A}$	$P_{A,A} + P_{B,A}$
	B	$P_{A,B}$	$P_{B,B}$	$P_{A,B} + P_{B,B}$
Reference	A,B	$P_{A,A} + P_{A,B}$	$P_{B,A} + P_{B,B}$	

Kappa is given by the following equation:

$$Kappa = \frac{P_O - P_E}{1 - P_E}$$

Where P_O represents the probability of the correct predictions of the rater, according to the reference and P_E represents the probability of agreement by chance between the rater and the reference.

$$P_O = P_{A,A} + P_{B,B}$$

$$P_E = \left(P_{A,A} + P_{A,B}\right) * \left(P_{A,A} + P_{B,A}\right) + \left(P_{B,A} + P_{B,B}\right) * \left(P_{A,B} + P_{B,B}\right)$$

As can it be easily seen, the value of Kappa is the difference between the P_O and P_E, divided by the difference of 1 and P_E, which refers to the maximum value that the numerator can achieve, resulting in the normalization of the metric by max.

4 Results and Discussion

4.1 Models' Evaluation

The aim of this work was the use of different machine learning methods for the discrimination of LIBS spectra originate from hard tissue remains of several individuals, with a potential use in the classification of entire samples to the individuals that they belong. This approach in the classification of hard tissues can potentially speed up the archaeological and forensic studies, overcoming the difficulties of classifying huge numbers of excavated hard tissue remains.

After the collection of an adequate number of hard tissue remains, forming three main datasets, the first step was the development of the machine learning models. Four different algorithms were employed and tested. In particular, k Nearest Neighbors (KNN), Support Vector Machine using a polynomial (SVMP) and a radial (SVMR) kernel, Random Forest (RF) and an Artificial Neural Network (ANN) were used. The results from every algorithm were compared for the selection of the most suitable model for each task. Principal Components Analysis (PCA) was considered non-suitable for the current task considering that its use requires the standardization of data by feature. That type of normalization would equalize the background signal with the spectral lines that appear in the LIBS spectra, leading to potentially false results based on differences in background noise. (Pořízka et al., 2017) Thus, PCA has been used in current work in a non-typical way, just for the visualization of the data in space, after the unit vector normalization of them.

Following the development of the models, a series of testing processes has been performed for the evaluation of their functionality and their robustness, prior to the analysis of the hard tissue datasets. The evaluation was performed using the well-known Iris dataset because of the extensive work with it in bibliography, its well separated classes and its small size. This dataset was created back in 1936 by Ronald Fisher and it consists of the measurement of petal and sepal length and width, in centimeters, from three different types of iris flowers (Iris setosa, Iris versicolor and Iris virginica). The data have been retrieved from measuring fifty plants from each of the three different types. This particular dataset is often referred as "Fisher's Iris Data", too. (Andrews & Herzberg, 1985)

As it was expected, three clearly separated clusters appeared in the PCA score plot (Figure 4.1), each one for a different class of the flower, while all the algorithms had excellent results for all the used metrics revealing that the developed models work sufficiently well (Figure 4.3, Table 4.1). Despite that first evidence, a second test was performed in order to evaluate the validity of the first result. In this second evaluation process, the labels of classes in the Iris dataset were disarranged prior to the analysis. This major change on the data, created a non-separable dataset, as it was verified by the PCA score plot, too, (Figure 4.2). The analysis of this newly formed dataset revealed that classification could not be achieved by any of the algorithms. The values of Kappa for every algorithm were defining for that failure of achieving sufficient

classification. The fact that the values of Kappa laid below zero reveals that the classification results were not better than the random labeling of unknown data (Figure 4.4, Table 4.2). The failure of the algorithms to achieve classification using the class-reassigned Iris data, combined with the excellent results of the original Iris data leads to the evaluation of the initial hypothesis that the developed models sufficient classification power.

Figure 4.1: *PCA score plots for the Iris dataset. a) The data have been normalized by standardization, b) the data have been normalized by unit vector normalization.*

Figure 4.2: *PCA score plots for the Iris dataset with rearranged class labels. a) The data have been normalized by standardization, b) the data have been normalized by unit vector normalization.*

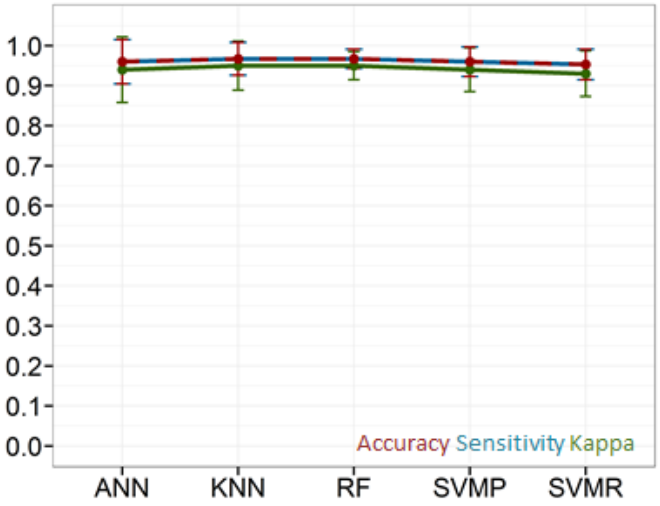

Figure 4.3: *Irish dataset, using all four features. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. All the algorithms have excellent classification results.*

Table 4.1: *Iris dataset results, using all four features.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.967	0.041	0.967	0.041	0.95	0.061
SVMR	0.953	0.038	0.953	0.038	0.93	0.057
SVMP	0.96	0.037	0.96	0.037	0.94	0.055
RF	0.967	0.024	0.967	0.024	0.95	0.035
ANN	0.96	0.055	0.96	0.055	0.94	0.082

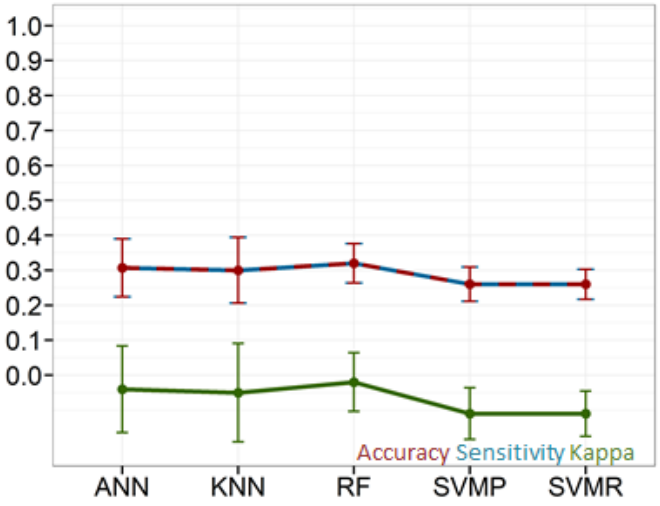

Figure 4.4: *Irish dataset, with rearranged class labels. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. Kappa values are lower than zero, revealing the inability of the models to perform better classification than the random selection.*

Table 4.2: *Iris dataset results, with rearranged class labels.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.3	0.094	0.3	0.094	-0.05	0.141
SVMR	0.26	0.043	0.26	0.043	-0.11	0.065
SVMP	0.26	0.049	0.26	0.049	-0.11	0.074
RF	0.32	0.056	0.32	0.056	-0.02	0.084
ANN	0.307	0.083	0.307	0.083	-0.04	0.124

4.2 Manual Feature Selection

4.2.1 Selection of 258 Pixels

After the evaluation of the correct behavior for the developed machine learning models, the Archaeological and Modern datasets were analyzed. The large number of pixels on each spectrum and thus the large amount of information to be processed by the classification models leads to a big amount of computational time and computational power needed. As a result, the reduction of the initial features prior to the analysis was considered necessary. Based on prior work (Siozos et al., 2021), the initial reduction was focused on specific spectral areas that correspond to specific spectral lines from biogenic elements of the hard tissues, leading to a significant reduction of the pixels/features(from 2048 to 258). Characteristic lines for calcium (Ca) and phosphorus (P), which are the building stones of hydroxyapatite and magnesium (Mg), sodium (Na) and strontium (Sr), which are the main elements that replace calcium in hydroxyapatite, have been chosen for the feature selection that was performed (Figure 4.5). This reduction on the number of the used features reduces the running time from several minutes to a couple of hours, depending on the size of the analyzed data and the algorithms that are used.

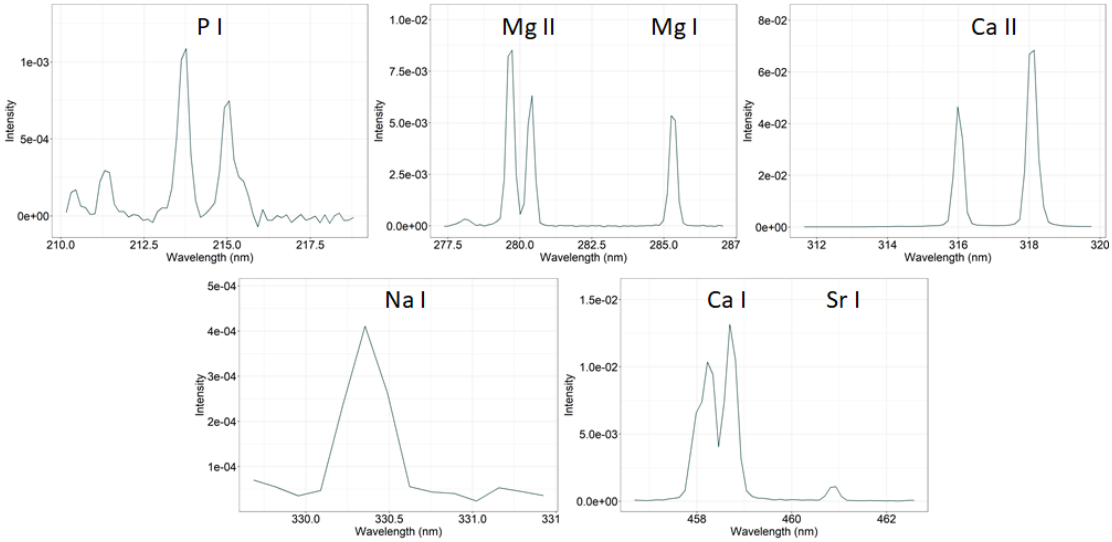

Figure 4.5: *Selected spectral areas with characteristic spectral lines from biogenic elements. The combined number of pixels from these areas was 258. Those areas were used for the initial manual feature selection, as they have been used in previous work, too. (Siozos et al., 2021)*

Firstly, the Archaeological dataset was tested (Figure 4.6) in order to compare the behavior of the models with a previous work (Siozos et al., 2021). The results from the algorithms showed a very good behavior of the Artificial Neural Network reaching a value over 0.85 in Accuracy, with the other two metrics reaching values over 0.8 supporting the achieved accuracy. These results showed a sufficiently good classification for this dataset, despite its lower values compared to earlier works (Figure 4.7, Table 4.3). The random forest reached almost an Accuracy value of 0.8 giving a sufficiently good result, while the rest of the algorithms resulted in metric values below 0.7, thus they consider unable to achieve a sufficient classification.

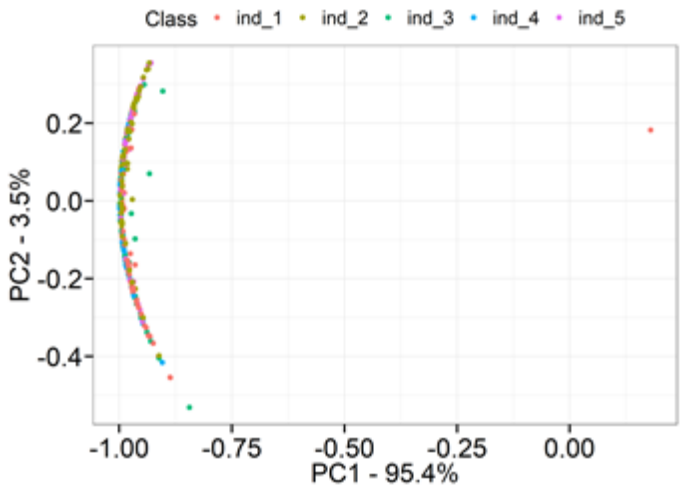

Figure 4.6: *PCA score plot for Archaeological Human Bones selecting 258 pixels and using unit vector normalization.*

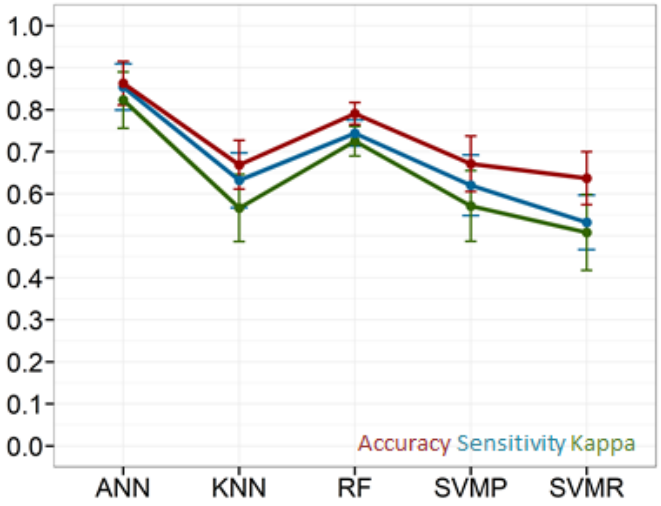

Figure 4.7: *Archaeological Human dataset, selecting spectral areas with biogenic spectral lines, corresponding in 258 pixels. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural network has significantly better results than any of the other algorithms.*

Table 4.3: *Metric values for the Archaeological Human data, using spectral areas with biogenic spectral lines, corresponding in 258 pixels.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.669	0.058	0.632	0.065	0.566	0.08
SVMR	0.637	0.063	0.532	0.065	0.508	0.09
SVMP	0.671	0.066	0.62	0.072	0.571	0.084
RF	0.791	0.026	0.744	0.031	0.725	0.035
ANN	0.863	0.052	0.854	0.055	0.823	0.067

Following the promising results from the Archaeological data, the algorithms were used for the analysis of the remaining datasets, under the same conditions. As it can be seen in Figure 4.9 and in Table 4.4, using the Domestic Sheep data (Figure 4.8) and the spectral areas that correspond to the previously mentioned 258 pixels as input vector for the algorithms, the machine learning algorithms gave lower results than expected, being unable to reach more than 0.5 Accuracy. Random Forest was the only machine learning algorithm able to reach almost 0.6. On the other hand, the result of the Artificial Neural Network was quite promising, with its Accuracy reaching a value more than 0.7, while Kappa value was almost 0.65 revealing the sufficient behavior of the model compared to the random labeling. Despite the fact that the developed models using these data were not able to achieve high results, the ANN algorithm was very promising for a potential successful classification using a most suitable feature selection. When the Modern Human data (Figure 4.10) were used under the same conditions (Figure 4.11, Table 4.5) the results were almost identical with those of the Domestic Sheep dataset. ANN had promising results with all the metric values reaching at least 0.7, while the rest of the algorithms had significantly lower results, with the two SVM algorithms having the lowest.

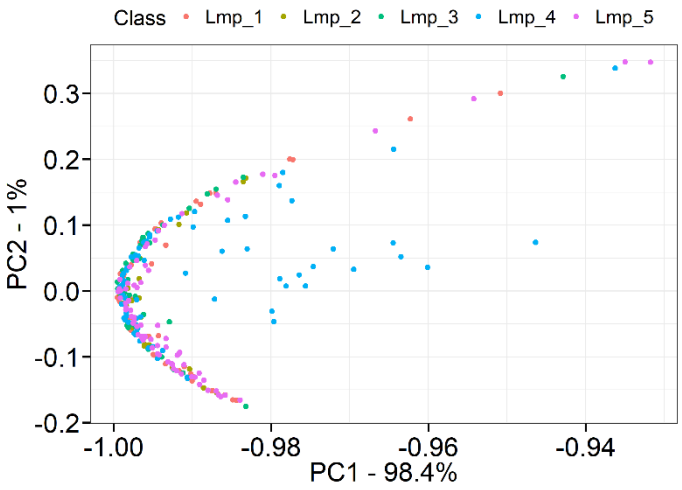

Figure 4.8: *PCA score plot for Modern Domestic Sheep Teeth, selecting 258 pixels and unit vector normalization.*

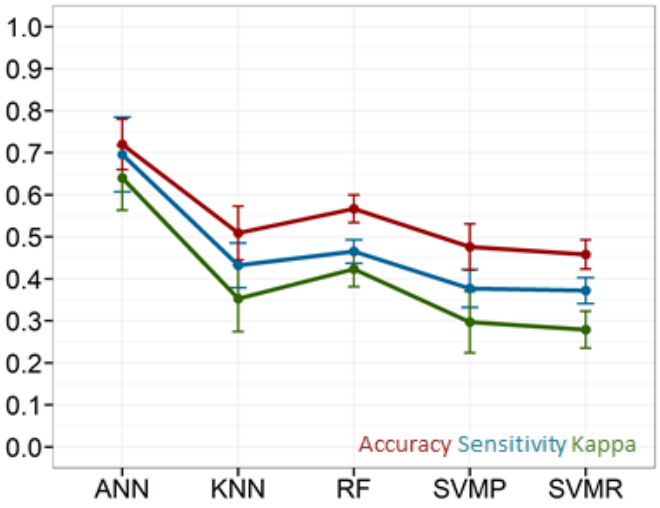

Figure 4.9: *Domestic Sheep dataset, using spectral areas with biogenic spectral lines, corresponding in 258 pixels. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural network has significantly better results than any of the other algorithms.*

Table 4.4: *Metric values for the Domestic Sheep data, selecting spectral areas with biogenic spectral lines, corresponding in 258 pixels.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.509	0.064	0.432	0.053	0.353	0.079
SVMR	0.458	0.035	0.372	0.031	0.279	0.044
SVMP	0.476	0.055	0.377	0.045	0.297	0.073
RF	0.567	0.033	0.465	0.028	0.423	0.042
ANN	0.72	0.06	0.696	0.089	0.64	0.077

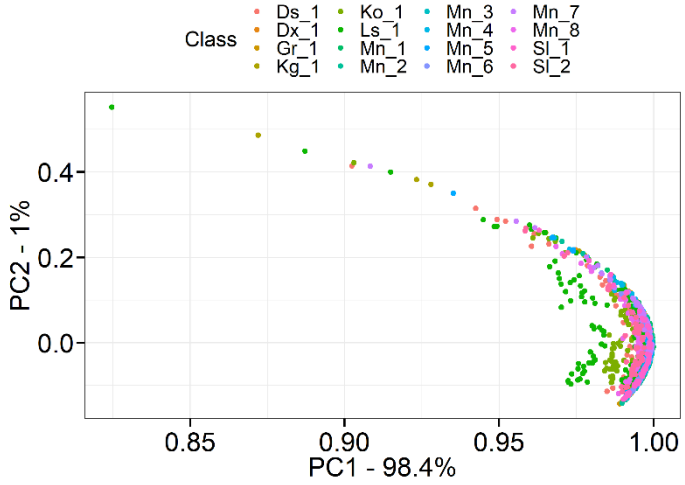

Figure 4.10: *PCA score plot for Modern Human Teeth, using 258 pixels and unit vector normalization.*

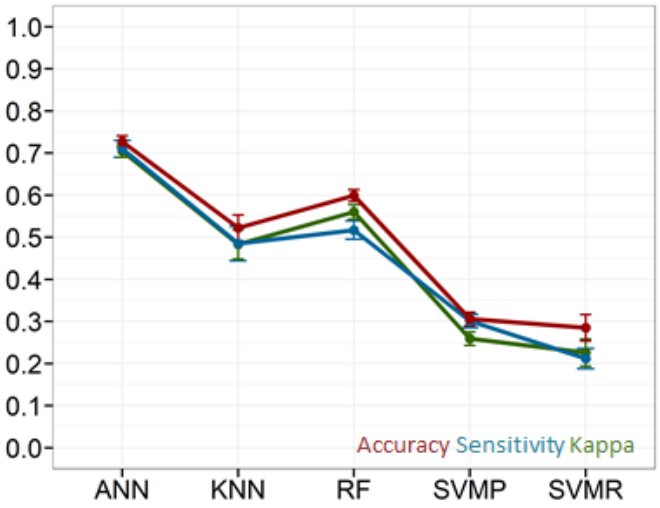

Figure 4.11: *Modern Human dataset, selecting spectral areas with biogenic spectral lines, corresponding in 258 pixels. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural network has significantly better results than any of the other algorithms.*

Table 4.5: *Metric values for the Modern Human data, selecting spectral areas with biogenic spectral lines, corresponding in 258 pixels.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.522	0.031	0.485	0.041	0.483	0.035
SVMR	0.285	0.031	0.212	0.024	0.226	0.032
SVMP	0.306	0.016	0.301	0.016	0.259	0.016
RF	0.599	0.014	0.517	0.022	0.56	0.017
ANN	0.727	0.014	0.71	0.02	0.705	0.015

The inability for sufficient discrimination in different individuals analyzing the two modern datasets raised the question whether the discrimination between species could be achieved easier. For that reason, the two modern datasets were combined

to each other in an attempt to create a larger dataset in which on the different samples to be grouped in the different species instead of individuals (Figure 4.12). Using the newly formed combined dataset, retaining the equal proportion of the two different classes, the analysis of those data with the developed algorithms had significantly good results, regardless of the algorithm that was used (Figure 4.13, Table 4.6). Once again, KNN and the SVM algorithms had lower results compared to RF and especially ANN. Accuracy value for Random Forest was almost 0.9, revealing a very good discrimination using this algorithm, while all the metric values for the ANN exceeded 0.9, revealing an excellent performance of the model in the discrimination between species.

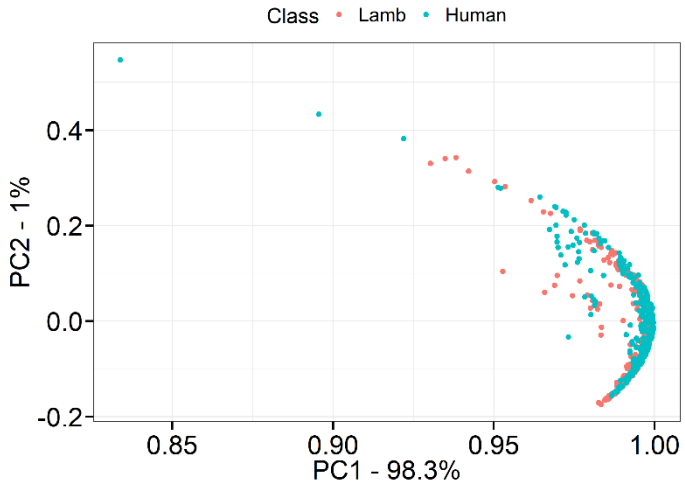

Figure 4.12: *PCA score plot for Modern Domestic Sheep and Human Teeth, using 258 pixels and unit vector normalization.*

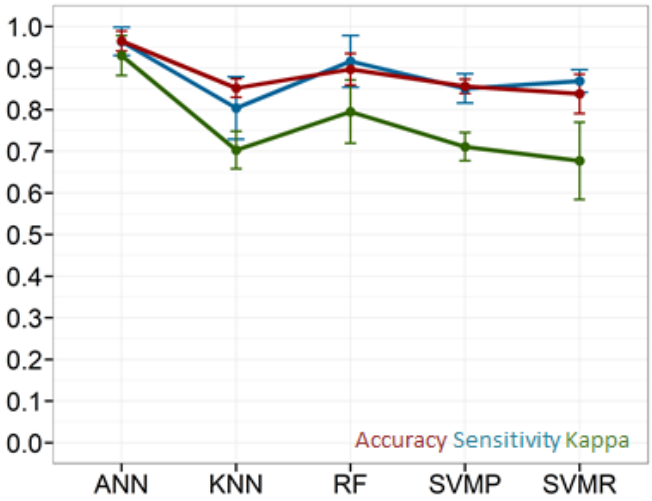

Figure 4.13: *Domestic Sheep and Human dataset, selecting spectral areas with biogenic spectral lines, corresponding in 258 pixels. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural network has significantly better results than any of the other algorithms.*

Table 4.6: *Metric values for the Domestic Sheep and Human data, selecting spectral areas with biogenic spectral lines, corresponding in 258 pixels.*

	Accuracy	*SD*	*Sensitivity*	*SD*	*Kappa*	*SD*
KNN	0.852	0.022	0.804	0.075	0.703	0.045
SVMR	0.838	0.047	0.869	0.027	0.677	0.093
SVMP	0.856	0.017	0.851	0.035	0.711	0.034
RF	0.897	0.038	0.916	0.062	0.795	0.076
ANN	0.965	0.024	0.964	0.034	0.93	0.048

4.2.2 Selection of 100 Pixels

The promising above results combined with the significant amount of running time led to a further manually feature reduction, for maximizing the resulted classification and minimizing the used features. The features from the background signal of the previously used areas, were discarded and resulting in narrower spectral areas as initial features. After the further feature reduction only 100 pixels/features from each spectrum were used (Figure 4.14).

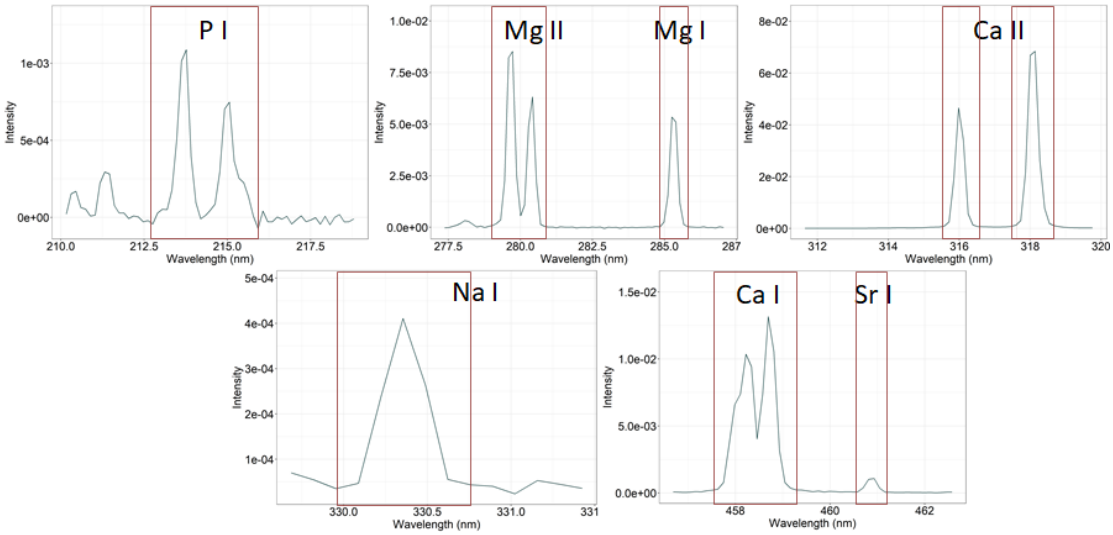

Figure 4.14: *Newly selected spectral areas. The background signal has been discarded, leading to a significant pixel reduction to 100, from the initial 258.*

Feature reduction significantly reduced the running time of the analysis. At the same time, the resulted classification either increased or remained the same depending on the used dataset. In particular, the already excellent classification to species (Figure 4.15) remained almost intact, resulting in excellent classification Accuracy values regardless the used algorithm (Figure 4.16, Table 4.7). ANN retained the higher results, with its three metric values exceeding 0.9 and Accuracy value exceeding even 0.95.

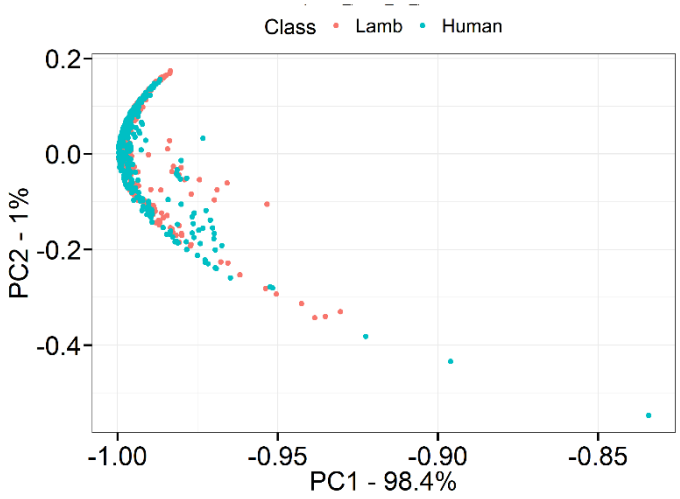

Figure 4.15: *PCA score plot for Modern Domestic Sheep and Human Teeth, using 100 pixels and unit vector normalization.*

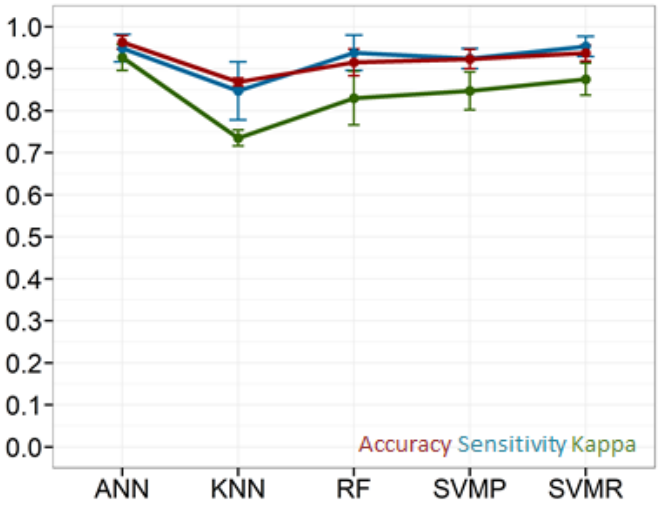

Figure 4.16: *Domestic Sheep and Human dataset, using further reduced biogenic spectral areas, corresponding in 100 pixels. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. Artificial Neural Network has slightly higher results, while all the algorithms except K Nearest Neighbors have accuracy higher than 0.9.*

Table 4.7: *Metric values for the Domestic Sheep and Human data, selecting further reduced biogenic spectral areas, corresponding in 100 pixels.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.868	0.01	0.847	0.069	0.735	0.019
SVMR	0.937	0.019	0.953	0.024	0.875	0.038
SVMP	0.923	0.023	0.924	0.024	0.847	0.045
RF	0.915	0.032	0.938	0.042	0.83	0.064
ANN	0.963	0.016	0.949	0.033	0.927	0.031

Two of the three remaining datasets, the Archaeological (Figure 4.17) and the Modern Domestic Sheep dataset (Figure 4.19), retain their results even after a similar reduction. For the Archaeological data ANN and RF have significantly higher results than the rest of the algorithms. Random Forest achieved an Accuracy value over 0.75 while the remaining metric values were around 0.7. ANN achieved an Accuracy value higher than 0.85 and while the remaining metric values exceeded 0.8 (Figure 4.18, Table 4.8). Using the Modern Domestic Sheep data an agreement of the produced results with those of the 258 pixels was observed. Once again, the ANN had the higher results, with its Accuracy and Sensitivity exceeding 0.7, followed by RF with an Accuracy value around 0.6. Those results remained close to the resulted values using 258 pixels (Figure 4.20, Table 4.9).

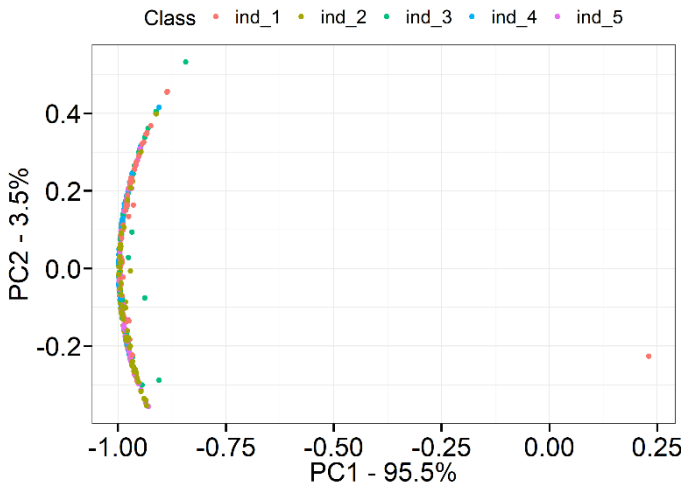

Figure 4.17: *PCA score plot for Archaeological Human Bones, using 100 pixels and unit vector normalization.*

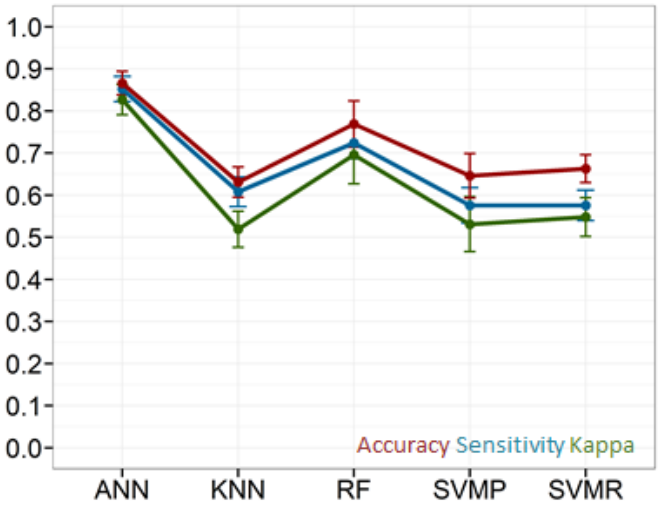

Figure 4.18: *Archaeological Human dataset, selecting further reduced biogenic spectral areas, corresponding in 100 pixels. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural network has significantly better results than any of the other algorithms.*

Table 4.8: *Metric values for the Archaeological Human data, with further reduced biogenic spectral areas, corresponding in 100 pixels.*

	Accuracy	*SD*	*Sensitivity*	*SD*	*Kappa*	*SD*
KNN	0.631	0.036	0.608	0.035	0.519	0.043
SVMR	0.663	0.033	0.576	0.036	0.548	0.046
SVMP	0.646	0.053	0.576	0.042	0.531	0.065
RF	0.769	0.055	0.724	0.038	0.696	0.069
ANN	0.866	0.028	0.852	0.03	0.827	0.036

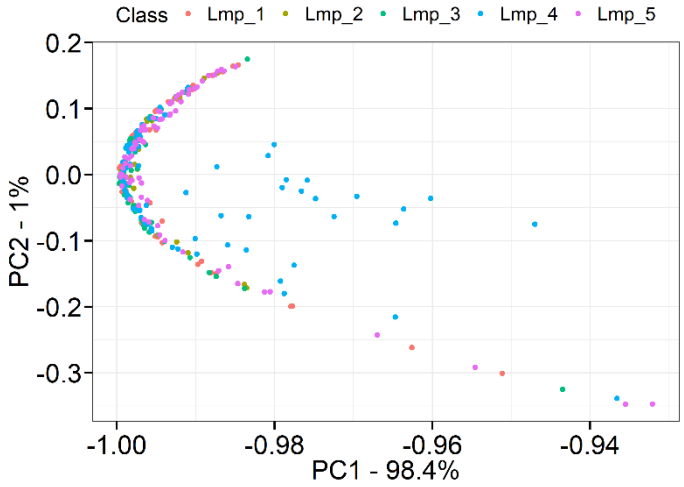

Figure 4.19: *PCA score plot for Modern Domestic Sheep Teeth, selecting 100 pixels and unit vector normalization.*

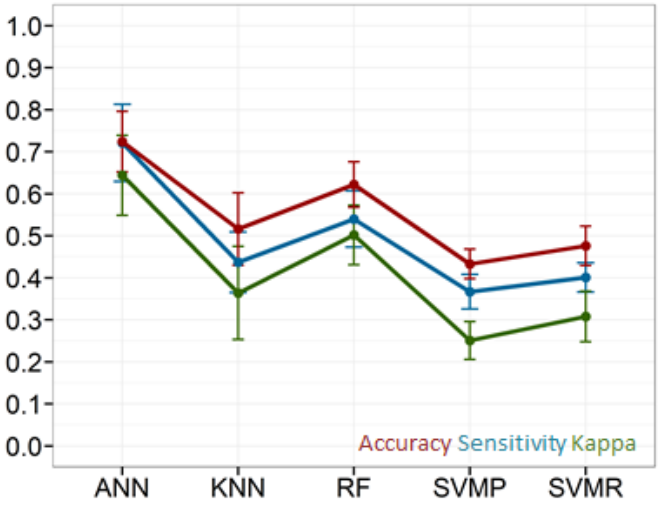

Figure 4.20: *Domestic Sheep dataset, selecting further reduced biogenic spectral areas, corresponding in 100 pixels. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural network has significantly better results than any of the other algorithms.*

Table 4.9: *Metric values for the Domestic Sheep data, selecting further reduced biogenic spectral areas*, corresponding in 100 pixels.

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.516	0.086	0.437	0.072	0.364	0.111
SVMR	0.476	0.047	0.401	0.035	0.308	0.06
SVMP	0.433	0.035	0.367	0.041	0.251	0.045
RF	0.622	0.054	0.54	0.067	0.502	0.071
ANN	0.724	0.072	0.721	0.092	0.644	0.095

The remaining Human Modern data (Figure 4.21) had a small but important increase in the resulted Accuracy value, exceeding 0.75 with the use of the ANN, while the rest of the metric values increased, as well. Random Forest followed ANN, with the remaining algorithms achieving insufficient results (Figure 4.22, Table 4.10).

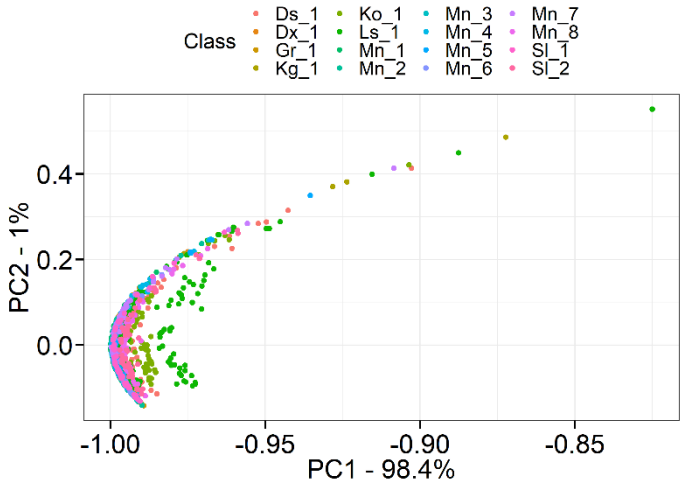

Figure 4.21: *PCA score plot for Modern Human Teeth, using 100 pixels and unit vector normalization.*

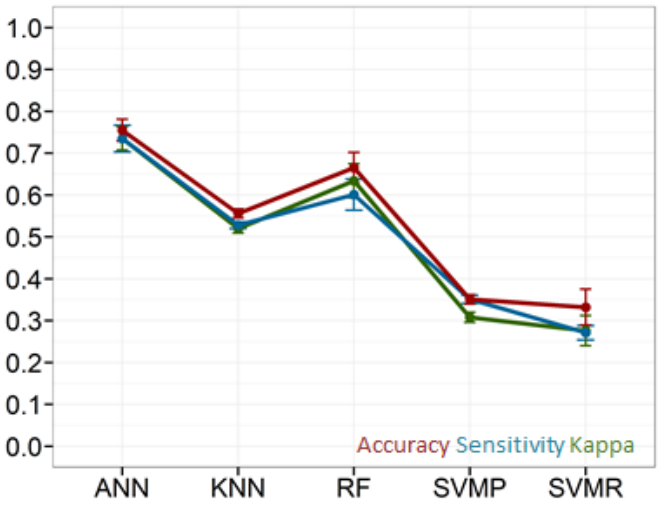

Figure 4.22: *Modern Human dataset, selecting further reduced biogenic spectral areas, corresponding in 100 pixels. The values of the three different metrics that were used, for each*

of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural network has significantly better results than any of the other algorithms.

Table 4.10: *Metric values for the Modern Human data, selecting further reduced biogenic spectral areas*, corresponding in 100 pixels.

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.556	0.011	0.528	0.008	0.52	0.011
SVMR	0.332	0.043	0.271	0.017	0.276	0.036
SVMP	0.351	0.011	0.351	0.01	0.308	0.012
RF	0.665	0.037	0.601	0.037	0.634	0.041
ANN	0.755	0.026	0.735	0.032	0.735	0.028

4.2.3 Selection of 130 Pixels

The significantly reduced running time, selecting only 100 pixels, created the opportunity to increase the inserted information to the models by adding new spectral areas to the already used. For that reason, combined with the already selected spectral areas, three more spectral areas were used (Figure 4.23). All the new areas corresponded in spectral lines from biogenic elements. In particular, the spectral lines of Ca in 370.6 nm and in 373.7 nm and the spectral line of Sr in 407.8 nm.

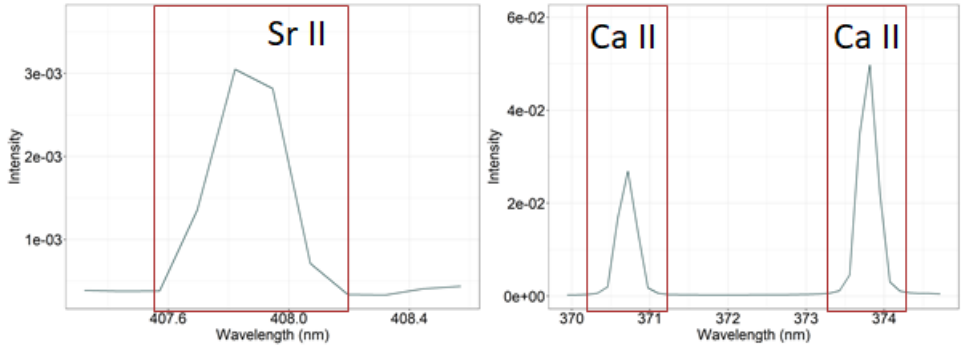

Figure 4.23: *Three more spectral areas were added in the previously selected (Figure 4.14), adding 30 more features.*

Thirty (30) new pixels were added in total in the previous 100 pixels, leading to performing models, using the Modern Domestic Sheep and the Modern Human data. In particular, using the Domestic Sheep dataset (Figure 4.24) the Accuracy values for all the models increased up to 0.1 (Figure 4.25, Table 4.11), depending on the algorithm, with the Artificial Neural Network reaching 0.8 and thus reached the limit for an efficient classification that has been set by previous works. Additionally, Sensitivity value for the ANN reached 0.8, too. Random Forest followed, reaching more than an Accuracy value of 0.65, with the rest of the model significantly improving their performance but retaining their insufficient behavior. Following this improvement, analyzing the Modern human data (Figure 4.26) the accuracy value for the ANN was increased by 0.07 exceeding 0.8, while the rest of its metric values

exceeded 0.8, as well. Exceeding the lowest threshold for an efficient classification for both Modern datasets was a key step for achieving the main aim of the project. Despite thee sufficiently good result of the ANN, the metric values for the remaining algorithms were significantly lower (Figure 4.27, Table 4.12).

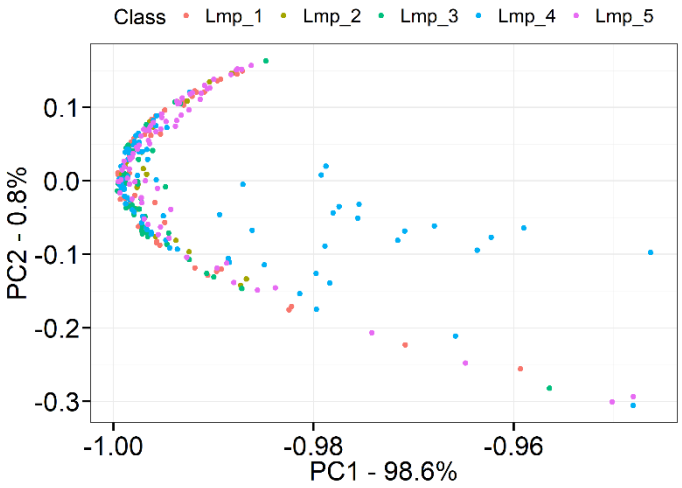

Figure 4.24: *PCA score plot for Modern Domestic Sheep Teeth, selecting 130 pixels and using unit vector normalization.*

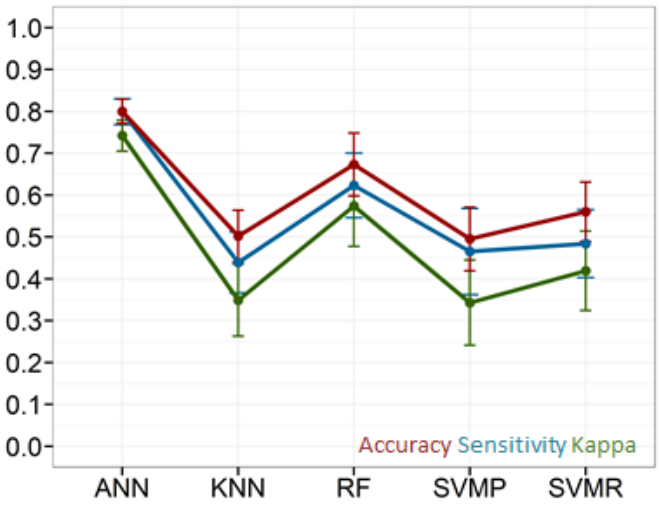

Figure 4.25: *Domestic Sheep dataset, selecting 130 pixels. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network has significantly better results than any of the other algorithms, followed by Random Forest.*

Table 4.11: *Metric values for the Domestic Sheep data, selecting 130 pixels.*

	Accuracy	*SD*	*Sensitivity*	*SD*	*Kappa*	*SD*
KNN	0.502	0.062	0.439	0.073	0.349	0.086
SVMR	0.56	0.071	0.484	0.081	0.419	0.095
SVMP	0.495	0.076	0.465	0.103	0.343	0.102
RF	0.673	0.075	0.623	0.077	0.574	0.097
ANN	0.8	0.029	0.799	0.031	0.742	0.037

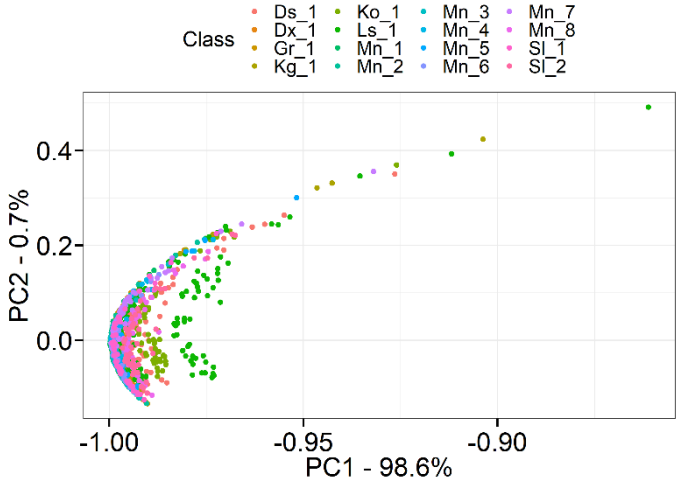

Figure 4.26*: PCA score plot for Modern Human Teeth, selecting 130 pixels and using unit vector normalization.*

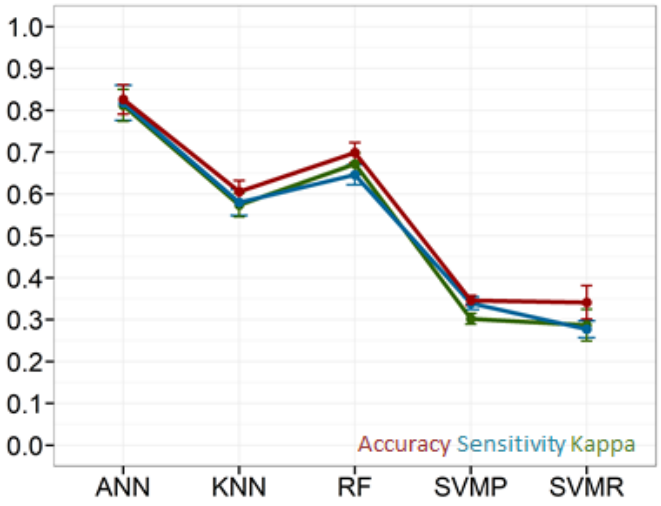

Figure 4.27: *Modern Human dataset, selecting 130 pixels. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network has significantly better results than any of the other algorithms.*

Table 4.12: *Metric values for the Modern Human data, selecting 130 pixels.*

	Accuracy	*SD*	*Sensitivity*	*SD*	*Kappa*	*SD*
KNN	0.606	0.026	0.58	0.031	0.574	0.028
SVMR	0.341	0.04	0.277	0.02	0.287	0.038
SVMP	0.346	0.012	0.339	0.015	0.302	0.013
RF	0.699	0.024	0.646	0.024	0.672	0.026
ANN	0.826	0.035	0.818	0.042	0.812	0.038

Furthermore, using the combined data (Figure 4.28) the models had excellent results with Accuracy values exceeding 0.94 (Figure 4.29, Table 4.13). The highest of the Accuracy values was achieved by ANN, which remained the best behaved algorithm reaching 0.97. On the other hand, the selection of the 130 pixels for the Archaeological data (Figure 4.30) occurred the impact of spectral lines by non-biogenic elements on the final results. It can be seen that adding the extra spectral areas for the analysis of the Archaeological data a small increase in the metric values of all used algorithms is revealed (Figure 4.31, Table 4.14). SVM and KNN had Accuracy values around 0.7, while the resulted Accuracy by using the Random Forest exceeded 0.8, providing a very good classification result. As it was expected, the ANN achieved a classification Accuracy over 0.87, exceeding the value of RF.

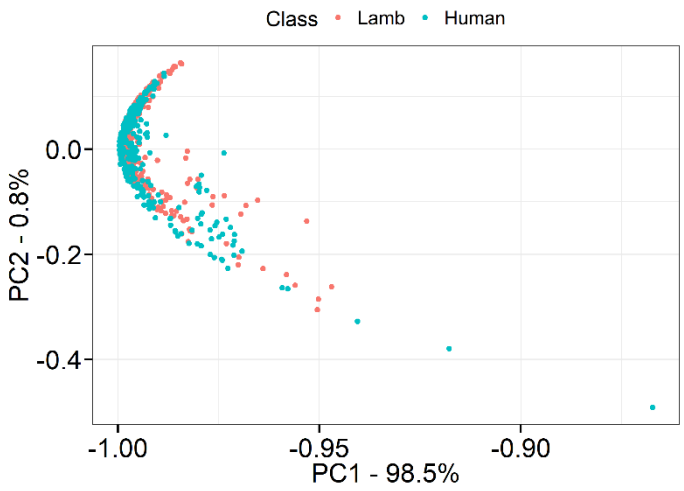

Figure 4.28: *PCA score plot for Domestic Sheep and Human Teeth, selecting 130 pixels and using unit vector normalization.*

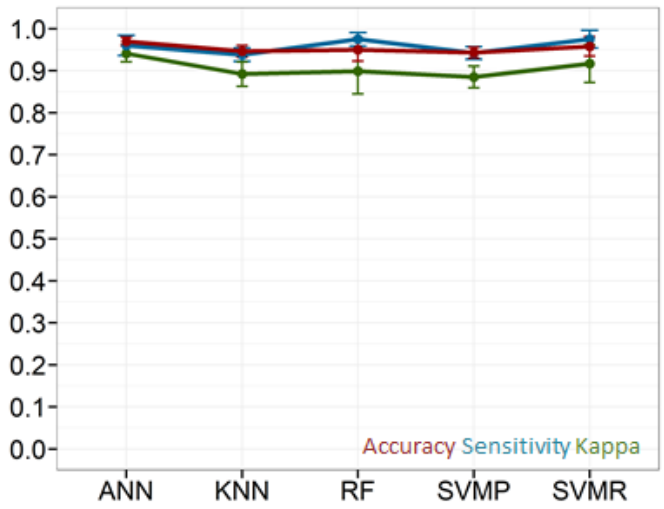

Figure 4.29: *Domestic Sheep and Human dataset, selecting 130 pixels. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. Artificial Neural Network has slightly higher results, while remaining algorithms give almost the same results.*

Table 4.13: *Metric values for the Domestic Sheep and Human data, selecting 130 pixels.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.946	0.014	0.938	0.016	0.892	0.029
SVMR	0.958	0.023	0.975	0.021	0.917	0.045
SVMP	0.943	0.013	0.942	0.015	0.885	0.026
RF	0.95	0.027	0.975	0.016	0.899	0.054
ANN	0.97	0.01	0.96	0.024	0.941	0.02

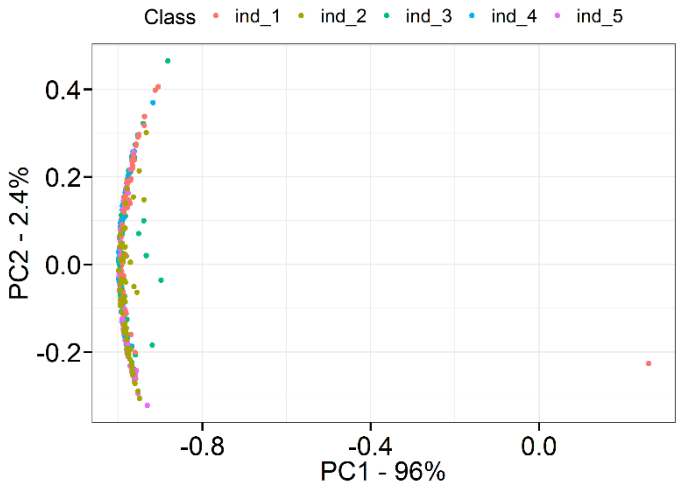

Figure 4.30: *PCA score plot for Archaeological Human Bones, selecting 130 pixels and using unit vector normalization.*

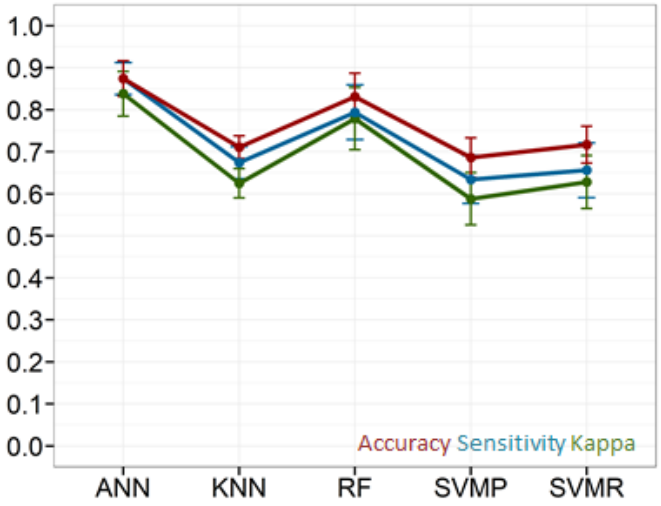

Figure 4.31: *Archaeological Human dataset, selecting 130 pixels. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network and Random Forest have significantly better results than any of the other algorithms, with the ANN having slightly higher results than RF.*

Table 4.14: *Metric values for the Archaeological Human data, using 130 pixels.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.711	0.027	0.674	0.038	0.625	0.035
SVMR	0.717	0.044	0.656	0.065	0.628	0.063
SVMP	0.686	0.047	0.634	0.057	0.588	0.062
RF	0.831	0.056	0.794	0.065	0.779	0.074
ANN	0.874	0.042	0.874	0.038	0.838	0.053

4.3 Feature Selection by Variance

After achieving a good classification Accuracy for all datasets, using the 130 selected pixels, further feature selection and further reduction of running time was attempted. As it was already discussed in the section 3.8, a feature selection function using variance as the selection factor was developed, in order to perform a drastic feature reduction to 10 final features. Two different approaches were examined. Firstly, the further feature reduction was performed starting from the 100 pixels and then starting from the 130 pixels that had been previously used.

4.3.1 Feature Selection by Variance on 100 Pixels

The heavy feature reduction, from 100 to 10, by using the newly developed feature selection function, neither improved nor retained the sufficient results of the algorithms by using the 100 preselected pixels. The results for the three datasets that were tested for discrimination between individuals had been drastically reduced. The Archaeological dataset (Figure 4.32) had the lowest results since none of the models achieved classification Accuracy over 0.55 (Figure 4.33, Table 4.15). The highest

Accuracy values were achieved by the Artificial Neural Network and the Random Forest, reaching slightly lower than 0.55. The remaining algorithms reached even lower Accuracy values. The ANN for the Modern Domestic Sheep dataset (Figure 4.34) managed to keep a sufficient Accuracy value, slightly lower than 0.7, but much reduced compared to the value of the model using all the 100 pixels. The rest of the algorithms achieved Accuracy values lower than 0.65 and some even lower than 0.6 (Figure 4.35, Table 4.16). For the Modern Human dataset (Figure 4.36), the results were equally low than the previous two datasets (Figure 4.37, Table 4.17). Once again, the ANN and the RF reached the highest Accuracy values, slightly over 0.6, but without reaching the previously achieved values using 100 pixels. The remaining algorithms had quite low results.

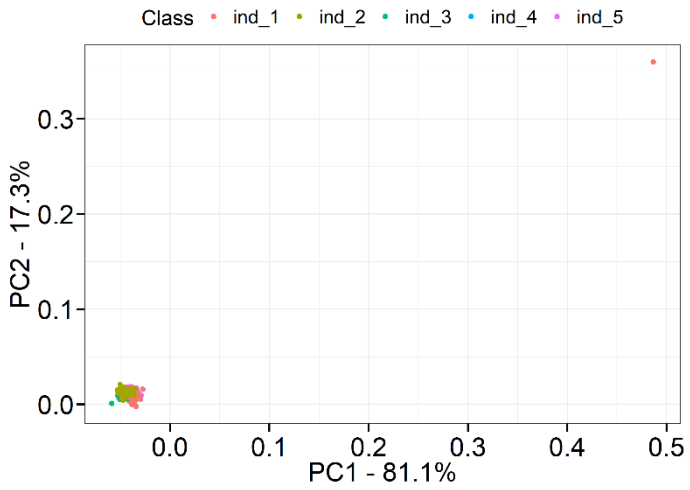

Figure 4.32: *PCA score plot for Archaeological Human Bones, selecting 10 from 100 pixels with the lowest in-class variance and the highest among classes variance.*

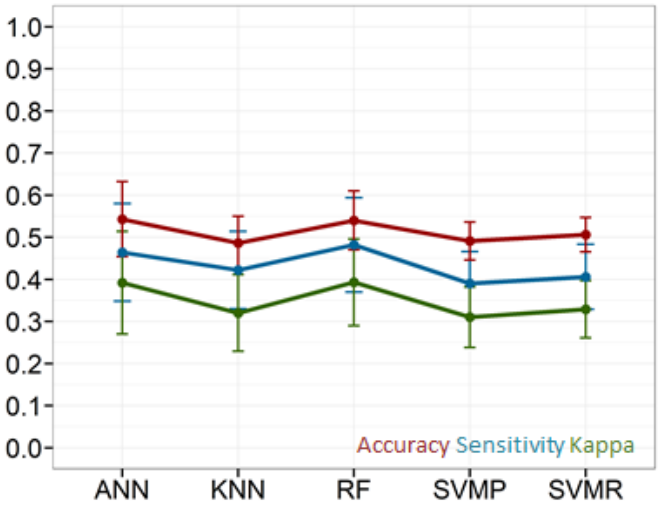

Figure 4.33: *Archaeological Human dataset, selecting 10 from 100 pixels with the lowest in-class variance and the highest among classes variance. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network and the Random Forest have slightly better results than any of the other algorithms.*

Table 4.15: *Metric values for the Archaeological Human data, selecting 10 from 100 pixels with the lowest in-class variance and the highest among classes variance.*

	Accuracy	*SD*	*Sensitivity*	*SD*	*Kappa*	*SD*
KNN	0.486	0.064	0.422	0.092	0.32	0.091
SVMR	0.506	0.041	0.406	0.077	0.329	0.068
SVMP	0.491	0.045	0.39	0.076	0.31	0.072
RF	0.54	0.07	0.482	0.112	0.393	0.103
ANN	0.543	0.089	0.464	0.116	0.392	0.122

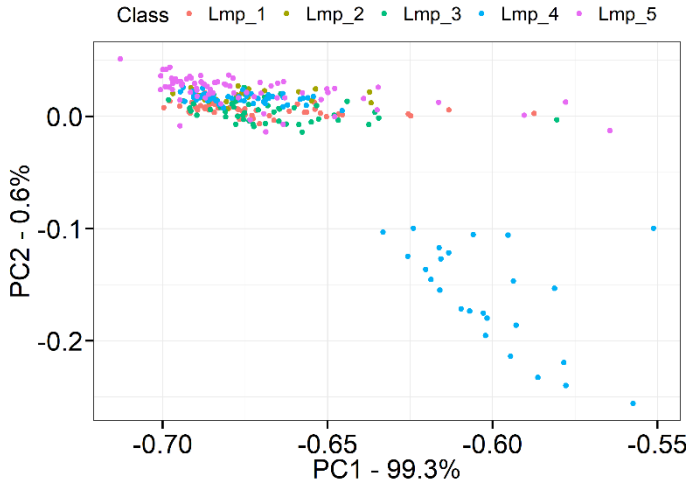

Figure 4.34: *PCA score plot for Modern Domestic Sheep Teeth, selecting 10 from 100 pixels with the lowest in-class variance and the highest among classes variance.*

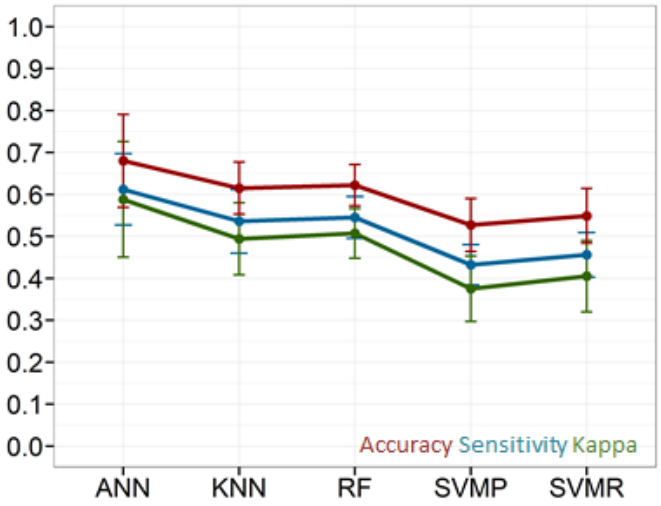

Figure 4.35: *Modern Domestic Sheep dataset, selecting 10 from 100 pixels with the lowest in-class variance and the highest among classes variance. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network has significantly better results than any of the other algorithms.*

Table 4.16: Metric values for the *Modern Domestic Sheep data, selecting 10 from 100 pixels with the lowest in-class variance and the highest among classes variance.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.615	0.062	0.536	0.076	0.494	0.086
SVMR	0.549	0.065	0.456	0.053	0.405	0.085
SVMP	0.527	0.063	0.432	0.048	0.375	0.078
RF	0.622	0.049	0.545	0.05	0.507	0.059
ANN	0.68	0.111	0.612	0.085	0.588	0.138

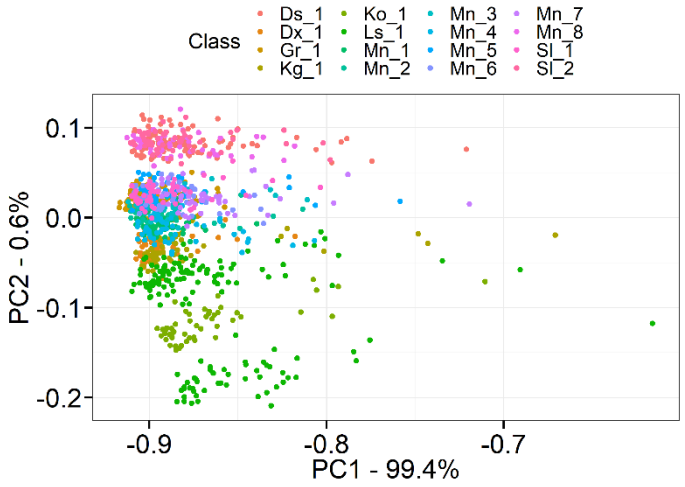

Figure 4.36: *PCA score plot for Modern Human Teeth, selected 10 from 100 pixels with the lowest in-class variance and the highest among classes variance.*

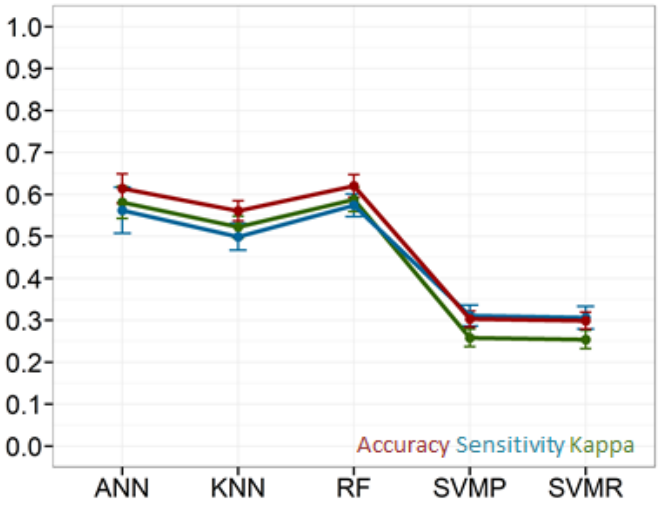

Figure 4.37: *Modern Human dataset, selected 10 from 100 pixels with the lowest in-class variance and the highest among classes variance. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network and Random forest have slightly better results than any of the other algorithms.*

Table 4.17: Metric values for the *Modern Human data, selected 10 from 100 pixels with the lowest in-class variance and the highest among classes variance.*

	Accuracy	*SD*	*Sensitivity*	*SD*	*Kappa*	*SD*
KNN	0.561	0.024	0.499	0.032	0.522	0.026
SVMR	0.299	0.02	0.306	0.027	0.254	0.022
SVMP	0.303	0.019	0.311	0.025	0.258	0.021
RF	0.62	0.027	0.574	0.027	0.588	0.029
ANN	0.614	0.035	0.562	0.055	0.581	0.038

The combined dataset (Figure 4.38) was the only dataset that maintained sufficient results regardless of the model that was used for its analysis. Its results were significantly lower than the previously achieved (using 100 pixels), but all the algorithms exceeded the threshold of 0.8 (Figure 4.39, Table 4.18). ANN had the highest performance with an Accuracy value of 0.87, while the rest of the algorithms had Accuracy values lower than 0.85.

Figure 4.38: *PCA score plot for Domestic Sheep and Human Teeth, selecting 10 from 100 pixels with the lowest in-class variance and the highest among classes variance.*

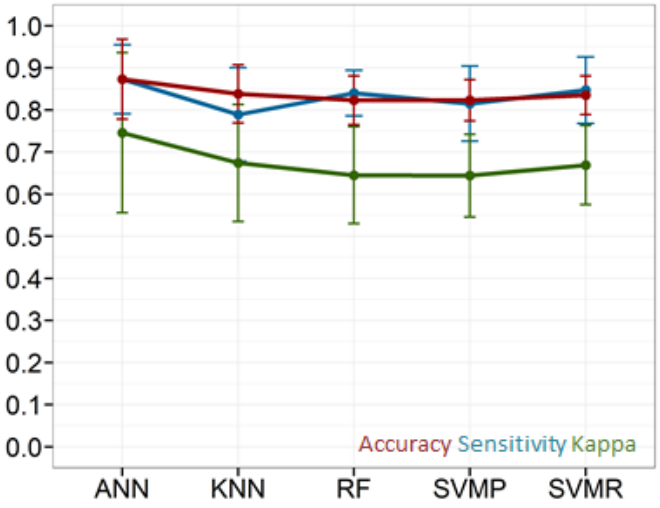

Figure 4.39: *Domestic Sheep and Human dataset, selecting 10 from 100 pixels with the lowest in-class variance and the highest among classes variance. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network has slightly better results than any of the other algorithms.*

Table 4.18: Metric values for the *Domestic Sheep and Human data, selecting 10 from 100 pixels with the lowest in-class variance and the highest among classes variance.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.838	0.069	0.789	0.111	0.674	0.139
SVMR	0.835	0.046	0.847	0.079	0.669	0.094
SVMP	0.823	0.049	0.815	0.089	0.644	0.098
RF	0.823	0.058	0.84	0.054	0.645	0.115
ANN	0.873	0.095	0.873	0.082	0.746	0.19

4.3.2 Feature Selection by Variance on 130 Pixels

The further feature selection starting from 130 resulted in lower metric values than using all the 130 pixels. The three datasets that were tested for discrimination between individuals, could not achieve reliable classification by any of the used algorithms. In particular, for the Archaeological dataset (Figure 4.40) the Random Forest reached the higher classification Accuracy with a value slightly lower than 0.55, while ANN and KNN followed with Accuracy values slightly lower than that of the RF (Figure 4.41, Table 4.19). For the Modern datasets, the results were higher than that of the Archaeological dataset but were lower than using the 130 pixels. Analyzing the Modern Domestic Sheep data (Figure 4.42) the higher Accuracy value was 0.68 by the ANN, while Random Forest had almost the same value. The rest of the algorithms achieved classification values lower than 0.6 (Figure 4.43, Table 4.20). In a similar situation, for the Modern Human dataset (Figure 4.44) none of the algorithms achieve a reliable classification (Figure 4.45, Table 4.21). The ANN had the higher Accuracy

value, almost 0.65, followed by the RF and KNN with 0.6. The SVM algorithm had significantly lower results.

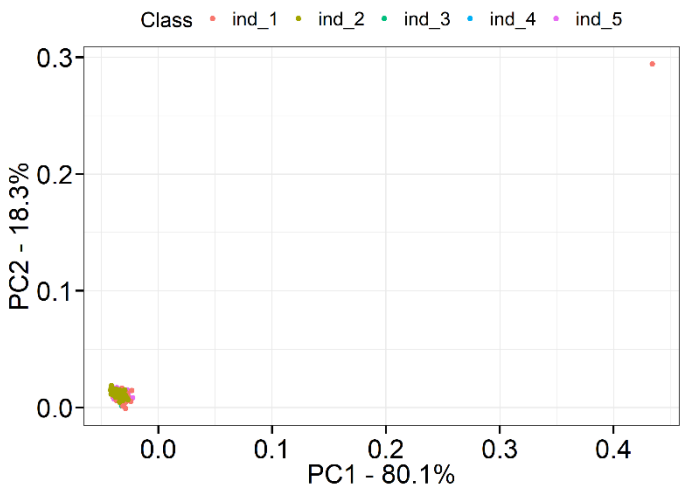

Figure 4.40: *PCA score plot for Archaeological Human Bones, selecting 10 from 130 pixels with the lowest in-class variance and the highest among classes variance.*

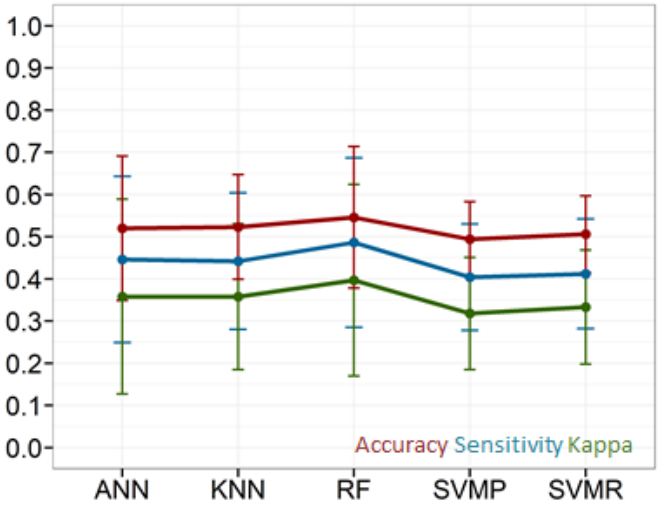

Figure 4.41: *Archaeological Human dataset, selecting 10 from 130 pixels with the lowest in-class variance and the highest among classes variance. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Random Forest has slightly better results than the other algorithms.*

Table 4.19: *Metric values for the Archaeological Human data, selecting 10 from 130 pixels with the lowest in-class variance and the highest among classes variance.*

	Accuracy	*SD*	*Sensitivity*	*SD*	*Kappa*	*SD*
KNN	0.523	0.124	0.442	0.162	0.358	0.173
SVMR	0.506	0.091	0.412	0.13	0.333	0.135
SVMP	0.494	0.089	0.404	0.126	0.318	0.133
RF	0.546	0.168	0.486	0.201	0.397	0.227
ANN	0.52	0.171	0.446	0.197	0.358	0.231

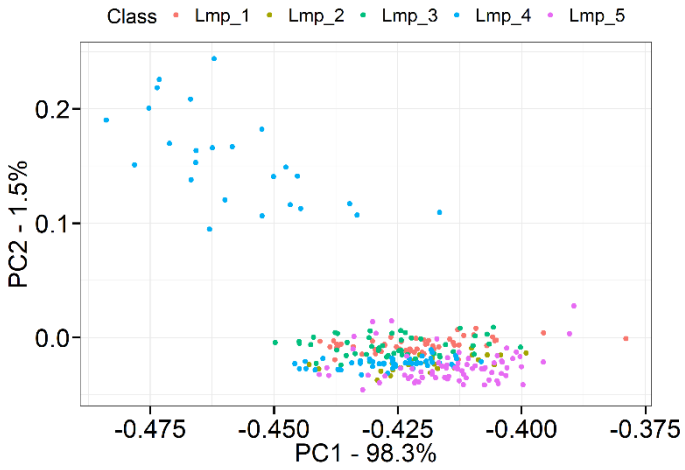

Figure 4.42: *PCA score plot for Modern Domestic Sheep Teeth, selecting 10 from 130 pixels with the lowest in-class variance and the highest among classes variance.*

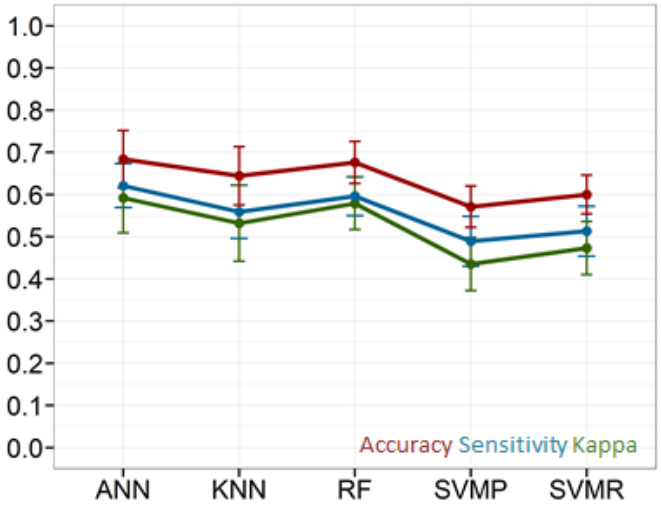

Figure 4.43: *Modern Domestic Sheep dataset, selecting 10 from 130 pixels with the lowest in-class variance and the highest among classes variance. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network and the Random Forest have slightly better results than the other algorithms.*

Table 4.20: *Metric values for the Modern Domestic Sheep data, selecting 10 from 130 pixels with the lowest in-class variance and the highest among classes variance.*

	Accuracy	*SD*	*Sensitivity*	*SD*	*Kappa*	*SD*
KNN	0.644	0.069	0.559	0.063	0.532	0.09
SVMR	0.6	0.046	0.513	0.059	0.473	0.063
SVMP	0.571	0.049	0.489	0.059	0.435	0.063
RF	0.676	0.05	0.596	0.046	0.579	0.062
ANN	0.684	0.068	0.621	0.052	0.592	0.083

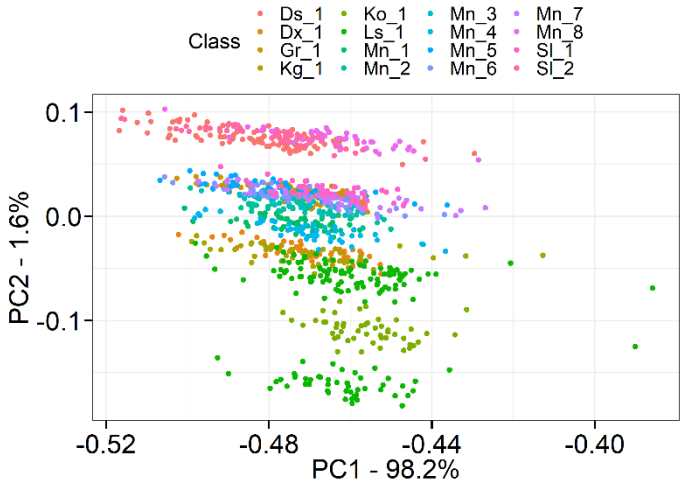

Figure 4.44: *PCA score plot for Modern Human Teeth, selecting 10 from 130 pixels with the lowest in-class variance and the highest among classes variance.*

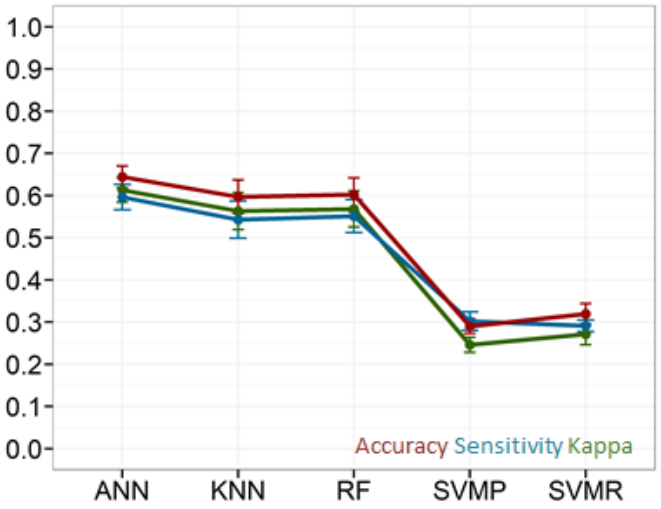

Figure 4.45: *Modern Human dataset, selecting 10 from 130 pixels with the lowest in-class variance and the highest among classes variance. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network has slightly better results than the other algorithms.*

Table 4.21: *Metric values for the Modern Human data, selecting 10 from 130 pixels with the lowest in-class variance and the highest among classes variance.*

	Accuracy	*SD*	*Sensitivity*	*SD*	*Kappa*	*SD*
KNN	0.597	0.04	0.543	0.044	0.563	0.044
SVMR	0.319	0.025	0.291	0.014	0.271	0.025
SVMP	0.29	0.017	0.302	0.022	0.246	0.018
RF	0.602	0.04	0.551	0.039	0.568	0.043
ANN	0.644	0.026	0.596	0.03	0.613	0.029

On the contrary, the behavior of the models for the discrimination between species on the combined dataset (Figure 4.46) was very good. All the models achieved Accuracy values between 0.89 and 0.93, with the RF achieving the higher results followed by the ANN and the KNN (Figure 4.47, Table 4.22). Despite the fact that the resulted classification was sufficiently good, the resulted metric values were significantly lower than those achieved by using 130 pixels.

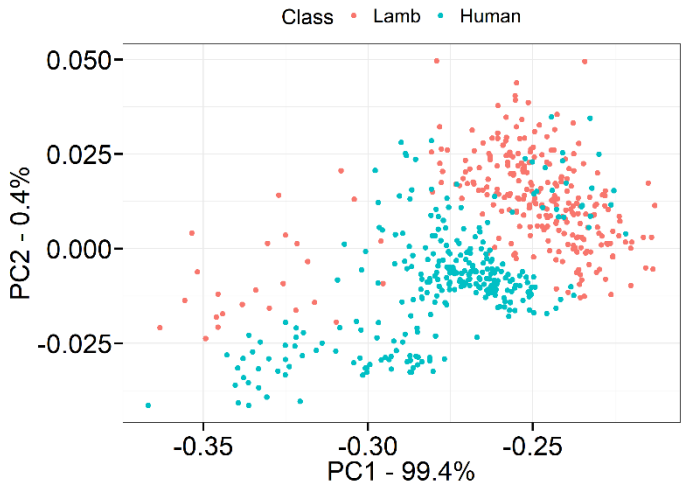

Figure 4.46: *PCA score plot for Domestic Sheep and Human Teeth, selecting 10 from 130 pixels with the lowest in-class variance and the highest among classes variance.*

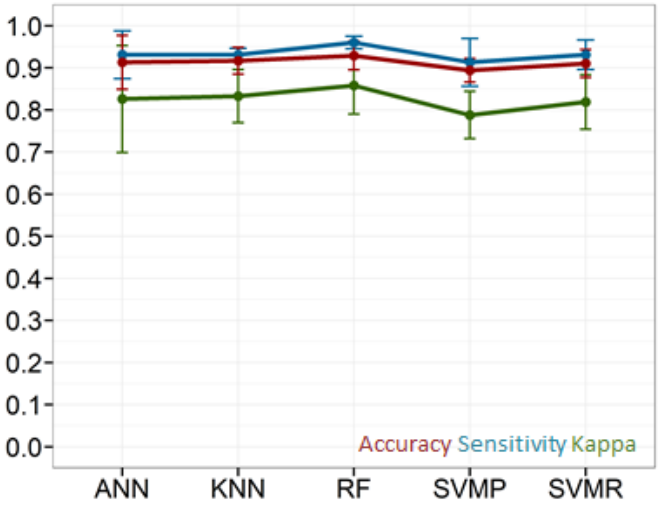

Figure 4.47: *Domestic Sheep and Human dataset, selecting 10 from 130 pixels with the lowest in-class variance and the highest among classes variance. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Random Forest has slightly better results than any of the other algorithms.*

Table 4.22: *Metric values for the Domestic Sheep and Human data selecting 10 from 130 pixels with the lowest in-class variance and the highest among classes variance.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.917	0.032	0.931	0.015	0.833	0.063
SVMR	0.91	0.033	0.931	0.035	0.819	0.065
SVMP	0.894	0.028	0.913	0.057	0.788	0.056
RF	0.929	0.034	0.96	0.015	0.858	0.068
ANN	0.913	0.064	0.931	0.057	0.826	0.127

4.4 Spectral Line Intensities

Following a different approach for a drastic feature selection, than the previously used, targeting a classification results for all four datasets, only the pixels with the maximum intensity from specific spectral lines were selected as important features (Figure 4.48). This feature selection has been already discussed in section 3.8. In particular, the spectral lines that were selected corresponded in spectral areas that were determined by the 258 and 100 pixels. Na's spectral line was excluded because of its low intensity.

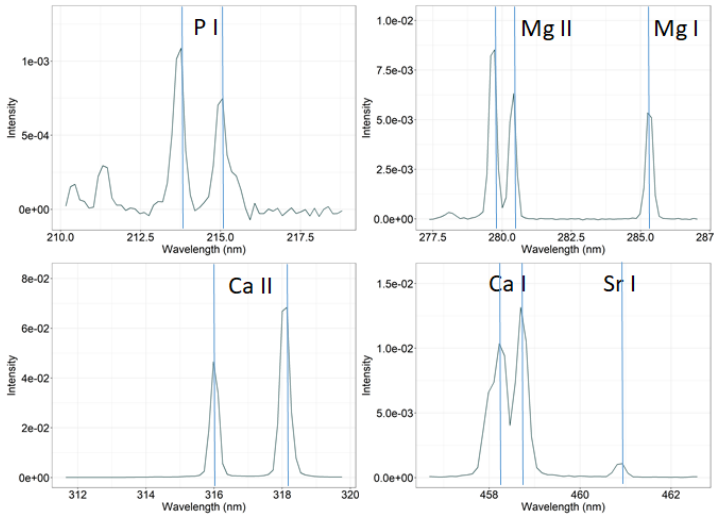

Figure 4.48: *Selecting only pixel with the highest intensity values from each spectral line.*

By using this feature selection, the classification results for all datasets were significantly decreased with the models failing to reach Accuracy values over 0.55 for classification in the different individuals. Using the Archaeological data (Figure 4.49), the highest Accuracy value of 0.55 was achieved by the ANN, while the rest of the models failed to achieve higher values than 0.5 (Figure 4.50, Table 4.23). Because of the low Accuracy values, the Kappa values were significantly low, as well. All the algorithms had Accuracy values lower than 0.6 during the analysis of the Domestic Sheep data (Figure 4.51, Figure 4.52, Table 4.24). The ANN was the highest scored algorithm having an Accuracy value slightly higher than 0.55, while the rest of the algorithms achieved Accuracy values lower than 0.55. In a similar situation, the Modern Human data (Figure 4.53) had the highest Accuracy value slightly over 0.45, using the ANN that was closely followed by the RF. The remaining algorithms scored very low results (Figure 4.54, Table 4.25). As a result, none of the algorithms was able for a dissent classification regardless the used data.

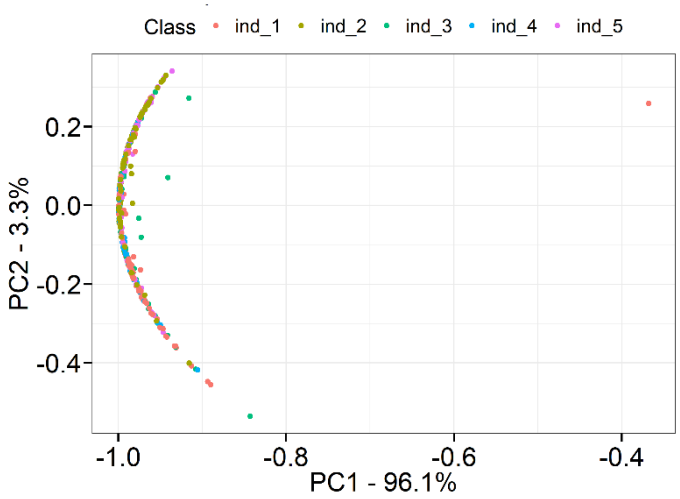

Figure 4.49: *PCA score plot for Archaeological Human Bones, selecting only the pixel with the highest intensity for specific biogenic spectral lines.*

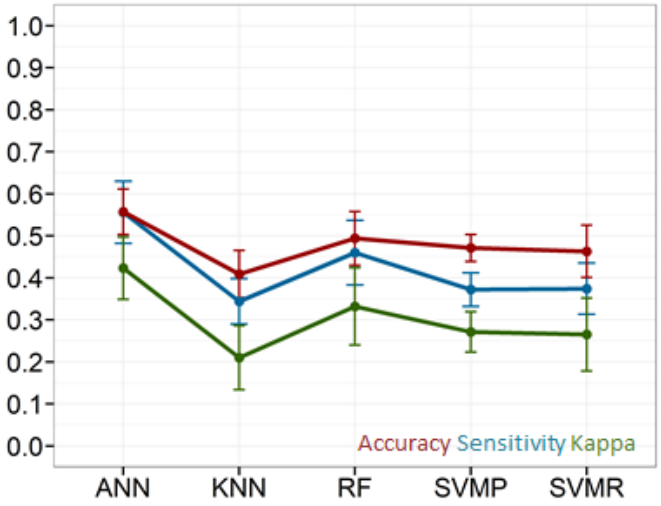

Figure 4.50: *Archaeological Human dataset, selecting only the pixel with the highest intensity for specific biogenic spectral lines. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network has significantly better results than any of the other algorithms.*

Table 4.23: *Metric values for the Archaeological Human data, selecting only the pixel with the highest intensity for specific biogenic spectral lines.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.409	0.056	0.344	0.054	0.21	0.076
SVMR	0.463	0.062	0.374	0.061	0.265	0.087
SVMP	0.471	0.032	0.372	0.04	0.271	0.048
RF	0.494	0.064	0.46	0.077	0.332	0.092
ANN	0.557	0.054	0.556	0.074	0.423	0.074

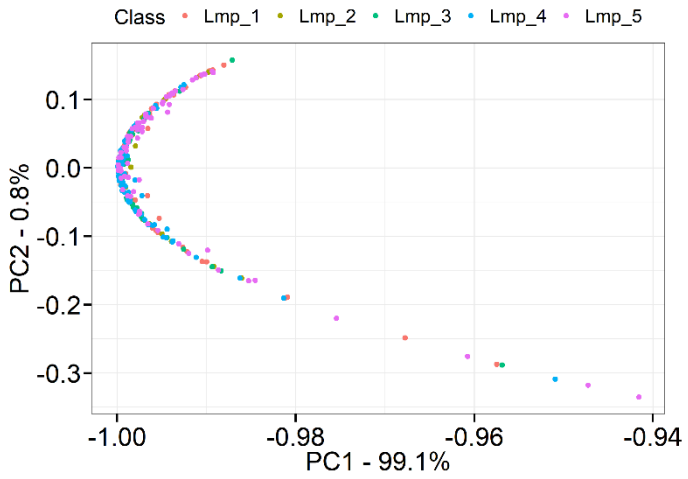

Figure 4.51: *PCA score plot for Modern Domestic Sheep Teeth, selecting only the pixel with the highest intensity for specific biogenic spectral lines.*

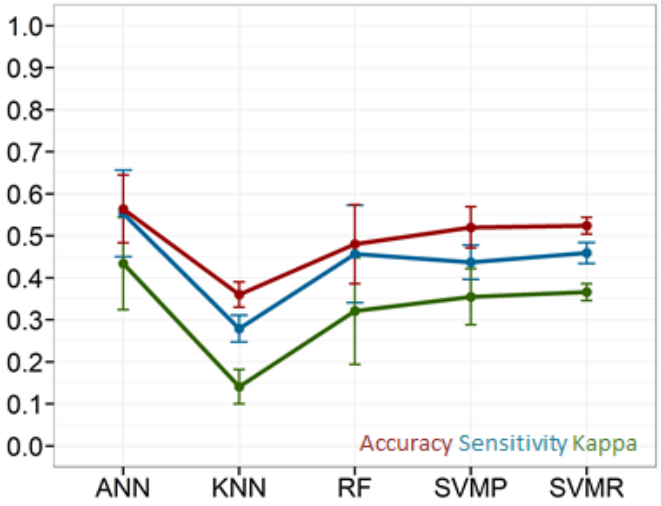

Figure 4.52: *Domestic Sheep dataset, selecting only the pixel with the highest intensity for specific biogenic spectral lines. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network has slightly better results than the other algorithms.*

Table 4.24: *Metric values for the Domestic Sheep data, selecting only the pixel with the highest intensity for specific biogenic spectral lines.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.36	0.03	0.279	0.032	0.141	0.041
SVMR	0.524	0.02	0.459	0.025	0.366	0.02
SVMP	0.52	0.049	0.437	0.041	0.355	0.067
RF	0.48	0.094	0.457	0.116	0.321	0.127
ANN	0.564	0.081	0.553	0.103	0.434	0.11

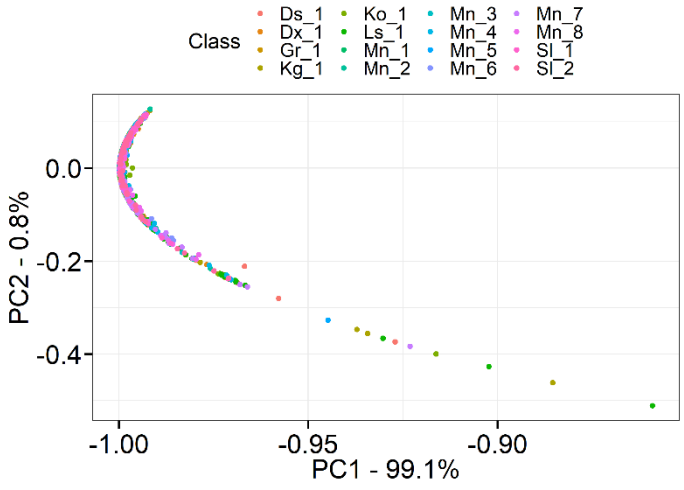

Figure 4.53: *PCA score plot for Modern Human Teeth, selecting only the pixel with the highest intensity for specific biogenic spectral lines.*

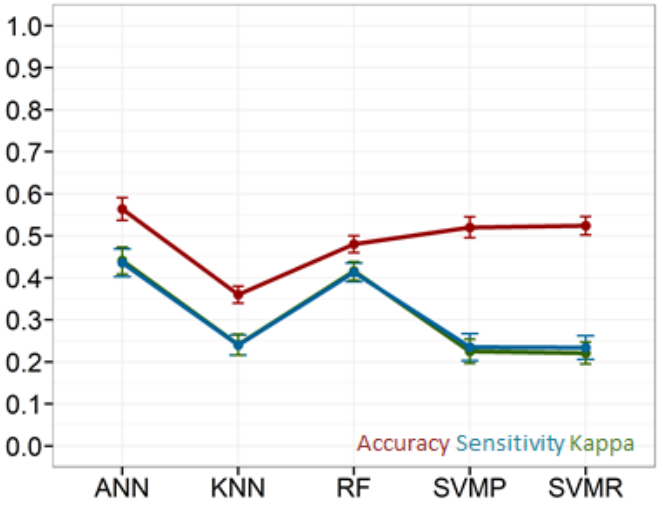

Figure 4.54: *Modern Human dataset, selecting only the pixel with the highest intensity for specific biogenic spectral lines. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network has slightly better results than the other algorithms.*

Table 4.25: *Metric values for the Modern Human data, selecting only the pixel with the highest intensity for specific biogenic spectral lines.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.307	0.02	0.24	0.024	0.241	0.025
SVMR	0.282	0.022	0.234	0.028	0.221	0.026
SVMP	0.285	0.025	0.235	0.032	0.225	0.029
RF	0.464	0.02	0.413	0.022	0.416	0.022
ANN	0.487	0.027	0.436	0.033	0.441	0.032

Using the combined dataset (Figure 4.55), all the metric values were significantly lower than those of the previous approaches (Figure 4.56, Table 4.26). The Accuracy values were slightly lower than 0.8 for all algorithms apart from ANN that slightly exceeded this value. This significantly decrease in the Accuracy values indicates that the created models have very low classification abilities, using this particular feature selection.

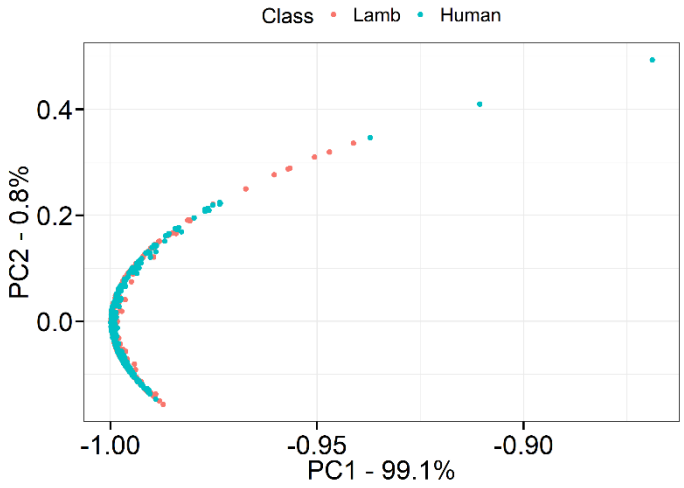

Figure 4.55: *PCA score plot for Domestic Sheep and Human Teeth, selecting only the pixel with the highest intensity for specific biogenic spectral lines.*

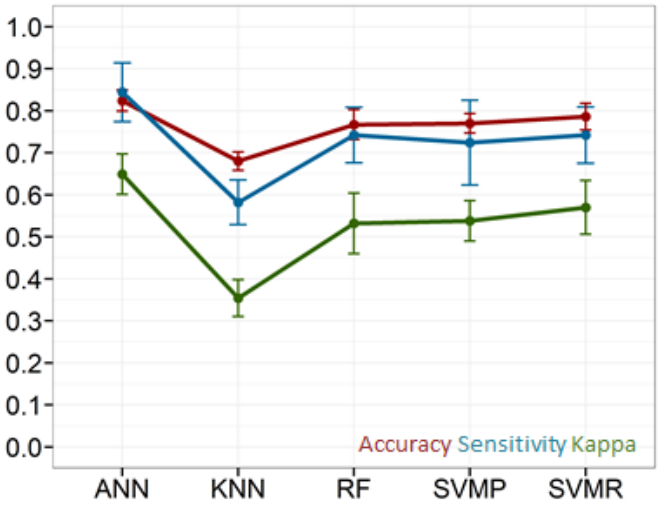

Figure 4.56: *Domestic Sheep and Human dataset, selecting only the pixel with the highest intensity for specific biogenic spectral lines. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network has slightly better results than any of the other algorithms.*

Table 4.26: *Metric values for the Domestic Sheep and Human data, selecting only the pixel with the highest intensity for specific biogenic spectral lines.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.68	0.022	0.582	0.053	0.354	0.044
SVMR	0.786	0.032	0.742	0.067	0.57	0.064
SVMP	0.77	0.023	0.724	0.101	0.538	0.048
RF	0.767	0.036	0.742	0.066	0.532	0.072
ANN	0.824	0.025	0.844	0.07	0.649	0.048

4.5 Feature Extraction by Calculating Integrals

In a different approach, a feature extraction technique was used. This method has already been discussed in section 3.8 and the resulting features were 10 integral values calculated by the same spectral lines that had been used (Figure 4.57) in section 4.4. The classification results were even lower than the previous case, resulting in algorithms that could not be used for the desirable task.

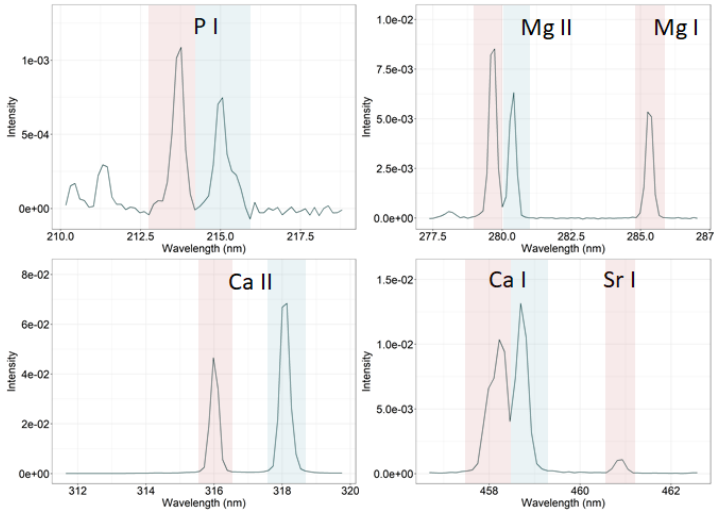

Figure 4.57: *Extracted features by the calculation of the integrals for biogenic spectral lines.*

All the Accuracy values were lower than 0.6 for the Archaeological (Figure 4.58, Figure 4.59, Table 4.27) and the Modern Domestic Sheep (Figure 4.60, Figure 4.61, Table 4.28) datasets, with the ANN achieving the higher values of 0.59 and 0.53, respectively. The results for the Modern Human dataset were even lower, considering that the highest Accuracy value was slightly higher than 0.4 (Figure 4.62, Figure 4.63, Table 4.29), achieved by ANN. Even the results for the combined data were significantly lower than the previous approaches, with all the algorithms except KNN achieving Accuracy values between 0.75 and 0.8 (Figure 4.64, Figure 4.65, Table 4.30). Those results proved the inability of the algorithms, using those extracted features, to create appropriate models that could achieve any reliable classification.

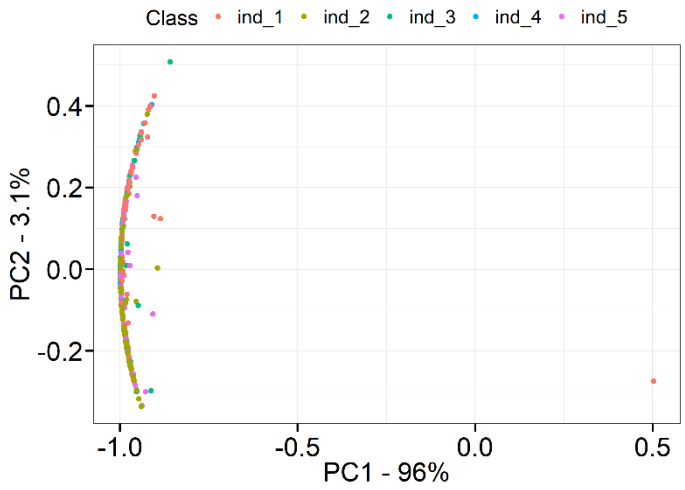

Figure 4.58: *PCA score plot for Archaeological Human Bones, using the integrals of specific biogenic spectral lines.*

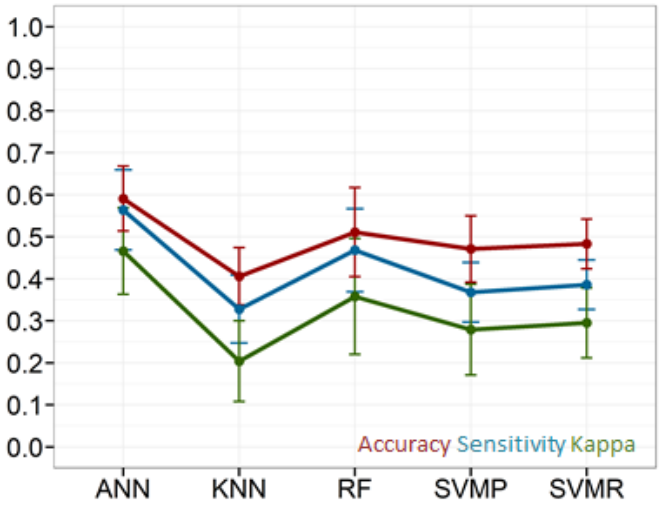

Figure 4.59: *Archaeological Human dataset, using the integrals of specific biogenic spectral lines. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network has significantly better results than any of the other algorithms.*

Table 4.27: *Metric values for the Archaeological Human data, using the integrals of specific biogenic spectral lines.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.406	0.068	0.328	0.081	0.204	0.096
SVMR	0.483	0.059	0.386	0.059	0.295	0.083
SVMP	0.471	0.079	0.368	0.071	0.279	0.108
RF	0.511	0.106	0.468	0.099	0.358	0.138
ANN	0.591	0.077	0.564	0.095	0.466	0.103

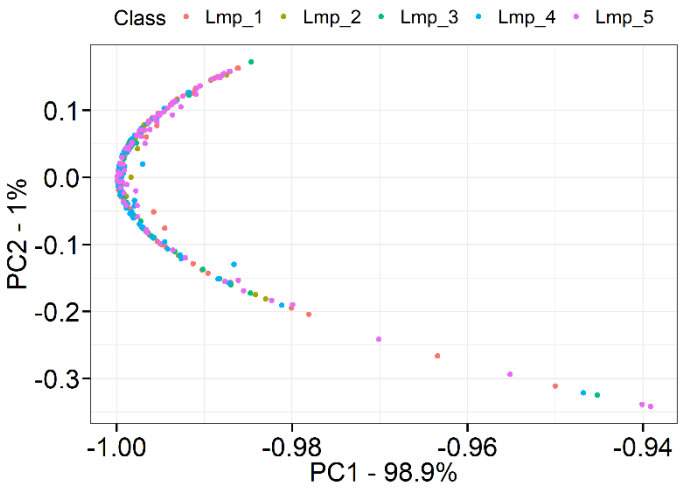

Figure 4.60: *PCA score plot for Modern Domestic Sheep Teeth, using the integrals of specific biogenic spectral lines.*

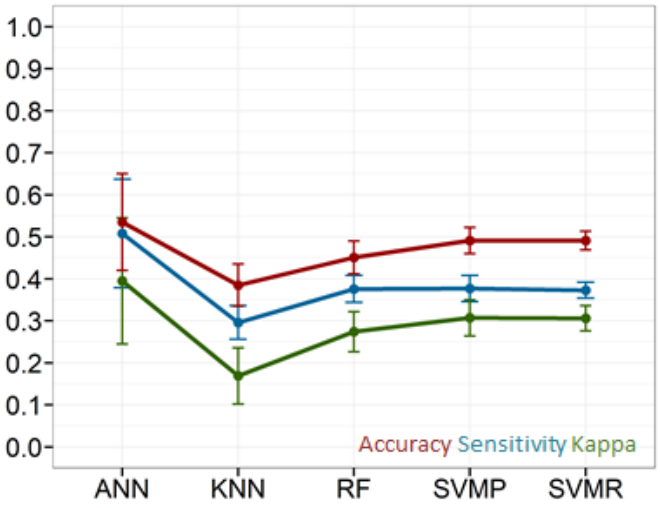

Figure 4.61: *Domestic Sheep dataset, using the integrals of specific biogenic spectral lines. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network has slightly better results than any of the other algorithms.*

Table 4.28: *Metric values for the Domestic Sheep data, using the integrals of specific biogenic spectral lines.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.385	0.05	0.296	0.04	0.169	0.067
SVMR	0.491	0.022	0.373	0.019	0.306	0.03
SVMP	0.491	0.031	0.377	0.031	0.307	0.043
RF	0.451	0.039	0.376	0.032	0.274	0.048
ANN	0.535	0.115	0.508	0.129	0.395	0.15

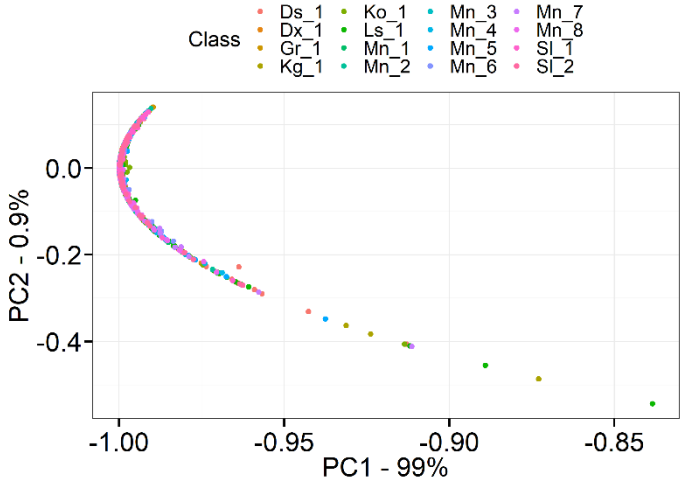

Figure 4.62: *PCA score plot for Modern Human Teeth, using the integrals of specific biogenic spectral lines.*

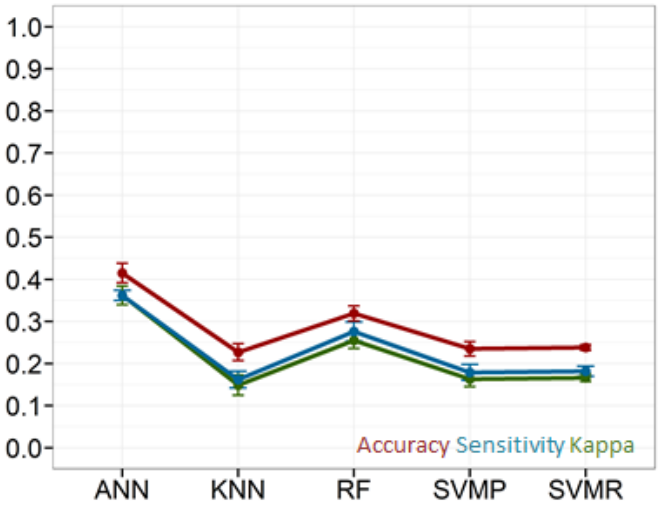

Figure 4.63: *Modern Human dataset, using the integrals of specific biogenic spectral lines. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The Artificial Neural Network has significantly better results than any of the other algorithms.*

Table 4.29: *Metric values for the Modern Human data, using the integrals of specific biogenic spectral lines.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.227	0.02	0.162	0.02	0.149	0.024
SVMR	0.238	0.007	0.182	0.012	0.166	0.009
SVMP	0.235	0.017	0.179	0.019	0.163	0.018
RF	0.319	0.018	0.276	0.023	0.255	0.019
ANN	0.415	0.023	0.362	0.012	0.362	0.022

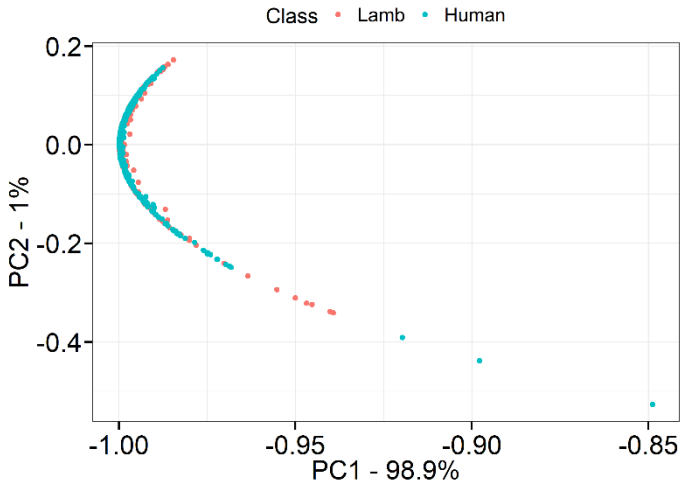

Figure 4.64: *PCA score plot for Domestic Sheep and Human Teeth, using the integrals of specific biogenic spectral lines.*

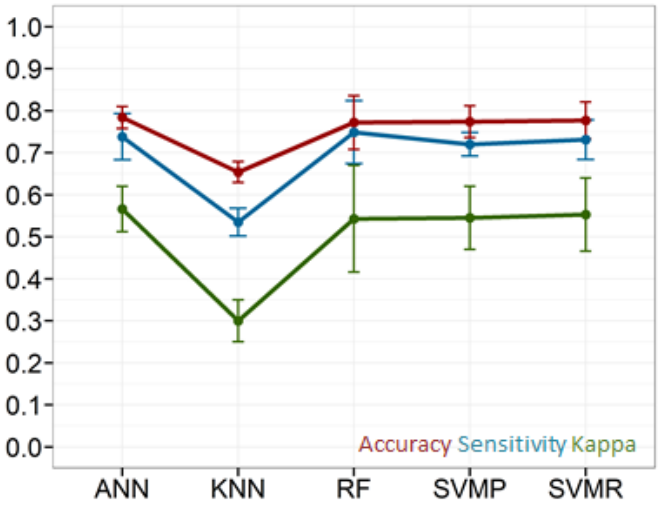

Figure 4.65: *Domestic Sheep and Human dataset, using the integrals of specific biogenic spectral lines. The values of the three different metrics that were used, for each of the algorithms, along with the standard deviation for each value, can be seen. The k Nearest Neighbors algorithm has lower metric values, while the rest of the algorithms give almost the same results for the three metric values.*

Table 4.30: *Metric values for the Domestic Sheep and Human data, using the integrals of specific biogenic spectral lines.*

	Accuracy	***SD***	***Sensitivity***	***SD***	***Kappa***	***SD***
KNN	0.654	0.025	0.535	0.033	0.3	0.05
SVMR	0.777	0.044	0.731	0.047	0.553	0.087
SVMP	0.774	0.038	0.72	0.028	0.545	0.075
RF	0.772	0.064	0.749	0.075	0.543	0.127
ANN	0.784	0.026	0.738	0.055	0.566	0.054

4.6 Results' Discussion

The different models that were used for the classification of hard tissue remains, under many circumstances and using several datasets, provided the opportunity for a comparison among them to identify the best fitted model for this particular classification task. Simultaneously, a comparison among the different feature sets that were used for each classification approach, resulted in the best feature selection procedure and the best set of features for the current work.

Among the three different metrics that were used for this work, Accuracy considered the most characteristic metric value for the evaluation of the results. For that reason and considering that the rest of the metric values consistently followed the same pattern, the comparison of the Accuracy values of the different classification approaches can provide a sufficient interpretation for the differences among the results. Thus, the Artificial Neural Network consistently appeared having better classification ability than the rest of the algorithms, scoring the highest results in the majority of the different approaches. Thus, ANN was considered the best possible choice for a decent classification result, among the used algorithms. The only approaches in which the ANN failed to reach the highest classification, for some of the datasets, were either in the feature selection by variance or when the Accuracy values were so low for all models that could not lead to a sufficient classification result. The ANN was systematically followed by the Random Forest algorithm that stood out as the best of the rest, showing great consistency but with lower results than those of the ANN.

Following the selection of ANN as the best classification algorithm for the current task, the comparison of the results for the different feature reduction approaches revealed that the Accuracy of the models can significantly vary depending on the features that are taken into consideration. The use of the newly built functions for feature selection and extraction resulted in models that failed to perform classification of unknown samples. Hence, the feature selection by the maximum intensity on specific spectral lines or by the variance, alongside the feature extraction by the calculation of the integrals for specific spectral lines, are unsuitable for feature reduction on the classification problems of the current work. On the other hand, the manual feature selection targeting specific spectral areas, with spectral lines that correspond in biogenic elements of hard tissues, works sufficiently well, as it was expected by a previous work in the field. (Siozos et al., 2021) The initial approach of 258 pixels, was suffered by lack of accuracy for the two modern datasets (Domestic Sheep dataset and Human dataset), during the classification on individuals, with the accuracy values of those two datasets lying slightly above 0.7. With a further significant feature reduction to 100 pixels of the initial 258, the classification results were increased but remained under the threshold of 0.8, for a sufficient classification. This threshold was reached and surpassed by the addition of 30 extra pixels, on the 100 previously mentioned pixels, with the accuracy values reaching 0.8 and 0.83, for the Modern Domestic Sheep and the Modern Human dataset, respectively. Additionally, the accuracy values of the ANN algorithm for the Archaeological and the Combined

Modern datasets, exceeding 0.85 and 0.95 respectively, remained almost the same despite the number of pixels that were used each time. Resulting to the best classification accuracy values for all datasets, the selection of the 130 pixels was considered better than the rest of the feature selection approaches and the one that could provide the most accurate and robust classification models (Figure 4.66, Figure 4.67, Figure 4.68, Figure 4.69).

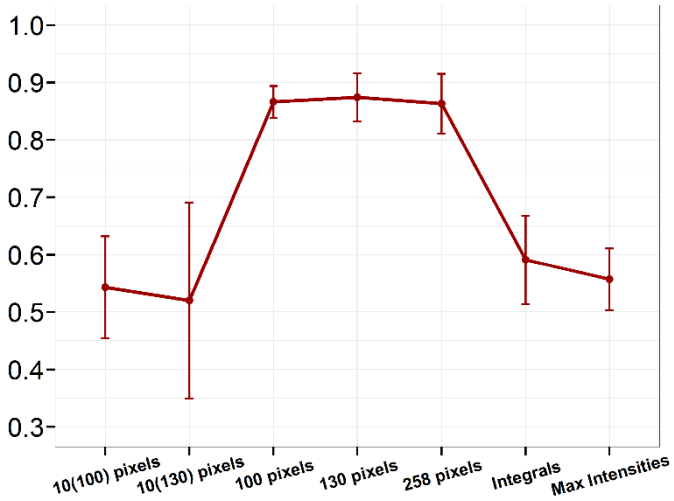

Figure 4.66: *Comparison of the Accuracy values among different used features using the Artificial Neural Network. These results refer to the discrimination of individuals using the Archaeological Human data. The Accuracy values of the manually selected features are significantly higher, with the one for the 130 pixels being the highest of all with a low standard deviation, too. On the contrary, the Accuracy value for the 100 pixels has the lowest standard deviation, achieving a very good classification behavior, as well.*

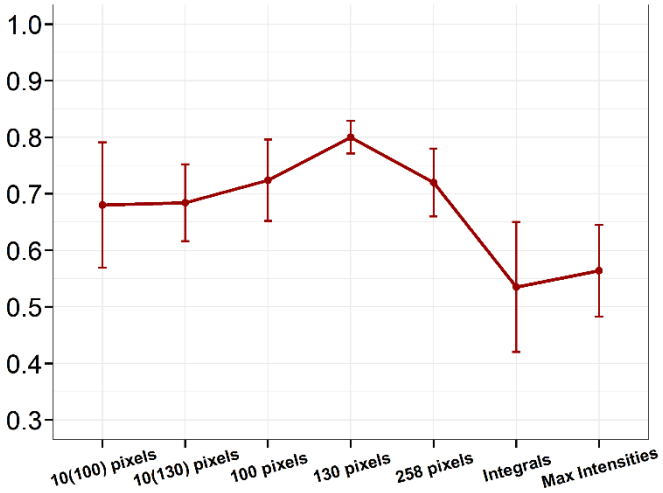

Figure 4.67: *Comparison of the Accuracy values among different used features using the Artificial Neural Network. These results refer to the discrimination of individuals using the Modern Domestic Sheep data. The Accuracy value of the 130 manually selected features is significantly higher than the rest, with the lowest value of standard deviation, too.*

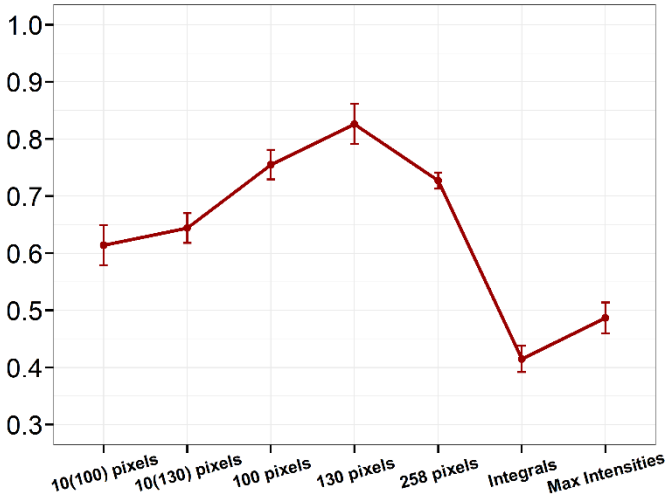

Figure 4.68: *Comparison of the Accuracy values among different used features using the Artificial Neural Network. These results refer to the discrimination of individuals using the Modern Human dataset. The Accuracy values of the manually selected features are significantly higher than the rest, with the 130 pixels reaching the highest value, by far. Using 258 pixels, the lowest standard deviation is achieved, but the Accuracy value is significantly lower than the one for the 130 pixels.*

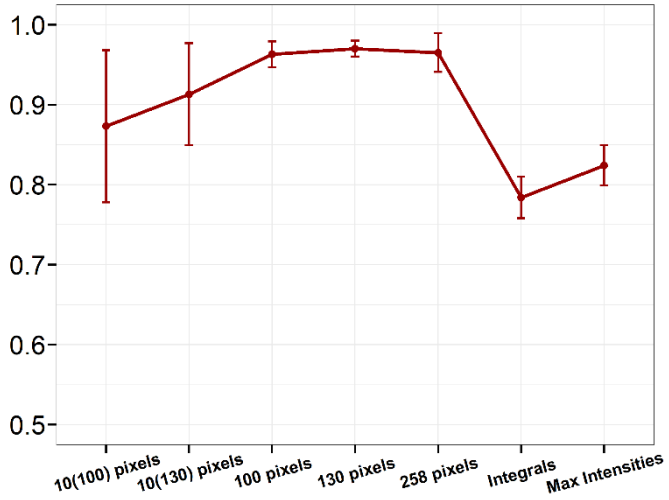

Figure 4.69: *Comparison of the Accuracy values among different used features using the Artificial Neural Network. These results refer to the discrimination between species using the combined (Domestic Sheep and Human) dataset. The Accuracy values of the manually selected features are significantly higher, with the one for the 130 pixels being the highest of all with the lowest standard deviation, too.*

Using the Artificial Neural Network and taking into consideration only the 130 pixels that were previously mentioned, the classification of LIBS spectra, coming from hard tissue remains of different individuals and species, was achieved. Depending on the dataset that was used, the classification could vary from sufficient to excellent. The classification of LIBS spectra on individuals for both Modern Domestic Sheep and Modern Human dataset was sufficient, exceeding the threshold of 0.8 on Accuracy that has been set from earlier works. Good classification on individuals was achieved in Archaeological data, even though the 30 extra pixels come from spectral areas with

many impurities and thus the results might not come strictly from biogenic elements. In addition, a very similar classification result was achieved without taking into consideration these extra spectral areas, for this particular dataset. Finally, the classification on LIBS spectra in different species, using the combined modern dataset, was excellent. The classification ability of the model for this specific task was great, reaching accuracy values exceeding 0.96, making it the best classification result that was achieved in the current work. Additionally, the feature selection processes resulted in a significant time reduction, with the running time varying from $30\ min$ (Modern Domestic Sheep dataset) to $60\ min$ (Modern Human dataset), depending on the dataset and the amount of data that the model had to proceed.

4.7 Conclusion

Archaeological and forensic excavations frequently reveal large numbers of hard tissue remains of animal or human origins. The discrimination between individuals and the classification of the remains in the individuals that they belong, especially in cases that the remains are severally fragmented and disarranged, can be very useful for the proper study of the artifacts, revealing information about the identity, life and death of each individual. The difficulty of the discrimination and classification task can vary depending on the condition of the biological remains and the excavation site. Thus, several methods are used to help archaeologist and forensic scientists to achieve their goal. Laser Induced Breakdown Spectroscopy is a micro-destructive, fast and transferable technique that can provide useful information about the elemental composition of the samples, with no prior treatment and minimal damage of them. LIBS can provide massive amounts of data in a very short time and thus is a suitable technique to be combined with machine learning methods. LIBS combined with machine learning and artificial neural networks can be a useful tool for archaeologist or forensic scientist, speeding up the process of discrimination/classification.

In the current work, LIBS data collected from four different sample sets were analyzed using several, widely used, machine learning algorithms implemented in the open source Programming Language R. Three sample sets consisted of modern human or/and animal teeth and one sample set consisted of archaeological bone fragments, were measured with the use of a customized LIBS microspectrometer. K Nearest Neighbors, Support Vector Machine using a polynomial and a radial kernel, Random Forest and a relatively simple Artificial Neural Network were used for the analysis of the collected data. Comparing the used algorithms, the results of ANN were significantly higher than the rest of the algorithms, in all the tested situations. RF followed with significantly lower results while the remaining algorithms were considered unable to provide a sufficient classification for the majority of classification tasks. Thus, the ANN was considered the most suitable algorithm to achieve the desirable classification. Because of the huge number of features resulting from the structure of the LIBS spectrum, several feature selection processes were tested for minimizing the computational time and maximizing the results. Seven different approaches were examined, with the one using 130 manually selected pixels provided the best results for the classification task. In particular, using the ANN algorithm and

the 130 selected features, an excellent classification accuracy on species was achieved for the dataset of Modern Human combined with Modern Domestic Sheep teeth, exceeding 95%, while the classification accuracies on individuals were reached and exceeded the threshold of 80% for the Modern Domestic Sheep teeth and the Modern Human teeth datasets, respectively, resulting in a decent classification for those two datasets. Furthermore, a very good classification was achieved for the Archaeological dataset, with the classification accuracy exceeding 85%, using the ANN algorithm and the 130 selected features. Hence, for the current work a robust and accurate classification model was developed, based on the comparison of several classification algorithms and many feature selection approaches. The model was tested in several situations for four different datasets and for two different tasks, successfully achieving the classification of LIBS spectra in the correct individual or species, with a classification accuracy varying from decently good to excellent.

Those promising results, combined with other familiar studies, can be used as starting point for further research on the development of a LIBS based combined with machine learning methods for the classification of hard tissue remains. The use of more advanced computer hardware will allow to test more parameters during the parameter selection and further increase the accuracy of the models. Increasing the number of testing samples, creating huge data sets, is another key point that can significantly improve the ability of the models to correctly discriminate individuals and classify samples. The fact that LIBS provides the opportunity for fast and accurate analysis resulting in a significant amount of data in a short period of time, makes it one of the most suitable techniques for creating huge datasets. Training the models with more and more data can potentially result in a major increase in their classification accuracy, especially for the ANN, providing the opportunity for potential classification of entire samples in the correct individual.

References

Aggarwal, C. (2018). *Neural Networks and Deep Learning*. Cham: Springer.

Agostini, P., Barjot, G., Bonnal, J., Mainfray, G., Manus, C., & Morellec, J. (1968). Multiphoton ionization of hydrogen and rare gases. *IEEE Journal Of Quantum Electronics*, 4(10), 667-669. doi: 10.1109/jqe.1968.1074955.

Alpaydin, E. (2010). *Introduction to machine learning*. Cambridge, Mass.: MIT Press.

Andrews, D.F., Herzberg, A.M. (1985) Iris Data. In: Data. Springer Series in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-5098-2_2.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, 25(2), 197-227. doi: 10.1007/s11749-016-0481-7.

Bishop, C. (1994). Neural networks and their applications. *Review Of Scientific Instruments*, 65(6), 1803-1832. doi: 10.1063/1.1144830.

Bishop, C. (2006). *Pattern recognition and machine learning*. New York: Springer.

Bloembergen, N. (1997). A Brief History of Light Breakdown. *Journal Of Nonlinear Optical Physics & Materials*, 6(4), 377-385. doi: 10.1142/s0218863597000289.

Boser, B., Guyon, I., & Vapnik, V. (1992). A training algorithm for optimal margin classifiers. *Proceedings Of The Fifth Annual Workshop On Computational Learning Theory - COLT '92*. https://doi.org/10.1145/130385.130401.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140. doi: 10.1007/bf00058655.

Breiman, L. (1996). Technical Note: Some Properties of Splitting Criteria. *Machine Learning*, 24(1), 41-47. doi: 10.1023/a:1018094028462.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. doi: 10.1023/a:1010933404324.

Castro, W., Hoogewerff, J., Latkoczy, C., & Almirall, J. (2010). Application of laser ablation (LA-ICP-SF-MS) for the elemental analysis of bone and teeth samples for discrimination purposes. *Forensic Science International*, 195(1-3), 17-27. doi: 10.1016/j.forsciint.2009.10.029.

Chaudhary, K., Rizvi, S. Z. H., & Ali, J. (2016). Laser-Induced Plasma and its Applications, in Plasma Science and Technology - Progress in Physical States and Chemical Reactions, Tetsu Mieno, IntechOpen. doi: 10.5772/61784.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational And Psychological Measurement*, 20(1), 37-46. doi: 10.1177/001316446002000104

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. https://doi.org/10.1007/bf00994018.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions On Information Theory, 13(1), 21-27. doi: 10.1109/tit.1967.1053964.

Cremers, D., & Radziemski, L. (2013). *Handbook of Laser-Induced Breakdown Spectroscopy* (2nd ed.). Chichester, West Sussex: John Wiley & Sons Ltd.

Cremers, D., Multari, R., & Knight, A. (2016). Laser-induced Breakdown Spectroscopy. *Encyclopedia Of Analytical Chemistry*. doi:10.1002/9780470027318.a5110t.pub3.

Cremers, D., Radziemski, L., & Loree, T. (1984). Spectrochemical Analysis of Liquids Using the Laser Spark. *Applied Spectroscopy*, 38(5), 721-729. doi: 10.1366/0003702844555034.

Cristianini, N., & Ricci, E. (2008). Support Vector Machines. *Encyclopedia Of Algorithms*, 928-932. https://doi.org/10.1007/978-0-387-30162-4_415.

Crossland, Z. (2000). Buried lives. *Archaeological Dialogues*, 7(2), 146-159. doi: 10.1017/s1380203800001707.

Currey J. (2008) *Collagen and the Mechanical Properties of Bone and Calcified Cartilage*. In: Fratzl P. (eds) *Collagen*. Boston: Springer. doi: 10.1007/978-0-387-73906-9_14.

Cutler, A., Cutler, D.R., Stevens, J.R. (2012) *Random Forests, Ensemble Machine Learning*. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-9326-7_5.

Daniel, G. E. (2019). *Archaeology. Encyclopedia Britannica*. https://www.britannica.com/science/archaeology.

Darvill, T. (2019). *Archaeology in the PPG16 era*. Oxford: OXBOW BOOKS.

Díaz-Uriarte, R., & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7(1), 3. doi: 10.1186/1471-2105-7-3

Fix, E., & Hodges, J. (1989). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *International Statistical Review*, 57(3), 238. doi: 10.2307/1403797.

Geltman, S. (1973). Free-free radiation in electron-neutral atom collisions. *Journal Of Quantitative Spectroscopy And Radiative Transfer*, 13(7), 601-613. doi: 10.1016/0022-4073(73)90019-8.

Gevin, G. (1997). An introduction to neural networks. London: UCL Press.

Giannelli, P. (2006). Forensic Science. *Journal Of Law, Medicine & Ethics*, 34(2), 310-319. doi: 10.1111/j.1748-720x.2006.00036.x

Gigosos, M. (2014). Stark broadening models for plasma diagnostics. *Journal Of Physics D: Applied Physics*, 47(34), 343001. doi: 10.1088/0022-3727/47/34/343001.

Gornushkin, I., King, L., Smith, B., Omenetto, N., & Winefordner, J. (1999). Line broadening mechanisms in the low pressure laser-induced plasma. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 54(8), 1207-1217. doi: 10.1016/s0584-8547(99)00064-6.

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal Of Machine Learning Research*, 3, 1157-1182.

Haglund, W., Connor, M., & Scott, D. (2001). The archaeology of contemporary mass graves. *Historical Archaeology*, 35(1), 57-69. doi: 10.1007/bf03374527.

Harilal, S., Bindhu, C., Issac, R., Nampoori, V., & Vallabhan, C. (1997). Electron density and temperature measurements in a laser produced carbon plasma. *Journal Of Applied Physics*, 82(5), 2140-2146. doi: 10.1063/1.366276.

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence. *Neuron*, 95(2), 245-258. doi: 10.1016/j.neuron.2017.06.011.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). New York: Springer-Verlag.

Hausmann, N., Prendergast, A., Lemonis, A., Zech, J., Roberts, P., Siozos, P., & Anglos, D. (2019). Extensive elemental mapping unlocks Mg/Ca ratios as climate proxy in seasonal records of Mediterranean limpets. *Scientific Reports*, 9(1). doi: 10.1038/s41598-019-39959-9.

Hausmann, N., Siozos, P., Lemonis, A., Colonese, A., Robson, H., & Anglos, D. (2017). Elemental mapping of Mg/Ca intensity ratios in marine mollusc shells using laser-induced breakdown spectroscopy. *Journal of Analytical Atomic Spectrometry*, 32(8), 1467-1472. doi: 10.1039/c7ja00131b.

Hecht, J. (2008). *Understanding lasers*. Piscataway, NJ: IEEE Press.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal Of Educational Psychology*, 24(6), 417-441. doi: 10.1037/h0071325.

Hunter, J. (2009). *Archaeology. Wiley Encyclopedia Of Forensic Science*. In: Jamieson, A., & Moenssens, A. (2009). *Wiley encyclopedia of forensic science*. Chichester, West Sussex, U.K.: John Wiley & Sons. doi: 10.1002/9780470061589.fsa059

Hunter, J., Brickley, M., Bourgeois, J., Bouts, W., Bourguignon, L., & Hubrecht, F. et al. (2001). Forensic archaeology, forensic anthropology and Human Rights in Europe. *Science & Justice*, 41(3), 173-178. doi: 10.1016/s1355-0306(01)71886-2.

Hunter, J., Heron, C., Janaway, R., Martin, A., Pollard, A., & Roberts, C. (1994). Forensic archaeology in Britain. *Antiquity*, 68(261), 758-769. doi: 10.1017/s0003598x00047451.

Jolliffe, I. (2002). *Principal component analysis*. New York: Springer.

Jolliffe, I., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions Of The Royal Society A: Mathematical, Physical And Engineering Sciences*, 374(2065), 20150202. doi: 10.1098/rsta.2015.0202.

Jurafsky, D., & Martin, J. (2009). Speech and language processing. Upper Saddle River, N.J.: Pearson Prentice Hall.

Kasem, M., Gonzalez, J., Russo, R., & Harith, M. (2014). Effect of the wavelength on laser induced breakdown spectrometric analysis of archaeological bone. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 101, 26-31. doi: 10.1016/j.sab.2014.07.010.

Ketkar N. (2017) Stochastic Gradient Descent. In: Deep Learning with Python. Berkeley, Apress. https://doi.org/10.1007/978-1-4842-2766-4_8.

Klepinger, L. (1984). Nutritional Assessment From Bone. *Annual Review Of Anthropology*, 13(1), 75-96. doi: 10.1146/annurev.an.13.100184.000451.

Koropoulis, A., Alachiotis, N., & Pavlidis, P. (2020). *Detecting Positive Selection in Populations Using Genetic Data, Statistical Population Genomics*. Humana Press, New York, U.S.A. https://doi.org/10.1007/978-1-0716-0199-0.

Krizhevsky, A., Sutskever, I., & Hinton, G. (2017). ImageNet classification with deep convolutional neural networks. *Communications Of The ACM*, 60(6), 84-90. doi: 10.1145/3065386.

Kumar, R., & Sharma, V. (2018). Chemometrics in forensic science. Trac Trends In Analytical Chemistry, 105, 191-201. doi: 10.1016/j.trac.2018.05.010.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. doi: 10.1038/nature14539.

Lee, K., Harris, C., Gunn, S., & Reed, P. (2001). Control Sensitivity SVM for Imbalanced Data A Case Study on Automotive Material. *Artificial Neural Nets And Genetic Algorithms*, 189-192. doi: 10.1007/978-3-7091-6230-9_46

Lever, J., Krzywinski, M., & Altman, N. (2016). Model selection and overfitting. *Nature Methods*, 13(9), 703-704. doi: 10.1038/nmeth.3968.

Liaw, A., & Wiener, M. (2002). Classification and regression by Random Forest. *R news*, 2(3), 18-22.

Lindley, D. (1977). A Problem in Forensic Science. *Biometrika*, 64(2), 207-213. doi: 10.2307/2335686.

Lynnerup, N., & Klaus, H.D. (2019). *Fundamentals of Human Bone and Dental Biology: Structure, Function, and Development*, In: Buikstra, J., *Ortner's Identification of Pathological Conditions in Human Skeletal Remains* (3rd ed.). Tempe: Arizona State University. doi: 10.1016/B978-0-12-809738-0.00004-1.

Ma, G., & Liu, X. (2009). Hydroxyapatite: Hexagonal or Monoclinic?. *Crystal Growth & Design*, 9(7), 2991-2994. doi: 10.1021/cg900156w.

Ma, J., Sheridan, R., Liaw, A., Dahl, G., & Svetnik, V. (2015). Deep Neural Nets as a Method for Quantitative Structure–Activity Relationships. *Journal Of Chemical Information And Modeling*, 55(2), 263-274. doi: 10.1021/ci500747n.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin Of Mathematical Biophysics*, 5(4), 115-133. doi: 10.1007/bf02478259.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2015). Misc Functions of the Department of Statistics, ProbabilityTheory Group (Formerly: E1071), TU Wien.

Miller, J., & Miller, J. (2014). *Stat Chemometrics Analytical Chem*. Pearson Australia Pty Ltd.

Mitchell, T. (1997). *Machine learning*. New York: MacGraw-Hill.

Moncayo, S., Manzoor, S., Ugidos, T., Navarro-Villoslada, F., & Caceres, J. (2014). Discrimination of human bodies from bones and teeth remains by Laser Induced Breakdown Spectroscopy and Neural Networks. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 101, 21-25. doi: 10.1016/j.sab.2014.07.008.

Mulak, P., Talhar, N. (2015). Analysis of Distance Measures Using K-Nearest Neighbor Algorithm on KDD Dataset. *International Journal of Science and Research*, 4(7), 2101-2104.

Müller, K., Smola, A., Rätsch, G., Schölkopf, B., Kohlmorgen, J., & Vapnik, V. (1997). Predicting time series with support vector machines. *Lecture Notes In Computer Science*, 999-1004. https://doi.org/10.1007/bfb0020283.

Ooi, H., Ng, S., & Lim, E. (2013). ANO Detection with K-Nearest Neighbor Using Minkowski Distance. *International Journal Of Signal Processing Systems*, 208-211. doi: 10.12720/ijsps.1.2.208-211.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, Series 6, 2(11), 559-572.

Pietrusewsky, M. (2007). *Metric Analysis of Skeletal Remains: Methods and Applications*. In: Katzenberg, M., & Saunders, S. (2007). *Biological anthropology of the human skeleton*. John Wiley & Sons. doi: 10.1002/9780470245842.ch16.

Pořízka, P., Klus, J., Hrdlička, A., Vrábel, J., Škarková, P., & Prochazka, D. et al. (2017). Impact of Laser-Induced Breakdown Spectroscopy data normalization on multivariate classification accuracy. *Journal Of Analytical Atomic Spectrometry*, 32(2), 277-288. doi: 10.1039/c6ja00322b

Random forests - classification description. Retrieved 11 January 2021, from https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

Rehse, S., Salimnia, H., & Miziolek, A. (2012). Laser-induced breakdown spectroscopy (LIBS): an overview of recent progress and future potential for biomedical applications. *Journal Of Medical Engineering & Technology*, 36(2), 77-89. doi: 10.3109/03091902.2011.645946.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533-536. doi: 10.1038/323533a0.

Samek, O., Beddows, D., Telle, H., Kaiser, J., Liška, M., Cáceres, J., & Gonzáles Ureña, A. (2001). Quantitative laser-induced breakdown spectroscopy analysis of calcified tissue samples. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 56(6), 865-875. doi: 10.1016/s0584-8547(01)00198-7.

Sarkar, M., & Leong, T. Y. (2000). Application of K-nearest neighbors algorithm on breast cancer diagnosis problem. *Proceedings. AMIA Symposium*, 759–763.

Siozos, P., Hausmann, N., Holst, M., & Anglos, D. (2021). Application of laser-induced breakdown spectroscopy and neural networks on archaeological human bones for the discrimination of distinct individuals. *Journal Of Archaeological Science: Reports*, 35, 102769. doi: 10.1016/j.jasrep.2020.102769.

Siozos, P., Philippidis, A., & Anglos, D. (2017). Portable laser-induced breakdown spectroscopy/diffuse reflectance hybrid spectrometer for analysis of inorganic pigments. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 137, 93-100. doi: 10.1016/j.sab.2017.09.005.

Stafe, M., Marcu, A., & Puscas, N. (2014). *Pulsed laser ablation of solids*. Berlin: Springer.

Takahashi, T., & Thornton, B. (2017). Quantitative methods for compensation of matrix effects and self-absorption in Laser Induced Breakdown Spectroscopy signals of solids. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 138, 31-42. doi: 10.1016/j.sab.2017.09.010.

Thyagarajan, K., & Ghatak, A. (2011). *Lasers (2nd ed.)*. Springer.

Tsamardinos, I., Rakhshani, A., & Lagani, V. (2014). Performance-Estimation Properties of Cross-Validation-Based Protocols with Simultaneous Hyper-Parameter Optimization. *Artificial Intelligence: Methods And Applications*, 1-14. doi: 10.1007/978-3-319-07064-3_1.

Türp, J. C., Alt, K. W. (1998) *Anatomy and Morphology of Human Teeth*. In: Alt K.W., Rösing F.W., Teschler-Nicola M. (eds) *Dental Anthropology.* Vienna: Springer. doi: 10.1007/978-3-7091-7496-8_6.

Vapnik, V. (1998). The Support Vector Method of Function Estimation. *Nonlinear Modeling*, 55-85. https://doi.org/10.1007/978-1-4615-5703-6_3.

Vapnik, V., Guyon, I., & Hastie, T. (1995). Support vector machines. *Machine Learning*, 20(3), 273-297.

Venables, W., & Ripley, B. (2010). *Modern applied statistics with S*. New York: Springer.

Veropoulos, K., Campbell, C., & Cristianini, N. (1999). Controlling the Sensitivity of Support Vector Machines. *Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden (IJCAI99)*, 55-60.

Villagran, X. S., Huisman, D. J., Mentzer, S. M., Miller, C. E., & Jans, M. M. (2009). *Bone and Other Skeletal Tissues*. In: Nicosia, C., & Stoops, G. (2017). *Archaeological soil and sediment micromorphology*. John Wiley & Sons Ltd. doi: 10.1002/9781118941065.ch1.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal Of Statistical Software*, 40(1). doi: 10.18637/jss.v040.i01

Wickham, H. (2016). *Ggplot2*. New York: Springer-Verlag.

Xia, S., Xiong, Z., Luo, Y., Dong, L., & Zhang, G. (2015). Location difference of multiple distances based k-nearest neighbors algorithm. *Knowledge-Based Systems*, 90, 99-110. doi: 10.1016/j.knosys.2015.09.028.

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal Of Physics: Conference Series*, 1168, 022022. doi: 10.1088/1742-6596/1168/2/022022.