# Diagnostic Decision Support System in Genetic Diseases: The FaceGP DDSS

**Kaya KURU[a], Yusuf TUNCA[b]**

[a] *IT Department, Gulhane Military Medical Academy (GATA), Turkey*
[b] **Department of Medical Genetics**, *Gulhane Military Medical Academy (GATA), Turkey*

**kkuru@gata.edu.tr**

## Abstract

*A face develops under the influence of many genes. Thus, facial appearance can be a significant clue in the initial identification of genetic anomalies associated with especially cognitive impairments. It may be possible to diagnose a good number of syndromes correctly by using computer-assisted face analysis. For dysmorphic syndromes –with - known genetic causes, cyto- and/or molecular genetic analysis is the appropriate route of investigation in order to confirm a diagnosis. In this study, in terms of helping non experienced practitioners in diagnosing process as well as supporting experts in their decisions, we established a methodology to ease the process and we refer to our method as FaceGP DDSS. In the methodology, digital facial image processing methods are used to reveal facial features with disorders indicating genotype-phenotype interrelation. A great number of genetic disorders indicating a characteristic pattern of facial anomalies can be typically identified by analyzing specific features with the aid of facial image processing methods such as PCA in order to determine reference values (eigenfaces) and train the system. Distance algorithms such as Euclidean, Mahalanobis are used to construct the correlation of the input image with the trained images in matching. Some image enhancement methods such as histogram equalization and median filter are implemented on detected degraded images to capture better features. This study proposes a novel computer-assisted and cost-effective method by merging several methods in the characterization of the facial dysmorphology, in particular a method relying primary on face image capture and manipulation to diagnose genetic diseases.*

## Introduction

Dysmorphology is the aspect of clinical genetics concerned with syndrome diagnosis in patients who have a combination of congenital malformations and unusual features, especially extraordinary faces, often with delayed motor and cognitive development. A face develops under the influence of many genes and in many cases; a face provides important information to diagnose a syndrome. Thus, facial appearance can be a significant clue in the initial identification of genetic anomalies associated with cognitive impairments. Making a diagnosis for a dysmorphic patient requires a high degree of experience and expertise since there are thousands of possible genetic conditions to be taken into consideration and each condition is in itself very rare. A specialists's evaluation generally seems the best path to ensure a proper genetic diagnosis is reached even though it is sometimes very difficult even for them. In some areas of the world, however, genetic diagnosis is generally performed by general practitioners, neurologists or pediatricians not particularly trained in dysmorphology, rather than trained geneticists due to the lack in the numbers, referring the images or terms standardized and specified in some limited number of catalogs and databases. This can lead to diagnostic inaccuracy which in turn curtails both the right cure of patients that will best suit their particular needs and access to critical services such as clinical trials and a patient's referral to supportive services including early intervention, physical or occupational therapy. Thus, delays in early treatment can have a significant impact on the child's well-being and can dramatically influence the chances of the child catching up to his/her peers and leading to a normal life. Therefore, reaching a thorough genetic diagnosis at an early stage is crucial. For dysmorphic syndromes with known genetic causes, molecular and/or cytogenetic analysis is the appropriate route of investigation in order to confirm a diagnosis. However, applying right analysis method throughout many probable analyses is very much dependent to the accurate diagnosis considered before genotyping is undertaken. Most physicians, neurologists or pediatricians are capable of noticing these early signs in a child, but are not equipped to perform a precise genetic diagnosis on their own especially for very rare diseases. A computer-assisted DSS specific to dysmormology may be very supportive for general practitioners as well as dysmorphologists to their decisions. In this study, in terms of both the feature extraction and helping non experienced practitioners in diagnosis process as well as to support experts in their decisions, we established an application to ease the process and we refer to our method as "Facial Genotype-Phenotype Diagnostic Decision Support System (FaceGP DDSS) in Dysmorphology".

## Methods

Our approach that we refer to as **FaceGP DDSS** to classifying a face is to compute the average face shape of groups of patients with different genetic conditions and then determine which average faces a patient's face is most similar to. The **FaceGP DDSS** methodology has been established in C programming language and all modalities and user interface are depicted in Figure 1 and Figure 2 respectively. We benefited OpenCV **(**Open Source **C**omputer **V**ision) library very much. The **FaceGP DDSS** methodology is designed for modeling and analyzing large sets of face images. We haven't use craniofacial landmarks or extracting feature vectors which have very limited number of data to describe a face and would be very limited to reveal the features of specific dysmorphic diseases. Instead; PCA which evaluates every part of the face is employed to extract the features of the faces. The most relevant information in the approach of the principal component analysis method that best describes a face is derived from the entire face image. Any particular face could be economically represented in terms of a best coordinate system named "eigenfaces". Thus a face can be represented by an ordered sequence of 50 or so numbers with sophisticated mathematical representations. This is a huge data compaction, reducing the representation of a face surface from as many as 75 000 parameters (20 000 2D points each with x and y coordinates) down to 50. The surface of each face can be reconstructed using a linear weighted sum of the PCA modes in return.

## Figures

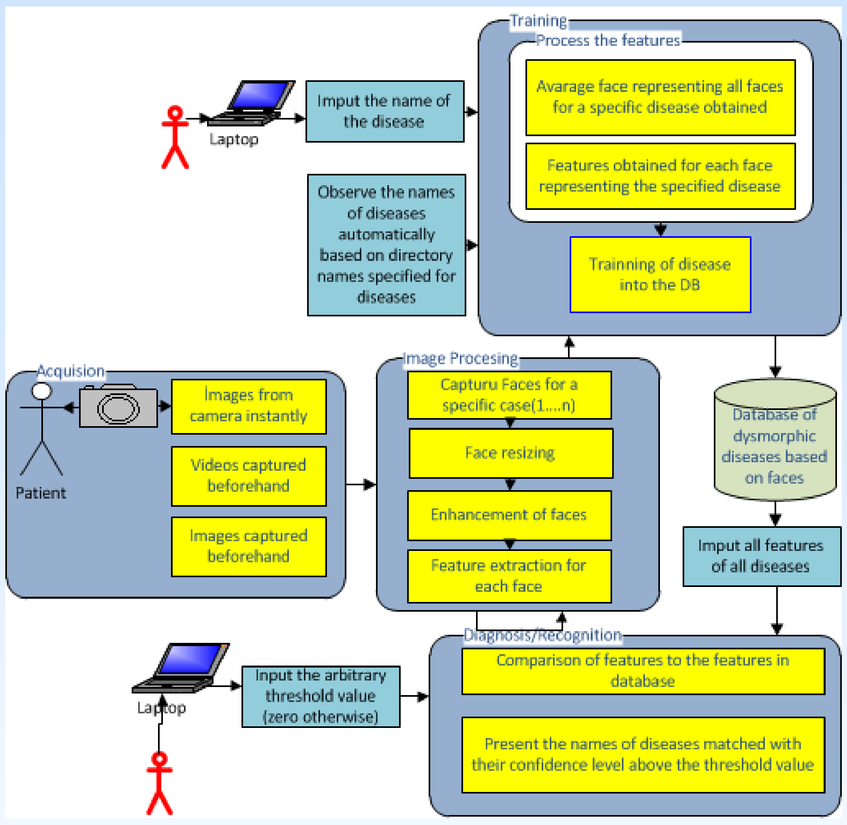


**Fig. 1. The modalities of acquisition, image processing, training and diagnosing/recognition:** An acquisition module can request face images from several different environments: The face image can be an digital image at the acquisition phase. Each image is resampled with bilinear interpolation to the same size and histogram equalization is performed at the image processing phase.
acquisition phase. *By image processing module*, each image is resampled with bilinear interpolation to the same size, resized according to a configurable dimension value and histogram equalization is performed at the image processing phase. The PCA procedure can then be run to generate the eigenfaces, eigenvalues, and average image. *In terms of training module, . During diagnosis/recognition module,* the dysmorphic face database is used by the face checking procedure each time a new dysmorphic face needs to be checked against the known dysmorphic diseases in terms of faces. The distance from the new face to each of the faces in the database is examined and the shortest distances above the threshold value are recorded. If the distances are smaller than a preset threshold set by the user, the corresponding face images in the database with this coefficient set are found. The image and the distance to it are both returned. The threshold value, the minimum distance for a face to be considered a match, is set to zero as a default value.
These steps are training of specific phenotypic diseases by using either on a cam or from sample images, storing the trained diseases in numbers for further analysis, identifying or diagnosing process either from a cam or from a sample image.

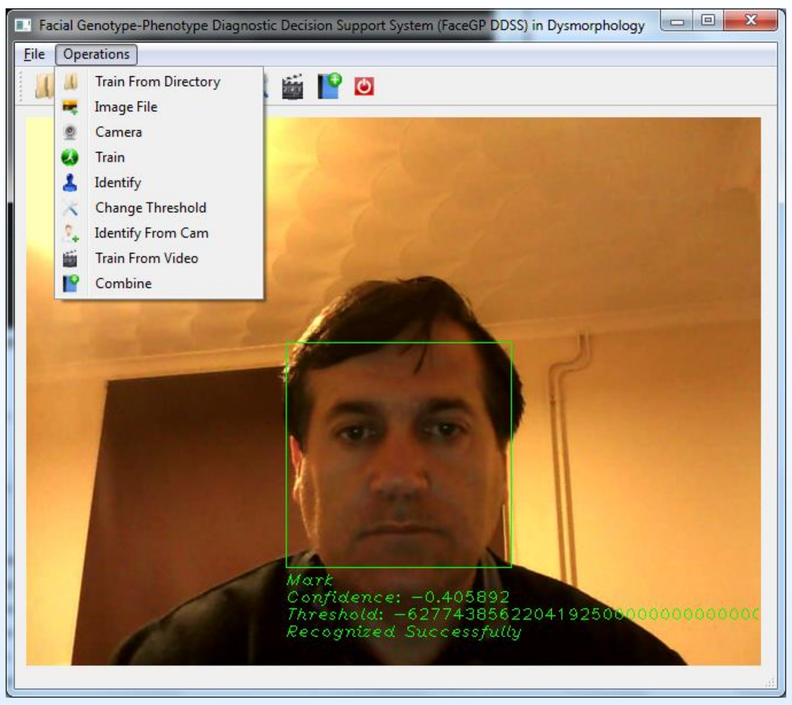


**Fig. 2.** Users could benefit the methodology with a user friendly interface without any manual intervention which may cause the users to avoid the use of any system.
Confidence values are defined as the resemblance or degree of proximity of eigenface values (how near they are?) between two sets of eigenface values obtained from the values of the trained diagnostic images and the identified image to be diagnosed. These values are used for assessing the reliability of the proposed diagnostic inference by the system.

Users can easily add new dysmorphic diseases by using their archives in which there are several sample images representing other diseases not defined in the system.

## Methods (Continue)

In the methodology the PCA procedure can be run to generate the eigenfaces, eigenvalues, and average image. The number of generated eigenfaces will be equal to the number of input images. The eigenface data is generated and is stored in XML file. We have tested the Mahalanobis distance and Euclidean distance in our study to find the genotype-phenotype correlations among 7 typical distance algorithms, namely city block, euclidean, weighted euclidean, sub/space method, multiple similarity method, bayes decision method and Mahalanobis. Euclidean distance matching process outperforms the process of Mahalanobis distance in finding the similar faces with a very limited study sample consisting of a total of 10 subjects. **The modalities of FaceGP DDSS** are **acquisition, image processing, training and diagnosing/recognition.** Images can be captured either from a camera instantly or from previously recorded digital images or videos at the acquisition phase. Each image is resampled with bilinear interpolation to the same size and resized according to a configurable dimension value, and PCA is then run to generate the eigenfaces, eigenvalues, and average image following histogram equalization and median filtering performed at the image processing phase to standardize cropped face images to remove illumination problem. Within training module, a sole average face representing all eigenfaces in terms of processed faces for a specific disease is created and written into a directory together with all processed eigenfaces obtained from eigenvalues. All these values are stored in an XML format with the referenced disease name provided by the user as well as all procedures may be as well triggered by the user to store more than one kind of disease which are stored in a directory with their directory names as being supposedly disease names. *During diagnosis/recognition module,* the dysmorphic face database is used by the face checking procedure each time a new dysmorphic face needs to be checked against the known dysmorphic diseases in terms of faces. The distance from the new face to each of the faces in the database is examined and the shortest distances in terms of the the threshold value are recorded. If the distances are smaller than a preset threshold set by the user, the corresponding face images in the database with this coefficient set are found. The image and the distance to it are both returned. The threshold value, the minimum distance for a face to be considered a match, is set to zero as a default value. Diagnosis is where the system compares the given patient's face features to all the other trained face features in the database and gives a ranked list of possible matches with respect to the confidence value above the threshold value supplied by the user.

## Results
## Discussion and Conclusion

Most DDSS studies on dysmorphology to date have focused on a limited number of specific diseases on an experimental level. This study aims to highlight that a computer can be very helpful in analysing dysmorphic disiases and can help and support experts for their decisions to diagnose through thousands dysmorphic diseases. A limitless number of dysmorphic diseases can be trained and tackled with by FaceGP DDSS methodology in diagnosing process simultaneously which is an handicap for other presented systems having a limited number of diseases, up to 10 in number. The **FaceGP DDSS** methodology is designed to aid investigators who are interested in expanding their study to include their findings that are even outside of their primary area of expertise. For example, a practitioner in a province who is not an expert on genetics or dysmorphology may catch a unique dysmorphologic disease and could present it to the knowledge of the scientific community. This will make it possible to easily combine data from multiple resources. Combining studies generates increased statistical power and the ability to detect both more subtle and more complex—and, perhaps, unexpected—genotype-phenotype dysmorphic associations. There are compelling reasons to promote these kinds of studies. The **FaceGP DDSS methodology** is designed primarily for investigators who wish to diagnose their patients with dysmorphic diseases quickly, effectively and successfully. Furthermore, it aims to support scientists for their studies who do not have expertise in the particular domain of dysmorphology. We need to extend this work to a wider environment by including domain experts from academic and government institutions. Based on user feedback, we expect to continue to update the functionality of the methodology. We plan to establish a process for updating content of dysmorphology. More specifically, the **FaceGP DDSS methodology** will make it easy for researchers to effectively expand the content of the current study to include their findings. The hope is that the **FaceGP DDSS methodology** will be widely adopted by the scientific community, fostering a new era of cooperation and collaboration and facilitating cross-study. More dysmorphological face images are needed to prove and improve the usefulnes of the methodology. Please provide these images to us by including you into the study.

**REFERENCES**

[1] Kuru K, Tunca Y, Niranjan M. Establishment of diagnostic decision support system (DDSS) in clinical diagnosis of genetic diseases: the facegp DDSS methodology and its applications. European Journal of Human Genetics. 2012 Jun 1;20(1):70..
[2] Kuru K, Tunca Y. Diagnostic Decision Support System in Dysmorphology. In Decision Support Systems 2012 Oct 17. IntechOpen.
[3] Kuru K, Girgin S, Arda K, Bozlar U, Akgün V. Developing diagnostic dsss based on a novel data collection methodology. In Knowledge Science, Engineering and Management: Third International Conference, KSEM 2009, Vienna, Austria, November 25-27, 2009. Proceedings 3 2009 (pp. 110-121). Springer Berlin Heidelberg.
[4] K. Kuru, M. Niranjan and Y. Tunca, "Establishment of a Diagnostic Decision Support System in Genetic Dysmorphology," 2012 11th International Conference on Machine Learning and Applications, Boca Raton, FL, USA, 2012, pp. 164-169, doi: 10.1109/ICMLA.2012.234.
[5] Kuru K, Girgin S, Arda K, Bozlar U. A novel report generation approach for medical applications: the SISDS methodology and its applications. International journal of medical informatics.