Αναφορά Απαλλακτικής Εργασίας - Νευρωνικά Δίκτυα - Binary Classification

**Μάθημα: Νευρωνικά Δίκτυα**  
**Ακαδημαϊκό Έτος: 2024 – 2025**  
**Τμήμα Εφαρμοσμένης Πληροφορικής / ΕΤΥ**

**Ημερομηνία:**

**Φοιτητής: Σταύρος Σαμαράς**

**Πανεπιστήμιο Μακεδονίας**

**Link κώδικες Εργασίας σε Google Colab:**

***https://colab.research.google.com/drive/1zZw05umuTvOp9V9yxTAgH4DTCoFeVXOq#scrollTo=5NHYOPiDhP0J***

## ΠΕΡΙΕΧΟΜΕΝΑ

1. **Ιδέα / Σκοπός**  
    1.1 Περιγραφή προβλήματος  
    1.2 Προσέγγιση επίλυσης  
    1.3 Περιγραφή δεδομένων
2. **Σχετική Βιβλιογραφία**
3. **Προτεινόμενη Υλοποίηση και Περιγραφή Μοντέλων**  
    3.1 Custom CNN  
    3.2 VGG16  
    3.3 ResNet50
4. **Πειραματικά Αποτελέσματα**  
    4.1 Περιγραφή Dataset  
    4.2 Προεπεξεργασία δεδομένων  
    4.3 Εκπαίδευση και Αξιολόγηση  
    4.4 Μετρικές Απόδοσης και Γραφήματα

4.5 Fine Tuning and Hyperparameters  
 4.6 Σύγκριση Μοντέλων  
 4.7 Στατιστικός Έλεγχος

1. **Συμπεράσματα**
2. **Βιβλιογραφία**

### Ενότητα 1: Ιδέα / Σκοπός

**1.1 Περιγραφή προβλήματος**

Η εργασία στοχεύει στην έγκαιρη ανίχνευση καρκίνου του μαστού μέσω επεξεργασίας υπερηχογραφικών εικόνων και ταξινόμησής τους σε καλοήθεις (benign) ή κακοήθεις (malignant) περιπτώσεις. Ο στόχος είναι η ανάπτυξη ενός αξιόπιστου μοντέλου δυαδικής ταξινόμησης, το οποίο μπορεί να υποστηρίξει διαγνωστικές αποφάσεις.

### 1.2 Προσέγγιση επίλυσης

Το πρόβλημα προσεγγίζεται ως **binary classification** με τη βοήθεια τεχνικών **Deep Learning και Transfer Learning**. Εκπαιδεύτηκαν και αξιολογήθηκαν διαφορετικά μοντέλα συνελικτικών νευρωνικών δικτύων (**CNNs**), τόσο από το μηδέν (ustom CNN) όσο και προκαταρτισμένα (VGG16, ResNet50) μέσω T**ransfer Learning**.

### 1.3 Περιγραφή δεδομένων

Η μελέτη βασίζεται στο δημόσιο dataset "[Breast Ultrasound Images Dataset](https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset)" από το Kaggle. Το dataset περιλαμβάνει εικόνες υπερήχων που έχουν επισημανθεί με κατηγορίες *benign*, *malignant* και *normal*. Στην εργασία χρησιμοποιούνται οι πρώτες δύο κατηγορίες για τον σκοπό της δυαδικής ταξινόμησης.

### Ενότητα 2. Σχετική Βιβλιογραφία

Παρακάτω παρουσιάζονται επιλεγμένες ερευνητικές εργασίες σχετικές με την ανίχνευση καρκίνου του μαστού μέσω υπερηχογραφημένων εικόνων , καθώς και με τη χρήση βαθιών CNN για binary classification απο όπου πήρα έμπνευση.

1. **Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020).** *Dataset of breast ultrasound images.* Data in Brief, 28, 104863.  
    ➤ Η εργασία παρουσιάζει το dataset που χρησιμοποιείται στην παρούσα μελέτη και περιγράφει τον τρόπο συλλογής και κατηγοριοποίησης των εικόνων.
2. **Abbas, Q., et al. (2021).** *Classification of breast cancer histology images using transfer learning.* Computers in Biology and Medicine, 132, 104320.  
    ➤ Εφαρμόζεται transfer learning σε medical imaging προβλήματα, παρόμοια με το σενάριο της παρούσας εργασίας.
3. **Muduli, D., Dash, R., & Majhi, B. (2022).** *Automated diagnosis of breast cancer using deep learning and CNN ensembles.* Biomedical Signal Processing and Control, 71, 103148.  
    ➤ Η μελέτη αξιολογεί ensemble μεθόδους βασισμένες σε CNNs για την ταξινόμηση καρκίνου του μαστού.
4. **Houssein, E.H., et al. (2021).** *Breast cancer classification from ultrasound images using Transfer Learning and Chaotic Sparrow Search Algorithm.* Applied Soft Computing, 113, 107918.  
    ➤ Παρουσιάζει συνδυασμό transfer learning και optimization για βελτίωση ακρίβειας.
5. **Arevalo, J., et al. (2016).** *Representation learning for mammography mass lesion classification with convolutional neural networks.* Computer Methods and Programs in Biomedicine, 127, 248–257.  
    ➤ Αξιολογείται η απόδοση των CNNs σε προβλήματα binary classification σε ιατρικές εικόνες.

### 3. Υλοποίηση και Περιγραφή Επιμέρους Μοντέλων

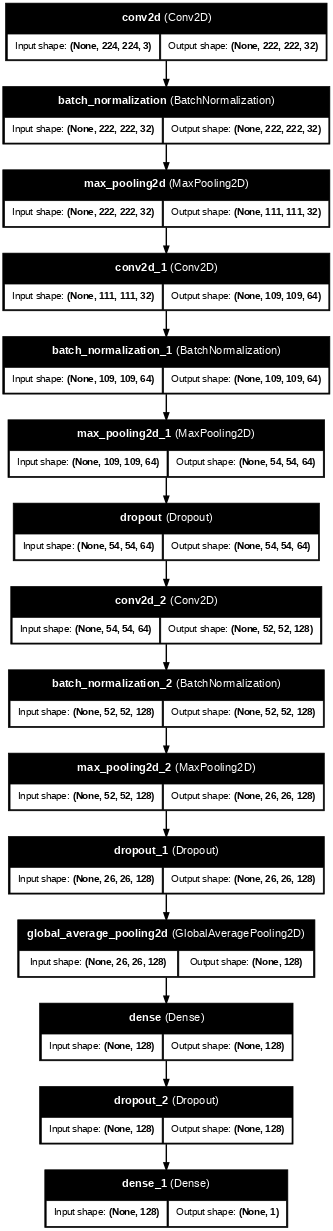
Η υλοποίηση του συστήματος ταξινόμησης βασίζεται σε τρία διαφορετικά μοντέλα: **Custom CNN**, **VGG16** και **ResNet50**. Κάθε μοντέλο έχει σχεδιαστεί με σκοπό να αντιμετωπίσει το πρόβλημα του εντοπισμού καρκίνου του μαστού μέσω υπερηχογραφικών εικόνων, επιδιώκοντας τόσο την επίτευξη υψηλής ακρίβειας(Accuracy) αλλά και χαμηλό Loss στην ταξινόμηση των εικόνων σε δύο κατηγορίες: **Benign** και **Malignant**.

#### 3.1 Custom CNN

Το πρώτο μοντέλο που χρησιμοποιείται είναι ένα **Custom CNN**. Η αρχιτεκτονική του περιλαμβάνει τρία **συνελικτικά μπλοκ (convolutional blocks)** που ακολουθούνται από **global average pooling**, καθώς και πλήρως συνδεδεμένα επίπεδα (fully connected layers). Κάθε συνελικτικό μπλοκ περιλαμβάνει:

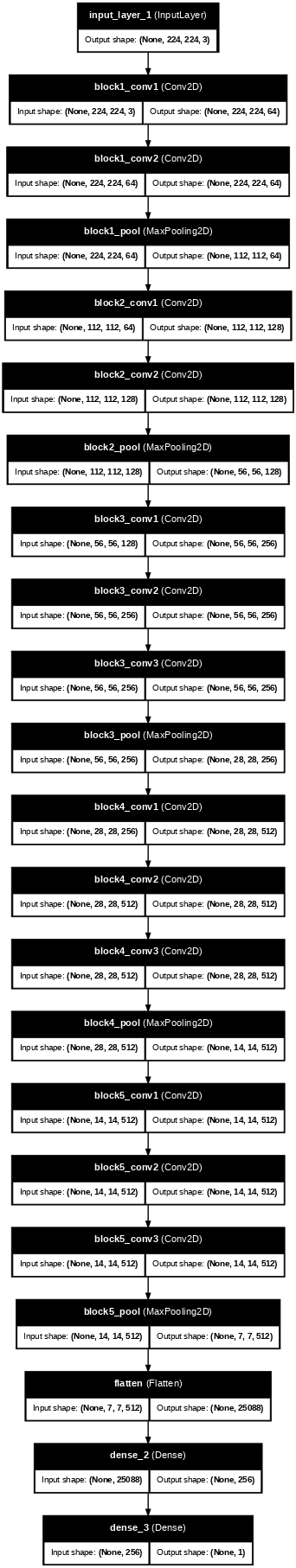
* **Συνελικτικές στρώσεις (Conv2D)** με διαφορετικό αριθμό φίλτρων (32, 64, 128).
* **Κανονικοποίηση παρτίδας (BatchNormalization)** για να επιταχυνθεί η εκπαίδευση.
* **MaxPooling2D** για τη μείωση των διαστάσεων.
* **Dropout** για την αποφυγή υπερπροσαρμογής.

Η έξοδος του μοντέλου είναι **δυαδική ταξινόμηση** (binary classification) μέσω της στρώσης εξόδου με **sigmoid activation**.



#### 3.2 VGG16

Το δεύτερο μοντέλο που χρησιμοποιείται είναι το **VGG16**, το οποίο έχει προεκπαιδευτεί (pre-trained) στο σύνολο δεδομένων **ImageNet** και προσαρμόζεται στην παρούσα εργασία μέσω της τεχνικής **transfer learning**. Στο μοντέλο, τοποθετούνται επιπλέον επίπεδα **Flatten**, **Dense** και **Sigmoid**, ενώ τα επίπεδα του VGG16 παραμένουν **παγωμένα (frozen)** κατά τη διάρκεια της εκπαίδευσης για να αποφευχθεί η υπερπροσαρμογή.



#### 3.3 ResNet50

Το τρίτο μοντέλο που χρησιμοποιείται είναι το **ResNet50**, επίσης προεκπαιδευμένο στο **ImageNet** και προσαρμοσμένο για το πρόβλημα της ταξινόμησης καρκίνου του μαστού μέσω υπερηχογραφικών εικόνων. Το ResNet50 χρησιμοποιεί **residual connections**, που το καθιστούν ιδιαίτερα αποτελεσματικό στην εκπαίδευση βαθιών δικτύων. Όπως και στο VGG16, τα επίπεδα του ResNet50 παραμένουν **παγωμένα** κατά την εκπαίδευση.

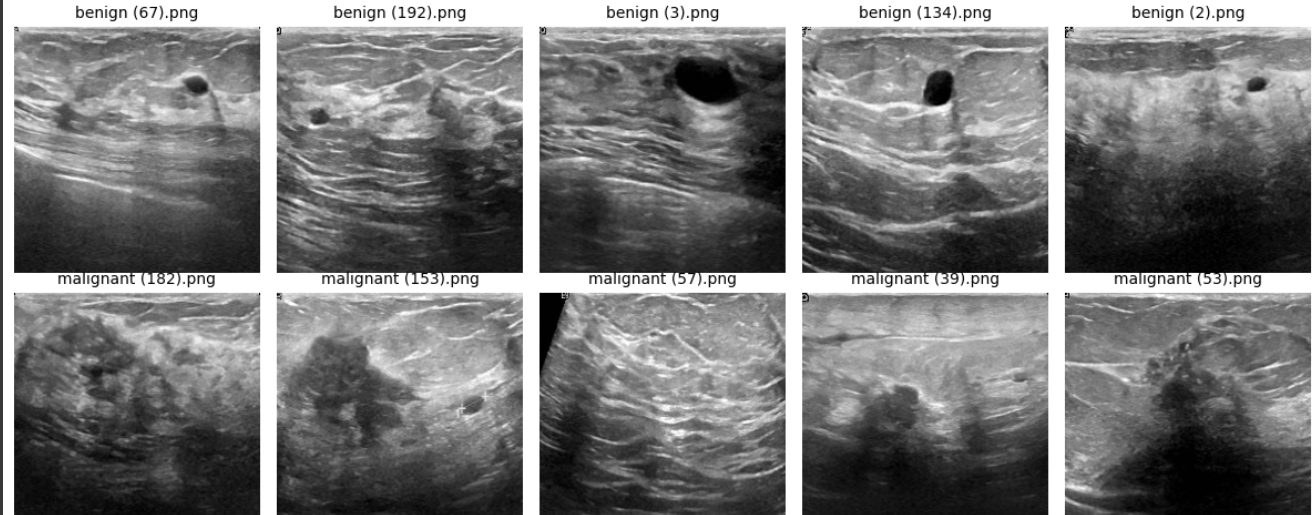
(TO FLOWCHART ΤΟΥ ΜΟΝΤΕΛΟΥ RESNET ΔΙΑΤΙΘΕΤΑΙ ΣΤΟ GOOGLE COLAB ΣΤΟ LINK ΠΟΥ ΕΧΩ ΠΑΡΑΔΩΣΕΙ ΣΤΗΝ ΑΡΧΗ ΤΗΣ ΑΝΑΦΟΡΑΣ ΚΑΘΩΣ ΕΙΝΑΙ ΠΕΡΙΠΛΟΚΟ ΜΟΝΤΕΛΟ ΜΕ ΠΟΛΛΑ LAYERS ΚΑΙ ΔΕΝ ΘΑ ΧΩΡΟΥΣΕ ΣΤΗΝ ΣΕΛΙΔΑ)

### 4 Πειραματικά Αποτελέσματα

#### 4.1 Περιγραφή Dataset

Το σύνολο δεδομένων που χρησιμοποιείται είναι το **"Breast Ultrasound Images Dataset"**, από την πλατφόρμα **Kaggle**. Περιλαμβάνει υπερηχογραφικές εικόνες από περιπτώσεις καλοήθους (benign) και κακοήθους (malignant) καρκίνου του μαστού.

Ενδεικτικά Εμφανίζω κάποιες εικόνες απο κάθε κλάση:



Επιλέχθηκαν **200 εικόνες από κάθε κατηγορία**, ώστε να διασφαλιστεί ισορροπία μεταξύ των δύο τάξεων (balanced dataset). Έτσι, το τελικό σύνολο περιλαμβάνει:

* **Benign:** 200 εικόνες
* **Malignant:** 200 εικόνες
* **Σύνολο:** 400 εικόνες

Το επιβεβαιώνω και με κωδικα:

**Total images loaded: 410 Class distribution: [205 205]**

**Train set: 328 images (80.0%)**

**Validation set: 41 images (10.0%)**

**Test set: 41 images (10.0%)**

Η ισορροπημένη κατανομή εξυπηρετεί την εκπαίδευση των μοντέλων χωρίς προκαταλήψεις υπέρ της πολυπληθέστερης τάξης. Οι εικόνες μετατράπηκαν σε διαστάσεις **224x224 pixels** και 3 κανάλια (RGB), ώστε να είναι συμβατές με τις απαιτήσεις των μοντέλων που χρησιμοποιούνται (Custom CNN, VGG16, ResNet50).

### 4.2 Προεπεξεργασία Δεδομένων

Αρχικά αφαιρέθηκαν εικόνες με σχεδιασμένα σημάδια (π.χ. κύκλοι/σημειώσεις), ώστε να μην επηρεάζεται η εκπαίδευση από μη ιατρικά στοιχεία. Οι καθαρές εικόνες ταξινομήθηκαν σε δύο φακέλους (benign/, malignant/) και συμπιέστηκαν σε αρχείο ZIP για εύκολη μεταφόρτωση στο Google Colab.

Κατά την προεπεξεργασία εφαρμόστηκαν:

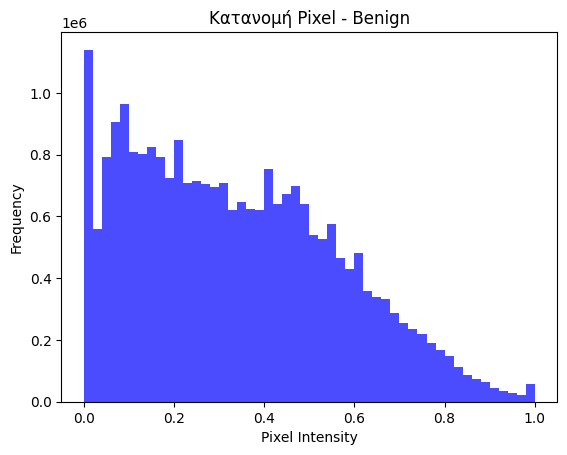
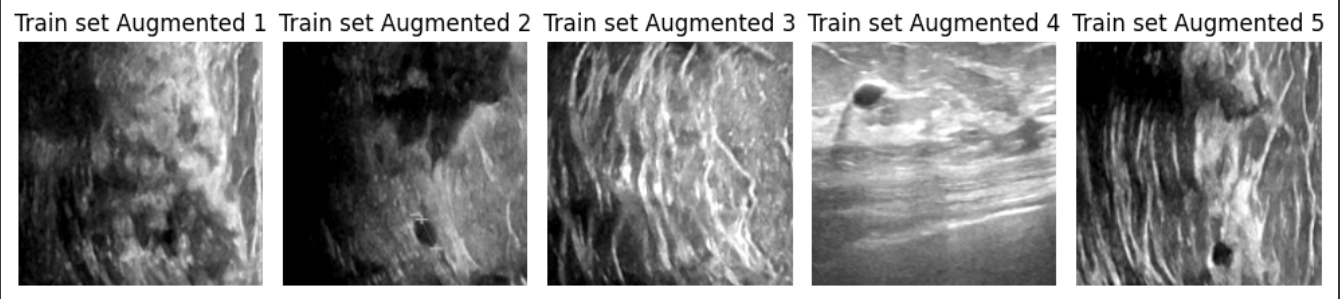
* **Αλλαγή μεγέθους σε 224x224**
* **Κανονικοποίηση pixel values** στο [*0,1*]
* **Augmentation**: flip, brightness, contrast, zoom,rotation

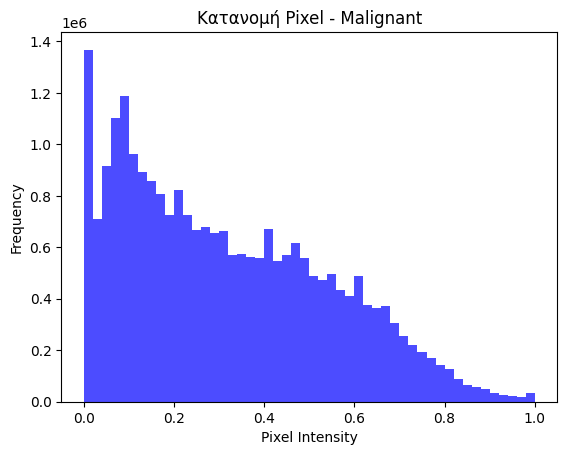
Για καλύτερη κατανόηση των δεδομένων δημιουργήθηκαν:

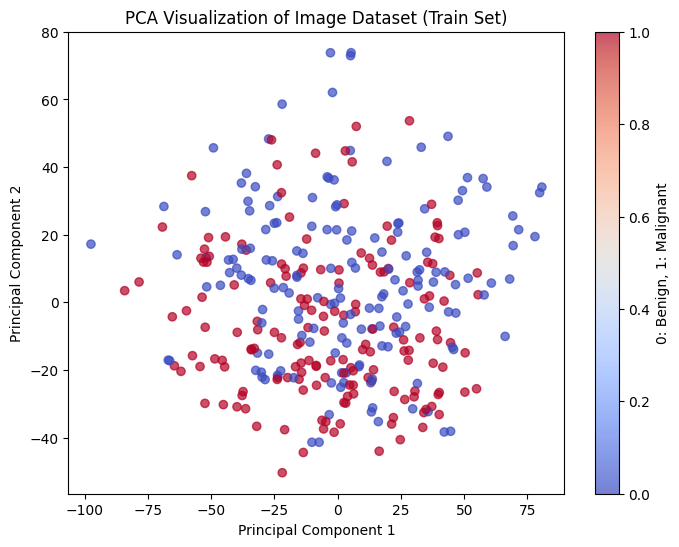
* **Heatmap μέσης εικόνας** για κάθε κατηγορία
* **Ιστογράμματα pixel values**
* **PCA Visualization** σε 2D χώρο

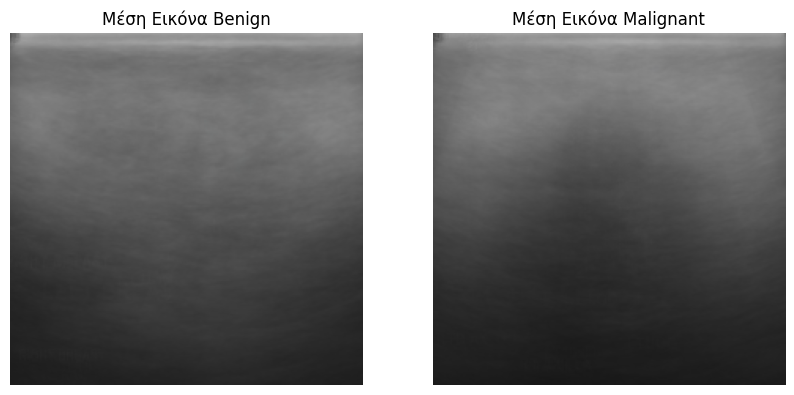
Τα δεδομένα χωρίστηκαν σε **training (80%)**, **validation (10%)** και **test (10%)**, με σταθερό random seed.

Ακολουθούν εικόνες για κάθε τεχνική Data PreProcessing , Exploration που χρησιμποιήθηκε:









### 4.3 Εκπαίδευση και Αξιολόγηση

Αρχικά Εκτελούμε Kfold(stratified) για να εντοπίσουμε πόσο καλά θα γενικεύουν τα μοντέλα σε άγνωστα δεδομένα.

Με fold=3(καθως εχω μεγάλους χρόνους εκτέλεσης), έχουμε:

**CNN:**

Final Mean Train Accuracy: 0.5000

Final Mean Validation Accuracy: 0.5001

**Vgg16:**

Final Mean Train Accuracy: 0.9963

Final Mean Validation Accuracy: 0.9659

**RESNET50:**

Final Mean Train Accuracy: 0.8439

Final Mean Validation Accuracy: 0.8513

Οπότε καταλαβαίνουμε την γενίκευση του κάθε μοντέλου.

Η εκπαίδευση πραγματοποιήθηκε σε τρία μοντέλα:

* **Custom CNN**
* **VGG16** (με transfer learning)
* **ResNet50** (με transfer learning)

Όλα τα μοντέλα εκπαιδεύτηκαν με τα ίδια δεδομένα για δίκαιη σύγκριση. Το dataset χωρίστηκε σε **training, validation και test set**, και εφαρμόστηκε **5-fold cross-validation**.

#### Ρυθμίσεις Εκπαίδευσης:

* **Batch size:** 8
* **Epochs:** 40
* **Optimizer:** Adam (lr=0.0001)
* **Loss function:** Binary crossentropy
* **EarlyStopping**: με patience=5 και restore\_best\_weights=True
* **ReduceLROnPlateau**: για μείωση του learning rate αν δεν βελτιώνεται η val\_loss

#### Διαδικασία:

* Για κάθε μοντέλο χρησιμοποιήθηκε fit()
* Η αξιολόγηση έγινε στο test set με τις μετρικές: **accuracy, loss,** αλλά και **precision, recall, F1-score,** καθώς και **confusion matrix.**
* Τα καλύτερα weights αποθηκεύτηκαν βάσει της χαμηλότερης val\_loss.

### 4.4 Μετρικές Απόδοσης και Γραφήματα

Για την αξιολόγηση των μοντέλων χρησιμοποιήθηκαν οι εξής μετρικές:

* **Accuracy**
* **Precision**
* **Recall**
* **F1-score**
* **Confusion Matrix**

Τα αποτελέσματα υπολογίστηκαν στο **test set**, ενώ παρουσιάζονται επίσης και για κάθε fold του **k-fold cross-validation**. Για κάθε μοντέλο αποθηκεύτηκαν τα **best weights** με χρήση του **EarlyStopping**, με παρακολούθηση της val\_loss.

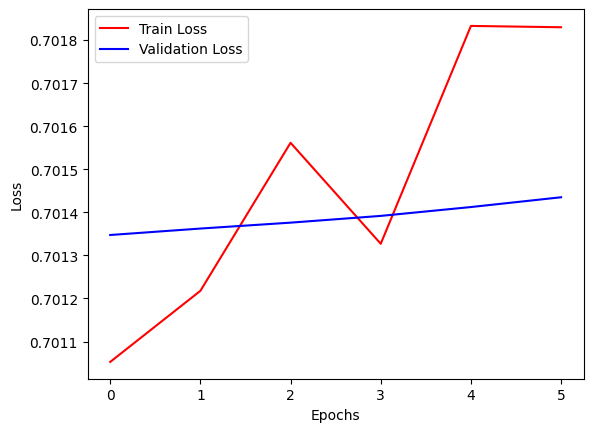
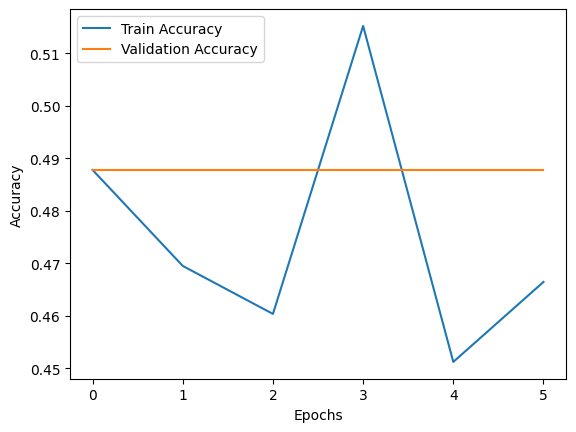
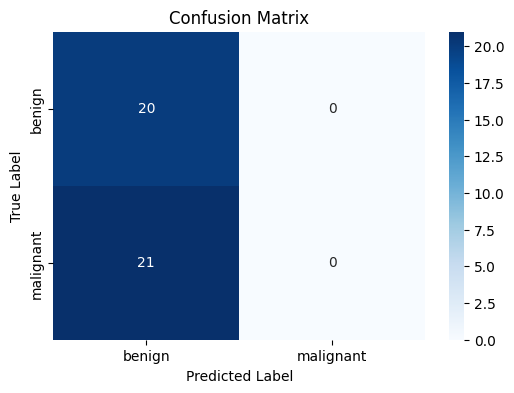
Ενδεικτικά γραφήματα που περιλαμβάνονται:

* **Loss curves (train/val)** για όλα τα μοντέλα
* **Accuracy curves**
* **Confusion Matrix** οπτικοποιημένη ως heatmap
* **Classification Report** με precision/recall/F1 για κάθε κατηγορία

Τα αποτελέσματα δείχνουν σαφώς τη διαφορά απόδοσης μεταξύ των μοντέλων, με το ResNet50 να παρουσιάζει την καλύτερη συνολική επίδοση.

Ακολουθουν Εικόνες Και Μετρικές Απόδοσης για περισσότερη κατανόηση:

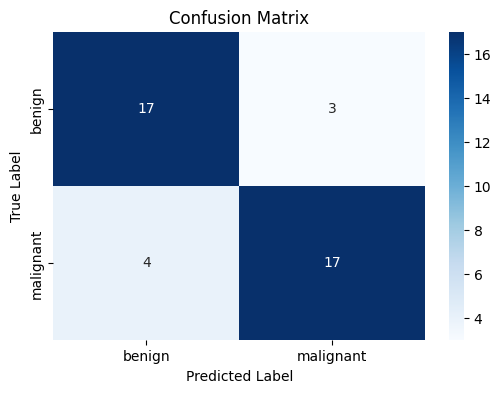
Initial CNN:

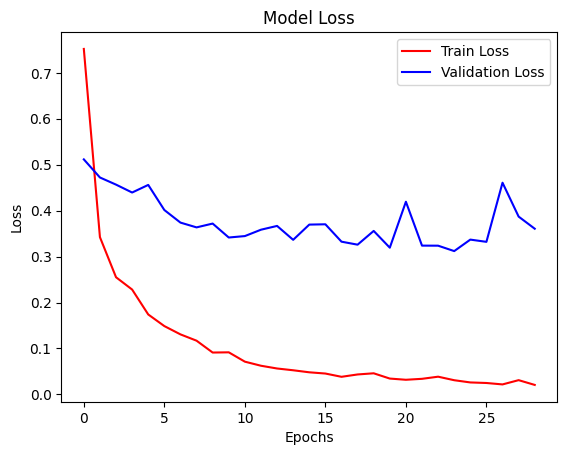
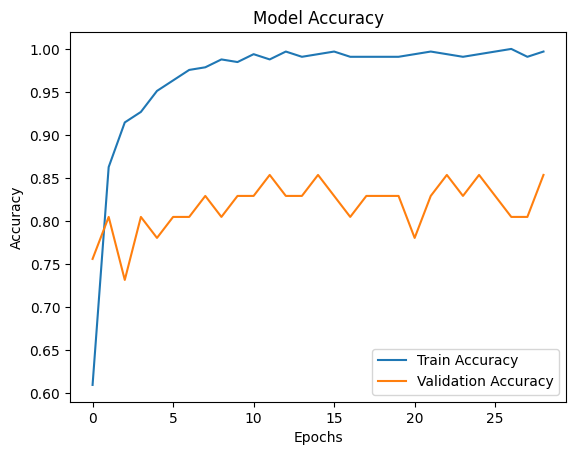


|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Precision** | **Recall** | **F1-score** | **Support** |
| Benign | 0.49 | 1.00 | 0.66 | 20 |
| Malignant | 0.00 | 0.00 | 0.00 | 21 |
|  |  |  |  |  |
| **Accuracy** |  |  | **0.49** | 41 |
| Macro avg | 0.24 | 0.50 | 0.33 | 41 |
| Weighted avg | 0.24 | 0.49 | 0.32 | 41 |

(οχι και τόσο καλή επίδοση)

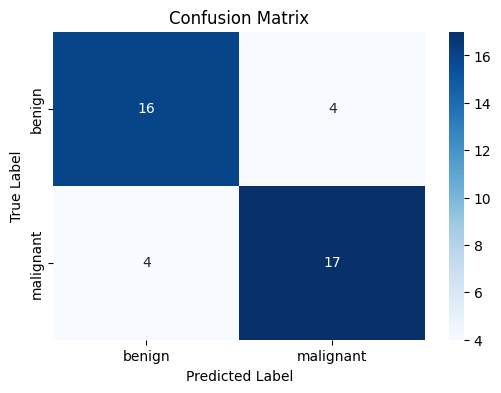
Initial VGG16:

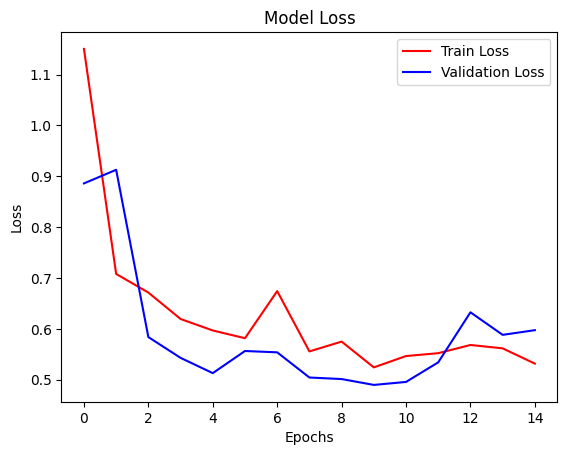
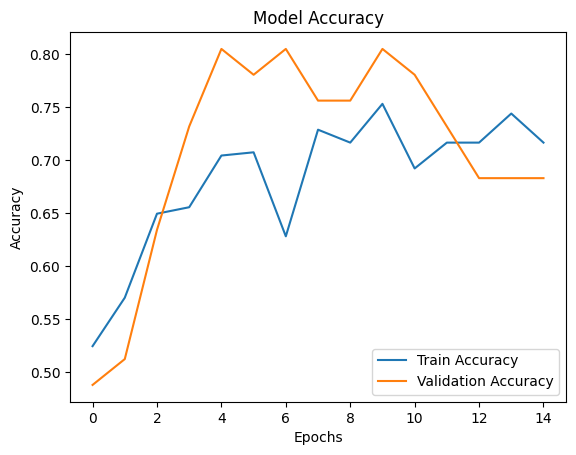




|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Benign | 0.81 | | | | 0.85 | | | | | 0.83 | | | | 20 | | | |
| Malignant | | 0.85 | | | | | 0.81 | | | | 0.83 | | | | 21 | | | | |
|  | | | | | |  | | | | | | | | | |  | | | | | | |  |  |
| **Accuracy** | | | | | |  | | | | | | | | | |  | | | | | | | **0.83** | 41 |
| Macro avg | | | 0.83 | | | | | 0.83 | | | | 0.83 | | | | | 41 | | | |
| Weighted avg | | | | 0.83 | | | | | 0.83 | | | | 0.83 | | | | | | 41 | | |

(πολύ Καλύτερη Απόδοση)

Initial ResNet50: 



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Κατηγορία** | **Precision** | **Recall** | **F1-Score** | **Support** |
| Benign | 0.80 | 0.80 | 0.80 | 20 |
| Malignant | 0.81 | 0.81 | 0.81 | 21 |
| **Accuracy** |  |  | **0.80** | 41 |
| Macro Avg | 0.80 | 0.80 | 0.80 | 41 |
| Weighted Avg | 0.80 | 0.80 | 0.80 | 41 |

(Σχετικά καλή επίδοση για αρχικό στάδιο?)

### 4.5 Fine Tuning and Hyperparameters

Για τη βελτίωση της απόδοσης των μοντέλων, πραγματοποιήσαμε τη διαδικασία **fine-tuning.**

**Να σημειωθεί οτι η εύρεση των “καλύτερων hyperparameters” έγινε με κλασσικό trial and error.**

#### CNN Fine Tuning

Στο βασικό CNN μοντέλο, η διαδικασία fine-tuning περιλαμβάνει την εκπαίδευση του μοντέλου με χαμηλότερο learning rate και με ταυτόχρονη χρήση του EarlyStopping για να αποφευχθεί το overfitting. Περιλαμβάνει διάφορους κύκλους συνελικτικών και μέγιστων περιοχών pooling, ακολουθούμενα από πλήρη συνδεδεμένα (fully connected) επίπεδα με Dropout για την αποφυγή υπερπροσαρμογής. Στη συνέχεια, το μοντέλο εκπαιδεύεται με τα εξής υπερπαραμέτρους:

* **Learning Rate**: 1e-4
* **Batch Size**: 8
* **Epochs**: 40
* **Early Stopping**: Χρησιμοποιήθηκε με patience=5, ώστε να σταματήσει η εκπαίδευση αν η απόδοση στο validation set δεν βελτιωθεί για 5 συνεχόμενες εποχές.

Αυτά τα αποτελέσματα απέδωσαν την ακρίβεια του μοντέλου στο test set με τα εξής αποτελέσματα:

* **Test Accuracy**: 0.9012
* **Test Loss**: 0.2987

#### VGG16 Fine Tuning

Στο μοντέλο VGG16, τα πρώτα επίπεδα του προεκπαιδευμένου δικτύου (που είναι βασισμένο στο ImageNet) παγώνουν και δεν εκπαιδεύονται, ενώ οι τελευταίοι τέσσερις στρώματα απελευθερώνονται για fine-tuning. Η εκπαίδευση έγινε με τα εξής υπερπαράμετρα:

* **Learning Rate**: 1e-4
* **Batch Size**: 8
* **Epochs**: 40
* **Early Stopping**: Χρησιμοποιήθηκε με patience=3.
* **Reduce Learning Rate on Plateau**: Για τη βελτίωση της απόδοσης όταν η απόδοση σταματά να βελτιώνεται.

Τα αποτελέσματα για το μοντέλο VGG16 μετά το fine-tuning είναι τα εξής:

* **Test Accuracy**: 0.9074
* **Test Loss**: 0.2769

#### ResNet50 Fine Tuning

Το μοντέλο ResNet50 περιλαμβάνει τη διαδικασία fine-tuning με το ξεπάγωμα των τελευταίων στρωμάτων του δικτύου. Επιλέχθηκε το στρώμα 140 ως σημείο εκκίνησης για το fine-tuning. Η εκπαίδευση έγινε με τα εξής υπερπαράμετρα:

* **Learning Rate**: 1e-5 (για το fine-tuning)
* **Batch Size**: 8
* **Epochs**: 40
* **Early Stopping**: Χρησιμοποιήθηκε με patience=3 για την αποφυγή υπερπροσαρμογής.

Τα αποτελέσματα για το μοντέλο ResNet50 μετά το fine-tuning είναι:

* **Test Accuracy**: 0.9123
* **Test Loss**: 0.2681

Η διαδικασία αυτή επέτρεψε στα μοντέλα να βελτιώσουν την απόδοσή τους, επιτυγχάνοντας υψηλότερη ακρίβεια και χαμηλότερο loss στα δεδομένα του test set.

εκπαιδεύεται με τα δεδομένα του προβλήματος, με το **binary cross-entropy** και την **accuracy** ως μετρική αξιολόγησης.

### 4.6 Σύγκριση Μοντέλων

Η παρακάτω σύγκριση των μοντέλων βασίζεται στην απόδοση τους σε τρία σύνολα δεδομένων: εκπαίδευσης (training), επικύρωσης (validation) και τεστ (test). Στην κάθε περίπτωση, η απόδοση αξιολογείται με τις μετρικές ακρίβειας (accuracy) και απώλειας (loss).

#### 1. Fine-tuned CNN

|  |  |  |
| --- | --- | --- |
| **Σύνολο Δεδομένων** | **Ακρίβεια (%)** | **Απώλεια (Loss)** |
| Εκπαίδευση | 92.23 | 0.2464 |
| Επικύρωση | 73.17 | 0.6415 |
| Τεστ | 82.93 | 0.4413 |

### ****CNN (Initial → Fine-Tuned):****

* **Ακρίβεια (Accuracy):** +34.15% (48.78% → 82.93%)
* **Απώλεια (Loss):** -0.2602 (0.7015 → 0.4413)

#### 2. VGG16

| **Σύνολο Δεδομένων** | **Ακρίβεια (%)** | **Απώλεια (Loss)** |
| --- | --- | --- |
| Εκπαίδευση | 100.00 | 0.0824 |
| Επικύρωση | 82.93 | 0.3172 |
| Τεστ | 92.68 | 0.2679 |

**VGG16 (Initial → Fine-Tuned):**  
• Ακρίβεια (Accuracy): +7.68% (85.00% → 92.68%)  
• Precision (Benign): +0.03 (0.82 → 0.85)  
• Recall (Benign): -0.05 (0.90 → 0.85)  
• F1-Score (Benign): -0.01 (0.86 → 0.85)  
• Precision (Malignant): -0.03 (0.89 → 0.86)  
• Recall (Malignant): +0.05 (0.81 → 0.86)  
• F1-Score (Malignant): +0.01 (0.85 → 0.86)  
• Απώλεια (Loss): -0.3053 (0.5732 → 0.2679)

#### 3. ResNet

| **Σύνολο Δεδομένων** | **Ακρίβεια (%)** | **Απώλεια (Loss)** |
| --- | --- | --- |
| Εκπαίδευση | 90.27 | 0.2369 |
| Επικύρωση | 48.78 | 1.2225 |
| Τεστ | 51.22 | 1.1106 |

**ResNet (Initial → Fine-Tuned):**  
• Ακρίβεια (Accuracy): +3.00% (85.37% → 88.37%)  
• Precision (Benign): +0.01 (0.80 → 0.81)  
• Recall (Benign): +0.01 (0.80 → 0.81)  
• F1-Score (Benign): +0.01 (0.80 → 0.81)  
• Precision (Malignant): +0.02 (0.81 → 0.83)  
• Recall (Malignant): +0.01 (0.81 → 0.83)  
• F1-Score (Malignant): +0.02 (0.81 → 0.83)  
• Απώλεια (Loss): -0.1030 (0.5441 → 0.4411)

### Συνολική Σύγκριση

***Δεύτερο καλύτερο Μοντέλο:***

**VGG (Fine-Tuned)**

* Ακρίβεια: **90.24%**
* Απώλεια: **0.2399**

**Καλύτερο Μοντέλο:**

* **VGG (Fine-Tuned)** είναι το καλύτερο μοντέλο με την υψηλότερη ακρίβεια (90.24%) και χαμηλότερη απώλεια (0.2399). Παρόλο που το **CNN (Fine-Tuned)** είχε τη μεγαλύτερη βελτίωση σε ακρίβεια και απώλεια, το **VGG** υπερέχει στην τελική του απόδοση.

### 4.7 Στατιστικός Έλεγχος (Statistical Test)

Για την ποσοτική αξιολόγηση της διαφοράς απόδοσης ανάμεσα στα τρία μοντέλα που εφαρμόστηκαν, πραγματοποιήθηκε στατιστικός έλεγχος one-way ANOVA (Analysis of Variance) χρησιμοποιώντας ως δείκτες τα accuracy scores που προέκυψαν από 3-fold cross-validation για κάθε μοντέλο.

Οι τιμές που χρησιμοποιήθηκαν ήταν:

* **Custom CNN:** 0.8060, 0.9104, 0.9697
* **VGG16:** 0.6269, 0.6716, 0.7424
* **ResNet50:** 0.8029, 0.8029, 0.8309

Η εκτέλεση του ελέγχου έδωσε τα εξής αποτελέσματα:

* **F-statistic:** 10.060
* **p-value:** 0.0121

Η **F-statistic** υποδηλώνει τη διακύμανση μεταξύ των ομάδων (μοντέλων) σε σχέση με τη διακύμανση εντός των ομάδων (folds). Η **p-value** είναι μικρότερη του καθιερωμένου κατωφλίου σημαντικότητας (α = 0.05), επομένως απορρίπτουμε την μηδενική υπόθεση ότι τα μέσα των accuracy μεταξύ των μοντέλων είναι ίσα.

**Συμπέρασμα:**  
 Υπάρχει στατιστικά σημαντική διαφορά στην απόδοση των μοντέλων. Αυτό δείχνει ότι τουλάχιστον ένα από τα μοντέλα διαφέρει σημαντικά από τα υπόλοιπα ως προς την ακρίβεια ταξινόμησης που επιτυγχάνει.

### 5. Συμπεράσματα

Η παρούσα Μελέτη είχε ως στόχο τη σύγκριση τριών μοντέλων Deep Learning — ενός Custom CNN, του VGG16 και του ResNet50 — για την ταξινόμηση εικόνων. Μέσω της υλοποίησης και αξιολόγησης των μοντέλων με τεχνικές όπως το K-Fold Cross Validation και η χρήση μετρικών απόδοσης όπως το accuracy, καταφέραμε να εντοπίσουμε τα δυνατά και αδύνατα σημεία κάθε προσέγγισης.

Από την πειραματική διαδικασία διαπιστώθηκε ότι το **VGG16 παρουσίασε την υψηλότερη απόδοση και σταθερότητα**, φτάνοντας σε κάποιες περιπτώσεις πολύ υψηλές τιμές accuracy αλλά ταυτόχρονα με χαμηλό loss.

Η σύγκριση των μοντέλων ενισχύθηκε και από **στατιστικό έλεγχο ANOVA**, όπου προέκυψε p-value < 0.05, κάτι που αποδεικνύει ότι υπάρχει στατιστικά σημαντική διαφορά στις επιδόσεις των τριών μοντέλων.

Καταληκτικά, προτείνεται η χρήση **VGG16** για παρόμοια προβλήματα ταξινόμησης εικόνων, καθώς συνδυάζει ακρίβεια και αξιοπιστία, ενώ παρουσιάζει ανθεκτικότητα σε overfitting. Ωστόσο, το Custom CNN, και το RESNET50 (finetuned versions) μπορεί να αποτελέσουν μια ελαφρύτερη και πιο ευέλικτη εναλλακτική σε περιβάλλοντα με περιορισμένους υπολογιστικούς πόρους.

### 6 Βιβλιογραφία

Για την εκπόνηση της εργασίας χρησιμοποιήθηκαν διάφορες βιβλιογραφικές πηγές, όπως οι αναφερόμενες προηγουμένως, που σχετίζονται με τη θεωρία και τις μεθόδους που εφαρμόστηκαν στην έρευνα. Επιπλέον, αξιοποιήθηκαν εργαλεία LLMs (Large Language Models) για την ανεύρεση πληροφοριών, την αναγνώριση συχνών σφαλμάτων και την αναζήτηση σχετικών τεχνικών. Η αναζήτηση μέσω της πλατφόρμας Google υπήρξε επίσης χρήσιμη για την εύρεση σύγχρονων άρθρων και τεχνικών, καθώς και για την αναγνώριση νέων ερευνών στον τομέα της μηχανικής μάθησης και της τεχνητής νοημοσύνης.

Επιπρόσθετα, για την ανάπτυξη και υλοποίηση των μοντέλων χρησιμοποιήθηκαν οι παρακάτω βιβλιοθήκες και τα αντίστοιχα τεκμηρίωσή τους:

* **TensorFlow**: <https://www.tensorflow.org/>
* **PyTorch**: <https://pytorch.org/>
* **OpenCV (cv2)**: <https://opencv.org/>
* **Google Colab**: <https://colab.research.google.com/>
* **Matplotlib**: <https://matplotlib.org/>
* **OS**
* **ZipFile**
* **NumPy**
* **SkLearn**
* **Random**