Detecting Education Level Using Facial Expressions

HabebaElRahman Hesham, Mourad Ahmed, Ola Mohamed, Youssef Mohamed

October 2019

Abstract

In our real-life, e-learning is improving and shows more. The examiners show every day a better way of detecting the level of education and the ground of education that student stands on. Although e-learning has some advantages in terms of information accessibility, time and place flexibility compared to formal learning, it does not provide enough face-to-face interactivity between an educator and learners. The main aim of proposed project is making the learners' facial expressions as well as their answers contribute in determining which next level they should be passed to. In this document, we are proposing a hybrid information system, which combines computer vision with machine learning algorithms for visual and interactive e-learning systems. By focusing on the the student's contentment with the overall experience through his emotional states and his way of answering.The proposed system will automatically decide the level which the student should be passed to.

1 Introduction

1.1 Background

The examiners suffered from not detecting the accurate decision for transferring the student to the next level, which shows a lot of disappointment in tracking the student level. Students or learners use the e-learning systems for several reasons that make them comfort and the connection between the learner and the material is more better through the days. Our proposed project aims to collect facial expressions and answers from the learners as an input to our system and combining it together to have an automated decision to predict the next level of students.

1.2 Motivation

The world market has all its attention focused on education and how we can enhance and develop it in the best and fast way. There is a lack of researches and systems that handle the education improvement and emotions of the learner.

This motivated us to work on a system that focuses on getting a better way of getting facial expression by getting the evaluation functions of this expression and enhance it the way of learning and upgrade the learner. 43 percent of US college students found digital study technologies extremely helpful for doing homework. (Source: Statista) Back in the day, you had to go to the college library to do your homework. That's not the case anymore for a growing number of college students, as statistics in 2018 speak volumes about the growth of online education. When the library buys an eBook, any given number of students can read it simultaneously. Whether they actually will is a different matter.

1.3 Problem Definitions

There are many problems concerning education that our project tries to solve it. Such as different rules that can't be applied to all conditions like if the answers are all correct and student expression is sad so the detecting the current educational level of the learner during test exam will be hard to be detected and we try to help the education organizations to know the accurate level for each student in each course. This system aims to detects the accurate level of the learners by using both together: facial expressions and their answers.

2 Project Description

Detecting the level of education using facial expressions during the exam and predict the next level for the student. The main goal is to help students to improve their learning and levels of success in the courses.

2.1 Objective

The aim of proposed system is improving the method that detects the current educational level by using facial expressions in addition to his answers. From his answers and facial expressions, the system will detect the weak points of the student and try to help him to improve these points. We repeat the test more than one time to get an accurate level. The system will declare which level is possible and accurate for the student. Finally, our system will make some operations such as evaluation functions to predict the next level in the course for each student.

2.2 Scope

- The proposed system works to improve the level of education in test exams by using the facial expression which is detected by webcam with quick and accurate results. A facial expression such as angry, happy, sad, surprise, fear, neutral.

- The system works on three phases: the first phase student enter the course and take placement test to detect his current level. If the there is a

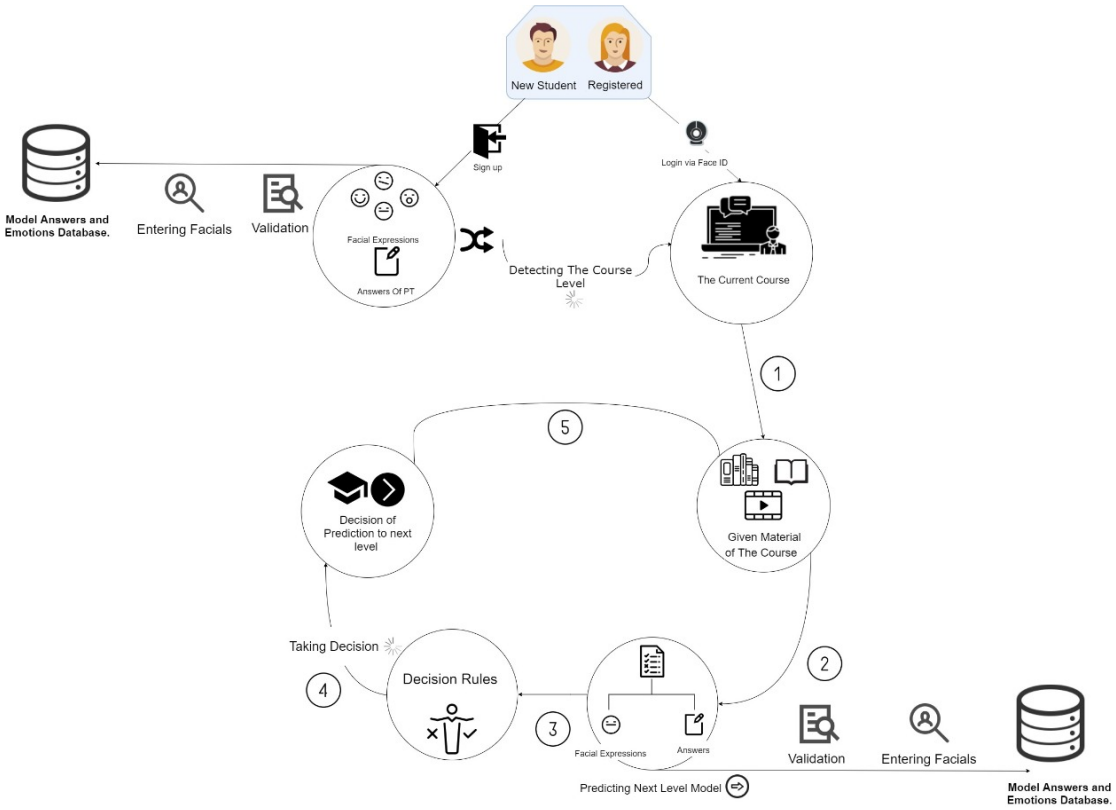
registered user that already takes place in the system so he will enter his specific level and continue his course, then after testing him many times during the period of the course and having a facial expression in each time, each time we use decision making techniques to conclude an overall facial expression to be companion with each test such as confusion matrix, analytic hierarchy process(AHP) and frame attention network. Those operations will detect the next level in his course.

- The system works based on machine learning techniques, the whole system will be offline after recording the facial expression and can not be used at dark mood.

2.3 Project Overview

Our proposed project aims to detect several types of facial expressions such as angry, happy, sad, surprise, disgust, neutral. The first step to being applied to the learner having the test online for the first time (entry-level) to detect his level for putting him in his specific place. The learner then will be tested 10 times for example during the whole period of the course, so we will have 7 facial expressions that recording through the webcam in each time he will be examined. Deep Learning Algorithm such as CNN classification applied at phase of recording facial expressions and have 2D array which has at the first array these expressions and the second array have the exams with their times, so the final output based on this array and his answers we will make some mathematical operations(decision making) such as confusion matrix [include 7 classes of facial expressions divided into rows and columns as a matrix], analytic hierarchy process[divided attributes and values into level depends on comparison and analysis] and frame attention network[divide the video into frames then these frame as a input so he can detect the facial expression as a output]. Those mathematical use to detect the next level in the course for the learner, so this phase will be offline. The result of confusion matrix and analytic hierarchy process(AHP) will be in the form of matrix. There are a lot of data-sets for facial expressions but by choosing CK+ we can do more because it shows more accurate results. They have a lot of data that can help us in choosing the right expressions in the data-set by identifying or matching them. In the first stage of explaining how our system will be working, we start with stage preprocessing: we collect the input data to our model from a recorded video by using the webcam and student's answers in the quizzes and saving these in database. In the second stage of our system is processing: firstly, By logging the learner the system by face ID, The given material of the course is qualifying the student to improve his knowledge. By testing the student and getting his facial expressions the system take all of them to take the decision using decision rules. Secondly, based on the students' answers the machine's role is to compare model answers and the answers of students. The final stage of our system is the output: which proposed results are to know the current education level of the students, after that during the course exams and quizzes our model should predict the next

level of the students and the the loop will be repeated.



3 Similar System

3.1 Similar System Information

3.1.1 Student Emotion Recognition System (SERS) for e-learning improvement based on learner concentration metric [2]

Motivation:
For enhancing the way of education there is a lot to do because education is a way of progress and success so the researchers focus on it. One of these ways to focus on the learner's face and emotions and what he feels about upgrading himself. Emotion plays an important role in analyzing the student's interest in classroom lectures. Out of the various ways to detect emotion, the quick way is to understand the emotional symptoms through facial expression. But the trends are changing from the instructor's teaching to self-pace e-learning with the introduction of new technologies.

Problem statement:
One of the main problems that face the researchers is it requires a very long training time. So there is slow progress in reaching any of classifying that emotion. All of that in the object detection algorithm called Viola-Jones for the computation speed. Efficient classifier called Ada Boost for choosing a small number of visual features from a very large set of potential features. It's sensitive to noisy data and outliers.

Contribution:
There are various face detection algorithms to extract the details of the face region. Some of the popular face detection algorithms are Viola-Jones, Local Binary Pattern (LBP), Ada Boost and Neural Network. This is the first face or object detection algorithm framed by Viola jones for solving the issue of face detection. It is projected in three significant ways namely through an integral image for the computation speed an efficient classifier called Ada Boost for choosing a small number of visual features from a very large set of potential features a process of cascade classifier for locating the required facial regions. High recognition accuracy and less false positive rate are its highlights. LBP is very effective to label the image features. LBP has advantages such as high-speed computation and assists the wide practice in the areas of image retrieval. Viola jones and LBP is considered here for the detection of the face because of its well-known characteristics.

Results:
The different concentration levels have been observed throughout the video i.e. High, medium and low. Thus after analyzing, it is concluded that at some consecutive frames the concentration level is either low/medium/high. Therefore it can be said that the student wasn't attentive in reading the content which was displayed at that time period.

Importance:
The researchers achieved the best for enhancing the way of concentration and get the result that the admin he/she needs. The quality could be achieved better based on the concentration level recognized using eye and head movement. The proposed system is efficient enough to detect the negative emotions like boredom or lack of interest of the student in e-learning environment.

3.1.2 Joint Fine-Tuning in Deep Neural Networks for Facial Expression Recognition [1]

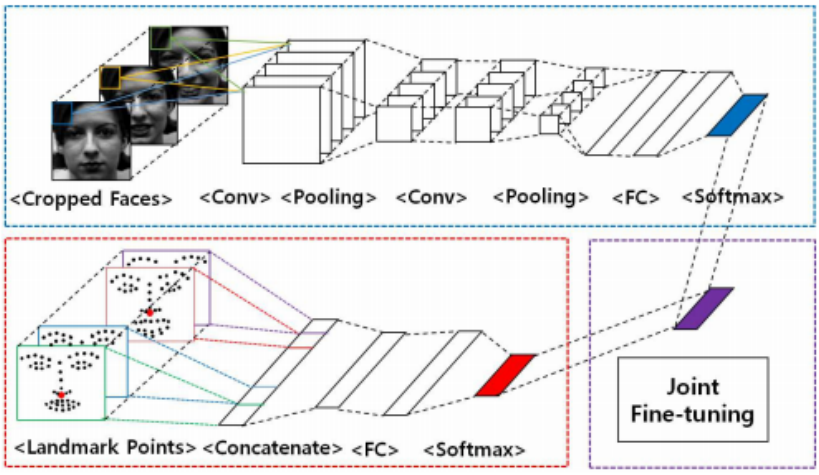
Motivation:
Recognizing emotion from a facial image is a classic problem in the field of computer vision, they found a solution that solves it easier and faster than ever. By reducing this effort, a deep learning technique, which is regarded as a tool

to automatically extract useful features from raw data. And by combining two algorithms (DTAN+DTGN=DTAGN), and by that get more accurate results for recognizing.

Problem Statement:
The algorithm was not successful enough, the researchers observed that the accuracy of fear was much lower than other emotions. In particular, most of the fear emotions were confused with surprise. The performance of deep learning techniques highly depends on the quality of training data, so our accuracy with fear was not good enough.

Contribution:

- Two deep network models are presented in order to extract useful temporal representations from two kinds of sequential data: image sequences and the trajectories of landmark points.
- We observed that the two networks automatically detect moving facial parts and action points, respectively.
- We presented a joint fine-tuning method integrating these two networks with different characteristics, and performance improvement was achieved in terms of the recognition rates.



Results:

There are a lot of results in joint fine-tuning method integrating (DTAGN), the researchers use three different databases (CK+, MMI, and Oulu-CASIA).

	Concatenation	Joint Fine-tuning
Accuracy (CK+)	94.5	**97.25**
Accuracy (Oulu)	75.63	**81.46**
Accuracy (MMI)	67.8	**70.24**

Joint fine-tuning method showed about 3 to 6 percent improvement in terms of the recognition rates than concatenation.

Importance:

The researchers achieved best recognition rates using the integrated deep network on the CK+ and Oulu-CASIA databases, show by numbers the improving in recognizing the emotions of face to get a more accurate point.

3.1.3 Formative Assessment In Higher Education: Detection And Improvement On Learning Level [6]

Motivation:
The objective of this paper to progress learning and levels of success in the subjects of Mathematics and develop a point of reflection around the instructing and learning process of some mathematical contents, Based on formative and summative assessments. The summative assessment is a way to know and calculate the benefits of the course for the student. And formative assessment is measuring the students' learning during the course.

Problem Statement:
The problem before the MatActiva project is how to improve the education level about this course and to detect the education level for each student in each subject of mathematics. After the project, the researchers decide to assess the students' knowledge using formative and summative evaluation. When they test the project, the analysis results show that the summative test is better. However, the student survey results show that the formative test is better to increase their knowledge.

Contribution:
The researchers study the purpose of each assessment and they reach that the aim of formative assessment is improving skills and extending their feedback and its duration during the course. And the aim of the summative test is measuring the level of education for each student and its duration at the end of the course.

Results:
The researchers achieve that the summative evaluation only not enough. But the use of formative test besides the summative test contribute to increasing students degree. At end of the project, the researchers take students' opin-

ions about the formative assessment and the result was positive for formative evaluation. And the researchers reach that the using summative and formative together is better than using each of them alone. The students' opinions show in this figure.

	Question	Positive opinion (%)
Settings	The FT should not have a limit on the response time	46
	The FT should allow only one attempt	13
	Make available the FT 10 days before the ST is enough	82
Formative/Summative relations	FT improve my final grade in Mathematica	100
	FT are useful to contents review	100
	FT stimulate and increase the theoretical knowledge	100
Feedback	The FT feedback by question is very useful	91
	The FT feedback by question it was enough to correct the mistakes	77
	The general FT feedback is very useful	91
Attitudes	The FT reinforce my motivation to study	93
	The FT promote my self-learning	96
	The FT contribute to identify my weakness	100
	The FT contribute to increase my self-confidence	89
	The FT change my way of learning	79

Importance:
This paper helps our proposed project to know how to improve the education level of the student. And help us to know the methods to detect and improve the education level.

3.1.4 Frame Attention Networks for Facial Expression Recognition in Videos [4]

Motivation:
the video-based facial expression recognition aims to classify a given video into several basic emotions such as happy, fear, sad, and because automatic facial expression has recently attracted increasing attention in academia and industry due to its wide range of the many applications nowadays, such as intelligent environments, multi-modal human-computing (HCI), etc.

Problem Statement:
The problem that the researchers faced in that project is through the great progress that has been made recently, facial expression recognition in the wild remains a challenging problem due to large head pose, illumination variance, occlusion, motion blur, etc.

Contribution:
The idea of the project is to have a good accuracy of many emotions based on the facial expression recognition and then calculate the probability of each emotion according to that it makes choosing a level of student on an institute

or an interviewee in a company, so the researchers are trying to solve the problem of the low accuracy and trying to get high accuracy, they decide to use and propose frame attention networks (FAN) for video-based facial expression recognition (FER). It takes a facial video with a variable number of face images as its input and to produce at the end a feature representation.

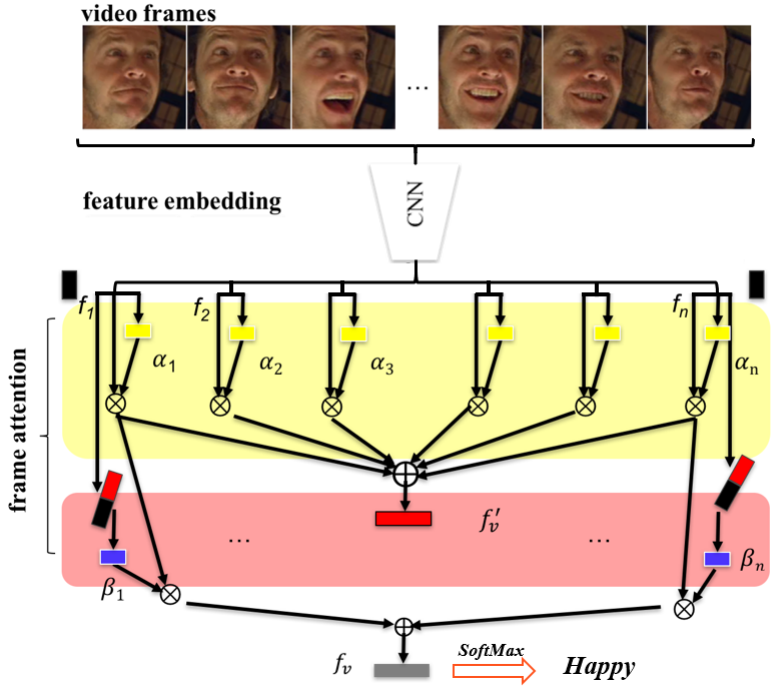


Fig.1. the proposed frame attention network architecture.

Results:
The researchers make an evaluation of the frame attention networks and they reach a result with a very high accuracy 99.69 percent also that accuracy thanks to the meta data-sets that they use as Cohen Kanade (ck+) and AFEW 8.0(Acted facial expression in the wild) to meet the real-world environment extracted from movies.

Importance:
Every paper has points of strength and points of weakness so here the most thing I see that the accuracy has reached very high and thanks to the data-sets and the algorithms they used to reach that accuracy.

3.1.5 Study of video-based facial expression and emotion recognition methods [5]

Motivation:
In real life, the facial expressions and emotions is just an interaction to external and internal events of a human being. The researchers thought that recognition of the users expression and emotions from video streaming is very important role, nowadays in many real-time applications as human-computer interaction based systems are used to immediately and accurately track the human activities from the videos, in another area recognition and tracking the human face expression and emotions from video streaming, is an objective of different purpose, such as physical fatigue.

Problem Statement:
The problem of the researchers that they faced is how to increase the accuracy and how to speed up the recognition of the emotions also they want to increase the efficiency. So the researchers decide to use some techniques to solve these problems to make it easy to identify and recognize easily.

Contribution:
The researchers proposed the system for an efficient face recognition and emotions recognition from video stream, to recognize the facial emotions in real-world natural situations, they used novel techniques called extreme sparse learning (ESL) which has the ability to jointly learn a dictionary (set of basics) and a nonlinear classification model.

Results:
The outcome of this paper is the current research gap and research challenges that are still open to solve for video-based facial detection. The researchers reached approximately accuracy of about 95 percent in facial emotion detection.

Importance:
The researchers' goal in this paper is to present a comparative study on different techniques of video-based facial and emotions recognition using different methods, those techniques and methods allow the facial recognition system to reach approximately 95 percent accuracy.

3.1.6 Real-time Personalized Facial Expression Recognition System Based on Deep Learning [3]

Motivation:
The researchers made this system to recognize human emotions based on facial expressions using webcam which can detect facial expressions from 2-3 meters distance using CNN classifier, facial expressions such as happiness, sadness, surprise, anger, disgust, neutral or any combination of those six emotions.

Problem statement:
The main problem is to detect the facial expression which is difficult to detect by a lot of details in the reactions of face, differences of illumination conditions and classification of the facial.

Contribution:
Research recommended to use personalized facial expression recognition method based on deep learning by using database facial expression databases (FED) such as CK+dataset and FER 2013 database, then training the deep network using CNN classifier based on three steps:

- General FED: There they input image in grayscale then use CNN classifier which has 3 core layers: 3 convolution layer, 3 max pooling, and 2 fully connected layers. They use 5x5 filters for each convolution layer which has activation function Rectified Linear Unit (ReLU) for preventing gradient disappearance problem. They use 3x3 filters for each max pooling layer to reduce the size of the input image to the half. For a fully connected layer, it has 1000 nodes which are a result of the aggregation max pooling layer. To regularizing the network, they use the dropout method (reducing overfitting in neural networks by preventing complex co-adaptations on training data). The top layer is Softmax which has 6 nodes which are detected 6 emotions. Despite that, the rate of this training might be low because of differences in facial expression so we go to the next step.

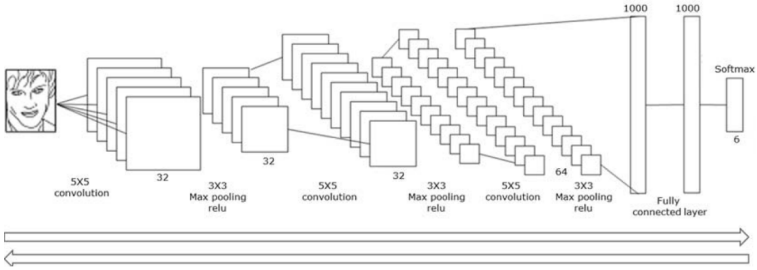


Fig. 1. The proposed CNN model for basic facial expression recognition

- Personal FED: There they registered information of the user, then obtain the facial expression by normal emotions. Then CNN will make training through personal facial expression then stored it as user's IDs.
- Personal facial expression recognition: There they detect face and recognizes the user, then get personized deep network by user id which is recognized from step 2 so they can detect the 6 facial expressions (angry, happy, sad, surprise, disgust and neutral) also can detect other expressions such as concentration, exited and bored.

Results:
The recognition results were good for 6 emotions and the accuracy achieved to

96 percent in 3m distance. The system can also detect bored emotion when neutral state at least 90 percent , the system can also detect exited emotion when happiness state more than 50 percent and the system can also detect concentration emotion from sad or surprise expression or mixed together.

Importance:
This technique is important due to its accuracy which helps us to can student monitoring the system.

3.2 Comparison with Proposed Project

Similar System	Datasets	Accuracy	Algorithms	Types of detect
Student Emotion Recognition System (SERS) for e-learning improvment	-----	89%	Viola Jones, Local Binary Pattern (LBP),Ada Boost and Neural Network	Concentration level recognized using eye and head movement.
Joint Fine-Tuning in Deep Neural Networks for Facial Expression Recognition	CK+ , Oulu-CASIA	CK+ = 97% Oulu-CASIA = 81.46%	DNN and CNN	Facial expression recognition by fast combining algorithms DTGAN
Frame Attention Networks for Facial Expression Recognition in Videos	Extend CK+ , AFEW 8.0	98%	Frame Attention Networks , CNN	Detection of expression emotions
Study of video-based facial expression and emotion recognition methods	------	95%	Novel facial motion , cylinder head model (CHM) , online statistical model (OSM)	Emotions of the user
Real-time Personalized Facial Expression Recognition System Based on Deep Learning	CK+ , FER 2013	96%	CNN	The personalized deep network using the recognized user ID from the personalized FED. And it recognized the facial expression. It can recognize 6 basic facial expressions such as happiness, sadness, etc.
Detection education level using facial expressions	CK+ , AFEW 8.0	------	Deep learning techniques such as CNN	Detect the educational level and the next level of the student based on facial expressions and his answers.

4 Project Management and Deliverables

4.1 Tasks and Time Plan

Task	Starting	Ending
Proposal evaluation	7/10/2019	10/10/2018
SRS evaluation	15/12/2019	15/12/2019
SSD Evolution	19/2/2020	19/2/2020
Evaluation implantation	After spring break	After spring break
Technical evaluation	7/5/2020	7/5/2020
Final Thesis	20/6/2020	20/6/2020
Cermoney	24/6/2020	24/6/2020

References

[1] Heechul Jung et al. "Joint fine-tuning in deep neural networks for facial expression recognition". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 2983–2991.

[2] LB Krithika. "Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric". In: *Procedia Computer Science* 85 (2016), pp. 767–776.

[3] Injae Lee et al. "Real-time personalized facial expression recognition system based on deep learning". In: *2016 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE. 2016, pp. 267–268.

[4] Debin Meng et al. "frame attention networks for facial expression recognition in videos". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2019, pp. 3866–3870.

[5] Husam Salih and Lalit Kulkarni. "Study of video based facial expression and emotions recognition methods". In: *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. IEEE. 2017, pp. 692–696.

[6] Babo, Lurdes and Azevedo, Jose Manuel and Torres, Cristina and Lopes, Ana Lopes. "Formative Assessment In Higher Education: Detection And Improvement On Learning Level." ResearchGate, ResearchGate GmbH, March 2017