Real-time Emotional State Detection from Facial Expression on Embedded Devices

Saeed Turabzadeh, Hongying Meng, Rafiq M. Swash
Department of Electronic and Computer Engineering
Brunel University London,
London, UK
Hongying.Meng@brunel.ac.uk

Matus Pleva, Jozef Juhar
Dept. of Electronics and Multimedia Telecommunications
Technical University of Kosice,
Kosice, Slovakia
Matus.Pleva@tuke.sk

***Abstract*—From the last decade, researches on human facial emotion recognition disclosed that computing models built on regression modelling can produce applicable performance. However, many systems need extensive computing power to be run that prevents its wide applications such as robots and smart devices. In this proposed system, a real-time automatic facial expression system was designed, implemented and tested on an embedded device such as FPGA that can be a first step for a specific facial expression recognition chip for a social robot. The system was built and simulated in MATLAB and then was built on FPGA and it can carry out real time continuously emotional state recognition at 30 fps with 47.44% accuracy. The proposed graphic user interface is able to display the participant video and two dimensional predict labels of the emotion in real time together.**

***Keywords*—FPGA, facial expression analysis, artificial intelligence, real-time implementation.**

I. INTRODUCTION

The capability of automatic emotion recognition could improve the human computer interaction (HCI) experience in adapting to consumers' mental state. The emotion recognition is the key for modern smart phone or robot HCI [1].

In the 19th century Charles Darwin recognized these general values of expressions in both humans and animals [2]:

- low spirits, anxiety, grief, dejection and despair
- joy, high spirits, love, tender feelings and devotion
- reflection, meditation, ill-temper and sulkiness
- hatred and anger
- disdain, contempt, disgust, guilt and pride
- surprise, astonishment, fear and horror
- self-attention, shame, shyness and modest

Paul Ekman and his colleagues classify the basic emotions, and their work had a significant impact on the current emotion analysis development [3].

II. RELATED WORK

This helped in the early 1990s [4] to growth the face detection and face tracking result in Human Computer Interaction (HCI and Affective Computing [5] evolution. The emotions of the user could be detected using advanced pattern recognition algorithms from extracted image (facial expression, gestures, etc.) or audio (speech) features. Recently Cambridge University introduced the emotional computer [6] and MIT the Mood Meter [7]. From 2011 we participated in several international emotion recognition challenges like AVEC or MediaEval [8][9].

In 2013, Cheng, J., Deng, Y., Meng, H. and Wang, Z propose the GP-GPU acceleration service for continuous face and emotion detection system [10]. For real-world scenario of continuously monitoring of movie scene promising results were achieved. The system was initially tested in MATLAB. It was proven that GPU acceleration can speed up the processing by 80 times comparing to CPU. This system can provide the detected emotional state every 1.5 second [10].

In 2015 the Microsoft Oxford API cloud service provides the recognition of emotions based on facial expressions [11]. This API provides the confidence across a set of emotions for each face in the image, as well as bounding box for the face.

The emotions detected are anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. These emotions are understood to be cross-culturally and universally communicated with facial expressions. Recognition is experimental and not always accurate [11].

As far as known, this is the first time that automatic emotional state detection has been successfully implemented on embedded device (FPGA). The proposed system is 20 times faster than the mentioned GPU implementation [10] and can analyze 30 frames per-second in real time.

The proposed Local Binary Pattern (LBP) algorithm uses alteration of the image into an arrangement of micro-patterns. The performance with face images is described in [12]. On the other hand the K-Nearest Neighbor algorithm (K-NN) is utilized for regression modelling [13]. This algorithm is a based-on learning, where the operation is only approximated nearby the training dataset and all calculation is delayed until regression. In this paper, these two techniques implementation on the FPGA and evaluation is presented. The system is able to display real time and automatic emotional state detection model on the connected monitor.

III. AFFECTIVE DIMENSION RECOGNITION

In affective dimensional space, it represents more details on how the emotional states other than six basic emotion categories (e.g. happy, sad, disgust, anger, surprise and fear) are dealt. The 3D model contains Arousal, Dominance and

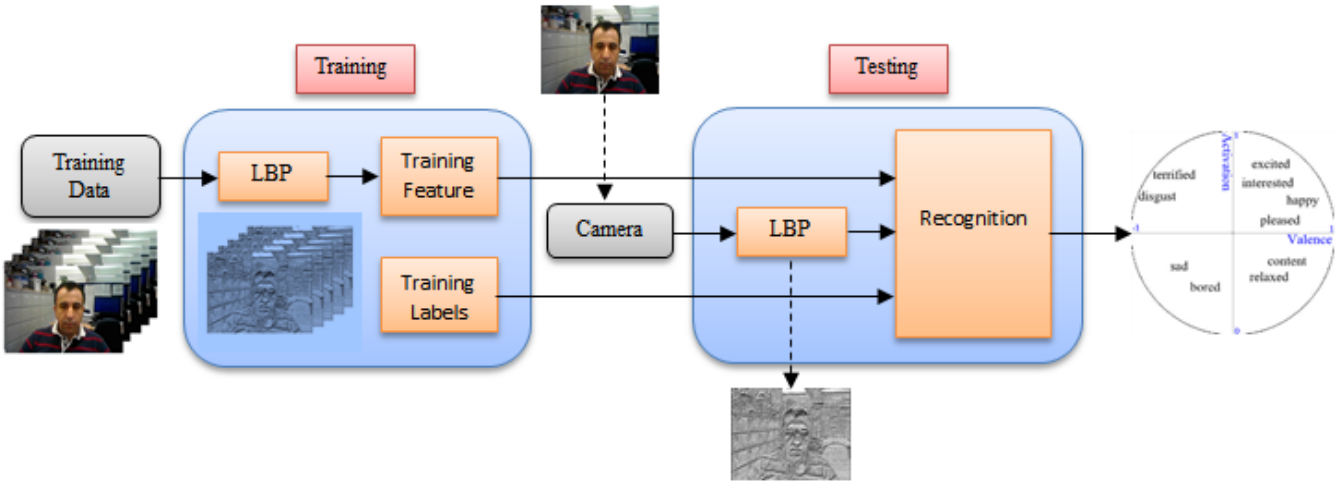

Fig 1. Real-time emotion detection system from facial expression on embedded devices

Valence. Therefore, the automatic emotional state detection system needs to comprehensively model the variations in visual and convert it to Arousal, Dominance and Valence scale from the captured video. From the machine learning point of view it is not a classification problem but a regression problem, because the predicted values are real numbers on each individual frame in image sequence.

A. System Overview

On Figure 1 the framework of automatic emotional state detection system is depicted. The inputs are the video frames captured from the camera. FeelTrace [14] was used to capture the videos and extract the training datasets. LBP features were extracted from captured videos. The K-Nearest Neighbor algorithm is used for regression. For training LBP are extracted from 10 videos image sequences. Video captures from the FPGA connected camera are processed and LBP features used for K-NN regression with dimensional emotion labels.

B. Image Feature Extraction

The LBP algorithm converts local sub-blocks of the grayscale converted image to alternation counts histogram. The histogram values of each block is horizontally concatenated together to form a single feature vector which is a descriptor for the entire image. There are numerous differences of LBP in present literature [15]. The method described in this paper uses features obtained by uniform LBP [16]. Uniform LBP uses dimensionally reduced feature vectors in order to save computational time and memory.

C. K-NN Algorithm for Regression Modelling

The K-Nearest Neighbor algorithm was utilized for regression modelling [13]. The LBP features of the K nearby images are the inputs of the network. The K-NN regression classifies the membership to a class. Usually K is a small positive integer. For example for K = 1 the test data are allocated to the class of that single nearest neighbor.

It would be more effective to choose differentia to the contributions of the neighbors, so that the closest neighbors play higher role to the average than the more distant ones [17].

The K-nearest neighbor regression is typically based on distance between a query point (test sample) and the specified samples (training dataset). In regression problems, K-NN predictions are founded on a voting system in which the winner is used to label the query.

D. Cross-Validation: Evaluating Estimator Performance

Cross validation is a well-known method for evaluation of the system on a smaller dataset when using all the data available for training and evaluation is necessary. The idea is to divide the dataset to smaller parts and exclude always another part for testing. Finally the results of all trained models are averaged and presented as cross-validated result of the algorithm [18] [19].

E. Pearson Product-Moment Correlation Coefficient

The Pearson Product-moment Correlation Coefficient (PPMCC) was used for linear correlation calculation between actual and predicted labels established by Karl Pearson [20]. There are various formulas for the correlation coefficient calculation. In this experiment the equation (1) below was used [21].

$$r = r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\,\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \qquad (1)$$

IV. Overview of the Classifier Fusion System

First of all, the MATLAB R2014a simulation of the correlation coefficient was realized to validate the results. Thereafter the MATLAB implementation was tested in real time with the camera. Finally, the Atlys™ Spartan-6 LX45 FPGA Development Board was used for FPGA implementation evaluation [22].

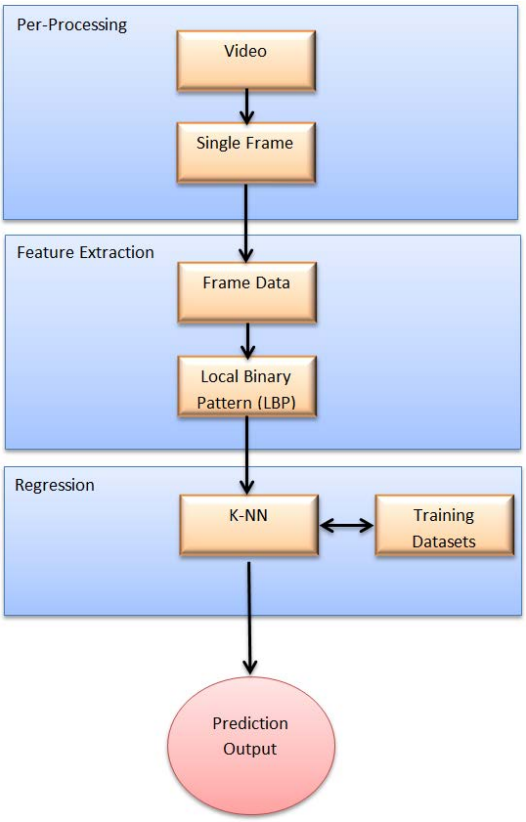

Fig 2. Workflow of the Proposed Methodology

The emotion detection was implemented using MATLAB R2014a with K-NN regression classifier from LIBSVM [23] and Xilinx ISE Design Suite 13.2 to generate VHDL code and upload the binary to FPGA board Atlys™ Spartan-6 FPGA Development Board and Spartan-6 FPGA Industrial Video Processing Kit [22]. Each FPGA output demonstrates the emotion prediction on each frame of the video from the camera. The predictions from the classifiers are used to create a single regression final output.

This FPGA implementation enables to run on real time video inputs with an ability of analysis 30 frames per-second which is 20 times faster than GP-GPU implementation [10]. The most emotional state recognition systems have suffered from the lack of real-time computing ability due to algorithm complexities. It prevents current systems to be used for real-world applications, especially for low-cost, low-power consumption and portable systems. The usage of FPGA provides a new platform for a real-time automatic emotional state detection and recognition system development [22].

V. EXPERIMENTAL RESULTS

A. Dataset

The dataset created for system evaluation was recorded with five adult human volunteers (3 males and 2 females) with 30 frames per second. Each person was asked to watch a 3 and a half minute video twice. The video contains 6 different scenarios of relaxation, funny, sadness, scary and disgusting.

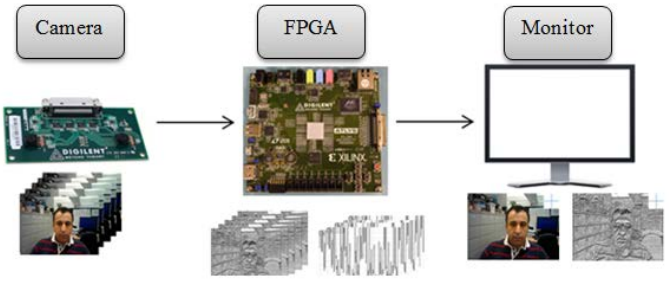

Fig 3. Data Workflow Diagram and Hardware used

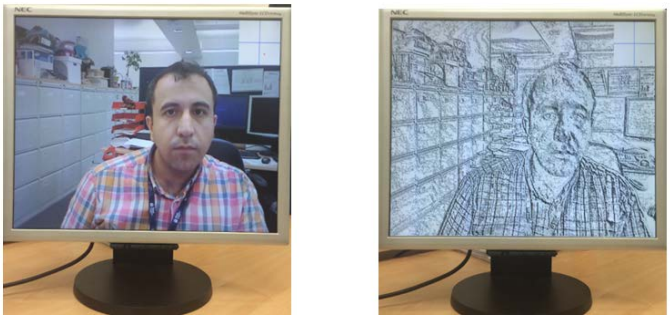

Fig 4. Real-time display of the LBP and K-NN from FPGA

This study tried to cover the emotions of FeelTrace. It was interesting that the volunteers had a dissimilar reaction to each part of the video [22]. So in over all the dataset contains 63,000 labelled samples.

Together with video the real-time features calculated on Xilinx Spartan-6 LX45 FPGA was captured by MATLAB Acquisition Toolbox for SVM training. The video captures was realized in RGB and converted to grayscale for LBP processing. The participant was captured in the office with real situations (office background, people passing, etc.).

B. FPGA Implementation

In the first part, the frames are extracted from the camera sensor captured video. In next section, the LBP features were calculated. The classification part used K-NN on the training dataset. The fourth part is for displaying the predicted labels on the attached screen. Check the block diagram and hardware realization on Figure 2 & 3.

The K-NN implementation produces the activation and valence values which were displayed on the output. After uploading of the current model's binary to the FPGA the display shows the live camera output and a modified two-dimensional FeelTrace axis on the top right corner of the monitor. The valence is presented on the horizontal axes and activation values on vertical. The predict label was on the display in red colour.

One of the FPGA board button was used for implementation of the changing of the camera view from RGB to LBP features as it could be seen on the Figure 4.

C. Performance Comparison

For cross validating the results the 5-fold and 2-fold cross validation was used. The 5-fold person independent (the

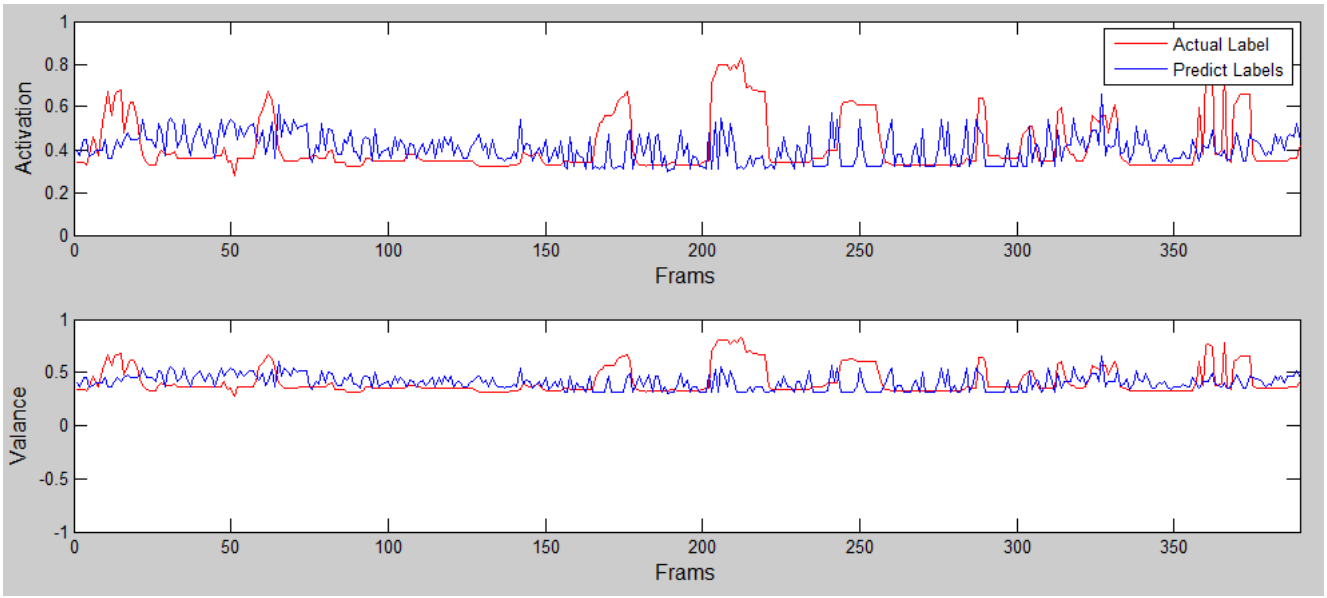

Fig 5. Example of Predict and Actual Valence and Activation Evaluation

system contains always the training data from more users) cross-validation the captures was divided in 5 subsets. For 2-fold cross validation always the second recording of the user goes to testing (see example depicted on Figure 5).

The acquired accuracy of the system is depicted on the Table I & II. The highest accuracy was achieved for 2-fold cross-validation for the K=5 with the accuracy of the 51.28% for MATLAB and 47.44% for FPGA implementation.

TABLE I. EMOTION DETECTION ACCURACY ACHIEVED FOR DIFFERENT K OF K-NN FOR MATLAB IMPLEMENTATION

Accuracy %		
K	**5-Fold (person independent)**	**2-Fold**
1	27.66	45.82
3	28.13	49.47
5	27.77	**51.28**

TABLE II. EMOTION DETECTION ACCURACY ACHIEVED FOR DIFFERENT K OF K-NN FOR FPGA IMPLEMENTATION

Accuracy %		
k	**5-Fold (person independent)**	**2-Fold**
1	25.52	43.84
3	26.81	45.93
5	26.07	**47.44**

The accuracy of the predict values was reduced on FPGA by 3.84% comparing to MATLAB implementation mainly because of using lower training datasets for the VHDL implementation. The MATLAB Simulink code runs slower mainly because of execution of the LBP algorithm and produce

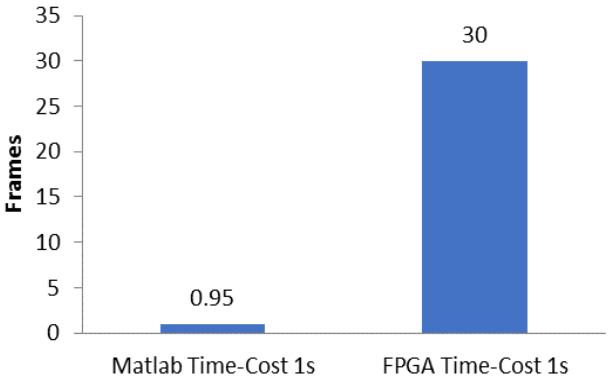

Fig 6. Comparing Computational Costs of MATLAB and FPGA implementation

0.95 frame result for every second. The FPGA implementation produce 30 frames per second real time outputs therefore we can say it is more than 30 times faster comparing to MATLAB Simulink implementation as depicted on Figure 6.

VI. CONCLUSION AND DISCUSSION

The main objective of this paper was to evaluate the FPGA implementation and efficiency of regression methods for an automatic emotional state detection and analysis. It can be concluded that the results prove that the FPGA implementation is ready to be used on embedded devices for human emotion recognition from live camera. The database of 5 users with 36,000 samples was recorded, labelled and cross validated during the experiment. The LBP method was implemented for comparison in MATLAB and FPGA too. The values of Activation and Valence for the database were extracted by FeelTrace from each frame.

The system could be improved by in the future work using different LPQ, EOH features extraction or different regression modelling methods such as Support Vector Machines. Regression classifier fusion systems, although difficult to design, have proven potential to overcome the problems facing human emotion research today [22]. The results presented in this paper using fusion system for automatic facial emotion recognition are promising. In the future also the audio features and speaker emotion detection models [24] will be tested and fused [25] with the frame level results.

ACKNOWLEDGMENT

The research presented in this paper was supported partially by the Slovak Research and Development Agency under the research projects APVV-15-0517 & APPV-15-0731 and by the Ministry of Education, Science, Research and Sport of the Slovak Republic under the project VEGA 1/0075/15.

REFERENCES

[1] Miwa, H., Itoh, K., Matsumoto, M., Zecca, M., Takanobu, H., Rocella, S., Carrozza, M.C., Dario, P. and Takanishi, A., 2004, September. *Effective emotional expressions with expression humanoid robot WE-4RII: integration of humanoid robot hand RCH-1*. In Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on, Vol. 3, pp. 2203-2208. IEEE.

[2] Darwin, C., 1998. *The expression of the emotions in man and animals*. Oxford University Press, USA.

[3] Ekman, P. and Keltner, D., 1970. *Universal facial expressions of emotion*. California Mental Health Research Digest, 8 (4), pp. 151-158.

[4] Suwa, M., Sugie, N. and Fujimora, K., 1978, November. *A preliminary note on pattern recognition of human emotional expression.* In International Joint Conference on Pattern Recognition, Vol. 1978, pp. 408-410.

[5] Picard, R.W. and Picard, R., 1997. *Affective computing*, Vol. 252. Cambridge: MIT press.

[6] El Kaliouby, R. and Robinson, P., 2005. *Real-time inference of complex mental states from facial expressions and head gestures.* In: Kisačanin B., Pavlović V., Huang T.S. (eds) Real-Time Vision for Human-Computer Interaction. Springer, pp. 181-200. Springer US.

[7] Yang, S. and Bhanu, B., 2012. *Understanding discrete facial expressions in video using an emotion avatar image*. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42 (4), pp.980-992.

[8] Meng, H. and Bianchi-Berthouze, N., 2011. *Naturalistic affective expression classification by a multi-stage approach based on Hidden Markov Models.* In: D'Mello S., Graesser A., Schuller B., Martin JC. (eds) Affective Computing and Intelligent Interaction. Lecture Notes in Computer Science, vol 6975. Springer, Berlin, Heidelberg, pp. 378-387.

[9] Meng, H., Romera-Paredes, B. and Bianchi-Berthouze, N., 2011, March. *Emotion recognition by two view SVM_2K classifier on dynamic facial expression features*. In Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, pp. 854-859. IEEE.

[10] Cheng, J., Deng, Y., Meng, H. and Wang, Z., 2013, April. *A facial expression based continuous emotional state monitoring system with GPU acceleration.* In Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, pp. 1-6. IEEE.

[11] https://azure.microsoft.com/en-us/services/cognitive-services/ (accessed June 1, 2017)

[12] Ahonen, T., Hadid, A. and Pietikainen, M., 2006. *Face description with local binary patterns: Application to face recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28 (12), pp. 2037-2041.

[13] Altman, N.S., 1992. *An introduction to kernel and nearest-neighbor nonparametric regression*. The American Statistician, 46 (3), pp. 175-185.

[14] Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M. and Schröder, M., 2000. *'FEELTRACE': An instrument for recording perceived emotion in real time.* In ITRW on SpeechEmotion-2000, pp. 19-24. ISCA.

[15] Ojala, T., Pietikainen, M. and Harwood, D., 1994, October. *Performance evaluation of texture measures with classification based on Kullback discrimination of distributions*. In Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing, Proceedings of the 12th IAPR International Conference on, Vol. 1, pp. 582-585. IEEE.

[16] Ojala, T., Pietikainen, M. and Maenpaa, T., 2002. *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 24 (7), pp. 971-987.

[17] Jaskowiak, P.A. and Campello, R.J.G.B., 2011, August. *Comparing correlation coefficients as dissimilarity measures for cancer classification in gene expression data.* In Proceedings of the Brazilian Symposium on Bioinformatics, pp. 1-8. Brasília.

[18] Kohavi, R., 1995, August. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. In proceedings of International Joint Conference on Artificial Intelligence, Vol. 14, No. 2, pp. 1137-1145.

[19] Picard, R.R. and Cook, R.D., 1984. *Cross-validation of regression models*. Journal of the American Statistical Association, 79 (387), pp.575-583.

[20] Pearson, K., 1895. *Note on regression and inheritance in the case of two parents.* Proceedings of the Royal Society of London, 58, pp.240-242.

[21] Cohen, J., 1988. *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

[22] Turabzadeh, S., 2015. *Automatic emotional state detection and analysis on embedded devices* (Doctoral dissertation, Brunel University London).

[23] Chang, C.C. and Lin, C.J., 2011. *LIBSVM: a library for support vector machines.* ACM Transactions on Intelligent Systems and Technology (TIST), 2 (3), p. 27.

[24] Mackova, L., Cizmar, A. and Juhar, J., 2015. *A study of acoustic features for emotional speaker recognition in I-vector representation*. Acta Electrotechnica et Informatica, 15 (2), pp.15-20. doi: 10.15546/aeei-2015-0011.

[25] Pleva, M., Bours, P., Ondas, S. and Juhar, J, 2017. *Improving static audio keystroke analysis by score fusion of acoustic and timing data.* Multimedia Tools and Applications. OnlineFirst March 29, 2017, OpenAccess, doi:10.1007/s11042-017-4571-7.