

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ**

**ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ**

**Φαφαλιός Στέφανος  
Μεταπτυχιακός Φοιτητής**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης  
Επόπτης Μεταπτ. Εργασίας: Καθηγητής, Ι. Τσαμαρδίνος**

**Παρασκευή , 13/12/2019, 16:00  
Αίθουσα K206, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

**“ Προς την ανάπτυξη ενός καθολικού μοριακού ταξινομητή”**

### **ΠΕΡΙΛΗΨΗ**

Παρά την πληθώρα των βιολογικών δεδομένων, είναι δύσκολο να συγκεντρωθεί μια ικανοποιητική ποσότητα αρκετά ομοιογενών δεδομένων αυτού του είδους, ώστε να προκύψουν ασφαλή συμπεράσματα που θα οδηγούσαν στη διάγνωση κάποιας νόσου. Επιπλέον, στο πρόβλημα διάγνωσης και θεραπείας ασθενειών, επιφέρει πρόσθετη πολυπλοκότητα η δυσκολία προσδιορισμού και ο εντοπισμός των πολύπλοκων χαρακτηριστικών τους, που εκτείνεται σε δεκάδες χιλιάδες. Διαμορφώνουμε μαθηματικά αυτά τα χαρακτηριστικά, χρησιμοποιώντας δημόσια διαθέσιμα δεδομένα από μελέτες ασθενειών, μέσω αναπαραστάσεων χαμηλότερων διαστάσεων. Στη συνέχεια, χρησιμοποιώντας εκλεπτυσμένες στατιστικές μεθόδους, κατάλληλες για δεδομένα πολλών διαστάσεων, είμαστε σε θέση να παράσχουμε μια λίστα πιο πιθανών ασθενειών για έναν συγκεκριμένο ασθενή, αλλά επίσης να συμπεριλάβουμε περιπτώσεις μη πρόβλεψης (περιπτώσεις στις οποίες η διάγνωση ασθένειας είναι αβέβαιη). Το όφελος αυτού του διαγνωστικού εργαλείου συνίσταται στη βοήθεια που μπορεί να παρέχει στους ιατρούς, υποδεικνύοντας τις πιθανότερες ασθένειες ενός ασθενή ώστε να επιτρέψει έγκαιρη-αξιόπιστη

διάγνωση και συνεπώς θεραπεία. Η ανάλυση είναι χρονικά αποδοτική. Μπορεί να πραγματοποιηθεί σε πραγματικό χρόνο, με ελάχιστη υπολογιστική ισχύ και χαμηλές απαιτήσεις μνήμης.

**Stefanos Fafalios**

**M.Sc. Thesis**

**Computer Science Department**

**University of Crete**

**Master's Thesis Supervisor: Professor, I. Tsamardinos**

**Friday, 13/12/2019, 16:00**

**Room K206, Computer Science Dept., University of Crete**

**“Towards a universal molecular classifier”**

### **ABSTRACT**

Despite the plethora of biological datasets, it is difficult to gather a satisfactory amount of as homogeneous as possible biological data and derive safe conclusions regarding disease diagnosis. On top of that, a challenging problem of disease diagnosis and their treatment is to identify their complex characteristics that span to tens of thousands. We mathematically shape these characteristics using publicly available datasets, from disease studies, via a low dimensional space representation. In essence, using sophisticated statistical methods, suitable for very high dimensional data, we are able to provide a list of the most probable diseases for a given patient, but also include don't-know-predictions (cases with uncertain disease diagnosis). The benefit of this diagnostic framework is that it can assist medical doctors by directing them towards the most probable diseases for a given patient and allows for timely disease diagnosis and treatment. The analysis is time-efficient and can be conducted in real time, with very little computational power and low ram and disk memory requirements.