

ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ
ΤΜΗΜΑ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
Π.Μ.Σ. ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ Η/Υ
(COMPUTER SCIENCE AND TECHNOLOGY)

ΑΝΙΧΝΕΥΣΗ ΣΦΑΛΜΑΤΩΝ ΜΕ ΧΡΗΣΗ ΑΛΓΟΡΙΘΜΟΥ QGA-SVM

Αχιλλέας Σούλης



Θεσσαλονίκη 15/6/2021

Εισαγωγή

Η χρήση των μηχανημάτων που στηρίζονται στην περιστρεφόμενη μηχανολογία (**rotating machinery**) έχει τύχει ευρύτατης εφαρμογής.

Η έγκαιρη **διάγνωση των πιθανών σφαλμάτων** που θα προκύψουν κατά την λειτουργία τους είναι μεγάλης σημασίας.

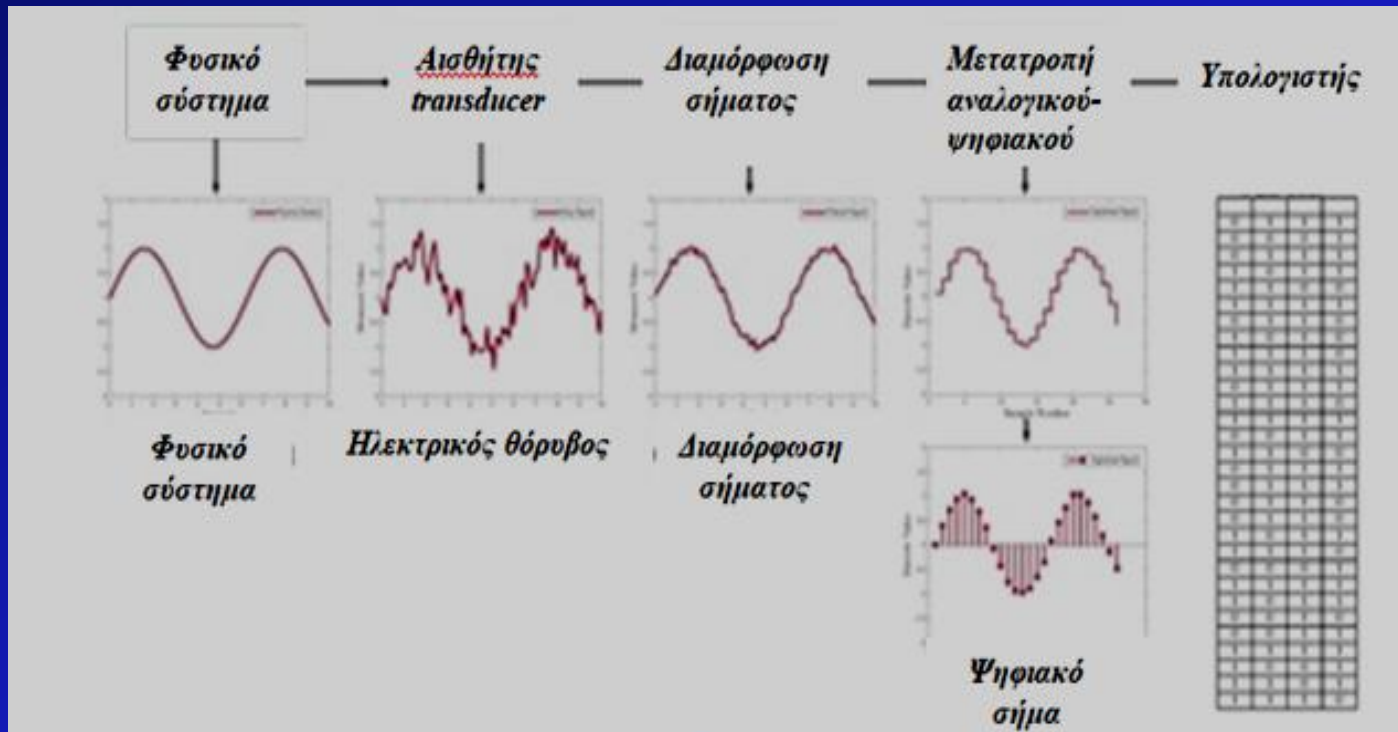
Στην παρούσα εργασία προτείνεται η μέθοδος διάγνωσης σφαλμάτων **QGA-SVM** που βασίζεται στην τεχνική της υποστηρικτικής μηχανής μέσω διανυσμάτων (**Supporting Vector Machine, SVM**) η οποία βελτιστοποιείται με τη χρήση κβαντικού γενετικού αλγορίθμου (**Quantum Genetic Algorithm, QGA**).

Τα αποτελέσματα της ανάλυσης θα δείξουν,

- α) εάν η προτεινόμενη μέθοδος έχει ορισμένες **φορές μεγαλύτερη ακρίβεια** στη διάγνωση σφαλμάτων από την παραδοσιακή μέθοδο SVM, αλλά και
- β) εάν η εύρεση του βέλτιστου αποτελέσματος επιτυγχάνεται **πιο γρήγορα και πιο εύκολα**.

Σύστημα συλλογής ψηφιακών δεδομένων

- τον αισθητήρα
- τον μετατροπέα (transducer)
- τον επεξεργαστή σήματος
- την παρουσίαση των μετρήσεων



Μηχανική εκμάθηση

Μηχανική μάθηση είναι πεδίο της επιστήμης των υπολογιστών. Αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής μάθησης στην τεχνητή νοημοσύνη.

Τύποι μηχανικής μάθησης:

- εποπτευόμενη μάθηση
- μη-εποπτευόμενη μάθηση
- Στο πεδίο της ανάλυσης των δεδομένων, η μηχανική μάθηση είναι μια μέθοδος που χρησιμοποιείται για την επινόηση πολύπλοκων μοντέλων και αλγορίθμων που οδηγούν στην κατανόηση και την πρόβλεψη προβλημάτων.
- Η μηχανική μάθηση είναι στενά συνδεδεμένη και συχνά συγχέεται με την υπολογιστική στατιστική.
- Η μηχανική μάθηση μερικές φορές συγχέεται με την εξόρυξη δεδομένων, όπου η τελευταία επικεντρώνεται περισσότερο στην εξερευνητική ανάλυση των δεδομένων, γνωστή και ως μη-εποπτευόμενη μάθηση

Χαρακτηριστικά μηχανικής εκμάθησης

Τα αναλυτικά μοντέλα επιτρέπουν στους ερευνητές να παίρνουν,

- αξιόπιστες αποφάσεις,
- ακριβή αποτελέσματα
- συσχετίσεις μεταξύ των δεδομένων

Η μηχανική μάθηση ερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από υπάρχοντα δεδομένα και να προβλέπουν σχετικά με αυτά όπως είναι,

- Decision Trees
- Naïve Bayes
- Multilayer Perceptron
- Support Vector Machines

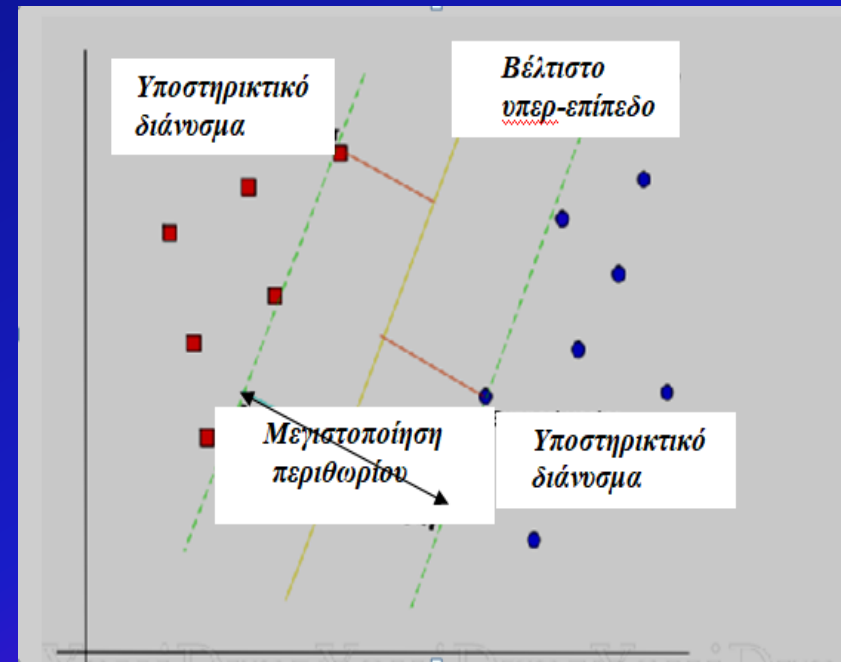
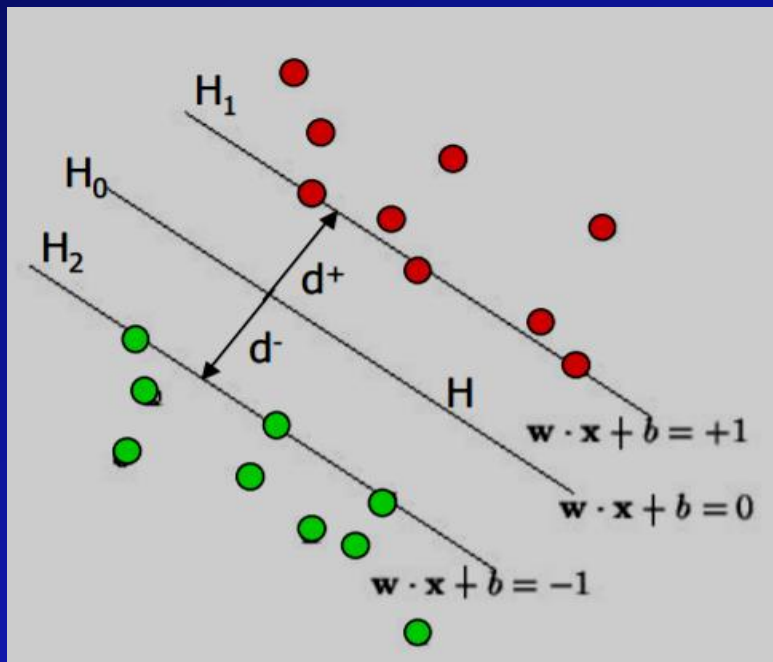
Υποστηρικτική μηχανή μέσω διανυσμάτων

Supporting Vector Machine, SVM

Τα διανύσματα υποστήριξης είναι τα σημεία δεδομένων που βρίσκονται πιο κοντά στην επιφάνεια διαχωρισμού ή υπερ-επίπεδο.

Η τεχνική SVM μεγιστοποιεί το περιθώριο μεταξύ των υποστηρικτικών διανυσμάτων

Η τεχνική SVM βρίσκει τη βέλτιστη λύση. Είναι ένα πρόβλημα βελτιστοποίησης και μπορεί να λυθεί με ανάλογες τεχνικές, όπως η τεχνική Lagrange.



Μεγιστοποίηση συνάρτησης, Dual problem

Primal problem:

$$\min L_P = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l a_i y_i (\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^l a_i$$

$$\text{s.t. } \forall i \ a_i \geq 0$$

$$\mathbf{w} = \sum_{i=1}^l a_i y_i \mathbf{x}_i, \quad \sum_{i=1}^l a_i y_i = 0$$

Dual problem:

$$\max L_D(a_i) = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i=1}^l a_i a_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{s.t. } \sum_{i=1}^l a_i y_i = 0 \ \& \ a_i \geq 0 \quad 0 \leq a_i \leq C$$

Κβαντική Μηχανική

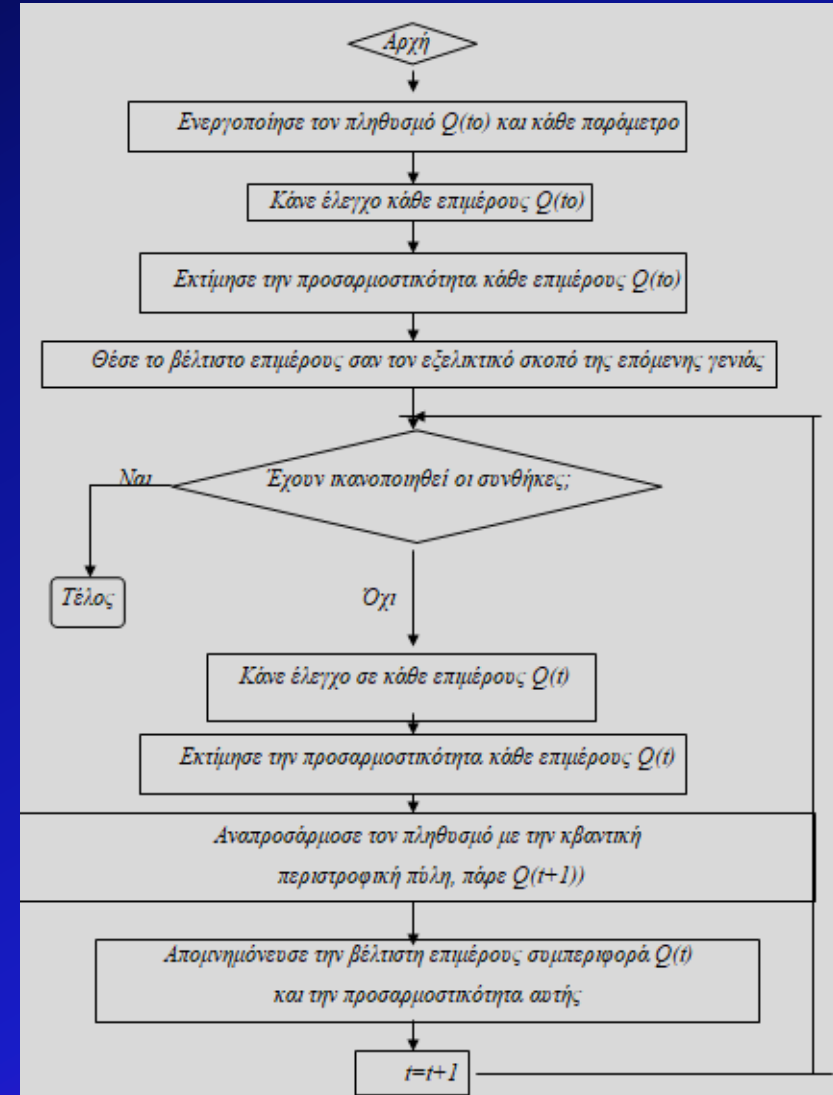
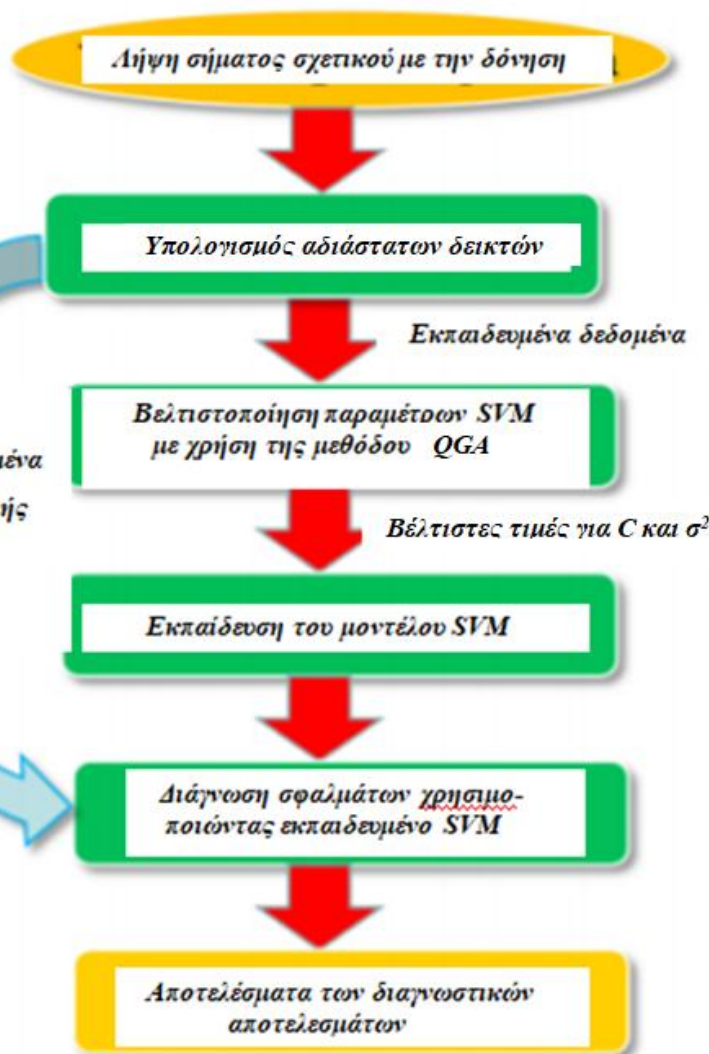
Χαρακτηριστικά Κβαντικής Μηχανικής

- **Υπέρθωση:** Ένα κλασικό bit μπορεί να έχει δύο διαφορετικές τιμές **0** ή **1**. Έχει πάντα διακεκριμένη τιμή ενώ το **qubit** μπορεί να δεχτεί απεριόριστα διαφορετικές τιμές.

Συγκεκριμένα $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$, με $\alpha^2 + \beta^2 = 1$

- **Διεμπλοκή:** είναι το φαινόμενο κατά το οποίο δύο σωματίδια ή ομάδες σωματιδίων που δημιουργούνται μαζί ή αλληλεπιδρούν συνενώνοντας τις κυματοσυναρτήσεις τους, μένουν σε κατάσταση διεμπλοκής μεταξύ τους, ασχέτως του χώρου που μεσολαβεί πλέον από το ένα στο άλλο.
- **Κβαντικές πύλες:** είναι ένα βασικό κβαντικό κύκλωμα που λειτουργεί σε έναν μικρό αριθμό qubits είναι δε τα δομικά στοιχεία των κβαντικών κυκλωμάτων, όπως οι κλασικές πύλες λογικής για συμβατικά ψηφιακά κυκλώματα.

Η μέθοδος διάγνωσης σφαλμάτων QGA-SVM



Κβαντική στροφή

x_i	$best_i$	$f(x) > f(best)$	$\Delta\theta_i$	$s(\alpha_i, \beta_i)$			
				$\alpha_i\beta_i > 0$	$\alpha_i\beta_i < 0$	$\alpha_i = 0$	$\beta_i = 0$
0	0	FALSE	0	0	0	0	0
0	0	TRUE	0	0	0	0	0
0	1	FALSE	$\Delta\theta_i$	+1	-1	0	± 1
0	1	TRUE	$\Delta\theta_i$	-1	+1	± 1	0
1	0	FALSE	$\Delta\theta_i$	-1	+1	± 1	0
1	0	TRUE	$\Delta\theta_i$	+1	-1	0	± 1
1	1	FALSE	0	0	0	0	0
1	1	TRUE	0	0	0	0	0

$$\begin{bmatrix} \alpha'_i \\ \beta'_i \end{bmatrix} = U(\theta_i) \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{pmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{pmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}$$

x_i αντιστοιχεί στον δυαδικό κώδικα του μεμονωμένου γονιδίου i χρωμοσώματος.

$best_i$ δυαδικός κώδικας που αντιστοιχεί στο καλύτερο άτομο είναι b_i .

$f(.)$ δηλώνει τη συνάρτηση φυσικής κατάστασης για να βρει την τιμή φυσικής κατάστασης.

Αν $\alpha_i\beta_i > 0$ υποδηλώνει ότι βρίσκεται στο 1^ο ή στο 3^ο τεταρτημόριο.

Αν $\alpha_i\beta_i < 0$ τότε το άτομο βρίσκεται στο δεύτερο ή το τέταρτο τεταρτημόριο.

Αν ο συνδυασμός είναι 0-1, τότε θέσε τη γωνία περιστροφή $\Delta\theta$.

Αν ο συνδυασμός είναι ίδιος π.χ. 0-0 and 1-1, επέλεξε γωνία περιστροφής μικρότερη από 0.2π .

Πειραματική διαδικασία και λήψη δεδομένων

Bearing data center drive end fault

Το πείραμα πραγματοποιήθηκε χρησιμοποιώντας έναν κινητήρα Reliance Electric 2.0 hp, και ελήφθησαν τα δεδομένα επιτάχυνσης σε τοποθεσίες που βρίσκονταν κοντά και μακριά από τα ρουλεμάν του κινητήρα. **Οι σημειακές αλλοιώσεις εισήχθησαν στα ρουλεμάν** χρησιμοποιώντας μια μηχανική ηλεκτρο-φόρτιση.

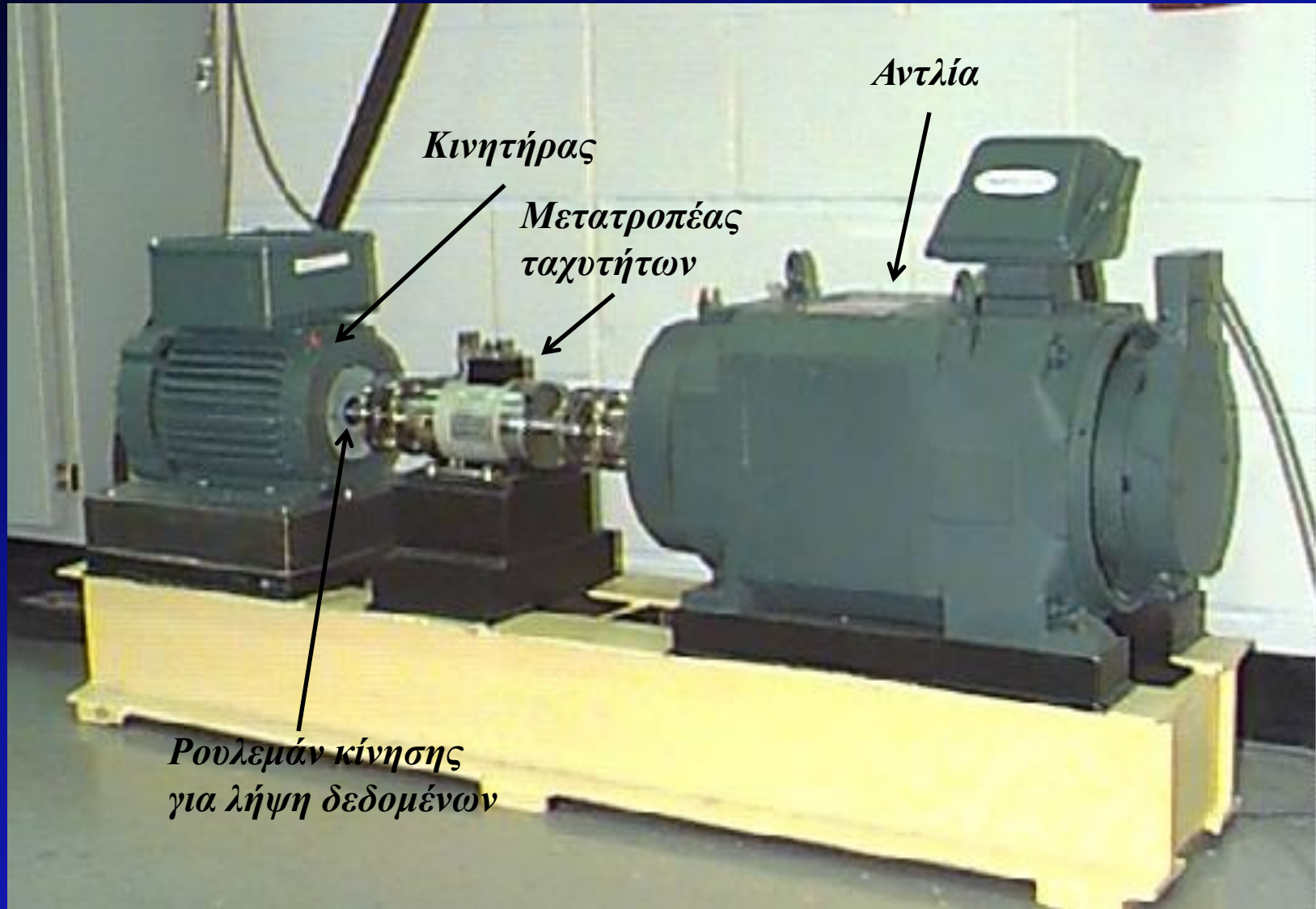
Οι σημειακές αλλοιώσεις σε διάμετρο ήταν, **0.007 ίντσες, 0.014 ίντσες, 0.021 ίντσες**, ενώ τα δεδομένα δόνησης καταγράφηκαν για φορτία κινητήρα ισχύος έως 3.0 hp και με ταχύτητες κινητήρα έως 1720 στροφές το λεπτό.

Οι αλλοιώσεις αυτές χαράχθηκαν ξεχωριστά σε τρία μέρη του ρουλεμάν, **στην εσωτερική κούρσα του ρουλεμάν, στο περίβλημα (σφαίρες) και στην εξωτερική πλευρά του ρουλεμάν**, σε σημεία αντίστοιχα με τις ώρες 12, 3 και 6 του ρολογιού.

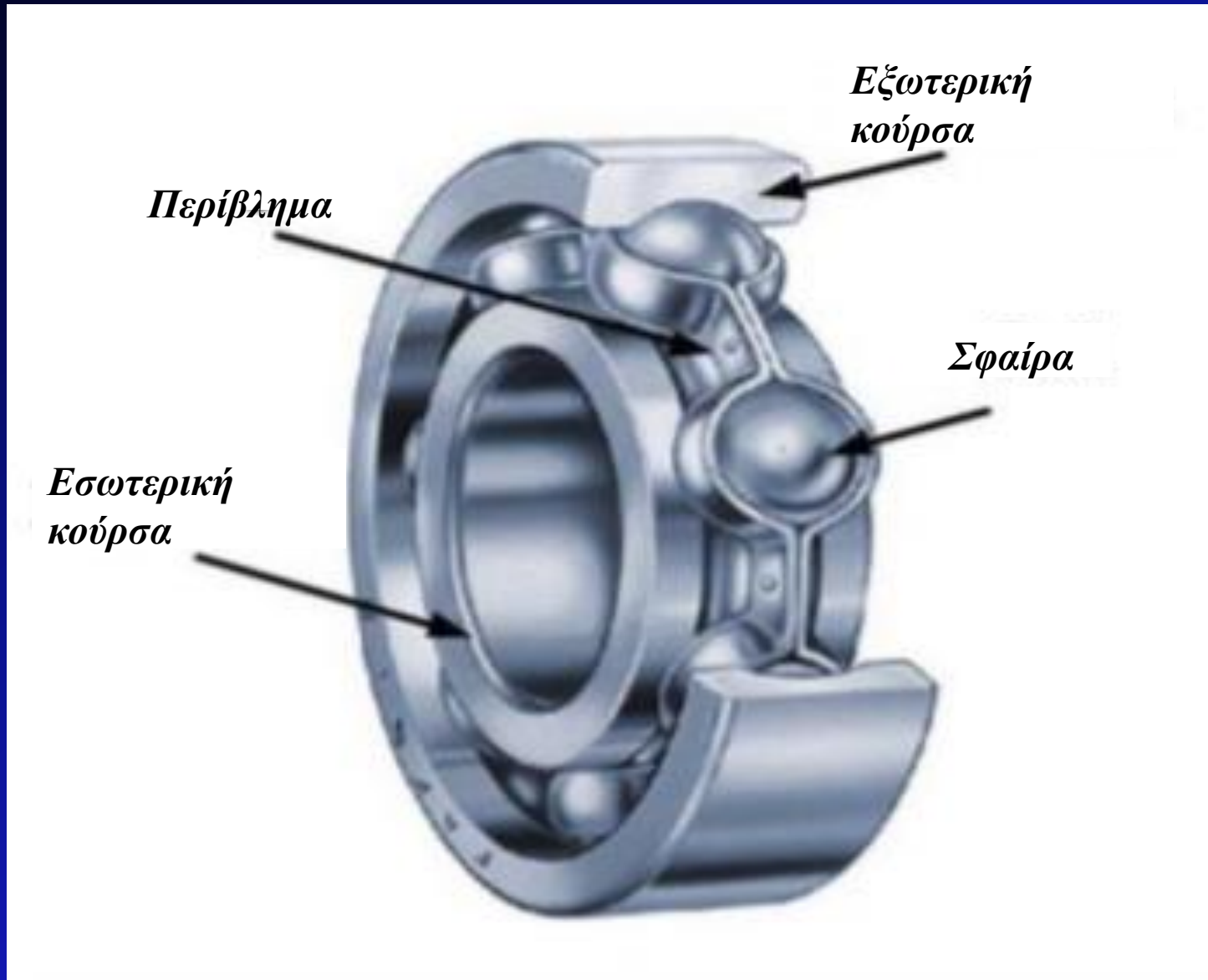
Τα σήματα της δόνησης, συλλέχθηκαν χρησιμοποιώντας μια συσκευή καταγραφής των δεδομένων, που αποτελούνταν **από 16 κανάλια**. Τα ψηφιακά δεδομένα συλλέχθηκαν στο **άκρο της κίνησης** με ρυθμό δειγμάτων **12.0 kHz και 48.0 kHz**. Η συλλογή διήρκεσε **100 δευτερόλεπτα**.

Διάταξη για μετρήσεις ρουλεμάν με αλλοιώσεις

Bearing data center drive end fault



Ρουλεμάν με τα τέσσερα κύρια μέρη αυτού



Κώδικας SVM. Εφαρμογή 1

Δεδομένα από: *Bearing data center drive end fault*

```
10
11 data_time = pd.read_csv("svm_data.txt")
12
13 data_time['fault'] = pd.Categorical(data_time['fault'])
14
15 ▼ train_data, test_data = train_test_split(data_time, test_size = 750, stratify = data_time['fault'],
16                                           random_state = 1234)
17 test_data['fault'].value_counts()
18
19
20 scaler = StandardScaler()
21 train_data_scaled = scaler.fit_transform(train_data.iloc[:, :-1])
22 test_data_scaled = (test_data.iloc[:, :-1].values - scaler.mean_) / np.sqrt(scaler.var_)
23 ▼ parameters = {"C": [1, 10, 50, 100, 300, 500],
24                'gamma': [0.01, 0.05, 0.1, 0.5, 1, 5],
25                'kernel': ["rbf"]}
26 tuned_svm_clf = GridSearchCV(SVC(), parameters, n_jobs = -1, cv = 10)
```

C και gamma παράμετροι για την συνάρτηση kernel rbf

SVM. Εφαρμογή 1

Αποτελέσματα δοκιμής: *Bearing data center drive end fault*

Test Confusion Matrix

C1	74	0	0	0	0	0	0	0	1	0
C2	1	66	0	0	0	0	2	0	5	1
C3	0	0	68	0	0	0	0	0	7	0
C4	0	0	0	75	0	0	0	0	0	0
C5	0	0	0	0	75	0	0	0	0	0
C6	0	0	0	0	0	75	0	0	0	0
C7	0	0	0	0	0	0	75	0	0	0
C8	0	0	0	0	0	0	0	75	0	0
C9	3	3	2	0	0	0	0	0	67	0
C10	0	0	0	1	0	0	0	0	0	74
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10

QGA-SVM. Εφαρμογή 2.

Δεδομένα από: *Zhu et al., 2018*

Στο πείραμα, η συνάρτηση *rbf* χρησιμοποιήθηκε ως ο πυρήνας του SVM. Έγινε βελτιστοποίηση των παραμέτρων του QGA-SVM και οι παράμετροι του γενετικού αλγορίθμου ορίστηκαν ως εξής:

το μέγεθος του πληθυσμού (*population_size*) ήταν 20,

ο αριθμός των γενεών ή ανακυκλώσεων (*iter_num*) ήταν 200,

η παράμετρος *C* ήταν στο εύρος [0, 100],

το εύρος της παραμέτρου σ^2 ή *gamma* ήταν [0, 1000],

ο αριθμός πολλαπλών επικυρώσεων (*cross validation*) ήταν 5 και,

τέλος, η πιθανότητα πλάτους γωνίας περιστροφής του κβαντικού γενετικού αλγορίθμου ήταν 0.009π.

Αποτελέσματα δοκιμής

Τα αποτελέσματα του αλγόριθμου QGA-SVM δίνουν *διαγνωστική ακρίβεια 0.925*.

QGA-SVM

Παρούσα μελέτη

Χρησιμοποιήθηκε υπολογιστικός αλγόριθμος QGA-SVM, με την τεχνική rbf.

Το σύνολο των δεδομένων χωρίζεται σε 1550 train size και με 750 test size.

Τυχαία ανάμειξη των δεδομένων, πριν το διαχωρισμό σε train/test, με τιμή 1234.

Οι παράμετροι που χρησιμοποιήθηκαν είναι,

population_size=20

chromosome_num=2

chromosome_length=17

max_value=15

min_value=0.01

iter_num=500

deta=0.1 * np.pi

Χρήση συνάρτησης rbf (radial basis function)

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

$$\sum_{i=1}^l a_i y_i = 0 \quad \& \quad a_i \geq 0 \\ 0 \leq a_i \leq C$$

Οι παράμετροι C και γ της συναρτήσεως του πυρήνα (kernel), βελτιστοποιούνται μέσω του κβαντικού αλγόριθμου QGA.

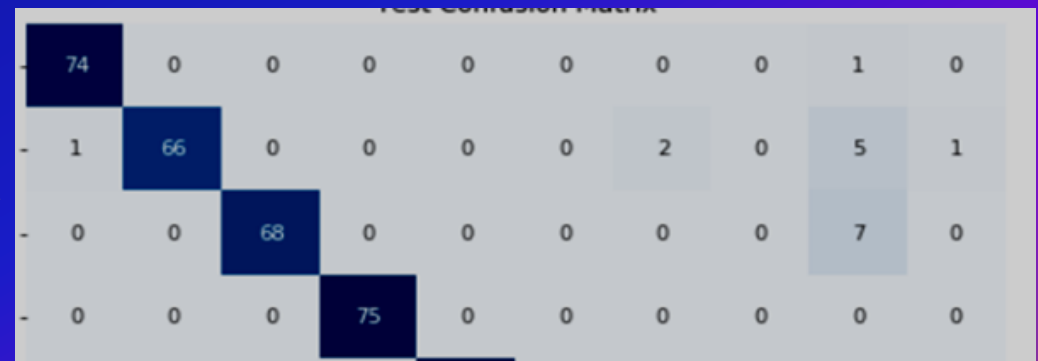
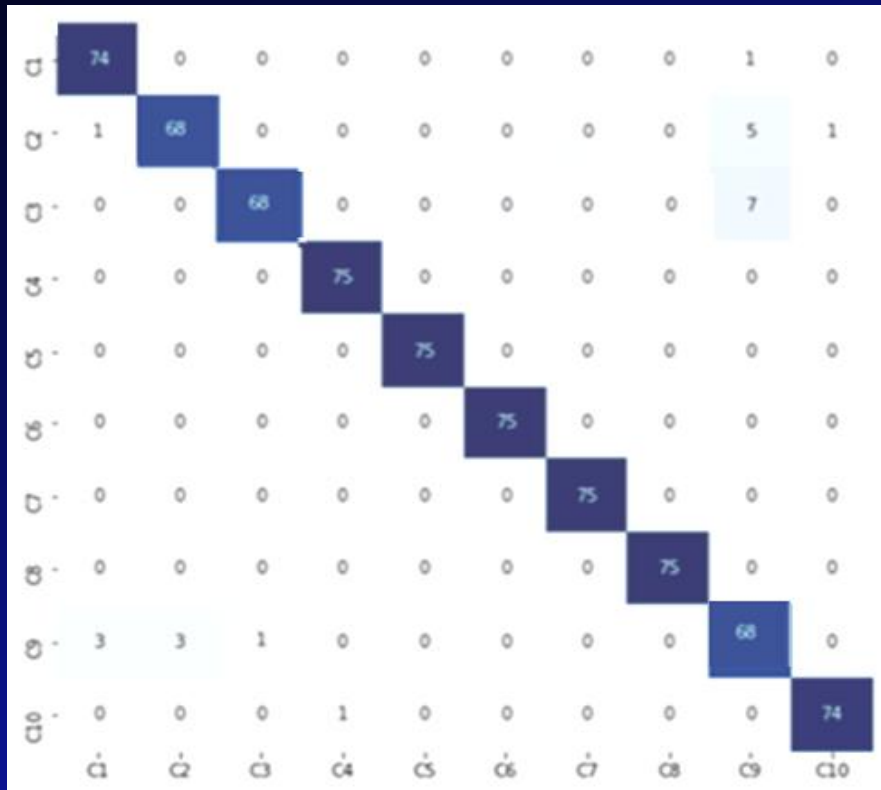
Στην τελική ανακύκλωση της παρούσας εργασίας με $f=0.9667$ η τιμή του C ισούται με 13.513 και του $\gamma=1/(2\sigma^2)$ είναι 0.383.

Μετά από διαδοχικές επαναληπτικές διαδικασίες για το ίδιο πρόβλημα, το αποτέλεσμα της συνάρτησης προσαρμοστικότητας f υπολογίσθηκε ως ίσο προς 0.9653 με τιμές $C=14.839$ και $\gamma=0.366$ μετά από 77 ανακυκλώσεις. Σε άλλους υπολογισμούς το ίδιο βέλτιστο της f βρέθηκε να είναι $C=11.471$ και $\gamma=0.382$ μετά από 9 ανακυκλώσεις.

Συμπερασματικά η τιμή του γ διαφοροποιείται ελάχιστα από υπολογισμό σε υπολογισμό ($=0.383$) ενώ η τιμή του C βρίσκεται γύρω στο 11.5-14.5 προκειμένου να βρεθεί η βέλτιστη τιμή της συνάρτησης προσαρμοστικότητας.

QGA-SVM. Παρούσα μελέτη

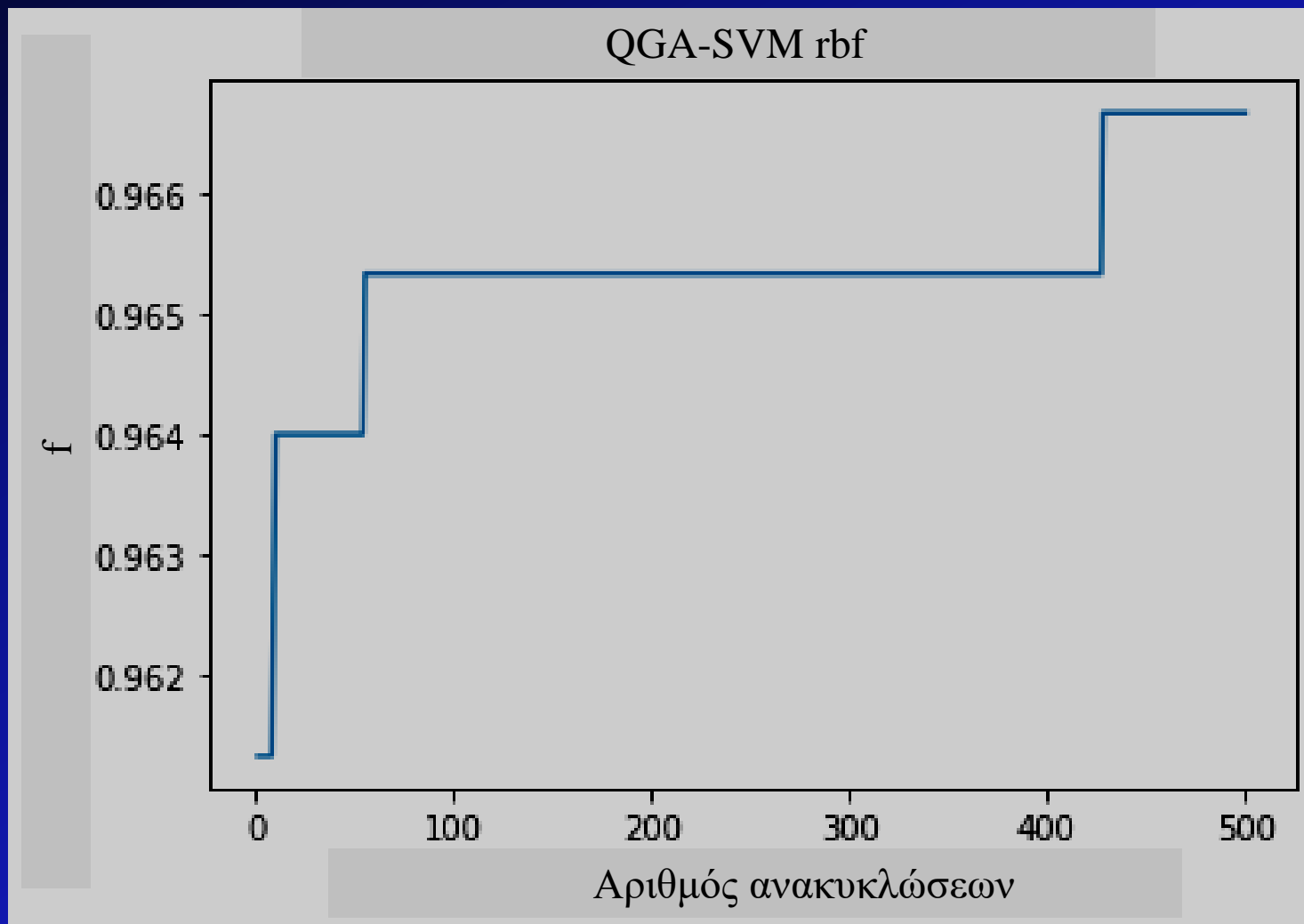
Αποτελέσματα δοκιμής



QGA-SVM. Εφαρμογή 1

Bearing data center drive end fault

Βελτιστοποίηση της συνάρτησης προσαρμοστικότητας f του αλγόριθμου QGA-SVM ($f=0.9667$)



Σύγκριση αποτελεσμάτων QGA-SVM με την τεχνική υπολογιστική τεχνική SVM

QGA-SVM. Παρούσα μελέτη

Μέθοδος	Αριθμός δεδομένων δοκιμής	Αριθμός σωστά διαγνωσμένων λαθών	Διαγνωστική ακρίβεια
SVM	2300	2220	0.9653
QGA-SVM	2300	2223	0.9667

Zhu et al., 2018

Μέθοδος	Αριθμός δεδομένων δοκιμής	Αριθμός σωστά διαγνωσμένων λαθών	Διαγνωστική ακρίβεια
SVM	40	34	0.85
QGA-SVM	40	37	0.925

QGA-SVM. Αποτελέσματα

```
iteration is : 500 ;Best parameters: [13.51314552307129, 0.3829430389404297] ;Best fitness 0.966666666666667
0.878666666666667
0.848
0.866666666666667
0.824
0.830666666666667
0.944
0.862666666666667
0.936
0.853333333333334
0.944
0.846666666666667
0.950666666666667
0.932
0.913333333333333
0.869333333333333
0.890666666666667
0.850666666666667
0.821333333333334
```

C

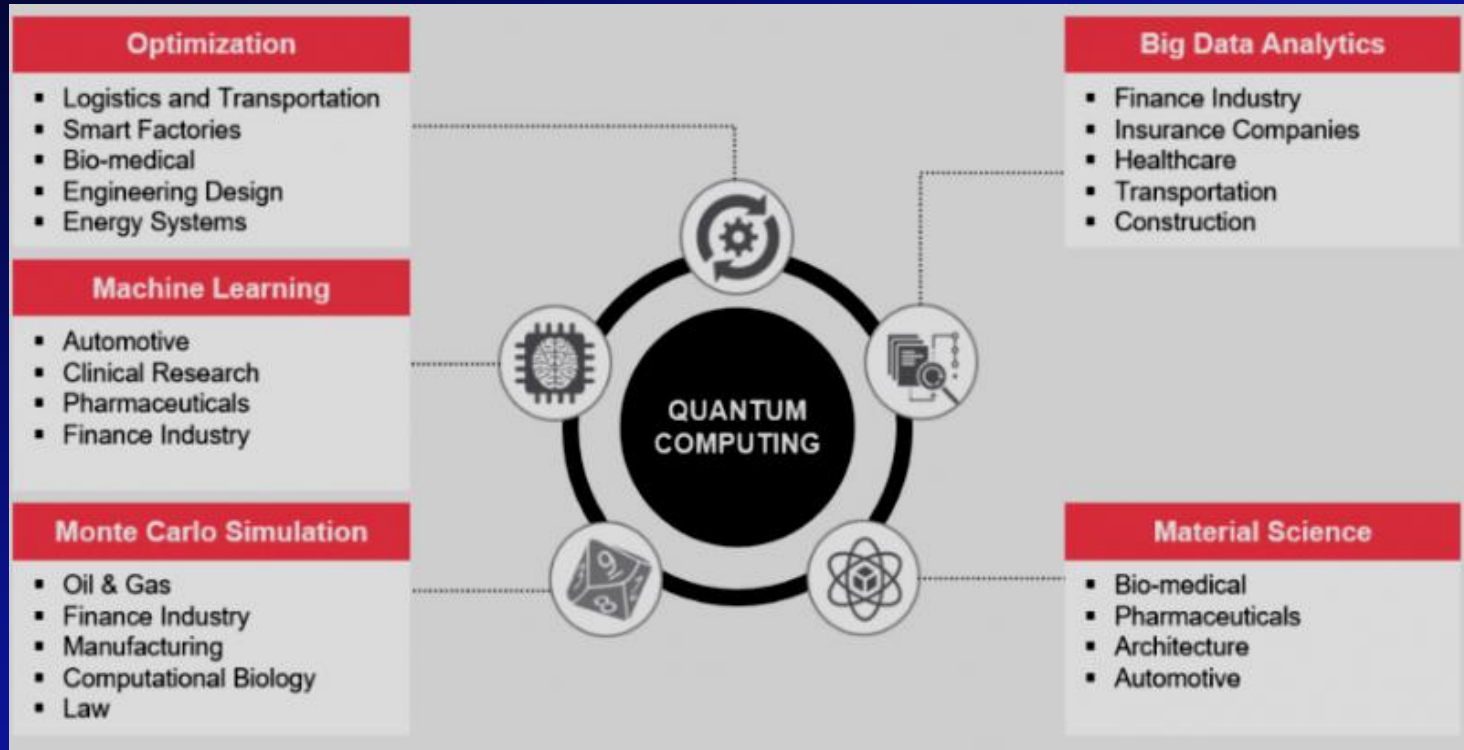
gamma

SVM. Αποτελέσματα

Bearing data center drive end fault

```
{'C': 50, 'gamma': 0.05, 'kernel': 'rbf'}
overall_test_accuracy, 0.9653333333333334
```

QGA-SVM. Εφαρμογές



QGA-SVM. Συμπεράσματα

- α) Η προτεινόμενη μέθοδος έχει **μεγαλύτερη ακρίβεια** στη διάγνωση σφαλμάτων από την παραδοσιακή μέθοδο SVM.
- β) Η εύρεση του βέλτιστου αποτελέσματος επιτυγχάνεται **πιο γρήγορα και πιο εύκολα** μέσω του αυτόματου υπολογισμού των C και σ^2 .

Σας ευχαριστώ