Masters Research Proposal - What are the effects of implementing a two-step system that includes a dimensionality reduction process for the detection of bone fractures

Alice Yang (597609)

School of Electrical and Information Engineering, University of the Witwatersrand, Private Bag 3, 2050, Johannesburg, South Africa

Abstract: The objective of this paper is to propose a two-step system for the detection of bone fractures in order to investigate the performance of the system compared to a single step system. The performance are defined by the speed of its execution, classification accuracy and error rate. The two-step system consists of a dimensionality reduction and a automated decision making process. The objective of the introduction of the dimensionality reduction process is to improve the performance of the automated decision making process during the training and decision making sessions.

Key words: Dimensionality Reduction, Neural Networks, Supervised Learning, Unsupervied Learning, Bone Fractures, Medical Images

1. Introduction

In medical the field there is an emphasis on the importance of medical diagnosis. In critical cases, an inaccurate diagnosis of a symptom can cost a patient's life. Many diagnosis are performed by trained medical doctors, however they are prone to making mistakes which can be influenced by external factors, such as fatigue and lack of training in the particular field. Automation of medical diagnosis was introduced in the early 1970s [1]. There are many algorithms available in diagnosing patients. The basic algorithm consists of "if... then..." statements. Although the algorithm is simple, the list of conditions in the medical field are endless and as a result the execution time to diagnosis a patient's symptoms is not realistic. There are many pattern recognition algorithms available which are categorized in two categories, namely, unsupervised and supervised. The reasoning behind making use of artificial intelligence is to reduce costs. In this proposal report, section 2 presents the background in the related topics, section 3 describes the problem, section 4 illustrates the proposal methodology, section 5 shows the time management for the proposed research and section 6 concludes the paper.

2. Background

There are many simple implementations of programs that are proposed to diagnosis common medical conditions. These implementations requires a vast majority of the medical knowledge to express the different descriptions of possible diseases to be stored in the program's database. Although the program is simple to implement, it has no method of measuring the severity of the patient's illness. Another downfall to using a database to diagnosis an illness, is that it can mean that the database can grow exponentially otherwise it would be outdated.

An alternative approach in medical diagnosis is to focus on one disease to ensure accuracy and quick response. The use of artificial intelligences can be focussed in one area. The focus of this paper is to detect bone fractures from X-Ray images. There are a selection of tools that can be applied in order to achieve the desired outcome for this investigation. These tools can be in the form of supervised learning, reinforcement, unsupervised learning or a combination of the two. Support Vector Machine and Neural Networks are common tools used in artificial intelligence (AI).

3. Literature Review

The review done by [2] for the automation of bone fracture detection indicated that there are various techniques in pattern recognition that comes in the form of unsupervised, reinforcement and supervised. The challenge found in classification are that there is an information overload and size and dimension are a concern. The problem with information overload is that it becomes to computationally expensive to find crucial information needed for the classification. The size and di-

mension of the information is a concern because large sizes of data contributes to the number of dimensionality, with high number of dimensions it presents a challenge in the classification process and in turn may require more computational resources.

3.1 Curse of Dimensionality

The main problematic factor affecting the performance in classification is the curse of dimensionality. The curse of dimensionality is a problem due to the sparsity of high dimensional spaces, in which an absurd amount of training data is needed in order to get low variance estimators [3].

Feature extraction with supervised learning algorithms may seem most desired compared to unsupervised learning algorithms, since supervised learning algorithms have more information about the problem and features. However, unsupervised methods do not suffer the curse of dimensionality as supervised learning does since it makes use of local measures to optimally estimate a single dimensional projection function [3].

3.2 Dimensionality Reduction

The collection of digital data has increased drastically over the past decade, which led to datasets having high dimensionality. The high dimensionality found in datasets affected the performance of data processing algorithms. This is known as the curse of dimensionality [4]. Dimensionality reduction is a process targeted at reducing the dimensionality of the considered dataset by reducing the number of random variables. This process can be divided into two stages, namely Feature Selection and Feature Extraction. Both these stages are crucial for automating bone fracture detection. There are two types of dimensionality techniques, convex and concave techniques. This section focuses on convex techniques. Convex techniques optimizes objective functions that do not contain any local optima, which means that the solution space is convex [5]. The objective function is usually in a generalized Rayleigh quotient, which can be expressed in the following form:

$$\phi(\mathbf{Y}) = \frac{\mathbf{Y}^T\mathbf{A}\mathbf{Y}}{\mathbf{Y}^T\mathbf{B}\mathbf{Y}} \tag{1}$$

The function expressed in the form of Equation 1 can be easily optimized by solving a generalized eigenproblem. Convex dimensionality reduction techniques can be subdivided into techniques that perform eigendecomposition on full matrix and sparse matrix. The following section focuses on the eigendecomposition of full matrix.

3.2.1 PCA/Classical Scaling Principal Component Analysis (PCA) is a linear technique that performs dimensionality reduction through the process of embedding the data into a linear subspace of low-dimensionality. The low-dimensional representation of the data describes the variance of the data [6]. During the construction of the low-dimensional representation, PCA does not discard information, instead it creates new characteristics to represent the original characteristics. This is done by searching for characteristics that show as much variation as possible. Additionally, PCA searches for characteristics that allows for the reconstruction of the original characteristics from the new characteristics. According to [7], maximizing the variance will result to minimizing the error. Hence, the construction of the low-dimensional representation is obtained by determining mapping $\mathbf{M}$ which maximizes the cost function. The cost function is expressed in Equation 2.

$$\phi(\mathbf{Y}) = \sum_{ij}(d_{ij}^2 - ||\mathbf{y}_i - \mathbf{y}_j||^2) \tag{2}$$

where

$||\mathbf{y}_i - \mathbf{y}_j||^2$ = squared Euclidean distance between low-dimensional datapoints $\mathbf{y}_i$ and $\mathbf{y}_j$

Eigenvectors and eigenvalues are solved using the eigenproblem expressed in Equation 3. Eigenvectors and eigenvalues are crucial in PCA since eigenvectors assist with determining the correlation between the data points whilst eigenvalues, λ dictates the weighted average of the variance for any projection.

$$cov(\mathbf{X})\mathbf{M} = \lambda\mathbf{M} \tag{3}$$

The disadvantage with PCA is that the size of the covariance matrix is dependent on the dimensionality of the original dataset. This means that the size of the covariance matrix is proportional to the number of dimensionalities presented by the dataset, which can result in the inability to

compute eigenvectors for very high-dimensional datasets [7].

Classical Scaling is an identical technique to PCA, however Classical Scaling searches for the linear mapping **M** that minimizes the cost function that is expressed in Equation 2. Furthermore, the low-dimensional representation is of the Gram Matrix in which the double-centering pairwise squared Euclidean distance matrix entries are obtained using Equation 4.

$$k_{ij} = -\frac{1}{2}(d_{ij}^2 - \frac{1}{n}\sum_l d_{il}^2 - \frac{1}{n}\sum_l d_{jl}^2 + \frac{1}{n^2}\sum_{lm} d_{lm}^2) \tag{4}$$

The disadvantage for both PCA and Classical Scaling is the cost function in Equation 2 focuses on retaining large pairwise distances d_{ij}^2, whereas retaining the small pairwise distance can be important for minimizing error.

3.2.2 Isomap Unlike PCA and Classical Scaling where the aim of the techniques is to retain the pairwise Euclidean distances, the Isomap technique attempts to preserves the pairwise geodesic distances between datapoints [8]. This means that the geodesic distance between x_i and x_j imitates as much of the Euclidean between the low-dimensional representation yi and yj as possible. The low-dimensional representation y_i and y_j are computed using the classical scaling technique, which results in a pairwise geodesic distance matrix. The drawback of the Isomap technique is that it constructs erroneous connections in the neighbourhood graph, G which can affect the performance of the Isomap.

3.2.3 Kernel PCA Kernel Principal Component Analysis (KPCA) is a technique that makes use kernel functions to map data into high dimensional space, by which the space is manipulated using linear PCA [9]. KPCA makes use of a mapping function by which it computes the kernel matrix K of datapoints x_i and x_j. The entries for the kernel matrix is defined by Equation 5.

$$k_{ij} = K(x_i, x_j) \tag{5}$$

where

$$K = \text{kernel function}$$
$$x_i \text{ and } x_j = \text{input datapoints}$$

The features defined by the kernel function have a zero-mean. Additionally, the eigenvectors of the covariance matrix $\mathbf{a}_i$ which is expressed in Equation 6 can be computed, since the eigenvectors of the kernel matrix are related.

$$\mathbf{a}_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{v}_i \tag{6}$$

The low-dimensional representation is obtained by the projection given by Equation 7.

$$\mathbf{y}_i = \left\{ \sum_{j=1}^{n} a_1^{(1)} K(\mathbf{X}_j, \mathbf{X}_i), ..., \sum_{j=1}^{n} a_d^{(j)} K(\mathbf{X}_j, \mathbf{X}_i) \right\} \tag{7}$$

where

$$a_1^{(j)} = \text{jth value in vector } \mathbf{a}_1$$
$$K = \text{kernel function}$$

A disadvantage with KPCA is that the size of the kernel matrix is proportional square number of instances in the dataset. However despite the obvious disadvantages, KPCA is applied to facial recognition, speech recognition and novelty detection [7].

3.2.4 MVU Maximum Variance Unfolding (MVU) is a technique that aims to preserve as much of the distances and angles between nearby points by studying the kernel matrix. As result it forms a neighbourhood graph, G [10]. A quadratic equation is formulated for the "unfolding" transformation. The "unfolding" transformation is where MVU maximizes the sum of the squared Euclidean distances between all the datapoints whilst preserving the distances within neighbouring points. This can be described in Equation 8.

$$\text{Maximize } \sum_{ij} ||\mathbf{y}_i - \mathbf{y}_j||^2 \tag{8}$$

$$\text{Subject to: } \begin{cases} (1)||\mathbf{y}_i - \mathbf{y}_j||^2 = ||\mathbf{x}_i - \mathbf{x}_j||^2, \eta_{ij} = 1 \\ (2)\sum_i \mathbf{y}_i = 0 \end{cases}$$

MVU reformulates the optimization as semidefinite programming (SDP) over matrix **K** by defining the inner product of $\mathbf{K}_{ij} = \mathbf{y}_i \cdot \mathbf{y}_j$. Thus the SDP can be written as follows:

Maximize trace($\mathbf{K}$) subjected to:

$$\begin{cases} (1)\mathbf{K} \geq 0 \\ (2)\sum_{ij} \mathbf{K}_{ij} = 0 \\ (3)\mathbf{K}_{ii} - 2\mathbf{K}_{ij} + \mathbf{K}_{jj} = ||\mathbf{x}_i - \mathbf{x}_j||^2, \eta_{ij} = 1 \end{cases} \tag{9}$$

From the first condition $\mathbf{K} \geq 0$, it indicates that matrix $\mathbf{K}$ is required to be positive semi-definite, since SDP is convex. Thus the low-dimensional representation is obtained by solving the SDP. The main weakness found in MVU is the addition of constraints during optimization, which can result to unsuccessful unfolding of manifold. Although, despite its main weakness it is applied to sensor location and DNA micro-array data analysis [7].

3.2.5 Diffusion Maps Diffusion Maps (DM) is a technique based on defining Markov Random Walk graph [11]. Markov Random Walk is a description of a random process that consists of the successful random steps applied to a mathematical space. The measurement for the proximity of the datapoints is attained by performing the random walk for a number of iterations. Through the number of iterations, the diffusion distances are obtained in which it is used for the representation of the data in the lower dimension. The purpose of the diffusion distance is to integrate over all the paths presented in the graph. For the construction the diffusion maps graph, the weights of the edges are computed using the Gaussian kernel function. The entries to form matrix $\mathbf{W}$ is expressed in Equation 10.

$$w_{ij} = e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}} \tag{10}$$

where

$$\sigma = \text{variance of the Gaussian}$$

The normalization of matrix $\mathbf{W}$ is necessary defining the forward transition probability matrix of dynamical process. The normalized matrix $\mathbf{W}$ results in matrix $\mathbf{P}^{(1)}$ which is formed using Equation 11.

$$p_{ij}^{(1)} = \frac{w_{ij}}{\sum_k w_{ik}} \tag{11}$$

By making use of the probability matrix for t iteration $\mathbf{P}^{(t)}$, the diffusion distance can be defined in Equation 12.

$$D^{(t)}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_k \frac{(p_{ik}^{(t)} - p_{jk}^{(t)})^2}{\psi(\mathbf{x}_k)^{(0)}}} \tag{12}$$

where

$$\psi(\mathbf{x}_i)^{(0)} = \frac{m_i}{\sum_j m_j}$$

m_i is the degree of the node $\mathbf{x}_i$ and can be further defined as follows:

$$m_i = \sum_j p_{ij}$$

The eigenproblem in which diffusion maps solves is expressed in Equation 13.

$$\mathbf{P}^{(t)}\mathbf{v} = \lambda\mathbf{v} \tag{13}$$

Since the graph is fully connected, the largest eigenvalue is trivial, therefore the corresponding eigenvalue is discarded. The low dimensional representation can be expressed in Equation 14.

$$\mathbf{Y} = \{\lambda_2\mathbf{v}_2, \lambda_3\mathbf{v}_3, ..., \lambda_{d+1}\mathbf{v}_{d+1}\} \tag{14}$$

where

$$\lambda = \text{eigenvalues}$$
$$\mathbf{v} = \text{eigenvectors}$$

The applications that diffusion maps have been applied to are shape matching and gene expression analysis.

3.3 Supervised Learning

The goal of supervised learning to to create a mapping function, such that when given new input data a prediction of the output data can be made. The mapping function is generated by with given data sets that consists of inputs with its corresponding outputs. Linear regression, random forest and SVM are popular supervised algorithms. Support Vector Machine (SVM) is a classifier mainly used for complex classification purposes. In order to separate new data into specified categories, SVM is trained with given labelled data making it supervised learning [12]. An implementation of SVM classifiers is done in [13] where the it is used to classify X-Ray images of the body into five categories, namely, head-neck, body, upper-limb and true-negative. Spatial Pyramid Histogram are used to decipher the

medical images which is then used to train the Chi-Kernel based SVM. In general, SVM is used for binary classification [12].

Neural Networks in another common tool selection for AI, whereby it consists of three distinguished layers, the input, hidden and output layer. A general neural network consists of one or two hidden layers. The network is trained before use, which deems it as supervised classification or predictions. In [14], the technique implemented is back-propagation neural network (BPNN). The network is trained using a Supervised Delta Learning Rule. The result from the developed technique is a binary outcome in which it indicates whether the subject is normal or abnormal.

Studies of improving the common tools in AI have been done by combining the tools. There are two types of combinations, hybrid and non-hybrid multiple classifiers. Hybrid multiple classifiers combines different types of AI algorithms, whereas non-hybrid multiple classifiers only consists of one base algorithm and is replicated multiple times. A hybrid multiple classifier is done by [15], in which the algorithms used are K-Nearest Neighbour, Back Propagation Neural Network (BPNN) and SVM. According to the authors, the each of the AI tools is trained with a different set of data. The decision is based on a voting scheme that is called fusion selection. However, the paper presented in [15] performs binary results. In other words, it indicates whether a fracture is present or absent.

Another AI algorithm to be considered for medical diagnosis is deep learning. According to the author in [16], deep learning has the ability to combat complex problems whereas SVM would not be able to, since SVM works better with small data sets that have few outliers. However, the downfall of deep learning is its learning process, in which it can become tedious and computational demanding. Thus for small data sets, it is more plausible to use an off-shelf classifier such as SVM.

3.4 Existing Solutions

There have been studies done which combine the algorithms of supervised and unsupervised. The authors of [17] proposed a method which combines both supervised and unsupervised algorithms, name.

The authors of [18] designed a system in which it comprises of Principal Component Analysis(PCA) and Neural Network for the classification of EEG Signals. PCA is a form of unsupervised learning whilst neural network is supervised learning and requires training. The authors compared the correctness of the classification using Neural Network to using Principal Component Analysis with Neural Network. The final outcome from the experimentation is that the correctness of the classification of the Principal Component Analysis with Neural Network is better than Neural Network alone.

4. Problem Analysis

The modern world has presented the technology to predict human speech as well as digitally categorizing handwritten numbers. However, this technology has not been fully utilized in the medical field. There are many debates surrounding the topic of whether machines are better than doctors. In many cases, the argument for doctors states that those who develop AI do not understand the complexity of medicine and those arguing for the machines say that not many understand how AI operates. However, at the end of the day there are still sick patients waiting to be treated with the right medication or medical care. In order to determine the type of medical care the patient needs a diagnosis must be performed. According to [19], where studies have been done around the United States, it was reported that doctors correctly diagnosed 55.3% for easier cases and 5.8% for more difficult cases. However, this is only one article and more investigation is needed to confirm these numbers. Even so, these numbers still indicate that doctors are not perfect and can mis-diagnosis patients. As a result of mis-diagnosis, the patient is subjected to stress or even life threatening situations should the disease be mis-treated. Therefore there is a large emphasis on accuracy and quick detection of diseases.

There are existing pattern recognition techniques available which have the ability to detect the presence of bone fracture. However, there are large amounts of datasets that is passed into the techniques. These datasets create dimensionality, the

more datasets there are the higher the number of dimensionality there are. High number of dimensionalities affects the performance of the techniques. This is known as the curse of dimensionality.

5. Proposed Methodology

The objective of this research is to propose a system which consists of two steps in order to determine the presence of a fracture within medical x-ray images. The first step is dimensionality reduction and the second step is the automation of the decision making process using neural networks. The dimensionality reduction step is intended to map the data from a high dimensionality to a low dimensionality without loosing any critical information needed for the input into the neural network. The system overview is illustrated in Figure 1.
In addition, to developing the two-step system an image pre-processing technique is introduced to the system. This can be seen in Figure 1.

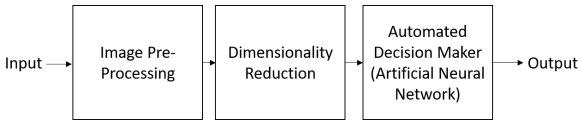

Figure 1 : Block Diagram illustrating an overview of the system with image pre-processing

The performance for the system is compared to the performance of a single step neural network. The performances are evaluated by accuracy of detection, error rate, and execution speed. The idea behind comparing the systems is to determine whether there are underlining features in the medical x-ray images that simplifies the training process for the neural network as well as obtaining accurate results in minimal time.

5.1 Image Pre-Processing

The medical image is pre-precessed to de-noise and sharpen the image. The image pre-processing system makes use of both sobel and canny techniques. The system produces a total of 26 images highlighting the defining features in the x-ray image in various ways. The types of images produced are listed in Table 1 along with the dimensions of each image.

Table 1 : Table showing the types of images produced by image pre-processing system along with the dimensions of each image

Image (jpg file)	Dimensions (pixels)
equalised gamma_corrected denoised sharpened hist1 hist2 bitnot bitor blur a_contour binary_high binary_inverse binary_low canny_contours_after_joining canny_contours_after_removing i_cannyx3 sobel_dire sobel_value u_hough v_hough w_hough	809×899
binary_gray circle circle_contours gray sobel	1200×900

Furthermore the image pre-processing system produces csv files for gradients, lines and contours for the i_cannyx3.jpg image. The dimensions of each extracted feature is shown Table 2.

Table 2 : Table showing dimensions of the extracted features from i_cannyx3.jpg image, where N_g is the total number of values for gradient, N_l is the total number of values for lines and N_c is the total number of values for contours.

Features	Dimensions	Description
gradient	$2 \times N_g$	hough line value, angle (in degrees)
lines	$4 \times N_l$	x_1, y_1 x_2, y_2
contours	$2 \times N_c$	x and y values

5.2 Data Structure

The data produced by the image pre-processing system can be structured in two various ways. The first structure consists of making use of all the images produced. This creates a three dimensional (3D) structure for one image. The structure can be represented by $1200 \times 900 \times 26$. A graphical representation of the structure can be seen in Figure 2.

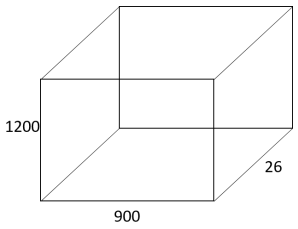

Figure 2 : Data structure using all 26 images produced by the image pre-processing systems

The second structure is constructed from the three extracted features from the i_crannyx3.jpg, which are gradient, lines and contours. A simplified graphical representation of the structure is shown in Figure 3.

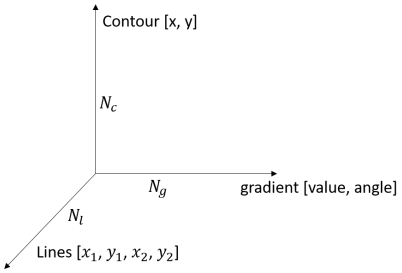

Figure 3 : Simplified data structure using the three extracted features from i_cannyx3.jpg

5.3 Dimensionality Reduction

For investigation purposes, dimensionality reduction is applied to both data structures to test the performance of the different data structures as well as determining whether there is a defining feature from the x-ray images to determine the presence of a bone fracture.

The first data structure for one image has a 3D structure, $1200 \times 900 \times 26$. Since there are 26 resulting images from one original x-ray image, the dimensionality reduction technique chosen is Principal Component Analysis (PCA) since this is linearly structured whereby only one data element is accessed as a time.

The first step is to determine the covariance matrix for original x-ray image. The covariance matrix is then used to determine the principal components. The number of principal components is proportional to the number of variables, n within an image. The resultant principal components, P_n are computed by determining the eigenvalues of the covariance matrix. The covariance matrix is in a symmetrical form. The eigenvalues found in the covariance matrix are the variances of the principal components. Since the number of principal component are proportional to the number of variables within an image, this means that there are n eigenvalues. All eigenvalues, λ are greater than or equal to zero. The largest eigenvalue corresponds to the first principal component. This applies to the next eigenvalue, until the i-th principal component. Therefore λ_i corresponds to the i-th eigenvector.

5.3.1 Step 1: The first step is to determine the covariance matrix for each original x-ray image, which has a three dimensional structure. Let the three dimensions be represented by x, y and z. Thus the covariance matrix, C can be expressed in Equation 15.

$$C = \begin{bmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{bmatrix} \quad (15)$$

The result of $cov(x,x)$, $cov(y,y)$ and $cov(z,z)$ are eigenvalues of matrix C which are the variances of the principal component.

5.3.2 Step 2: The eigenvectors $v_1, v_2, v_3, ..., v_i$ corresponding to the eigenvalues $\lambda_1, \lambda_2, \lambda_3, ..., \lambda_i$ are calculated. Assuming that the eigenvalues are in descending order such that: $\lambda_1 \geq \lambda_2 \geq \lambda_3, \geq , ..., \geq \lambda_i$, then λ_1 is the first principal component and v_1 consists of the main characteristics for the given data.

5.3.3 Step 3: The principal components that expresses the 26 images can be expressed as fol-

lows:

$$P_i = \lambda_{i1} Z_1 + \lambda_{i2} Z_2 + ... + \lambda_{in} Z_n \qquad (16)$$

where Z_n represents the 26 various images of the original x-ray image.

5.3.4 Step 4: The low-dimension matrix is constructed by selecting the principal components that consists of the most variations of the data. Therefore the result mapping covariance matrix can be expressed as follows:

$$M = [\lambda v_1, \lambda v_2, ..., \lambda v_n] \qquad (17)$$

5.3.5 Step 5: The low-dimension matrix is then constructed as follows:

$$Y = M^T \times \text{Original Data} \qquad (18)$$

5.4 Neural Network

An artificial neural network consists of three different layers: input layer, hidden layer and output layer. There are weighted connections which link the three layers together. The weights are factors which assist with the automated decision making process. The weights values are determined through a training process. The chosen neural network training method is back-propagation. The comparison

The automation of bone fracture detection can be done by utilizing a neural network technique. The implementation of the neural network technique consists of two stages. The first stage is the training of the technique and the second is the classification of the bone fracture. The neural network technique has the ability to process the raw data from the selected features, however since the raw data has high dimensionality it has the potential to hinder the performance of both stages in the neural network technique. This will be observed during execution. The neural network technique in consideration is Back Propagation Neural Network, however this choice of technique is subjected to change.

6. Time Management

Table 3 presents the time management for the proposed research presented.

Table 3 : Table showing the time management of the implementation of a bone fracture detection two step system.

Date	Task
1 March	1st Draft Hand-In
2 March	Investigate best suited dimensionality reduction technique for selected features
8 March	Present 1st iteration of source code for the two step system
15 March	Present 2nd iteration with automated decision making process implemented
22 March	Submit Proposal

7. Conclusion

To conclude, the system proposed for the detection of bone fracture consists of two steps. The first step is dimensionality reduction and the second step is the automation of decision making. The objective of implementing a dimensionality reduction process is to improve the performance of the automated bone fracture detection. The performance is measured by speed, classification accuracy and error rate.

REFERENCES

[1] A. N. Ramesh, C. Kambhampati, J. R. T. Monson, and P. J. Drew. "Artificial intelligence in medicine." *Annals of the Royal College of Surgeons of England*, vol. 86, no. 5, pp. 334–338, 2004.

[2] S. K. Mahendran and S. S. Baboo. "Automatic Fracture Detection Using Classifiers- A Review." vol. 8, no. 6, pp. 340–345, 2011.

[3] N. Intrator. "Feature Extraction Using an Unsupervised Neural Network." vol. 4, no. 1, pp. 98–107.

[4] I. K. F. Center, A. Scienti, and L. L. N. L. P. O. Box. "A survey of dimension reduction techniques." *Computing*, , no. 1, pp. 1–18, 2002.

[5] S. Boyd and L. Vandenberghe. *Convex Optimization*, vol. 25. 2010.

[6] I. T. Jolliffe, J. Cadima, and J. Cadima. "Principal component analysis : a review and recent developments Subject Areas : Author for correspondence :." 2016.

[7] L. Van Der Maaten, E. Postma, and J. Van den Herik. "Dimensionality reduction: a comparative." *J Mach Learn Res*, vol. 10, pp. 66–71, 2009.

[8] Z. Zhang, S. Member, T. W. S. Chow, and S. Member. "M-Isomap : Orthogonal Constrained Marginal Isomap for Nonlinear Dimensionality Reduction." pp. 1–12, 2012.

[9] W. Cui, S. Chen, T. Yu, and L. Ren. "Feature extraction of X-ray chest image based on KPCA." *Proceedings of 2nd International Conference on Computer Science and Network Technology, ICCSNT 2012*, pp. 1263–1266, 2012.

[10] S. Jiang, J. Zhang, and C. Kuang. "A linear maximum variance unfolding algorithm and its application in image recognition." *2011 International Conference on Electric Information and Control Engineering, ICEICE 2011 - Proceedings*, , no. 61070137, pp. 36–39, 2011.

[11] R. R. Coifman and S. Lafon. "Diffusion maps." *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5 – 30, 2006.

[12] P. Rebentrost, M. Mohseni, and S. Lloyd. "Quantum support vector machine for big data classification." *Physical Review Letters*, vol. 113, no. 3, pp. 1–5, 2014.

[13] S. Saha, A. Mahmud, A. A. Ali, and M. A. Amin. "Classifying digital X-ray images into different human body parts." In *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 67–71.

[14] S. Jyothi and K. Vanisree. "Congenital Heart Septum Defect Diagnosis on Chest X-Ray Features Using Neural Networks." In *2016 Second International Conference on Computational Intelligence Communication Technology (CICT)*, pp. 265–269.

[15] N. Umadevi and S. N. Geethalakshmi. "Multiple classification system for fracture detection in human bone x-ray images." In *2012 Third International Conference on Computing Communication Networking Technologies (ICCCNT)*, pp. 1–8.

[16] Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning." *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[17] V. E. Neagoe and R. M. Stoica. "A new neural network-based approach for automatic annotation of remote sensing imagery." In *2014 IEEE Geoscience and Remote Sensing Symposium*, pp. 1781–1784.

[18] R. Kottaimalai, M. P. Rajasekaran, V. Selvam, and B. Kannapiran. "EEG signal classification using Principal Component Analysis with Neural Network in Brain Computer Interface applications." In *2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN)*, pp. 227–231.

[19] A. N. D. Meyer, V. L. Payne, D. W. Meeks, R. Rao, and H. Singh. "Physicians Diagnostic Accuracy, Confidence, and Resource Requests: A Vignette Study." vol. 173, no. 21, pp. 1952–1958.