European Journal of Endocrinology, 2023, **188**, 1–8
https://doi.org/10.1093/ejendo/lvad005
Advance access publication 13 January 2023
Original Research

Real-time detection of acromegaly from facial images with artificial intelligence

Muhammed Kizilgul,[1,*] Rukiye Karakis,[2] Nurettin Dogan,[3] Hayri Bostan,[1] Muhammed Mutlu Yapici,[4] Umran Gul,[1] Bekir Ucan,[1] Elvan Duman,[5] Hakan Duger,[1] Erman Cakal,[1] and Omer Akin[6]

[1]University of Health Sciences, Diskapi Yildirim Beyazit Training and Research Hospital, Department of Endocrinology, Ankara, Turkey
[2]Sivas Cumhuriyet University, Faculty of Technology, Software Engineering Department, Sivas, Turkey
[3]Selçuk University, Faculty of Technology, Computer Engineering Department, Konya, Turkey
[4]Ankara University, Elmadağ Vocational School, Computer Technologies Department, Ankara, Turkey
[5]Burdur Mehmet Akif Ersoy University, Faculty of Technology, Software Engineering Department, Burdur, Turkey
[6]TOBB ETU, Faculty of Science and Literature, Mathematics Department, Ankara, Turkey

***Corresponding author:** Ziraat Mah. Şehit Ömer Halisdemir Cad. No: 20, Diskapi Yildirim Beyazit Training and Research Hospital, 06130, Ankara, Turkey. Email: muhammed.kizilgul@saglik.gov.tr, muhammedkzgl@gmail.com

Abstract

Objective: Despite improvements in diagnostic methods, acromegaly is still a late-diagnosed disease. In this study, it was aimed to automatically recognize acromegaly disease from facial images by using deep learning methods and to facilitate the detection of the disease.

Design: Cross-sectional, single-centre study

Methods: The study included 77 acromegaly (52.56 ± 11.74, 34 males/43 females) patients and 71 healthy controls (48.47 ± 8.91, 39 males/32 females), considering gender and age compatibility. At the time of the photography, 56/77 (73%) of the acromegaly patients were in remission. Normalized images were obtained by scaling, aligning, and cropping video frames. Three architectures named ResNet50, DenseNet121, and InceptionV3 were used for the transfer learning-based convolutional neural network (CNN) model developed to classify face images as "Healthy" or "Acromegaly". Additionally, we trained and integrated these CNN machine learning methods to create an Ensemble Method (EM) for facial detection of acromegaly.

Results: The positive predictive values obtained for acromegaly with the ResNet50, DenseNet121, InceptionV3, and EM were calculated as 0.958, 0.965, 0.962, and 0.997, respectively. The average sensitivity, specificity, precision, and correlation coefficient values calculated for each of the ResNet50, DenseNet121, and InceptionV3 models are quite close. On the other hand, EM outperformed these three CNN architectures and provided the best overall performance in terms of sensitivity, specificity, accuracy, and precision as 0.997, 0.997, 0.997, and 0.998, respectively.

Conclusions: The present study provided evidence that the proposed AcroEnsemble Model might detect acromegaly from facial images with high performance. This highlights that artificial intelligence programs are promising methods for detecting acromegaly in the future.

Keywords: acromegaly, artificial intelligence, deep learning, detection

Significance

In the present study, we aimed to automatically recognize acromegaly disease from facial images by using deep learning methods and to facilitate the diagnosis of the disease. The main contribution of the current paper is summarized in the following: (a) for the first time, transfer learning-based deep learning and deep learning-based machine learning techniques were used from facial images in the detection of acromegaly disease, (b) a high-accuracy transfer learning-based deep learning framework we call AcroEnsemble, in which three deep learning techniques are combined, has been proposed, (c) the proposed framework is benchmarked against deep learning and machine learning methods under various statistical measures, (d) a unique acromegaly dataset has been created for the Turkish nation.

Introduction

Acromegaly is a rare and slowly progressing disease typically caused by overproduction of growth hormone (GH) and, consequently, insulin-like growth factor 1 (IGF-1), usually resulting from a pituitary adenoma. Acromegaly is associated with a variety of cardiovascular, respiratory, endocrine, metabolic, and musculoskeletal comorbidities, as well as increased mortality.[1] A recent study of 3173 patients reports that the delay between the onset of the first symptoms and the diagnosis of acromegaly is ~8–10 years.[2] Despite efforts for earlier recognition of acromegaly and improved diagnostic tests, the delay from onset of symptoms to diagnosis of the disease could not

Received: October 12, 2022. **Revised:** November 21, 2022. **Editorial Decision:** January 5, 2023. **Accepted:** January 5, 2023

have been changed in the last three decades.[3] This diagnostic delay may cause psychosocial deterioration and increase the risk of complications.[1,4] Therefore, there is a need for new techniques that can increase diagnostic awareness and facilitate the detection of acromegaly.

Patients with acromegaly have typical facial changes such as enlargement of the nose, chin, forehead, cheekbones, soft tissue swelling (lip, nose, ear enlargement), and skin thickening. Since the face change happens so slowly, it is often overlooked, especially by the patient himself and those who see the patient frequently.[5] Acromegaly in patients with subtle features can be detected by computer software using facial features on patients' photographs.[6] Machine learning (ML) systems contribute to the earlier recognition of patients with acromegaly.[7]

In this study, we aimed to automatically detect acromegaly disease from facial images. For this purpose, the effectiveness of acromegaly recognition using deep learning (DL) methods and ML approaches was examined. On the basis of experimental results, we proposed a DL model called AcroEnsemble.

Materials and methods

Patients and controls

The study was conducted in the Endocrinology and Metabolism Department of Health Science University, Diskapi Yildirim Beyazit Training and Research Hospital. We included 77 acromegaly (52.56 ± 11.74, 34 males/43 females) patients who were followed up in our clinic and 71 healthy individuals (48.47 ± 8.91, 39 males/32 females) as the control group, considering gender and age compatibility ($P > .05$). An unequivocally increased serum IGF-1 concentration with typical clinical manifestations, nonsuppressed GH in an oral glucose tolerance test and the presence of a pituitary tumor were used to diagnose acromegaly.[8] Disease control was described as the patient having normal IGF-1 levels according to age and sex.

The present study was conducted ethically in accordance with the World Medical Association Declaration of Helsinki. Approval for the current study was granted by the Ethics Committee of the University of Health Sciences, Diskapi Yildirim Beyazit Training and Research Hospital (approval date: 20.09.2021/no: 120/07). Informed consent was obtained from all participants in the study.

Biochemical evaluations

Serum GH and IGF-1 levels were measured by IMMULITE 2000 Immunoassay System (Siemens Healthcare GmbH, Erlangen, Germany). The method for GH measurement was based on a solid-phase, two-site chemiluminescent immunometric assay (Siemens Healthcare Diagnostics Products—Glyn Rhonwy Llanberis, Gwynedd LL55 4EL, United Kingdom). GH measurement was standardized using the recombinant WHO standard second IS 98/574. The method for IGF-1 assay was a solid-phase, enzyme-labeled chemiluminescent immunometric assay (Siemens Healthcare Diagnostics Products—Glyn Rhonwy Llanberis, Gwynedd LL55 4EL, United Kingdom). Serum IGF-1 levels were compared with the age-sex adjusted normal reference values obtained from the manufacturer's instructions for use.

Learning models

DL is the process of extracting a hierarchy of features from raw input images using neural networks that have many layers. The most distinguishing characteristic of DL methods is that they obtain the complex hierarchy of image features by using their self-learning ability rather than handcrafted features such as face landmarks. It means that deep networks are able to create multiple levels of abstraction to represent a data. Thus, they can be used in solving heuristic problems that do not have a specific formula or how they are solved by humans cannot be defined exactly. Contrary to traditional ML methods, the performance of DL methods does not remain constant after a certain point but increases depending on the size of the data. However, working with large datasets in many problem areas, such as data on rare diseases in healthcare, is a significant challenge. Transfer learning, which is inspired by the idea of leveraging prior knowledge to solve problems encountered for the first time, can help in some situations where data is limited. The knowledge obtained from huge datasets is transferred to solve specific problems such as acromegaly detection. The weight transfer process can be defined as the transfer of expert knowledge of a DL network specialized in a different classification problem to a new DL network. This allows the DL network to find the optimum weights and bias values of the classifier to solve the problem more easily.

The recognition of acromegaly has been addressed in this study using both transfer learning-based DL models such as ResNet50,[9] DenseNet121,[10] and InceptionV3[11] and traditional ML techniques such as k-nearest neighbor (kNN), support vector machine (SVM), decision tree (DT), and random forest (RF). The effectiveness of DL models based on transfer learning and ML approaches was thoroughly compared. Based on experimental studies, a deep convolutional neural network (CNN) framework was proposed with transfer learning for acromegaly recognition. The proposed framework takes advantage of the performances of three pre-trained networks, ResNet50, DenseNet121, and InceptionV3, which we call AcroEnsemble. The general architecture of the proposed method is given in Figure 1.

Face classification

In the initial phase, video recordings of the participants were captured with the different mobile phone cameras at the clinic by the authors. In order to prevent image shifting during recordings, the participants were seated on a chair, and the background was adjusted as a single color. 180-degree images were recorded from the right profile to the left profile of the patients for an average of 20 s. For each participant, an average of 10 images were obtained and saved from the video. The images were composed of recorded images from the right profile to the left profile at an angle of ~20 degrees. Thus, we obtained an average of 10 images consisting of the right profile, left profile, cross profile, and opposite side of each patient. Finally, all images were cropped according to the head border of the individual and resized to 256×256. Aspect ratios were preserved to avoid distortion of images during resizing. Zero padding was applied to the blank areas in the image.

After the normalization process, images were divided into training and test sets at 80% and %20, respectively. A 5-fold cross-validation method was used to achieve more consistent and unbiased performance results. The initial phase is followed by training of the pre-trained CNN models ResNet50,[9] DenseNet121,[10] and InceptionV3.[11] Two fully connected layers containing 1024 and 1024 neurons, respectively, and a softmax layer with two neurons for the "Healthy" or "Acromegaly" classes were added to the classification

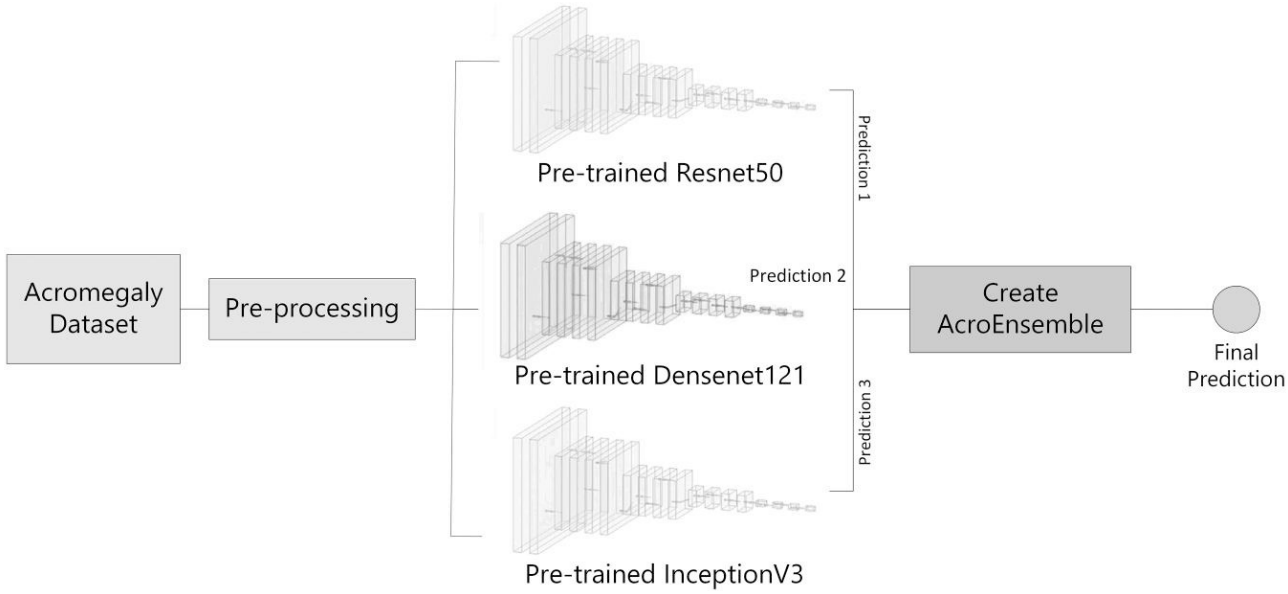

Figure 1. The general architecture of the proposed framework.

layers of these pre-trained models. The weights of feature extraction layers in the pre-trained models were used to initialize the weights of the acromegaly recognition models. The transfer learning architecture of the proposed framework is shown in Figure 2. The experimental setup of the current study was provided in Supplementary.

Statistical analysis

Statistical analysis of baseline characteristics and medical data was performed using SPSS software (version 23.0, SPSS, IBM Corporation, NY, USA). Data distribution was examined with the Kolmogorov-Smirnov test. Normally distributed data were expressed using mean ± standard deviation (SD) values, while non-normally distributed data were expressed using median with interquartile range (IQR) values. Categorical variables were reported as frequencies and percentages (%).

The confusion matrix evaluates ML classification by calculating true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values. In the current study, the acromegaly classification performance of DL models

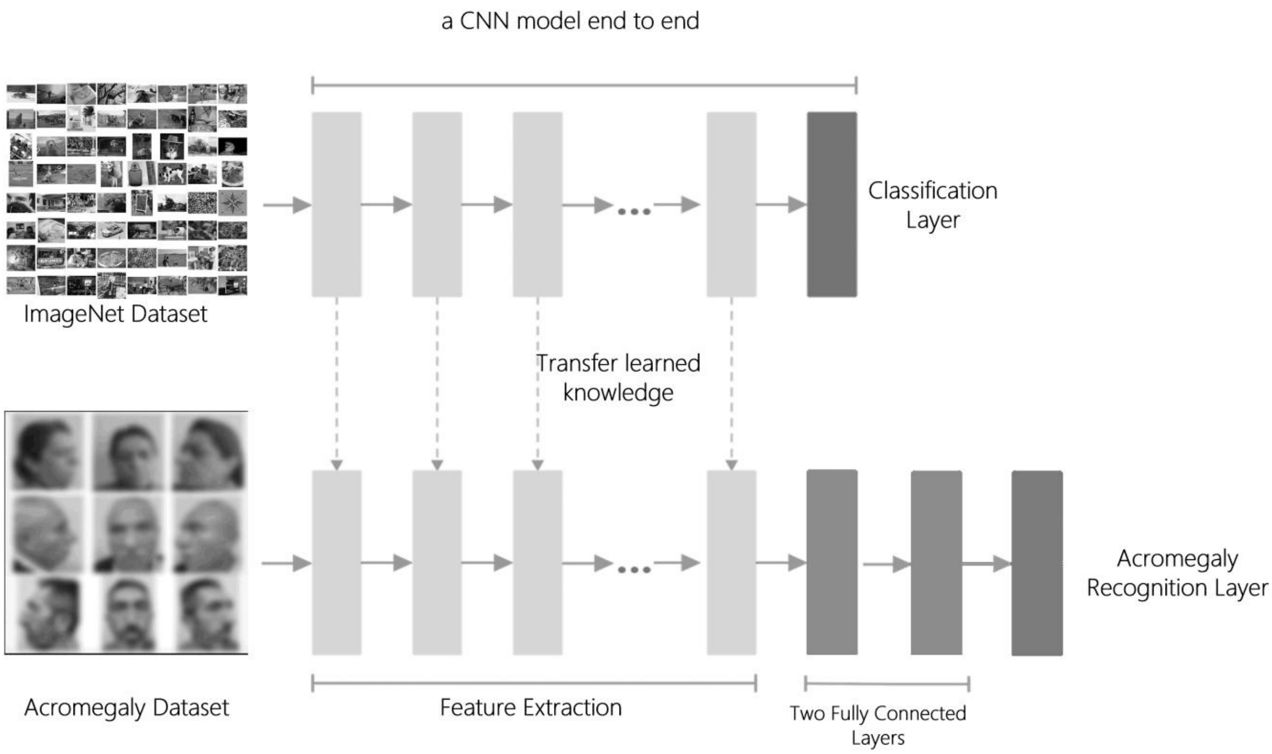

Figure 2. Transfer learning architecture. CNN, convolutional neural network.

was evaluated between the predicted outputs of the networks and the real outputs, which are receiver operating characteristic curve (ROC) statistical metrics such as accuracy $Acc = (TP + TN)/(TP + TN + FP + FN)$), area under the curve (AUC), specificity ($\text{Spec} = TN/(TN + FP)$), precision ($P = TP/(TP + FP)$), recall ($Re = TP/(TP + FN)$), F1-score ($\text{F1} = 2\frac{Precision.Recall}{Precision+Recall}$), and pearson correlation coefficient $\left(R = \frac{\sum(x = \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2}\sqrt{\sum(y - \bar{y})^2}}\right)$. In addition, the performances of DL models were evaluated with negative predictive value (NPV) and positive predictive value (PPV). These performance values were obtained class-based using a one-hot encoding method. In this context, first, the acromegaly and healthy class labels for predicted and real outputs were assigned as 1 and 0, respectively. Second, the ROC performance of the DL methods was obtained by setting the healthy and acromegaly class labels as 1 and 0, respectively. The overall performance metrics of DL methods were found by calculating the macro-average of the values obtained for the acromegaly patients and healthy controls.

Results

Fifty-six (72.7%) of the acromegaly patient group had controlled disease at the time of photography. The median time from symptom onset to diagnosis was 4 (2–6) years, while the time from diagnosis to photography was 70 (33–120) months. Forty (51.9%) of the patients have been followed up with medical treatment after surgery for pituitary adenoma. While the median IGF-1 levels of the patients were 2.58×upper limit of normal (ULN) (1.95–3.33) at the time of diagnosis, these levels decreased to 0.75 xULN (0.59–1.09) at the time of photography ($P < .001$). The characteristics of acromegaly patients are summarized in Table 1.

Table 1. Characteristics of acromegaly patient population.

Variables	Patients
N (%)	77 (100.0)
Age, mean ± SD, years	52.56 ± 11.74
Female gender, *n* (%)	43 (55.8)
Disease control, *n* (%)	
Remission	56 (72.7)
Active	21 (27.3)
Time from symptom onset to diagnosis, median (IQR), years	4.0 (2.0–6.0)
Time from diagnosis to photography, median (IQR), months	70.0 (33.0–120.0)
Treatment modality, *n* (%)	
Surgery alone	31 (40.3)
Surgery plus medical	40 (51.9)
Medical alone	6 (7.8)
Laboratory at the time of diagnosis ($n = 60$)	
Basal GH, median (IQR), mcg/L	8.47 (4.35–17.45)
IGF-1, median (IQR), xULN	2.58 (1.95–3.33)
Laboratory at the time of photography ($n = 77$)	
Basal GH, median (IQR), mcg/L	0.82 (0.39–1.99)
IGF-1, median (IQR), xULN	0.75 (0.59–1.09)

GH, growth hormone, IGF-1, insulin-like growth factor 1; IQR, interquartile range; ULN, upper limit of normal; SD, standard deviation.

The acromegaly dataset was divided into training and test sets at 80% and %20, respectively. The samples were divided into five training and testing subsets using the repeated random subsampling validation (RSV) technique. The test datasets include images of 14 acromegaly (7 males, 7 females) patients and 14 controls (7 males, 7 females). Train datasets contain images of 57 controls (32 males, 25 females) and 63 patients with acromegaly (27 males, 36 females). The performance of the DL and ML models was compared using the same RSV sub data sets.

The average performance results calculated for the DenseNet121, ResNet50, and InceptionV3 DL models to differentiate acromegaly from the healthy condition are given in Table 2. Although the results calculated in its three models are high, the DenseNet121 model achieved the highest performance values on the test data.

The results obtained by the model we call AcroEnsemble, which combines the 3 DL models in the study, are shown in Table 2. When the results were examined, it was revealed that the ensemble process improved the test performance.

In Figure 3, NPV (Figure 3A) and PPV (Figure 3B) values obtained by DenseNet121, ResNet50, InceptionV3, and AcroEnsemble models for five different sub-datasets are shown. When the results are examined, it was seen that each DL model obtains performance values close to each other for the subtest data sets, while the AcroEnsemble model achieves better results.

In addition to the DL models, the performance of the proposed framework was compared with four different ML classifiers: kNN, SVM, DT, and RF. The features obtained from the flatten layer of the DenseNet121 model were used in the input of the ML models in this scenario. The performance results of the ML models on the test data sets are given in Table 3. The best results among ML methods were obtained with the SVM model.

The distribution plots of the NPV (Figure 4A) and PPV (Figure 4B) values obtained by the ML models for the subdatasets are illustrated in Figure 4. Although it was observed that each DL model obtained performance values close to each other for the subtest datasets, the results of the SVM model were comparatively better.

Tables 2 and 3 shows the overall average performance results of the DL and DL-based ML models. Accordingly, the worst results were obtained for the kNN classifier. The worst results were obtained by the kNN classifier with an accuracy of 0.862. Among the ML methods, the performance of the SVM model in distinguishing acromegaly from the healthy state is relatively high. However, the best results were obtained with the AcroEnsemble model, which combined the 3 DL models.

The AcroEnsemble model found the AUC and accuracy values for men and women with acromegaly to be 0.989, 0.989, 1.00, and 1.00, respectively. These performance values for control men and women were calculated as 1.00, 1.00, 0.989, and 0.989, respectively.

Discussion

Acromegaly is characterized by occult onset, a slow progression that eventually leads to comorbidities and shortened life expectancy.[1] The prevalence of acromegaly around the world is 28–137 per million, while the incidence ranges from 2 to 11

Table 2. The average performance results of the transfer-based deep learning models and AcroEnsemble model.

Performance metrics		AUC	Spec[c]	Acc[d]	R[e]	Pre[f]	Recall	F1-score	NPV	PPV
DenseNet-121	A[a]	0.966	0.972	0.965	0.931	0.975	0.959	0.967	0.955	0.975
	H[b]	0.966	0.959	0.965	0.931	0.955	0.972	0.964	0.975	0.955
	Overall	0.966	0.966	0.965	0.931	0.965	0.966	0.965	0.965	0.965
ResNet50	A[a]	0.959	0.966	0.958	0.917	0.968	0.952	0.960	0.948	0.968
	H[b]	0.959	0.952	0.958	0.917	0.948	0.966	0.956	0.968	0.948
	Overall	0.959	0.959	0.958	0.917	0.958	0.959	0.958	0.958	0.958
Inception-V3	A[a]	0.960	0.945	0.961	0.922	0.952	0.975	0.963	0.972	0.952
	H[b]	0.960	0.975	0.961	0.922	0.972	0.945	0.958	0.952	0.972
	Overall	0.960	0.960	0.961	0.922	0.962	0.960	0.961	0.962	0.962
AcroEnsemble	A[a]	0.997	0.995	0.997	0.995	0.995	1.000	0.997	1.000	0.995
	H[b]	0.997	1.000	0.997	0.995	1.000	0.995	0.997	0.995	1.000
	Overall	0.997	0.997	0.997	0.995	0.997	0.997	0.997	0.997	0.997

[a]Acromegaly patients.
[b]Healthy controls.
[c]Specificity.
[d]Accuracy.
[e]Regression.
[f]Precision.
AUC, area under the curve, NPV, negative predictive value, PPV, positive predictive value.

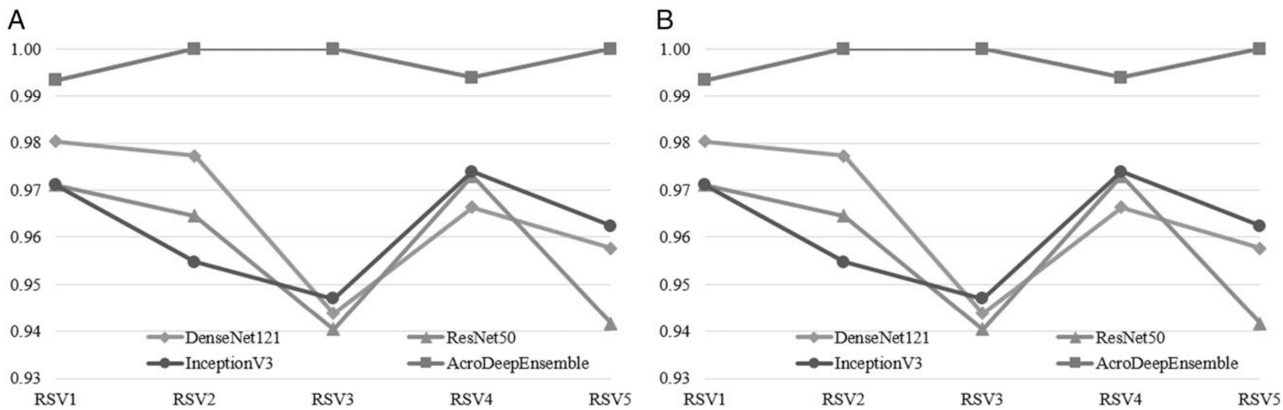

Figure 3. Performance results of the deep learning models for five testing subsets, (A) NPV: negative predictive value, (B) PPV: positive predictive value.

cases per million people per year.[12] The diagnosis of acromegaly is still delayed in most patients, and this delay corresponds to an increase in morbidity and mortality.[13] At the time of diagnosis, approximately 70% of somatotroph adenomas are macroadenomas and the cavernous sinus is usually invaded by the tumor,[14] which reduces the possibility of cure. Acromegaly is clearly underdiagnosed in the general population, and with active screening, many previously undetected cases of acromegaly can be found.[15] The true prevalence of acromegaly is probably higher than currently thought and may reach 4-10 cases per 10,000 of the general population.[16] The variety of disease symptoms leading the patients to refer to clinicians from different specialties (dentists, hand surgeons, ophthalmologists, gynecologists, etc.) and the relatively insidious facial changes may contribute to the delay in diagnosing acromegaly.[7] Therefore, new approaches that can increase awareness, especially in patients and non-endocrinologist clinicians, may prevent further delay in the diagnosis of the

Table 3. Performance results of machine learning classifiers.

ML models	AUC	Spec[a]	Acc[b]	R[c]	Pre[d]	Recall	F1-Score	NPV	PPV
kNN	0.869	0.740	0.862	0.757	0.780	0.997	0.874	0.997	0.780
SVM[e]	**0.992**	**0.991**	**0.991**	**0.983**	**0.990**	**0.992**	**0.991**	**0.992**	**0.990**
DT	0.844	0.845	0.844	0.690	0.834	0.844	0.837	0.858	0.834
RF	0.928	0.920	0.926	0.857	0.916	0.936	0.923	0.942	0.916

[a]Specificity.
[b]Accuracy.
[c]Regression.
[d]Precision.
[e]The best results among ML methods were obtained with the SVM model.
AUC, area under the curve; DT, decision tree; kNN, k-nearest neighbor; ML, machine learning, NPV, negative predictive value; PPV, positive predictive value; RF, random forest; SVM, support vector machine.

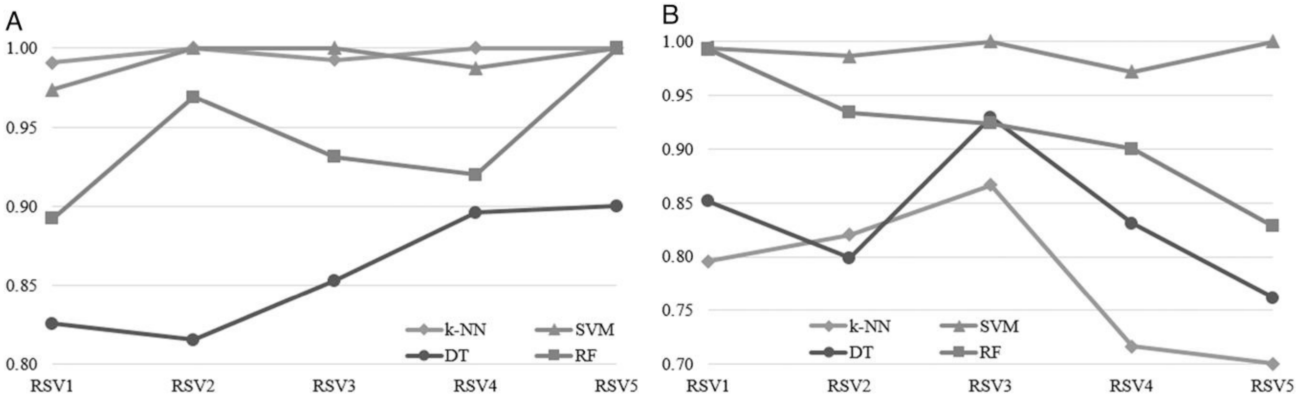

Figure 4. Performance results of the deep learning-based machine learning models for five testing subsets, (A) NPV, negative predictive value, (B) PPV, positive predictive value. DT, decision tree, kNN, k-nearest neighbor, RF, random forest, SVM, support vector machine.

disease. We aimed to use ML systems analyzing facial pictures to screen acromegaly in adults, which achieved up to 98% accuracy.

A new imaging technique, three-dimensional (3D) stereophotography may provide more detailed facial features in patients with acromegaly. A study of 39 acromegalic patients demonstrated that 3D stereophotography is an accurate and reliable tool for analyzing facial characteristics.[17] Meng et al.[18] reported that the addition of ML to 3D imaging could accurately identify and predict facial variables in acromegaly patients. However, these studies were not an acromegaly patient recognition study, and it may not be possible to use it in instant and real-time applications since the image must be taken from 3 different angles in an experimental environment for an adequate assessment. ML has the potential to become a real game-changer in many areas of medicine by opening up new ways to handle complex healthcare problems. ML is a subset of artificial intelligence (AI) that focuses on using algorithms to learn from the data without the need for further programming. CNN, as a typical kind of feed-forward artificial neural network, has been widely applied in video, audio, and image recognition. Face recognition technologies have been used in the diagnosis of diseases. Kruszka et al.[19] reported that a facial recognition technology used for the diagnosis of 22q11.2 deletion syndrome had a sensitivity and specificity of more than 96%. Abnormal facial bone hyperosteogeny and soft tissue hypertrophy chronically deform the faces of acromegaly patients, causing self-confidence loss and psychological trauma while disturbing quality of life.[17] The facial changes seen in acromegaly are due to chronic exposure to increased GH and IGF-1 on the bone, cartilage, and soft tissue.[20] Miller et al. compared the facial photographs of 24 patients with acromegaly with 25 normal subjects by a computer program, which achieved an accuracy of 86%.[21] Schneider et al.[6] analyzed face photographs of 57 acromegaly patients and 60 controls by similar computer software, with an accuracy of 71.9% for patients and 91.5% for controls. Classification accuracy by software is higher than by medical experts or general internists, particularly in patients with mild features of acromegaly.[6] These studies were performed with a small sample size and in a specific population and can not be adaptable to other populations. Using CNN as software is an advantage of our study.

In a recent study, ML methodology was used in the detection of acromegaly, and a sensitivity of 96%, a specificity of 96%, a PPV of 96%, and an NPV of 95% were obtained by the analysis of facial photographs.[7] They used various ML methods (Linear models, kNN, SVM, RF, CNN) to generate a better-integrated one (Ensemble Method) for the first time. However, there are some limitations to this study. They did not include photographs of side views. We had side view photographs and video images. Another limitation of their study was using a specific population. Hence the results could not be extrapolated to other populations. We have shown that similarly successful results can be obtained with the ML-based face recognition method in a different society and with smaller sample size.

Previous studies have found that male acromegaly patients have more remarkable and typical changes, particularly in the nose and lip contours, than females.[6,17,18] Schneider et al.[6] found that by using only frontal views, the rates of correct classification in female and male acromegaly patients were 67.4% and 82.6%, respectively, and when side views were added, the correct classification rates increased to 81.0% and 86.2% in females and males, respectively. Meng et al. showed that the disease may be predicted with high accuracy in both genders by performing linear discriminant analysis, which is a multivariate classification technique from ML.[18] Also, we proposed a unique AcroEnsemble model in the current study and characterized acromegaly features with very high accuracy in both genders. These results highlight that CNN-based software can create separate fascial data systems for male and female genders and identify the disease with high accuracy in both genders.

Although ML methods have been used in other studies before, the methods used in these studies differ from each other. Since the face datasets used in other studies are not shared as open source, it is not possible to access those datasets, so we created our dataset. In our study, in order to avoid the complex feature selection process of ML and increase its success, the feature extraction capability of DL models was utilized. Although it has been observed that the success of ML methods has been improved, DL methods with transfer learning have obtained the best results.

The main contribution of the current paper is summarized in the following: a) for the first time, transfer learning-based DL and DL-based ML techniques were used from facial images in the detection of acromegaly disease, b) a transfer learning-based DL framework with high accuracy has been proposed, c) the proposed framework is benchmarked against DL and ML methods under various statistical measures, d) a unique acromegaly dataset has been created for the Turkish nation.

There are several limitations of our study. First, the current study has a relatively small sample size. Second, in the vast majority of patients, current photos and videos were used instead of photos and videos at the time of diagnosis. In addition, the median time from symptom onset to diagnosis and from diagnosis to photography in the study population was 4 years and 70 months, respectively, indicating that the patients included in the study were cases with substantial acromegalic facial features rather than mild cases. However, considering that three-quarters of the patients in the study had controlled disease, it should be taken into account that there may have been also some improvements in the facial features of the patients over time, compared to the time of diagnosis.[22] Third, the performance of AI methods was not evaluated by comparing it with clinician performance or a questionnaire method. On the other hand, there are some important strengths of our study. In previous studies, the working time of the methods has not been discussed and given. Our work can handle images in real-time. Hence, it can generate tags for more than 30 photos per second. Since it does not involve feature extraction, it is not affected by the success of the feature extraction method. With transfer learning, the results of intensive training processes were utilized. With the video data set, not only one or several angles, but also an average of 10 image frames from 180-degree angles were studied.

Acromegaly facial recognition technologies can be used in clinic settings. Individuals who voluntarily accept automatic screening and the system gives a positive result can be referred to health professionals for IGF-1 testing and, if necessary, pituitary MRI. In this way, early diagnosis can be achieved in underdiagnosed diseases such as acromegaly. This way may provide hope for earlier diagnosis which is the one of the most significant problems in the management of acromegaly.

In conclusion, it is clear that there is a need for improved screening strategies for the diagnosis of acromegaly. Our data demonstrated that ML methods are highly effective in acromegaly face detection. The future development of smartphone-based applications may guide the clinician in the selection of patients for further testing and specialist referral. Additionally, using this application by people can direct possible acromegaly patients to the hospital, which will pave the way for the diagnosis of the disease. However, privacy protection, which is a significant obstacle to AI applications, with ethical design thinking should always be kept in mind. A series of actions and rules designed to protect privacy are deserved to be made.

Funding

This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

Author contribution statement

M.K., N.D. and O.A. designed and conceptualized the study. M.K., R.K., H.B., M.M.Y., U.G., E.D., and H.D. provided data. R.K., N.D., M.M.Y. and E.D. performed and studied deep learning and machine learning methods. M.K., R.K., E.D. and H.B. wrote the first draft of the manuscript. All authors provided intellectual input and read, revised and approved the final version of the manuscript.

Conflicts of interest: None declared.

References

1. Colao A, Grasso LFS, Giustina A, *et al.* Acromegaly. *Nat Rev Dis Primers*. 2019;5(1):20. https://doi.org/10.1038/s41572-019-0071-6
2. Petrossians P, Daly AF, Natchev E, *et al.* Acromegaly at diagnosis in 3173 patients from the liège acromegaly survey (LAS) database. *Endocr Relat Cancer*. 2017;24(10):505-518. https://doi.org/10.1530/ERC-17-0253
3. Reid TJ, Post KD, Bruce JN, Nabi Kanibir M, Reyes-Vidal CM, Freda PU. Features at diagnosis of 324 patients with acromegaly did not change from 1981 to 2006: acromegaly remains under-recognized and under-diagnosed. *Clin Endocrinol (Oxf)*. 2010;72(2):203-208. https://doi.org/10.1111/j.1365-2265.2009.03626.x
4. Siegel S, Streetz-van der Werf C, Schott JS, Nolte K, Karges W, Kreitschmann-Andermahr I. Diagnostic delay is associated with psychosocial impairment in acromegaly. *Pituitary*. 2013;16(4):507-514. https://doi.org/10.1007/s11102-012-0447-z
5. Vilar L, Vilar CF, Lyra R, Lyra R, Naves LA. Acromegaly: clinical features at diagnosis. *Pituitary*. 2017;20(1):22-32. https://doi.org/10.1007/s11102-016-0772-8
6. Schneider HJ, Kosilek RP, Günther M, *et al.* A novel approach to the detection of acromegaly: accuracy of diagnosis by automatic face classification. *J Clin Endocrinol Metab*. 2011;96(7):2074-2080. https://doi.org/10.1210/jc.2011-0237
7. Kong X, Gong S, Su L, Howard N, Kong Y. Automatic detection of acromegaly from facial photographs using machine learning methods. *eBioMedicine*. 2018;27:94-102. https://doi.org/10.1016/j.ebiom.2017.12.015
8. Kizilgul M, Duger H, Nasiroglu NI, *et al.* Efficacy of cabergoline add-on therapy in patients with acromegaly resistance to somatostatin analogs treatment and the review of literature. *Arch Endocrinol Metab*. 2022;66(3):278-285. https://doi.org/10.20945/2359-3997000000481
9. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2016:770-778. https://doi.org/10.1109/CVPR.2016.90
10. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In IEEE Conference on Computer Vision and Pattern Recognition. 2017:2261-2269. https://doi.org/10.1109/CVPR.2017.243
11. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In IEEE Conference on Computer Vision and Pattern Recognition. 2016:2818-2826. https://doi.org/10.1109/CVPR.2016.308
12. Lavrentaki A, Paluzzi A, Wass JA, Karavitaki N. Epidemiology of acromegaly: review of population studies. *Pituitary*. 2017;20(1):4-9. https://doi.org/10.1007/s11102-016-0754-x
13. Esposito D, Ragnarsson O, Johannsson G, Olsson DS. Prolonged diagnostic delay in acromegaly is associated with increased morbidity and mortality. *Eur J Endocrinol*. 2020;182(6):523-531. https://doi.org/10.1530/EJE-20-0019
14. Ershadinia N, Tritos NA. Diagnosis and treatment of acromegaly: an update. *Mayo Clin Proc*. 2022;97(2):333-346. https://doi.org/10.1016/j.mayocp.2021.11.007
15. Schneider HJ, Sievers C, Saller B, Wittchen HU, Stalla GK. High prevalence of biochemical acromegaly in primary care patients with elevated IGF-1 levels. *Clin Endocrinol (Oxf)*. 2008;69(3):432-435. https://doi.org/10.1111/j.1365-2265.2008.03221.x
16. Ribeiro-Oliveira A Jr, Barkan A. The changing face of acromegaly—advances in diagnosis and treatment. *Nat Rev Endocrinol*. 2012;8(10):605-611. https://doi.org/10.1038/nrendo.2012.101

17. Guo X, Meng T, Huang J, *et al.* 3D Facial analysis in acromegaly: gender-specific features and clinical correlations. *Front Endocrinol (Lausanne)*. 2018;9:722. https://doi.org/10.3389/fendo.2018.00722
18. Meng T, Guo X, Lian W, *et al.* Identifying facial features and predicting patients of acromegaly using three-dimensional imaging techniques and machine learning. *Front Endocrinol (Lausanne)*. 2020;11:492. https://doi.org/10.3389/fendo.2020.00492
19. Kruszka P, Addissie YA, McGinn DE, *et al.* 22q11.2 Deletion syndrome in diverse populations. *Am J Med Genet*. 2017;173(4):879-888. https://doi.org/10.1002/ajmg.a.38199
20. Wagenmakers MA, Roerink SH, Maal TJ, *et al.* Three-dimensional facial analysis in acromegaly: a novel tool to quantify craniofacial characteristics after long-term remission. *Pituitary*. 2015;18(1):126-134. https://doi.org/10.1007/s11102-014-0565-x
21. Miller RE, Learned-Miller EG, Trainer P, Paisley A, Blanz V. Early diagnosis of acromegaly: computers vs clinicians. *Clin Endocrinol (Oxf)*. 2011;75(2):226-231. https://doi.org/10.1111/j.1365-2265.2011.04020.x
22. Du F, Chen Q, Wang X, *et al.* Long-term facial changes and clinical correlations in patients with treated acromegaly: a cohort study. *Eur J Endocrinol*. 2021;184(2):231-241. https://doi.org/10.1530/EJE-20-0962