**İSTANBUL UNIVERSITY-CERRAHPAŞA**
**INSTITUTE OF GRADUATE STUDIES**

**Ph.D. THESIS**

# MATHEMATICAL MODEL-BASED CLINICAL DECISION SUPPORT SYSTEM ALGORITHM DESIGN STUDY THAT CAN SUPPORT THE DIAGNOSIS OF CELIAC DISEASE

**Elif KESKİN BİLGİÇ**

**SUPERVISOR**
**Assist. Prof. Dr. İnci ZAİM GÖKBAY**

**Department Of Biomedical Engineering**

**Biomedical Engineering Program**

**July, 2024**

## THESIS ACCEPTANCE AND APPROVAL

This thesis titled **"MATHEMATICAL MODEL-BASED CLINICAL DECISION SUPPORT SYSTEM ALGORITHM DESIGN STUDY THAT CAN SUPPORT THE DIAGNOSIS OF CELIAC DISEASE "**, prepared by **Elif KESKİN BİLGİÇ** under the supervision of **Assist. Prof. Dr. İnci ZAİM GÖKBAY** and **Assoc. Prof. Dr. Yusuf KAYAR** , was accepted as a **Ph.D. Thesis** after being **consensus** approved by our jury as a result of the exam held on **15/08/2024**.

**Examining Committee Members**

| | | Signature | Decision |
|---|---|---|---|
| **SUPERVISOR** | Assist. Prof. Dr. İnci ZAİM GÖKBAY<br>Istanbul University<br>Informatics Department | | ☐ Accept<br>☐ Reject |
| **MEMBER** | Assoc. Prof. Dr. Aysel ERSOY<br>İstanbul Cerrahpaşa University<br>Electric Electronic Engineering Department | | ☐ Accept<br>☐ Reject |
| **MEMBER** | Assist. Prof. Dr. Ahmet Naci ÜNAL<br>Bahçeşehir University<br>Software Engineering Department | | ☐ Accept<br>☐ Reject |
| **MEMBER** | Assist. Prof. Dr. Rana ORTAÇ KABAOĞLU<br>İstanbul Cerrahpaşa University<br>Electric Electronic Engineering Department | | ☐ Accept<br>☐ Reject |

| | | |
|---|---|---|
| **MEMBER** | Assoc. Prof. Dr. Zerrin ÖNAL<br>İstanbul University<br>Division of Medical Sciences<br>Department | ☐ Accept<br>☐ Reject |

# DECLARATION

I hereby declare that this thesis is my own work, I have not had engaged in any unethical behavior at any of the stages from planning to the writing of the thesis, I have obtained all the information in this thesis within the limits of the academic and scientific ethical rules, I have cited all the information and comments being obtained in this thesis study, and I have included these references in the list of references. During the analysis and writing stages of this thesis, I accept that I have abided by the patent and copyright rights and therefore I take all kinds of legal responsibility.

Elif KESKİN BİLGİÇ

(Signature)

*To My Daughter and Husband …*

# FUNDING

## MATHEMATICAL MODEL-BASED CLINICAL DECISION SUPPORT SYSTEM ALGORITHM DESIGN STUDY THAT CAN SUPPORT THE DIAGNOSIS OF CELIAC DISEASE

No financial support was received from any institution for this thesis.

# ACKNOWLEDGEMENT

Completing this thesis was made possible by the presence of many valuable people who have always been by my side, offering their support and encouragement at every moment of my life. I extend my deepest gratitude and heartfelt thanks to everyone who guided, inspired, and stood by me throughout this long and challenging journey.

First and foremost, I express my deepest gratitude to my esteemed advisor, Dr. İnci Zaim Gökbay. You witnessed both the hardest and happiest moments of my life and never withheld your moral support. This was more than just mentorship. I am deeply grateful for your presence. Your endless patience, wisdom, and encouragement have guided me. Even when I doubted myself, you believed in me, and for that, I am forever thankful. Your guidance has shaped not only this thesis but also my development as a researcher and an individual.

To my valuable co-advisor, Assoc. Prof. Dr. Yusuf Kayar, I extend my profound love and respect. Your invaluable insights, thoughtful feedback, and continuous support were crucial in developing my research. Your dedication to my success is something I will always remember with gratitude.

I sincerely thank the staff at Van Training and Research Hospital for providing the necessary resources and datasets for my research. The diligence and dedication with which they collected the data made this study possible.

I am very grateful to my esteemed Principal, Aytekin Alpay, for his unwavering support and continuous understanding throughout my doctoral studies.

My dear family, words cannot express the depth of my gratitude for your love and unwavering support. My first thanks go to my greatest supporter, my dear mother-in-law, Zeynep Bilgiç. Beyond taking care of my daughter, I am grateful for the help and moral support you provided. I extend my sincere thanks to my father-in-law, Yaşar Bilgiç, for his sacrifices during this time.

To my dear husband, Engin Erdem Bilgiç, and my daughter, Elisa Ece Bilgiç, your patience, understanding, and love have been my refuge. This success is as much yours as it is mine.

My father Hasan Keskin's biggest wish was to see me as a doctor. Now is the time to witness this. I learned the importance of education, learning, and self-improvement from you. Your vision and cultural richness always broadened my horizons. You are one of the greatest warriors against ignorance. Thank you for opening a vast library in your heart for me and for beautifying my path with your quiet encouragement. My dear mother and siblings, your belief in me and encouragement were my pillars during the toughest times. Your moral support was my source of motivation.

Finally, I thank everyone who touched my life and contributed to this thesis, big or small. Your kindness, support, and belief in me have been the driving force behind my success.

With heartfelt gratitude,

July, 2024 Elif KESKİN BİLGİÇ

# CONTENTS

**Page**

**THESIS ACCEPTANCE AND APPROVAL** ..... **ii**
**DECLARATION** ..... **iv**
**FUNDING** ..... **vi**
**ACKNOWLEDGEMENT** ..... **vii**
**CONTENTS** ..... **ix**
**LIST OF FIGURES** ..... **xii**
**LIST OF TABLES** ..... **xiv**
**LIST OF SYMBOLS AND ABBREVIATIONS** ..... **xv**
**GENİŞLETİLMİŞ ÖZET** ..... **xvii**
**ABSTRACT** ..... **xxii**
**1. INTRODUCTION** ..... **1**
**2. CONCEPTUAL FRAMEWORK** ..... **4**
2.1. CELIAC DISEASE ..... 5
2.1.1. Celiac Disease Pathophysiology ..... 5
2.1.2. Celiac Disease Diagnosis and Challenges ..... 6
2.2. INNOVATIVE APPROACHES TO CLINICAL DIAGNOSIS OF CELIAC ..... 7
*2.3.* HARNESSING AI IN CELIAC DISEASE ..... 9
2.3.1. Facial Recognition Technology ..... 9
2.3.2. Deep Learning and Machine Learning ..... 9
2.3.3. Background Research and Related Works ..... 10
2.4. TRANSFER LEARNING ..... 14
2.4.1. Designing VGG16 Model with Transfer Learning ..... 15
2.4.2. Designing MobileNetV2 Model with Transfer Learning ..... 18
2.5. DESIGNING OF ML MODEL FOR CELIAC MARSH LEVEL CLASSIFICATIONS ..... 20
**3. METHOD** ..... **22**
3.1. DATA ANONYMIZATION ..... 22

3.2. CLASSIFICATION OF CELIAC PATIENT FACE IMAGES WITH TRANSFER LEARNING ..... 23
3.2.1 Collection And Preparation of Datasets for VGG16 And MOBILENET2 ..... 23
3.3. TRANSFER LEARNING WITH VGG16 ..... 24
3.3.1. VGG16 model working stages ..... 24
3.3.2. Preprocessing of Image ..... 24
3.3.3. Configuration of Model ..... 24
3.3.4. Custom Layers Addition ..... 24
3.3.5. Compilation of Model ..... 25
3.3.6. Augmentation of Data ..... 25
3.3.7. Data Generation ..... 25
3.3.8. Training Cycle ..... 25
3.3.9. Metrics and Evaluation ..... 26
3.3.10. Visualization Step ..... 26
3.3.11. Classification Report and Confusion Matrix ..... 26
3.4. TRANSFER LEARNING WITH MOBILENETV2 ..... 26
3.4.1. Environment and Library Setup ..... 26
3.4.2. Uploading and Viewing Images ..... 27
3.4.3. Preparing the Data Set ..... 27
3.4.4. Model Properties and Hyperparameters ..... 27
3.4.5. Training the Model ..... 28
3.4.6. Model Evaluation ..... 28
3.4.7. Exporting and Downloading the Model ..... 28
3.5. TRADITIONAL MACHINE LEARNING AND DEEP LEARNING FOR MARSH LEVEL DETECTION ..... 29
3.5.1. Data Collection and Dataset Preparation ..... 29
3.5.2. Import Necessary Libraries ..... 30
3.5.3. Data Loading and Preparation ..... 30
3.5.4. Model Definitions ..... 31
3.5.5. Model Training ..... 31
3.5.6. Model Evaluation ..... 31
3.5.7. Prediction Interface ..... 31
**4. RESULTS ..... 33**
4.1. RESULTS OF CLASSIFICATION OF CELIAC PATIENT FACE IMAGES WITH VGG16 ..... 33

4.2. RESULTS OF CLASSIFICATION OF CELIAC PATIENT FACE IMAGES WITH MOBİLENETV2 ..... 35
4.3. RESULTS OF CLASSIFICATION OF CELIAC PATIENT CLINICAL DATASETS WITH PYTORCH AND TRADITIONAL MACHINE LEARNING ALGORITHMS ..... 47
**5. DISCUSSION ..... 52**
**6. CONCLUSION AND RECOMMENDATIONS ..... 58**
**REFERENCES ..... 61**
**APPENDICES ..... 70**
**FIRST PAGE OF THE PLAGIARISM REPORT ..... 96**
**ETHICS COMMITTEE PERMISSION .....** Hata! Yer işareti tanımlanmamış.
**INSTITUTIONAL PERMISSIONS .....** Hata! Yer işareti tanımlanmamış.
**CURRICULUM VITAE .....** Hata! Yer işareti tanımlanmamış.

## LIST OF FIGURES

**Page**

**Figure 2.1.** The VGG16 Architectural Structure……..…15

**Figure 2.2.** Removing the "Top" part, VGG16 uses pre-trained output layers to convolve the data……16

**Figure 2.3.** Fully Connected layer for both training and prediction…….17

**Figure 2.4.** Conceptual VGG16 Model…….18

**Figure 2.5.** The model retrained and recreated by Model Maker…….19

**Figure 2.6.** Conceptual MobileNetV2 model……20

**Figure 2.7.** Conceptual Model for marsh level classification…….……..21

**Figure 3.1.** Number of Female and male participating…….……...23

**Figure 3.2.** Diet status of CD patients……..……23

**Figure 3.3.** Celiac Male and Female……..……..30

**Figure 3.4.** Diet status of celiac patients……..……30

**Figure 3.5.** Prediction interface ranging from text boxes to dropdown menus……32

**Figure 4.1.** Displaying train loss and validation loss and train accuracy, validation accuracy versus number of epochs……35

**Figure 4.2.** Displaying train and validation loss and train and validation accuracy versus number of epochs and precision, recall and F1 score……..….38

**Figure 4.3.** First Test image…….....39

**Figure 4.4.** Second Test image……..……...40

**Figure 4.5.** Third Test image……41

**Figure 4.6.** Fourth Test image.....42

**Figure 4.7.** Fifth Test image……...……...43

**Figure 4.8.** Sixth Test image…….……44

**Figure 4.9.** Seventh Test image……..……...45

**Figure 4.10.** Eight Test image……..……..…46

**Figure 4.11.** Ninth Test image……..…….47

**Figure 4.12.** Precision, recall, and F1-score metrics…….49

**Figure 4.13.** PyTorch model's confusion matrix……...49

**Figure 4.14.** Decision Trees, Random Forests, and Gradient Boosting's confusion matrix ……...50

**Figure 4.15.** Precision, recall, and F1-score metrics…….……50

**Figure 4.16.** Precision, recall, and F1-score metrics…….51

**Figure 4.17.** Naive Bayes's confusion matrix……...……51

## LIST OF TABLES

**Page**

**Table 4.1.** Trainable and Non-Trainable Parameters……..…….34

**Table 4.2.**Trained model's testing and validation accuracies…….…….34

**Table 4.3.** The performance Metrices of model……...36

**Table 4.4.** Trainable and Non-Trainable Parameters…….37

**Table 4.5.** Trained model's testing and validation accuracies…….…….37

# LIST OF SYMBOLS AND ABBREVIATIONS

| Symbols | Description |
|---|---|
| **X** | : Feature Space |
| **P(X)** | : Probability Distribution |
| **T** | :Task |
| **D** | : Field |
| **Y** | : Label Space |
| **f(·)** | : Objective Prediction Function |
| xi, yi | : Goals of Transfer Learning |
| **DS** | : Source Domain |
| **TS,TT** | : Learning Task |
| **DT** | : Target Domain |
| **fT(·)** | : Target Prediction Function |

| Abbreviations | Description |
|---|---|
| **CD** | : Celiac Disease |
| **ÇH** | : Çölyak hastalığı |
| **ML** | : Machine Learning |
| **DL** | : Deep Learning |
| **TL** | : Transfer Learning |
| **DS_Face** | : Dataset Containing Face Images |
| **DS_Marsh** | : Dataset Containing Clinical Tests |
| **CNN** | : Convolutional Neural Network |
| **TTG2** | : Tissue Transglutaminase 2 |
| **LgA** | : Immunoglobulin Antibody |
| **CDSS** | : Clinical Decision Support System |
| **FCM** | : Fuzzy Cognitive Map |
| **PCD** | : Potential of Celiac Disease |
| **tTG** | : Tissue Transglutaminase Antibodies |

**MMM** : Mediapipe Model-Marker
**BMI** : Body Mass Index
**GIS** : Gastrointestinal Symptoms
**HCT** : Hematocrit
**HGB** : Hemoglobin
**IGM** : Immunoglobulin M
**IGG** : Immunoglobulin G
**IGA** : Immunoglobulin A

# GENİŞLETİLMİŞ ÖZET

## DOKTORA TEZİ

### ÇÖLYAK HASTALIĞININ TEŞHİSİNE DESTEK VEREBİLECEK MATEMATİK MODEL TABANLI KDS ALGORİTMASI TASARIMI

**Elif KESKİN BİLGİÇ**

**İstanbul Üniversitesi-Cerrahpaşa**

**Lisansüstü Eğitim Enstitüsü**

**Biyomedikal Mühendisliği Anabilim Dalı**

**Biyomedikal Mühendisliği Programı**

**Danışman: Dr. Öğr. Üyesi İnci ZAİM GÖKBAY**

**II. Danışman: Doç. Dr. Yusuf KAYAR**

**Giriş**: Çölyak hastalığı (ÇH), ince bağırsağı etkileyen, buğday, arpa ve çavdarda bulunan bir protein olan glutenin tüketilmesiyle tetiklenen bir otoimmün bozukluktur. ÇH'li bireyler gluten tüketimleri sonrası ince bağırsaklarının iç kısmının zarar görmesine bağlı olarak, çeşitli semptomlar ve potansiyel uzun vadeli komplikasyonlar yaşamalarına neden olan bağışıklık tepkisi yaşamaktadır. ÇH'nin kesin tanısı tipik olarak bağırsak hasarının boyutunu belirleyen, ince bağırsak biyopsisi ile yapılmaktadır. Çölyak tanısının koyulması, bazı klinik belirtileri ve semptomlarının diğer otoimmün hastalıklar ile örtüşmesi nedeniyle zordur. Erken teşhis, bağırsak geçirgenliğinin hızla artmasını önleyerek hastalığın şiddetli ilerlemesini önleyici niteliktedir. ÇH tanısında yaşanan zorlukların üstesinden gelmek ve hastaların en kısa sürede uzmana ulaşmalarını sağlamak amacıyla teşhise destek verebilecek nitelikte sistemlerin geliştirilmesi gelişen teknolojinin de yardımıyla gittikçe önemli hale gelmektedir. Yapılan araştırmalar, video kapsül görüntü işleme alanında bağırsak epiteli deformasyonuna odaklanırken, glüten duyarlılığının yüz görünümü üzerindeki etkisine alan yazında halen açık olan bir araştırma konusudur. Hâlbuki, çölyak, bağırsakta iltihaplanma, bağışıklık sistemi değişiklikleri ve mikro-biyom disbiyozu gibi mekanizmalar yoluyla cildi etkileyen

bozulmalarla ilişkilidir. Ayrıca, ÇH, gastrointestinal sistemi etkilemesi ve besinlerin emilim bozukluğuna yol açabilmesi nedeniyle cilt yapısında ve yüz özelliklerinde değişimlere neden olmaktadır. Bu durum gözlemlenebilir yüz özellikleri veya desenleri olan hastalıkların teşhisini ve tanımlanmasını geliştirmek için kullanılan yüz sınıflandırmasının ÇH teşhisine destek vermesinde umut vaat etmektedir.

Her ne kadar alan yazında, hastalık öyküsü ve semptomlarına ilişkin verilerin kullanıldığı erken teşhise destek verme amacında çalışmalar bulunsa da yüz görüntülerinin de kullanıldığı hibrit modeller bulunmamaktadır. Buradan hareketle bu tez çalışması iki farklı araştırma sorusu altında planlanmıştır. İlk araştırma sorusu ÇH yüz üzerindeki değişimlere odaklanmış ve *"Transfer öğrenme modelleri VGG16 ve MobileNetV2, çölyak hastalığının tanısına yardımcı olmak amacıyla çölyak hastalarının yüz görüntülerini sınıflandırmada ne kadar etkilidir ve hastalığının erken teşhisine anlamlı performansta destek verir mi?"* biçiminde kurgulanmıştır. İkinci araştırma sorusunda ise *"Birinci basamak sağlık merkezlerinde yapılan rutin kan testi sonuçları ve demografik verilerin kullanılmasıyla çölyak hastalığı Marsh (0-4) düzeyleri makine öğrenme modelleriyle belirlenebilir mi?"* odaklanılmış ve yöntem olarak derin öğrenme modelleri (PyTorch) ve geleneksel makine öğrenme algoritmalarının performansları karşılaştırılmıştır. Her iki araştırma sorusuyla alan yazına yeni bir bakış açısı sunulması hedeflenmiştir. Hali hazırda böyle bir veri setinin olmaması nedeniyle Türkiye, Van Eğitim ve Araştırma hastanesinde gastroenteroloji kliniğinde takip edilen ÇH tanılı hastalardan *2022/09-02* etik izni kapsamında veri toplanmıştır.

**Amaç**: Birçok otoimmün hastalıkla benzer semptomları bulunan ÇH'nın erken teşhisinin yapılabilmesi, hastaların en kısa sürede uzmana yönlendirilmesini sağlama amacını güden bir sistem tasarımının yapılabilirliği ve klinikte etkin kullanımının incelenmesinin önemi aşikardır. Buradan hareketle bu tezde ilk olarak *yüz görüntülerinin yer aldığı veri tabanının transfer öğrenme yöntemleriyle analizi sonucu çölyak hastalığının erken teşhisine anlamlı performans gösterebilme yeteneğinin sorgulanması amaçlanmaktadır. İkinci olarak ise birinci basamak sağlık birimlerinde diğer hastalıklarla benzer semptomatik özellikleri nedeniyle uzmana yönlendirilmede gecikilmesinin önüne geçilmesinin hedeflendiği rutin kan testi ve demografik verilerle, Karar Ağacı, Rastgele Orman, Gradient Boosting, Naive Bayes ve derin öğrenme(pytorch) sınıflandırma algoritmaları kullanılarak hastalığın teşhisine ve Marsh seviyelerinin belirlenmesine destek verecek bir yöntemin analizinin yapılması da amaçlanmaktadır.*

**Yöntem**: Tez çalışmasında iki farklı veri setinden faydalanarak iki amaca uygun yöntem uygulaması yapılarak yorumlanmıştır. İlk araştırma sorusu için, çölyak hastalarının yüz görüntüleri transfer öğrenme yöntemi ile sınıflandırılmıştır. Önceden eğitilmiş iki farklı model (VGG16 ve MobileNetV2) kullanılmış ve en yüksek sınıflandırma performansına sahip model seçilmiştir. İkinci araştırma sorusu için, çölyak hastalarında bağırsak atrofisi (marsh level) belirleme hedefiyle, klinik belirti ve semptomları derin öğrenme ve geleneksel makine öğrenme algoritmaları ile sınıflandırılmış, ve yine en yüksek doğruluk oranına sahip model belirlenmiştir. Her iki aşamada da performans ölçütleri yüksek olan modeller ilerleyen zamanlarda hibrit veri setlerinin kullanılabileceği modellerin geliştirilmesi için seçilmiştir.

Yüz görüntülerinin yer aldığı veri setinde (DS_Face), ÇH tanılı ve eşit sayıda kontrol grubuna sahip 200 yetişkine ait yüz görüntüsü bulunmaktadır. DS_Face veri setinde yer alan ÇH tanısı bulunan görüntülerin %70'i kadın, %30'u erkek iken kontrol grubunun dağılımında ise %60 kadın, %40 erkektir. ÇH tanılı yüz görüntülerindeki hastaların 28'ine yeni tanı konmuş, 72'sine önceden tanı almıştır. Önceden tanı konulmuş hastaların 25'i glutensiz diyete uymuyor, 47'si ise kısmen diyete bağlı kalıyordu.

Rutin kan testi ve demografik bilgilere dayanarak marsh seviyesinin belirlenmesi amacıyla oluşturulan ikinci veri seti (DS_Marsh); 182 kişiye ait hasta öyküsü, demografik bilgi ve kan testi sonuçlarından oluşmaktadır. DS_Marsh veri seti; 132 kadın ve 50’si erkek hasta dağılımına sahiptir. Bu veri setinde glütensiz diyete uyum dağılımı ise uymayan 72 ve kısmi uyum gösteren 106 kişi biçimindedir.

**Bulgular:** DS_Face veri setinde glüten diyetine uymayan, kısmi uyan, yeni tanılı hastalar ve kontrol grubunun yüz görüntülerinin önceden eğitilmiş VGG16 modelinin kullanılmasıyla yapılan sınıflandırılmasında %73’lük doğruluk oranı elde edilmiştir. Hassasiyet, hatırlama ve F1 puanı değerleri ise sırasıyla 0.54, 0.56 ve 0.52 olarak bulunmuştur. Medipipe model marker’ın kullandığı önceden eğitilmiş MobileNet.v2 Modelinde yaklaşık 0,44'lük bir test kaybına ve test veri kümesinde %85'lik bir doğruluk oranı elde edilmiştir. Hassasiyet, hatırlama ve F1 puanı değerleri sırasıyla 0.80, 0.81 ve 0.80’ dir. Sonuç olarak DS_Face veri setinde MobileNet.v2 modeli, VGG16 modeline kıyasla ÇH tanısında daha güçlü bir performans göstermiştir.

DS_Marsh veri setinde PyTorch modeli, Marsh seviyelerini belirlemede %80 doğruluk oranında performans göstermiştir. Hassasiyet, geri çağırma ve F1 puanları ise sırasıyla 0.81,

0.80 ve 0.70 olarak ölçülmüştür. Herbiri tam performansla, %100 doğruluk oranı, sınıflandırma kapasitesi gösteren Karar Ağacı, Rastgele Orman ve Gradient Boosting sınıflandırıcılarıyla hassasiyet, geri çağırma ve F1 puanları 1.0, 1.0, 1.0 olarak elde etmiştir. Diğerleriyle kıyaslandığında %55 doğruluk oranı gösteren Naive Bayes sınıflandırıcısının hassasiyet, geri çağırma ve F1 puanları ise sırasıyla 0.67, 0.44 ve 0.53 olarak sıralanmaktadır.

**Sonuçlar:** Birinci araştırma sorusunun incelenmesi için oluşturulan DS_Face veri setinde VGG16 Modeli ve Mobilenet.V2 Modelinin performansları karşılaştırılmıştır. VGG16 Modeli test görüntülerinin yaklaşık dörtte üçünü doğru şekilde sınıflandırmıştır. Mobilenet.V2 Model ise, test görüntülerinin %85 oranında doğru sınıflandırmıştır. VGG16 modelinin, mimari yapısı ve aşağıda belirtilen bazı diğer özellikleri , sınıflandırma performansının MobilenetV2 modelden daha düşük olmasına neden olmuştur. VGG16 model mimarisi gereği daha fazla parametreye sahip daha derin bir ağdır. MobileNetV2 ise daha az parametreyle, küçük veri kümelerinde daha verimli olacak şekilde tasarlanmıştır. Buda, MobilenetV2'nin DS_Face veri setinde daha iyi performans göstermesine neden olmuştur. Ayrıca, MobileNetV2, Mediapipe model marker ile kullanıldığı için, mediapipe model marker, modelin performansını artıran ve daha yüksek doğruluğa katkıda bulunan ek ön işleme ve özellik çıkarımı sağlamıştır. Ve son olarak, daha karmaşık bir model olan VGG16, özellikle veri kümesi yeterince büyük değilse aşırı uyum sağlamış olabilir. Sonuç olarak, Mobilenet.V2 gibi yüz sınıflandırma modellerinin yeni ÇH tanısı almış, önceden ÇH tanısı almış fakat glüten diyetine uymayan hastaların cilt ve yüz özellikleri üzerinde ortaya çıkan değişimler üzerindeki performansları, bu hastalarda erken tanıya destek verebilme yeteneğine sahip karar destek modellerinin geliştirilebilirliğine ilişkin somut bulgular sunmaktadır.

İkinci araştırma sorusunun incelenmesi için oluşturulan DS_Marsh veri setinde sınıflandırma için kullanılan modeller, sırasıyla pytorch %80 üzerinde doğruluk gösterirken, Karar Ağacı, Rastgele Orman ve Gradient Boosting ise 100% oranınında performans göstermiştir. Bu rutin kan testi bulguları ve demografik bilgilerin birinci basamak sağlık birimlerinde dahi kullanılabilmesine olanak sağlayacak sistem tasarımlarının hastaların yaşam kalitesi, sağlık sistemindeki hasta yükü ve ekonomiye olan katkıları nedeniyle umut vaat edicidir.

İki araştırma sorusuyla hastanın yaşam kalitesinin arttırılması, sağlık hizmetlerinin yükünde ve ekonomide harcamaları azaltmayı amaçlayarak planlanan bu çalışma, karar vermeyi geliştirmek ve hasta sonuçlarına fayda sağlamak için tıbbi teşhislerde veriye dayalı yöntemlerin önemini vurgulamaktadır. İlerleyen çalışmalarda hem yüz görüntülerinin hem rutin kan bulgusu

ve demografik bilgilerin kullanılarak erken teşhis için uzmana yönlendirmeyi hızlandıracak hibrit sistem tasarımlarında, Mediapipe model marker'ın önceden eğitilmiş MobileNetV2 modeli ve Karar Ağacı, Rastgele Orman ve Gradient Boosting sınıflandırıcılarının kullanılmasının yüksek performans göstereceği sonucuna varılmıştır.

Temmuz 2024 , 3 sayfa.

**Anahtar kelimeler:** Çölyak Hastalığı Teşhisi, Görüntü sınıflandırma, Derin ve makine öğrenimi, Transfer öğrenimi, VGG16 ve Mediapipe.

# ABSTRACT

## Ph.D. THESIS

### MATHEMATICAL MODEL-BASED CLINICAL DECISION SUPPORT SYSTEM ALGORITHM DESIGN STUDY THAT CAN SUPPORT THE DIAGNOSIS OF CELIAC DISEASE

**ELİF KESKİN BİLGİÇ**

**İstanbul University-Cerrahpaşa**

**Institute of Graduate Studies**

**Department of Biomedical Engineering**

**Biomedical Engineering Program**

**Supervisor: Assist. Prof. Dr. İNCİ ZAİM GÖKBAY**

**Co-Supervisor: Assoc. Prof. Dr. YUSUF KAYAR**

**Background**: Celiac disease (CD) is one of autoimmune disorder. It is caused by consumption of gluten which is a protein found in barley, rye and wheat. Individuals with CD experience an immunity response due to damage to the lining of their small intestines after consuming gluten, causing them to experience a variety of sign and symptoms also potential long-term complications. Currently definitive diagnosis of CD is typically made by small bowel biopsy, which determines the extent of intestinal damage. Diagnosing celiac disease is difficult because some of its signs and symptoms resemble with other autoimmune diseases. Early diagnosis prevents severe progression of the disease by preventing rapid increase in intestinal permeability. In order to overcome the difficulties experienced in the diagnosis of CD and to enable patients to reach a specialist as soon as possible, it is becoming increasingly important to develop systems that can support diagnosis, with the help of developing technology. While research focuses on intestinal epithelium deformation in the field of video capsule image processing, the effect of gluten sensitivity on facial appearance is still an open research topic in

the literature. When looking at the development of celiac disease in the body, it is associated with diseases that impact the skin through intestinal inflammation, immune system problems and microbiome dysbiosis. In addition, CD causes changes in skin structure and facial features because it affects the gastro-intestinal system and can cause malabsorption of nutrients. This holds promise for supporting the diagnosis of CD through face image classification by detecting observable facial features or some patterns. Although there are studies in the literature that aim to support early diagnosis using data on disease history and symptoms, there are no hybrid models that use facial images. Based on this, this thesis study was planned under two different research questions. The first research question focused on the changes on the CD face images and asked, “How effective are the transfer learning models VGG16 and MobileNetV2 in classification CD facial images to aid in the diagnosis, and do they support the early diagnosis of CD with significant performance?” It is structured in the format. The second research question asked: "Can CD Marsh (0-4) levels be determined with machine learning models using routine blood test results and demographic data performed in primary health centers?" focused on and compared the performance of deep learning models (PyTorch) and traditional machine learning algorithms. With both research questions, it is aimed to present a new perspective on the literature. Since such a data set does not currently exist, data was collected from patients diagnosed with CD followed in the gastroenterology clinic at Van Training and Research Hospital, Turkey, within the scope of ethical permission 2022/09-02.

**Purpose**: It is obvious that it is important to examine the feasibility of designing a system that aims to ensure early diagnosis of CD, which has similar symptoms to many autoimmune diseases, and to ensure that patients are referred to a specialist as soon as possible, and to examine its effective use in the clinic. Based on this, this thesis first aims to question the ability of transfer learning with two pretrained models to classify the database containing facial images to show meaningful performance in the early diagnosis of CD. Secondly, it is aimed to prevent delays in referring celiac disease to a specialist in primary health care units due to its similar symptomatic features with other diseases. For this purpose, it is aimed to analyze a method that will support the diagnosis of the disease and the determination of Marsh levels by classifying routine blood test and demographic data with Decision Tree, Random Forest, Gradient Boosting, Naive Bayes and deep learning (Pytorch) algorithms. It is also aimed to analyze a method that will support the determination of the celiac level severity.

**Method**: In the thesis study, two different data sets were used and interpreted by applying two methods suitable for the purpose. For the first research question, facial images of celiac patients were classified using the transfer learning method. Two different pre-trained models (VGG16 and MobileNetV2) were used. Then model with the highest classification performance was selected. For the second research question, with the aim of determining intestinal atrophy (marsh level) in celiac patients, clinical signs and symptoms were classified with deep learning (Pytorch) and traditional machine learning algorithms, and the model with the highest accuracy rate was determined. Models with high performance criteria in both stages were selected for the development of models in which hybrid data sets could be used in the future.

In the dataset containing face images (DS_Face), there was a CD diagnosis and an equal number of control groups. There are 200 adult facial images. While 70% of the images diagnosed with CD in the DS_Face dataset is female and 30% are male, the distribution of the control group is 60% female and 40% male. Of the patients in the facial images diagnosed with CD, 28 were newly diagnosed and 72 were previously diagnosed. Of the previously diagnosed patients, 25 did not comply with the gluten-free diet. 47 of them are partially adhered to the diet.

The second data set (DS_Marsh) was created to determine the marsh level based on routine blood test and demographic information; It consists of patient history, demographic information and blood test results of 182 people. DS_Marsh dataset; It has a distribution of 132 female and 50 male patients. In this data set, the distribution of compliance with the gluten-free diet is as follows: 72 people who do not comply and 106 people who partially comply.

**Results**: In the DS_Face dataset, an accuracy rate of 73% was reached in the classifying of facial images of non-compliant, partially adherent, newly diagnosed patients and the control group using the pre-trained VGG16 model. Precision, F1 and recall score values were found to be 0.54, 0.52 and 0.56, respectively. Moreover, the pre-trained MobileNet.v2 model used by the Medipipe model marker had a test loss of approximately 0.44 and an accuracy rate of 85% on the test dataset with 0.80 precision, 0.81 recall and 0.80 F1 scores. As a result, in the DS_Face dataset, the MobileNet.v2 model showed a stronger performance in CD diagnosis compared to the VGG16 model.

On the DS_Marsh dataset, the PyTorch model performed with 80% accuracy in determining Marsh levels. Recall, precision, and F1 scores were measured as 0.80, 0.81 and 0.70, respectively. Recall, precision, and F1 cores were achieved as 1.0, 1.0, 1.0 with the Decision

Tree, Random Forest and Gradient Boosting classifiers, each of which showed full performance, 100% accuracy rate and classification capacity. The recall, precision, and F1 scores of the Naive Bayes classifier, which shows a 55% accuracy rate compared to others, are listed as 0.44, 0.67 and 0.53, respectively.

**Conclusion**: The performances of the VGG16 Model and Mobilenet.V2 Model were compared in the DS_Face data set created to examine the first research question. Model VGG16 correctly classified almost one third of the test images. Mobilenet.V2 Model classified 85% of the test images correctly. The VGG16 model's classification performance was lower than the MobilenetV2 model due to its architectural structure and some other features mentioned below. VGG16 is a deeper network with more parameters due to its model architecture and more suitable for big dataset. MobileNetV2, on the other hand, is designed to be more efficient on small data sets with fewer parameters. This caused MobilenetV2 to perform better on the DS_Face dataset. Additionally, since MobileNetV2 was used with the Mediapipe model marker, the mediapipe model marker provided additional preprocessing and feature extraction that improved the performance of the model and contributed to higher accuracy. And finally, VGG16, a more complex model, may have overfitted, especially if the dataset is not large enough. In conclusion, the performance of facial classification models such as Mobilenet.V2 on the changes that occur on the skin and facial features of patients newly diagnosed with CD and previously diagnosed but not complying with the gluten diet provides concrete findings regarding the development of decision support models capable of supporting early diagnosis in these patients.

The models used for classification in the DS_Marsh dataset created to examine the second research question, respectively, showed over 80% accuracy in Pytorch, while Decision Tree, Random Forest and Gradient Boosting showed 100% performance. System designs that will allow these routine blood test findings and demographic information to be used even in primary health care units are promising due to their contribution to the quality of life of patients, the patient load in the healthcare system, and the economy.

This study, planned with two research questions to improve the patient' life quality and reduce the burden of healthcare services and expenses in the economy, emphasizes the importance of data-based methods in medical diagnoses to improve decision-making and benefit patient outcomes. In future studies, it has been concluded that the use of the Mediapipe model marker with pre-trained MobileNetV2 model and Decision Tree, Random Forest and Gradient

Boosting classifiers will provide high performance in hybrid system designs that will accelerate referral to a specialist for early diagnosis by using both facial images, routine blood findings and demographic information.

July 2024, 3 pages.

**Keywords:** Diagnosis of Celiac Disease, Face Image classification, Deep and machine learning, Transfer learning, Mediapipe (MobileNetV2),and VGG16

# 1. INTRODUCTION

Celiac disease (CD) is the most common seen food-sensitive intestinal disease in humans. CD is an autoimmune disease triggered by the eating of gluten in predisposed individuals genetically. It affects around 1% of the global population, although many cases remain undiagnosed [1]. When individuals with CD ingest gluten, their immune system invades the inner lining of the small intestine, causing to a range of symptoms including gastrointestinal distress, malabsorption of nutrients, and a variety of extraintestinal manifestations [2]. Early diagnosis of CD is crucial to catch growth in children and also prevent long-lasting complications. Furthermore, if patients of CD are not diagnosed in time, they are more likely to develop some certain type of cancers. These are small intestine, esophageal cancers, colon cancers, B-cell and T-cell non-Hodgkin lymphomas, oropharyngeal and hepatobiliary cancers, and lastly pancreatic cancer [3]. Despite advances in medical research, the diagnosis of celiac disease remains challenging. The current diagnostic process involves a combination of serological tests and invasive intestinal biopsies, which are not only uncomfortable for the patient but also may not be immediately accessible in all healthcare settings. Furthermore, the variability in symptoms presentation often leads to delayed diagnosis, resulting in prolonged patient suffering and an increased risk of complications [4]. When we take these limitations into consideration, this area requires ongoing research to advance the accuracy and topicality of CD diagnosis. CD diagnosis is challenging because of its various clinical manifestations and symptoms common with other diseases [5]. There is a critical need for non-invasive and rapid diagnostic tools that is more accessible to improve early detection and management of celiac disease.[6]. At the present time, there is a growing interest for discovering novel clinical decision support systems that can aid [7]in CD. However, very less studies present about effect of gluten irritations in intestine on facial appearance. CD is closely related with disturbance in intestinal health which can indirectly affect the skin with many mechanisms including inflammation, microbiome dysbiosis and immune system variation[8, 9]. Adult CD patients tend to have a certain facial appearance, featured by pale complexion, itchy rashes, cracking at the corners of the lips, facial edema, dark circles under the eyes, gaunt appearance and a hollow look[10, 11]. In this manner, CD primarily impact the intestinal tract resulting in malabsorption of nutrient, it also induces in changes in facial and skin characteristics [12]. Facial image classification

promises in assistance to the diagnosis and defining of specific diseases, particularly those with clear and discernable facial features or patterns like CD [13–16].

Doing this with Deep learning is very complex as it acquires a big enough dataset to train a deep CNN. Training such a large dataset from scratch requires significant computational and recollection resources, as well as extensive datasets that are often not available for certain type of diseases. However, the risk of coming upon overfitting issues becomes more pervasive when small datasets are processed [17]. To overcome this challenge, transfer learning has been developed as a viable solution.

In this study, firstly transfer learning with pre-trained models namely VGG16 and MobileNetV2 to classify celiac patient face images to aid diagnosis has been used. Ability of transfer learning with these pretrained models for classifying celiac disease patient from nonceliac have been investigated. These two architectures are designed to extract specific features from face for classification. Although the fine facial features of CD patients are difficult to distinguish with the naked eye, these models trained with millions of images in Imagenet have the ability to capture them with their special sensitive parameters [18,19].

Secondly, we wanted to determine the intestinal atrophy (marsh level from 0-4) of CD by classifying of serological test and demographic data with machine learning algorithms to aid diagnosis and guidance of current patients in first step health centers. CD exhibits varied clinical presentations and symptoms that overlap with other conditions. Using some sign, symptoms and serological test to find out CD severity level can be time saving during diagnosis. CD diagnosis often involves biopsy to assess the severity of small intestinal mucosal damage, commonly categorized using the Marsh classification system [20]. Traditional methods for interpreting biopsy images and classifying Marsh levels rely heavily on manual analysis, which can be time-consuming and liable to subjectivity. Although guidelines exist for diagnosing celiac disease, assessing its severity is challenging. This challenge arises from inter observer variability, the non-specific nature of histopathological properties, and the lack of certain markers that associate with the solemnity of CD or its associated comorbid conditions [21, 22]. Recent advancements in machine learning techniques offer promising ways to automate biopsy image analysis and improving the correctness of Marsh level detection in CD [23]. Although there are some research about marsh level detection in CD by using serological test [24] but there are very few studies for classification of serological tests and demographic data with machine learning algorithms for identification of intestinal atrophy.

This thesis aims to design a hybrid decision support system which can help in the diagnosis of CD by using patients' facial photographs and clinical data. With integrating facial recognition and ML, the system seeks to enhance the diagnostic process, ensuring that patients are directed to the appropriate specialist promptly. The goal is to prevent complications arising from delayed diagnosis, improving patient outcomes by facilitating timely intervention.

The study is conducted in two stages: First, CD *Facial Image Classification*: The first stage involves the classification of facial images dataset (DS_Face) of celiac patients using the transfer learning method [25]. Two pre-trained models, VGG16 [15,26]and MobileNetV2 [27], were employed to identify the model with the highest classification performance. Second, CD *Clinical Data Analysis*: In the second stage, clinical signs and symptoms dataset (DS_Marsh) of celiac disease was classified using DL and traditional ML algorithms to establish the Marsh level, which indicates the severity of intestinal damage. The model that showed highest correctness rate was selected[22]. Two different datasets DS_Face and DS_Marsh have been used for two different methods for one purpose.

A collection of DS_Face dataset including 100 patient images diagnosed as CD was gathered by the Gastroenterology Department of Van Education and Research Hospital between 2022 and 2023, according to *2022/09-02* established protocols. Furthermore, set of non-celiac images consisting of 100 cases were procured from the same hospital. This research study received backing from Health Science University in Turkey.

DS_Marsh dataset was obtained from Van Training and Research Hospital, Gastroenterology Department polyclinic from patients between 2015-2019. Research Ethics Committee of the University of Health Sciences approved this study with decision no. 586 dated 10.09.2018.

The proposed decision support system represents a novel approach to the diagnosis of CD. The system has the potential to bring novelty the diagnostic process by leveraging non-invasive facial image analysis and integrating them with clinical data analysis. This approach not only aims to reduce the reliance on invasive procedures but also seeks to ensure that patients receive timely and accurate diagnoses, thereby reducing the burden of undiagnosed celiac disease and improving overall patient care. By addressing the challenges of current diagnostic methods and proposing an innovative solution, this thesis aims to conduce significantly to the early identification and management of celiac disease.

# 2. CONCEPTUAL FRAMEWORK

## 2.1. CELIAC DISEASE

This thesis aims to design a hybrid decision support system to help in the diagnosis of CD by usage of patients' facial photographs and rutin blood test and demographic data. The system aims to prevent complications arising from delayed diagnosis by ensuring prompt referral to the appropriate specialist. The conceptual framework integrates principles from medical imaging, transfer learning, machine learning, and clinical diagnostics to achieve this goal in this thesis.

### 2.1.1. Celiac Disease Pathophysiology

Hippocrates said that" All diseases begin in the gut" twenty-five hundred years ago. It was emphasized that although there is not much information about the pathology of many chronic inflammatory diseases, it may be related to gut health. Twenty-three thousand human genes are inadequate to clarify all possibility of human pathophysiology, with the inclusion of whether we develop disease and when and why we develop it [28]. The health of the intestines, which is the center of the digestive system and defined as the second brain, affects the entire metabolism [29]. An undesirable situation in the intestine is that it is permeable. Intestinal permeability or leakage is a condition related to the opening of the tight ligaments that make up the intestinal mucosa [30]. These "tight ligaments" are transit points that allow only digested food, minerals and vitamins to enter the bloodstream from the intestines. If the tight intercellular bonds in the intestines are broken, undigested food particles, harmful substances, toxins and pathogens can pass into the bloodstream [31, 32]. Leaking pathogens or poorly digested nutrients enter the bloodstream, and the immune system attacks these unknown substances and attacks its own tissue [33]. Zonulin is the most important of the intestinal tight junctions [33]. Among the few potential gut lumen stimulants that can stimulate the release of zonulin in the gut, bacterial growth and exposure to celiac disease-causing gluten have been identified as the two most potent triggers [34]. Gliadins, which are indigestible gluten fragments, bind to CXCR3 receptors and can release zonulin. Although gluten can stimulate the release of zonulin in healthy individuals and celiac patients, quantity and lenght of time of zonulin produced is very important in CD [35]. This leads to a considerable increase in intestinal permeability whilst displayed by the capacity of the zonulin inhibitor [36]. CD is the most encountered food-sensitive gastrointestinal case induced by gluten irritability. CD also causes leaky gut which opens the

door to all kinds of autoimmune diseases in humans [37] . Looking at the genetic background of celiac disease, it is seen that most celiac patients possess HLA-DQ2 (DQA105 and DQB102 alleles) or HLA-DQ8 (DQA103 and DQB103:02 alleles) haplotypes [38]. These genetic markers are necessary in diagnosis, but they are not sufficient alone. More than 40 non-HLA genetic loci also contribute to disease susceptibility [39]. In genetically predisposed people, gluten consumption stimulates an immune response in which gluten peptides bind to HLA-DQ2 or HLA-DQ8, activating CD4+ T cells and causing intestinal inflammation [40]. Genetic testing for these markers helps with diagnosis and family screening but does not provide a definitive diagnosis of the disease.

CD influences about 1% of the global population. However, differences in incidence in the world cannot be elucidated by genetic and environmental risk factors [1] Most of the survey studies were carried out in Europe. Findings about incidence differ among countries. Sweden, Turkey, Finland, United Kingdom, Czech Republic, Italy and Portugal have high prevalence for CD in Europe, while celiac disease is less common in Estonia, Russia, Iceland and Poland [41]. The total cost to the Clinical Commissioning Group of managing CD patients was over £5 million. In addition, the cost has increased due to other diseases that develop together with celiac disease. A significant number of patients (29%) have iron deficiency anemia and complications related to iron deficiency[42].

### 2.1.2. Celiac Disease Diagnosis and Challenges

Celiac diseases diagnosis is challenging in many ways. Mostly immunoglobulin antibody (lgA) and Tissue transglutaminase 2 (TTG2) markers are check at first suspected. TTG2, which is an intestinal enzyme, does gluten peptides toxic through transamination and deamidation reactions [43] . Plasma cells secrete IgA antibodies against components of the mucosal layer and deamidated gluten peptides. Furthermore, IgA molecules are produced as antibodies against endomysium, transglutaminase 2, and deamidated gliadin peptides [44]. Determination of lgA molecules passing into the bloodstream is useful for the diagnosis of CD [45]. However, the use of IgA markers in the diagnosis of celiac disease poses difficulties due to IgA deficiency in 2-3% of patients, leading to false negative results in IgA-based tests. Amount of antibody vary with disease stage and gluten exposure, and cross-reactivity can complicate results. Correct diagnosis often requires serological testing combined with clinical evaluation, genetic testing, and intestinal biopsy [46]. Diagnosing celiac disease presents several challenges that make timely and accurate detection difficult. One of the biggest hurdles is the reliance on invasive biopsies,

which are usually needed to endorse the diagnosis by examining the extent of intestinal damage. This procedure can be uncomfortable for patients and may prevent timely testing. Additionally, symptom variability poses a significant challenge[47]. Celiac disease presents differently among patients, ranging from gastrointestinal symptoms to extra-intestinal symptoms such as anemia and dermatitis herpetiformis[48] . This variability can lead to misdiagnosis or delayed diagnosis due to overlap of symptoms with other conditions. Delayed diagnosis caused to malnutrition and growth retardation. Mood disorders such as irritability and introversion may accompany these symptoms [49]. Early diagnosis of CD is crucial to catch growth in children and prevent long-lasting complications. Moreover, they are more likely to develop certain type of the cancer in delayed diagnosis. These are oropharyngeal and esophageal cancers, T-cell and B-cell non-Hodgkin lymphomas, small intestine and colon cancers, pancreatic cancer and hepatobiliary cancers [47]. Therefore, it leads to the emergence of much more serious pictures than the decrease in the quality of life [50].

CD patients frequently exhibit distinct facial characteristics such as pale complexion due to anemia, mouth (dermatitis herpetiformis), itchy rashes with small blisters around the hairline and cracking at the corners of the lips (cheilitis). They also have dental enamel defects like discoloration, pitting, or grooving of teeth, facial edema that is swelling around the eyes, dark circles around the eyes from fatigue and nutrient deficiencies, and a gaunt appearance with reduced facial fat and a hollow look due to weight loss [51,53]. Facial recognition models can identify subtle signs that may be overlooked in other conditions, such as facial pallor, dermatitis herpetiformis, cheilitis, facial edema, dark circles and weight loss.

In addition, Celiac disease mainly causes damage to the small intestinal mucosa, causing villus atrophy [54]. Knowing the level of intestinal atrophy (marsh level) provides several important benefits in the preliminary diagnosis phase. It allows diagnosis of the disease in its early stages. Intestinal atrophy is an important indicator of celiac disease [55]. In this way, patients can quickly start a gluten-free diet, relieving symptoms and improving their quality of life. Determining the level of atrophy helps assess the severity of the disease and create appropriate treatment plans. Treatment approaches can be determined according to mild, moderate or severe atrophy [24].

## 2.2. INNOVATIVE APPROACHES TO CLINICAL DIAGNOSIS OF CELIAC

This research lay foundation for implementing a system that combines face image classification and clinical data analysis for intestinal atrophy identification in CD. It could significantly

increase the diagnostic accuracy of celiac disease. By integrating these visual cues with clinical data, transfer and machine learning algorithms can more reliably identify patterns suggestive of celiac disease. This combined approach can speed up timely specialist referral by flagging potential cases earlier in the diagnostic process, reducing reliance on invasive biopsies and solving the problem of symptom variability. Such a system would not only streamline the path to diagnosis, but also improve patient outcomes by facilitating earlier and more accurate detection.

We aim to develop an integrated system that combines face image classification with clinical data to improve early detection of celiac disease, allowing rapid intervention. Timely and accurate diagnosis leads to rapid improvement of symptoms and improved quality of life, while continuous monitoring ensures the effectiveness and compliance of treatment. This proactive management reduces hospitalizations, lowers health care costs, and provides better long-term health outcomes; It highlights the important benefits of early and definitive medical intervention.

The future decision-support system, which utilizes integrated clinical and facial analysis data, has the potential to be used to guide treatment and management as well. For example, it can provide guidance on referrals to general practitioners, gastroenterologists or other specialists. By continuous analysis, it can identify symptoms that indicate celiac disease. When it detects significant risk factors or symptoms, it can generate alerts and recommendations indicating the need for expert evaluation. This approach allows potential celiac disease patients to be referred to advanced diagnostic procedures (endoscopy, blood tests) for early detection and intervention. It also may play a role in monitoring the effectiveness of the treatment and the patient's adaptation to the gluten-free diet. By analyzing facial and clinical data, it can track changes in patients' symptoms and physical indicators. For example, symptoms such as facial pallor, reduction in dermatitis herpetiformis, and weight stabilization may be signs of adherence to a gluten-free diet. Persistent or worsening symptoms may indicate noncompliance or the need for dietary adjustment.

## *2.3.* HARNESSING AI IN CELIAC DISEASE

### 2.3.1. Facial Recognition Technology

In recent years, interest in facial recognition technologies has increased day by day. Facial recognition technology is increasingly used in medical diagnostics to identify signs of disease by analyzing facial features [56]. This non-invasive method involves capturing high-resolution facial images, standardizing them, and using algorithms to extract and analyze key features such as color and texture[16]. By comparing these features to known disease models, machine learning models provide preliminary diagnoses or flag potential health problems[14]. Applications include detecting endocrine disorders (e.g., Cushing's syndrome), genetic disorders (e.g., Down syndrome), neurological conditions (e.g., Parkinson's disease), endocrine disorders (e.g., Cushing's syndrome), skin conditions, nutritional deficiencies, and infectious diseases[57]. While this technology requires strict data confidentiality, accuracy, and clinical validation, it offers early detection, continuous monitoring through mobile applications, and cost-effective screening. As a result of a comprehensive literature review, no study was found to design a system that helps diagnose celiac disease by classifying celiac patient photographs. There is also a lack of research focusing on diagnostic assistance systems that aid diagnosis or aim to provide primary care physicians with guidance on referrals to specialists. Additionally, there are no studies using transfer learning.

### 2.3.2. Deep Learning and Machine Learning

Machine learning (ML) is a field where algorithms can make easy predictions, based on the machine's capability to learn through large scale data exposure. Categorically, machine learning is classified based on learning style and function similarity. The three primary categories encompass unsupervised, supervised, and reinforcement learning [58]. In supervised learning, algorithms learn from human-provided training data, including labels for each input element and feedback to establish relationships between input and output data. Once trained, the algorithm is applied to new data for accurate predictions. Supervised learning broadly falls into two categories: classification and regression [58]. Classification algorithms are essential in this domain such as Random Forest, Decision Trees, Naïve Bayes, Support Vector Machines, Bayesian networks, Hidden Markov Models, and Neural Networks, [59]. Neural Networks, mimicking the human brain, consist of three layers: input, hidden (computational), and output [60]. Deep learning (DL), in the sense of deep network, is a subset of ML that utilize neural networks with many layers. Unlike task-specific algorithms, it focuses on learning data

representations. Deep learning is essentially an advanced form of neural networks [61]. The primary distinction lies in the depth and complexity of the network. In Neural Networks, there is input layer, hidden layer(s) and output layer. In Deep Learning, there is an input layer, many hidden layers (deep architecture) and output layer [62]. In short, deep learning works with a deep architecture consisting of many neural networks. Extending from Neural Networks, Deep Neural Networks exhibit deep learning capabilities, progressively extracting complex features for desired output at each layer. CNN is a discriminative supervised model trained using backpropagation and gradient descent [63, 64]. Recent advances in CNN architecture prioritize efficient resource usage by taking into account factors such as memory, power budget, and execution speed. Notably MobileNetV1 and MobileNetV2, developed by Google Inc., are architectures tailored for vision applications on mobile devices [65]. Research conducted in recent years highlights VGG16. They also mentioned effectiveness of MobileNetV2 as a feature extractor for object detection, image classification and segmentation. [27].

### 2.3.3. Background Research and Related Works

Many studies have previously been carried out on developing systems to help diagnose celiac disease. Gadermayr and friends [23] provides a conspectus of current advances in computer-based diagnosis of CD by utilizing upper intestinal endoscopy. They propose a frame to attain completely patient-specific diagnosis and comprises expert knowledge into automated decision-making system. The authors showed a feature descriptor adapted for the classification of video capsule endoscopy images in their work [24]. They also presented a system that uses deep CNNs to qualify small intestinal movement.

Lately, DL methods especially CNN used GoogLeNet, AlexNet, and VGG16 to classify endoscopic images. One of the study found a DL-based method that aims to recalibrate modules to differentiate images containing important regions for CD from healthy individuals [25]. They used k-nearest neighbor modules and support Vector Machine for linear discriminant analysis. And they accomplished fascinating results of 95.94% accuracy, 97.20% sensitivity, and 95.63% specificity using a 10x 10-fold cross-validation strategy. In addition to this, Amirkhani and friends [26]proposed a method by utilizing the fuzzy cognitive map (FCM) and a probabilistic fuzzy C-means clustering algorithm to distinguish CD cases. Objectives of this research wanted to develop an decision-support system that could classify CD patients into three groups: A, B1, and B2, demonstrating the newest classification method. They obtained accuracy rates of 88%,

90%, and 91% for the respective analysis classes. Their results disclosed an fascinating average classification accuracy of 96% for six different intestinal movement events.

Furthermore, valuable studies on automatic detection and diagnosis of celiac disease using computer-aided techniques have contributed to this field. The use of computer-assisted techniques, particularly those leveraging artificial intelligence and machine learning, holds significant promise in improving the diagnosis of celiac disease. One research conducted to focuses on automatic identification of CD by usage ML algorithms applied to capsule endoscopy images. Their algorithm achieved 94.1% accuracy to identify subtle signs of villous atrophy that may not be obvious by visual inspection alone. There are many other research about biopsy image analysis with deep learning [66]. One of them investigated the use of a DL model to determine CD in duodenal biopsy slides. Their model achieved high accuracy rates with a significant area under the receiver operating characteristic curve greater than 0.95 for all classes, indicating excellent classification performance [67]. One further study used discrete wavelet transform and nonlinear features to analyze video capsule endoscopy images. Their system achieved an accuracy of 86.47% with 10-fold cross-validation; They suggests that such image analysis techniques can significantly improve diagnostic performance and reduce misdiagnoses. CNNs also used for Endoscopy Images analysis in many studies [68]. In a study, researchers developed a convolutional neural network algorithm that can detect villous atrophy in duodenal endoscopy images with 99.30% sensitivity. Incorporating such AI algorithms into endoscopic equipment could greatly increase detection rates of celiac disease during routine examinations. When we look at early research on automated classification for celiac disease, we can see some research has been done in the past [69]. The initial study has been done to demonstrated the potential of computer-aided classification of duodenal images using various texture classification techniques and Fourier feature vectors. Their work highlighted the high specificity of these methods in detecting mucosal damage characteristic of untreated celiac disease [70].Machine Learning is also one of the commonly used tools for biopsy image interpretation.

Furthermore, in other research they explored the use of ML to automate the analysis of biopsy images. Their technique achieved high accuracy in classifying villous abnormalities, demonstrating the potential to aid pathologists in the process of diagnosis and facilitate early treatment of celiac disease patients. There is impressive real-time diagnosis with confocal endomicroscopy [71]. In this area, they introduced a method for automatic image classification

of CD using confocal endomicroscopy. Their computer-based method demonstrated accuracy comparable to human observers, demonstrating the potential for real-time, noninvasive diagnosis during endoscopy procedures [72]. Other studies in this field are that artificial intelligence based clinical decision support systems.

A study in this field, with a concept closest to our study, was conducted in 2011. They developed a model that classifies all signs and symptoms of CD with ML algorithms, without blood tests or facial images. They developed a clinical decision support system (CDSS) integrated with a Bayesian classifier that achieved 80% accuracy during training and 84.2% accuracy during validation. This CDSS showed good agreement with physicians' diagnoses, highlighting its capabilities as a valuable tool in clinics [73]. Collectively, these studies highlight the promising role of ML in improving the diagnosis process and governing of CD, supporting the transition to precision medicine.

Other studies that used ML methodologies to advance the diagnosis of CD using are as follows.

Piccialli and friends [74]used a variety of ML techniques, including Random Forests, Extremely Random Trees, Boosted Trees, and Logistic Regression, in their study to foresee the potential progression of celiac disease (PCD) in genetically predisposed patients with positive serology but no mucosal damage. All models achieved an accuracy of over 75%, with specificity scores exceeding 75%, and the best model, Boosted Trees, achieved an accuracy of 80% and a sensitivity and specificity of 58%. This predictive ability helps in the precise management of PCD patients.

Özer and Arda [75], in their study to demonstrate the potency of ML in standardizing and improving the diagnostic process for CD, compared different ML algorithms to estimate the diagnostic certainty parameters of serological tests for CD and compared the effectiveness of decision trees that achieved 88.7% accuracy. he emphasized.

In contrast, Hujoel and friends [76] attempted to identify undiagnosed CD cases using ML models based on possibilities of accepted indicators for testing. Despite using a variety of approaches, including logistic regression, random forest, and support vector machines, the models performed poorly, with AUCs ranging from 0.49 to 0.53, indicating the need for further improvement.

Collectively, these studies highlight the promising role of ML in improving CD diagnosis and management, supporting the transition to precision medicine. However, our understanding is

that there is an absence of improvement in clinical decision support systems particularly designed for CD diagnosis by using patient face images based on non-invasive diagnostic tests to reduce relying on biopsies. This gap in this field is likely due to the paucity of sweeping data and the complex structure of the diagnostic challenge.

As 2nd stage of this thesis focused on celiac marsh level classification using deep and machine learning, we would like to give some space for related works in this area. The following studies have been conducted to determine the marsh level of celiac disease. They proposed a fully automated histological classification system that provides an effective and objective method to characterize histological features related to CD severity, associated with Marsh scores [77]. Celiac marsh level classification is tried to be detected with many other laboratory techniques rather than AI tools.

Many studies showed that clinical and laboratory correlates of marsh classification. Studies by Bekdaş and friends [55] ,and Kalhan and friends [78] highlight significant associations between Marsh classification and clinical manifestations. Patients with higher Marsh grades exhibit more severe small intestinal mucosal injury, correlating with increased gastrointestinal symptoms and higher levels of tTG IgA seropositivity. These findings underscore the clinical relevance of Marsh classification in predicting disease severity and guiding management strategies. Furthermore, serological markers, particularly tissue transglutaminase antibodies (tTG), have been extensively studied in relation to Marsh classification. One of the study demonstrated a strong relation among the serum tTG levels and the degree of upper intestinal damage, suggesting that elevated tTG titers can predict mucosal atrophy without the need for biopsy in certain cases [79]. Similarly, Kaur and friends and Punia and friends [80] reported high sensitivity and specificity of tTG antibodies in identifying Marsh grade ≥2 lesions, supporting their role in non-invasive diagnosis approaches. In addition, when considering global perspectives and clinical implications for celiac marsh level classification, some studies appeared. Studies across diverse populations, including North India [78] and [81], highlight variations in clinical presentation and diagnostic accuracy based on Marsh classification. These findings underscore the global applicability of Marsh grading in assessing disease severity and tailoring management strategies.

In conclusion, the Marsh classification remains important in the diagnosis and treatment of CD by integrating histopathological evaluation with serological markers to guide clinical decisions.

In the second stage of our study, a machine learning model was designed to determine the marsh level with some blood tests and demographic information that are not difficult to access. We believe that future research may improve outcomes for individuals affected by this complex autoimmune condition by further improving diagnostic algorithms and therapeutic approaches.

## 2.4. TRANSFER LEARNING

In the context of Deep Neural Networks, training a whole CNN from scratch is rare with big enough dataset. Deep Neural Network with a large number of parameters have difficulties to be trained on small datasets, affecting generalization ability. It often results in overfitting [82]. TL has revolutionized facial recognition technology by leveraging pre-trained models to enhance performance in specific applications [83]. When we look at the history of transfer learning, its official definition, first expressed by Pan and Yang, includes the concepts of domain and task. In this framework, a field encompasses a feature space X represented as X = {x1, ..., xn} ∈ X and the marginal probability distribution P(X). Given a given field, D = {X is denoted as P(X)}, a task is denoted as T = {Y, f(·)}, where Y is a label space and f(·) is an objective prediction function. Learning a task consists of pairs {xi, yi}, where xi ∈ X and yi ∈ Y. The goal of TL with a source domain DS and learning task TS and a target domain DT and learning task TT is to improve the learning of the target prediction function fT(·) in DT by exploiting the information obtained from DS and TS [84]. In this approach, knowledge obtained from solution of one problem is transferred and implemented to a different but associated problem. In facial image classification and recognition, TL permits models trained on big datasets of general facial features to be fine-tuned for specialized tasks, such as detecting specific facial characteristics indicator of diseases [85]. By starting with a pre-trained model (e.g., on a dataset of general facial images), and fine-tuning it with smaller, disease-specific datasets, transfer learning improves accuracy and efficiency. This method significantly reduces the amount of labeled data and computational resources needed for training, making it particularly valuable in medical diagnostics where datasets may be limited [15]. Transfer learning thus plays a crucial role in advancing the capabilities of facial recognition technology for precise and reliable disease diagnosis and monitoring [85].

This research firstly goal to explore whether TL algorithms can accomplish adequate accuracy in classification facial images of CD to aid primary care physicians in making referrals to specialists for the diagnosis of the disease. TL was used with VGG16 and MobileNetV2

architecture for classification of DS_Face dataset. The architects designed for both models are different.

### 2.4.1. Designing VGG16 Model with Transfer Learning

VGG16 is a model originally trained on large-scale image datasets such as ImageNet. It is famous for its simplicity and efficiency in extracting hierarchical features from images through its deep convolutional layers [86]. TL enables efficient adaptation and fine-tuning of the pre-trained VGG16 model with new data. This approach significantly reduces the time and computational resources required for training from scratch. VGG16 accelerates the learning process on the new task by using features learned from the original task. And due to its structure, it can generally achieve good performance with less data. (**Figure 2.1**)[87].

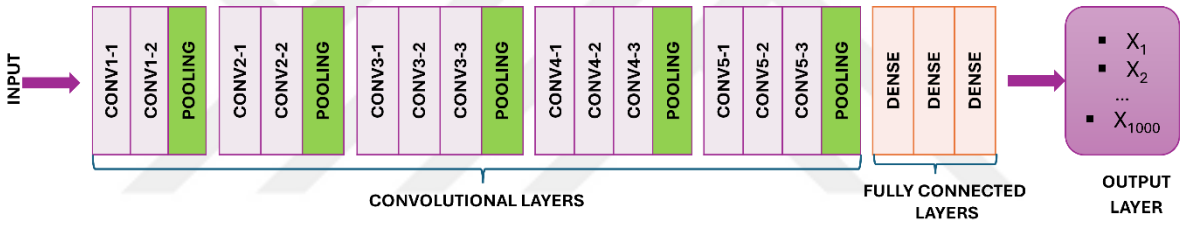



Figure 2.1. The VGG16 Architectural Structure.

VGG16 is a widely used convolutional neural network. The architecture of VGG16 has 16 layers, including 13 convolutional layers and 3 fully connected layers. It uses 3x3 convolutional filters to process data consistently. VGG16's maximum pooling layers help reduce dimensions and increase the robustness of the model. The architectural network ends with fully connected layers for decision making and a SoftMax layer for classification. On the other hand, although successful, VGG16's complexity of approximately 138 million parameters poses computational challenges. To exploit VGG16, understanding its architectural structure and elements is crucial in deep learning research and applications [14]. VGG16, which transfer learning uses, speeds up the training process by building on underlying logic built into pre-existing models and requires less data for retraining [88].

In our study, VGG16's architecture was used to create a new dataset from input images. Specifically, the model's convolutional and pooling layers were imported and the "upper part" containing the fully connected layer was skipped. Images were passed through these layers of

the VGG16 model to obtain a feature stack representing the detected visual features. This 3D feature stack was then converted into a NumPy array, allowing seamless integration with various modeling techniques (**Figure 2.2**)[89]

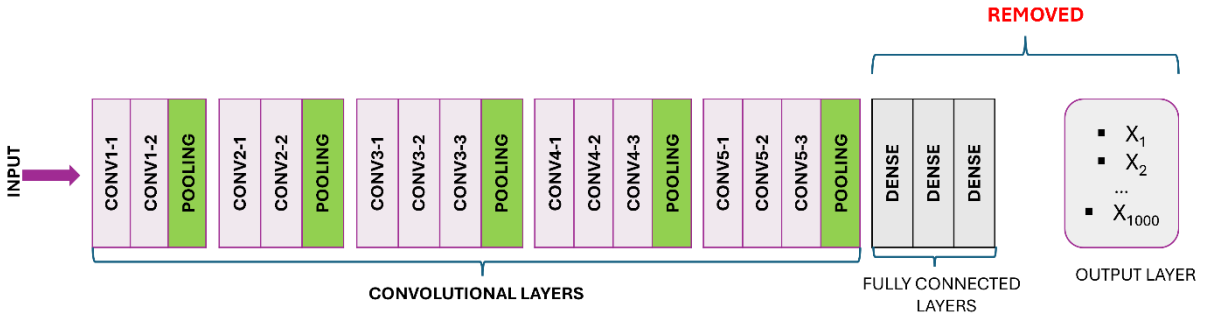



Figure 2.2. Removing the "Top" part, VGG16 uses pre-trained output layers to convolve the data

VGG16's pre-trained layers produced transformed visual features by processing the image dataset through its convolutional layers and weights. In this way, VGG16 was used for feature extraction. These pre-trained layers were not trained directly when classifying the DS_Face dataset. Frozen, pre-trained layers maintain standard convolution of visual features, while unfrozen (or 'trainable') pre-trained layers are adjusted to fit our specific dataset by updating based on predictions from the fully connected layer. (**Figure 2.3**) [87]. At the end, the final convolutional and pooling layers are unfrozen to allow training, and a fully connected layer is added for both training and prediction.

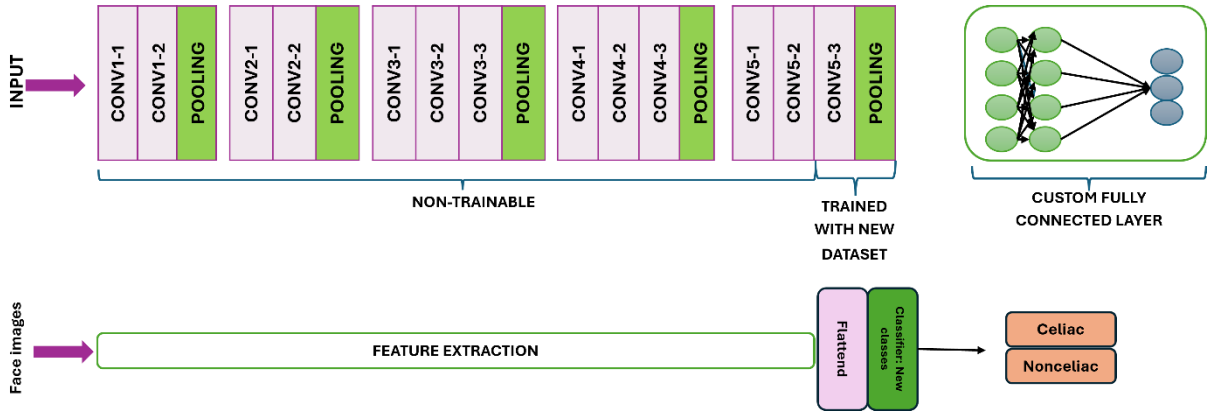



Figure 2.3. Fully Connected layer for both training and prediction.

The VGG16 model removes the last few layers responsible for sorting data into specific categories and retrains those layers using new data. To further improve performance, fine-

tuning is applied to the upper layers of the underlying network to optimize the performance of the model by incorporating special features[86]. When using a pre-trained model such as VGG16, only additional classifier layers are trained. This method offers several advantages. First, using the efficiency of TL. It utilizes pre-trained weights to extract general features and improves the efficiency of transfer learning. Secondly, it avoids overfitting. Freezing pre-trained layers improves generalization and minimizes the risk of overfitting to smaller data sets. The third and final one ensures that the training is fast. It reduces overall training time and computational costs by focusing resources on new layers. As a result, simply tuning additional classifier layers rather than retraining the entire model saves time and requires fewer computational resources.

The reason for choosing the VGG16 model for CD face image classification is its proven success in image classification tasks and its deep architecture that captures complex features in images. At the same time, using VGG16's pre-trained weights on the ImageNet dataset provides a solid foundation for transfer learning with a smaller dataset. This approach significantly reduces training time and computational resources while improving the model's ability to generalize from a relatively small dataset by leveraging features learned from extensive and diverse images.

VGG16's built-in layers are deployed for feature extraction parts. In the conceptual model of VGG16 (**Figure 2.4**), the centipede image dataset was processed using convolutional layers and the weights of these layers, and transformed visual features were created. These stored layers were directly classified; instead, it is preserved to be used to preserve standard visual feature convolution. However, non-frozen (or 'trainable') pre-generated layers have been updated with feedback from fully connected layers and adapted to make predictions suitable for celiac and non-celiac classes.

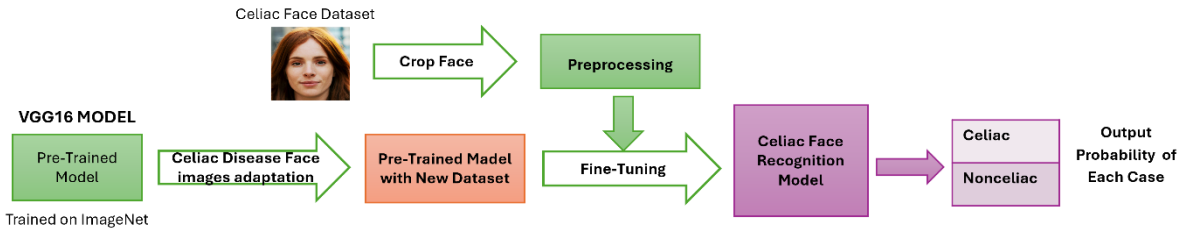



Figure 2.4. Conceptual VGG16 Model

### 2.4.2. Designing MobileNetV2 Model with Transfer Learning

At this stage, Mediapipe model-marker (MMM) was used with pre-trained model MobileNetv2. MMM is a tool developed by Google and used to create computer vision applications [83]. TensorFlow Lite offers ease of use by providing a high-level API to create models quickly and easily. It supports data types such as image classification, object detection, pose estimation and segmentation [84]. It utilizes pre-trained models using transfer learning techniques, so faster and more effective models can be trained with less data. MMM provides high performance on mobile and edge devices with TensorFlow Lite's optimization techniques [85]. It can be easily integrated with Google's other tools and platforms, and is especially suitable for use in Android and iOS applications. Model Maker is a powerful tool for developers who want to create machine learning models quickly and effectively, offering rapid prototyping, high performance, and flexibility [86]. However, it may require knowledge of machine learning and TensorFlow Lite, and data preparation may be time-consuming. MMM empowers users with versatile adaptability by operating across a spectrum of ML applications, including object detection, gesture recognition, and classification tasks for various data modalities such as images, text, and audio [87]. Through the TL technique, by leveraging the underlying logic already built into pre-trained models, MMM speeds up the training process and requires less data for effective retraining [83]. The core mechanism of MMM involves the strategic removal of the final layers responsible for categorizing data into distinct classes, followed by their reconstruction using the provided dataset (**Figure 2.5**) [88]. Furthermore, the tool offers advanced options for fine-tuning model layers, enabling users to enhance both accuracy and performance according to specific requirements.

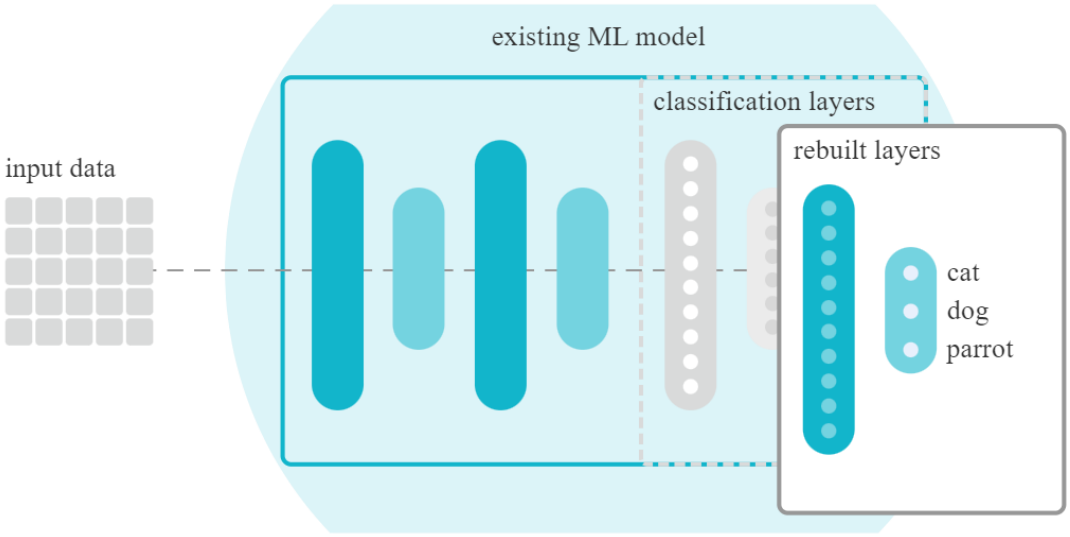



Figure 2.5. The model retrained and recreated by Model Maker [90].

In this work, second pretrained model is MobileNetV2 which used to classify celiac patient face photographs. MMM uses a transfer learning approach based on MobileNetV2 model that is a well-known pre-trained model with its mobile vision efficiency and fine-tunes it to classify celiac disease face images [91, 92]. In this process, the last layers of the model are replaced with new layers suitable for the new data set, resulting in a compact and special model. This approach provides more efficient inference and better performance in recognizing certain categories in celiac disease images [93]. MMM used a transfer learning approach to customize existing machine learning models for the celiac dataset. MMM fine-tuned CD face image classification using MobileNetV2 (**Figure 2.6**). A compact and specific model was obtained by replacing the last layers with new layers suitable for the dataset [90]. This enables efficient inference and better performance in recognizing specific categories in CD images [93].

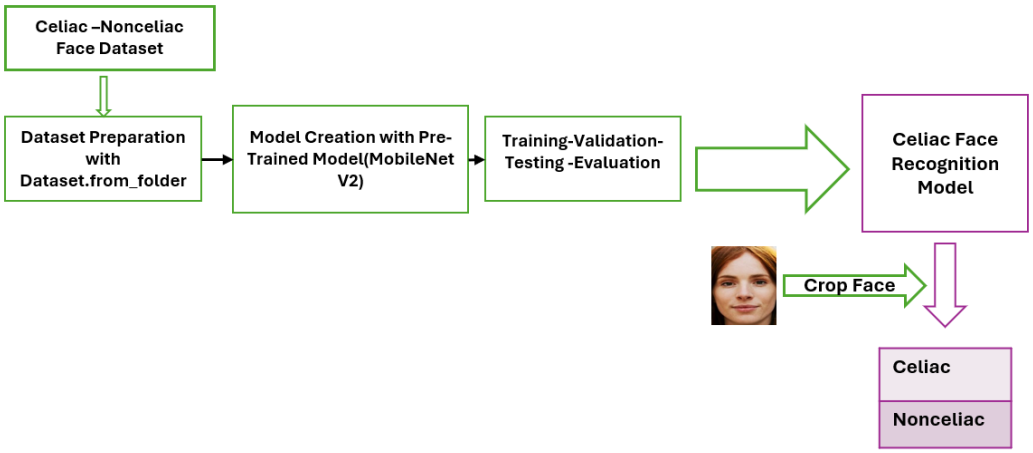



Figure 2.6. Conceptual MobileNetV2 model

## 2.5. DESIGNING OF ML MODEL FOR CELIAC MARSH LEVEL CLASSIFICATIONS

Diagnosis of CD involves assessing the severity of intestinal mucosal damage, which is categorized using the Marsh classification system [94]. Traditional methods for interpreting biopsy images and classifying Marsh levels rely heavily on manual analysis; this can be time wasting and tend be to subjectivity [41]. Recent advances in ML techniques offer promising ways to automate biopsy image analysis and improve the accuracy of Marsh level detection on CD. There are many studies investigating the application of ML algorithms to this task, aiming to facilitate the diagnostic process and improve patient care [23]. In the second stage, this study used both conventional ML and deep learning DL to predict CD Marsh levels using clinical data. Decision trees, logistic regression, random forests, and support vector machines are interpretable and effective traditional ML methods for smaller data sets but can be challenged with complex models [95]. In contrast, DL models such as CNN, artificial neural networks, recurrent neural networks, and autoencoders, although they require significant computational resources, can be used on larger datasets without extensive feature engineering [96]. It is successful in capturing complex relationships, but a large data set is required. A structured approach (**Figure 2.7**) is followed for celiac marsh level classification: It starts with data loading and preparation, where the dataset is loaded, missing values are handled, numerical features are normalized, and target labels are encoded. Pattern definition then requires designing

a neural network architecture (Pytorch, Decision Tree, random forests, Gradient Boosting and Naïve bayes) adapted to capture complex patterns in celiac data. Model Training then involves optimizing the neural network parameters using prepared data to minimize prediction errors. Finally, last step evaluates the performance of model on the entire dataset to determine its effectiveness in accurately classifying swamp levels. This systematic framework combines rigorous data processing, advanced model design, rigorous training, and comprehensive evaluation to develop a robust classification system for marsh levels.

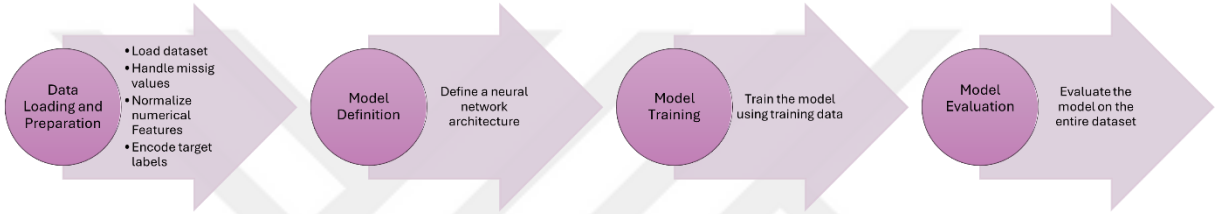



Figure 2.7. Conceptual Model for marsh level classification

# 3. METHOD

The study was completed in two stages. In the first stage, facial images of celiac patients were classified using the transfer learning method. In this process, transfer learning was applied using two different pre-trained models (VGG16 and MobileNetV2) on DS_Face dataset and the model with higher classification performance was selected. In the second stage, clinical signs and symptoms dataset (DS_Marsh,) were classified with deep and traditional machine learning algorithms to determine the Marsh level of CD, and the model with a high accuracy rate was selected. The most successful models were selected in both stages. While the same dataset (DS_Face) was used for two pre-trained models, VGG16 and MobileNetV2, for classifying facial images, a different dataset (DS_Marsh) was used to determine the Marsh level. Because we could not obtain the clinical data of the patients whose facial images we obtained at the same time.

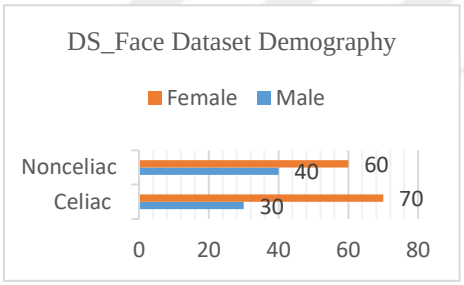
## 3.1. DATA ANONYMIZATION

To maintain the highest standards of privacy and data protection, we have implemented extensive measures to anonymize the dataset used in our research. Anonymization involves the transformation of personal data in such a way that individuals cannot be identified, directly or indirectly, by any means reasonable to use. We completed the process of anonymizing the dataset by removing direct identifiers and generalizing. Removal of Direct Identifiers: All direct identifiers such as names, addresses, social security numbers, and other unique identifiers have been removed from the dataset. Generalization: Specific data points have been replaced with broader categories to avoid identifying individuals. For example, exact ages are grouped into age ranges. By ensuring the anonymization of data, we fulfill our ethical obligations and increase trust in the individuals and communities represented in our research. The facial image dataset was anonymized by placing masks over the entire face to cover the original features.

## 3.2. CLASSIFICATION OF CELIAC PATIENT FACE IMAGES WITH TRANSFER LEARNING

### 3.2.1 Collection And Preparation of Datasets for VGG16 And MOBILENET2

To recognize the VGG16 model dataset, the DS_Face dataset contains two categories. Using transfer learning with VGG16 and MobileNetV2 achieves satisfactory performance even with small dataset. The DS_Face dataset consists of two groups, celiac and non-celiac, each containing 100 images. DS_Face dataset was obtained from Van Training and Research Hospital, Department of Gastroenterology (**Figure 3.1**). Newly diagnosed and previously diagnosed patients were included in the study. Among previously diagnosed patients, those who did not follow a gluten-free diet and those who followed a partial diet were included in the dataset. Information showing the dietary status of celiac patients is in (**Figure 3.2**).



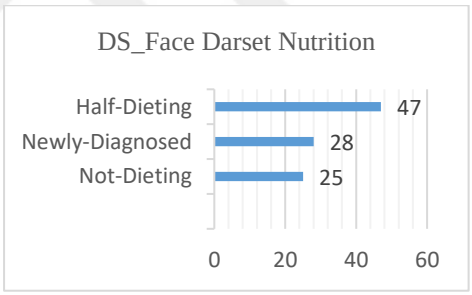




Figure 3.1. Number of Female and male participating

Figure 3.2. Diet status of CD patients

Patients included in the DS_Face dataset were informed by the specialist physician in accordance with the ethics committee rules. Then, the patients were asked to sign a volunteer form to participate in the study. Participants were seated in a fixed chair in front of a white wall. They were photographed using a Sony DSC-S2100 12.1 MP camera mounted on a tripod positioned 50 cm away. Their faces were shot only vertically, in front profile, making sure they were not wearing glasses or hats. The lens of the tripod was placed at the same height as the patient's face.

## 3.3. TRANSFER LEARNING WITH VGG16

### 3.3.1. VGG16 model working stages

The model was fine-tuned with VGG16 using the custom DS_Face dataset for image classification. The work phases respectively include image preprocessing, model configuration, adding custom layers, compilation, data augmentation, data generation, training, evaluation, visualization of results, and generation of a classification report. TensorFlow and Keras were used to develop and train the deep learning model. The Glob library was used to retrieve file paths from the dataset directories, while Matplotlib and Seaborn were used to plot the training/validation metrics and confusion matrix. Scikit-learn was used to generate classification reports and confusion matrices.

### 3.3.2. Preprocessing of Image

All facial images taken are original. First, the face images in the DS_Face dataset were captured at 1600x2000 pixels. They were then cropped to 400x600 pixels to focus on the face only. For transfer learning, pre-trained models resize images to 224x224 pixels, which is the input size required by VGG16. Training and validation dataset paths were specified. Since VGG16 was trained on ImageNet using the BGR format, the image color channels were reordered from RGB (Red, Green, Blue) to BGR (Blue, Green, Red).

### 3.3.3. Configuration of Model

The VGG16 model comes with pre-trained weights from the ImageNet dataset. The upper fully connected layers of VGG16 are excluded (include_top=False). All pre-trained layers are frozen to prevent their weights from updating during training, thus preserving the features learned from ImageNet.

### 3.3.4. Custom Layers Addition

To adapt the VGG16 model to our classification task, its 3D output was flattened into a 1D tensor using Flatten () (vgg.output). Two fully connected layers were then added. A hidden layer is with 1000 neurons and ReLU activation (Dense (1000, activation='relu')(x)). An output layer is consisting of neurons matching the number of classes, using softmax activation for classification as celiac and non-celiac categories (Intensive(len(folders), activation='softmax')(x)). The modified model was defined with VGG16 input and custom

output layers (Model (inputs=vgg.input, Outputs=prediction)) and summarized with model.summary().

### 3.3.5. Compilation of Model

For model compilation, categorical cross-entropy was used as the loss function, suitable for multi-class classification. Adam optimizer' known for its adaptive learning rate to improve the training process was used. Compilation parameters were set as shown below: model. compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy']).

### 3.3.6. Augmentation of Data

For data augmentation, the model was trained with a batch size of 32 by processing 32 samples before updating the internal parameters. The training took place over 10 periods; each epoch represented a complete pass through the entire training dataset. The learning rate implicitly determined by the Adam optimizer was 0.001. Data augmentation techniques were used to increase the robustness of the model and avoid overfitting, including rescaling pixel values by 1. /255, random truncation transformations within the range of 0.2, random zooming within the range of 0.2, and horizontal flipping of images.

### 3.3.7. Data Generation

Data generators were used to bulk generate images and labels directly from the directory. And efficient memory usage was also ensured. The training set was created by specifying the training path, 224x224 target size, batch size, and categorical class mode and using the train_datagen.flow_from_directory command. Similarly, the test set was created using the test_datagen.flow_from_directory command with validation path, 224x224 target size, batch size, and categorical class mode.

### 3.3.8. Training Cycle

For greater control and potential customization, a custom training cycle was used to iterate over periods and groups, as well as to manually manage the training and validation processes. In each epoch, the model was trained in batches using model.train_on_batch(X_batch, y_batch) to obtain the training loss and accuracy. Validation loss and accuracy were evaluated at the end of each epoch using model.evaluate(test_set,steps=validation_steps).

### 3.3.9. Metrics and Evaluation

The model.evaluate(test_set) function to calculate test loss and results was used to assess the model's performance on the test set using . For future use, the trained model was saved with the file name facefeatures_new_model.h5.

### 3.3.10. Visualization Step

Graphs showing training and validation losses and accuracy per epoch were created to visualize the training process. The training and validation loss was shown by the first subplot using plt. plot(epochs_range, training_losses, label='Training Loss') and plt.plot(epochs_range, validation_losses, label='Validation Loss'). The training and validation accuracy was displayed with the second subplot by usage of plt.plot(epochs_range, training_accuracies, label='Training Accuracy') and plt.plot(epochs_range, validation_accuracies, label='Validation Accuracy'). These plots were arranged side by side using plt.subplot and displayed with plt.show().

### 3.3.11. Classification Report and Confusion Matrix

Predictions were made on the test set using Model.predict(test_set,steps=validation_steps). And the predicted classes were determined by np.argmax(Y_pred, axis=1). For each class, a classification report was generated using the elaboration precision, recall, and F1 score with the command class_report(y_true, y_pred,target_names=target_names). To further evaluate classifier performance in distinguishing classes, a confusion matrix was created using confusion_matrix(y_true, y_pred). This matrix was visualized with a heatmap using sns.heatmap(conf_matrix, annot=True, fmt='d', xticklabels=target_names, yticklabels=target_names) and displayed with plt.show().

## 3.4. TRANSFER LEARNING WITH MOBILENETV2

### 3.4.1. Environment and Library Setup

Google Colab environment was used. And to facilitate access to larger datasets and model storage, access to the dataset uploaded to Google Drive was made easier by using, Drive.mount('/content/drive') code. Python version verified with !python --version and upgraded using the pip package manager !pip install --upgrade pip. Additionally, the mediapipe-model-maker library was installed with !pip install mediapipe-model-maker to ensure compatibility and take advantage of the latest features. Then, using the google.colab

import files, import os, import tensorflow as tf and asset tf.__version__.startswith('2') commands, the base libraries were imported and TensorFlow version 2.x was verified. TensorFlow 2.x offers significant usability and performance improvements over TensorFlow 1.x. MediaPipe Model Maker makes it easy to create and train machine learning models.

### 3.4.2. Uploading and Viewing Images

To verify that the DS_Face data was loaded correctly and to ensure an intuitive understanding of the dataset, images were loaded from the specified directory and several samples for each label were displayed using Matplotlib. The code started by defining the image path and extracting tag names from subdirectories within the specified path. For each tag, the code accessed the relevant directory and selected a predefined number of sample images (num_examples = 2). These images were then plotted using plt.subplots to create a shape with multiple subplots. Each image was read using plt.imread and displayed without axis labels for clarity. This visualization step, where each subgraph is named with its own label, ensures that the data is loaded correctly and provides a clear visual representation of the dataset, facilitating further processing and analysis.

### 3.4.3. Preparing the Data Set

The DS_Face dataset was structured using image_classifier. Dataset.from_folder(image_path) which creates a dataset from images stored in the specified folder. Then DS_Face was split into training, validation and test sets by usage of data.split(0.8) and remaining_data.split(0.5), respectively. This section is important as it facilitates effective training, hyperparameter tuning, and evaluation of the performance of model on unseen data. Training data was utilized to optimize the parameters of the model, while validation data helps fine-tune and optimize model performance. The test data set serves to evaluate the ability of model to generalize on completely DS_Face dataset, providing a reliable and unbiased assessment of the model's effectiveness across different scenarios. This systematic approach to dataset preparation enables robust model development and reliable performance evaluation in machine learning applications.

### 3.4.4. Model Properties and Hyperparameters

The MobileNetV2 model specification was chosen using image_classifier. SupportedModels.MOBILENET_V2 due to the optimal balance between accuracy and

efficiency, which is particularly advantageous for deployment on mobile and edge devices. Hyperparameters were set using image_classifier.HParams with import_dir=exported_model", ensuring that the trained model can be saved and accessed for future use. This approach allows efficient management and deployment of the image classifier model in line with the requirements of practical applications where resource efficiency and model portability are crucial.

### 3.4.5. Training the Model

The image classification model was trained using image_classifier. ImageClassifier.create() with the specified training and validation datasets (train_data and validation_data) and predefined options. Including a validation dataset during training is crucial as it allows continuous evaluation for performance of model on unseen data. This practice helps monitor the generalization ability of the model and facilitates early detection of overfitting, in which the model works well on training data but cannot make generalization to new examples. Leveraging the validation dataset, adjustments can be made to the model's parameters and architecture to optimize performance and ensure robustness across diverse datasets and real-world applications.

### 3.4.6. Model Evaluation

Model evaluation was done using model.evaluate(test_data) in Python, which calculates loss and accuracy metrics on the specified test dataset. The results, including test loss and accuracy, were then printed for analysis. This evaluation step is critical since it provides an objective and unbiased assessment for performance of the model on new data. The retained model was tested on a separate test set from the training and validation data used during model development. This ensured that the reported measurements accurately reflected the model's ability to generalize to real-world scenarios. This rigorous evaluation process is necessary to verify the effectiveness and reliability of the image classification model across different datasets and deployment environments.

### 3.4.7. Exporting and Downloading the Model

The trained model was exported to TensorFlow Lite format using model.export_model() and then downloaded using files.download('exported_model/model.tflite'). Exporting the model to the TensorFlow Lite format is crucial to optimizing its deployment on mobile and edge devices,

where computational resources are often limited. This format conversion ensures that the model can efficiently perform real-time inference while maintaining high accuracy. TensorFlow Lite makes the model lightweight and compatible with a variety of hardware platforms, facilitating seamless integration into applications requiring fast and effective image classification capabilities. This approach aligns with the practical considerations of deploying machine learning models in resource-constrained environments and increases accessibility and usability in a variety of deployment scenarios.

## 3.5. TRADITIONAL MACHINE LEARNING AND DEEP LEARNING FOR MARSH LEVEL DETECTION

### 3.5.1. Data Collection and Dataset Preparation

DS_Marsh dataset for CD marsh level classification was collected from Van Training and Research Hospital, Gastroenterology Department polyclinic. Patients were informed by a physician in adherence to ethical guidelines and provided voluntary consent through a signed form. The dataset includes 182 adult individuals presenting with clinical manifestations, symptoms, and blood test findings. Of these patients, 132 (73%) were female with an average age of 34 years (ranging from 18-65 years), while 50 (27%) were male with an average age of 32 years (ranging from 18-57 years) (**Figure 3.3**). These patients demonstrated Marsh levels spanning from Marsh classification '0' to '4'. Additionally, among them, there were those not adhering to a gluten-free diet numbering at72 individuals, whereas106 displayed partial adherence as indicated in (**Figure 3.4**).

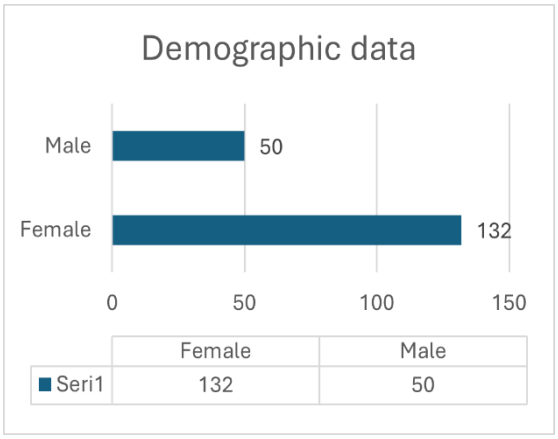



Figure 3.3. Celiac Male and Female

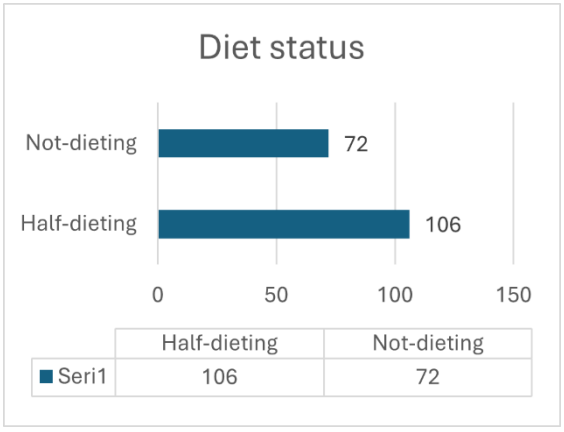



Figure 3.4. Diet status of celiac patients

Patients with suspected CD were included in the study, and data were collected on gender, presence of CD in the family, and Body Mass Index (BMI). Participants completed a questionnaire assessing gastrointestinal (GIS) symptoms, including diarrhea, reflux, abdominal pain, nausea, bloating, soft stool, urgent defecation, heartburn, and flatulence. Additionally, we documented several blood parameters such as Hematocrit (HCT), Hemoglobin (HGB), Vitamin D, B12, Folate, Ferritin, magnesium, serum iron, iron binding capacity, Immunoglobulin M (IGM), Immunoglobulin G (IGG), and Immunoglobulin A (IGA). Both GIS symptoms and serological tests are markedly abnormal in celiac patients compared to non-celiac individuals. Our approach aims to enhance diagnostic accuracy while minimizing the number of questions and tests, thereby saving time and resources.

### 3.5.2. Import Necessary Libraries

Jupyter notebook environment was used. Leveraged several key Python libraries to perform data processing, visualization, machine learning model training, and user interaction. Pandas library is used to load and manipulate data while numpy supports numeric operations and array operations. Matplotlib.pyplot creates plots such as confusion matrices to visualize data and results, and seaborn provides advanced statistical visualizations, although it is not used directly in this code. The Sklearn.preprocessing module includes LabelEncoder to convert categorical labels to numeric form and StandardScaler to normalize data. To handle missing data, SimpleImputer in sklearn.impute replaces missing values with column averages. Machine learning models used include DecisionTreeClassifier from sklearn.tree, RandomForestClassifier and GradientBoostingClassifier from sklearn.ensemble, and GaussianNB from sklearn.naive_bayes. These models are trained on data and their performance is evaluated using metrics such as accuracy_score, confusion_matrix, and classification_report from sklearn.metrics. Interactive elements are created using ipywidgets, which provide drop-down menus, buttons, and other input forms for user interaction, and IPython.display is used to display these widgets in the notebook. This integration enables comprehensive data analysis, model training, evaluation, and user interaction in a seamless and interactive manner.

### 3.5.3. Data Loading and Preparation

To prepare the dataset for analysis, we initially loaded it from a CSV file using the pandas library. Missing values were addressed by assigning them with the mean, to ensure data integrity. Subsequently, numerical features underwent normalization to maintain consistency in scale across the dataset, enhancing the efficacy of subsequent analyses. Following

normalization, features were segregated from the target variable, streamlining the modeling process. For categorical target variables, encoding techniques like Label Encoding were employed to convert them into numerical representations suitable for analysis. These meticulous preprocessing steps guarantee that the dataset is primed for rigorous analysis and modeling, laying a robust foundation for insightful conclusions and accurate predictions.

### 3.5.4. Model Definitions

To establish a comprehensive suite of classification models, leveraging diverse algorithms to explore the dataset's predictive potential. These models include PyTorch as deep learning neural network and Random Forest, Decision Tree, Gradient Boosting, and Naive Bayes classifiers, each offering unique strengths in handling various data characteristics.

### 3.5.5. Model Training

Each classification model on the preprocessed dataset were meticulously trained, harnessing their distinct algorithms and capabilities to extract meaningful patterns and relationships from the data. Following established best practices, we partitioned the dataset into training and validation subsets to facilitate unbiased model evaluation and performance assessment.

### 3.5.6. Model Evaluation

Each classification model's performance was assessed using accuracy calculations, confusion matrices, and classification reports. Accuracy measured the overall predictive success, while confusion matrices provided insights into specific classification outcomes. Classification reports detailed precision, recall, and F1-score metrics for each class, offering a comprehensive perspective for performance of the model. These evaluations guided the selection of the most effective model for real-world deployment, ensuring reliable predictions tailored to the dataset's nuances.

### 3.5.7. Prediction Interface

To provide a clear specification of the columns required for new data input, including their respective data types such as numerical, categorical, or textual. Following this, we developed user-friendly widgets tailored to each data type, ensuring intuitive interaction for users. These widgets, ranging from text boxes to dropdown menus, enable users to input values for each feature efficiently (**Figure 3.5**). This approach enhances the accessibility and usability of the

prediction system, accommodating diverse user preferences and facilitating seamless integration into various applications and workflows.

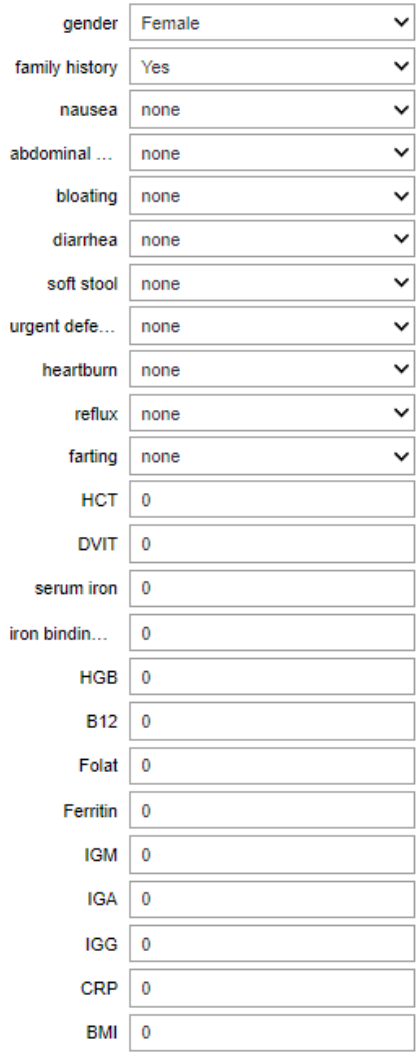



Figure 3.5. Prediction interface ranging from text boxes to dropdown menus

# 4. RESULTS

## 4.1. RESULTS OF CLASSIFICATION OF CELIAC PATIENT FACE IMAGES WITH VGG16

DS_Face dataset consists of two groups: celiac and non-celiac. Celiac patients are new and previously diagnosed patients. Among previously diagnosed patients, only patients who were not on a gluten diet and those who were on a partial gluten diet were included. VGG16 customized with the DS_Face dataset made over 70% correct classification. VGG16 was used as the fixed feature extractor. VGG16's stacked layers were trained on DS_Face datasets. The weights and biases in the model architecture together constitute the parameters. Parameters that remain unchanged throughout the training process are called untrainable parameters. List is shown below (**Table 4.1**) indicating the number of trainable and non-trainable parameters for the customized model.

Table 4.1. Trainable and Non-Trainable Parameters

| Model | Total Parameters | Trainable | Non-Trainable |
|---|---|---|---|
| Customized Model | 39,805,690 | 25,091,002 | 14,714,688 |

The customized model was trained and validated by adjusting the learning rate and number of epochs. The test and validation accuracies of the trained models are presented in (**Table 4.2**). Total Running Time: 484.59 seconds.

Table 4.2. Trained model’s testing and validation accuracies

| No. of Epoch | Learning rate | Testing Loss | Testing Accuracy | Validation Loss | Validation Accuracy |
|---|---|---|---|---|---|
| Epoch 1/10 | 15s 3s/step | 6.2137 | 0.5000 | 6.2114 | 0.5000 |
| Epoch 2/10 | 16s 3s/step | 2.0953 | 0.5217 | 2.1691 | 0.5100 |
| Epoch 3/10 | 15s 3s/step | 1.7656 | 0.5669 | 1.7505 | 0.5500 |
| Epoch 4/10 | 16s 3s/step | 0.8504 | 0.7587 | 0.9814 | 0.7300 |
| Epoch 5/10 | 15s 3s/step | 0.9370 | 0.7054 | 0.7952 | 0.7400 |

| | | | | | |
|---|---|---|---|---|---|
| Epoch 6/10 | 16s 3s/step | 0.6226 | 0.7319 | 0.6401 | 0.7100 |
| Epoch 7/10 | 15s 3s/step | 0.6660 | 0.7668 | 0.7453 | 0.7400 |
| Epoch 8/10 | 15s 3s/step | 0.5976 | 0.7720 | 0.6835 | 0.7400 |
| Epoch 9/10 | 20s 5s/step | 0.5129 | 0.7398 | 0.5192 | 0.7400 |
| Epoch 10/10 | 16s 3s/step | 0.6034 | 0.7305 | 0.6018 | 0.7300 |

It clearly shows that the accuracy rate increases as the number of epochs increases for the same network in (**Table 4.2**). Additionally, there is a noticeable relationship between training accuracy and learning rate.

Graphs are necessary to evaluate a model's performance during training. Graphs showing training loss, validation loss, training accuracy and validation accuracy are shown below (**Figure 4.1)**. The reduction in training loss indicates increased learning and efficiency with the training data, while validation loss reflects the ability of model to make generalization to unseen data. Training accuracy measures how well the model classifies or predicts training data and generally shows improvement over time. On the other hand, validation accuracy, indicates the performance of the model for generalization on unseen data; Consistent or increased validation accuracy throughout training suggests effective generalization and helps prevent overfitting.

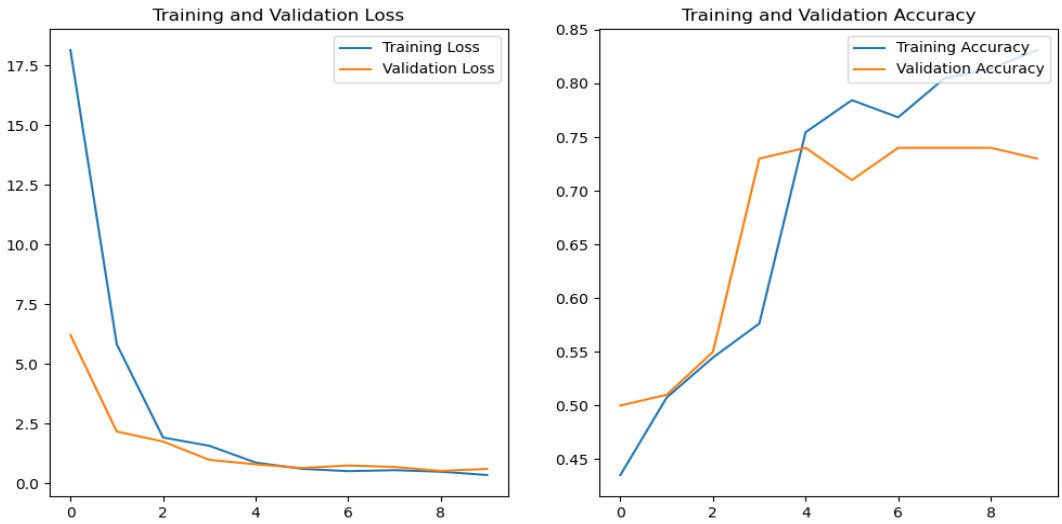


Figure 4.1. Displaying train loss and validation loss and train accuracy, validation accuracy versus number of epochs

It is shown below (**Table 4.3**) the performance evaluation of the classification model in detail. For evaluation, it was carried out using metrics such as precision, recall, F1 score and accuracy. The overall accuracy of the model slightly exceeded random chance, successfully distinguishing more than 70% of celiac patients from non-celiac patients. This evaluation was performed on an independent test set, allowing the model's performance to be evaluated with new, unseen data. Macro and weighted averages for precision, recall, and F1 score were consistent with measurements for individual classes and provided a comprehensive view of the model's performance across all categories. Although the model's performance is relatively balanced between the two classes, further improvements can be made, especially in increasing recall for both classes. These findings provide a basis for further refinement and optimization of the classification model to increase its accuracy in diagnosing CD.

Table 4.3. The performance Metrices of model

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Celiac | 0.57 | 0.56 | 0.52 | 50 |
| Nonceliac | 0.54 | 0.70 | 0.61 | 50 |
| Accuracy | | | 0.55 | 100 |
| Macro Avg | 0.55 | 0.55 | 0.54 | 100 |
| Weighted Avg | 0.55 | 0.55 | 0.54 | 100 |

## 4.2. RESULTS OF CLASSIFICATION OF CELIAC PATIENT FACE IMAGES WITH MOBİLENETV2

We used the Mediapipe model marker, which is a package of the Mediapipe library and used a previously trained MobilenetV2 model for classifying of the DS_Face dataset. The MobilenetV2 model accomplished a test accuracy of 0.85 with a test loss of 0.46 in 10 epochs for classifying CD face images. The model performed consistently; testing accuracy hovered around 0.85 across periods, while validation accuracy remained relatively constant with minor fluctuations. The training process was efficient, with the trainable parameters being 2,562 and the training time of each epoch varying from 7 to 13 seconds per step. The highest test accuracy of 0.8813 was reached in the 3rd epoch, indicating the best performance of the model. The lowest test loss occurred in period 3 with 0.3967, indicating the optimal fit of the model at this

stage. Verification accuracy remained consistently high at around 0.85 in the 10th epoch. The validation loss showed small changes, with the lowest value at 0.4866 in epoch 6 (**Figure 4.2**).

The overall results indicate strong performance with minor fluctuations in loss and accuracy measurements, reflecting the stability and reliability of the model in classifying CD face images.

In the first phase of the study, TL's MobileNetV2 was used as a fixed feature extractor with additional layers trained on DS_Face datasets. To improve network performance, certain underlying model layers were configured to be updated during training. As shown below (**Table 4.4**), in the neural network architecture, weights and biases entirely constitute parameters, and fixed parameters were classified as untrainable.

Table 4.4. Trainable and Non-Trainable Parameters

| Model | Total Parameters | Trainable | Non-Trainable Parameters |
|---|---|---|---|
| Customized Model, MobilenetV2 | 2260546 | 2562 | 2257984 |

A customized model suitable for the new task was trained and validated by adjusting the learning rate and number of epochs. The test and validation accuracies of the trained models are presented below (**Table 4.5**).

Table 4.5. Trained model's testing and validation accuracies.

| No. of Epoch | Learning rate | Testing Loss | Testing Accuracy | Validation Loss | Validation Accuracy |
|---|---|---|---|---|---|
| Epoch 1/10 | 13s 98ms/step | 0.4274 | 0.8625 | 0.4972 | 0.8000 |
| Epoch 2/10 | 10s 129ms/step | 0.4466 | 0.8250 | 0.4982 | 0.8000 |
| Epoch 3/10 | 8s 95ms/step | 0.3967 | 0.8813 | 0.5083 | 0.8500 |
| Epoch 4/10 | 9s 118ms/step | 0.4708 | 0.8062 | 0.4968 | 0.8500 |
| Epoch 5/10 | 8s 100ms/step | 0.4265 | 0.8438 | 0.5048 | 0.8500 |
| Epoch 6/10 | 7s 89ms/step | 0.4239 | 0.8750 | 0.4866 | 0.8500 |

| | | | | | |
|---|---|---|---|---|---|
| Epoch 7/10 | 10s 128ms/step | 0.4151 | 0.8625 | 0.4988 | 0.8500 |
| Epoch 8/10 | 8s 102ms/step | 0.5008 | 0.8125 | 0.5022 | 0.8500 |
| Epoch 9/10 | 10s 126ms/step | 0.4622 | 0.8125 | 0.4956 | 0.8500 |
| Epoch 10/10 | 8s 101ms/step | 0.4215 | 0.8625 | 0.5036 | 0.8500 |

As shown in the table above, as the number of epochs increases for the same network, the accuracy rate also increases. Additionally, a relationship between training accuracy and learning rate is clearly seen (**Figure 4.2**).

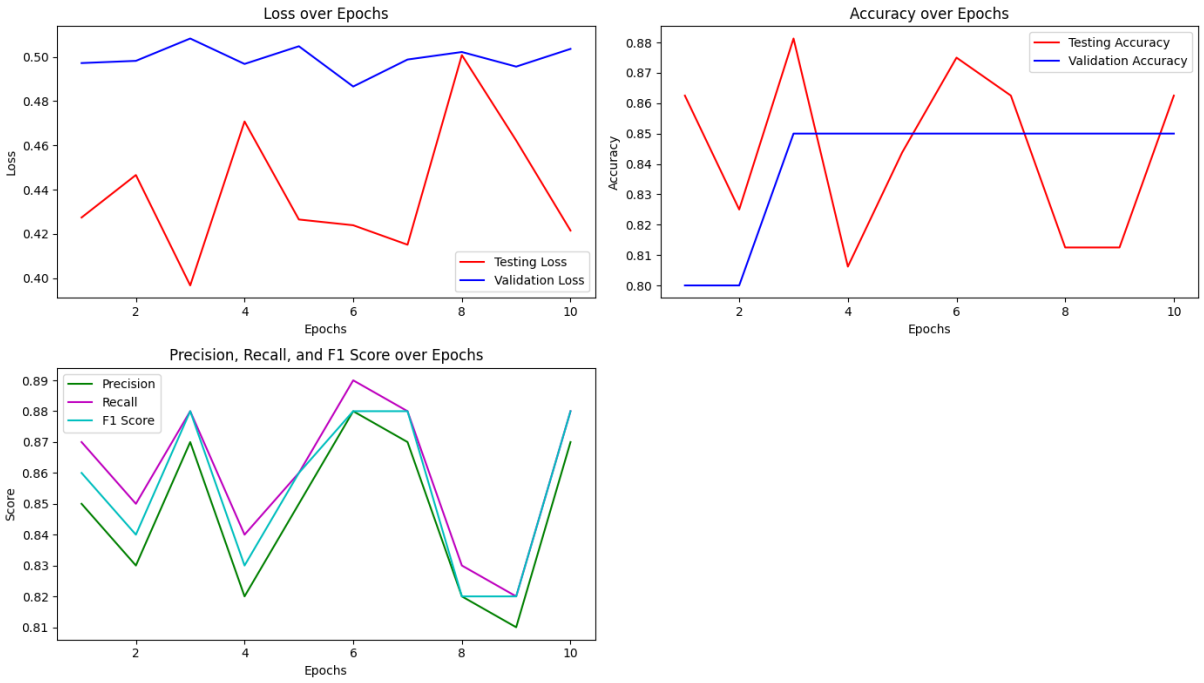



Figure 4.2. Displaying train and validation loss and train and validation accuracy versus number of epochs and precision, recall and F1 score.

The model yielded a recall of 0.80, precision of 0.80, and an F1 score of 0.80, indicating a balanced performance in recision and recall (**Figure 4.2**). These metrics suggest that the model is equally skilled at correctly classifying both true positives and minimizing false negatives, which is critical for maintaining reliability and accuracy in facial image classification tasks.

The exported model was tested by uploading 9 images with an application in Mediapipe Studio [97]. Model predicted uploaded images as celiac and non- celiac with percentage ratio. Prediction of model on Test images in Mediapipe Studio are as follows (**Figure 4.3, Figure 4.4, Figure 4.5, Figure 4.6, Figure 4.7, Figure 4.8, Figure 4.9, Figure 4.10, Figure 4.11**).

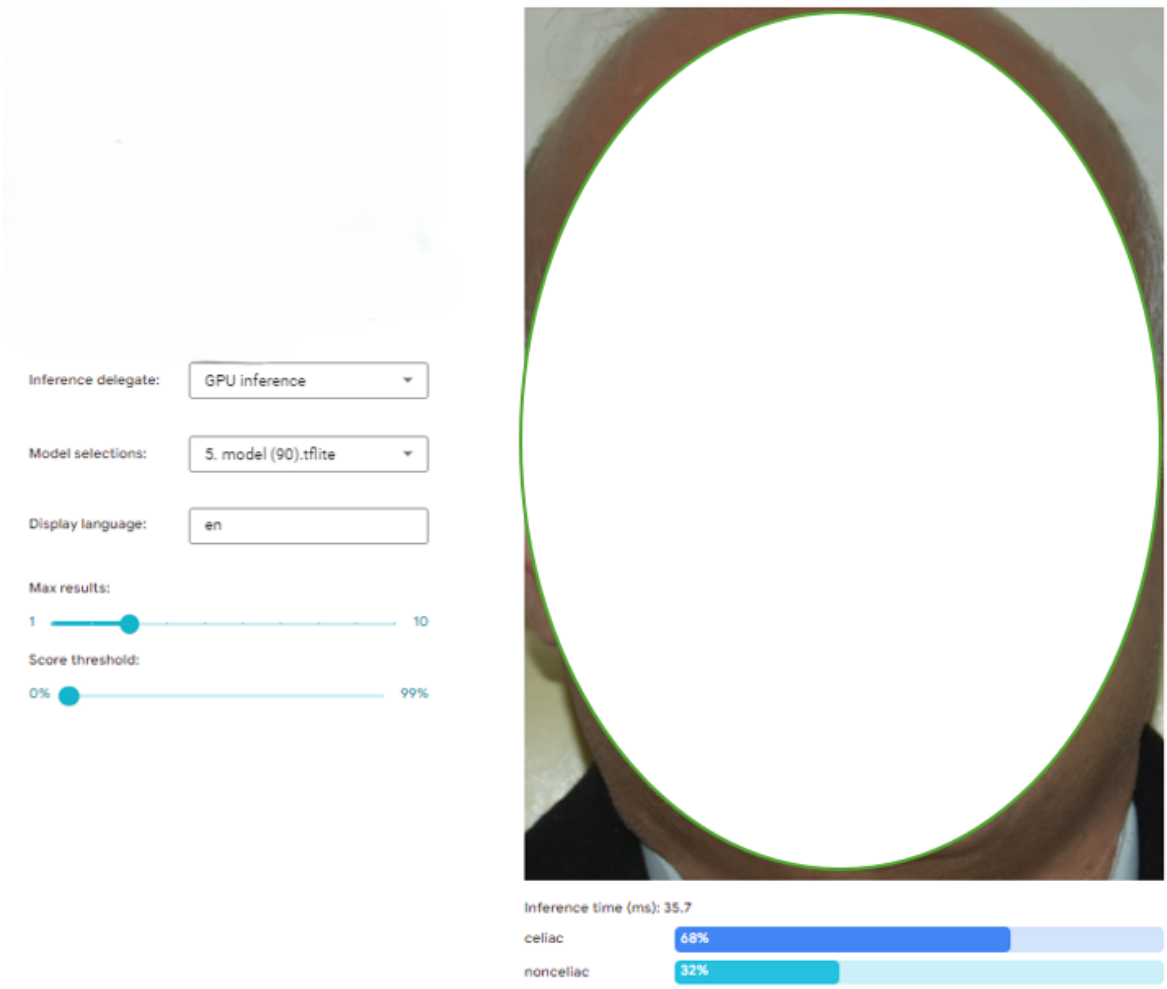



Figure 4.3 First Test image

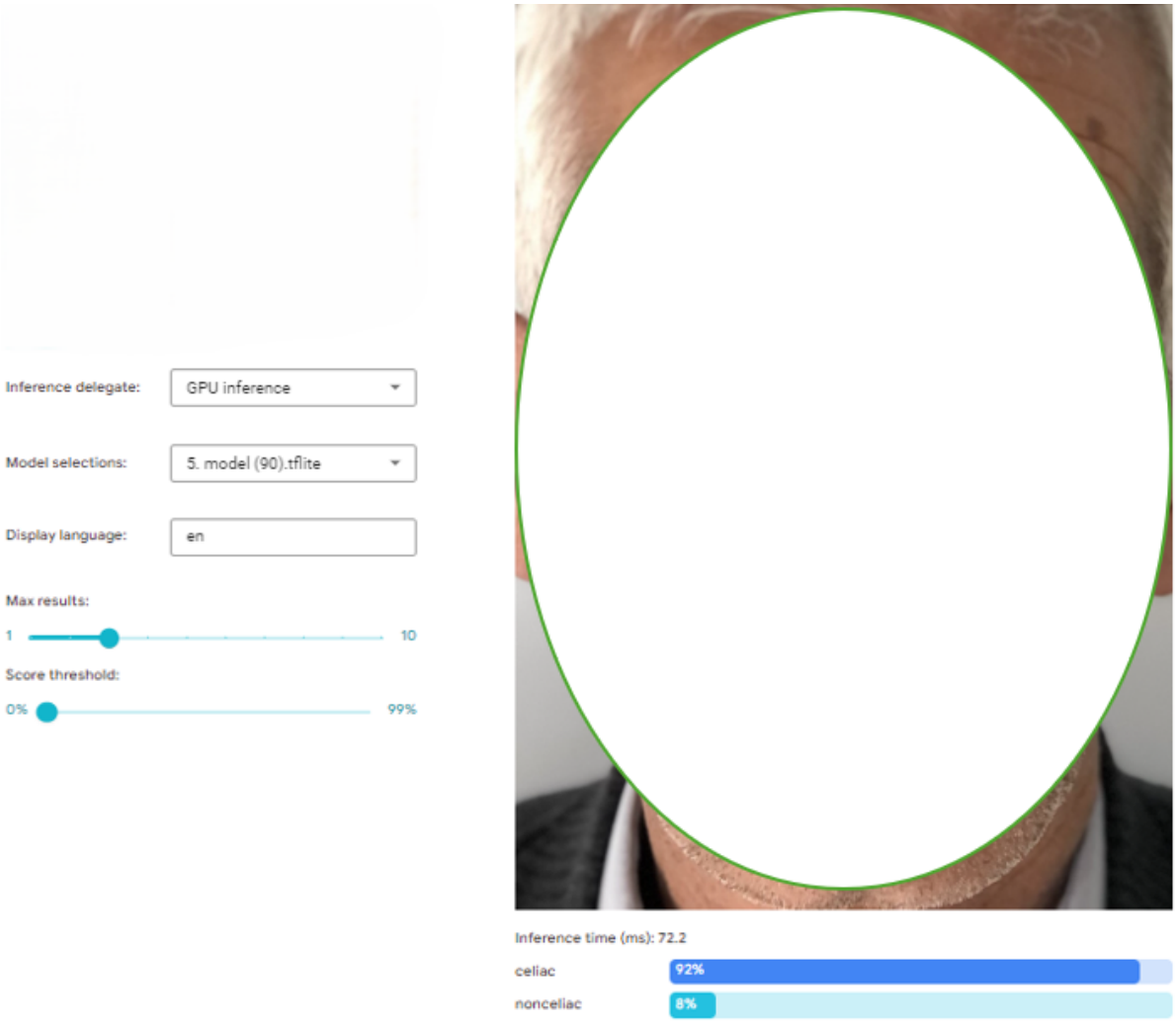



Figure 4.4 Second Test image

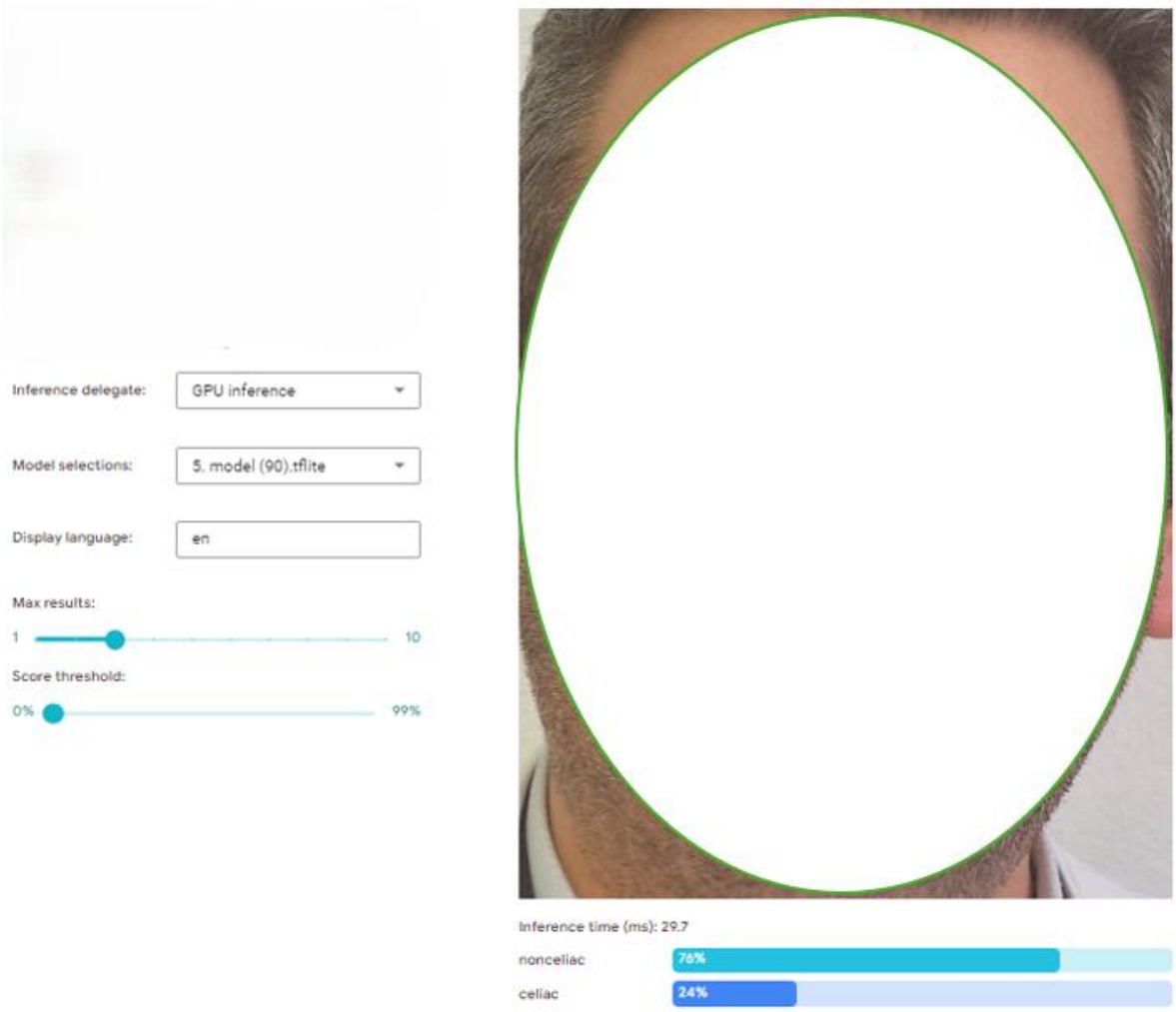



Figure 4.5 Third Test image

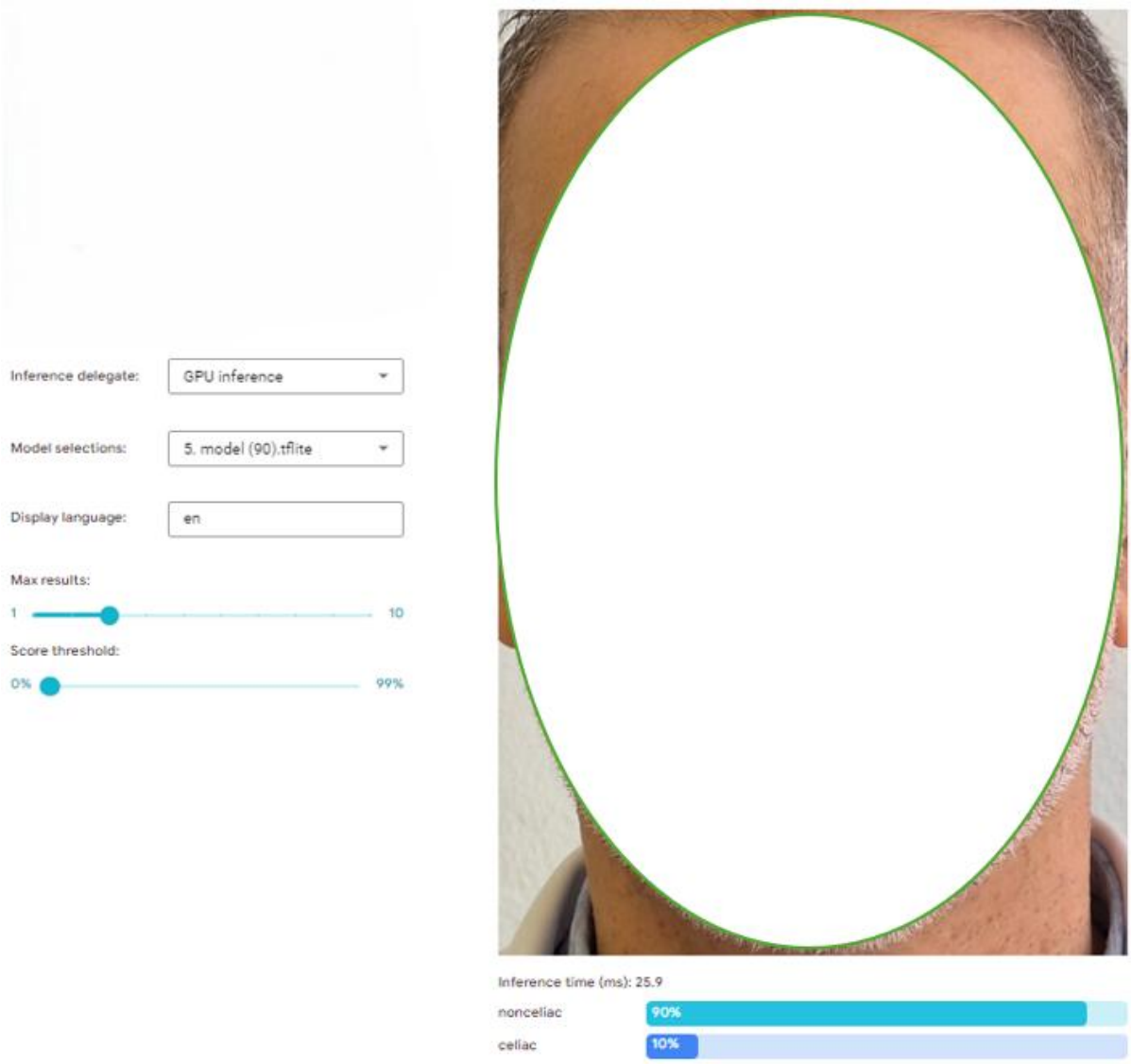



Figure 4.6 Fourth Test image

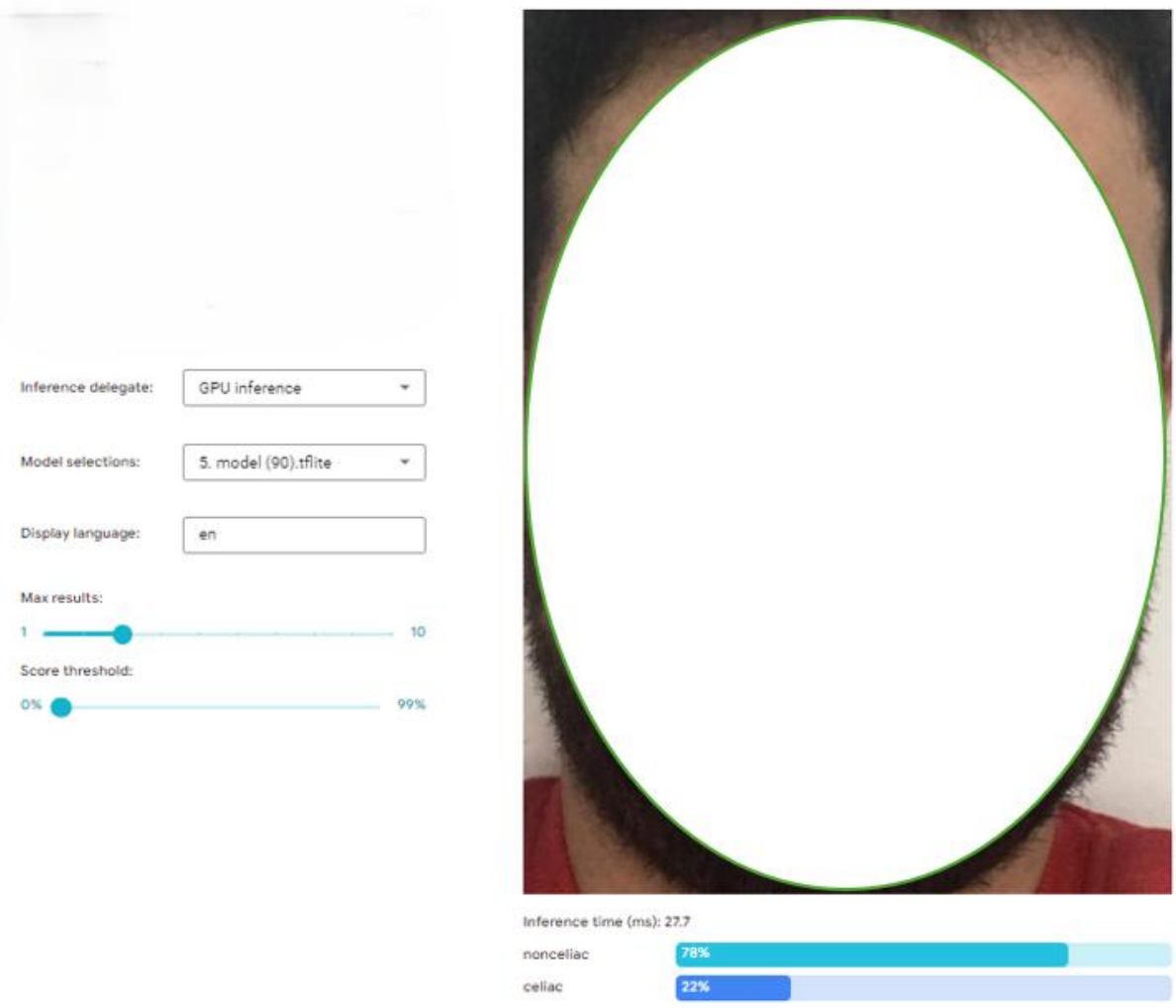



Figure 4.7 Fifth Test image

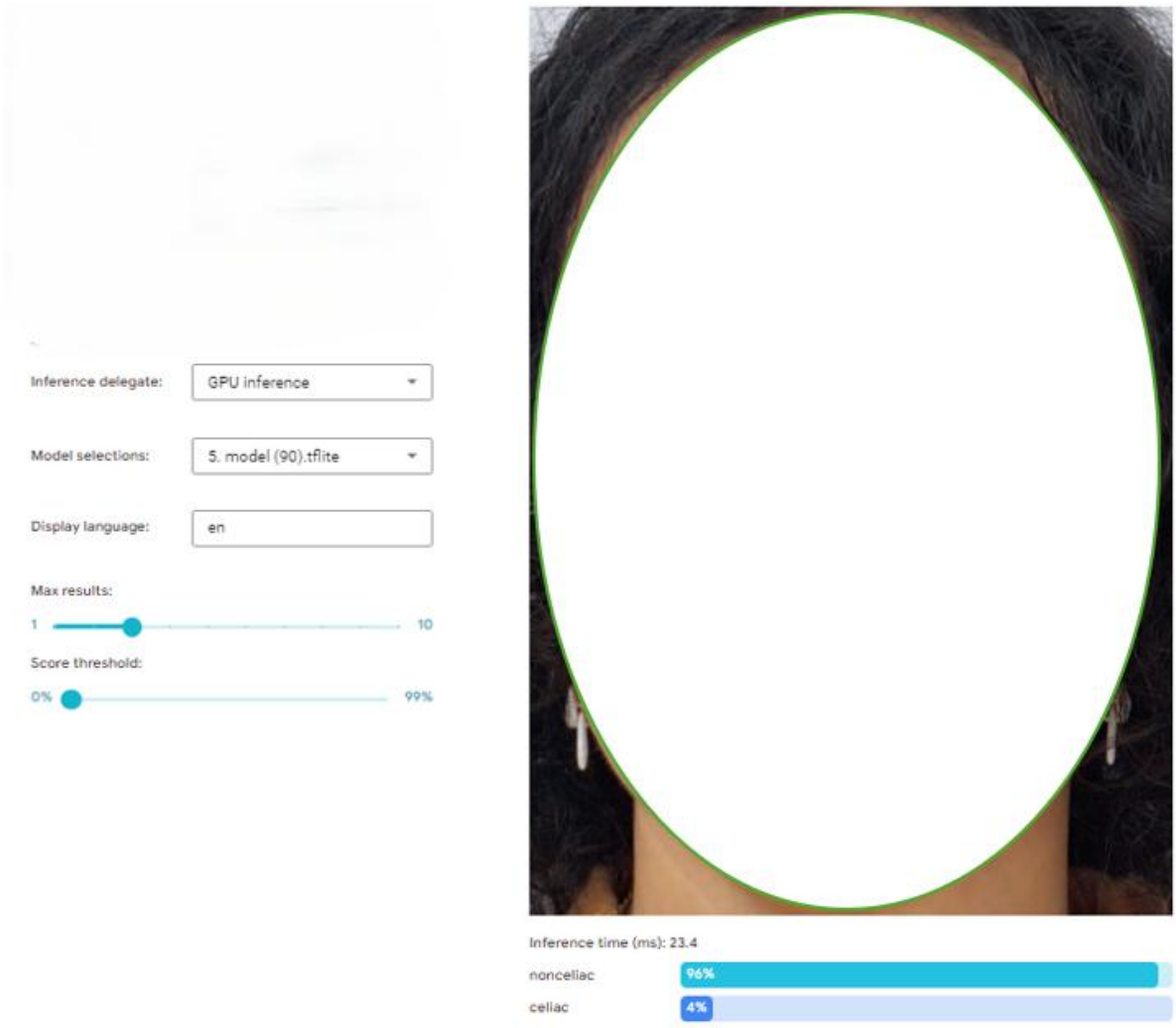



Figure 4.8 Sixth Test image

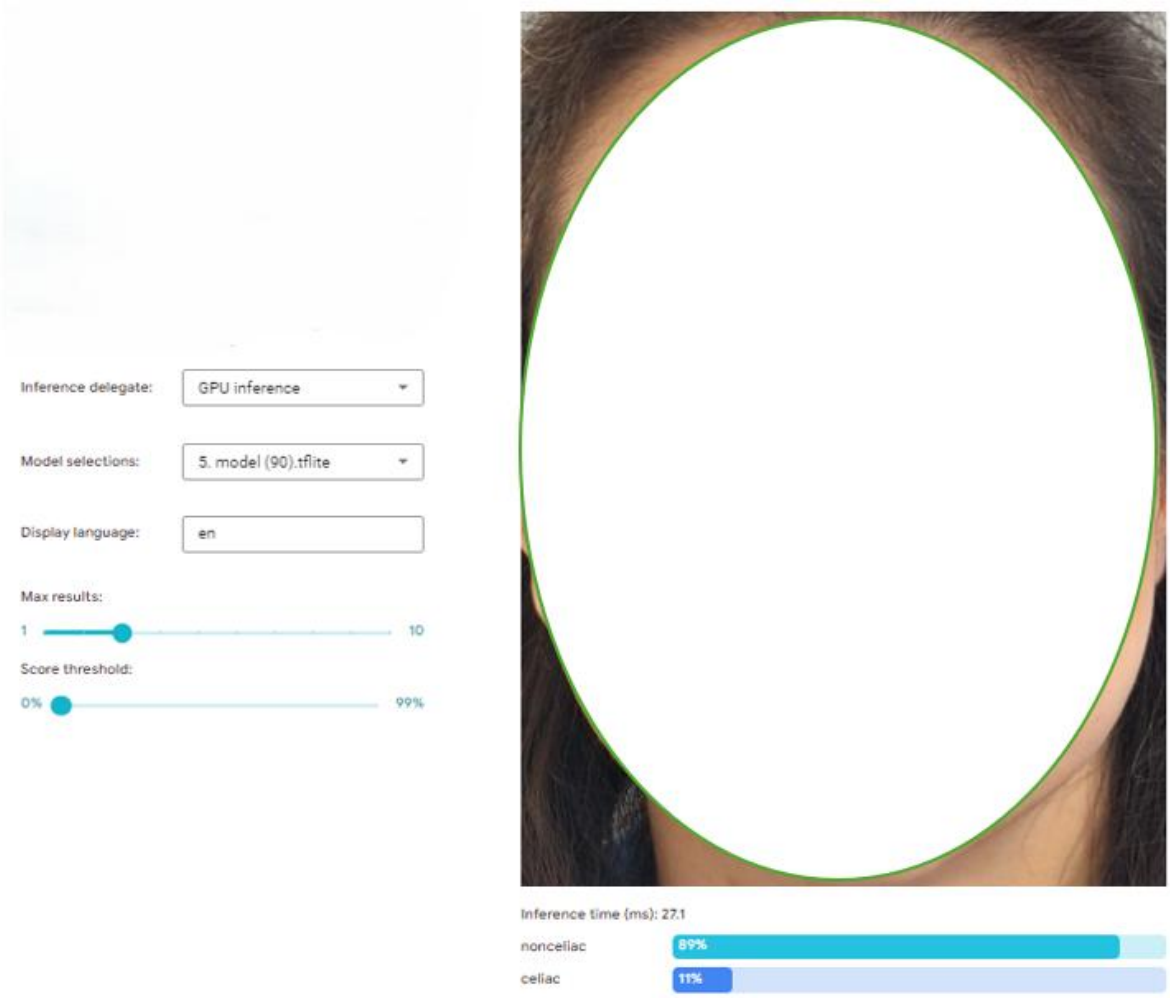



Figure 4.9 Seventh Test image

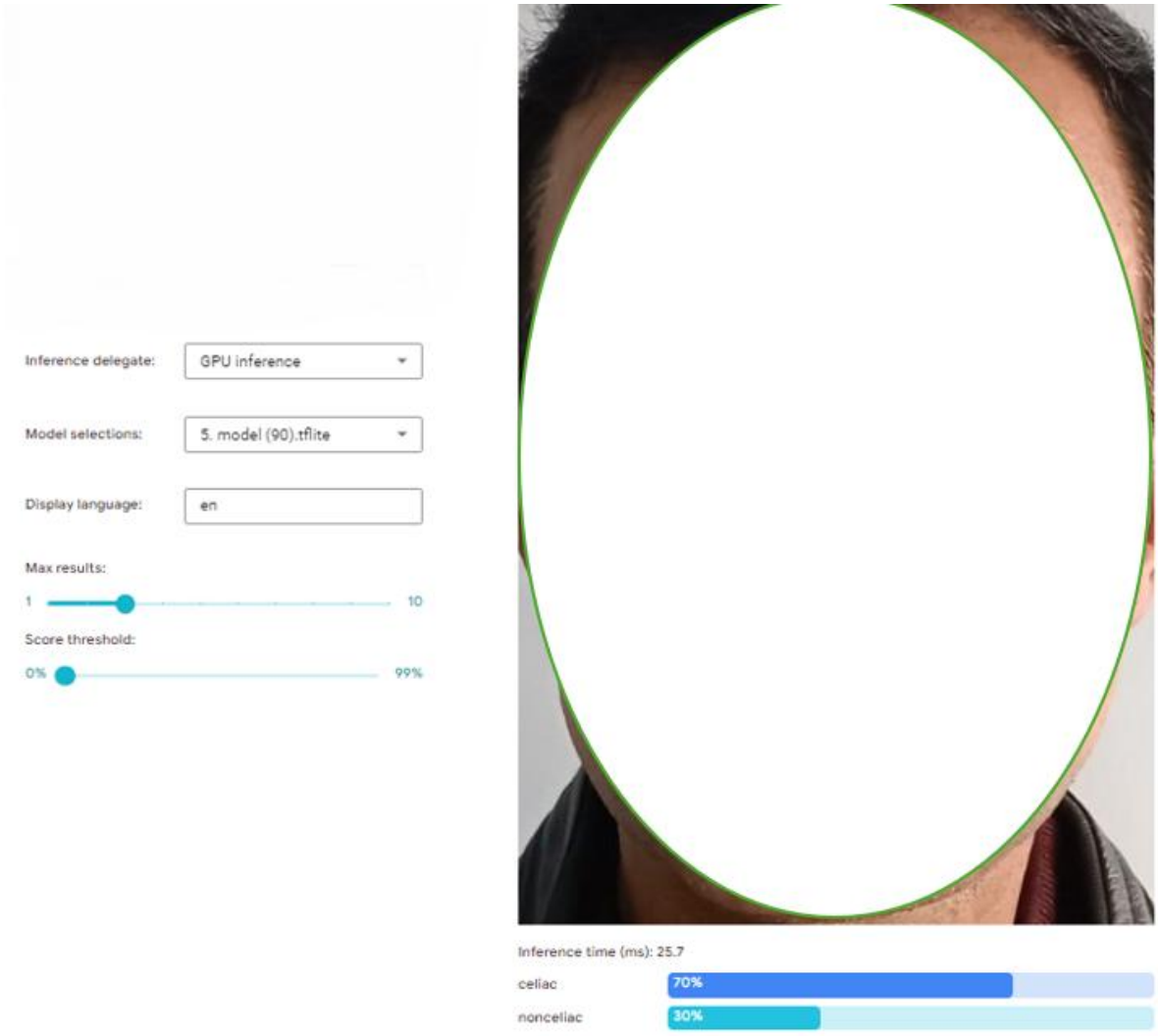



Figure 4.10 Eight Test image

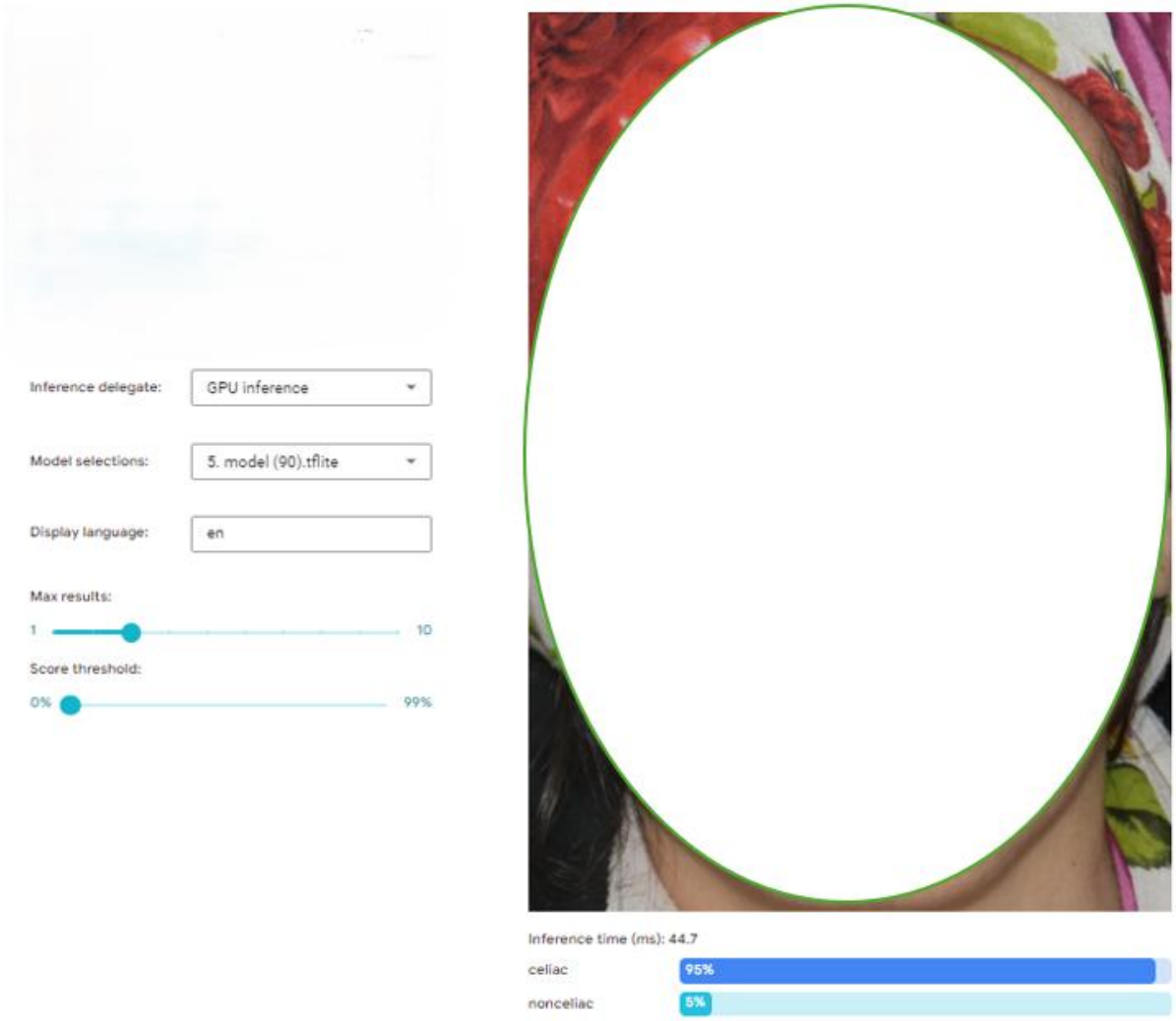



Figure 4.11 Ninth Test image

## 4.3. RESULTS OF CLASSIFICATION OF CELIAC PATIENT CLINICAL DATASETS WITH PYTORCH AND TRADITIONAL MACHINE LEARNING ALGORITHMS

In this study, we employed a comprehensive array of classification models on DS_Marsh dataset, including a deep learning neural network implemented in PyTorch, as well as traditional classifiers such as Random Forest, Decision Tree, Gradient Boosting, and Naive Bayes. Our objective was to identify the optimal model for predicting celiac marsh levels by using rutin blood test and demographic data. Celiac marsh level or intestinal atrophy level is a critical task in diagnosing and managing celiac disease. The results of our analysis reveal distinct performance characteristics across the different models.

In the evaluation of classification models, the PyTorch model demonstrated (**Figure 4.12 and 4.13**) an 80% accuracy in identifying Marsh levels during testing, exhibiting commendable recall, precision, and F1-score metrics of 0.81, 0.80, and 0.70, respectively. Conversely, the Decision Tree and Gradient Boosting classifiers, Random Forest displayed flawless accuracy, each achieving 100%, with corresponding recall, precision and F1-score metrics of 1.00 (Figure 4.14. and 4.15). However, the Naive Bayes classifier yielded comparatively lower performance, attaining an accuracy of 55% and precision, recall, and F1-score metrics of 0.67, 0.44, and 0.53, respectively (**Figure 4.16 and 4.17**). These findings highlight the strengths of various classifiers in Marsh level identification, emphasizing the robustness of ensemble methods such as Random Forests, Decision Trees and Gradient Boosting, while also underscoring the importance of considering model suitability for specific datasets and tasks.

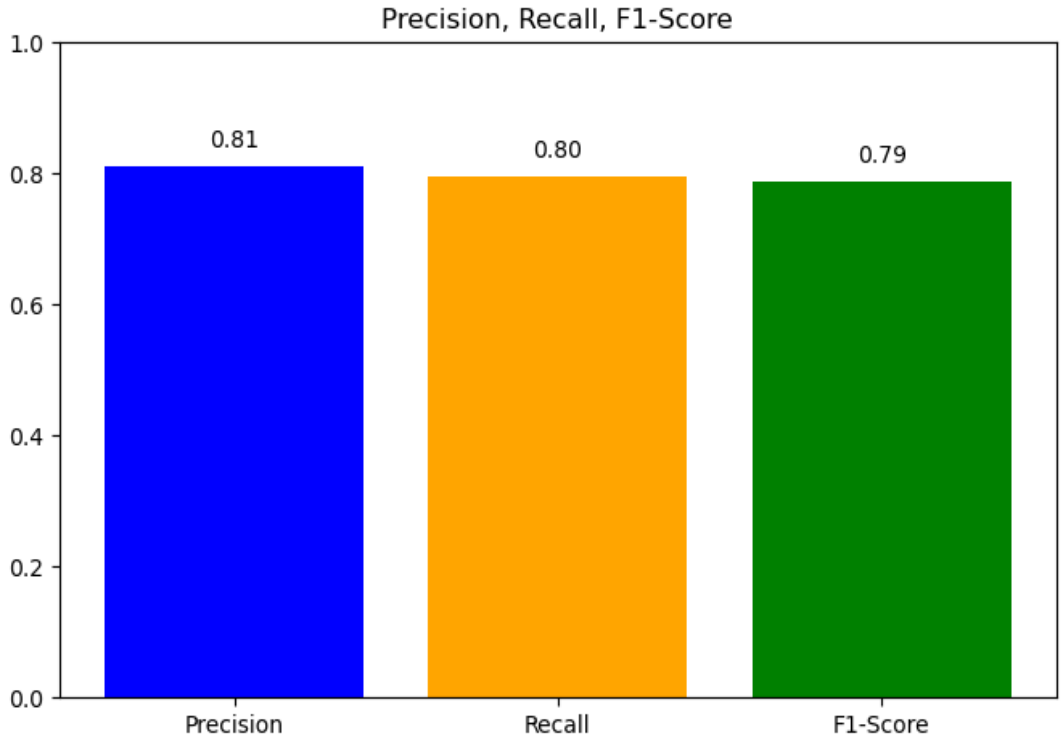



Figure 4.12. Precision, recall, and F1-score metrics

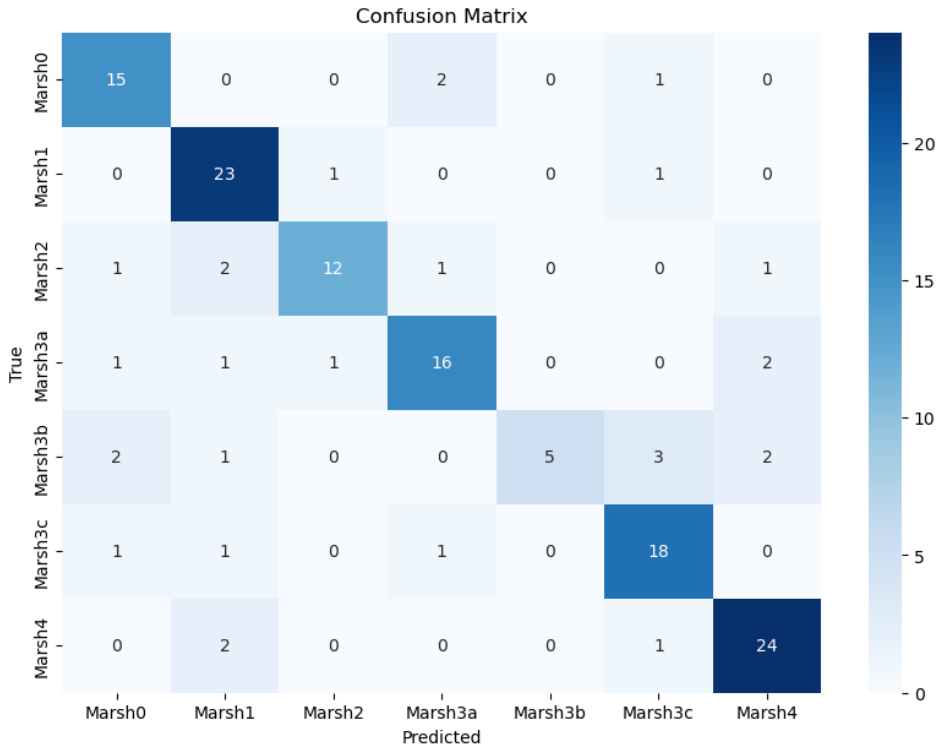



Figure 4.13. PyTorch model's confusion matrix

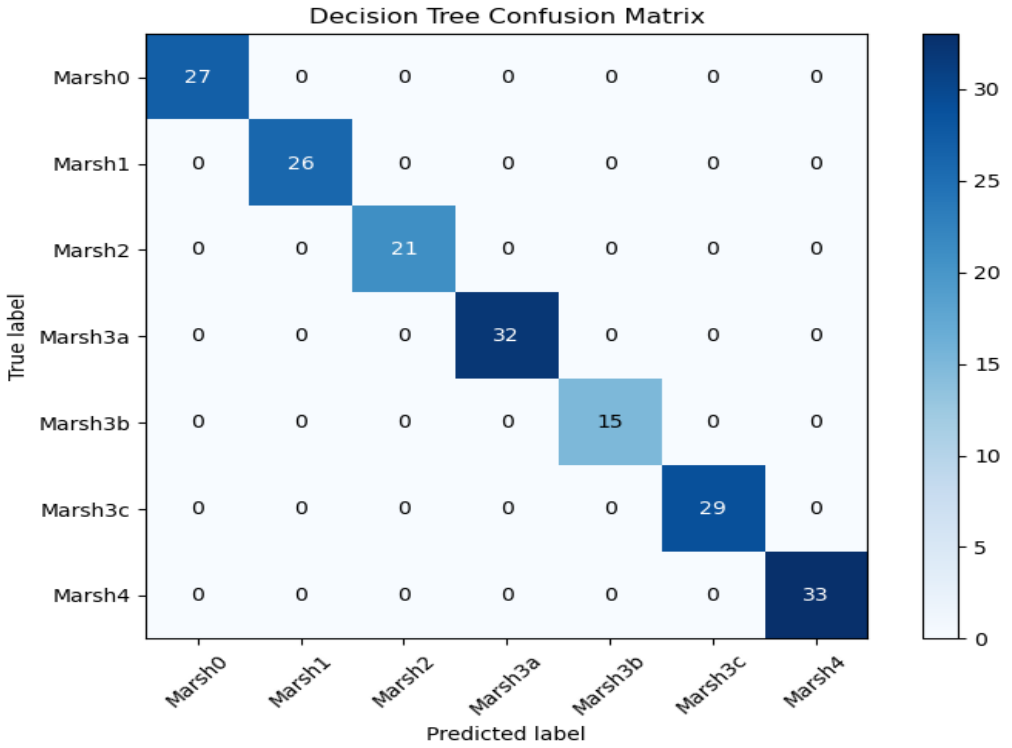



Figure 4.14. Decision Trees, Random Forests, and Gradient Boosting's confusion matrix

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Marsh0 | 1.00 | 1.00 | 1.00 | 27 |
| Marsh1 | 1.00 | 1.00 | 1.00 | 26 |
| Marsh2 | 1.00 | 1.00 | 1.00 | 21 |
| Marsh3a | 1.00 | 1.00 | 1.00 | 32 |
| Marsh3b | 1.00 | 1.00 | 1.00 | 15 |
| Marsh3c | 1.00 | 1.00 | 1.00 | 29 |
| Marsh4 | 1.00 | 1.00 | 1.00 | 33 |
| accuracy | | | 1.00 | 183 |
| macro avg | 1.00 | 1.00 | 1.00 | 183 |
| weighted avg | 1.00 | 1.00 | 1.00 | 183 |

Figure 4.15. Precision, recall, and F1-score metrics

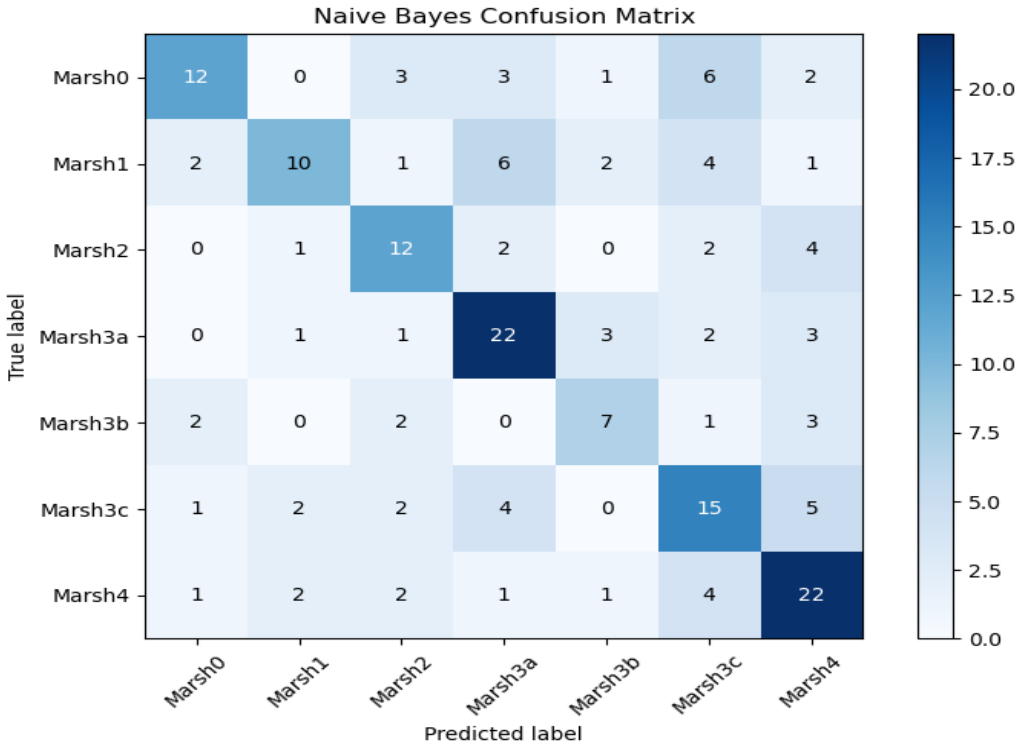



Figure 4.16. Precision, recall, and F1-score metrics

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Marsh0 | 0.67 | 0.44 | 0.53 | 27 |
| Marsh1 | 0.62 | 0.38 | 0.48 | 26 |
| Marsh2 | 0.52 | 0.57 | 0.55 | 21 |
| Marsh3a | 0.58 | 0.69 | 0.63 | 32 |
| Marsh3b | 0.50 | 0.47 | 0.48 | 15 |
| Marsh3c | 0.44 | 0.52 | 0.48 | 29 |
| Marsh4 | 0.55 | 0.67 | 0.60 | 33 |
| accuracy | | | 0.55 | 183 |
| macro avg | 0.55 | 0.53 | 0.54 | 183 |
| weighted avg | 0.56 | 0.55 | 0.54 | 183 |

Figure 4.17. Naive Bayes's confusion matrix

# 5. DISCUSSION

This thesis study has two research questions and two aims. Firstly, it is aimed to investigate whether VGG16 and MobilenetV2 pre-trained models perform significantly well to classify the facial images of celiac disease by using transfer learning to help diagnosis or not. Secondly, it was aimed to analyze a method that would support diagnosis of the disease by determination of Marsh levels through routine blood tests and classification of demographic data on purpose of preventing delays in referral to a specialist in primary health care units. Random Forest, Decision Tree, Naive Bayes, Gradient Boosting and deep learning (pytorch) algorithms were used for classifying of CD face images and as a result of comparisons, the best performing model was selected.

This approach fills a significant gap in the existing literature, which predominantly focuses on the analysis of endoscopic and biopsy images for the diagnosis of CD Our findings lay the groundwork for further investigation of non-invasive diagnostic methods using facial image classification and rutin blood tests with demographic information.

Compared to existing diagnostic techniques, previous studies have showed the effectiveness of artificial intelligence and ML in diagnosing CD, mostly through endoscopic and biopsy images. For example, Molder and friends [69] and Stoleru and friends [66] highlighted the high accuracy of computer-assisted techniques in the analysis of endoscopic images. Some other studies has been done about the use of deep learning for endoscopy image analysis, which significantly increased diagnostic accuracy and efficiency. Similarly, Stoleru and friends achieved a remarkable accuracy of 94.1% in detecting celiac disease using a machine learning algorithm on video capsule endoscopy images. However, no study has been conducted on facial images that could help diagnose the disease.

DL methods are generally used in biopsy image analysis. The effectiveness of deep learning models in analyzing biopsy images for celiac disease diagnosis is also well documented. Wei and friends reported a 95.3% accuracy rate in detecting celiac disease using a deep learning model on biopsy slides. Moreover, a computer-based detection system developed by Koh and friends which achieved 86.47% accuracy for capsule endoscopy images [71]. These studies underscore the high potential of DL in improvement of diagnostic accuracy in clinical settings.

In this field, image processing, classification or analysis using transfer learning has not been done. Our work is also pioneering in this field.

There are great number of research that have been conducted to improve the diagnosis and management of CD using machine learning methodologies. The methodology they mostly use is standard machine learning. Clinical signs, symptoms, signs and serological tests were used, but facial images were not used. This study is the first to use facial image analysis method in the design of a support system for the CD diagnosis. In addition, the use of transfer learning in this field was also made for the first time in our study.

Facial image classification for diagnostic aid system design studies in celiac disease have not been conducted before. Our work stands out by applying transfer learning to classify facial images for CD diagnosis a method that has not been previously investigated in the literature. While endoscopic and biopsy image analyzes reported higher accuracies, our study achieved a moderate accuracy of 73% with VGG16 and a stronger accuracy of 85% with MobileNetv2. This shows that facial image classification has potential but requires further refinement and validation to reach the performance levels of traditional diagnostic methods. When we compare the model performances, the MobileNetV2 model achieved higher success.

The VGG16 model classified celiac face images in the DS_Face dataset with an accuracy rate of 73%. And it achieved precision, recall, and F1 scores of 0.54, 0.56, and 0.52, respectively. While the model demonstrated moderate accuracy in detecting subtle facial features associated with celiac disease, the balanced but moderate sensitivity and recall indicate that further improvements are required. Performance limitations are partly due to the small size of the dataset. The dataset varies in terms of dietary compliance. The high number of newly diagnosed patients in the dataset who are not on a gluten diet will directly affect the success of the model. As a result, the need for a larger and more diverse data set is emphasized.

Transfer learning with VGG16 provided a practical solution that allowed reasonable performance with a limited dataset. However, resizing images from 1600x2000 pixels to 224x224 pixels as required by VGG16 can introduce artifacts and hide fine details. Although VGG16 was still able to extract robust features, the reduction of detail in the patient's facial photographs negatively affected the model performance. Additionally, since we aim to develop a web-based system for patients to upload images for predictive purposes, variability in image sizes and possible pixelation may further affect the accuracy of the model.

On the other hand, the MobileNetv2 model showed a stronger performance, achieving 85% accuracy and 0.44 test loss with precision, recall, and F1 score values of 0.80, 0.81, and 0.80, respectively. The difference in performance between VGG16 and MobileNetV2 in classifying the DS_Face dataset of celiac face images can be attributed to several factors. VGG16, which is a deeper network with more parameters, may not perform best on smaller datasets. On the other hand, MobileNetV2, which was designed for efficiency with fewer parameters and deeply separable convolutions, showed superior performance on the DS_Face dataset. MobileNetV2's robustness to changes in preprocessing, better convergence during fine-tuning, and superior feature extraction capabilities have made it more suitable for face image classification. Additionally, using it with the Mediapipe model pointer MobileNetV2 improved performance by enabling improved preprocessing or feature extraction. The complexity of VGG16 also made it more prone to overfitting, reducing its ability to generalize. These factors collectively explain why MobileNetV2 outperforms VGG16 on this task.

As a results, MobileNetv2 is more suitable for this type of image analysis due to its architecture and training efficiency. This study represents the first attempt to diagnose celiac disease through facial image classification, reflecting the novelty and complexity of this approach. And it stands out as an innovative attempt in employing a TL approach by using the pre-trained MobilNet.V2 model to classify facial images of both CD and non-CD individuals. This assists in diagnosing celiac disease or guiding referrals to specialists by primary care physicians. Although MobileNetV2 showed very good success, a higher number of newly diagnosed CD patients in the DS_Face dataset could have significantly increased the effectiveness of the model, and it would even be better if the dataset consisted only of newly diagnosed CD cases. Furthermore, a larger data set size could have more increased the success of the model. Consequently, expanding the dataset in size and diversity can increase the generalizability and robustness of the models.

We believe the potential impact of our study is that facial image analysis will begin to be used as a complementary tool for CD diagnosis, offering a non-invasive and easily accessible method for preliminary screenings. Facial image analysis may enable earlier detection and treatment of celiac disease by complementing traditional diagnostic methods and thus may reduce the risk of associated complications.

In conclusion, first part of this study successfully applied transfer learning using VGG16 and MobileNetv2 to classify facial images for celiac disease diagnosis and achieved moderate to

strong accuracy. Although the results are promising, it is important to validate and improve this new diagnostic approach. The analysis in this study represents a considerable contribution to this field with provided a potential non-invasive tool that could complement existing diagnostic methods for diagnosing celiac disease.

In second half of the research was celiac marsh level detection. Celiac disease is a complex autoimmune condition triggered by gluten consumption, affecting individuals worldwide with a variety of clinical manifestations. Accurate classification of disease severity through Marsh grading is crucial to guide clinical management and improve patient outcomes. Traditional diagnostic methods rely on invasive procedures such as duodenal biopsies and serological tests, highlighting the need for more effective and accurate diagnostic approaches. Our study brought methodological innovation to determine the celiac marsh level. It innovatively uses machine learning models, including traditional classifiers such as Decision Trees, Random Forests, Gradient Boosting, and Naive Bayes, as well as a deep learning neural network implemented in PyTorch. These models were trained and validated using a dataset of 182 adults diagnosed at Marsh levels 0 to 4, including clinical symptoms and serological markers. The study aimed to create a robust classification system that could predict Marsh levels with high accuracy and reduce reliance on potentially invasive procedures. Comparing existing research in this area reveals interesting contrasts and similarities. Studies focusing on serological markers such as IGA, IGG, IGM, hemoglobin, plasma iron binding capacity, ferritin, B12, HGB, HCT consistently highlight their correlation with Marsh grading and disease severity [24, 78]. Many reseach showed a strong relationship between serological marker levels and Marsh classification which is closely consistent with the findings of our study. Serological markers were crucial in our machine learning models and contributed to the high accuracy rates observed in classifiers such as Decision Trees and Gradient Boosting, which achieved excellent accuracy in predicting Marsh levels. By achieving up to 100% accuracy in certain models, it suggests a promising path for the development of automated diagnostic tools that can complement traditional biopsy and serological testing methods. Additionally, the use of PyTorch as DL neural networks represents a new application in the diagnosis of celiac disease. While the PyTorch neural network exhibits promising prediction accuracy, its weaknesses include dependence on large and diverse datasets and the need for extensive computational resources and expertise for effective training. In contrast, the weaknesses of Naive Bayes lie in its assumption of feature independence, which may not be valid for all datasets, and its limited ability to handle numerical features or continuous data; this potentially leads to lower prediction

accuracy on complex datasets like ours. Through our comparative evaluation, we gathered valuable information about the relative strengths and weaknesses of each model. By adopting a data-driven approach, our study contributes to increasing diagnostic accuracy and improving patient management strategies in the context of celiac disease. This methodology offers robustness and scalability, paving the way for future research on integrating AI-supported diagnostic tools into clinical practice. Comparative analysis with traditional classifiers further enriches our understanding of the strengths and restrictions of different ML approaches in healthcare applications. While machine learning models demonstrated high accuracy in our dataset, external validation in larger cohorts and different demographic settings would strengthen the generalizability of these findings. The primary contribution of the study lies in demonstrating the effectiveness of machine learning in increasing diagnostic accuracy for the severity of celiac disease. This approach not only increases diagnostic speed and accuracy, but also potentially reduces patient discomfort and healthcare costs associated with invasive procedures.

Future research directions may investigate the integration of additional clinical and genetic markers into machine learning models to further increase prediction accuracy. Longitudinal studies may also investigate the utility of these models in monitoring disease progression and treatment response over time, potentially leading to personalized treatment strategies for celiac disease patients.

In conclusion, second part of this study advances the field by proposing an innovative and reliable classification system to predict celiac Marsh levels using advanced machine learning techniques. Our approach not only increases diagnostic accuracy through clinical symptoms and serological markers, but also contributes to integrating the broader goal of personalized medicine in autoimmune disorders.

This thesis study served to develop a system that will help diagnose celiac disease and showed that it is possible to diagnose patients from facial images and serological tests by using all types of ML (transfer learning and deep learning) together. Ultimately, the study underscores the transformative potential of AI-driven diagnostics in optimizing healthcare delivery and patient outcomes in celiac disease management.

# 6. CONCLUSION AND RECOMMENDATIONS

This study, conducted as part of a doctoral thesis, successfully explored the application of transfer learning using VGG16 and MobileNetv2 models to classify facial images for assisting of the diagnosis of celiac disease. A comprehensive review of existing studies reveals a significant gap in research on diagnostic assistance systems that use transfer learning to classify photographs of CD patients. This area of research aimed at supporting the diagnosis of celiac disease or providing referral guidance to primary care physicians has been understudied. The retrained model achieved an accuracy of 73%, demonstrating its potential as a diagnostic tool for identifying celiac disease. However, this result also shows that there is significant room for improvement, especially considering that some images are difficult to classify even for human experts. While VGG16 achieved an accuracy of 73%, MobileNetv2 showed a higher accuracy of 85%, demonstrating its suitability for this type of image analysis. A TL approach with MMM is employed to classify CD face images. MobileNetV2 is a pre-trained model serves as a feature extractor. This model accomplished an accuracy of 85%. The proposed model worked well on new data with efficient learning from the CD face image dataset. The test loss is 0.4365, which means that, on average, the difference between the prediction of model and in the test dataset the actual labels are moderate. Additionally, the test accuracy is 0.85, indicating that most samples in the test dataset were classified correctly. This suggests that the model performs quite well on the given test data. These findings highlight the potential of facial image classification as a non-invasive diagnostic tool, but further improvements and validation are required to match the performance of traditional diagnostic methods.

The results make significant contributions to the scientific community by bringing an innovative approach to celiac disease diagnosis. Unlike previous studies that focused on endoscopic and biopsy image analyses, this study pioneers the use of facial images, opening new avenues for non-invasive diagnostic techniques. The moderate success achieved with VGG16 and the stronger performance of MobileNetv2 highlight the feasibility of this method, while also pointing to the need for more complex models and larger, more diverse datasets to improve accuracy and reliability. A complementary diagnostic tool to facial image analysis its introduction offers many potential benefits. It provides a non-invasive and accessible method

for preliminary screenings that may be particularly useful in clinical settings and telemedicine platforms, especially in regions where access to specialized medical care is limited. In conclusion, while this study demonstrates the promise of facial image classification for diagnosing celiac disease, it also highlights difficulties and restrictions that need to be addressed. Future research needs to focus on improving the complexity of the model, expanding datasets, and integrating this approach with traditional diagnostic methods to create a comprehensive and robust diagnostic tool. The innovative use of facial image analysis in this research demonstrates a significant advance in the field, offering a potential non-invasive diagnostic tool that could transform current diagnostic practices and contribute to improved patient care.

To improve the diagnostic accuracy and reliability of facial image classification in celiac disease, upcoming research should work on several important areas. Firstly, developing more complex models or incorporating additional types of data, such as clinical data, can increase diagnostic accuracy. Secondly, collecting a larger and more diverse dataset of facial images will improve model training and validation and provide better generalizability across different populations. Finally, integrating facial image analysis with traditional diagnostic methods can create a comprehensive diagnostic tool that provides a holistic approach to diagnosing celiac disease. This multifaceted approach will provide a more robust and reliable method for early diagnosis and treatment, as well as improving diagnostic accuracy. Our work pioneers the use of facial image classification in the diagnosis of CD and contributes to the field by exploring an unconventional diagnostic method. It paves the way for future research to improve this approach and integrate it with other diagnostic tools for a more comprehensive diagnosis. The innovative use of facial image analysis represents a significant advance in the field, offering a potential non-invasive diagnostic tool that could transform current diagnostic practices.

The second phase of the study significantly advanced the field of celiac disease diagnosis by providing a new and reliable method for estimating Marsh levels using deep and advanced machine learning techniques. Our study comprehensively evaluated a range of classification models, including both modern deep learning approaches like the PyTorch neural network and traditional algorithms such as Decision Trees, Random Forests, Gradient Boosting, and Naive Bayes, for predicting celiac marsh levels. While each model exhibited distinct strengths and weaknesses, our findings provide valuable insights into their relative performance and suitability for this critical diagnostic task. Integration of clinical symptoms and serological

markers have demonstrated a high degree of accuracy. In particular, some models have achieved up to 100% accuracy, demonstrating their potential to complement traditional diagnostic methods and reduce dependence on invasive procedures. In this context, the innovative use of deep learning neural networks represents a pioneering application in celiac disease diagnosis and offers robustness and scalability, paving the way for future research into AI-supported diagnostic tools.

Despite the promising results, upcoming research should address unresolved issues and increase the generalizability of our findings. They also should focus on external validation in larger and more diverse cohorts, integrating additional clinical and genetic markers, and investigating the usefulness of these models in monitoring disease progression and response to treatment over time. Contributing to the goal of personalized medicine in autoimmune disorders, our approach represents a significant step forward in the noninvasive and effective diagnosis of CD.

As a result, VGG16 and MobilenetV2 models used by transfer learning showed significant accuracy in diagnosing celiac disease with only the patient's facial image. By using routine blood tests and demographic data in primary health care units, ML methods have shown the capability of significant diagnosis aid system, follow-up and management of the disease by determining intestinal atrophy (Marsh level).

Going forward, further research may explore hybrid approaches that will leverage the strengths of multiple models and improve prediction performance. MobilenetV2 can be used for facial image processing. Decision Trees, Random Forests and Gradient Boosting can be used to detect disease severity. With such a hybrid decision support system, diagnosis of celiac disease could be faster and more accurate. Patients can be diagnosed by directing them to the right unit without wasting time, before the symptoms worsen and the disease caused by celiac disease opens up. Overall, our study highlights the importance of taking a data-driven approach to medical diagnosis, facilitating informed and timely decision-making and improving patient outcomes, using artificial intelligence principles.

# REFERENCES


1. Caio, G., Volta, U., Sapone, A., Leffler, D. A., De Giorgio, R., Catassi, C., & Fasano, A. (2019). Celiac disease: A comprehensive current review. In *BMC Medicine* (Vol. 17, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12916-019-1380-z

2. Holtmeier, W., & Caspary, W. F. (2006). Celiac disease. *Orphanet Journal of Rare Diseases*, *1*, 3. https://doi.org/10.1186/1750-1172-1-3

3. Malamut, G., Cording, S., & Cerf-Bensussan, N. (2019). Recent advances in celiac disease and refractory celiac disease. In *F1000Research* (Vol. 8). F1000 Research Ltd. https://doi.org/10.12688/f1000research.18701.1

4. Elli, L., Branchi, F., Tomba, C., Villalta, D., Norsa, L., Ferretti, F., Roncoroni, L., & Bardella, M. T. (2015). Diagnosis of gluten related disorders: Celiac disease, wheat allergy and non-celiac gluten sensitivity. *World Journal of Gastroenterology*, *21*(23), 7110–7119. https://doi.org/10.3748/wjg.v21.i23.7110

5. Kayar, Y., & Dertli, R. (2019). Association of autoimmune diseases with celiac disease and its risk factors. *Pakistan Journal of Medical Sciences*, *35*(6), 1548–1553. https://doi.org/10.12669/pjms.35.6.821

6. Sadeghi, A., Rad, N., Ashtari, S., Rostami-Nejad, M., Moradi, A., Haghbin, M., Rostami, K., Volta, U., & Zali, M. R. (2019). The value of a biopsy in celiac disease follow up: assessment of the small bowel after 6 and 24 months treatment with a gluten free diet. *Revista Española de Enfermedades Digestivas*, *112*. https://doi.org/10.17235/reed.2019.5947/2018

7. Keskin, E. (2019). *CLINICAL DECISION SUPPORT SYSTEMS IN DIAGNOSIS OF*. *November*, 239–250.

8. Russo, E., Di Gloria, L., Cerboneschi, M., Smeazzetto, S., Baruzzi, G. P., Romano, F., Ramazzotti, M., & Amedei, A. (2023). Facial Skin Microbiome: Aging-Related Changes and Exploratory Functional Associations with Host Genetic Factors, a Pilot Study. *Biomedicines*, *11*(3). https://doi.org/10.3390/biomedicines11030684

9. Sinha, S., Lin, G., & Ferenczi, K. (2021). The skin microbiome and the gut-skin axis. *Clinics in Dermatology*, *39*(5), 829–839. https://doi.org/10.1016/j.clindermatol.2021.08.021

10. Telega, G., Bennet, T. R., & Werlin, S. (2008). Emerging New Clinical Patterns in the Presentation of Celiac Disease. *Archives of Pediatrics & Adolescent Medicine*, *162*(2), 164–168. https://doi.org/10.1001/ARCHPEDIATRICS.2007.38

11. Finizio, M., Quaremba, G., Mazzacca, G., & Ciacci, C. (2005). Large forehead: A novel sign of undiagnosed coeliac disease. *Digestive and Liver Disease*, *37*(9), 659–664. https://doi.org/10.1016/j.dld.2005.04.014

12. Durazzo, M., Ferro, A., Brascugli, I., Mattivi, S., Fagoonee, S., & Pellicano, R. (2022). Extra-Intestinal Manifestations of Celiac Disease. What Should We Know in 2022? In *Journal of Clinical Medicine* (Vol. 11, Issue 1). MDPI. https://doi.org/10.3390/jcm11010258

13. Sernicola, A., & Alaibac, M. (2023). Editorial: Cutaneous manifestations of systemic disease. In *Frontiers in Medicine* (Vol. 10). Frontiers Media SA. https://doi.org/10.3389/fmed.2023.1236570

14. Akhand, M. A. H., Roy, S., Siddique, N., Kamal, M. A. S., & Shimamura, T. (2021). Facial emotion recognition using transfer learning in the deep CNN. *Electronics (Switzerland)*, *10*(9). https://doi.org/10.3390/electronics10091036

15. Singh, A., & Kisku, D. R. (2018). Detection of Rare Genetic Diseases using Facial 2D Images with Transfer Learning. *Proceedings of the 2018 8th International Symposium on Embedded Computing and System Design, ISED 2018*, 26–30. https://doi.org/10.1109/ISED.2018.8703997

16. Jin, B., Cruz, L., & Goncalves, N. (2020). Deep Facial Diagnosis: Deep Transfer Learning from Face Recognition to Facial Diagnosis. *IEEE Access*, *8*, 123649–123661. https://doi.org/10.1109/ACCESS.2020.3005687

17. Dānishgāh-i Ṣanʿatī-i Amīr Kabīr, Institute of Electrical and Electronics Engineers, & International Iranian Conference on Biomedical Engineering (2nd : 2017 : Tehran, I. (n.d.). *2017 24th Iranian Conference on Biomedical Engineering and 2017 2nd International Iranian Conference on Biomedical Engineering (ICBME).*

18. Alhayani, B. S. A., & Rane, M. (2020). Face Recognition System By Image Processing. *International Journal of Electronics and Communication Engineering & Technology (IJECET)*, *5*(5), 80–90. https://doi.org/10.31224/osf.io/75qba

19. Cires¸an, D. C. C., Meier, U., Masci, J., Gambardella, L. M., & Urgen Schmidhuber, J. ¨. (n.d.). *Flexible, High Performance Convolutional Neural Networks for Image Classification*.

20. Vécsei, A., Amann, G., Hegenbart, S., Liedlgruber, M., & Uhl, A. (2011). Automated Marsh-like classification of celiac disease in children using local texture operators. *Computers in Biology and Medicine*, *41*(6), 313–325. https://doi.org/10.1016/j.compbiomed.2011.03.009

21. Ehsan, L., Khan, M. N., Sali, R., & Kowsari, K. (n.d.). *Prediction of Celiac Disease Severity and Associated Endocrine Morbidities through Deep Learning-based Image Analytics*. https://doi.org/10.1101/2021.01.20.21250194

22. Pachisia, A. V., Kumari, A., Mehta, S., Ahmed, A., Chauhan, A., Agarwal, A., Dwarkanathan, V., Rajpoot, S., Prasad, S., Kumar, S., Sinha, S. K., Sharma, D., Rajput, M., Das, P., Falodia, S., Kochhar, R., Ramakrishna, B. S., Ahuja, V., & Makharia, G.

(2024). Validation of no-biopsy pathway for the diagnosis of celiac disease in Asian adults: a multicenter retrospective study. *Journal of Gastroenterology and Hepatology*, *39*(3), 489–495. https://doi.org/10.1111/JGH.16436

23. Tursi, A., Brandimarte, G., & Giorgetti, G. M. (2003). Prevalence of antitissue transglutaminase antibodies in different degrees of intestinal damage in celiac disease. *Journal of Clinical Gastroenterology*, *36*(3), 219–221. https://doi.org/10.1097/00004836-200303000-00007

24. Ganji, A., Esmaeilzadeh, A., Bahari, A., Ghafarzadegan, K., Afzal Aghayee, M., Mosanen Mozafari, H., Hayatbakhsh, A., Ghavami Ghanbarabadi, V., Ravarian, B., & Rahimi, leili. (2016). Correlation Between Cut-off Level of Tissue Transglutaminase Antibody and Marsh Classification Brief Report. *Middle East Journal of Digestive Diseases*, *8*(4), 318–322. https://doi.org/10.15171/mejdd.2016.42

25. Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., & Abdul Azim, M. (n.d.). *Transfer learning: a friendly introduction*. https://doi.org/10.1186/s40537-022-00652-w

26. Ritahani Ismail, A., Syed Qamrun Nisa, Shaharuddin, S. A., Masni, S. I., & Suharudin Amin, S. A. (2024). Utilising VGG-16 of Convolutional Neural Network for Medical Image Classification. *International Journal on Perceptive and Cognitive Computing*, *10*(1), 113–118. https://doi.org/10.31436/ijpcc.v10i1.460

27. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (n.d.). *MobileNetV2: Inverted Residuals and Linear Bottlenecks*.

28. Pevzner. (2017). 乳鼠心肌提取 HHS Public Access. *Physiology & Behavior*, *176*(3), 139–148. https://doi.org/10.1016/j.cgh.2017.12.022.Machine

29. Li, X. (2016). 乳鼠心肌提取 HHS Public Access. *Physiology & Behavior*, *176*(3), 139–148. https://doi.org/10.1007/s13679-016-0191-1.The

30. Fasano, A. (2020). All disease begins in the (leaky) gut: Role of zonulin-mediated gut permeability in the pathogenesis of some chronic inflammatory diseases. In *F1000Research* (Vol. 9). F1000 Research Ltd. https://doi.org/10.12688/f1000research.20510.1

31. Obrenovich, M. E. M. (n.d.). *microorganisms Leaky Gut, Leaky Brain?* https://doi.org/10.3390/microorganisms6040107

32. Fasano, A. (2012). Leaky gut and autoimmune diseases. *Clinical Reviews in Allergy and Immunology*, *42*(1), 71–78. https://doi.org/10.1007/s12016-011-8291-x

33. Fasano, A., Not, T., Wang, W., Uzzau, S., Berti, I., Tommasini, A., & Goldblum, S. E. (2000). Zonulin, a newly discovered modulator of intestinal permeability, and its expression in coeliac disease. *Lancet*, *355*(9214), 1518–1519. https://doi.org/10.1016/S0140-6736(00)02169-3

34. Asmar, R. El, Panigrahi, P., Bamford, P., Berti, I., Not, T., Coppa, G. V., Catassi, C., & Fasano, A. (2002). Host-dependent zonulin secretion causes the impairment of the small

intestine barrier function after bacterial exposure. *Gastroenterology*, *123*(5), 1607–1615. https://doi.org/10.1053/gast.2002.36578

35. Lammers, K. M., Lu, R., Brownley, J., Lu, B., Gerard, C., Thomas, K., Rallabhandi, P., Shea-Donohue, T., Tamiz, A., Alkan, S., Netzel-Arnett, S., Antalis, T., Vogel, S. N., & Fasano, A. (2008). Gliadin Induces an Increase in Intestinal Permeability and Zonulin Release by Binding to the Chemokine Receptor CXCR3. *Gastroenterology*, *135*(1), 194-204.e3. https://doi.org/10.1053/j.gastro.2008.03.023

36. Carrera-Bastos, P., Picazo, óscar, Fontes-Villalba, M., Pareja-Galeano, H., Lindeberg, S., Martínez-Selles, M., Lucia, A., & Emanuele, E. (2018). Serum Zonulin and endotoxin levels in exceptional longevity versus precocious myocardial infarction. *Aging and Disease*, *9*(2), 317–321. https://doi.org/10.14336/AD.2017.0630

37. Mu, Q., Kirby, J., Cm, R., & Luo, X. M. (2017). Leaky Gut As a Danger Signal for Autoimmune Diseases. *Leaky Gut As a Danger Signal for Autoimmune Diseases. Front. Immunol*, *8*, 598. https://doi.org/10.3389/fimmu.2017.00598

38. Torres, M. I., López Casado, M. A., & Ríos, A. (2007). New aspects in celiac disease. *World Journal of Gastroenterology*, *13*(8), 1156–1161. https://doi.org/10.3748/wjg.v13.i8.1156

39. Ensari, A., & Marsh, M. N. (2019). Diagnosing celiac disease: A critical overview. In *Turkish Journal of Gastroenterology* (Vol. 30, Issue 5). https://doi.org/10.5152/tjg.2018.18635

40. Durham, J., & Temples, H. S. (2018). Celiac Disease in the Pediatric Population. *Journal of Pediatric Health Care*, *32*(6). https://doi.org/10.1016/j.pedhc.2018.04.023

41. Forchielli, M. L., Diani, L., Labriola, F., Bolasco, G., Rocca, A., Salfi, N. C., Leone, A., Miserocchi, C., Andreozzi, L., Levi della Vida, F., Pessina, A. C., Lima, M., & Pession, A. (2020). Gluten deprivation: What nutritional changes are found during the first year in newly diagnosed coeliac children? *Nutrients*, *12*(1). https://doi.org/10.3390/nu12010060

42. Greenham, O., Johnson, L., & Johnson, M. (2023). A general health economics review of the hidden costs involved in discharging coeliac patients from hospital-based specialty clinics to community-based management. *Gastroenterology and Hepatology from Bed to Bench*, *16*(2), 173–180. https://doi.org/10.22037/ghfbb.v16i2.2715

43. Di Sabatino, A., Vanoli, A., Giuffrida, P., Luinetti, O., Solcia, E., & Corazza, G. R. (2012). The function of tissue transglutaminase in celiac disease. In *Autoimmunity Reviews* (Vol. 11, Issue 10, pp. 746–753). https://doi.org/10.1016/j.autrev.2012.01.007

44. Roca, M., Donat, E., Marco-Maestud, N., Masip, E., Hervás-Marín, D., Ramos, D., Polo, B., & Ribes-Koninckx, C. (2019). Efficacy Study of Anti-Endomysium Antibodies for Celiac Disease Diagnosis: A Retrospective Study in a Spanish Pediatric Population. *Journal of Clinical Medicine*, *8*(12), 2179. https://doi.org/10.3390/jcm8122179

45. Losurdo, G., Leo, M. Di, Santamato, E., Arena, M., Rendina, M., Luigiano, C., Ierardi, E., & Leo, A. Di. (2021). Serologic diagnosis of celiac disease: May it be suitable for

adults? In *World Journal of Gastroenterology* (Vol. 27, Issue 42, pp. 7233–7239). Baishideng Publishing Group Inc. https://doi.org/10.3748/wjg.v27.i42.7233

46. Karaahmet, F. (2018). Çölyak hastalığı ' nda teşhis süresi Length of diagnosis in celiac disease. *Ege Journal of Medicine*, *57*(X), 12–15.

47. Rubin, J. E., & Crowe, S. E. (n.d.). *Celiac Disease*. https://doi.org/10.7326/AITC202001070

48. Green, P. H. R. (2005). The many faces of celiac disease: Clinical presentation of celiac disease in the adult population. *Gastroenterology*, *128*(4 SUPPL. 1), S74–S78. https://doi.org/10.1053/j.gastro.2005.02.016

49. *Çölyak hastaliğinda aile hekimleri için tani, tedavi ve izlem rehberi*. (2019).

50. Rodrigo, L. (2019). Celiac Disease: A Common Unrecognized Health Problem with a Very Delayed Diagnosis. In *Medicina (Kaunas, Lithuania)* (Vol. 56, Issue 1). NLM (Medline). https://doi.org/10.3390/medicina56010009

51. Valitutti, F., Cucchiara, S., & Fasano, A. (n.d.). *Celiac Disease and the Microbiome*. https://doi.org/10.3390/nu11102403

52. Zanchi, C., Ventura, G., Di Leo, G., Orzes, N., Ronfani, L., Not, T., & Ventura, A. (2013). Leonardo da Vinci Meets Celiac Disease. *Journal of Pediatric Gastroenterology and Nutrition*, *56*(2), 206–210. https://doi.org/10.1097/MPG.0b013e31827114a4

53. Kayar, Y., Dertli, R., Sürmeli, N., & Bilgili, M. A. (2019). Extraintestinal Manifestations Associated with Celiac Disease. *East J Med*, *24*(4), 478–483. https://doi.org/10.5505/ejm.2019.26122

54. *Treatment of Celiac Disease*. (n.d.).

55. Bekdas, M., Unal, F., & Demircioglu, F. (2017). Investigation of celiac disease according to marsh classification in childhood. *Bangladesh Journal of Medical Science*, *16*(2), 259–265. https://doi.org/10.3329/bjms.v16i2.26700

56. Wu, D., Chen, S., Zhang, Y., Zhang, H., Wang, Q., Li, J., Fu, Y., Wang, S., Yang, H., Du, H., Zhu, H., Pan, H., Shen, Z., & Feng, M. (2021). *Personalized Medicine Article Facial Recognition Intensity in Disease Diagnosis Using Automatic Facial Recognition*. https://doi.org/10.3390/jpm11111172

57. Hargrave, R., Maddock, R. J., & Stone, V. (2002). Impaired recognition of facial expressions of emotion in Alzheimer's disease. *Journal of Neuropsychiatry and Clinical Neurosciences*, *14*(1), 64–71. https://doi.org/10.1176/jnp.14.1.64

58. Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. In *Eurasip Journal on Advances in Signal Processing* (Vol. 2016, Issue 1). Springer International Publishing. https://doi.org/10.1186/s13634-016-0355-x

59. Zheng, J., Shen, F., Fan, H., & Zhao, J. (2013). An online incremental learning support vector machine for large-scale data. *Neural Computing and Applications*, *22*(5), 1023–1035. https://doi.org/10.1007/s00521-011-0793-1

60. Haykin, S., York, N., San, B., London, F., Sydney, T., Singapore, T., Mexico, M., Munich, C., Cape, P., Hong, T., & Montreal, K. (2009). *Neural Networks and Learning Machines Third Edition*. http://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009.pdf

61. Tulasi Krishna, S. (2019). *Deep Learning and Transfer Learning Approaches for Image Classification*. https://www.researchgate.net/publication/333666150

62. Bengio, Y. (2012). *Deep Learning of Representations for Unsupervised and Transfer Learning* (Vol. 27). http://www.causality.inf.ethz.ch/unsupervised-learning.php

63. Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. In *Zeitschrift fur Medizinische Physik* (Vol. 29, Issue 2, pp. 102–127). Elsevier GmbH. https://doi.org/10.1016/j.zemedi.2018.11.002

64. Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. In *IEEE Access* (Vol. 7, pp. 53040–53065). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2019.2912200

65. Wang, J., Cao, B., Yu, P. S., Sun, L., Bao, W., & Zhu, X. (2018). *Deep Learning Towards Mobile Applications*. http://arxiv.org/abs/1809.03559

66. Stoleru, C. A., Dulf, E. H., & Ciobanu, L. (2022). Automated detection of celiac disease using Machine Learning Algorithms. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-07199-z

67. Vicnesh, J., Wei, J. K. E., Ciaccio, E. J., Oh, S. L., Bhagat, G., Lewis, S. K., Green, P. H., & Acharya, U. R. (2019). Automated diagnosis of celiac disease by video capsule endoscopy using DAISY Descriptors. *Journal of Medical Systems*, *43*(6), 1–9. https://doi.org/10.1007/s10916-019-1285-6

68. Koh, J. E. W., Hagiwara, Y., Oh, S. L., Tan, J. H., Ciaccio, E. J., Green, P. H., Lewis, S. K., & Rajendra Acharya, U. (2019). Automated diagnosis of celiac disease using DWT and nonlinear features with video capsule endoscopy images. *Future Generation Computer Systems*, *90*, 86–93. https://doi.org/10.1016/j.future.2018.07.044

69. Molder, A., Balaban, D. V., Molder, C. C., Jinga, M., & Robin, A. (2023). Computer-Based Diagnosis of Celiac Disease by Quantitative Processing of Duodenal Endoscopy Images. *Diagnostics*, *13*(17). https://doi.org/10.3390/diagnostics13172780

70. Vécsei, A., Fuhrmann, T., Liedlgruber, M., Brunauer, L., Payer, H., & Uhl, A. (2009). Fourier feature vectors. *Computer Methods and Programs in Biomedicine*, *95*, 68–78. https://doi.org/10.1016/j.cmpb.2009.02.017

71. Koh, J. E. W., De Michele, S., Sudarshan, V. K., Jahmunah, V., Ciaccio, E. J., Ooi, C. P., Gururajan, R., Gururajan, R., Oh, S. L., Lewis, S. K., Green, P. H., Bhagat, G., & Acharya, U. R. (2021). Automated interpretation of biopsy images for the detection of celiac disease using a machine learning approach. *Computer Methods and Programs in Biomedicine*, *203*. https://doi.org/10.1016/j.cmpb.2021.106010

72. IEEE Engineering in Medicine and Biology Society, IEEE Signal Processing Society, IEEE International Conference on Service-Oriented Computing and Applications, IEEE International Symposium on Biomedical Imaging 11 2014.04.29-05.02 Beijing, & ISBI 11 2014.04.29-05.02 Beijing. (n.d.). *IEEE 11th International Symposium on Biomedical Imaging (ISBI), 2014 April 29, 2014 - May 2, 2014, Renaissance Beijing Capital Hotel, Beijing, China*.

73. Tenório, J. M., Hummel, A. D., Cohrs, F. M., Sdepanian, V. L., Pisa, I. T., & De Fátima Marin, H. (2011). Artificial intelligence techniques applied to the development of a decision-support system for diagnosing celiac disease. *International Journal of Medical Informatics*, *80*(11), 793–802. https://doi.org/10.1016/j.ijmedinf.2011.08.001

74. Piccialli, F., Calabrò, F., Crisci, D., Cuomo, S., Prezioso, E., Mandile, R., Troncone, R., Greco, L., & Auricchio, R. (123 C.E.). Precision medicine and machine learning towards the prediction of the outcome of potential celiac disease Potential Celiac Patients (PCD) bear the Celiac Disease (CD) genetic predisposition, a significant production of antihuman transglutaminase antibodies, but no morphological changes in the. *Scientific Reports |*, *11*, 5683. https://doi.org/10.1038/s41598-021-84951-x

75. Özer, Ö., & Arda, N. (2022). Comparison of Different Machine Learning Algorithms to Predict the Diagnostic Accuracy Parameters of Celiac Serological Tests. *SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES*, *5*(1). https://doi.org/10.35377/saucis.05.01.1094043

76. Hujoel, I. A., Murphree Jr, D. H., Van Dyke, C. T., Seon Choung, R., Sharma, A., Murray, J. A., & Rubio-Tapia, A. (n.d.). *Machine Learning in Detection of Undiagnosed Celiac Disease*. https://doi.org/10.1016/j.cgh.2017.12.022

77. Kumar, R., Wang, H. C., Mukundan, B., Gupta, S. K., & Kumari, C. S. (2023). Effective machine learning-based skin disease diagnosis using PyTorch. *Journal of Physics: Conference Series*, *2595*(1). https://doi.org/10.1088/1742-6596/2595/1/012008

78. Kalhan, S., Joseph, P., Sharma, S., Dubey, S., Dudani, S., & Dixit, M. (2011). Comparative study of histopathological Marsh grading with clinical and serological parameters in celiac iceberg of north India. *Indian Journal of Pathology and Microbiology*, *54*(2), 279–283. https://doi.org/10.4103/0377-4929.81593

79. Marsh, M. N. (1992). *Gluten, Major Histocompatibility and the Smal1 Intestine Complex,*. 330–354.

80. Punia, P., Bala, K., Verma, M., Nandi, A., Mahlotra, P., Singh, S., Garg, S., Parmar, A., & Kumar, D. (n.d.). *Feasibility of a "No-Biopsy" Approach for the Diagnosis of Celiac Disease in Symptomatic Adults*. https://doi.org/10.7759/cureus.54578

81. Alhabbal, A., & Khamis, I. A. (2022). Evaluating the Correlation between High Titers of Tissue-Transglutaminase Antibody with the Grade of Severity of Villous Atrophy in Syrian Patients with Celiac Disease. *Russian Journal of Gastroenterology, Hepatology, Coloproctology*, *32*(1), 34–40. https://doi.org/10.22416/1382-4376-2022-32-1-34-40

82. Bengio, Y. (2012). *Deep Learning of Representations for Unsupervised and Transfer Learning*. *27*, 17–37. http://www.causality.inf.ethz.ch/unsupervised-learning.php

83. Desai, C. (2021). Image Classification Using Transfer Learning and Deep Learning. *International Journal of Engineering and Computer Science*, *10*(9), 25394–25398. https://doi.org/10.18535/ijecs/v10i9.4622

84. Gadermayr, M., Wimmer, G., Kogler, H., Ecsei, A. V, Merhof, D., & Uhl, A. (2018). *Automated classification of celiac disease during upper endoscopy: Status quo and quo vadis*. https://doi.org/10.1016/j.compbiomed.2018.04.020

85. Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: a literature review. In *BMC Medical Imaging* (Vol. 22, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12880-022-00793-7

86. Vimal, C., & Shirivastava, N. (2022). Face and Face-mask Detection System using VGG-16 Architecture based on Convolutional Neural Network. In *International Journal of Computer Applications* (Vol. 183, Issue 50).

87. *LearnDataSci*. (n.d.), *Hands-on Transfer Learning with Keras and the VGG16 Model*, from https://www.learndatasci.com/tutorials/hands-on-transfer-learning-keras/, Retrieved April 3, 2024

88. Cires¸an, D. C. C., Meier, U., Masci, J., Gambardella, L. M., & Urgen Schmidhuber, J. ¨. (n.d.). *Flexible, High Performance Convolutional Neural Networks for Image Classification*.

89. Su, Z., Liang, B., Shi, F., Gelfond, J., Šegalo, S., Wang, J., Jia, P., & Hao, X. (2021). Deep learning-based facial image analysis in medical research: A systematic review protocol. In *BMJ Open* (Vol. 11, Issue 11). BMJ Publishing Group. https://doi.org/10.1136/bmjopen-2020-047549

90. *Google for Developers*. (n.d.), *Image classification model customization guide | MediaPipe |*, from https://developers.google.com/mediapipe/solutions/customization/image_classifier, Retrieved February 19, 2024

91. Wang, X., Qian, H., Ciaccio, E. J., Lewis, S. K., Bhagat, G., Green, P. H., Xu, S., Huang, L., Gao, R., & Liu, Y. (2020). Celiac disease diagnosis from videocapsule endoscopy images with residual learning and deep feature extraction. *Computer Methods and Programs in Biomedicine*, *187*, 105236. https://doi.org/10.1016/j.cmpb.2019.105236

92. Amirkhani, A., Mosavi, M. R., Mohammadi, K., & Papageorgiou, E. I. (2018). A novel hybrid method based on fuzzy cognitive maps and fuzzy clustering algorithms for grading celiac disease. *Neural Computing and Applications*, *30*(5), 1573–1588. https://doi.org/10.1007/s00521-016-2765-y

93. Garg, S., Saxena, A., & Gupta, R. (2023). Yoga pose classification: a CNN and MediaPipe inspired deep learning approach for real-world application. *Journal of*

*Ambient Intelligence and Humanized Computing*, *14*(12), 16551–16562. https://doi.org/10.1007/s12652-022-03910-0

94. Grotra, R., Vijay, P., Gupta, R. K., & Malhotra, B. (2024). Positive association of serum IgG4 levels with modified marsh classification system in children with gluten enteropathy: a single centre experience. *International Journal of Contemporary Pediatrics*, *11*(5), 545–548. https://doi.org/10.18203/2349-3291.ijcp20241034

95. Bond, K., & Sheta, A. (2021). Medical Data Classification using Machine Learning Techniques. *International Journal of Computer Applications*, *183*(6), 1–8. https://doi.org/10.5120/ijca2021921339

96. Khanna, A., Gupta, D., Bhattacharyya, S., Snasel, V., Platos, J., Aboul, ·, & Hassanien, E. (n.d.). *Advances in Intelligent Systems and Computing 1087* (Vol. 1). http://www.springer.com/series/11156

97. *MediaPipe Studio*. (n.d.), from https://mediapipe-studio.webapps.google.com/demo/image_classifier?hl=tr, Retrieved May 20, 2024

# APPENDICES

## APPENDIX-1: SOURCE CODES

### VGG16 CODES:

```
import tensorflow as tf
from tensorflow.keras.layers import Input, Flatten, Dense
from tensorflow.keras.models import Model, load_model
from tensorflow.keras.applications.vgg16 import VGG16, preprocess_input
from tensorflow.keras.preprocessing.image import ImageDataGenerator
import numpy as np
from glob import glob
import matplotlib.pyplot as plt
from sklearn.metrics import classification_report, confusion_matrix
import seaborn as sns
import time

# Re-size all the images to this
IMAGE_SIZE = [224, 224]

train_path = r'C:\Users\Ebiom\Desktop\dataset8\train'
valid_path = r'C:\Users\Ebiom\Desktop\dataset8\test'

# Record the start time
start_time = time.time()

# Add preprocessing layer to the front of VGG
vgg = VGG16(input_shape=IMAGE_SIZE + [3], weights='imagenet', include_top=False)

# Don't train existing weights
for layer in vgg.layers:
    layer.trainable = False

# Useful for getting number of classes
folders = glob(r'C:\Users\Ebiom\Desktop\dataset8\train\*')

# Our layers - you can add more if you want
x = Flatten()(vgg.output)
```

```
x = Dense(1000, activation='relu')(x)
prediction = Dense(len(folders), activation='softmax')(x)

# Create a model object
model = Model(inputs=vgg.input, outputs=prediction)

# View the structure of the model
model.summary()

# Tell the model what cost and optimization method to use
model.compile(
   loss='categorical_crossentropy',
   optimizer='adam',
   metrics=['accuracy']
)

# Define ImageDataGenerator without repeat
train_datagen = ImageDataGenerator(
   rescale=1. / 255,
   shear_range=0.2,
   zoom_range=0.2,
   horizontal_flip=True
)

test_datagen = ImageDataGenerator(rescale=1. / 255)

training_set = train_datagen.flow_from_directory(train_path,
                                                 target_size=(224, 224),
                                                 batch_size=32,
                                                 class_mode='categorical')

test_set = test_datagen.flow_from_directory(valid_path,
                                            target_size=(224, 224),
                                            batch_size=32,
                                            class_mode='categorical')

# Define number of epochs
epochs = 10

# Define steps per epoch and validation steps
steps_per_epoch = len(training_set)
validation_steps = len(test_set)
```

```
# Initialize lists to store metrics
training_losses = []
training_accuracies = []
validation_losses = []
validation_accuracies = []

# Training loop
for epoch in range(epochs):
    print(f"Epoch {epoch + 1}/{epochs}")

    # Training
    training_loss = 0.0
    training_accuracy = 0.0
    training_batches = 0

    # Loop over batches in the training set
    for _ in range(steps_per_epoch):
        try:
            X_batch, y_batch = next(training_set)
            loss, accuracy = model.train_on_batch(X_batch, y_batch)
            training_loss += loss
            training_accuracy += accuracy
            training_batches += 1
        except StopIteration:
            # Restart the generator if it reaches the end
            training_set.reset()

    # Calculate average training loss and accuracy
    training_loss /= training_batches
    training_accuracy /= training_batches

    print(f"Training Loss: {training_loss:.4f}, Accuracy: {training_accuracy:.4f}")

    # Store training loss and accuracy
    training_losses.append(training_loss)
    training_accuracies.append(training_accuracy)

    # Validation
    validation_loss, validation_accuracy = model.evaluate(test_set, steps=validation_steps)

    print(f"Validation Loss: {validation_loss:.4f}, Accuracy: {validation_accuracy:.4f}")

    # Store validation loss and accuracy
```

```
    validation_losses.append(validation_loss)
    validation_accuracies.append(validation_accuracy)

# Evaluate accuracy on the test set
test_loss, test_accuracy = model.evaluate(test_set)
print("Test Accuracy:", test_accuracy)

# Save the model
model.save('facefeatures_new_model.h5')

# Record the end time
end_time = time.time()

# Calculate the total running time
running_time = end_time - start_time
print(f"Total Running Time: {running_time:.2f} seconds")

# Plot training and validation loss and accuracy
epochs_range = range(epochs)

plt.figure(figsize=(12, 6))

plt.subplot(1, 2, 1)
plt.plot(epochs_range, training_losses, label='Training Loss')
plt.plot(epochs_range, validation_losses, label='Validation Loss')
plt.legend(loc='upper right')
plt.title('Training and Validation Loss')

plt.subplot(1, 2, 2)
plt.plot(epochs_range, training_accuracies, label='Training Accuracy')
plt.plot(epochs_range, validation_accuracies, label='Validation Accuracy')
plt.legend(loc='upper right')
plt.title('Training and Validation Accuracy')

plt.show()

# Get the ground truth labels and predictions
Y_pred = model.predict(test_set, steps=validation_steps)
y_pred = np.argmax(Y_pred, axis=1)

# Get true labels
y_true = test_set.classes
```

```
# Print classification report
print('Classification Report')
target_names = list(test_set.class_indices.keys())
print(classification_report(y_true, y_pred, target_names=target_names))

# Print confusion matrix
print('Confusion Matrix')
conf_matrix = confusion_matrix(y_true, y_pred)
print(conf_matrix)

# Plot confusion matrix
plt.figure(figsize=(10, 8))
sns.heatmap(conf_matrix, annot=True, fmt='d', xticklabels=target_names,
yticklabels=target_names)
plt.xlabel('Predicted')
plt.ylabel('True')
plt.show()
```

**MMM MobileNetV2 CODES**

```
from google.colab import drive

drive.mount('/content/drive')

!python --version

!pip install --upgrade pip

!pip install mediapipe-model-maker

from google.colab import files

import os

import tensorflow as tf

assert tf.__version__.startswith('2')


from mediapipe_model_maker import image_classifier


import matplotlib.pyplot as plt

image_path = '/content/drive/MyDrive/Colab Notebooks/face_images-20240115T205912Z-
001/face_images/face_images'
```

```
image_path = os.path.join(os.path.dirname(image_path), 'face_images')
print(image_path)
labels = []
for i in os.listdir(image_path):
  if os.path.isdir(os.path.join(image_path, i)):
    labels.append(i)
print(labels)
NUM_EXAMPLES = 2

for label in labels:
  label_dir = os.path.join(image_path, label)
  example_filenames = os.listdir(label_dir)[:NUM_EXAMPLES]
  fig, axs = plt.subplots(1, NUM_EXAMPLES, figsize=(10,2))
  for i in range (NUM_EXAMPLES):
    axs[i].imshow(plt.imread(os.path.join(label_dir, example_filenames[i])))
    axs[i].get_xaxis().set_visible(False)
    axs[i].get_yaxis().set_visible(False)
  fig.suptitle(f'Showing {NUM_EXAMPLES} examples for {label}')

plt.show()
data = image_classifier.Dataset.from_folder(image_path)
train_data, remaining_data = data.split(0.8)
test_data, validation_data = remaining_data.split(0.5)
spec = image_classifier.SupportedModels.MOBILENET_V2
hparams = image_classifier.HParams(export_dir="exported_model")
options = image_classifier.ImageClassifierOptions(supported_model=spec,
hparams=hparams)
model = image_classifier.ImageClassifier.create(
    train_data = train_data,
```

```
    validation_data = validation_data,
    options=options,
)
loss, acc = model.evaluate(test_data)
print (f'Test loss:{loss}, Test accuracy:{acc}')
model.export_model()
!ls exported_model
files.download('exported_model/model.tflite')
pip install matplotlib
import matplotlib.pyplot as plt
from sklearn.metrics import precision_score, recall_score, f1_score
import numpy as np

# Data for the epochs
epochs = list(range(1, 11))
testing_loss = [0.4274, 0.4466, 0.3967, 0.4708, 0.4265, 0.4239, 0.4151, 0.5008, 0.4622, 0.4215]
testing_accuracy = [0.8625, 0.8250, 0.8813, 0.8062, 0.8438, 0.8750, 0.8625, 0.8125, 0.8125, 0.8625]
validation_loss = [0.4972, 0.4982, 0.5083, 0.4968, 0.5048, 0.4866, 0.4988, 0.5022, 0.4956, 0.5036]
validation_accuracy = [0.8000, 0.8000, 0.8500, 0.8500, 0.8500, 0.8500, 0.8500, 0.8500, 0.8500, 0.8500]

# Simulated precision, recall, and F1 score data
# These should be calculated based on your model's predictions and true labels
precision = [0.85, 0.83, 0.87, 0.82, 0.85, 0.88, 0.87, 0.82, 0.81, 0.87]
recall = [0.87, 0.85, 0.88, 0.84, 0.86, 0.89, 0.88, 0.83, 0.82, 0.88]
f1_score_vals = [0.86, 0.84, 0.88, 0.83, 0.86, 0.88, 0.88, 0.82, 0.82, 0.88]
```

```
# Plotting Loss and Accuracy
plt.figure(figsize=(14, 8))

plt.subplot(2, 2, 1)
plt.plot(epochs, testing_loss, 'r', label='Testing Loss')
plt.plot(epochs, validation_loss, 'b', label='Validation Loss')
plt.title('Loss over Epochs')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()

plt.subplot(2, 2, 2)
plt.plot(epochs, testing_accuracy, 'r', label='Testing Accuracy')
plt.plot(epochs, validation_accuracy, 'b', label='Validation Accuracy')
plt.title('Accuracy over Epochs')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend()

# Plotting Precision, Recall, and F1 Score
plt.subplot(2, 2, 3)
plt.plot(epochs, precision, 'g', label='Precision')
plt.plot(epochs, recall, 'm', label='Recall')
plt.plot(epochs, f1_score_vals, 'c', label='F1 Score')
plt.title('Precision, Recall, and F1 Score over Epochs')
plt.xlabel('Epochs')
plt.ylabel('Score')
plt.legend()
```

```
plt.tight_layout()
plt.show()
from sklearn.metrics import precision_score, recall_score, f1_score

# Example predictions and true labels
y_true = [0, 1, 1, 0, 1, 1, 0, 0, 1, 0]
y_pred = [0, 1, 1, 1, 0, 1, 0, 0, 1, 0]

precision = precision_score(y_true, y_pred)
recall = recall_score(y_true, y_pred)
f1 = f1_score(y_true, y_pred)

print (f'Precision: {precision}, Recall: {recall}, F1 Score: {f1}')
```

**MACHINE LEARNING CODES FOR CD MARSH LEVEL CLASSIFICATION:**

```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.impute import SimpleImputer
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
import ipywidgets as widgets
from IPython.display import display
```

```
# Step 1: Data Loading and Preparation

# Load the dataset
df = pd.read_csv('RevisedOlddataonlynewlydiagnosedhalfdiet (1).csv')

# Sample data creation for demonstration purposes
# Replace this with loading your actual dataset
df = pd.DataFrame(np.random.rand(183, 25), columns=[
    "gender", "family history", "nausea", "abdominal pain", "bloating", "diarrhea",
    "soft stool", "urgent defecation", "heartburn", "reflux", "farting", "HCT",
    "DVIT", "serum iron", "iron binding capacity", "HGB", "B12", "Folat", "Ferritin",
"magnesium",
    "IGM", "IGA", "IGG", "CRP", "BMI"
])
df['Disease Level'] = np.random.choice([
    'Marsh0', 'Marsh1', 'Marsh2', 'Marsh3a', 'Marsh3b', 'Marsh3c', 'Marsh4'
], size=183)

# Handle Missing Values
imputer = SimpleImputer(strategy='mean')
df.iloc[:, :-1] = imputer.fit_transform(df.iloc[:, :-1])

# Normalize Numerical Features
scaler = StandardScaler()
df.iloc[:, :-1] = scaler.fit_transform(df.iloc[:, :-1])

# Separate features and target
X = df.drop('Disease Level', axis=1).values
y = df['Disease Level']
```

```
# Encode target labels
label_encoder = LabelEncoder()
y_encoded = label_encoder.fit_transform(y)

# Step 2: Model Definitions

# Define the models
models = {
    "Decision Tree": DecisionTreeClassifier(),
    "Random Forest": RandomForestClassifier(),
    "Gradient Boosting": GradientBoostingClassifier(),
    "Naive Bayes": GaussianNB()
}
# Step 3: Model Training
# Train each model
trained_models = {}
for model_name, model in models.items():
    model.fit(X, y_encoded)
    trained_models[model_name] = model

# Evaluate each model and plot confusion matrix
def plot_confusion_matrix(cm, classes, title='Confusion Matrix', cmap=plt.cm.Blues):
    plt.figure(figsize=(8, 6))
    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(classes))
    plt.xticks(tick_marks, classes, rotation=45)
```

```
    plt.yticks(tick_marks, classes)

    fmt = 'd'
    thresh = cm.max() / 2.
    for i, j in np.ndindex(cm.shape):
        plt.text(j, i, format(cm[i, j], fmt),
                 ha="center", va="center",
                 color="white" if cm[i, j] > thresh else "black")

    plt.ylabel('True label')
    plt.xlabel('Predicted label')
    plt.tight_layout()

for model_name, model in trained_models.items():
    y_pred = model.predict(X)
    accuracy = accuracy_score(y_encoded, y_pred)
    cm = confusion_matrix(y_encoded, y_pred)
    print(f'{model_name} accuracy: {accuracy:.2f}')
    plot_confusion_matrix(cm, classes=label_encoder.classes_, title=f'{model_name}
Confusion Matrix')
    plt.show()
    report = classification_report(y_encoded, y_pred, target_names=label_encoder.classes_)
    print(report)

# Step 4: Prediction Interface

# Define the columns for new data input
columns = [
    "gender", "family history", "nausea", "abdominal pain", "bloating", "diarrhea",
```

```
    "soft stool", "urgent defecation", "heartburn", "reflux", "farting", "HCT",
    "DVIT", "serum iron", "iron binding capacity", "HGB", "B12", "Folat", "Ferritin", "magnesium",
    "IGM", "IGA", "IGG", "CRP", "BMI"
]
# Define the type and options for each column
column_types = {
    "gender": "dropdown",
    "family history": "dropdown",
    "nausea": "dropdown",
    "abdominal pain": "dropdown",
    "bloating": "dropdown",
    "diarrhea": "dropdown",
    "soft stool": "dropdown",
    "urgent defecation": "dropdown",
    "heartburn": "dropdown",
    "reflux": "dropdown",
    "farting": "dropdown",
    "HCT": "float",
    "DVIT": "float",
    "serum iron": "float",
    "iron binding capacity": "float",
    "HGB": "float",
    "B12": "float",
    "Folat": "float",
    "Ferritin": "float",
    "magnesium": "float",
    "IGM": "float",
    "IGA": "float",
```

```
    "IGG": "float",
    "CRP": "float",
    "BMI": "float",
}
column_options = {
    "gender": ['Female', 'Male'],
    "family history": ['Yes', 'No'],
    "nausea": ['none', 'mild', 'moderate', 'severe'],
    "abdominal pain": ['none', 'mild', 'moderate', 'severe'],
    "bloating": ['none', 'mild', 'moderate', 'severe'],
    "diarrhea": ['none', 'mild', 'moderate', 'severe'],
    "soft stool": ['none', 'mild', 'moderate', 'severe'],
    "urgent defecation": ['none', 'mild', 'moderate', 'severe'],
    "heartburn": ['none', 'mild', 'moderate', 'severe'],
    "reflux": ['none', 'mild', 'moderate', 'severe'],
    "farting": ['none', 'mild', 'moderate', 'severe'],
}
# Create widgets for each column
widgets_dict = {}
for col in columns:
    if column_types[col] == "dropdown":
        options = column_options.get(col, [])
        widgets_dict[col] = widgets.Dropdown(options=options, description=col)
    elif column_types[col] == "int":
        widgets_dict[col] = widgets.IntText(description=col)
    elif column_types[col] == "float":
        widgets_dict[col] = widgets.FloatText(description=col)

# Create a dropdown for model selection
```

```
model_selector = widgets.Dropdown(
    options=list(models.keys()),
    description='Model'
)

# Create a button to submit the form
button = widgets.Button(description="Predict")

# Display the widgets
form_items = [model_selector] + [item for item in widgets_dict.values()]
form = widgets.VBox(form_items)
display(form)
display(button)

# Define the prediction function
def on_button_clicked(b):
    # Extract values from widgets
    input_data = []
    for col in columns:
        value = widgets_dict[col].value
        if column_types[col] == "dropdown" and value in ['Yes', 'No']:
            value = 1 if value == 'Yes' else 0
        elif column_types[col] == "dropdown" and col == "gender":
            value = 1 if value == 'Male' else 0
        elif column_types[col] == "dropdown":
            value = ['none', 'mild', 'moderate', 'severe'].index(value)
        input_data.append(value)

    input_data = np.array(input_data).reshape(1, -1)
```

```
    # Normalize the new input data
    input_data = scaler.transform(input_data)

    # Handle missing values in the new input data
    input_data = imputer.transform(input_data)

    # Get the selected model
    selected_model_name = model_selector.value
    selected_model = trained_models[selected_model_name]

    # Make predictions
    predicted = selected_model.predict(input_data)
    predicted_class = label_encoder.inverse_transform(predicted)
    print (f'Predicted Class with {selected_model_name}: {predicted_class[0]}')

# Attach the prediction function to the button
button.on_click(on_button_clicked)
```

**DEEP LEARNING (PYTORCH) CODES FOR MARSH CLASSIFICATION**

```
# Import necessary libraries
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.impute import SimpleImputer
import torch
import torch.nn as nn
import torch.optim as optim
```

```
from torch.utils.data import DataLoader, TensorDataset

from sklearn.metrics import precision_score, recall_score, f1_score, confusion_matrix, ConfusionMatrixDisplay

import matplotlib.pyplot as plt

import seaborn as sns

import ipywidgets as widgets

from IPython.display import display


# Load the dataset

# Replace this with your actual data loading method

df = pd.read_csv('RevisedOlddataonlynewlydiagnosedhalfdiet (1).csv')


# Sample data creation for demonstration purposes

# Replace this with loading your actual dataset

df = pd.DataFrame(np.random.rand(183, 25), columns=["gender", "family history", "nausea", "abdominal pain",

"bloating", "diarrhea", "soft stool", "urgent defecation", "heartburn", "reflux", "farting", "HCT", "DVIT", "serum iron", "iron binding capacity",

"HGB", "B12", "Folat", "Ferritin", "magnesium", "IGM", "IGA", "IGG", "CRP", "BMI"

])

df['Disease Level'] = np.random.choice(['Marsh0', 'Marsh1', 'Marsh2', 'Marsh3a', 'Marsh3b', 'Marsh3c', 'Marsh4'], size=183)


# Data Cleaning and Preparation

# 1. Handle Missing Values

imputer = SimpleImputer(strategy='mean')

df.iloc[:, :-1] = imputer.fit_transform(df.iloc[:, :-1])


# 2. Normalize Numerical Features

scaler = StandardScaler()

df.iloc[:, :-1] = scaler.fit_transform(df.iloc[:, :-1])
```

```
# Separate features and target
X = df.drop('Disease Level', axis=1).values
y = df['Disease Level']

# Encode target labels
label_encoder = LabelEncoder()
y_encoded = label_encoder.fit_transform(y)

# Convert to PyTorch tensors
X_tensor = torch.tensor(X, dtype=torch.float32)
y_tensor = torch.tensor(y_encoded, dtype=torch.long)

# Create a DataLoader
dataset = TensorDataset(X_tensor, y_tensor)
train_loader = DataLoader(dataset, batch_size=32, shuffle=True)

# Define a simple neural network
class SimpleNN(nn.Module):
    def __init__(self, input_dim, output_dim):
        super (SimpleNN, self).__init__()
        self.fc1 = nn.Linear(input_dim, 128)
        self.fc2 = nn.Linear(128, 64)
        self.fc3 = nn.Linear(64, output_dim)

    def forward (self, x):
        x = torch.relu(self.fc1(x))
        x = torch.relu(self.fc2(x))
        x = self.fc3(x)
        return x
```

```
# Initialize the neural network
input_dim = X.shape[1]
output_dim = len(label_encoder.classes_)
model = SimpleNN(input_dim, output_dim)

# Define loss function and optimizer
criterion = nn.CrossEntropyLoss()
optimizer = optim.Adam(model.parameters(), lr=0.001)

# Train the model
num_epochs = 20
for epoch in range(num_epochs):
    for batch_X, batch_y in train_loader:
        optimizer.zero_grad()
        outputs = model(batch_X)
        loss = criterion (outputs, batch_y)
        loss.backward()
        optimizer.step()
    print (f'Epoch {epoch+1}/{num_epochs}, Loss: {loss.item():.4f}')

# Evaluate the model
model.eval()
with torch.no_grad():
    outputs = model(X_tensor)
    _, predicted = torch.max(outputs, 1)
    accuracy = (predicted == y_tensor).float().mean()
    print (f'Test accuracy: {accuracy:.2f}')

# Calculate Precision, Recall, F1-Score, and Confusion Matrix
```

```
y_true = y_tensor.numpy()
y_pred = predicted.numpy()

precision = precision_score(y_true, y_pred, average='weighted')
recall = recall_score(y_true, y_pred, average='weighted')
f1 = f1_score(y_true, y_pred, average='weighted')
conf_matrix = confusion_matrix(y_true, y_pred)

print (f'Precision: {precision:.2f}')
print (f'Recall: {recall:.2f}')
print (f'F1-Score: {f1:.2f}')

# Visualization
def plot_metrics():
    # Plot confusion matrix
    plt.figure(figsize=(10, 7))
    sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues',
xticklabels=label_encoder.classes_, yticklabels=label_encoder.classes_)
    plt.xlabel('Predicted')
    plt.ylabel('True')
    plt.title('Confusion Matrix')
    plt.show()

    # Plot Precision, Recall, F1-Score
    metrics = {'Precision': precision, 'Recall': recall, 'F1-Score': f1}
    plt.figure(figsize=(8, 5))
    plt.bar(metrics.keys(), metrics.values(), color=['blue', 'orange', 'green'])
    plt.ylim(0, 1)
    for i, v in enumerate(metrics.values()):
        plt.text(i, v + 0.02, f"{v:.2f}", ha='center', va='bottom')
```

```
    plt.title('Precision, Recall, F1-Score')
    plt.show()

plot_metrics()

# Define the columns for new data input
columns = ["gender", "family history", "nausea", "abdominal pain",
"bloating", "diarrhea", "soft stool", "urgent defecation", "heartburn", "reflux", "farting",
"HCT", "DVIT", "serum iron", "iron binding capacity",
"HGB", "B12", "Folat", "Ferritin", "magnesium", "IGM", "IGA", "IGG", "CRP", "BMI"
]
# Define the type and options for each column
column_types = {
    "gender": "dropdown",
    "family history": "dropdown",
    "nausea": "dropdown",
    "abdominal pain": "dropdown",
    "bloating": "dropdown",
    "diarrhea": "dropdown",
    "soft stool": "dropdown",
    "urgent defecation": "dropdown",
    "heartburn": "dropdown",
    "reflux": "dropdown",
    "farting": "dropdown",
    "HCT": "float",
    "DVIT": "float",
    "serum iron": "float",
    "iron binding capacity": "float",
    "HGB": "float",
    "B12": "float",
```

```
    "Folat": "float",
    "Ferritin": "float",
    "magnesium": "float",
    "IGM": "float",
    "IGA": "float",
    "IGG": "float",
    "CRP": "float",
    "BMI": "float",
}
column_options = {
    "gender": ['Female', 'Male'],
   "family history": ['Yes', 'No'],
    "nausea": ['none', 'mild', 'moderate', "severe"],
    "abdominal pain": ['none', 'mild', 'moderate', "severe"],
    "bloating": ['none', 'mild', 'moderate', "severe"],
    "diarrhea": ['none', 'mild', 'moderate', "severe"],
    "soft stool": ['none', 'mild', 'moderate', "severe"],
    "urgent defecation": ['none', 'mild', 'moderate', "severe"],
    "heartburn": ['none', 'mild', 'moderate', "severe"],
    "reflux": ['none', 'mild', 'moderate', "severe"],
    "farting": ['none', 'mild', 'moderate', "severe"],
}
# Create widgets for each column
widgets_dict = {}
for col in columns:
    if column_types[col] == "dropdown":
        options = column_options.get(col, [])
        widgets_dict[col] = widgets.Dropdown(options=options, description=col)
    elif column_types[col] == "int":
        widgets_dict[col] = widgets.IntText(description=col)
```

```
    elif column_types[col] == "float":
        widgets_dict[col] = widgets.FloatText(description=col)

# Create a button to submit the form
button = widgets.Button(description="Predict")

# Display the widgets
form_items = [item for item in widgets_dict.values()]
form = widgets.VBox(form_items)
display(form)
display(button)

# Define the prediction function
def on_button_clicked(b):
    # Extract values from widgets
    input_data = []
    for col in columns:
        value = widgets_dict[col].value
        if column_types[col] == "dropdown" and value in ['Yes', 'No']:
            value = 1 if value == 'Yes' else 0
        elif column_types[col] == "dropdown" and col == "gender":
            value = 1 if value == 'Male' else 0
        elif column_types[col] == "dropdown":
            value = ['none', 'mild', 'moderate', 'severe']. index(value)
        input_data.append(value)

    input_data = np.array(input_data).reshape(1, -1)

    # Normalize the new input data
    input_data = scaler.transform(input_data)
```

```
    # Handle missing values in the new input data
    input_data = imputer.transform(input_data)

    # Make predictions
    model.eval()
    input_tensor = torch.tensor(input_data, dtype=torch.float32)
    with torch.no_grad():
        outputs = model(input_tensor)
        _, predicted = torch.max(outputs, 1)
        predicted_class = label_encoder.inverse_transform(predicted.numpy())
        print(f'Predicted Class: {predicted_class[0]}')

# Attach the prediction function to the button
button.on_click(on_button_clicked)
```

## APPENDIX-2: CONSENT FORM EXAMPLE

**Sağlık Bilimleri Üniversitesi**

**Van Eğitim ve Araştırma Hastanesi Klinik Araştırmaları Etik Kurulu**

**BİLGİLENDİRİLMİŞ GÖNÜLLÜ ONAM FORMU**

Sizi **Elif Keskin Bilgiç** tarafından İstanbul Üniversitesi Biyomedikal Mühendisliği ABD Doktora programı kapsamında **Dr. Öğrt Üyesi İnci ZAİM GÖKBAY** danışmanlığında yürütülen ve klinik çalışmaları **DR: Yusuf KAYAR denetiminde gerçekleşecek**" **Çölyak Hastalığının Teşhisine Destek Verebilecek Matematik Model Tabanlı KDS Algoritması Tasarımı**" başlıklı doktora tez çalışmasının araştırmasına katılmaya davet ediyoruz. Bu araştırmanın amacı; Çölyak Hastalığı gibi özelleşmiş, erken teşhisinin hayatı, kalitesinden ödün vermeden yaşamaya elverişli bir hale getirmeye destek verebilecek iki fazlı bir mobil uygulama tasarımı yapmaktır. Çalışmaya hastanın yakınmalarına istinaden biyopsi yapılarak teşhisi kesinleşmiş ve çölyak benzeri semptomlar gösteren 18 yaş üzerindeki kadın ve erkekler dahil edilecektir. Van Eğitim Araştırma Hastanesi Gastroenteroloji Kliniğine başvuran hastalardan gönüllülük esasına dayanarak oluşturulacak ortamda anamnez ile birlikte yüz fotoğrafları alınacaktır. Çalışmada kullanılacak örneklem büyüklüğü 100 tane kadın erkek Çölyak Teşhisi konulan bireyler ve 50 tane Çölyak Hastalığına benzer yakınmalarla polikliniğe başvuran ancak kan testleri veya biyopsi sonucunda hastalık teşhisi almamış bireylerden oluşacaktır. Hastalarımızdan beyaz bir duvar önüne yerleştirilmiş sabit (dönme özelliği olmayan standart tip) sandalye üzerine oturtulup, 1 m uzağa tripot üzerine yerleştirilmiş Sony DSC-S2100 12.1MP marka kamera ile yüz fotoğrafları alınacaktır.

Hastadan yüz fotoğrafı alındıktan sonra ilgili personel tarafından, önceki adımda oluşturularak mobil uygulamaya geçirilmiş anketin doldurulması istenecektir. Tüm elde edilen bilgiler 500 GB kapasiteli hard diskte analiz için kaydedilecektir.

Araştırmaya sizin dışınızda 150 kişinin katılması öngörülmektedir.[1] Bu çalışmaya katılmak tamamen **gönüllülük** esasına dayanmaktadır. Çalışmanın amacına ulaşması için sizden beklenen, kimsenin baskısı veya telkini altında olmadan, size en uygun gelen cevapları içtenlikle verecek şekilde cevaplamanızdır. Bu formu okuyup onaylamanız, araştırmaya katılmayı kabul ettiğiniz anlamına gelecektir. Ancak, çalışmaya katılmama veya katıldıktan sonra herhangi bir anda çalışmayı bırakma hakkına da sahipsiniz. Bu çalışmadan elde edilecek bilgiler tamamen araştırma amacı ile kullanılacak olup kişisel bilgileriniz **gizli tutulacaktır**; ancak verileriniz yayın amacı ile kullanılabilir. İletişim bilgileriniz ise sadece izninize bağlı olarak ve farklı araştırmacıların sizinle iletişime geçebilmesi için "ortak katılımcı havuzuna" aktarılabilir. Eğer araştırmanın amacı ile ilgili verilen bu bilgiler dışında şimdi veya sonra daha fazla bilgiye ihtiyaç duyarsanız araştırmacıya şimdi sorabilir veya ……e-posta adresi ve ……numaralı telefondan ulaşabilirsiniz. Araştırma tamamlandığında genel/size özel sonuçların sizinle paylaşılmasını istiyorsanız lütfen araştırmacıya iletiniz.

---

Yukarıda yer alan ve araştırmadan önce katılımcıya verilmesi gereken bilgileri okudum ve katılmam istenen çalışmanın kapsamını ve amacını, gönüllü olarak üzerime düşen sorumlulukları anladım. Çalışma hakkında yazılı ve sözlü açıklama aşağıda adı belirtilen araştırmacı/araştırmacılar tarafından yapıldı. Bana, çalışmanın muhtemel riskleri ve faydaları sözlü olarak da anlatıldı. Kişisel bilgilerimin özenle korunacağı konusunda yeterli güven verildi.

Bu koşullarda söz konusu araştırmaya kendi isteğimle, hiçbir baskı ve telkin olmaksızın katılmayı kabul ediyorum.

Katılımcının [2]**:**
Adı-Soyadı: .....
İmzası: İletişim Bilgileri: e-posta: Telefon:

İletişim bilgilerimin diğer araştırmacıların benimle iletişime geçebilmesi için "ortak araştırma havuzuna" aktarılmasını; ☐ kabul ediyorum ☐ kabul etmiyorum (lütfen uygun seçeneği işaretleyiniz)

Velayet veya Vesayet Altında Bulunanlar İçin;

Veli veya Vasisinin

Adı-Soyadı: .....
İmzası:

Araştırmacının

Adı-Soyadı: Elif Keskin Bilgiç
İmzası:

Şahidin:[3]
Adı-Soyadı: .....
İmzası:

---

[1]Bu cümle yalnızca bir örnek olup bu cümlede araştırmanın amacının ve gerekiyorsa nasıl yapılacağının (örneğin psikometrik test mi, öyle ise kaç soru sorulacağı veya kaç ölçekten oluştuğu; ses kaydı, görüntü alımı, gözlem gibi işlemleri mi içerdiği ve ne kadar süreceği gibi) 3 cümleyi geçmeyecek şekilde kısaca anlatılması beklenilmektedir.

[2]İmza bölümünde ideal olan katılımcının kendisinin imzasının alınmasıdır. Bu durumda onam formunu katılımcı ve araştırmacı imzalar. Katılımcının araştırmaya bireysel olarak katılmayı kabul edip onam formunu imzalamayı istemediği durumlarda şahide ihtiyaç doğar ve bu durumda araştırmacı ve katılımcı yerine şahidin imzalarının olması yeterlidir.

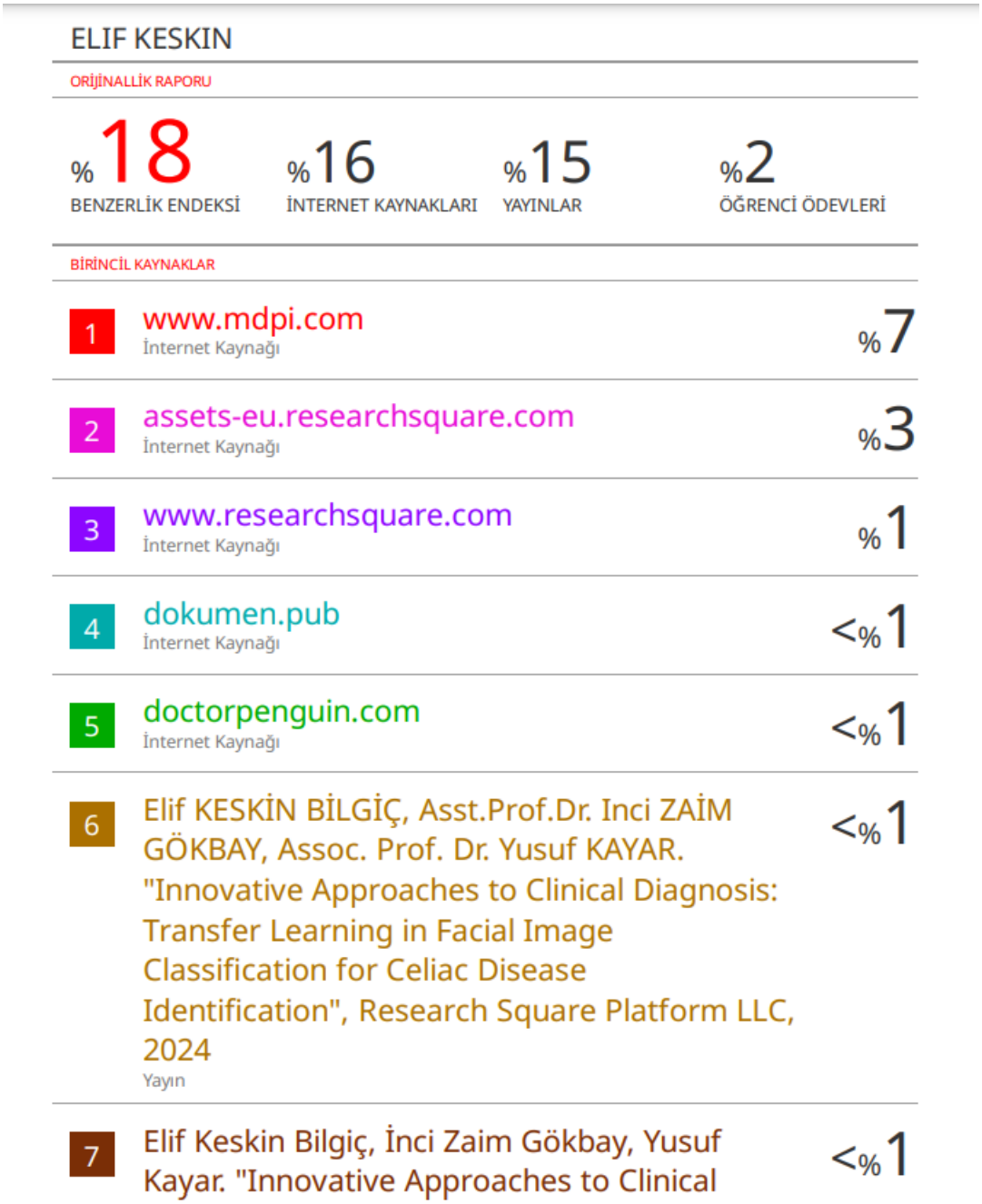
Verilerin yüz yüze iletişim içermeyen; a) Internet ortamında toplanması durumunda katılımcıların uygulama materyallerine erişebilmesi için, online sistemde sunulan bilgilendirilmiş onam formunu okuyup araştırmaya katılmayı onayladıklarına dair ilgili kutucuğu işaretlemeleri gerekmektedir. Bu işaretleme katılımcıların onam imzaları yerine geçer. Katılımcılar onam formunun sonundaki "araştırmaya katılmayı kabul ettiklerine dair" ilgili kutucuğu işaretlemedikleri takdirde onay vermemiş sayılırlar ve bu durumda araştırmaya devam edilmez. b) Telefonla uygulamalarda ise araştırmacı araştırma sorularına geçmeden önce Bilgilendirilmiş Onam Formundaki bilgileri katılımcıya sesli olarak okur. Bu durumda katılımcının sözlü onayı imza yerine geçer. Telefonda bu sözlü onay alınmadığı takdirde uygulamaya geçilmez. Hem Internet, hem telefon hem de benzeri yüz yüze iletişimin olmadığı ortamlarda yapılan uygulamalarda katılımcı onay vermediği takdirde bir şahidin onayına başvurulmaksızın uygulamaya devam edilmez.

Eğer veriler okullarda, kurumlarda vb. ortamlarda aynı anda birden fazla kişiden grup uygulaması şeklinde toplanacaksa, yine tercihen tüm katılımcıların onam formlarını bireysel olarak imzalamaları istenir. Ancak katılımcı sayısının fazlalığı ve bununla birlikte zamanın kısıtlılığı gibi durumlar söz konusu olduğunda araştırmacı tüm gruba onam formundaki bilgileri tek seferde sözlü olarak okumayı ve bir imza listesi dolaştırarak katılımcıların araştırmaya katılmayı kabul ettiklerine dair bu listeye imza atmalarını tercih edebilir. Grup çalışmasında da tercih edilen katılımcının kendisinin imzasıdır, ancak araştırıcının etik kurula tanımlaması gereken ender durumlarda ise şahit, grup adına da imza atabilir. Fakat grup ortamında herkes çalışmaya katılmayı kabul etmeyebilir. Bu durumda sadece araştırmaya katılmayı isteyenlerin çalışmaya alınması ve bu kişiler adına toplu imza alınması gerekmektedir. (Çalışmanızda şahidin imzasını grup adına kullanmak istiyorsanız etik kurula koşullarını açıklamanız gerekmektedir).

[3]Şahit Kriterleri: Çalışmanın bir üyesi olmayan, araştırmacı tarafından belirlenen ve araştırmanın bulguları üzerinde herhangi bir olumlu/olumsuz etki yaratma olasılığı bulunmayan tarafsız yetişkinlerdir. Katılımcı araştırmaya katılmayı kabul edip onam formunu imzalamayı istemediği durumlarda araştırmacı onam formundaki bilgileri katılımcıya sözlü olarak okur. Katılımcı onayladığını sözlü olarak beyan ederse şahit de bu sözlü onam sürecine yazılı onam formunu imzalamak sureti ile şahitlik ettiğini beyan etmiş olur.

NOT: Araştırmacıdan, onam formunun imza kısmında bulunan ikili seçenekten çalışmasına uygun olan alternatifi yazması ve formda yer alan boşlukları çalışmasına uyarlamak yoluyla onam formuna son halini vermesi ve bu şekliyle formu göndermesi beklenilmektedir.

## FIRST PAGE OF THE PLAGIARISM REPORT



ELIF KESKIN

ORİJİNALLİK RAPORU

| %18 | %16 | %15 | %2 |
|---|---|---|---|
| BENZERLİK ENDEKSİ | İNTERNET KAYNAKLARI | YAYINLAR | ÖĞRENCİ ÖDEVLERİ |

BİRİNCİL KAYNAKLAR

| | | |
|---|---|---|
| 1 | www.mdpi.com<br>İnternet Kaynağı | %7 |
| 2 | assets-eu.researchsquare.com<br>İnternet Kaynağı | %3 |
| 3 | www.researchsquare.com<br>İnternet Kaynağı | %1 |
| 4 | dokumen.pub<br>İnternet Kaynağı | <%1 |
| 5 | doctorpenguin.com<br>İnternet Kaynağı | <%1 |
| 6 | Elif KESKİN BİLGİÇ, Asst.Prof.Dr. Inci ZAİM GÖKBAY, Assoc. Prof. Dr. Yusuf KAYAR. "Innovative Approaches to Clinical Diagnosis: Transfer Learning in Facial Image Classification for Celiac Disease Identification", Research Square Platform LLC, 2024<br>Yayın | <%1 |
| 7 | Elif Keskin Bilgiç, İnci Zaim Gökbay, Yusuf Kayar. "Innovative Approaches to Clinical | <%1 |

| **Publications** |
| --- |
| 1-Keskin Bilgiç, E., Zaim Gökbay, İ., & Kayar, Y. (2024). Innovative Approaches to Clinical Diagnosis: Transfer Learning in Facial Image Classification for Celiac Disease Identification. *Applied Sciences*, *14*(14), 6207. https://doi.org/10.3390/app14146207 https://www.mdpi.com/2076-3417/14/14/6207/pdf<br>2-Keskin, E. (2019). *CLINICAL DECISION SUPPORT SYSTEMS IN DIAGNOSIS OF*. *November*, 239–250, DOI:10.13140/RG.2.2.18411.08481<br>3-Katana, D., & Keskin, E. (2018). *Investigation of the therapeutic effects of L-Dopa and Lawsone in wound healing*, https://www.researchgate.net/publication/328148225 |