ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

Machine Learning Techniques For Medical Diagnosis Of Diabetes Using Iris Images

MMI 700: Research Methods

Ümit Mert Çağlar

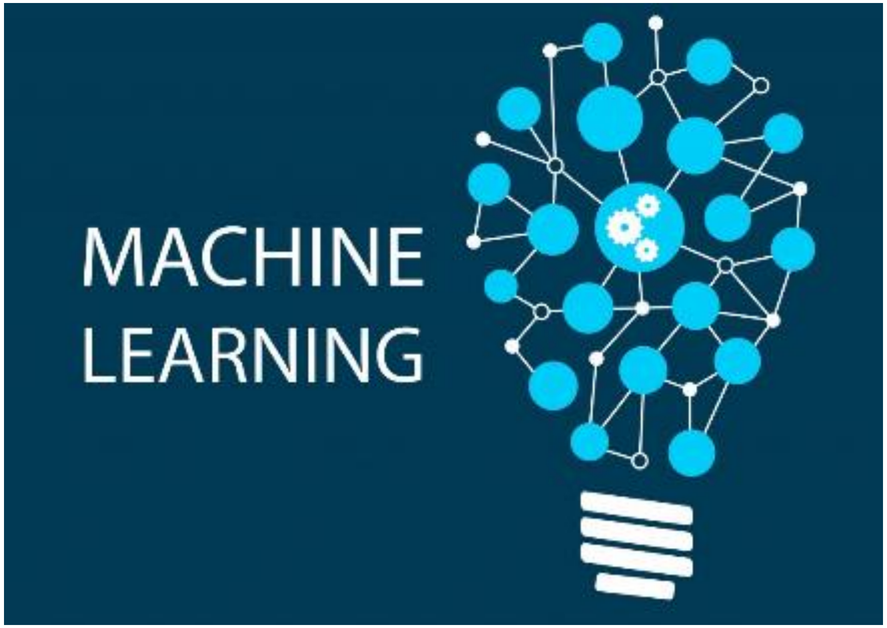
Lecturer:
Assoc. Prof. Hüseyin Hacıhabiboğlu

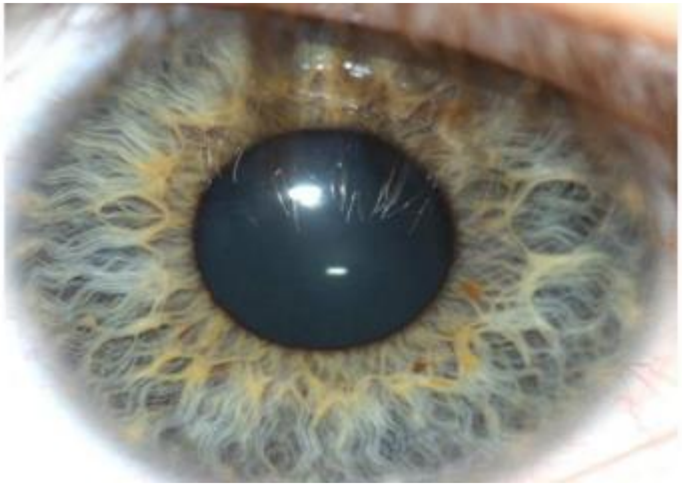
Graduate School of Informatics, METU

Mert Çağlar

Introduction

- Introduction
- Abstract
- Discussion
- Results
- Conclusion





Introduction

Machine learning techniques for medical diagnosis of diabetes using iris images

Iris
Diabetes
Classification
Segmentation
Feature extraction
Disease diagnosis

Piyush Samant, Ravinder Agarwal
Thapar University, Patiala, India

Accepted 10 January 2018

- **Circular Hough transform** used for segmentation,
- **Daugman's rubber sheet model** used for normalization,
- **Gabor filter** and **2-D DWT** based features

https://www.sciencedirect.com/science/article/pii/S0169260717304649

Mert Çağlar

Introduction

- Diabetes:
  - 381.8 million people are affected
  - excessive blood sugar level
  - late diagnosis leads to advert effects
  - early diagnosis is a challenge

Prevalence of diabetes worldwide (per 1000 inhabitants).

Mert Çağlar

Abstract

- Diabetes is about pancreatic enzyme Insulin
- Complementary-Alternative Medicine (CAM)
- Digital image processing
- Best classification accuracy 89.63% RF classifier
- Effective and diagnostically significant model for non-invasive and automatic diabetes diagnosis.

Classification

Identifying to which category an object belongs to.

Applications: Spam detection, Image recognition.
Algorithms: SVM, nearest neighbors, random forest, … — Examples

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.
Algorithms: SVR, ridge regression, Lasso, … — Examples

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes
Algorithms: k-Means, spectral clustering, mean-shift, … — Examples

Dimensionality reduction

Reducing the number of random variables to consider.

Applications: Visualization, Increased efficiency
Algorithms: PCA, feature selection, non-

Model selection

Comparing, validating and choosing parameters and models.

Goal: Improved accuracy via parameter tuning
Modules: grid search, cross validation,

Preprocessing

Feature extraction and normalization.

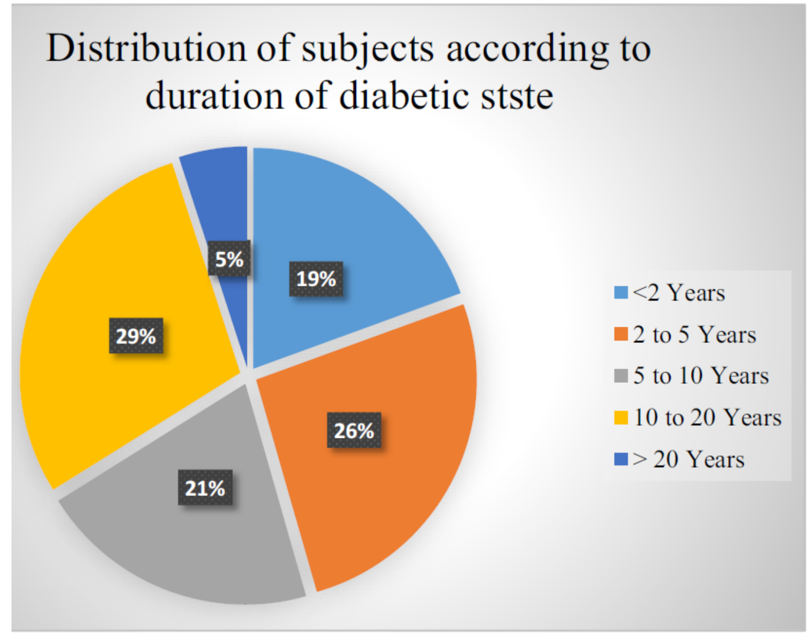
Application: Transforming input data such as text for use with machine learning algorithms.
Modules: preprocessing, feature extraction. — Examples

Mert Çağlar

Discussion

Subject Selection and Data Collection

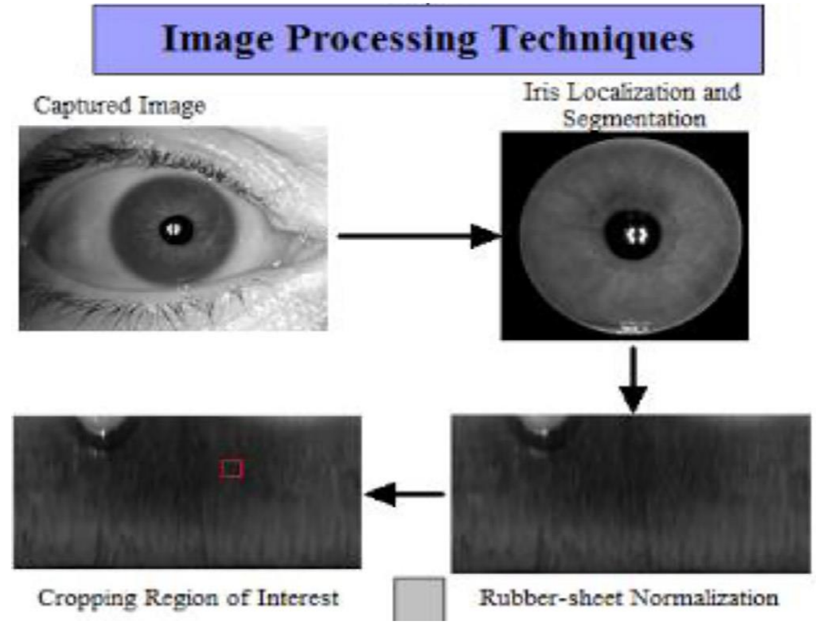
- Non-contact IR imaging
- Varied diabetic states of 1-25 years

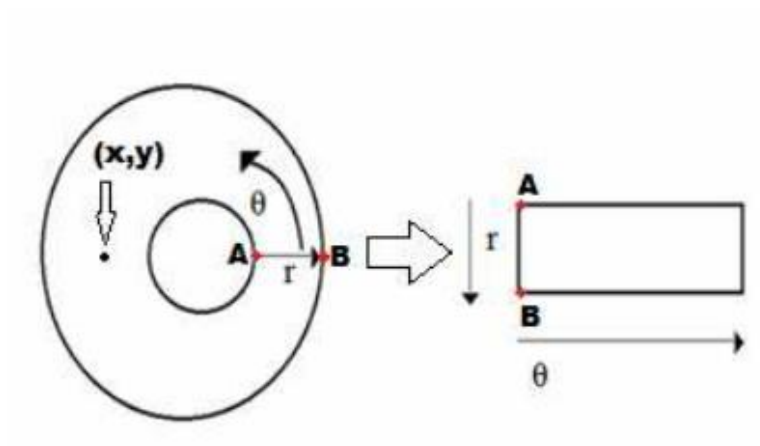


	No. of male	No. of female	Gender ratio	Average age	Standard deviation	Total
Diabetic subjects	102	78	1.307	53.32	8.56	180
Healthy subjects	91	67	1.35	52.86	9.98	158

Discussion

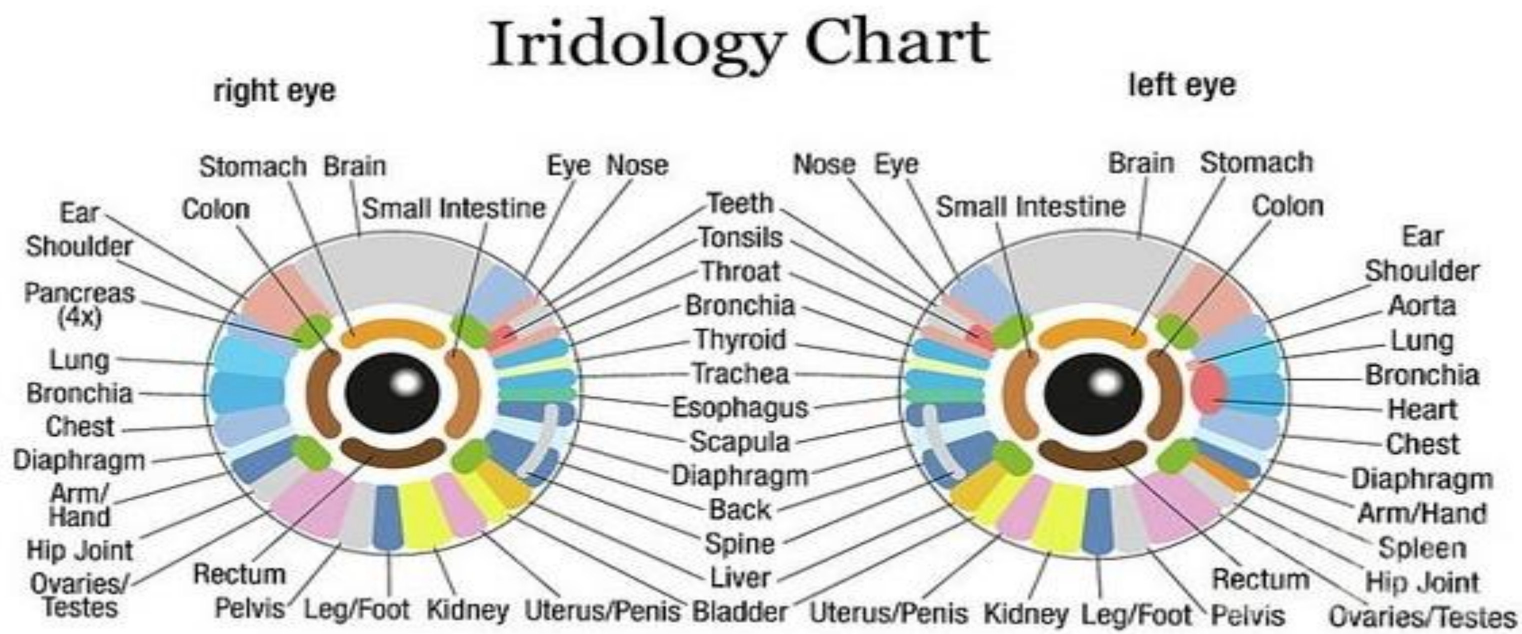
- Input: Gray IR images 640 ×480 (VGA) of each iris
- Iris segmentation (Circular Hough transform)
- Rubber-sheet normalization
- Output: Homogeneous 2D array of 360 ×720





Discussion

- The Region of Interest (ROI) cropped from the rubber sheet normalized iris.
- Green areas in the chart indicate pancreas (source of Insulin enzyme which balances blood sugar)



Mert Çağlar

Discussion

Textural features	Formulas
Contrast	$\sum_{i,j} \lvert i-j \rvert^2 P(i,j)$
Correlation	$\sum_{i,j} (i-\mu_i)(j-\mu_j)P(i,j)$
Energy	$\sum_{i,j} \mathrm{P}(\mathrm{i},\ \mathrm{j})^2$
Homogeneity	$\sum_{i,j} \frac{\mathrm{P(i,j)}}{1+\lvert \mathrm{i-j} \rvert}$

Feature Extraction

Statistical Features

Texture Based Features

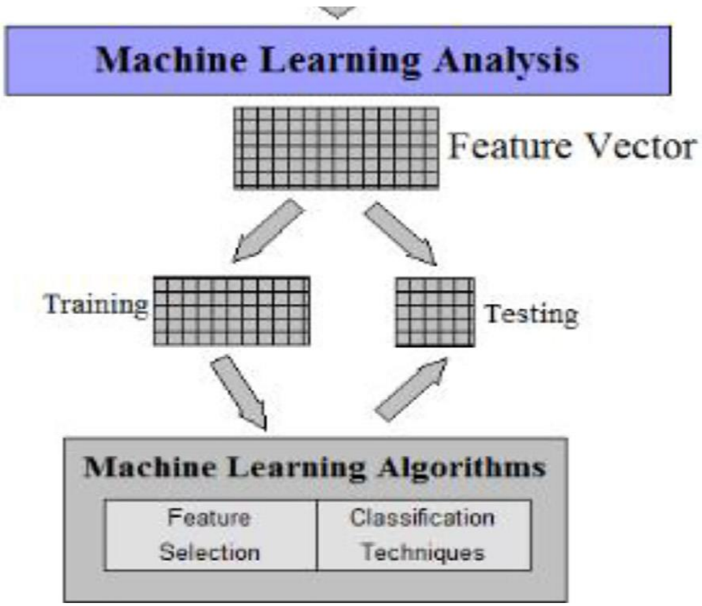
Wavelet Based Features

Discrete Wavelet Transforms (DWT)

Statistical features	Formulas
Mean intensity	$\frac{1}{N}\sum_{i=1}^{N} X(i)$
Standard deviation	$\left(\frac{1}{(N-1)}\sum_{i=1}^{N}\left(X(i)-\bar{X}\right)^2\right)^{1/2}$
Entropy	$\sum_{i=1}^{N_1} P(i).\log_2 P(i)$
Histogram intensity features	Five histogram intensity levels

Mert Çağlar

Results



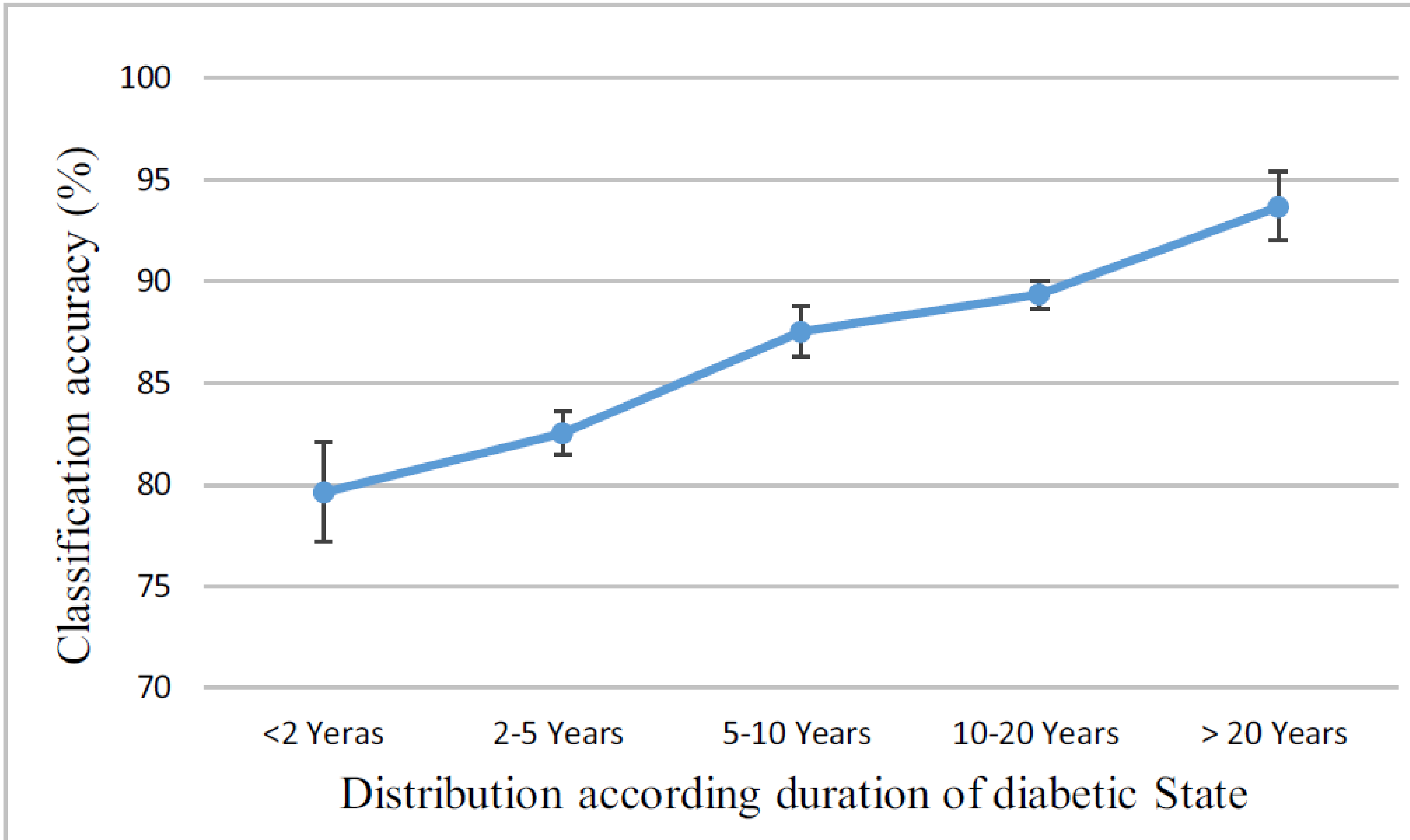
- Binary Tree Model (BT),
- Support Vector Machine (SVM)
- Adaptive Boosting Model (AB)
- Generalized Linear Models (GL)
- Neural Network Model (NN)
- Random Forest (RF)

Classifiers have been trained using the repeated 10 fold cross validation technique.

Feature selection method	Formula for relevance index or scoring criteria	Specification
Fisher score	$J_{fisher}(X_k) = \frac{\sum_{m=1}^{2} n_m(\mu_{k,m}-\mu_k)^2}{\sum_{m=1}^{2} n_m\sigma_{k,m}^2}$	X_k=feature to be evaluated, μ_k=overall mean of feature to be evaluated, m=no of samples in mth class, $\mu_{k,m}$=mean of (X_k) on mth class and $\sigma_{k,m}^2$=variance of (X_k) on mth class.
t-test	$J_{ttest}(X_k) = \frac{\mu_1-\mu_2}{\sqrt{\frac{\sigma_1^2}{n_1}-\frac{\sigma_2^2}{n_2}}}$	μ_1 and μ_2=means of two classes $\sigma_1$and σ_2=variance of two classes
Chi-square test	$J_{chi-square}(X_k)=\sum_{i=1}^{r}\sum_{m=1}^{2}\frac{(n_{im}-\mu_{im})^2}{\mu_{im}}$ where, $\mu_{im} = \frac{n_{*m}n_{i*}}{N}$	n_{im}=no of samples with ith feature value in mth class, n_{i*}=no of samples with ith feature value n_{*m} =no of samples in class m and N no of samples

Mert Çağlar

Results



Mert Çağlar

Questions & Conclusion

- Diabetes is a nasty and hard to diagnose disease
- Digital image processing and machine learning algorithms can create complementary-alternative medical possibilities
- Introduce similar methods to commonly available smartphones/selfie industry
- Identify other diseases from iridology chart and DSP ML methods
- Improve the results with worldwide data collection
- What do you think about the topic?