New Methods on Image Analysis, Two Applications: An Augmented Reality Application for Heartbeat Sound Analysis And A MRI Brain Injury Image Analysis

L J Lechuga¹, a, T Yamamoto², b, O López³, c and G Lechuga⁴, d
¹Japanese International Cooperation Agency (JICA) trainee program at the Kanazawa Institute of Technology 7-1 Ohgigaoka Nonoichi Ishikawa / 921 - 8501, Japan
²Human interaction laboratory Kanazawa Institute of Technology 7-1 Ohgigaoka Nonoichi Ishikawa / 921 - 8501, Japan
³Industrial Engineering Department ITESM calle del Puente 222, Ejidos de Huipulco 13480 CDMX, Mexico
⁴Computer Vision Detection ENGIE 1, place Samuel de Champlain 92400 Courbevoie- France

ajulian.lechuga305@gmail.com; btyama@neptune.kanazawa-it.ac.jp; collopez@tec.mx; dgisela.lechuga@gmail.com

Abstract. Traditional statistical methods have become insufficient when applied to image analysis. The increasing size of data volume and its complexity demands new statistical approaches and algorithms. Current methods imply losing intrinsic data structures, for example when data comes from multiway arrays. In this work we concentrate in two applications i) A pre-diagnostic smartphone application for detection of cardiovascular abnormalities through the analysis of heartbeat sounds and the use of augmented reality for displaying valuable information to the end user in an immersive experience. Using the latest augmented reality smartphone applications, a digital stethoscope, heartbeat audios and classification using neural networks, we measure a user's heartbeat and output in real time, a pre-diagnostic of their current cardiovascular health. ii) A study concerning comatose patients, based on a Diffusion Tensor Image Magnetic Resonance Imaging (MRI) dataset that predicts long-term outcome for patients having suffered a brain traumatic injury. MRI images were obtained from 104 comatose patients, 65 with positive outcome and 39 with negative outcome, 39 controls were used. The fact that each volumetric image led into a 143x255726x4 tensor input, is used to briefly explain how new multiway methods could be useful in image analysis methods.

1. Introduction
According to data from the WHO (World Health Organization) 17.9 million people die each year from cardiovascular diseases (CVD), representing 31% of all deaths worldwide and more than 75% of these occur in developing countries. Most of these health issues are caused by smoking, alcohol intake, bad eating habits, lack of physical activity and in general no prevention and diagnostic measures from the patient.

Traumatic brain injury is one of the leading causes of death and disability in the industrialized world, generally requiring resuscitation in an intensive care unit and prolonged rehabilitation [1]. It is estimated that a head injury occurs every seven seconds and hospital emergency rooms treat 1 million people for
brain injuries every year. Traumatic brain injury may occur at any age, but the peak incidence is among people between the ages of 15 and 24. Men are affected three to four times more often than women. Motor vehicle accidents are the leading cause, accounting for approximately 50 percent of all cases. Falls produce the most brain injuries in people older than 60 and younger than 5 [2].

In both cases traditional procedures of prevention or diagnosis, have been improved by innovative technologies and new statistical tools and methods linked to novel technologies.

In order to solve these problematics, in this work we present our development of a pre-diagnostic smartphone application for detection of cardiovascular diseases through the analysis of heartbeat sounds and the use of augmented reality technology for displaying valuable information to the end user in an immersive experience. Using augmented reality allows the creation of powerful immersive experiences for the smartphone handler. In this way we can give relevant information to the user in a simple intuitive way, using the latest technology available in virtual experiences.

Using the latest toolkit for augmented reality smartphone applications, a digital stethoscope, labelled databases of heartbeat audios and the implementation of neural networks, we measure a user's heartbeat and output, in real time, a pre-diagnostic of their current cardiovascular health.

On the other hand the most used validated scoring system to assess the outcome of traumatic brain injury is the International Mission for Prognosis and Analysis of Clinical Trials score, which is based in a multivariate model that combines clinical, biochemical and computed tomography variables to provide a probabilistic estimate of the outcome at 6 months, but not for long-term assessment. In this is way it is important to obtain additional information about the current state of comatose patients, in order to obtain a more thorough evaluation [3]. A relevant application of medical analysis is to be able to establish accurate predictions from 3-D images datasets. In the scope of this work the objective is to present an example for the classification of multi-way brain MRI in order to predict the recovery of patients that suffered from traumatic brain injury.

2. Augmented reality application for heartbeat sound analysis

2.1 Overview
During the last decade the development of technology has seen an exponential interest in two areas of engineering and science: artificial intelligence and augmented reality. These technologies have been applied, and are constantly developed, in different areas, such as: industry, health, retail, food industry or videogames. The development and ease of use has made these technologies available to anyone and any industry that can profit from applying them. By combining these two growing areas of computer science and a low-cost implementation with smartphone hardware the possibilities for applications are endless. Augmented reality allows for immersive and shared virtual experiences while machine learning, represents the state of the art in learning models for the analysis of static data patterns. The interaction between the user, media and technology will generate new developments that allow for experiences that transcend the physical connection between a human and the computer.

2.2 Technology
For the development of the AR application we decided to use Apple’s ARKit tool since it allows to create very complex application with a fairly easy implementation of code. Another option was to use Android’s ARCore but unfortunately the detection features and overall functionality were not as powerful as the toolkit from Apple. The code was written using Xcode and Swift which is Apple’s native programming language. The final application was tested using an iPhone 8 with the latest version of iOS 12.

The classification tasks were implemented using three approaches: multilayer, deep and convolutional neural networks. In order to model, train and test each algorithm we used Google’s open-source library TensorFlow. This library has a different set of tools to develop different machine learning models and deploy robust applications in a small amount of time.
Our general programming language for the project was Python. This was used for the TensorFlow library and the pre-processing of the audio signals, transformation into numerical data and further analysis. We also made use of Django web framework to develop a small API for data transmission.

Finally, for the acquisition of audio signals from the user we used a 3M Littman Electronic Stethoscope. This stethoscope has ambient noise reduction and sound improvement capabilities which facilitates the acquisition of clean and clear recording from the heart.

2.2.1 Heartbeat sounds database. Due to the complexity of contact with medical patients in order to get physiological data it was necessary to obtain an open-source database with labelled audio files of heartbeat sounds in order to train our classification model. We were able to obtain a database spawned from the “Classification of normal/abnormal heart sound recordings: The Physionet/Computing in Cardiology Challenge 2016”. This challenge aims to encourage the development of algorithms focused on heart sound diagnostic.

The heart sound recordings were sourced from several contributors around the world, collected at either a clinical or nonclinical environment, from both healthy subjects and pathological patients. The training set consists of five databases (A through E) containing a total of 3,126 heart sound recordings, lasting from 5 seconds to just over 120 seconds.

The recordings were collected from the typical four locations: aortic area, pulmonic area, tricuspid area and mitral area. In both training and test sets, heart sound recordings were divided into two types: normal and abnormal heart sound recordings. The normal recordings were from healthy subjects and the abnormal ones were from patients with a confirmed cardiac diagnosis. All recordings were resampled to 2,000 Hz. Before being able to model and train our classification algorithm it was necessary to prepare the dataset in order to filter and transform the audio files into raw data points so that the computer can analyse them. It was necessary to transform the information from audio into values that the computer is able to interpret. The Fast Fourier Transform (FFT) is a signal processing method that is used widely in engineering in order to analyse the frequencies that constitute a given signal, in our case a given heart sound. On the other hand, the spectrogram method is a visual representation of the frequency spectrum of a signal, this basically means that you can transform sound into an image.

By using the FFT we transformed our audio files from the time domain into the frequency domain. By doing this, the audio file can then be transformed into numerical values representing the fundamental frequencies that constitute a heartbeat sound.

For the purpose of our classification tasks we performed two types of transformation of the audio files. We applied a raw FFT to each audio file and generated a csv file from the frequency/amplitude values, which is basically a row/column file. This csv files were the input of a Multilayer Perceptron Neural Network and a Deep Neural Network (DNN), which use numerical data for training. Figure 1 shows the result from performing the FFT to an audio file.

![Figure 1. Processing of audio files](image)

The second transformation was to create spectrogram images from each original audio file and in this way change our sound classification problem into an image classification problem by using
Convolutional Neural Networks (CNN). The images are then transformed into numeric matrices and are the input of the neural network. Figure 2 shows the result from obtaining the spectrogram from an audio file.

![Figure 2. Spectrogram obtained from audio files](image)

Each of these transformations was done using Python scripts with the use of Pandas and Matplotlib libraries.

2.3 The model

In order to classify the database of heartbeat sounds it was needed to develop a specific machine learning model in order to analyse the huge amount of data that was obtained from the processing of the audio files. For this purpose, we opted to use three different models in order to compare the level of accuracy that was obtained from each one and ultimately use the one with better performance so we could have a more reliable pre-diagnostic tool. These approaches were: multilayer perceptron, deep and convolutional neural networks.

The multilayer perceptron neural network (MPL) is probably one of the simplest, but not less powerful, approaches to machine learning. This type of model makes use of an input layer, several hidden layers and a fully connected layer of perceptrons. The method used for training this type of neural network is called backpropagation which aims to reduce an error function in each of the training steps.

Our MLP consists of 2 hidden layers containing 64 perceptrons each and the output layer of 2 neurons. This approach was made in order to familiarize ourselves with the use of TensorFlow and to validate that our data had the correct format to be used in training a model. Deep Learning is a sub branch of machine learning that attempts to model high-level abstractions in data by using multiple processing layers. By using multiple linear and non-linear transformations the model can detect patterns, characteristics and features in big amounts of data.

Our Deep neural network (DNN) was performed using TensorFlow and following a previous model for classification which was modified for our purposes. The DNN consists of 3 hidden layers with decreasing number of neurons in each layer and an output layer of 2 neurons to obtain the classification as normal or abnormal. This outlay of the neural network was selected based on a symmetric backpropagation algorithm used in classification tasks. The training was performed using a batch size of 100 blocks of data for each run that the algorithm was trained. The total amount of steps of training varied between 2,000 and 10,000 steps for comparison between the time of training, percentage of success for the error function and the final accuracy of the model. [4]

The convolutional neural networks are very similar to the previous deep neural network, except for the fact that these architectures make the explicit assumption that the inputs are images, which allows us to perform different operations on the numeric matrices of pixel information. These then make the forward function more efficient to implement and vastly reduce the number of parameters in the network.
In order to build a convolutional net, we stack three main types of layers to build them: Convolutional Layer, Pooling Layer, and Fully-Connected Layer.

The images generated from the spectrogram of the audio files were resized to the scale of 248x188 pixels in 3 channels of colour (RGB) in order to obtain symmetric filtering in the convolution and pooling layers.

2.4 Augmented reality application
We developed a basic API for connection between a cloud database that will store the medical records and diagnostics and the smartphone’s application. The processing is, up to this point, made using a computer and the information gathered is sent to the API with secure protocols in order to store the information of the patient safely. This API is then connected to the AR application developed on iOS in order to get the information of a specific patient and display the information requested using AR elements through the smartphone’s camera.

2.4.1 Application system operation. The user has their heartbeat recorded with the digital stethoscope to gather information about their current cardiovascular state. The audio file generated from the user is then displayed using Littman’s software. This information is pre-processed by a Python script to obtain the frequency values of the audio file and it is then used as input values for our classification model in order to obtain a basic diagnostic that can help the user detect any anomaly in the user’s cardiovascular health.

Once the prediction is made, the information is uploaded to our web API and then it is obtained by the smartphone to be displayed interactively with augmented reality elements. The results are shown to the user with virtual elements added to the view of their smartphone's camera through our augmented reality application. The whole process takes less than 2 minutes to complete from the moment the users get their heartbeat readings until the information is displayed in the application. Figure 3 shows the complete flow of operation for the diagnostic application.

![Figure 3. Flow of operation](image)

2.5 The results
The training of each of the models implemented for the classification tasks was done with the following characteristics:

- The dataset was divided into 70% training set and 30% testing set.
- The training was done increasing the amount of epoch, training cycle, in order to detect if there was an improvement when increasing the value of if we would only face overfitting problems.
- The number of datasets used for training, the database consists of datasets A through E, was increased gradually to perform faster trials of training and then increased in order to improve the models.
We present only the best results which were obtained with the DNN, table 1.

Table 1. Results of testing DNN

| # Steps | Training time | Test accuracy |
|---------|---------------|---------------|
| 5000    | ~2 minute     | 72.41%        |
| 10,000  | ~2 minute     | 84.18%        |
| 15,000  | ~2 minutes    | 75.72%        |
| 20,000  | ~2 minutes (8 minutes total) | 85.03% |

3. A MRI brain injury image analysis

3.1 Overview

Diffusion tensor imaging (DTI) is an advanced medical imaging modality that reveals in vivo information of fibrous structures of the body, such as the white matter of the brain. Being recently implemented, just a decade ago, it is one of the newest magnetic resonance imaging modalities. But it has already become a very important tool in the study of brain pathologies. Since it has been proven effective in studying a range of white matter disorders including brain injury, brain tumours and Alzheimer’s disease. A diffusion tensor measures a 3D diffusion process and has six interrelated tensor components. Tensor methods are extensions of standard matrix decompositions and data modelling methods, such as the matrix Singular Value Decomposition (SVD), Principal Component Analysis (PCA), Independent Component Analysis (ICA) to higher dimensions.

Given the size of current datasets and the multiway structure into which they are encoded classical statistical tools cannot longer be applied without altering their intrinsic organization; leading to a potential loss of information. In recent years the increasing size of emerging data is posing unprecedented demands for new statistical methods. Correspondingly most data analysis methods in their standard definition do not take into account the inherent structure that is present in the data. In the given case that the data comes as a multiway array, also known as tensor, the data is described by more than two dimensions. Consider for example the situation were the data is contained in a three dimensional matrix $I \times J \times K$, where the dimension $I$ corresponds to the samples, $J$ to the set of features and $K$ to different instances. In order to apply a given a statistical method it is required to re-structure the data into a $I \times JK$ two way array, where the original structure is lost.

An illustrative example of this can be given by a multi-modal brain Magnetic Resonance Imaging (MRI) data set where $K$ neuroimaging modalities (each characterized by roughly $J = 250,000$ pixels), are collected on a set of $I$ patients, forming a $I \times J \times K$ multiway array. In this context, an individuals x voxels x modalities array can be considered and yields a three-way array. When considering the standard statistical approach, the size of the parameter to be estimated would be of size $Jx250,000xK$, which is computationally burdensome and leads to a challenging interpretation. Hence, dedicated modelling algorithms able to cope with the inherent properties of tensor structures are therefore mandatory for harnessing their complexity and provide relevant information.

Tensor decompositions have a wide range of applications in data analysis. These decompositions have been successful in many areas; and may be used to classify any objects, or object attributes. Most data are the result of several causal factors of data formation and are well suited for multimodal data tensor analysis; for example, medical images are the result of a person’s geometry, imaging equipment parameters, etc. While we can directly measure the greyscale (or colour) values in an image, we are often more interested in the information associated with the causal factors, such as the person’s health. The causal factors are a set of latent/hidden variables and the goal is to estimate them.

Next we present an example that make use of these multiway methods this study use a real Diffusion Tensor (DTI) Magnetic Resonance Imaging (MRI) dataset. Being the objective of this application to make a long term prediction of the outcome of comatose patients after a traumatic brain injury. This example uses the classification methods of Fisher Discriminant Analysis (FDA) and Multiway Fisher
Discriminant Analysis (MFDA).

3.2 The data
Multi-modal MRI diffusion images were acquired on individuals divided into 3 classes: 39 controls, 65 comatose patients with a positive outcome and 39 comatose patients with a negative outcome (I = 143). Each volumetric image has 91x109x91 voxels which are then reshaped into a 1x 902 629 vector which after removing the zeros corresponding to areas outside the brain we obtain J = 255 726. Four types of diffusion images were measured (K = 4): fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (L1) and radial diffusivity (Lt), figure 4, acquired from the entire brain for both patients and controls.

![Types of diffusion images](image)

**Figure 4.** Types of diffusion images

3.3 Results
When considering the entire brain and the 3 classes previously presented and FDA is applied to X, it results in weight matrices (since we obtain 2 weight vectors per modality).

The following Figure 5, shows the resulting weight matrices, represented in images, corresponding to the modalities FA, MD, L1 and Lt respectively. These volumes are difficult to interpret since there is no focalized region used for the discrimination.

![Weight matrices](image)

**Figure 5.** Weight matrices corresponding to the modalities FA, MD, L1 and Lt respectively
Applying MFDA to X, results in only 2 weight vectors $w_J^1$ and $w_J^2$ associated to the J variables which integrate the weight of the variables, and 2 weight vectors $w_K^1$ and $w_K^2$ associated to the different modalities. This yields a single volumetric image integrating 4 modalities per weight vector instead of one for each modality as in FDA (Remark that this represents a great gain, since now the obtained weight vectors are reduced by a quart).

In the figure 6 we show the obtained MFDA weights $w_J^1$, $w_J^2$, $w_K^1$ and $w_K^2$ showing the contribution of each modality for the construction of each volumetric image. Since specific and smooth regions are selected, MFDA model is easier to interpret. MFDA seems to supply more concrete information on the location of discriminative voxels by identifying specific areas within the white matter without any given a priori which is consistent with the evidence that damage in this region is a distinctive feature of traumatic brain injury.

![Figure 6. Multiway FDA weights](image)

|   | FA       | MD        | Lt        | L1        |
|---|----------|-----------|-----------|-----------|
| $w_K^1$ | 0.9969   | -0.0013   | 2.9 e-04  | -0.0018   |

An extended developed and detailed example on long-term outcome forecast, on brain injuries and its level of accuracy can be seen in [5].

4. Conclusion

Augmented reality is the next step in terms of immersive experiences in a broad amount of industries. The approach that we made by using the latest technologies for displaying information showed excellent results, even though it is not the most complex experience that can be achieved. This application is part of a bigger project that aims to create low-cost diagnostic medical tools using the latest technology, in order to spread the use in developing nations. Both applications made use of new technological tools and recent adapted statistical methods to new technologies and data. Even though the software tools that we used for the development made several tasks simpler, the areas of machine learning and augmented reality present a very steep learning curve. Before being able to develop complex systems with these technologies, a big learning background is required in order to implement the solutions that you design. Nonetheless, the combination of these two technologies will transform the way we do and live things in the future.

Acknowledgements

The Japanese International Cooperation Agency (JICA) and the Mexican National Council for Science and Technology (CONACYT) are acknowledged for the financial support in the development of these projects. “The authors would also like to acknowledge the financial and the technical support of Writing Lab, TecLabs, Tecnologico de Monterrey, Mexico, in the production of this work.”
References

[1] Grubb A, Walsh P, Lambe N, Murrells T and Robinson. Views on management of patients in persistent vegetative state 1996. *S. Survey of British clinicians’ The Lancet*, vol. 348, no. 9019, pages 35–40.

[2] Judd K. The DANA guide to brain health. A practical family reference from medical experts 2007. *J. of neurology, neurosurgery, and psychiatry*, vol. 78, no. 7, page 779.

[3] Galanaud D, Perlberg V, Gupta R, Stevens R D, Sanchez P. Assessment of white matter injury and outcome in severe brain trauma a prospective multicenter cohort. 2012. *J. of the American Society of Anaesthesiologists*, vol. 117, no. 6, pages 1300–10.

[4] Chollet F. Deep Learning with Python. Manning publications company 2017, pages 105–143.

[5] Lechuga G. Contributions à l’analyse des données multivoie : algorithmes et applications 2017. *Theses.fr/2017SACL038*. 