Please cite the Published Version

Additional Information: This is an Open Access article which appeared in F1000Research and has undergone an open peer-review process.

Check for updates

RESEARCH ARTICLE

REVISED **Non-invasive health prediction from visually observable features** [version 2; peer review: 1 approved, 1 approved with reservations]

Fan Yi Khong[1], Tee Connie [1], Michael Kah Ong Goh [1], Li Pei Wong[2], Pin Shen Teh[3], Ai Ling Choo[4]

[1]Faculty of Information Science and Technology, Multimedia University, Melaka, Melaka, 75450, Malaysia
[2]School of Computer Sciences, Universiti Sains Malaysia, Penang, Penang, 11800, Malaysia
[3]Department of Operations, Technology, Events and Hospitality Management, Faculty of Business and Law, Manchester Metropolitan University, Manchester, Manchester, M15 6BH, UK
[4]iRadar Sdn. Bhd., Melaka, Melaka, 75450, Malaysia

Abstract
Background: The unprecedented development of Artificial Intelligence has revolutionised the healthcare industry. In the next generation of healthcare systems, self-diagnosis will be pivotal to personalised healthcare services. During the COVID-19 pandemic, new screening and diagnostic approaches like mobile health are well-positioned to reduce disease spread and overcome geographical barriers. This paper presents a non-invasive screening approach to predict the health of a person from visually observable features using machine learning techniques. Images like face and skin surface of the patients are acquired using camera or mobile devices and analysed to derive clinical reasoning and prediction of the person's health.
Methods: In specific, a two-level classification approach is presented. The proposed hierarchical model chooses a class by training a binary classifier at the node of the hierarchy. Prediction is then made using a set of class-specific reduced feature set.
Results: Testing accuracies of 86.87% and 76.84% are reported for the first and second-level classification. Empirical results demonstrate that the proposed approach yields favourable prediction results while greatly reduces the computational time.
Conclusions: The study suggests that it is possible to predict the health condition of a person based on his/her face appearance using cost-effective machine learning approaches.

Keywords
Machine learning, Health prediction, Remote screening and diagnosis

Open Peer Review

Approval Status ✓ ?

	1	2
version 2 (revision) 02 Mar 2022		
version 1 13 Sep 2021	✓ view	? view

1. **Andrews Samraj** [iD], Mahendra Engineering College, Namakkal, India
   Mahendra Engineering College, Namakkal, India

2. **Prabina Kumar Meher** [iD], ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the Research Synergy Foundation gateway.

Corresponding author: Tee Connie (tee.connie@mmu.edu.my)

Author roles: Khong FY: Data Curation, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation; **Connie T**: Conceptualization, Project Administration, Resources, Supervision, Validation, Writing – Review & Editing; **Goh MKO**: Investigation, Software, Validation, Writing – Review & Editing; **Wong LP**: Formal Analysis, Validation, Visualization, Writing – Review & Editing; **Teh PS**: Validation, Visualization, Writing – Review & Editing; **Choo AL**: Validation, Writing – Review & Editing

How to cite this article: Khong FY, Connie T, Goh MKO *et al.* **Non-invasive health prediction from visually observable features [version 2; peer review: 1 approved, 1 approved with reservations]** F1000Research 2022, **10**:918 https://doi.org/10.12688/f1000research.72894.2

First published: 13 Sep 2021, **10**:918 https://doi.org/10.12688/f1000research.72894.1

> **REVISED** **Amendments from Version 1**
>
> Following are the changes from previous version of the article:
>
> - More work related to the study has been added in the Literature Review
> - More details about how the features are extracted from the images are provided in the Proposed Solution
> - A comparison with state-of-the-art methods has been presented in the Experiment
>
> These changes are reflected in a revised Table 1 and new Table 6.
>
> **Any further responses from the reviewers can be found at the end of the article**

Introduction

As technology advances, machine learning techniques have been growing in popularity over the past years. Machine learning techniques have proven to be effective in solving many modern problems in different domains. There is an increased research interest in applying machine learning methods for clinical informatics and healthcare systems.[1-4] Meanwhile, facial recognition technology has been vastly utilized in various fields. For instance, it has been applied to unlock phones, find wanted fugitives and diagnose diseases. There have been many kinds of research done on disease diagnosis using facial images.[5-8] Systems that only use facial features to diagnose illnesses are beneficial for remote medical diagnosis.

In this research, a machine learning approach was developed to detect the health condition of a person based on facial features. The purpose of the health prediction system was to identify images as 'healthy', or 'ill' with either 'fever', 'sore throat', or 'runny nose' symptoms. Facial images containing healthy and ill faces (fever, sore throat and runny nose) were collected. Then, discriminative facial features were extracted from the images using different feature extraction techniques. These features were used to train several machine learning classifiers for health prediction.

Literature review

In this section, various types of approaches to health prediction using facial features are studied and reviewed to learn about their respective advantages and disadvantages. These approaches are separated into two categories: conventional approaches and deep learning approaches.

Conventional approaches

In 2013, Zhao *et al.*[1] introduced an approach to classify Down Syndrome through image-based facial dysmorphology. Facial features were extracted using Contourlet transform-based and local binary pattern- based (LBP) local texture features, as well as geometric features using landmarks of facial anatomy. The support vector machine (SVM) classifier, this technique has produced an accuracy of 97.92%.

A survey done by[2] about genetic disorders diagnosis based on facial images, Saraydemir *et al.*[3] presented an approach to identify subjects with Down Syndrome from healthy subjects using facial image. Gabor wavelet transform (GWT) was implemented for feature extraction purposes. Then, linear discriminant analysis (LDA) and principal component analysis (PCA) were carried out for the reduction of dimension. 96% and 97.34% accuracy were produced.

A research conducted by[5] developed an approach for identifying Down Syndrome based on analysis of facial landmarks on 2D images. An independent component analysis-based hierarchical constrained local model (HCLM) was introduced to identify the landmarks of a face. The method was also tested on a mixed-syndromes dataset, and the highest accuracy achieved was 97%.

Another study related to health prediction systems using facial features that uses traditional machine learning methods, is an acromegaly identification using facial images proposed by.[6] A few conventional methods such as SVM, generalized linear models (LM), k-NN, RF of randomized trees (RT) as well as other deep learning methods were used to train the model. The best performance was attained by the SVM method with a 95% PPV and 88% NPV, and with an accuracy of 91%. With frontalized faces, k-NN worked best with 89% PPV and 93% NPV, also with an accuracy of 91%.

Deep learning approaches

In 2018, Sajid *et al.*[7] developed a palsy grading system based on unsymmetrical facial features using deep learning. A convolutional neural network (CNN) was proposed to extract features that exhibited palsy symptoms from a large number of facial images. The results of the model on the improved dataset showed a recognition rate of 92.6%.

A facial analysis framework introduced by[8] called DeepGestalt, to identify rare genetic syndromes using deep learning. The training process of the DeepGestalt model consisted of two steps. Firstly, an overall representation of the face was learned by the model. The binary classification problem of identifying Angelman Syndrome (AS) and Cornelia De Lange Syndrome (CdLS) patients achieved an accuracy of 92% and 96.88%, respectively.

In year 2020,[9] proposed to detect cancer using the facial features of patients. They used the network architecture of a residual neural network (ResNet) which comprised 27 convolution layers and two fully connected (FC) layers. Transfer learning was also applied for convolution layers 1-5 by directly obtaining the weights from a pre-trained face encoding model developed by.[10] To describe the distinguishing traits of non-cancer and cancer datasets, they used gradient-weighted class activation mapping (grad-CAM) for the model that they trained. The accuracy rate produced by this approach was 82%.

Apart from that,[11] developed a technique to detect Down Syndrome automatically based on facial images with deep convolutional neural network (DCNN). Firstly, they trained a DCNN on a large dataset to acquire an overall face encoding network. The network architecture consists of ten convolutional layers activated by ReLU along with three FC layers. This method achieved an accuracy of 95.87%.

Also in 2020,[12] developed a study to diagnose and classify the severity of acromegaly at different severity levels using facial images with deep learning. CNN was used in this method. For facial recognition, the pre-trained Inception ResNet V1 was utilized to extract features. The total prediction accuracy achieved by this method was 90.7%.

Besides, Forte *et al.*[13] presented a deep learning approach to assess a patient's health by using facial and bodily cues. To increase the dataset size, a synthetic dataset containing acutely ill images were generated using a neural transfer CNN network. After that, four CNN models were trained on different parts of the faces and the features were concatenated into a final feature and fed to a staked CNN. The proposed model was tested using a dataset that was made up of images of volunteers injected with lipopolysaccharide.

On the other hand, Onyema *et al.*[14] performed facial recognition for patients monitoring using ResNet. Facial emotions is believed to be closely related to the patient's state of mind. The seven universal emotions including happy, sad, fear, anger, surprise and neural were investigated. Data augmentation was applied to increase the diversity of the data. An accuracy of 70% was achieved using the proposed approach.

Recently, Connie *et al.*[15] proposed an explainable AI approach for providing explanations for the predictions made by an AI model for health application. A transfer learning approach with VGGFace model was applied to process the facial images. After that, an outcome whether the face belongs to a sick person was derived. Explainable AI (XAI) was used to provide explanation why the outcome, e.g. sick or healthy face, was produced. Different XAI techniques including Integrated Gradient, Explainable region-based AI (XRAI) and Local Interpretable Model-Agnostic Explanations (LIME) were investigated in the paper. The proposed approach had helped to increase the accountability of the healthcare system. A summary of works related to this study (hand-crafted features based methods), together with the pros and cons of each method, is presented in Table 1.

Methods
Proposed solution
A two-level classification approach is presented in this paper for health prediction based on facial features. Figure 1 shows the processes of how a prediction model was developed. First, facial images of healthy and ill (fever, sore throat and runny nose) persons were collected. Then, these images were pre-processed to clean, standardize and normalize the data. There are two levels of classification. The first-level classification is responsible for classifying samples into 'healthy' and 'ill' classes, while the second-level classification is in charge of classifying the 'ill' samples' into 'fever', 'sore throat', and 'runny nose' classes. Therefore, there are two levels of model training in the proposed solution. In this research, conventional machine learning methods were adopted.

Table 1. A summary of works related to this study.

| Author	Method	Database	Classes	Recognition Rate	Pros	Cons
Zhao, Q., Rosenbaum, K., Sze, R., Zand, D., Summar, M., & Linguraru, M. G.[1]	■ Geometric + SVM ■ Texture + SVM ■ Combined + SVM	Self-collected dataset	Down syndrome + Normal	97.92%	1. Contourlets preserve important wavelet features and provide a high level of anisotropy and directionality 2. LBP features are robust against illumination changes and takes less computational time	Facial anatomical landmarks and texture features need to be defined manually, requires more time and effort
Saraydemir, Ş., Taşpınar, N., Eroğul, O., Kayserili, H., & Dinçkan, N.[3]	■ GWT + PCA & LDA + SVM ■ GWT + PCA & LDA + k-NN	■ University Medicine Faculty Department of Medical Genetics ■ Down Syndrome Association of Turkey and Istanbul	Down syndrome + Healthy	97.34%	1. Dataset is small to produce robust results 2. Resistant to biases due to pose, illumination, and expression variances	Manual normalization requires more effort and time than automated approaches
Ferry, Q., Steinberg, J., Webber, C., FitzPatrick, D. R., Ponting, C. P., Zisserman, A., & Nellåker, C.[4]	PCA + AAM + k-NN	■ Publicly available resources ■ Scientifically published pictures of patients	Eight genetic disorders + Healthy	99.5%	1. Robust to artificial variations such as lighting, pose, and image quality 2. Provides consistent computational descriptions of facial gestalt	1. AAMs involve complex texture mapping and image warping operations which are susceptible to errors 2. AAMs have low performance on unseen faces
Zhao, Q., Okada, K., Rosenbaum, K., Kehoe, L., Zand, D. J., Sze, R., Summar, M., & Linguraru, M. G.[5]	Features: ■ Geometric ■ LBP ■ Geometric + LBP ■ GWT ■ Geometric + GWT Classifiers: ■ SVM-RBF ■ Linear SVM ■ k-NN ■ RF ■ LDA	Self-collected dataset	Down syndrome + Healthy	96.7%	1. CLMs are more generative and discriminative on unseen appearance	1. ICA requires large datasets to train to produce good results 2. Optimization can converge to local minima or false locations
			Mixed syndromes + Healthy	97%	2. CLMs are more constant to global illumination variation and occlusion	
Kong, X., Gong, S., Su, L., Howard, N., & Kong, Y.[6]	■ k-NN ■ SVM ■ RF	■ SCUT-FBP dataset ■ Neurosurgery inpatient departments of hospitals in China ■ Self-collected dataset	Acromegaly + Normal	95%	SVM performs well on extracted facial features	1. A possibility of bias caused by the selection of samples may occur 2. It is not known whether the outcome is generalizable to different populations
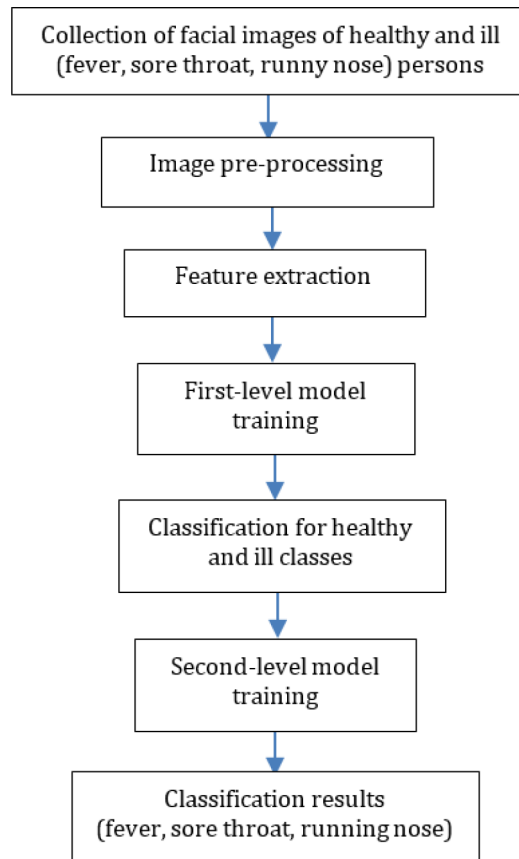
Figure 1. Proposed framework.

The feature extraction methods used were local binary pattern (LBP), Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Gabor filter. LBP is a straightforward texture analysis method that constructs binary numbers by thresholding the neighbours of every pixel in an image. For every pixel, its eight neighbours are examined to see whether their intensity is higher than the particular pixel. The threshold results from the eight neighbours are used to construct an eight-digit binary number. If the intensity of the neighbour is less than or equal to the pixel, then the first digit of the binary number would be 0, otherwise, it would be 1. Then, the texture of the image is represented by a histogram of these numbers.

On the other hand, PCA is a dimensionality reduction method that works by finding out patterns and correlations that best represent the data in a least-square sense. Higher-dimensional data are projected to a lower-dimensional space. It is an unsupervised technique that does not consider labels. It seeks directions that maximize variance and are efficient for representation.

LDA is also a dimensionality reduction tool. Higher-dimensional data are projected to a lower-dimensional space. It works by finding the projection that best separates the data of two or more classes in a least-square sense. It is a supervised technique that seeks directions that maximizes the distance between classes and are efficient for discrimination.

On the contrary, Gabor filter is a technique used for texture analysis, edge detection, feature extraction and more. These filters have been claimed that they stimulate the visual system of some mammals. They can filter any particular frequencies in an image in the region of analysis. For example, they recognize some specific frequencies and ignore the rest. To analyse the texture from an image, a collection of Gabor filters containing different orientations are applied.

In this study, the pre-processed images are converted to grayscale before the features are extracted from the images. After that, the features from the images are extracted in two different ways. For LBP and Gabor filter features, the feature extraction procedure follows the order of: loading original and augmented images, extracting features from the whole dataset, separating the original and augmented images' extracted features from the dataset, splitting the extracted features

of the original dataset into training and testing sets, adding the extracted features of the augmented dataset into the training set, and shuffling, and scaling the training and testing sets (as required depending on the model's performance).

On the other hand, for PCA and LDA features, the feature extraction procedure follows the order of: loading original and augmented images, splitting the original dataset into training and testing sets, adding the augmented dataset into the training set, and shuffling, scaling the training and testing sets (as required depending on the model's performance), and extracting features from the training and testing sets.

The classifiers used were SVM, NN, KNN, and RF. A total of 16 combinations among the feature extraction techniques and classifiers mentioned were experimented with to find the best-performing model.

Datasets
In this study, a total of 733 facial images of healthy and ill persons were collected. Among 733 images, 233 are images of ill persons who had either fever, sore throat, or runny nose and 500 were images with healthy or normal persons. 420 out of the 500 healthy images contained normal faces of people from ages 1 to 50, while the remaining 80 images were healthy throat images. Images of healthy throat and ill persons were manually collected from various online sources, while images of healthy faces were obtained from the UTKFace database.[16] The number of images for each class and subclass is listed in Table 2.

Results and discussion
In this section, the experimental results for the different models that consist of the combinations of four feature extraction methods and four classifiers are presented, analysed and discussed. The testing accuracies of the first and second-level classification of each model were recorded for 10 runs.[17]

SVM variants
The first experiment validates the performance of the SVM variant. Table 3 demonstrates the results of the SVM variant for the first and second classification tests. Among LBP, PCA, LDA and Gabor filter features, PCA features performed

Table 2. Number of images for each class and subclass.

Class	Subclass	Number of images
Healthy	-	500
Il	Fever	78
	Sore throat	80
	Running nose	75

Table 3. Experimental results of SVM variants.

Methods	1st Level classification testing	2nd Level classification testing
LBP + SVM	80.88	73.32
PCA + SVM	85.85	64.05
LDS + SVM	85.37	63.01
GABOR FILTER + SVM	81.29	63.45

Table 4. Experimental results of NN variants. The best results are highlighted in bold.

Methods	1st Level classification testing	2nd Level classification testing
LBP + NN	86.87	**76.84**
PCA + NN	**91.84**	66.96
LDA + NN	86.87	63.19
GABOR FILTER + NN	86.05	66.93

the best with SVM in the first-level classification. It achieves a promising result of 85.85% average testing accuracy with minimal overfitting. On the other hand, the LBP features performed the best with SVM in the second-level classification, with an average testing accuracy of 73.32%. The SVM variants generally produced results with the least overfitting among all the classifiers.[18]

NN variants

The experimental results for the NN variants are depicted in Table 6. Among all the feature extraction techniques, PCA features worked best with NN in the first-level classification. It achieved an average testing accuracy as high as 91.84%. On the other hand, the LBP features performed best with NN in the second-level classification with an average testing accuracy of 76.84%. In the second-level classification, the LBP model was also the only model that stood out among the other NN variants.[19]

KNN variants

The performance of the KNN variants for the first and second level classifications is given in Table 5. Among all the feature extraction techniques, again, PCA features worked best with KNN in the first-level classification, with an average testing accuracy as high as 90.34%. The same model also performed best in the second-level classification among all the KNN variants, as it obtained an average testing accuracy of 70.03%.

RF variants

The experimental results for the RF variants are displayed in Table 6. Among all the features extraction techniques, once again, at 88.57% average testing accuracy, PCA features performed the best with RF in the first-level classification. This model also scored best in the second-level classification among all the RF variants as it obtained an average testing accuracy of 74.15%.

First-level classification results

According to the experimental results of all the models shown in Tables 3 to 6, two models achieved over 90% average testing accuracies in the first-level classification. These models are the PCA+NN and PCA+KNN model.

PCA+NN

The model that achieved the highest accuracy in the first-level classification was PCA+NN. It obtained a 91.84% average testing accuracy. The high accuracy could be due to the fact that PCA effectively reduces the dimensions of data and it is able to capture the important correlations and patterns that best characterize the data. The misclassified samples were plotted during one of the runs of the finalized PCA+NN model. Out of the 147 samples, there were 15 misclassified samples.

Table 5. Experimental results of KNN variants. The best results are highlighted in bold.

Methods	1st Level classification testing	2nd Level classification testing
LBP + NN	83.54	65.33
PCA + KNN	**90.34**	**70.03**
LDA + KNN	86.53	63.78
GABOR FILTER + KNN	72.26	63.02

Table 6. Experimental results of RF variants. The best results are highlighted in bold.

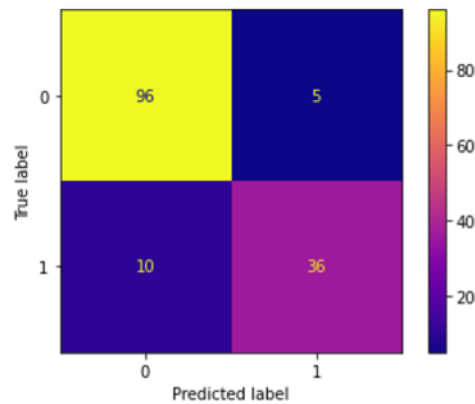| Methods	1st Level classification testing	2nd Level classification testing
LBP + RF	83.61	67.95
PCA + RF	**88.57**	**74.15**
LDA + RF	85.31	64.15
GABOR FILTER + RF	87.89	62.41

Figure 2. Confusion matrix of PCA+KNN at first-level classification.

PCA+KNN

The PCA+KNN model obtained the second-highest accuracies in the first-level classification after PCA+NN. Its performance was as good as that of PCA+KNN as it achieved a 90.34% average testing accuracy. Figure 2 shows the confusion matrix after running the first-level classification of PCA+KNN model. There is not much difference between the performance of PCA+NN and PCA+KNN as both of them were able to perform equally well.

Overall first-level performance

Apart from PCA+NN and PCA+KNN, the overall results of the first-level classification were rather good as most of the models achieved an average testing accuracy of 80% and above. Even though the other models overfit more than PCA+NN and PCA+KNN, their results were still considered rather satisfactory. The symptoms shown on the faces of ill people or sore throats are important features to help the model classify healthy and ill samples.[20]

Second-level classification results
Based on the results given in Tables 3 to 6, a total of four models achieved average testing accuracies between 70% and 77% in the second-level classification. These models were the LBP+NN, PCA+RF, LBP+SVM, and PCA+KNN model.

LBP+NN

The model that achieved the highest accuracy in the second-level classification was LBP+NN. It obtained an average testing accuracy of 76.84%. Its performance was considered rather satisfactory, as most of the other models only obtained testing accuracies between 60% and 68% on average. The reason that LBP+NN could perform well could be that the LBP features were invariant to illumination and were highly discriminative.

PCA+RF

The PCA+RF model performed nearly as well compared to LBP+NN with an average testing accuracy of 74.15% in the second-level classification. It performed well due to the previously mentioned benefits of the combination of PCA and RF being a classifier with outstanding predictive capabilities. Figure 3 shows the confusion matrix produced after running the second-level classification of PCA+RF model.

The confusion matrix was generated during one of the runs of the finalized PCA+RF model. The 0 label represents the 'fever' class, 1 represents the 'sore throat' class and 2 represents the 'running nose' class. It can be seen that the top two misclassified classes were the 'fever' (0) and 'runny nose' (2) classes, with 15 fever samples misclassified as runny nose and seven runny nose samples misclassified as fever. The reason for this occurrence is the same as for the LBP+NN model's case. The total number of samples misclassified by PCA+RF for this run was 31 samples, with only one additional misclassified sample compared to the LBP+NN model. Hence, PCA+RF was able to produce results as good as LBP+NN in the second-level classification.
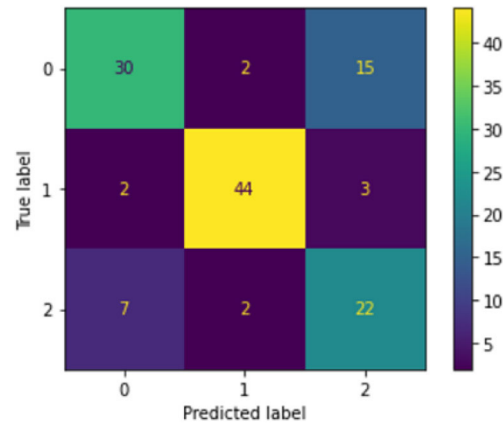
Figure 3. Confusion matrix of PCA+RF at second-level classification.

Table 7. A comparison with state-of-the-art methods.

Methods	1 st Level classification testing
LBP + NN (First level classification)	86.87
PCA + RF (First level classification)	88.57
CNN[7]	92.71
VGGFace[21]	96.25

LBP+SVM and PCA+KNN

Other than the LBP+NN and PCA+RF, the two remaining models that achieved over 70% average testing accuracies were LBP+SVM and PCA+KNN. The LBP+SVM model obtained a 73.32% average testing accuracy in the second-level classification. The reason behind its performance is the robustness of LBP as well as the fact that SVM is effective in situations where the number of dimensions is larger than the number of samples. In the model's second-level classification, the number of testing samples was always lesser than the number of dimensions.

Best model for the health prediction system
Among all models, the LBP+NN variant had the best overall performance in the first and second-level classifications. It achieved the highest average testing accuracy of 76.84% in the second-level classification. It also performed considerably well in the first-level classification with lesser overfitting than the other models with similar performances, as it showed 94.38% and 86.87% average training and testing accuracies, respectively.

Comparison with other methods
A comparison of the proposed methods with state-of-the-art approaches is presented in Table 7. It can be observed that the deep learning approaches including CNN[1] and VGGFace[2] outperform the proposed methods that rely on hand-crafted features. Nevertheless, the proposed approach has a great advantage as compared to the deep learning approaches in terms of computational speed. For example, it only took 0.0015 seconds to train the PCA+RF classifier, while it takes more than five minutes to perform training using the deep learning models. Therefore given a scenario where speed is a critical requirement and there is not many training samples available, the proposed method appears to be a more favourable choice.

Conclusions
This paper presents a health prediction system using facial features evaluated using different machine learning models. Datasets containing facial images of healthy and ill (fever, sore throat and runny nose) persons were collected. The facial features of the images were extracted using LBP, PCA, LDA and Gabor filter feature extraction techniques. The features were trained using SVM, NN, KNN and RF classifiers. Among the 16 models, the LBP+NN model yielded the best overall performance for both the first and second-level classifications. It obtained average testing accuracies of 86.87% and 76.84% for the first and second-level classification, respectively.

Data availability

Underlying data

UTKFace Large Scale Face Dataset: https://susanqq.github.io/UTKFace/.

As it is impossible to obtain the consent for the face images retrieved from the UTKFace dataset, the images cannot be shared in this article.

Software availability

Source code available from: https://doi.org/10.5281/zenodo.5266406.[22]

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

References

1. Zhao Q, Rosenbaum K, Sze R, *et al.*: **Down syndrome detection from facial photographs using machine learning techniques**. Novak CL, Aylward S, Eds.; 2013; p. 867003.

2. Rai EL, Werghi MC, Al Muhairi N, *et al.*: **Using facial images for the diagnosis of genetic syndromes: A survey.** *2015 Int Conf Communications, Signal Processing, and Their Applications (ICCSPA'15).* 2015; 1–6.
Publisher Full Text

3. Saraydemir Ş, Taşpınar N, Eroğul O, *et al.*: **Down Syndrome Diagnosis Based on Gabor Wavelet Transform.** *J Med Syst.* 2012; **36**(5): 3205–3213.
PubMed Abstract | **Publisher Full Text**

4. Ferry Q, Steinberg J, Webber C, *et al.*: **Diagnostically relevant facial gestalt information from ordinary photos.** *ELife.* 2014; **3**: e02020.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Zhao Q, Okada K, Rosenbaum K, *et al.*: **Digital facial dysmorphology for genetic screening: Hierarchical constrained local model using ICA.** *Med Image Anal.* 2014; **18**(5): 699–710.
PubMed Abstract | **Publisher Full Text**

6. Kong X, Gong S, Su L, *et al.*: **Automatic Detection of Acromegaly From Facial Photographs Using Machine Learning Methods.** *EBioMedicine.* 2018; **27**: 94–102.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Sajid M, Shafique T, Baig M, *et al.*: **Automatic Grading of Palsy Using Asymmetrical Facial Features: A Study Complemented by New Solutions.** *Symmetry.* 2018; **10**(7): 242.
Publisher Full Text

8. Gurovich Y, Hanani Y, Bar O, *et al.*: **DeepGestalt - Identifying Rare Genetic Syndromes Using Deep Learning.** *ArXiv:1801.07637 [Cs].* 2018.
Publisher Full Text

9. Liang B, Yang N, He G, *et al.*: **Identification of the Facial Features of Patients With Cancer: A Deep Learning–Based Pilot Study.** *J Med Internet Res.* 2020; **22**(4): e17234.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

10. Schroff F, Kalenichenko D, Philbin J: **FaceNet: A unified embedding for face recognition and clustering.** *IEEE Conf Computer Vision Pattern Recognition (CVPR).* 2015; **2015**: 815–823.
Publisher Full Text

11. Qin B, Liang L, Wu J, *et al.*: **Automatic Identification of Down Syndrome Using Facial Images with Deep Convolutional Neural Network.** *Diagnostics.* 2020; **10**(7): 487.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. Kong Y, Kong X, He C, *et al.*: **Constructing an automatic diagnosis and severity-classification model for acromegaly using facial photographs by deep learning.** *J Hematol Oncol.* 2020; **13**(1): 88.
Publisher Full Text

13. Forte C, Voinea A, Chichirau M, *et al.*: **Deep Learning for Identification of Acute Illness and Facial Cues of Illness.** *Front Med.* 2021; **8**: 661309.
PubMed Abstract | **Publisher Full Text**

14. Onyema EM, Shukla PK, Dalal S, *et al.*: **Enhancement of Patient Facial Recognition through Deep Learning Algorithm: ConvNet.** *J Healthc Eng.* 2021; **2021**: 8.
Publisher Full Text | **PubMed Abstract**

15. Connie T, Tan YF, Goh MKO, *et al.*: **Explainable Health Prediction from Facial Features with Transfer Learning.** *J Intell Fuzzy Syst..* 2022; **42**(3): 2491–2503.
Publisher Full Text

16. Zhang Z, Song Y, Qi H: **Age Progression/Regression by Conditional Adversarial Autoencoder.** *IEEE Conf Computer Vision Pattern Recognition (CVPR).* 2017: 4352–4360.

17. Jordan J: **Evaluating a machine learning model.** 2017.
Reference Source

18. Abdiansah A, Wardoyo R: **Time Complexity Analysis of Support Vector Machines (SVM) in LibSVM.** *Int J Computer Applications.* 2015; **128**(3): 28–34.
Publisher Full Text

19. Karis NS, Rafiqah N, Nursabillilah, *et al.*: **Local Binary Pattern (LBP) with application to variant object detection: A survey and method.** *12th Int Colloquium on Signal Processing & Its Applications (CSPA).* 2016). 2016.
Publisher Full Text

20. British Columbia Health Link BC: **Facts about Influenza.** 2021.
Reference Source

21. Parkhi OM, Vedaldi A, Zisserman A: **Deep Face Recognition.** *Procedings of the British Machine Vision Conference 2015. British Machine Vision Association, Swansea.* 2015: 41.1–41.12.
Publisher Full Text

22. Khong FY, Connie T, Michael GKO, *et al.*: **gkomix88/ HealthPrediction: Non-invasive Health Prediction from Visually Observable Features (HealthPrediction).** *Zenodo.* 2021.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ❓

Version 1

Reviewer Report 14 February 2022

❓ **Prabina Kumar Meher** (iD)

ICAR-Indian Agricultural Statistics Research Institute, New Delhi, Delhi, India

The authors developed a machine learning-based approach for predicting the health status of an individual by using the image data of their face. Below are my comments for further improvement of the article.

1. The authors must clearly describe how the features were generated from the image dataset.

2. The deep learning method convolutional neural network (CNN) is one of the most appropriate methods for prediction using image data. The authors should employ the CNN for the same and compare the accuracy with that of machine learning algorithms.

3. The author should perform a comparative analysis with the existing methods to claim the superiority of the method.

4. The author must try to establish an online prediction tool for the real use of the developed approach.

5. More work related to this subject must be discussed.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Application of Machine learning for prediction/classification using biological data in agricultural and biomedical science.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 16 Feb 2022
Tee Connie, Multimedia University, Melaka, Malaysia

Dear Reviewer,

Thank you very much for your time and efforts in reviewing our manuscript "Non-Invasive Health Prediction from Visually Observable Features". According to your valuable comments and suggestions, more analysis has been conducted. In this response letter, we list the specific concerns and questions raised by the reviewer and provide our itemized response.

Point 1: The authors must clearly describe how the features were generated from the image dataset.

Response 1: We thank the reviewer for the careful review and comment. We understand that it is important to describe how the features were generated from the images. In the study, four feature extraction methods namely local binary pattern (LBP), Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Gabor filter were applied. The details how the features are extracted from the image have been added.

Point 2: The deep learning method convolutional neural network (CNN) is one of the most appropriate methods for prediction using image data. The authors should employ the CNN for the same and compare the accuracy with that of machine learning algorithms.

Response 2: This is a very good suggestion. In fact, we had also performed experiments using CNN on the image data. However, as the scope of the paper is more towards the analysis and comparison of conventional feature extraction and classification methods rather than the deep learning approach, we did not include the details for CNN in the paper.

Anyway, to better illustrate the use of CNN on the health dataset, we provide some

information about the experiments using CNN in this response letter here.

Point 3: The author should perform a comparative analysis with the existing methods to claim the superiority of the method.

Response 3: We thank the reviewer for the suggestion. It is important to provide a comparative analysis with the existing methods to claim the superiority of the method. To better illustrate the performance of the proposed methods as compared to state-of-the-art techniques, we have added a new section in Experiments.

Point 4: The author must try to establish an online prediction tool for the real use of the developed approach.

Response 4: We thank the for the suggestion. It is advantageous to establish an online prediction tool for the real use of the developed approach and we will consider this in our future endeavors.

Point 5: More work related to this subject must be discussed.

Response 5: Thank you for the suggestion. In addition to the related works that have been discussed in the Literature Review section in the paper, more studies related to the use of face images for health prediction are provided in the article.

Competing Interests: No competing interests were disclosed.

Reviewer Report 14 September 2021

https://doi.org/10.5256/f1000research.76503.r94202

✔ **Andrews Samraj** iD
¹ Department of Computer Science and Engineering, Mahendra Engineering College, Namakkal, Tamil Nadu, India
² Mahendra Engineering College, Namakkal, Tamil Nadu, India

Non-invasive health prediction is the essential need of the hour due to the pandemic situation in health care. The approach and modeling using image data set of human faces with sound health and with illness are taken for research and the findings were presented in this work.

Though the testing accuracies are low, an increase in the number of samples may help in improving.

The modifications suggested:
1. In methods, the two-level classification has to be depicted clearly.

2. Tables are not visible. Improve font size

3. Results and discussion should only explain the results, assumptions, and parameters. So explain the variants in the Methods section.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Image processing, Non invasive signal and image analysis, Soft cyborgs, Wearable Technology, Bionics, AI-ML

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research