

Εκπαίδευση Τεχνητών Νευρωνικών Δικτύων ανά Πρότυπο Εισόδου

Β.Π. Πλαγιανάκος¹, Γ.Δ. Μαγουλάς² και Μ.Ν. Βραχάτης¹

¹ Πανεπιστήμιο Πατρών, Τμήμα Μαθηματικών, Πάτρα 26500, Ελλάδα.

² Department of Information Systems and Computing, Brunel University, Uxbridge UB8 3PH, U.K.
e-mail: vpp@math.upatras.gr, George.Magoulas@brunel.ac.uk, vrahatis@math.upatras.gr

Περίληψη— Σε αυτή την εργασία θα ασχοληθούμε με την εκπαίδευση ΤΝΔ ανά πρότυπο εισόδου (online training). Αυτή η προσέγγιση στην εκπαίδευση ΤΝΔ θεωρείται κατεξοχήν κατάλληλη για περιπτώσεις όπου η εκπαίδευση ανά ομάδα προτύπων εισόδου απαιτεί σημαντικό χρόνο και μεγάλο αποθηκευτικό χώρο, όπως συμβαίνει όταν έχουμε μεγάλα σύνολα προτύπων ή/και δίκτυα. Επιπρόσθετα, η προσέγγιση αυτή συχνά βοηθά στην αποφυγή των τοπικών ελάχιστων και παρέχει μια πιο φυσική προσέγγιση για την εκπαίδευση ΤΝΔ σε μη στάσιμα περιβάλλοντα, δηλαδή την εκπαίδευση και επίλυση προβλημάτων που το σύνολο των προτύπων μεταβάλλεται με τον χρόνο.

Εδώ θα παρουσιάσουμε μια μέθοδο για την αυτόματη προσαρμογή ενός κοινού ρυθμού εκπαίδευσης για όλα τα βάρη, που χρησιμοποιείται στην εκπαίδευση ανά πρότυπο εισόδου. Η προτεινόμενη τεχνική αποτελεί τροποποίηση της στοχαστικής μεθόδου της πιο απότομης καθόδου (stochastic gradient descent), σε συνδυασμό με την μέθοδο που προτείνεται στην εργασία [1]. Η νέα πτυχή της προσέγγισής μας συνίσταται στο γεγονός ότι λαμβάνονται υπόψη ήδη υπολογισμένες πληροφορίες σχετικά με τις παραγώγους και το ρυθμό εκπαίδευσης προηγούμενων επαναλήψεων. Η χρήση επιπλέον πληροφορίας επιτρέπει καλύτερη προσαρμογή του κοινού ρυθμού εκπαίδευσης και οδηγεί σε γρηγορότερη εκπαίδευση.

Ο προτεινόμενος αλγόριθμος έχει εφαρμοστεί, εξεταστεί και συγκριθεί με άλλες μεθόδους εκπαίδευσης ανά πρότυπο εισόδου, αλλά και μεθόδους εκπαίδευσης ανά ομάδα προτύπων εισόδου (batch training). Τα αποτελέσματα που παρουσιάζουμε δείχνουν ότι συμπεριφέρεται προβλέψιμα, είναι αποδοτικός και ότι έχει αρκετά ικανοποιητική μέση απόδοση [8, 9, 15, 16].

1 Εισαγωγή

Στο ερευνητικό πεδίο των ΤΝΔ, τα δίκτυα που η εκπαίδευσή τους βασίζεται στην μέθοδο της οπισθοδρομικής διάδοσης του σφάλματος (Backpropagation) [21] είναι τα πιο δημοφιλή. Αποδοτικοί αλγόριθμοι εκπαίδευσης τέτοιων ΤΝΔ είναι ένα βασικό θέμα της επίκαιρης έρευνας και για αυτόν τον σκοπό έχουν προταθεί πολυάριθμοι αλγόριθμοι.

Μια κοινή προσέγγιση εκπαίδευσης είναι η ελαχιστοποίηση του σφάλματος των δικτύων, το οποίο αποτελεί ένα μέτρο της απόδοσής τους. Το σφάλμα αυτό υπολογίζεται συνήθως ως η διαφορά μεταξύ του πραγματικού διανύσματος εξόδου του δικτύου και του επιθυμητού διανύσματος εξόδου (εκπαίδευση με επίβλεψη). Ο γρήγορος υπολογισμός ενός συνόλου βαρών που ελαχιστοποιεί αυτό το σφάλμα είναι μια μάλλον δύσκολη εργασία λόγω του ότι, γενικά, ο αριθμός των βαρών του δικτύων είναι ιδιαίτερα μεγάλος και η συνάρτηση σφάλματος παράγει μια περίπλοκη επιφάνεια. Η συνάρτηση σφάλματος συνήθως έχει πολλά τοπικά ελάχιστα και μεγάλες επίπεδες περιοχές (flat regions), που ακολουθούνται από πολύ στενές περιοχές με μεγάλη κλίση.

Οι αλγόριθμοι εκπαίδευσης με επίβλεψη μπορούν να χωριστούν σε δύο βασικές κατηγορίες:

- αλγόριθμοι εκπαίδευσης ανά πρότυπο εισόδου (on-line ή stochastic training), και
- αλγόριθμοι εκπαίδευσης ανά ομάδα προτύπων εισόδου (batch ή off-line training).

Η εκπαίδευση ανά ομάδα προτύπων εισόδου είναι η κλασική προσέγγιση, όπου ένα σύνολο παραδειγμάτων λαμβάνεται και χρησιμοποιείται προκειμένου να εκπαιδευτεί το ΤΝΔ, προτού αυτό χρησιμοποιηθεί σε κάποια εφαρμογή. Αντίθετα, στην εκπαίδευση ανά πρότυπο εισόδου τα στοιχεία που συγκεντρώνονται κατά τη διάρκεια της κανονικής λειτουργίας του συστήματος χρησιμοποιούνται για την συνεχή εκπαίδευση και προσαρμογή του ΤΝΔ.

Η εκπαίδευση ανά ομάδα προτύπων εισόδου αντιμετωπίζεται με τη θεωρία της βελτιστοποίησης χωρίς περιορισμούς, αφού η πραγματική κλίση της συνάρτησης σφάλματος και πληροφορίες από ολόκληρο το σύνολο χρησιμοποιούνται. Η εκπαίδευση ανά ομάδα προτύπων εισόδου αντιμετωπίζεται με την ελαχιστοποίηση της συνάρτησης σφάλματος E που ορίζεται ως το άθροισμα, για όλα τα πρότυπα εισόδου, των τετραγώνων των διαφορών της πραγματικής εξόδου του ΤΝΔ από την επιθυμητή έξοδο:

$$E(w) = \frac{1}{2} \sum_{p=1}^P \sum_{j=1}^{N_L} (y_{j,p}^L - t_{j,p})^2 = \sum_{p=1}^P E_p, \quad (1)$$

όπου P είναι ο συνολικός αριθμός προτύπων, $y_{j,p}^L$ η έξοδος του j νευρώνα που ανήκει στο L στρώμα, N_L ο αριθμός των νευρώνων του στρώματος εξόδου, $t_{j,p}$ η επιθυμητή έξοδος του j νευρώνα εξόδου στο πρότυπο p , και E_p το σφάλμα του δικτύου ανά πρότυπο εισόδου. Στην εκπαίδευση ανά πρότυπο εισόδου τα βάρη του δικτύου ενημερώνονται μετά από την παρουσίαση κάθε πρότυπο εισόδου, το οποίο μπορεί να επιλεγεί από το σύνολο εκπαίδευσης με ή χωρίς επανατοποθέτηση. Αυτό αντιστοιχεί στην ελαχιστοποίηση του στιγμιαίου σφάλματος E_p του ΤΝΔ, όπως φαίνεται στη Σχέση (1).

Η εκπαίδευση ανά πρότυπο εισόδου μπορεί να επιλεγεί για προβλήματα που έχουν είτε πολύ μεγάλο αριθμό προτύπων (πιθανά και κάποιο αριθμό περιπτώσεων ή λανθασμένων προτύπων), είτε γιατί προσπαθούμε να προσεγγίσουμε ένα αργά μεταβαλλόμενο σύστημα. Αν και η εκπαίδευση ανά ομάδα προτύπων εισόδου φαίνεται να είναι ταχύτερη για μικρά σύνολα προτύπων και δίκτυα, η εκπαίδευση ανά πρότυπο εισόδου είναι σαφώς γρηγορότερη για μεγάλα σύνολα προτύπων και ΤΝΔ με πολλά βάρη (πολλές χιλιάδες βάρη και πολώσεις), βοηθά σημαντικά στην αποφυγή τοπικών ελαχίστων και παρέχει μια φυσική προσέγγιση για την εκπαίδευση ΤΝΔ σε μη στατικά προβλήματα, δηλαδή προβλήματα που το σύνολο προτύπων μεταβάλλεται αργά με τον χρόνο.

Λαμβάνοντας υπόψη την έμφυτη αποδοτικότητα της στοχαστικής μεθόδου της πιο απότομης καθόδου, έχουν προταθεί πρόσφατα διάφοροι αλγόριθμοι [1, 22, 24–26]. Δυστυχώς, η εκπαίδευση ανά πρότυπο εισόδου πάσχει από κάποια μειονεκτήματα όπως για παράδειγμα η ευαισθησία στις παραμέτρων εκπαίδευσης [22]. Ένα άλλο μειονέκτημα είναι ότι πιο προηγμένες μέθοδοι βελτιστοποίησης, όπως οι μέθοδοι συζυγών κλίσεων (conjugate gradient), οι μέθοδοι μεταβλητής μετρικής (variable metric), η μέθοδος της προσομοιωμένης απόπτωσης (simulated annealing) κ.α., μπορούν να εφαρμοστούν μόνο σε σταθερές επιφάνειες σφάλματος, και έτσι υπάρχει δυσκολία στη χρησιμοποίησή τους στην εκπαίδευση ανά πρότυπο εισόδου [22].

Εντούτοις, η εκπαίδευση ανά πρότυπο εισόδου έχει πολλά πλεονεκτήματα σε σχέση με την εκπαίδευση ανά ομάδα προτύπων εισόδου. Οι μέθοδοι εκπαίδευσης ανά πρότυπο εισόδου παρουσιάζουν σθεναρότητα ως προς λανθασμένα και περιττά πρότυπα, και έτσι παραλείψεις στο σύνολο προτύπων εκπαίδευσης μπορούν να διορθωθούν κατά τη διάρκεια της εκπαίδευσης του ΤΝΔ. Επιπλέον, τα πρότυπα εκπαίδευσης μπορούν συχνά να παραχθούν εύκολα και σε μεγάλες ποσότητες όταν το σύστημα είναι σε λειτουργία, ενώ είναι συνήθως λιγότερα πριν την αρχή της εκπαίδευσης. Γενικά, η εκπαίδευση ανά πρότυπο εισόδου είναι απαραίτητη εάν χρειαζόμαστε συστήματα με δυνατότητα εκμάθησης, σε αντιδιαστολή με τα απλά εκπαιδευμένα συστήματα [27].

2 Αλγόριθμοι Εκπαίδευσης ανά Πρότυπο Εισόδου

Παρά την αφθονία μεθόδων για εκπαίδευση ΤΝΔ, υπάρχουν μόνο λίγες μέθοδοι που μπορούν να χρησιμοποιηθούν αποδοτικά για εκπαίδευση ανά πρότυπο εισόδου. Για παράδειγμα, οι κλασικοί αλγόριθμοι

εκπαίδευσης ανά ομάδα προτύπων εισόδου δεν μπορούν απευθείας να χειριστούν μη στατικά σύνολα προτύπων. Ακόμα και όταν μερικοί από αυτούς χρησιμοποιούνται στην εκπαίδευση ανά πρότυπο εισόδου, συχνά παρουσιάζεται το πρόβλημα της «καταστροφικής παρέμβασης», δηλαδή η εκπαίδευση του ΤΝΔ στα νέα πρότυπα παρεμποδίζεται υπερβολικά από τα προηγούμενα, οδηγώντας σε κορεσμό και την αργή σύγκλιση [27].

Οι μέθοδοι που μπορούν να χρησιμοποιηθούν για εκπαίδευση ανά πρότυπο εισόδου, είναι εκείνες που μπορούν με επιτυχία να χειριστούν μεταβλητά με το χρόνο σύνολα προτύπων (time-varying training sets), ενώ συγχρόνως απαιτούν λίγους σχετικά πρόσθετους υπολογιστικούς πόρους (μνήμη και επεξεργαστική ισχύ) προκειμένου να υποβληθεί σε επεξεργασία κάθε πρόσθετο πρότυπο. Παραδείγματα τέτοιων μεθόδων είναι οι παραλλαγές της στοχαστικής μεθόδου της πιο απότομης καθόδου [1].

Η πρώτη μέθοδος που θα μελετήσουμε, ονομάζεται ALAP₁, και χρησιμοποιεί σε κάθε επανάληψη ένα κοινό ρυθμό εκπαίδευσης για όλα τα βάρη:

$$\eta_i^k = \eta_i^{k-1} + \gamma \langle \nabla E_{p-1}(w^{k-1}), \nabla E_p(w^k) \rangle, \quad (2)$$

όπου $\eta_i^0 = c$ για όλα τα βάρη του δικτύου (το c είναι μια μικρή θετική σταθερά), και $\langle \cdot, \cdot \rangle$ συμβολίζουμε το εσωτερικό γινόμενο στον \mathbb{R}^n .

Οι άλλες δύο μέθοδοι που περιγράφονται στην εργασία [1] (ALAP₂ και ALAP₃), χρησιμοποιούν διαφορετικό ρυθμό εκπαίδευσης για κάθε βάρος. Αυτό το χαρακτηριστικό γνώρισμα καθιστά αυτούς τους αλγόριθμους εκπαίδευσης ανά πρότυπο εισόδου ικανούς να χρησιμοποιούν παραλλαγές της κατεύθυνσης της πιο απότομης καθόδου και να κινούνται κατά μήκος μιας κατεύθυνσης που δεν συμπίπτει απαραίτητως με αυτή, με αποτέλεσμα συχνά να επιταχύνεται η διαδικασία ελαχιστοποίησης.

Ο αλγόριθμος ALAP₂ χρησιμοποιεί τον ακόλουθο τύπο για την προσαρμογή του ρυθμού εκπαίδευσης:

$$\eta_i^k = \eta_i^{k-1} \left[1 + \gamma \partial_i E_{p-1}(w^{k-1}) \partial_i E_p(w^k) \right]. \quad (3)$$

Ο αλγόριθμος ALAP₂ χρησιμοποιεί μια κανονικοποιημένη έκδοση του κανόνα προσαρμογής του ALAP₂, που δίνεται από τον τύπο:

$$\eta_i^k = \eta_i^{k-1} \left[1 + \gamma \frac{\partial_i E_{p-1}(w^{k-1}) \partial_i E_p(w^k)}{u_i^k} \right], \quad (4)$$

όπου u_i^k αποτελεί τον εκθετικό μέσο όρο του τετραγώνου της μερικής παραγώγου του στιγμιαίου σφάλματος, $\partial_i E_p(w^k)$, που δίνεται από τον τύπο:

$$u_i^k = \mu u_i^{k-1} + (1 - \mu) \left[\partial_i E_p(w^k) \right]^2,$$

όπου μ και γ είναι θετικές σταθερές. Ενδεικτικά, οι τιμές $\mu = 0.9$ και $\gamma = 0.01$ είναι κατάλληλες για τις περισσότερες των περιπτώσεων [1].

Στη συνέχεια προτείνουμε τον παρακάτω νέο αλγόριθμο για την προσαρμογή ενός κοινού ρυθμού εκπαίδευσης για όλα τα βάρη, στα πλαίσια της στοχαστικής μεθόδου της πιο απότομης καθόδου:

$$\begin{aligned} \eta^{k+1} = \eta^k + \gamma_1 \langle \nabla E_{p-1}(w^{k-1}), \nabla E_p(w^k) \rangle \\ + \gamma_2 \langle \nabla E_{p-2}(w^{k-2}), \nabla E_{p-1}(w^{k-1}) \rangle. \end{aligned} \quad (5)$$

Το βασικό γνώρισμα του προτεινόμενου αλγόριθμου είναι ότι αξιοποιεί πληροφορίες σχετικές με την κλίση, από την τρέχουσα καθώς επίσης και τις δύο προηγούμενες παρουσιάσεις προτύπων [8, 9, 16].

Αυτό φαίνεται να παρέχει κάποια σταθεροποίηση στην προσαρμογή των τιμών του ρυθμού εκπαίδευσης και βοηθά την στοχαστική μέθοδο της πιο απότομης καθόδου να επιτύχει γρήγορη σύγκλιση και υψηλό ποσοστό επιτυχίας. Μια αλγοριθμική περιγραφή του προτεινόμενου αλγορίθμου δίνεται στον Αλγόριθμο 1.

Σε αυτό το αλγοριθμικό μοντέλο, η είναι ο ρυθμός εκπαίδευσης, και γ_1 και γ_2 είναι μετα-ρυθμοί εκπαίδευσης (meta-learning rates). Σαν συνθήκη τερματισμού μπορεί να χρησιμοποιηθεί το σφάλμα ταξινόμησης ή ένα άνω όριο στους υπολογισμούς της συνάρτησης σφάλματος.

Αλγόριθμος 1: Ο προτεινόμενος αλγόριθμος σε ψευδοκώδικα.

ΣΤΟΧΑΣΤΙΚΗ ΜΕΘΟΔΟΣ ΤΗΣ ΠΙΟ ΑΠΟΤΟΜΗΣ ΚΑΘΟΔΟΥ ΜΕ ΜΕΤΑΒΛΗΤΟ ΡΥΘΜΟ ΕΚΠΑΙΔΕΥΣΗΣ

- 0: Αρχικοποίησε w^0 , η^0 , γ_1 , και γ_2 .
- 1: **Repeat**
- 2: Θέσε $k = k + 1$.
- 3: Επέλεξε τυχαία το p πρότυπο από το σύνολο προτύπων.
- 4: Υπολόγισε $E_p(w^k)$ και μετά $\nabla E_p(w^k)$.
- 5: Υπολόγισε τα νέα βάρη:
$$w^{k+1} = w^k - \eta^k \nabla E_p(w^k).$$
- 6: Υπολόγισε το νέο ρυθμό εκπαίδευσης:
$$\eta^{k+1} = \eta^k + \gamma_1 \langle \nabla E_{p-1}(w^{k-1}), \nabla E_p(w^k) \rangle + \gamma_2 \langle \nabla E_{p-2}(w^{k-2}), \nabla E_{p-1}(w^{k-1}) \rangle.$$
- 7: **Until** Συνθήκη Τερματισμού.
- 8: **Return** τα τελικά βάρη w^{k+1} .

3 Προσομοιώσεις και Αποτελέσματα

Σε αυτή την Ενότητα παρουσιάζουμε τα αποτελέσματα του προτεινόμενου αλγόριθμου εκπαίδευσης ανά πρότυπο εισόδου σε γνωστά προβλήματα εκπαίδευσης ΤΝΔ. Για τον λόγο αυτό, ο Αλγόριθμος-1 έχει αξιολογηθεί και έχει συγκριθεί με στοχαστικές μεθόδους καθώς επίσης και με μεθόδους εκπαίδευσης ανά ομάδα προτύπων εισόδου. Πιο συγκεκριμένα, στις προσομοιώσεις έχουμε συγκρίνει τον Αλγόριθμος-1 με τις 3 στοχαστικές μεθόδους προσαρμογής του ρυθμού εκπαίδευσης που έχουν προταθεί από τους Almeida, Langlois, Amaral και Plankhon στην εργασία [1] (ALAP₁, ALAP₂ και ALAP₃), και την μέθοδο οπισθοδρομικής διάδοσης του σφάλματος για εκπαίδευση ανά πρότυπο εισόδου (On-line BP) [21]. Επιπλέον, για τις συγκρίσεις έχουμε εξετάσει και τις μεθόδους εκπαίδευσης ανά ομάδα προτύπων εισόδου: την μέθοδο οπισθοδρομικής διάδοσης του σφάλματος για εκπαίδευση ανά ομάδα προτύπων εισόδου [21] (Batch BP), και την μέθοδο οπισθοδρομικής διάδοσης του σφάλματος με προσαρμοστικό ρυθμό εκπαίδευσης και ορμή (momentum) για εκπαίδευση ανά ομάδα προτύπων εισόδου [29] (Batch ABP).

Οι αλγόριθμοι εξετάστηκαν χρησιμοποιώντας τα ίδια αρχικά βάρη, που αρχικοποιήθηκαν με τη μέθοδο των Nguyen-Widrow [12], και έλαβαν την ίδια ακολουθία προτύπων εισόδου. Για κάθε πρόβλημα που περιγράφεται κατωτέρω, παρουσιάζουμε έναν πίνακα που συνοψίζει την απόδοση των αλγορίθμων για τις προσομοιώσεις που έφθασαν σε σύγκλιση. Η εκπαίδευση θεωρείται επιτυχής όταν το ΤΝΔ δεν παρουσιάζει λάθη ταξινόμησης στο σύνολο εκπαίδευσης.

Οι αναφερόμενες στους παρακάτω πίνακες παράμετροι είναι: *Min* ο ελάχιστος αριθμός παρουσιάσεων προτύπων, μ η μέση τιμή των παρουσιάσεων προτύπων, *Max* η ανώτατη τιμή παρουσιάσεων προτύπων, και Επιτυχία (%) ο αριθμός των επιτυχημένων προσομοιώσεων σε σύνολο 100 δοκιμών. Εάν ένας αλγόριθμος αποτύχει να συγκλίνει μέσα σε ένα προκαθορισμένο όριο υπολογισμών της συνάρτησης σφάλματος, θεωρείται ότι αποτυγχάνει να εκπαιδεύσει το ΤΝΔ, και οι παρουσιάσεις προτύπων κατά την δοκιμή αυτή δεν συμπεριλαμβάνονται στη στατιστική ανάλυση των αποτελεσμάτων.

Οι τιμές των παραμέτρων γ_1 και γ_2 επιλέχθηκαν να είναι $\gamma_1 \ll \gamma_2 = 1$. Φαίνεται ότι η επιλογή τιμών για τους μετά-ρυθμούς εκπαίδευσης γ_1 και γ_2 δεν είναι κρίσιμη για την επιτυχή εκπαίδευση. Εντούτοις, μπορεί να επιτευχθεί γρηγορότερη σύγκλιση, εάν γίνει μια ακριβής προσαρμογή τους ανάλογα με το πρόβλημα. Τέτοια προσαρμογή δεν κρίθηκε αναγκαία στα πειράματά μας, γιατί βασικός σκοπός μας ήταν η συγκριτική μελέτη των μεθόδων και όχι η βέλτιστη απόδοσή τους. Αφ' ετέρου, μεγάλη προσπάθεια έχει γίνει για να προσδιοριστούν κατάλληλες τιμές για τις ευρετικές παραμέτρους των μεθόδων εκπαίδευσης ανά ομάδα προτύπων BP και ABP. Η εμπειρία μας με προσομοιώσεις των μεθόδων αυτών δείχνει ότι η συμπεριφορά των αλγορίθμων, που περιγράφεται στα παραδείγματα που ακολουθούν, είναι χαρακτηριστική.

3.1 Το πρόβλημα του Αποκλειστικού-ΕΙΤΕ

Το πρόβλημα του Αποκλειστικού-ΕΙΤΕ έχει θεωρηθεί ιστορικά ως κλασικό πρόβλημα εκπαίδευσης ΤΝΔ. Στην περίπτωση μας έχει χρησιμοποιηθεί ένα 2-2-1 ΤΝΔ (έξι βάρη και τρεις πολώσεις). Το δίκτυο είναι βασισμένο σε νευρώνες με λογιστικές συναρτήσεις ενεργοποιήσεις. Το όριο υπολογισμών της συνάρτησης σφάλματος ήταν 4000, δηλαδή επιτράπηκαν μόνο 4000 παρουσιάσεις προτύπων. Τα συγκριτικά αποτελέσματα εμφανίζονται στον Πίνακα 1.

Από τα αποτελέσματα είναι εμφανές ότι ο προτεινόμενος αλγόριθμος ξεπερνά σαφώς τους αλγόριθμους $ALAP_1$, $ALAP_2$ και $ALAP_3$, και τις δύο εκδοχές της μεθόδου οπισθοδρομικής διάδοσης του σφάλματος (Batch BP και On-line BP), αλλά η μέθοδος ABP έχει ελαφρώς υψηλότερο ποσοστό επιτυχίας. Αυτό αναμενόταν δεδομένου ότι, γενικά, οι αλγόριθμοι εκπαίδευσης ανά ομάδα προτύπων εισόδου είναι πολύ καλοί με προβλήματα που έχουν μικρά σύνολα προτύπων ή/και μικρές τοπολογίες δικτύων, αλλά είναι πιο αργοί από τις μεθόδους εκπαίδευσης ανά πρότυπο εισόδου.

Πίνακας 1: Αποτελέσματα από το πρόβλημα του αποκλειστικού ΕΙΤΕ

Αλγόριθμος	<i>Min</i>	μ	<i>Max</i>	Επιτυχία (%)
Batch BP	176	1693.9	3840	17%
Batch ABP	144	1430.4	3708	49%
On-line BP	72	724.2	2972	43%
$ALAP_1$	56	736.1	3900	38%
$ALAP_2$	40	816.9	3960	37%
$ALAP_3$	52	1000.5	3636	43%
Αλγόριθμος-1	44	680.2	3388	48%

3.2 Το πρόβλημα αναγνώρισης των αριθμών

Στο δεύτερο πείραμα, ένα δίκτυο με 64 νευρώνες εισόδου, 6 κρυφούς νευρώνες και 10 νευρώνες εξόδου (444 βάρη και 16 πολώσεις) εκπαιδεύεται για να αναγνωρίσει τους αριθμούς από το 0 έως το 9, εκτυπωμένους με πλάγιους χαρακτήρες (*italics*) και μέγεθος 8×8 εικονοκύτταρα (pixel) [11]. Το δίκτυο είναι βασισμένο σε νευρώνες με την λογιστική συνάρτηση ενεργοποίησης. Η συνθήκη τερματισμού για όλους τους αλγόριθμους είναι να έχουν μηδέν λάθος ταξινόμησης στο σύνολο εκπαίδευσης, μέσα σε 1000 υπολογισμούς της συνάρτησης σφάλματος. Λεπτομερή αποτελέσματα σχετικά με την απόδοση των αλγορίθμων εκπαίδευσης παρουσιάζονται στον Πίνακα 2.

Πίνακας 2: Αποτελέσματα από το πρόβλημα αναγνώρισης αριθμών

Αλγόριθμος	<i>Min</i>	μ	<i>Max</i>	Επιτυχία (%)
Batch BP	210	500.8	980	90%
Batch ABP	420	789.2	990	51%
On-line BP	230	507.7	950	99%
$ALAP_1$	130	475.5	990	90%
$ALAP_2$	190	433.6	860	90%
$ALAP_3$	210	486.3	990	96%
Αλγόριθμος-1	170	436.3	870	99%

Σχολιάζοντας τα αποτελέσματα βλέπουμε ότι η μέθοδος οπισθοδρομικής διάδοσης του σφάλματος για εκπαίδευση ανά πρότυπο εισόδου (On-line BP) είχε πολύ υψηλό ποσοστό επιτυχίας, αλλά οι μέθοδοι $ALAP_1$,

ALAP₂ και ALAP₃ είχαν γρηγορότερη σύγκλιση. Παρόλα αυτά, η προτεινόμενη μέθοδος και η μέθοδος οπισθοδρομικής διάδοσης του σφάλματος για εκπαίδευση ανά πρότυπο εισόδου είχαν σχεδόν τέλειο ποσοστό επιτυχίας (99%). Επιπλέον, ο Αλγόριθμος-1 επιδεικνύει πολύ γρήγορη σύγκλιση, δεδομένου ότι χρειάστηκε κατά μέσον όρο μόνο 436 παρουσιάσεις προτύπων εισόδου προκειμένου να εκπαιδεύσει το ΤΝΔ. Συνεπώς, η γενική απόδοση του Αλγόριθμου-1 κρίνεται πολύ ικανοποιητική.

3.3 Το πρόβλημα αναγνώρισης των κεφαλαίων γραμμάτων

Για το πρόβλημα πρόβλημα αναγνώρισης των κεφαλαίων γραμμάτων, 26 πίνακες με τα κεφαλαία γράμματα του αγγλικού αλφάβητου παρουσιάζονται σε ένα 35-30-26 ΤΝΔ (1830 βάρη, 56 πολώσεις). Κάθε γράμμα προσδιορίζεται από δυαδικές τιμές (0 ή 1) σε ένα πλέγμα (grid) του μεγέθους 5 × 7. Το ΤΝΔ βασίστηκε σε νευρώνες με λογιστικές συναρτήσεις ενεργοποίησης. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 3.

Για ακόμα μια φορά, η προτεινόμενη μέθοδος (Αλγόριθμος-1) είχε πολύ υψηλό ποσοστό επιτυχίας (96%) και ήταν γρηγορότερη από όλες τις άλλες μεθόδους που εξετάσαμε. Κατά μέσον όρο χρειάστηκε μόνο 749 παρουσιάσεις προτύπων προκειμένου να ολοκληρωθεί επιτυχώς η εκπαίδευση.

Πίνακας 3: Αποτελέσματα από το πρόβλημα αναγνώρισης κεφαλαίων γραμμάτων

Αλγόριθμος	<i>Min</i>	μ	<i>Max</i>	Επιτυχία (%)
Batch BP	4498	21375.9	41860	79%
Batch ABP	3588	3815.7	4212	98%
On-line BP	1404	1861.1	2418	87%
ALAP ₁	494	1519.4	2548	72%
ALAP ₁	338	756.6	1846	94%
ALAP ₁	338	754.5	2418	79%
Αλγόριθμος-1	364	749.6	1872	96%

4 Υβριδικές Μέθοδοι για την Επανεκπαίδευση ΤΝΔ

Σε αυτή την Ενότητα θα παρουσιάσουμε μια νέα υβριδική μέθοδο που βελτιώνει την γενίκευση των ΤΝΔ, σε προβλήματα που το σύνολο προτύπων μεταβάλλεται αργά με τον χρόνο. Αυτή η προσέγγιση συμπίπτει με τον τρόπο που οι γενετικοί και εξελικτικοί αλγόριθμοι επεκτείνονται και εξερευνούν χώρους που μεταβάλλονται αργά [2, 23, 28]. Η νέα μέθοδος αποτελείται από δύο Φάσεις [10, 17]. Στην Φάση 1, ο Αλγόριθμος-1 εφαρμόζεται για να εκπαιδεύσει το ΤΝΔ ανά πρότυπο εισόδου. Μετά το τέλος της εκπαίδευσης, το ΤΝΔ επανεκπαίδευεται ανά πρότυπο εισόδου χρησιμοποιώντας Διαφοροεξελικτικούς Αλγόριθμους (ΔΕΑ) [18-20].

Με τον τρόπο αυτό συνδυάζουμε την ταχύτητα της στοχαστικής μεθόδου της πιο απότομης καθόδου, που εφαρμόζεται για την εκπαίδευση ανά πρότυπο εισόδου, με ένα Εξελικτικό Αλγόριθμο, που αποτελεί μέθοδο ολικής βελτιστοποίησης. Με την ιδέα της επανεκπαίδευσης είμαστε σε θέση να συνεχίσουμε την εκπαίδευση ενός ήδη εκπαιδευμένου ΤΝΔ. Ο λόγος για να γίνει αυτό είναι ότι πιθανά νέα πρότυπα είναι διαθέσιμα και επιθυμούμε να εκπαιδεύσουμε το ΤΝΔ και με αυτά, χωρίς όμως να πρέπει να αρχίσουμε την διαδικασία της εκπαίδευσης από την αρχή.

Επισημαίνουμε τέλος ότι η διαδικασία της εκπαίδευσης και επανεκπαίδευσης ανά πρότυπο εισόδου είναι πολύ σημαντική για την επίλυση πολλών πραγματικών προβλημάτων, όπως για παράδειγμα ο έλεγχος της κατεύθυνσης ενός κινούμενου οχήματος ανάλογα με τις συνθήκες που επικρατούν στην πορεία του [4] και η αναγνώριση και εξαγωγή προτύπων από εικόνες που προέρχονται από συσκευές λήψης κάτω από μεταβαλλόμενες συνθήκες αντίληψης (φωτεινότητα, σκιές, συνθήκες φωτισμού και αντανάκλασεις) [3, 5, 13,

14, 30]. Ο Αλγόριθμος 2 περιγράφει την παραπάνω διαδικασία. Ο νέος αυτός αλγόριθμος δοκιμάστηκε σε δύο δύσκολα προβλήματα ταξινόμησης: το πρόβλημα ταξινόμησης υφής και το πρόβλημα της αναγνώρισης ανωμαλιών σε κολονοσκοπήσεις.

Αλγόριθμος 2: Γενικό μοντέλο του υβριδικού αλγόριθμου

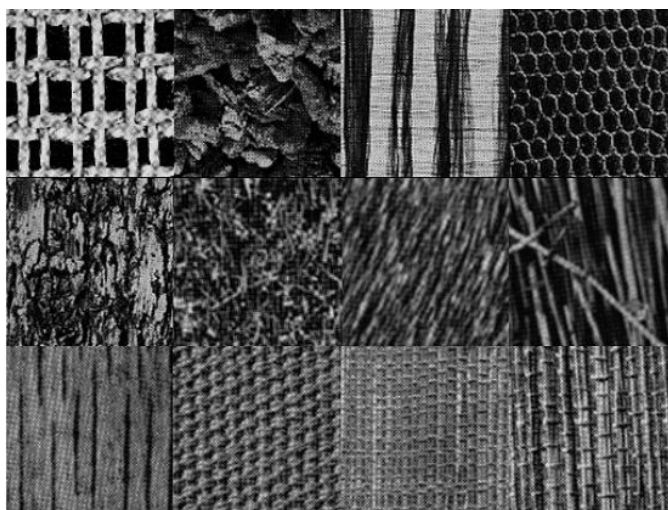
Φάση 1 - «Εκπαίδευση»	
Βήμα 0α:	Αρχικοποίησε w^0, η^0, γ_1 και γ_2 .
Βήμα 1α:	Repeat για κάθε πρότυπο p .
Βήμα 2α:	Υπολόγισε $E_p(w^k)$ και μετά $\nabla E_p(w^k)$.
Βήμα 3α:	Υπολόγισε τα νέα βάρη: $w^{k+1} = w^k - \eta^k \nabla E_p(w^k)$.
Βήμα 4α:	Υπολόγισε το ρυθμό εκπαίδευσης για το επόμενο πρότυπο ($p + 1$): $\eta^{k+1} = \eta^k + \gamma_1 \langle \nabla E_{p-1}(w^{k-1}), \nabla E_p(w^k) \rangle +$ $+ \gamma_2 \langle \nabla E_{p-2}(w^{k-2}), \nabla E_{p-1}(w^{k-1}) \rangle$.
Βήμα 5α:	Until Συνθήκη Τερματισμού.
Βήμα 6α:	Return τα τελικά βάρη w^{k+1} στη Φάση 2.
Φάση 2 - «Εξέλιξη»	
Βήμα 0β:	Αρχικοποίησε τον πληθυσμό στην γειτονιά του w^{k+1} .
Βήμα 1β:	Repeat για κάθε πρότυπο p .
Βήμα 2β:	For $i = 1$ to NP
Βήμα 3β:	MUTATION(w_i^k) \rightarrow Mutant_Vector.
Βήμα 4β:	CROSSOVER(Mutant_Vector) \rightarrow Trial_Vector.
Βήμα 5β:	If $E_p(\text{Trial_Vector}) \leq E_p(w_i^k)$, αποδοχή του Trial_Vector.
Βήμα 6β:	EndFor
Βήμα 7β:	Until Συνθήκη Τερματισμού.

4.1 Το πρόβλημα ταξινόμησης υφής

Στο πρόβλημα ταξινόμησης υφής ένα 16–8–12 TND εκπαιδεύεται να αναγνωρίζει 12 εικόνες διαφορετικής υφής (βλ. Σχήμα 1). Το δίκτυο βασίζεται σε νευρώνες με λογιστικές συναρτήσεις ενεργοποίησης και τα βάρη αρχικοποιήθηκαν με τυχαίους αριθμούς από το διάστημα $(-1, 1)$. Η συνθήκη τερματισμού της εκπαίδευσης ανά πρότυπο εισόδου ήταν να επιτύχουμε σφάλμα ταξινόμησης $CE \leq 3\%$. Τότε, ξεκινά η δεύτερη Φάση, για την επανεκπαίδευση ανά πρότυπο εισόδου με χρήση ΔΕΑ. Τελικά, η ικανότητα γενίκευσης του TND δοκιμάστηκε σε ένα σύνολο από 320 πρότυπα, που δεν συμμετείχαν στο σύνολο εκπαίδευσης. Το TND αναγνώρισε σωστά 304 από τα 320, δηλαδή είχε γενίκευση 95%. Στο ίδιο πρόβλημα, ο Αλγόριθμος-1 χωρίς την χρήση της επανεκπαίδευσης με ΔΕΑ, είχε ποσοστό 93%, ενώ άλλες μέθοδοι εκπαίδευσης ανά ομάδα προτύπων (π.χ. η μέθοδος BPVS [11]) έχουν ποσοστό περίπου 90%.

4.2 Το πρόβλημα αναγνώρισης ανωμαλιών σε κολονοσκοπήσεις

Η κολονοσκόπηση (ή κολοσκόπηση) είναι η εξέταση του παχέος εντέρου με εύκαμπτο ενδοσκόπιο και χαρακτηρίζεται ως εξέταση ελάχιστα εισβολής (minimally invasive procedure). Το ενδοσκόπιο είναι ένας εύκαμπτος λεπτός σωλήνας που στο μπροστινό άκρο του έχει προσαρμοσμένη μια συσκευή λήψης και μαγνητικής εγγραφής εικόνων (video camera), η οποία στέλνει το σήμα σε μία μικρή τηλεόραση. Το ενδοσκόπιο εισέρχεται στο σώμα του ασθενούς και εικόνες από το παχύ έντερο εμφανίζονται στην οθόνη. Ο γιατρός ελέγχει τη κατεύθυνση της συσκευής χρησιμοποιώντας διακόπτες και τροχούς.



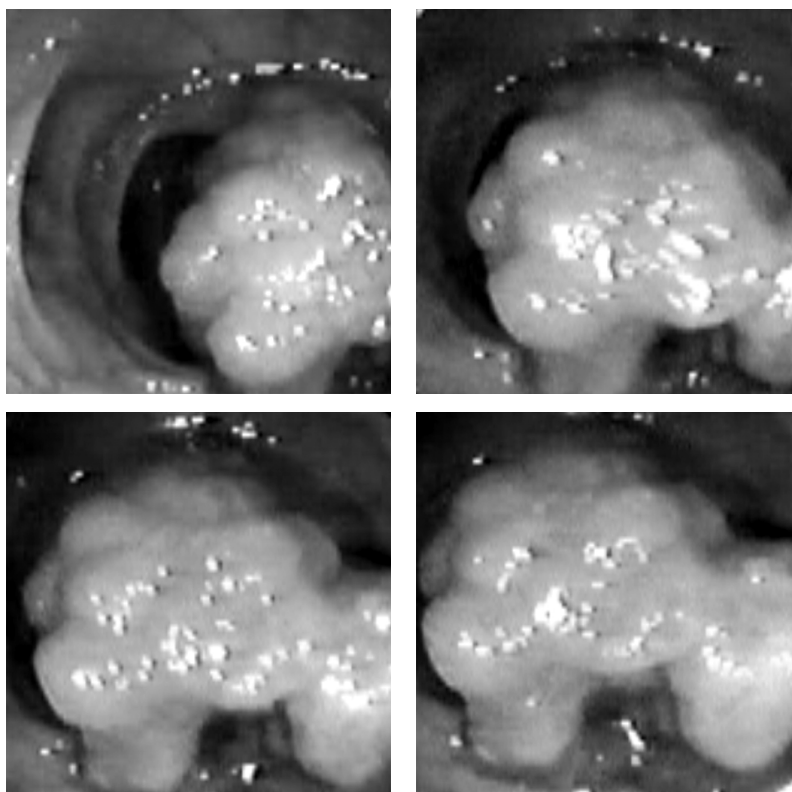
Σχήμα 1: Οι 12 εικόνες υφής που χρησιμοποιήθηκαν για την εξαγωγή των προτύπων εκπαίδευσης

Η χρήση ΤΝΔ για την αυτόματη ανακάλυψη κακοήθων όγκων σε εικόνες από ενδοσκόπια παρουσιάζει αρκετές δυσκολίες, καθώς οι εικόνες αυτές έχουν διαφορές στη φωτεινότητα, είναι από διαφορετική οπτική γωνία ανάλογα με τον γιατρό, και υπάρχουν διαφοροποιήσεις στη διάχυση του φωτός. Τέλος, από τη φύση του το πρόβλημα αυτό είναι ένα πρόβλημα που τα δεδομένα αλλάζουν κατά την διάρκεια της εξέτασης. Για να αντιμετωπίσουμε τα παραπάνω προβλήματα, προτείνουμε την εκπαίδευση ανά πρότυπο εισόδου και στην συνέχεια επανεκπαίδευση ανά πρότυπο εισόδου με χρήση ΔΕΑ. Η υβριδική μέθοδος φαίνεται να καταφέρνει να τροποποιεί κατάλληλα τα βάρη του ΤΝΔ, και κατά τη διάρκεια της εξέτασης, λαμβάνοντας υπόψη και τα δεδομένα του συνόλου εκπαίδευσης αλλά και τη γνώση από τα νέα πρότυπα.

Για το πρόβλημα αυτό χρησιμοποιήσαμε ένα 16-30-2 ΤΝΔ (540 βάρη και 32 πολώσεις), όπου οι νευρώνες του είχαν λογιστικές συναρτήσεις ενεργοποίησης. Το ΤΝΔ πρέπει να διαχωρίσει φυσιολογικές από «ύποπτες» περιοχές μιας εικόνας, δηλαδή να ξεχωρίσει περιοχές που περιέχουν τμήματα φυσιολογικού ιστού εντέρου από περιοχές που περιέχουν τμήματα όγκων. Χρησιμοποιήσαμε 4 σχεδόν διαδοχικές εικόνες από το ενδοσκόπιο (βλ. Σχήμα 2) και κάθε εικόνα χωρίστηκε σε περιοχές 16×16 εικονοκυττάρων. Συνολικά κάθε εικόνα χωρίστηκε σε περίπου 4000 περιοχές, από τις οποίες δημιουργήσαμε τα σύνολα εκπαίδευσης και ελέγχου. Μια πολύ σημαντική φάση της αναγνώρισης ανωμαλιών σε κολonosκοπήσεις είναι η επιλογή της διαδικασίας εξαγωγής προτύπων. Στα πειράματά μας χρησιμοποιήσαμε την μέθοδο εξαγωγής προτύπων που πρότεινε ο Haralick στην εργασία [6]. Περισσότερες τεχνικές λεπτομέρειες σχετικά με την εξαγωγή των προτύπων εκπαίδευσης βρίσκονται στην εργασία [7].

Έτσι δημιουργήσαμε 4 σύνολα εκπαίδευσης (ένα για κάθε εικόνα), που το καθένα αποτελείται από 300 τυχαία επιλεγμένα πρότυπα (από τα 4000 συνολικά πρότυπα για κάθε εικόνα). Στη συνέχεια εκπαιδεύσαμε το παραπάνω ΤΝΔ ανά πρότυπο εισόδου με τον Αλγόριθμο-1, χρησιμοποιώντας μόνο τα πρότυπα από την πρώτη εικόνα. Η ιδέα ήταν να δοκιμάσουμε το εκπαιδευμένο ΤΝΔ σε πρότυπα από τις επόμενες εικόνες, ώστε να διαπιστώσουμε την ικανότητα γενίκευσής του ειδικότερα όταν το σύνολο εκπαίδευσης μεταβάλλεται με τον χρόνο.

Η εκπαίδευση τερματίστηκε όταν το ΤΝΔ είχε σφάλμα ταξινόμησης στο σύνολο εκπαίδευσης, $CE \leq 3\%$. Πρέπει εδώ να σημειωθεί ότι η φάση αυτή της εκπαίδευσης ήταν πολύ γρήγορη: χρειάστηκαν μόνο 40 επαναλήψεις. Το εκπαιδευμένο ΤΝΔ επανεκπαιδεύεται ανά πρότυπο εισόδου με χρήση ΔΕΑ. Στη φάση αυτή χρησιμοποιούμε σαν σύνολο εκπαίδευσης και τα 4 αρχικά σύνολα εκπαίδευσης, δηλαδή το νέο σύνολο εκπαίδευσης αποτελείται από 1200 πρότυπα. Έτσι μπορούμε να δούμε την συμπεριφορά του ΤΝΔ και των αλγόριθμων εκπαίδευσης όταν τα δεδομένα μεταβάλλονται με τον χρόνο, αφού οι άλλες 3 εικόνες είναι αυτές



Σχήμα 2: Διαδοχικές εικόνες από ενδοσκόπιο

που λάβαμε από το ενδοσκόπιο μετά την πρώτη. Επιτρέπουμε στον ΔΕΑ να κάνει μόνο δύο επαναλήψεις με κάθε πρότυπο, έτσι ώστε να αποφύγουμε την «καταστροφική παρέμβαση» των προτύπων από τις 3 νέες εικόνες με τα ήδη γνωστά πρότυπα της πρώτης εικόνας. Για τον έλεγχο του τελικού ΤΝΔ χρησιμοποιήσαμε όλα τα διαθέσιμα πρότυπα, δηλαδή 4000 πρότυπα από κάθε μια από τις 4 εικόνες. Τα αποτελέσματα γενίκευσης παρουσιάζονται στον Πίνακα 4.

Πίνακας 4: Αποτελέσματα από το πρόβλημα αναγνώρισης ανωμαλιών σε κολonosκοπήσεις

	Χωρίς Εξέλιξη (Φάση 1 μόνο)	Με Εξέλιξη (Νέος Υβριδικός Αλγόριθμος)
Εικόνα 1	83.77%	91.91%
Εικόνα 2	77.18%	83.57%
Εικόνα 3	82.84%	93.09%
Εικόνα 4	87.60%	89.24%

Στον πίνακα αυτό βλέπουμε π.χ. ότι το ΤΝΔ που εκπαιδεύσαμε με πρότυπα από την πρώτη εικόνα, πριν την επανεκπαίδευση, είχε 83.77% επιτυχία αναγνώρισης, ενώ μετά την επανεκπαίδευση το ποσοστό αναγνώρισης για την πρώτη εικόνα ανέβηκε στο 91.91%. Είναι προφανές ότι το επανεκπαιδευμένο ΤΝΔ έχει μεγαλύτερο ποσοστό επιτυχίας και στις 4 εικόνες που εξετάσαμε. Μετά το τέλος της εκπαίδευσης, το ΤΝΔ είναι σε θέση να αναγνωρίζει αυτόματα «ύποπτα» τμήματα στις εικόνες που παρέχει το ενδοσκόπιο. Τέλος, αξίζει να σημειώσουμε ότι τα ποσοστά γενίκευσης που επιτύχαμε με την μέθοδο που περιγράφηκε είναι ανάλογα με τα καλύτερα ποσοστά που έχουν βρεθεί σε ΤΝΔ που εκπαιδεύτηκαν με μεθόδους εκπαίδευσης

ανά ομάδα προτύπων εισόδου [7].

5 Συμπεράσματα

Σε αυτή την εργασία, προτάθηκε ένας νέος αλγόριθμος εκπαίδευσης ανά πρότυπο εισόδου για ΤΝΔ. Οι αλγόριθμοι εκπαίδευσης ανά πρότυπο εισόδου είναι σε θέση να εκπαιδεύσουν αποδοτικά μεγάλα ΤΝΔ με βάρη πλήθους της τάξης των χιλιάδων και ταιριάζουν καλύτερα για την εκπαίδευση μεγάλων συνόλων προτύπων, ή συνόλων που περιέχουν επαναλαμβανόμενα ή αργά μεταβαλλόμενα με τον χρόνο πρότυπα. Επίσης παρέχουν την δυνατότητα της συνεχούς εκπαίδευσης με νέα δεδομένα που δεν ήταν διαθέσιμα την χρονική στιγμή της πρώτης εκπαίδευσης του ΤΝΔ.

Τα αποτελέσματα της προσομοίωσης δείχνουν ότι ο προτεινόμενος αλγόριθμος παρέχει γρήγορη και σταθερή εκπαίδευση σε σύγκριση με άλλες μεθόδους εκπαίδευσης ανά πρότυπο εισόδου, καθώς επίσης και μεθόδους εκπαίδευσης ανά ομάδα προτύπων εισόδου, και επομένως παρέχει μεγαλύτερη πιθανότητα επιτυχημένης και γρήγορης εκπαίδευσης για προβλήματα του πραγματικού κόσμου.

Επίσης, παρουσιάσαμε μια νέα υβριδική μέθοδο και εξετάσαμε την απόδοσή της σε δύο πραγματικές εφαρμογές. Τα αποτελέσματα γενίκευσης των ΤΝΔ που αρχικά εκπαιδεύτηκαν από την προτεινόμενη μέθοδο εκπαίδευσης ανά πρότυπο εισόδου (Φάση 1) και στη συνέχεια εκπαιδεύτηκαν ξανά με ένα εξελικτικό αλγόριθμο (Φάση 2) είναι ικανοποιητικά και ανάλογα με τα καλύτερα αποτελέσματα μεθόδων που εκπαιδεύουν ανά ομάδα προτύπων εισόδου. Η προτεινόμενη υβριδική μέθοδος ανταποκρίθηκε με επιτυχία στο μη στατικό πρόβλημα της αναγνώρισης ανωμαλιών σε κολονοσκοπήσεις και αποδείχθηκε σθεναρή και προβλέψιμη. Τέλος, αξίζει να αναφερθεί ότι δεν παρατηρήθηκε το φαινόμενο της «καταστροφικής παρέμβασης» μεταξύ των προτύπων που προέρχονταν από διαφορετικές εικόνες.

Αναφορές

- [1] L.B. Almeida, T. Langlois, J.D. Amaral, and A. Plankhov, “Parameter adaptation in stochastic optimization”, In *On-line Learning in Neural Networks*, edited by D. Saad, Cambridge University Press, 111–134, 1998.
- [2] P. Angeline, “Tracking extrema in dynamic environments”, In the proceedings of the *Sixth Annual conference on Evolutionary Programming VI*, Springer, 335–345, 1997.
- [3] S.W. Baik and P. Pachowicz, “Adaptive object recognition based on the radial basis function paradigm”, In the proceedings of the *IEEE International Joint Conference on Neural Networks (IJCNN'99)*, CD-ROM Proceedings, Paper No.215, Session 9.4, Washington, U.S.A., 1999.
- [4] S. Baluja, “Evolution of an artificial neural network based autonomous land vehicle controller”, *IEEE Transactions on System, Man and Cybernetics-Part B*, vol. 26, 450–463, 1996.
- [5] A.D. Doulamis, N.D. Doulamis, and S.D. Kollias, “On-line retrainable neural networks: improving the performance of neural networks in image analysis problems”, *IEEE Transactions on Neural Networks*, vol. 11, 137–155, 2000.
- [6] R. Haralick, K. Shanmugan, and I. Dinstein, “Textural features for image classification”, *IEEE Transactions on System, Man and Cybernetics*, vol. 3, 610–621, 1973.
- [7] S. Karkanis, G.D. Magoulas, and N. Theofanous, “Image recognition and neuronal networks: intelligent systems for the improvement of imaging information”, In the proceedings of the *Minimally Invasive Therapy and Allied Technologies*, 225–230, 2000.

- [8] G.D. Magoulas, V.P. Plagianakos, and M.N. Vrahatis, "Adaptive stepsize algorithms for on-line training of neural networks", *Nonlinear Analysis, Theory, Methods and Applications*, vol. 47, 3425–3430, 2001.
- [9] G.D. Magoulas, V.P. Plagianakos, and M.N. Vrahatis, "Hybrid methods using evolutionary algorithms for on-line training", In the proceedings of the *INNS-IEEE International Joint Conference on Neural Networks (IJCNN'2001)*, Washington D.C., U.S.A., 2001.
- [10] G.D. Magoulas, V.P. Plagianakos, and M.N. Vrahatis, "Improved neural network-based interpretation of colonoscopy images through on-line learning and evolution", In the proceedings of the *EUNITE 2001 Conference*, Tenerife, Spain, 2001.
- [11] G.D. Magoulas, M.N. Vrahatis, and G.S. Androulakis, "Effective back-propagation with variable stepsize", *Neural Networks*, vol. 10, 69–82, 1997.
- [12] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural network by choosing initial values of the adaptive weights", In the proceedings of the *IEEE First International Joint Conference on Neural Networks*, vol. 3, 21–26, 1990.
- [13] P.W. Pachowicz and S.W. Baik, "Adaptive rbf classifier for object recognition in images sequences", In the proceedings of the *IEEE International Joint Conference on Neural Networks (IJCNN'2000)*, vol. VI-600, Como, Italy, 2000.
- [14] N.G. Panagiotidis, D. Kalogeras, S.D. Kollias, and A. Stafylopatis, "Neural network-assisted effective lossy compression of medical images", *Proc. IEEE*, vol. 84, 1474–1487, 1996.
- [15] V.P. Plagianakos, G.D. Magoulas, and M.N. Vrahatis, "Global learning rate adaptation in on-line neural network training", In the proceedings of the *Second International ICSC Symposium on Neural Computation (NC'2000)*, Berlin, Germany, 2000.
- [16] V.P. Plagianakos, G.D. Magoulas, and M.N. Vrahatis, "Learning rate adaptation in stochastic gradient descent", In *Advances in Convex Analysis and Global Optimization, Honoring the memory of C. Caratheodory (1873-1950)*, edited by N. Hadjisavvas and P. Pardalos, chap. 27, Kluwer Academic Publishers, Dordrecht, The Netherlands, 433–444, 2001.
- [17] V.P. Plagianakos, G.D. Magoulas, and M.N. Vrahatis, "Tumor detection in colonoscopic images using hybrid methods for on-line neural network training", In the proceedings of the *Neural Networks and Expert Systems in Medicine and HealthCare*, Milos Island, Greece, 2001.
- [18] V.P. Plagianakos, D.G. Sotiropoulos, and M.N. Vrahatis, "Integer weight training by differential evolution algorithms", In *Recent Advances in circuits and systems*, edited by N. Mastorakis, World Scientific, 1998.
- [19] V.P. Plagianakos and M.N. Vrahatis, "Neural network training with constrained integer weights", In the proceedings of the *Congress of Evolutionary Computation (CEC'99)*, edited by M. S. X. Y. P.J. Angeline, Z. Michalewicz and A. Zalzala, IEEE Press, Washington D.C., U.S.A., 2007–2013, 1999.
- [20] V.P. Plagianakos and M.N. Vrahatis, "Training neural networks with 3-bit integer weights", In the proceedings of the *Genetic and Evolutionary Computation Conference (GECCO'99)*, edited by W. Banzhaf, J. Daida, A. Eiben, M. Garzon, V. Honavar, M. Jakiela, and R. Smith, Morgan Kaufmann, Orlando, U.S.A., 910–915, 1999.
- [21] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning internal representations by error propagation", In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, edited by D. Rumelhart and J. McClelland, MIT Press, Cambridge, Massachusetts, 318–362, 1986.
- [22] D. Saad, *On-line learning in neural networks*, Cambridge University Press, 1998.

- [23] R. Salomon and P. Eggenberger, "Adaptation on the evolutionary time scale: a working hypothesis and basic experiments", In *Third European Conference on Artificial Evolution (AE'97), Lecture Notes in Computer Science vol. 1363*, Springer, Nimes, France, 1998.
- [24] N.N. Schraudolph, "Online local gain adaptation for multi-layer perceptrons", Technical Report IDSIA-09-98, IDSIA, Lugano, Switzerland, 1998.
- [25] N.N. Schraudolph, "Local gain adaptation in stochastic gradient descend", Technical Report IDSIA-09-99, IDSIA, Lugano, Switzerland, 1999.
- [26] R.S. Sutton, "Adapting bias by gradient descent: an incremental version of delta-bar-delta", In the proceedings of the *Tenth National Conference on Artificial Intelligence*, MIT Press, 171-176, 1992.
- [27] R.S. Sutton and S.D. Whitehead, "Online learning with random representations", In the proceedings of the *Tenth International Conference on Machine Learning*, Morgan Kaufmann, 314-321, 1993.
- [28] F. Vavak and T.C. Fogarty, "A comparative study of steady state and generational genetic algorithms", In *Evolutionary Computing: AISB Workshop, Lecture Notes in Computer Science vol. 1143*, Springer, 1996.
- [29] T.P. Vogl, J.K. Mangis, J.K. Rigler, W.T. Zink, and D.L. Alkon, "Accelerating the convergence of the back-propagation method", *Biological Cybernetics*, vol. 59, 257-263, 1988.
- [30] G.G. Wilkenson, "Open questions in neurocomputing for earth observation", In the proceedings of the *First COMPARES Workshop*, York, U.K., 1996.