Software Engineering Department

ORT Braude College

Course 61766: Extended Project in Software Engineering

Lie Detection Using Convolutional Neural Network

In Partial Fulfillment of the Requirements for

Final Project in Software Engineering (Course 61401)

Karmiel – June 2018

Elias Nijim 206160624

Vladimir Balagula 323792770

Supervisor:

Prof. Zeev Volkovich

Contents

***Abstract**: There are a lot of studies that are trying to reveal the nature of human lie and connect it to physical and Behavioral Sciences. In recent years, a new method has been found useful for that purpose. This method based on non-invasive physiology sensing such as voice pitch variation analyzed by deep learning model of Convolutional Neural Network. This study presents methods that we used to build voice lie detector. Those methods include MFCC algorithm as preparation input of Convolution Layer. The second step defines the architecture of the neural network and the origin of it. The study also reveals teaching mode for a network with backpropagation algorithm. We explain how convolution works combined with MAXPOOL Layer, and which layers our model contains. Also we provide all the documentation data necessary for the implementation of the program.*

We expect that our Convolutional Neural Network will find different patterns using multiple filters and will predict lie with high accuracy.

***Keywords:** Convolutional Neural Networks, MFCC, Voice Lie Detector, Back Propagation.*

1. INTRODUCTION

Nowadays there are various of studies that are working to explain the nature of lying and trying to attach it to physical and Behavioral Sciences. Those studies use a lot of different techniques but, none of them reach the 100 percent of accuracy.

The best technique available today is the lie detector (Polygraph), which was invented in the early 19th century and has the accuracy of 90 percent in exposing lies [1]. Since the Polygraph measures physiological signals as blood pressure, pulse and sweat a person can be trained to lie without all of these symptoms. [2].

In recent years, a new method has been developed to identify lies. This method based on non-invasive physiology sensing such as voice pitch variation and heart rate. All the data were analyzed using a neural network. The researchers were able to bring about 82% accuracy of identifying high-stress situations and 71% accuracy of lying detection, which is lower than the Polygraph. [3]

In our project, we present a new approach to identify a lie. We use a convolutional neural network (CNN); this method gives the option to analyses the voice deeper. We hope that this method will give the best result and the CNN will find the dependency between human voice and lie.

2. BACKGROUND

2.1. Sound

Sound is a type of energy that travels in waves through a medium as air or water. The waves are generated by vibration source which pushes particles of air together and creates a pressure wave. Sound waves move out from the vibration source to all directions. This type of wave is referred to as a longitudinal wave. Though, there are two types of waves (Longitudinal and transverse), gas (the air) can sustain only longitudinal waves because transverse waves require a shear force to maintain them, in this case light is an example of a transverse wave, while sound is a longitudinal wave. When the sound wave reaches the ear, it transforms from longitudinal wave into an electrical impulses which are sent into the brain

There are two main elements which characterize vibration and the way it sounds - the amplitude and the frequency.

Amplitude is the size of the vibration, in other words, the height of the wave determines how loud the sound will be. Amplitude is important when balancing and controlling the loudness of sounds, a larger amplitude will produce a louder sound, and moderate amplitude will perform quite tones.

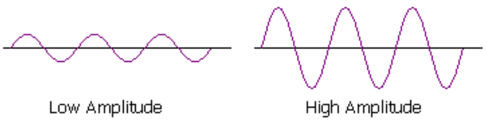

Figure 1 -Low and High Amplitude

Frequency is the speed of the vibration, and this determines the pitch of the sound and determined by the wavelength. Frequency is measured as the number of wave cycles that occur in one second. A cycle is a repeating pattern in given amount of time – a period.

The unit of frequency measurement is Hertz (Hz for short). (The hertz unit is named after Heinrich Hertz, a famous 19th century physicist.). [4]

Pitch of the sound is also determined by the frequency of the wave but from different viewpoints. While frequency measures the cycle rate of the physical waveform, pitch is how high or low it sounds when you hear it. High pitch is generated by high frequency wave and low-pitch generated by low frequency wave.

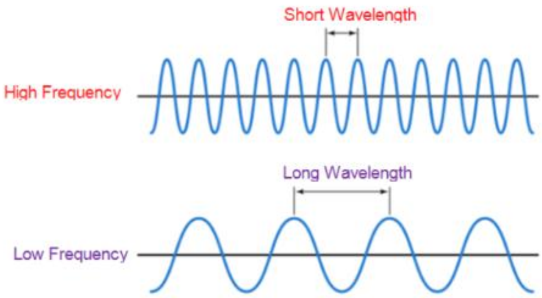

Figure 2 - High/Low Frequency Wave
http://www.techplayon.com/wavelength-frequency-amplitude-phase-defining-waves/

2.2. Analog Signal

To process and represent sound wave into a digital signal in the computer it must be converted into an analog signal by using one of the analogs to digital converter.

Analog signal is a continuous wave that changes over time; it can be used to measure different types of waves such as sound and temperature. One kind of analog signal is "sin wave" which cannot be broken. The most common way to convert an analog signal into a digital signal is via a microphone.

2.3. Digital Signal

Digital signal is a non-continuous signal that was converted into a pattern of bits to represent data as a sequence of discrete values. Digital signal input can be represented in Time-Frequency Domain graph where the Y-axis describes the amplitude and the X-axis describes the time (Figure 3). [5]

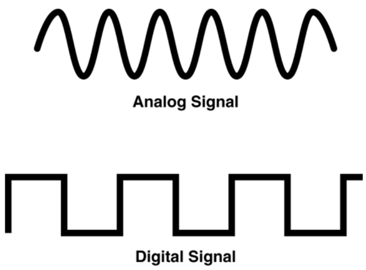

Figure 3 - Analog Signal compares to Digital Signal
https://www.allaboutcircuits.com/technical-articles/an-introduction-to-digital-signal-processing/

To do a deeper analysis of the time domain signal we can use the frequency domain. The frequency domain signal can be analyzed with the use of spectrum analyzer. In addition, time domain signal can be transformed to frequency domain signal by calculating Fast Fourier Transform (FFT).

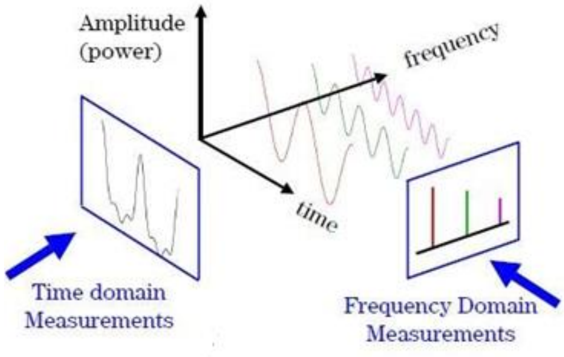

Figure 4 - depicts time domain and frequency domain measurements
http://www.test-and-measurement-world.com/Measurements/Time-domain-measurements-vs-Frequency-domain-measurements.html

2.4. MFCC

Mel Frequency Cepstral Coefficient (MFCC) is commonly used in the speech recognition process and in the analyzing speech signal. The pre-step in the speech recognition system is to identify the components of the signal [6].

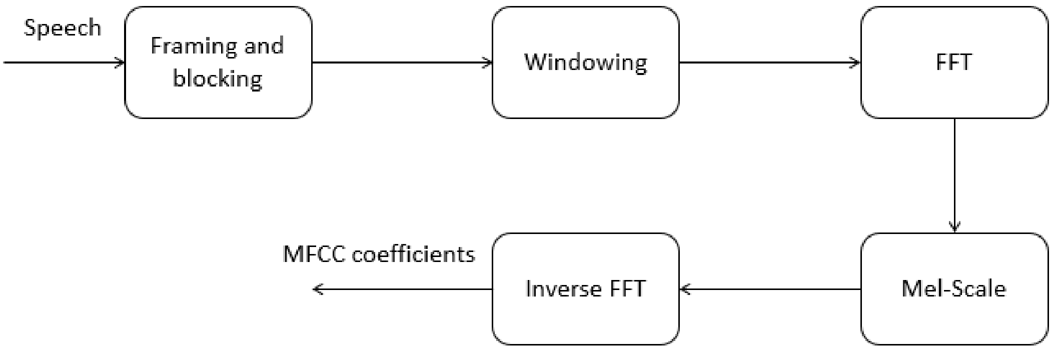

Figure 5 - MFCC Processing

2.4.1. Framing and Blocking

The first step of MFCC procedure is to divide the signal into continuous smaller frames by dividing the speech signal to “X” signal frames. To reach the best results, we must find the best divider number that contains enough information, many researchers found that the common divider number that divides the speech signal to the adequate number of frames is 256 frames, this number of frames is enough to include all the data which gives the most accurate result, increasing or decreasing the divider number can manipulate with the data amount in the frames which can lead to reducing the results accuracy.

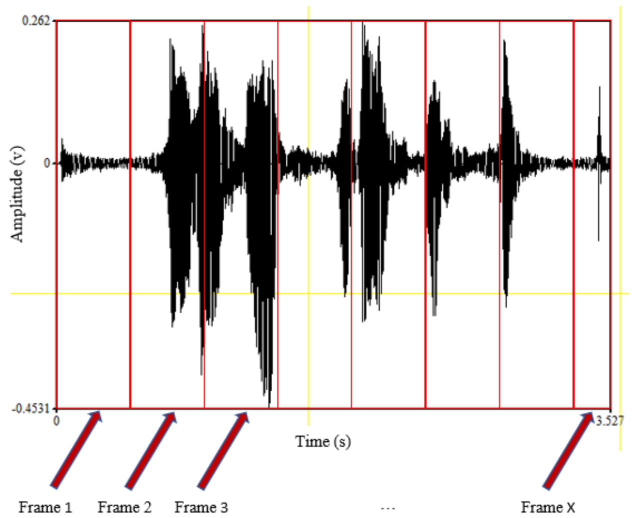

Figure 6 - Framing

2.4.2. Windowing

The purpose of windowing is minimizing the disruptions at the beginning and at the end of a certain frame and its function is multiplied as well as the window is one, the multiplication result of the signal in window gives a signal sample, The appropriate window function that brings the best results is Hamming Window (Equation1).

$$\omega(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)$$

Equation 1 - Hamming Window

N = number of blocks in each signal frame.

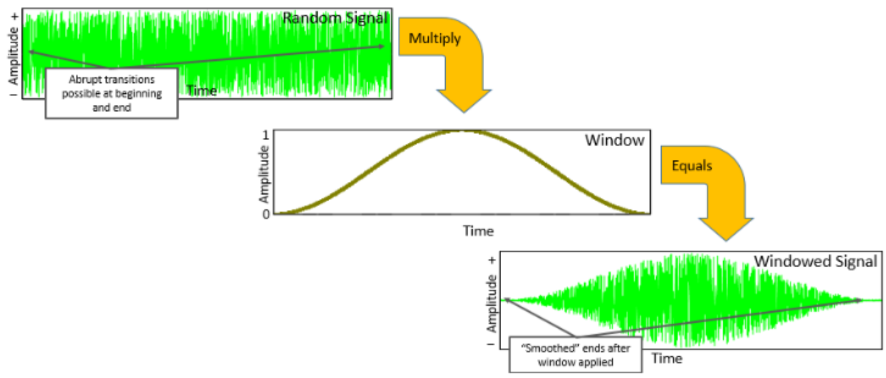

Figure 7 - Hamming window
https://community.plm.automation.siemens.com/t5/Testing-Knowledge-Base/Window-Types-Hanning-Flattop-Uniform-Tukey-and-Exponential/ta-p/445063

2.4.3. Fast Fourier Transform – FFT

Fast Fourier Transform is a rapid computation algorithm to convert time domain signal wave samples to frequency domain spectrum samples (Equation 2).

$$X_k = \sum_{n=0}^{N-1} x_n e^{\frac{-i2\pi kn}{N}}$$

Equation 2 - FFT formula

X_k - Frequency domain sample x_n – Time domain sample N – FFT size K = {0, 1, 2…N-1}

These transformations provides the location of the speech wave signal in time domain graph by breaking the complicated signal into small sin waves and express it as picks on another graph.

2.4.4. Mel Scale

The linear scale is not pursued by the human perception of frequency contents of sounds for speech signal. The Mel scale frequency is a linear frequency and a scale of pitches, it is the way of spacing the filter and calculating how much wider it should be, consequently when the frequency gets higher the filters get wider as well. A popular formula to convert hertz into Mel (Equation 3)

$$m = 2595\log_{10}(1 + \frac{f}{700})$$

Equation 3 - Hertz to Mel formula

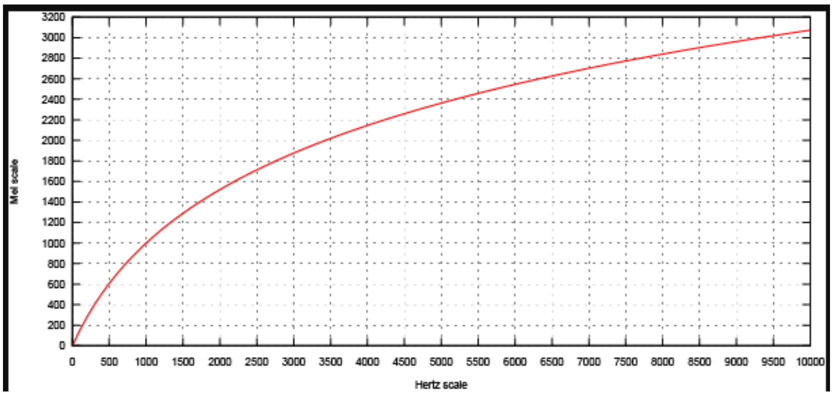

Figure 8 - Mel and Hertz scale
https://cpb-us-w2.wpmucdn.com/u.osu.edu/dist/3/41388/files/2015/02/Mel-Hz_plot-2juu180.png

Mel spectrum is generated from result of the log calculation on the energies at each block. The popular usage of Mel scale is to calculate coefficients

2.4.5. Inverse Fast Fourier Transform – IFFT

Inverse fast Fourier transform is a fast algorithm to perform backward of Fourier transform, which revokes the process of fast Fourier transform. IFFT converts frequency domain wave signal into its original time domain wave signal using IIFT formula (Equation 4)

$$X_n = \frac{1}{N} \sum_{n=0}^{N-1} x_k e^{\frac{-i2\pi kn}{N}}$$

Equation 4 - IFFT formula

X_k - Frequency domain sample x_n – Time domain sample N – FFT size K ={0,1,2...N-1}.

2.5. Neural Network

2.5.1. Biological Neural Network

The neuronal system is a group of nerve cells that divide into a central nervous system (the spine and brain) and peripheral (which is divided into a motor and a sensor). Neurons are nerve cells responsible for passing the information along the human body; each neuron is composed of Dendrites and one Oxon.

The Dendrites are branches on the cell body that initiate information transfer by intake chemical signal from outside the cell and convert them to an electrical signal which passes through the Oxon. At the end of the Oxon, there are neurotransmitter molecules that packed in vesicles, waiting for the electrical process reaching the end of the axon. When it does, there is an increase in the concentration of intracellular calcium, which increases the movement of the full neurotransmitters to the exocytosis (the exit of the chemicals out of the cell). The strength of the signal works according to the principle of "the more neurotransmitters there are, the signal is more powerful". The external neurotransmitters are injected into the synaptic space and are absorbed by specific receptors on the membrane of the next neuron or the target cell to pass on the information. [7]

2.5.2. Artificial Neural Network

Artificial Neural network – is a mathematical model also known as a neural network which based on biological neuron system and is part of artificial intelligence discipline. The main idea of artificial neuron is taken from the biological neuron. The artificial network has same properties as the biological system including modifying signals by weights at the receiving synapse, summing the weighted inputs and transmitting outputs to other neurons.

The main advantage of the artificial network is the ability to improve itself by learning from the mistakes. The learning phase is inserting data with mapped output, which gives the system options of finding relation coefficients between neurons inputs. [7]

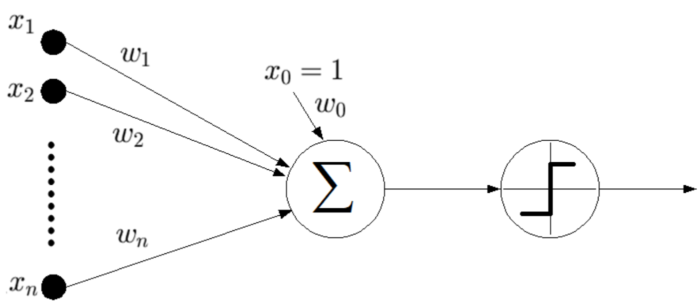

Figure 9 - Diagram of an Artificial Neuron
https://www.fungglobalretailtech.com/research/introduction-deep-learning-neural-networks/

$$v_j = \sum_{i=1}^{m} w_{ji}^{T} y_i$$

Equation 5 - Neuron input

$$y_j = f(v_j)$$

Equation 6 - The neuron output

The w_i on the diagram is a weight for each input for each neuron. All the weights are summarized and converted to a specific output using activation function. The whole network contains a big number of neurons which communicate one with the other by sending and receiving outputs. A large number of neurons can solve complex tasks.

2.5.3. Back-Propagation

In order to complete the goal tasks of the neural network which includes processing the input from convolutional layer into an accurate output, the network must adjust weights on each node. To achieve this, in our project we use backpropagation algorithm based on gradient descent. The algorithm calculates the error between produced output and desired output and corrects it to increase the accuracy. [8]

The mathematics stands at the base of this algorithm take in a matter of relative errors (Equation 7) and instantaneous error energy of neuron (Equation 8). Where "d" is the expected result and "y" is neural network result

$$e_i(n) = d_i - y_i$$

Equation 7 - Relative Error

$$E_i(n) = \frac{1}{2} e^2(n)$$

Equation 8 - Error Energy of Neuron

Partial derivation of the above by using chain rules gives (Equation 9)

$$\frac{dE(n)}{dw_{ij}} = \frac{dE(n)}{de_j(n)} \frac{de_j(n)}{dy_j(n)} \frac{dy_j(n)}{dv_j(n)} \frac{dv_j(n)}{dw_{ji}(n)}$$

Equation9 - Pertial Deviation

Positioning derivations of Equation (8) by e(n), derivation of Equation (7) by y(n), derivation of Equation (6) by v(n) and derivation of Equation (5) by w(n) in the Equation (8) will gives Equation (10)

$$\frac{dE(n)}{dw_{ij}} = -e_j(n) f_j'\left(v_j(n)\right) y(n)$$

Equation 10 - Derivation Error Energy by weight

Correction of the weights defined by delta rule (formula 7) where "η" is a learning rate.

$$\Delta w_{ij}(n) = -\eta \frac{dE(n)}{dw_{ij}}$$

Equation 11 - Weight correction by delta rule

The negative value of η indicates a direction for gradient descent to minimize the error. Using the equations (10) and (11) we define that delta rule depends on learning rate, local gradient δ and y_j (Equation 12).

$$\Delta w_{ij} = \eta \delta y_j(n)$$

Equation 12 Weight correction

2.5.4. Convolution

The main idea of the convolutional layer is to derive meaningful information using pattern filters on the input.

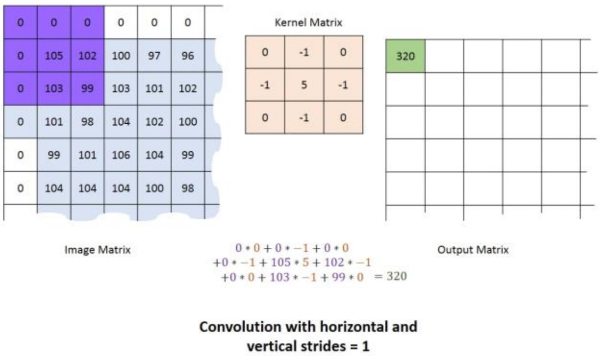

Figure 10 - Filter Scanning
http://machinelearninguru.com/computer_vision/basics/convolution/convolution_layer.html

The input to the Convolution layer is two-dimension matrix. The convolution use different filters to extract various features. The filter is a smaller matrix which moves through the image, multiply each pixel value with the filter on it and summarize it. The result is smaller image with the new weight.

In the polling level (MAXPOOL layer), new matrix is built with maximum values from every filtered cell. The process is repeated multiple times depending on the user's definition. [9]

3. ARCHITECTURE

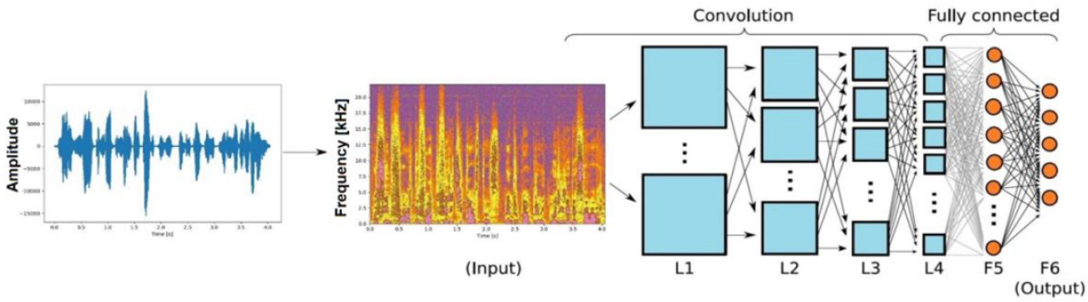

Figure 11 - Convolutional Neural Network

In this project, we chose free software PRAAT to capture the voice in wav format. The next step, MFCC obtained from the multi-spectrogram as a table of real values. The MFCC coefficients are the convolution layer input. We choose the middle 400 lines from 12-real values in a .csv. The data for the network is a list of arrays 400x12 normalized values, which can be analyzed as an image.

We adopt the architecture from the emotion identification using CNN paper [10] to solve our problem. This CNN has 200 convolutional filters 5x5 with RELU activation function because we don't know which filters is better to choose for our CNN model we learn which filters give the best result and pick them. One of the benefits of RELU (Equation 13) function is fast calculation and thus has much better performance than other activation functions. The convolutional layer has been repeated twenty times to get more detailed data. After it, coming Flatten layer which converts two dimensional output to one dimensional input to fully connected layer.

$$f(x) = max(0, x)$$

Equation 13 - ReLu function

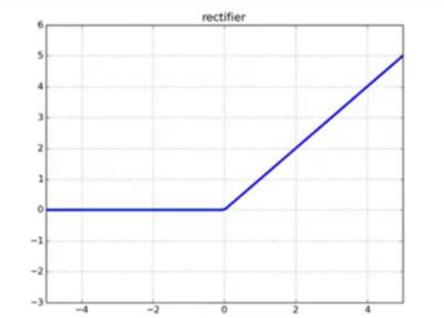

Figure 12- ReLu function graph
https://towardsdatascience.com/activation-functions-in-neural-networks-58115cda9c96

The CNN has 400x12 neurons as input. In the final stage, there is a fully connected layer with 1000 neurons, followed by lie detection classifier. Our CNN implements using Python and Tensor-Flow as the back-end with KERAS library. KERAS is a library created especially for "deep learning" for research and developing. Tensor-Flow is an open source library for dataflow programming and computation. The final layer using SOFTMAX activation function (Equation 14) to display two possible outputs: the first option gives the percentage of lying, and the second output is the percentage of truth.

$$\mathrm{out}_j = \frac{e^{x_j^T w_j}}{\sum_{i=1}^{k} e^{x_i^T w_i}}$$

14 - *SoftMax function*

1. Convolutional layer
2. MAXPOOL layer
3. Flatten layer
4. Fully connected layer with rectifier activation function (RELU)
5. Fully connected layer with SOFTMAX activation function

4. EXPECTED RESULTS

We expect that our Convolutional Neural Network will find different patterns using multiple filters and will predict lie with high accuracy. Our program will give the option to teach models using specific language due to suspect that patterns can be depend on language. Hopefully, this program will be useful for validating the person lie or changing the meaning of the words by different pronounce. The main advantage of this approach is that there is no need for special sensors only microphone to record person voice.

5. SOFTWARE ENGINEERING DOCUMENTATION.

5.1. Use Case

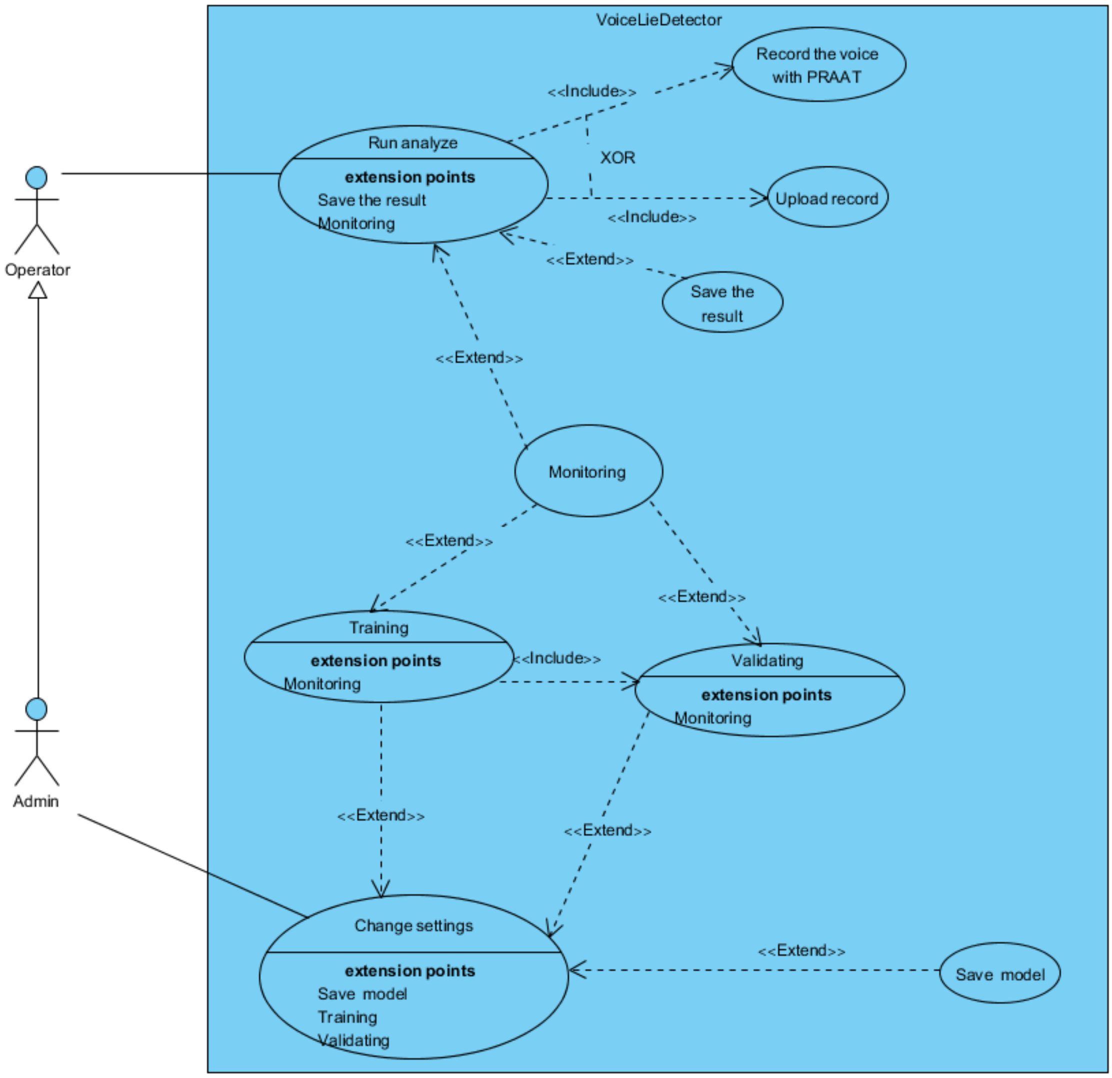

Figure 13- Lie Detector use case diagram

UC1: Run analyze

- Goal – run the model on a recorded voice to get lie probability.
- Preconditions – learning model and voice file exist.
- Possible errors – microphone is off.
- Limitations – voice record length.
- Pseudo code:

Actor	System
Click on "Run analyze" button	Open "Analyze" window
Choose data source record or voice file option	
Click on "Detect lie"	If voice file selected Display choose file from file explorer. Else Open PRAAT program
	Insert voice file to the CNN

Actor	System
	Display an answer of the prediction model
Click on "Save result"	

Table 1- Run analyze use case

UC2: Change settings

- Goal – train, set up and validate the CNN.
- Preconditions – administrator log in.
- Possible errors – data set is corrupted.
- Limitation – processing time.
- Pseudo code:

Actor	System
Click on "Settings"	Display textbox for password and user
Insert credentials	If credentials are correct Display settings window
Set up learning rate, window length, window step, filter amount, FFT size	
Set data source path	
Set training and validation ratio	
Click start "Teach & Validate"	Start training
	Display correct prediction
	Finish training display complete message and starting to validate the model
	Display statistical data during validation
	Display message validation complete
Save model	Save the machine learning configuration

Table 2 - Change settings Use Case

5.2. Class Diagram

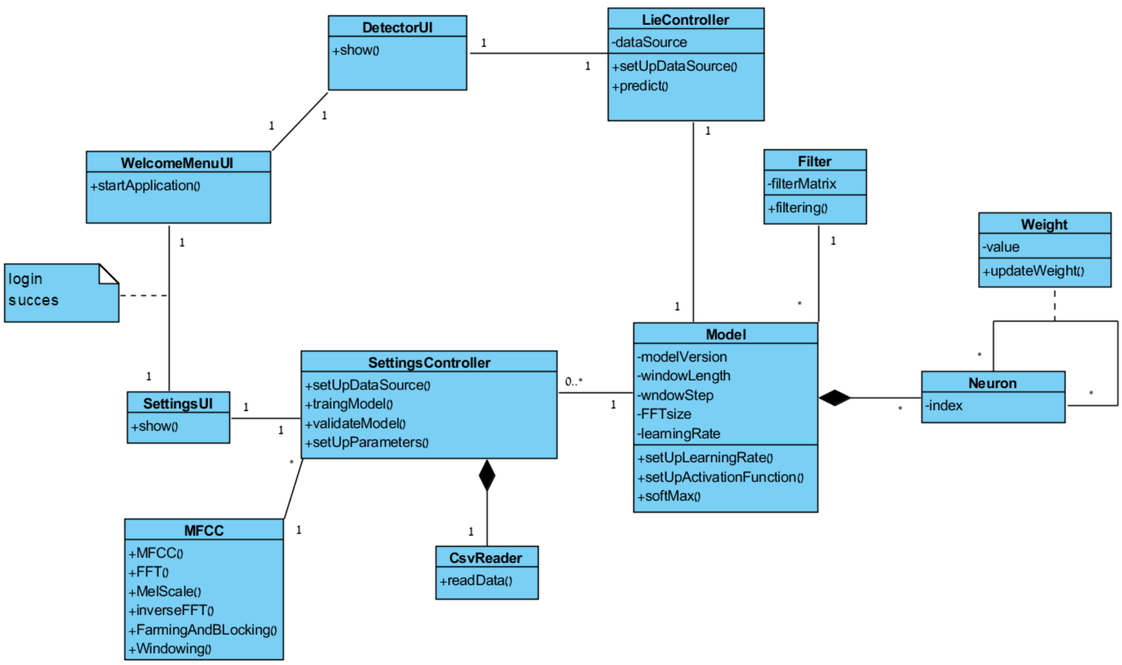

Figure 14 - Class Diagram

5.3. User Interface

5.3.1. Windows

Welcome window: the first displayed window once the program starts, gives the user opportunity to configure the model or to detect lie using existing model.

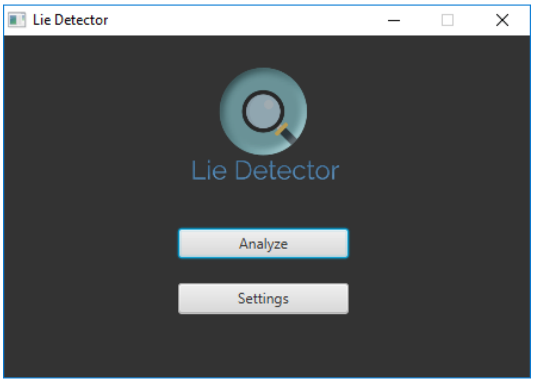

Figure 15 – Welcome window

Login: once the user clicks on settings, the program asks for admin permission.

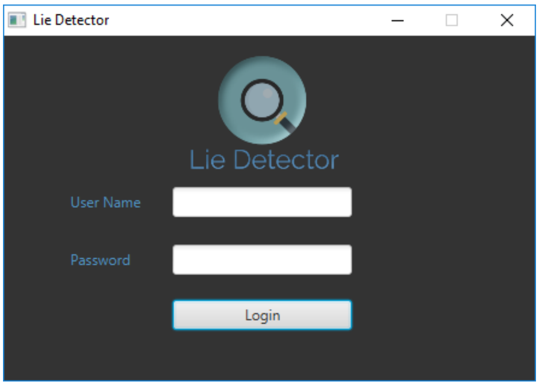

Figure 16 – Login Window

Settings: Admin is capable to configure the model by setting new parameters and giving different data sources, in addition, the admin can monitor the model process.

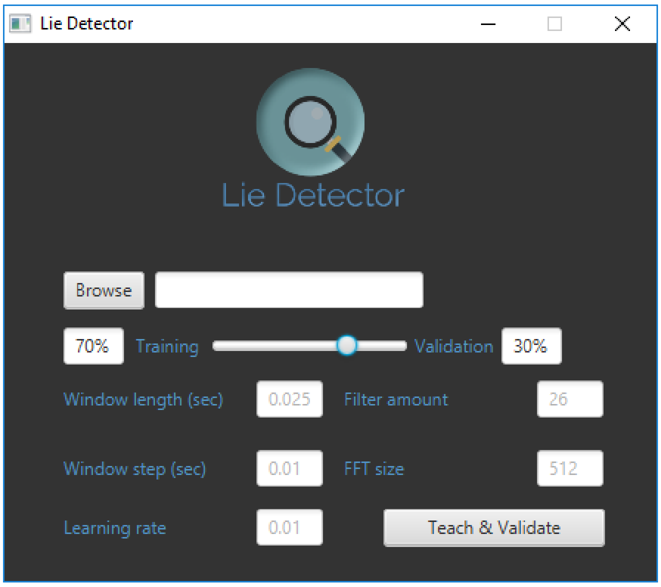

Figure 17 - Settings Window

After finishing learning the system asks for saving the model.

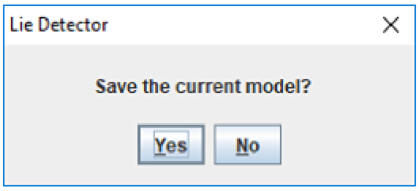

Figure 18 - Save Model Window

Detect Lie: User can choose to upload local .wav, .csv file or to record voice with external program called PRAAT

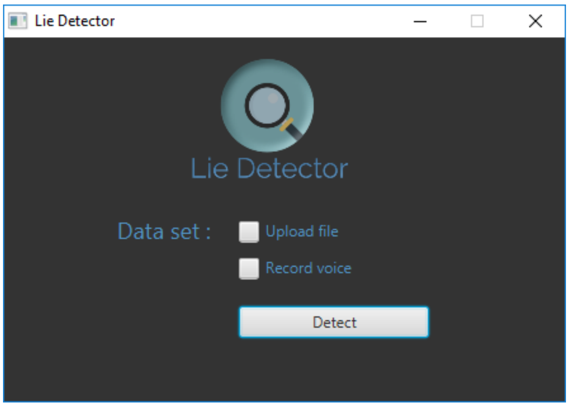

Figure 19 - Prediction Window

Monitor: User can monitor the results through this window using graphs.

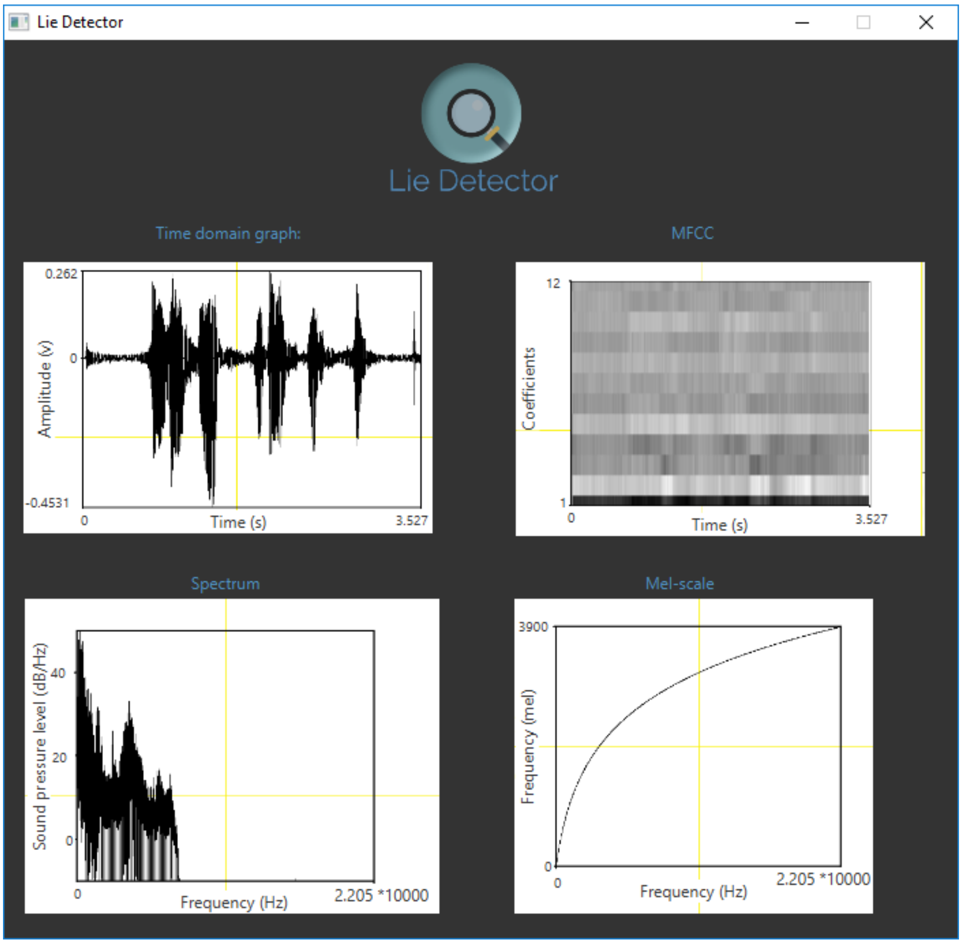

Figure 20 - Monitoring Window

5.3.2. Errors

If file not in .wav or .csv format:

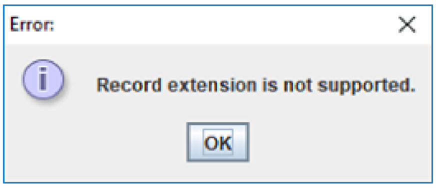

Figure 21 - Wrong Format Error Bar

If the user doesn't choose the file:

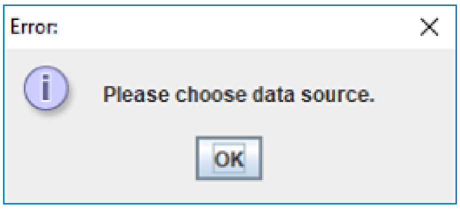

Figure 22 - No Data source Selected Error Bar

If user insert wrong credentials

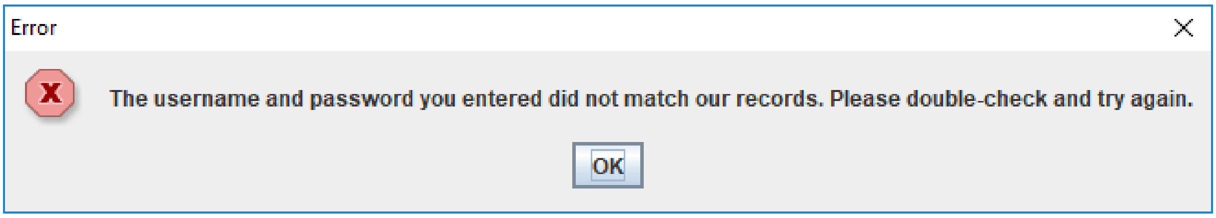

Figure 23 - Wrong Credentials Error Bar

If file that user chooses doesn't exist:

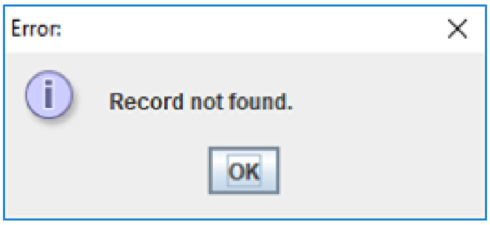

Figure 24 - Record Not Found error Bar

5.4. Testing Plan

Test name	Description	Expected result	Actual Result	Comments
		Math MFCC		
FFT calculation	Execute function using the same data in MATLAB FFT and in the project	The same result with small error	Equals	
Mel-Scale calculation	Execute function using the same data in MATLAB Mel-Scale and in the project	The same result	Equals	Input from FFT function

Test name	Description	Expected result	Actual Result	Comments
Inverse FFT calculation	Execute function using the same data in MATLAB inverse FFT and in the project	The same result with small error	Equals	Input from Mel-Scale
FFT calculation	Run twice the FFT function with different parameters each time	Different values	Not Equals	
Inverse FFT calculation	Run twice the Inverse FFT function with different parameters each time	Different values	Not Equals	Input the same data from Mel-Scale
External application PRAAT				
Open PRAAT	1)Choose data source as voice record 2)Press detect lie	PRAAT opens	Call to external program, window opens	External application event
Close PRAAT	1)Click "Close PRAAT"	PRAAT closed but project still running	Close external program and return to main program	
Save wav file	1) Choose data source as voice record 2)Press detect lie from the project 3)Press Start record 4)Press finish record 5)Check the wav file	File has been saved	File stored locally	
Check voice record	1) Choose data source as voice record 2)Press detect lie from the project 3)Press Start record 4)Press finish record 5)Check the wav file	Hearing human voice from the wav	Wav played	Possible error microphone doesn't work, microphone not selected or PRAAT not work correctly
GUI				
Go to Settings	Click on settings	Asking for password	Pass	
Login	Insert credential and press "Login"	Open settings windows	Pass	
Add file to predict	1)Choose voice file option 2)Press predict	File explorer opens to give an option to choose file	Pass	
Predict without data source choose	Click on "Detect"	Asking to choose data source before	Pass	

Test name	Description	Expected result	Actual Result	Comments
Display monitoring at Learning phase	1)Choose data source to use for training model 2)Run "Teach & Validate" 3)Open monitoring	Opens Monitoring window with success/fail prediction and error value of each epoch	Pass	
Display save option on settings	When finished validating asking for store the model	Pop-up window asking for storing model	Model stored	while the system finished learning phase.
Display save option on prediction	When finished predicting	Pop-up window asking for storing data	Data stored on Disk	

Table 3 – Testing Plan

6. BIBLIOGRAPHY

[1] P. Trovillo, "History of Lie Detection," *Journal of Criminal Law and Criminology,* pp. 848-881, 1939.

[2] R. Warner, "What's the Most Accurate Lie Detector? It's Not What You Think," *Huufpost,* 2017.

[3] N. Srivastava and S. Dubey, "LIE DETECTION SYSTEM USING ARTIFICIAL NEURAL NETWORK," *Journal of Global Research in Computer Science,* pp. 9-13, 2014.

[4] B. Hollis, "Physics of Sound," *The Method Behind The Music,* 2004.

[5] D. Krambeck, "An Introduction to Digital Signal Processing," *All About Circuits,* 2015.

[6] S. Gupta, J. Jafar and A. Fatimah, "FEATURE EXTRACTION USING MFCC," *Signal & Image Processing : An International Journal (SIPIJ),* vol. 4, no. 4, pp. 102-104, 2013.

[7] S. Haykin, Neural Network and Learning Machines, 3rd ed., 2008.

[8] L. Fausset, Fundamentals of Neural Network, 1994.

[9] D. Gupta, "Architecture of Convolutional Neural Networks (CNNs) demystified," *Analytics Vidhya,* 2017.

[10] E. Franti, I. Ispas, V. Dragomir, M. Dascalut, E. Zoltan and I. Stoica, "Voice Based Emotion Recognition with Convolutional Neural Networks for Companion Robots," *The CNN designing and training,* pp. 11-12, 2017.