Response to Reviewer's Comments

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 have 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 are added in the Proposed Solution section as follows:

The four feature extraction methods used for this study are 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 is projected to a lower-dimensional space. It is an unsupervised technique that does not take labels into account, it seeks directions that maximize variance and are efficient for representation.

LDA is also a dimensionality reduction tool. Higher-dimensional data is 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 then 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.

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 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 below:

CNN is one of the most widely used deep learning architecture to perform prediction/classification for image data. The CNN architecture shown in Figure 1 is used to process the health images.

```
Model: "sequential_61"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_430 (Conv2D)          (None, 194, 194, 64)      9472
_________________________________________________________________
max_pooling2d_409 (MaxPoolin (None, 97, 97, 64)        0
_________________________________________________________________
conv2d_431 (Conv2D)          (None, 97, 97, 128)       73856
_________________________________________________________________
max_pooling2d_410 (MaxPoolin (None, 48, 48, 128)       0
_________________________________________________________________
conv2d_432 (Conv2D)          (None, 48, 48, 256)       295168
_________________________________________________________________
max_pooling2d_411 (MaxPoolin (None, 24, 24, 256)       0
_________________________________________________________________
flatten_56 (Flatten)         (None, 147456)            0
_________________________________________________________________
dense_224 (Dense)            (None, 128)               18874496
_________________________________________________________________
dropout_168 (Dropout)        (None, 128)               0
_________________________________________________________________
dense_225 (Dense)            (None, 64)                8256
_________________________________________________________________
dropout_169 (Dropout)        (None, 64)                0
_________________________________________________________________
dense_226 (Dense)            (None, 2)                 130
=================================================================
Total params: 19,261,378
Trainable params: 19,261,378
Non-trainable params: 0
_________________________________________________________________
```

Figure 1. Architecture of CNN model.

The accuracy obtained using this network model was 0.9062. To further improve the result, hyperparamete tuning was performed. First, the activitation function was investigated. Several activation functions like sigmoid function, Tanh, Elu and also Softmax activation function were tested. From Table 1, the softmax activation function gave the best result among the others.

Table 1. Hyperparameter tuning for activation functions.

Activation Function	Accuracy
Sigmoid	0.8931
Tanh	0.6233
Elu	0.6233
Softmax	0.9271

Besides, we have also investigated the appropriate dropout rate for the model training. We observe from Table 2 that an accuracy of 0.9271 is obtained in 5 epochs using a dropout rate of 0.3. Droping out some of the nodes during training has indeed helped to increase the model accuracy by avoiding the model from learning unimportant features.

Table 2. Hyperparameter tuning for dropout rates.

Dropout Rate	Accuracy
0.3	0.9271
0.5	0.9110
0.7	0.6233

Figure 2 presents the loss and epoch relationship for the CNN model.

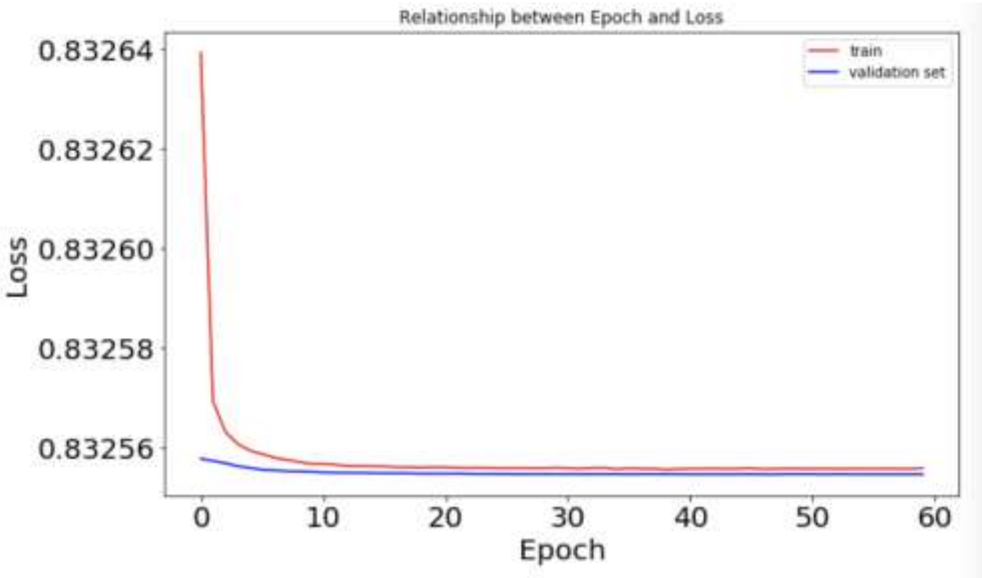

Figure 2. Relationship between epoch and loss for CNN model.

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 Experiment as below:

A comparison of the proposed methods with state-of-the-art approaches is presented in Table 3. 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 favorable choice.

Table 3. A comparison with state-of-the-art methods.

Method	*Accuracy*
LBP+NN (First level classification)	*0.8687*
PCA+RF (First level classification)	*0.8857*
CNN [1]	*0.9271*
VGGFace [2]	*0.9625*

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature. 521 (2015) 436–444. doi:10.1038/nature14539.

[2] O.M. Parkhi, A. Vedaldi, and A. Zisserman, Deep Face Recognition, in: Procedings of the British Machine Vision Conference 2015, British Machine Vision Association, Swansea, 2015: p. 41.1-41.12. doi:10.5244/C.29.41.

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 endeavor.

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 as follows:

Forte et al. [1] 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. [2] 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 [3] 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, together with the pros and cons of each method, is presented in Table 4.*

Table 4. 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.	▪ *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.	▪ *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.	*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.	*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* *2. CLMs are more constant to global illumination variation and occlusion*	*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%*		
Kong, X., Gong, S., Su, L., Howard, N., & Kong, Y.	▪ *OpenCV + Dlib + LM* ▪ *OpenCV + Dlib + k-NN* ▪ *OpenCV + Dlib + SVM* ▪ *OpenCV + Dlib + RT*	▪ *SCUT-FBP dataset* ▪ *Neurosurgery inpatient departments of hospitals in China* ▪ *Self-collected dataset* ▪ *Previously published studies*	*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*

References

[1] Forte C, Voinea A, Chichirau M, Yeshmagambetova G, Albrecht LM, Erfurt C, Freundt LA, Carmo LOe, Henning RH, Horst ICCvd, Sundelin T, Wiering MA, Axelsson J and Epema AH (2021) Deep Learning for Identification of Acute Illness and Facial Cues of Illness. Front. Med. 8:661309. doi: 10.3389/fmed.2021.661309

[2] Onyema EM, Shukla PK, Dalal S, Mathur MN, Zakariah M, Tiwari B (2021) Enhancement of Patient Facial Recognition through Deep Learning Algorithm: ConvNet. Journal of Healthcare Engineering. https://doi.org/10.1155/2021/5196000

[3] Tee Connie, Yee Fan Tan, Michael Kah Ong Goh, Hock Woon Hon and Zulaikha Kadim, Li Pei Wong, "Explainable Health Prediction from Facial Features with Transfer Learning", Journal of Intelligent & Fuzzy Systems, Vol. 42, no. 3, pp. 2491-2503, 2022. DOI: 10.3233/JIFS-211737

[4] 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. 24835178 10.1016/j.media.2014.04.002

[5] Saraydemir Ş, Taşpınar N, Eroğul O, et al.: Down Syndrome Diagnosis Based on Gabor Wavelet Transform. J Med Syst .2012;36(5):3205–3213. 22127522 10.1007/s10916-011-9811-1

[6] Ferry Q, Steinberg J, Webber C, et al.: Diagnostically relevant facial gestalt information from ordinary photos. ELife .2014;3:e02020. 24963138 10.7554/eLife.02020 PMC4067075

[7] 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. 24835178 10.1016/j.media.2014.04.002

[8] Kong X, Gong S, Su L, et al.: Automatic Detection of Acromegaly From Facial Photographs Using Machine Learning Methods. EBioMedicine .2018;27:94–102. 29269039 10.1016/j.ebiom.2017.12.015 PMC5828367