Object Recognition of Handwritten and Printed Musical Notation Using Neuro-Fuzzy

1

Lintag, joshua f., Ateneo de Davao University   
regodon, rhyle abram p., Ateneo de Davao University

From the earlier times until today, sheet music has always been an integral part of a musician’s life, whether that musician is a professional, or a student, they would always write their sheet music onto paper. This paper would either be lost, destroyed, or would need adjustments on spot. After long advancements of technology, Optical Music Recognition was developed in order to be of great help to these musicians. Although Optical Music Recognition is efficient in recognizing handwritten or printed sheet music, the methods and algorithms used for it barely scratched the surface of OMR. This research and study is aimed to help Optical Music Recognition using Neural Networks tuned with Fuzzy Logic parameters, known as the Neuro-Fuzzy System.

General Terms: ANN – Artificial Neural Networks

OMR – Optical Music Recognition

NFIS – Neuro-Fuzzy Inference System

FL – Fuzzy Logic

Additional Key Words and Phrases:

Music Sheets

# INTRODUCTION

## Background of the Study

One of the difficult problems faced by computer vision today is the ability to be able to aid musicians when they write their sheets for music. Musicians today write their sheet music onto paper and later use an application like Sibelius, Musecore, or Finale to have a digital copy of their sheet music. The problem is that it is a repetitive process for the musicians and a very tedious one because they would have to do it all over again. Some sheet music are also complex so it would take hours and hours for them to transcribe sheet music from paper.

One of the approaches that would help solve this problem is the use of Optical Music Recognition (OMR). This branch of image processing and computer vision is similar to Optical Character Recognition (OCR), but in a study conducted by Novotny and Pokorny, they stated that OCR proves to be more difficult because unlike the OMR, OCR only focuses on character recognition while OMR focuses on music notation recognition. Therefore, this would prove that Optical Music Recognition can not only be useful, but also more efficient.

For this research, it is necessary to use a Machine Learning algorithm like Neuro-Fuzzy. Using techniques modified from vanilla neural networks, Neuro-Fuzzy is a hybrid neural network tuned with Fuzzy Logic (Fuzzy Sets, Fuzzy Rules) in order to have a human-like reasoning. [[1](#Nov15)]

## Problem Statement

This research and study seeks to identify music notation using optical music recognition with a machine learning algorithm. Primarily, this study would answer the following questions:

The specific problems of the study are as follows:

1. What are the factors needed to detect a handwritten and printed sheet music?
2. Where does Neuro-Fuzzy algorithm stand when compared to other similar machine learning algorithms in terms of accuracy?
3. How can Neuro-Fuzzy machine learning algorithm be used to identify sheet music?
4. What is the accuracy ratings of the Neuro-Fuzzy machine learning algorithm when applied to reading handwritten and printed sheet music?

## Objectives

The whole objective of this research is to know how a machine learning algorithm works to identify handwritten sheet music. During the research, this study would hope to:

The specific objectives of the study are as follows:

1. To discover the factors need to detect handwritten and printed sheet music.
2. To Find out accuracy ratings of Neuro-Fuzzy and other different machine learning algorithms.
3. To implement the Neuro-Fuzzy machine learning algorithm to read handwritten and printed sheet music.
4. To determine the accuracy of Neuro-Fuzzy machine learning algorithm applied to handwritten and printed sheet music.

## Significance of the Study

Machine learning is one of the most efficient approaches in optical music recognition. This would be of great help to musicians, songwriters, and composers since now they would be able to easily convert their handwritten music sheets into digital format. With the help of Machine Learning, Optical Music Recognition would be faster and more efficient. Furthermore, it will also contribute to researchers of OMR. With all the reasons stated above, this study will be of help to build a better algorithm and to aid musicians and researchers alike.

## Scope and Limitations

This research is only limited to handwritten musical notations without the staff line. Furthermore, this research would only use the symbols found on table 1. In the entirety of this research, only the Neuro-Fuzzy Machine Learning algorithm along with image recognition would be used.

# REVIEW OR RELATED LITERATURE

2

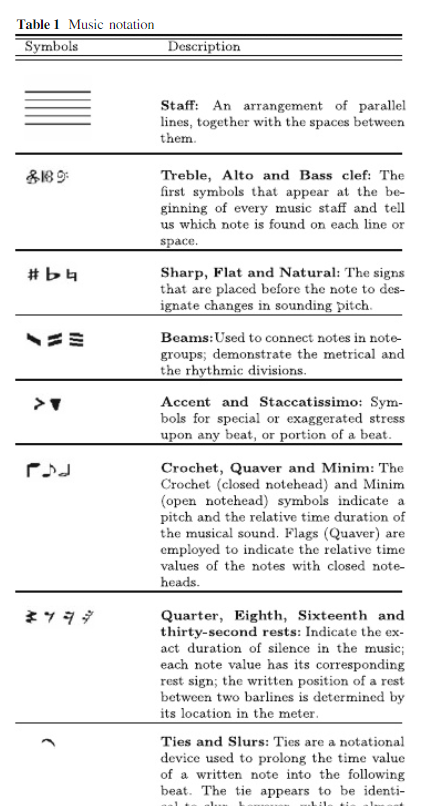
## Optical Music Recognition

In a study conducted by Novotny and Pokorny, they stated that “Optical music recognition (OMR) refers to a discipline that investigates music score recognition systems. This is similar to well-known optical character recognition systems, except OMR systems try to automatically transform scanned sheet music into a computer-readable format.” [[1](#Nov15)]

Optical Music Recognition (OMR) is a form of Optical Character Recognition (OCR) that is used to understand musical scores and convert it to an audio file or an editable file. OMR has been studied for a decade, and a lot of researchers have addressed these problems in different ways and through different algorithms. These methods and algorithms, however, still only scratch the surface of OMR, and researchers are still bothered with plenty of roadblocks. however readable in human perception, can be easily mistaken by the program [[2](#Gon08)].

With the power of machine learning, we can use Optical Music Recognition (OMR) to interpret and understand printed or handwritten scores of music. The images taken will be scanned and using OMR, the musical notation would be recognized so the machine will be able to fully understand the image. In this way, Optical Music Recognition (OMR) is almost the same as Optical Character Recognition (OCR) but OCR proves to be more difficult [[3](#Zen14)].

**Table 1 – taken from** [[4](#Reb09)]



### Contemporary Approaches

In a research conducted by Rebelo, et al. in 2009, they used 4 recognition processes in order to classify both handwritten and printed musical symbols [[5](#Adi14)].

The 4 recognition algorithms used to approach this problem were:

1. Hidden Markov Models
2. K-nearest neighbor
3. Neural Networks
4. Support Vector Machines

Respectively, Hidden Markov Models, even though it is rarely used for OMR, are hidden symbols generated by a stochastic process. The researches planned to use the Hidden Markov Models algorithm because it can be used for recognition and segmentation both at the same time.

The next approach the researchers used was the K-Nearest Neighbor algorithm, which is one of the simplest machine learning algorithms. The algorithm uses Instance-Based Learning to find the nearest neighbor which is classified by votes to the class the object belongs to.

In the third approach they used a multi-layer perceptron neural networks, also known as MLP. This type of a feed-forward network is structured with layers of nodes or units (neurons) and connects or links make up the successive layers in between. Network with K values for each class consisting of 1 for correct and 0 otherwise were used.

The final approach Rebelo, et al. used for their research were Support Vector Machines (SVM).

The main objective of an SVM is to look for a decision boundary found on a plane between the points of two classes in the training data.

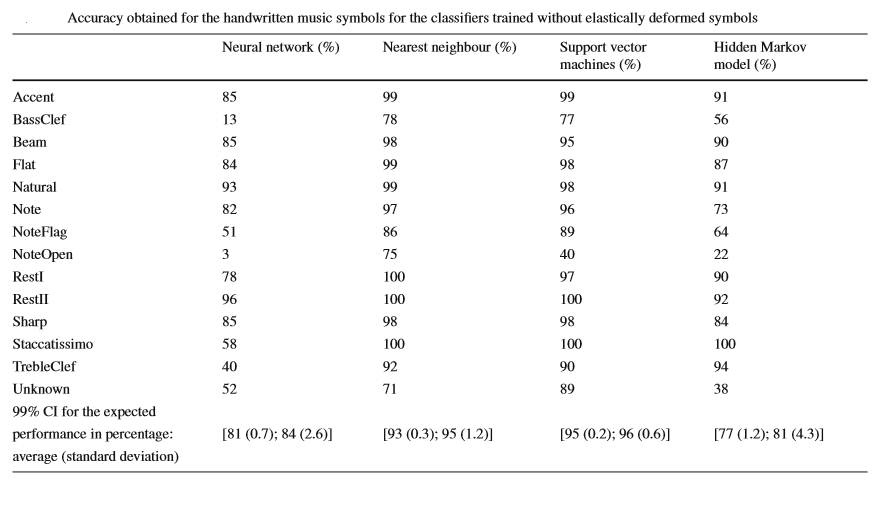
****

Table 3a. Handwritten music symbol accuracy

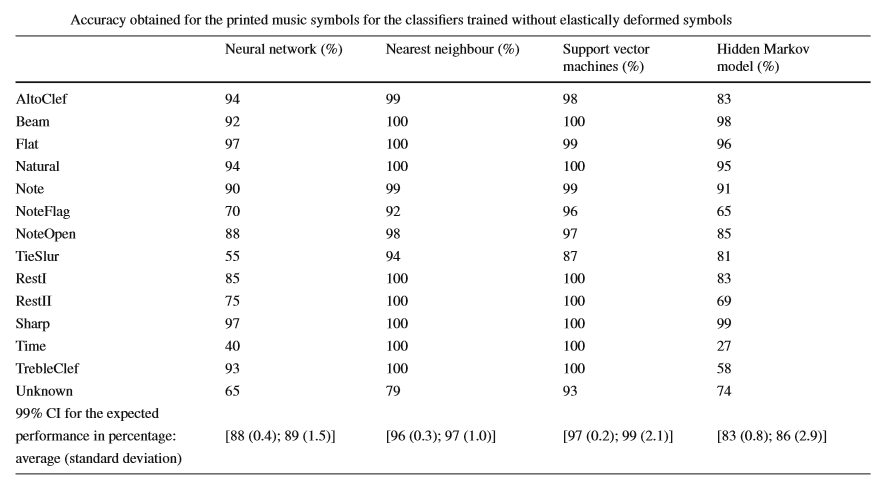
****

Table 3b. Printed music symbol accuracy

## Fuzzy Logic

Based on the mathematical theory of fuzzy sets, Fuzzy Logic was created by Zadeh in 1965 as an extension of Boolean logic which is generalized from the classical set theory of mathematics. The verification by notion of degree was introduced to enable a condition to have a state other than true or false providing flexible reasoning, thus giving Fuzzy Logic a human-like reasoning, e.g. the speed is “too fast or a little slow”. Below is an example of fuzzy logic formalizing human-reasoning: [[6](#Der13)]

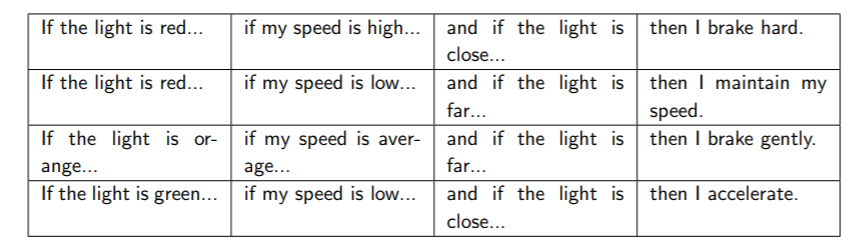


Figure 2.2 Human-like reasoning table by using Fuzzy Logic

## Digital Image Processing

Digital images contain a limited number of elements called picture elements, commonly referred to as pixels, and these pixels each have their own value and location. Digital Image Processing is the act of using digital computers in handling pixels that are grouped by regions. These regions are divided within the process to complete the image processing much easier and more identifiable to a computer. Machines that handle digital image processing can handle a broad span of the electromagnetic spectrum, unlike human vision which is limited only to the visual spectrum. The images that can be processed by these machines can cover rays that even humans cannot perceive with the naked eye. This is a promising fact for the future of digital image processing going forward for it can be used in a wide range of applications that have not yet been fully explored by man [[2](#Gon08)].

According to [[6](#Cha17)], There are 4 phases of Image processing. The first step is the pre-processing of the image. The next step is to segmentation. The third step is feature extraction. And finally, classification and recognition.

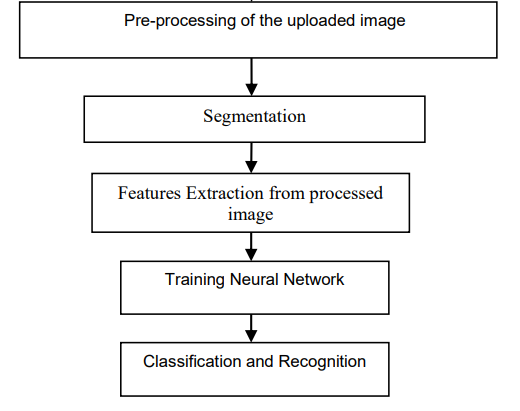


Figure 2.2.1. The four phases of Image Processing.

### Pre-processing

This phase renders the image to prepare it for segmentation. It does that by dissecting the background from the patterns seen in the image. This is done through noise reduction, normalization, and compression. Pre-processing also defines patterns that are closely packed together. It is then turned from a grayscale image to a binary image by means of binarization. Normalization makes the data set uniform. Compression reduces the size of the image [[7](#Dal08)].

#### Binarization

The binarization method is the method responsible for turning a grayscale image to a binary image. This enhances the edges of the image and makes the edges of characters distinct from the background by turning the gray levels of an image (from 0 to 256 gray levels) to binary. This method is the one responsible to make details and edges of characters more distinct [[6](#Cha17)].

#### Noise Reduction

Image noise in Optical Character Recognition adds additional markings on characters such as lines, gaps, marks, etc. Additional alterations to the image include rounding of corners, distortion of the character, dilation, and erosion. All these unwanted markings and distortions of characters would impose problems and it is necessary to correct this flaws before processing the information contained in the image. One of the three major groups of techniques on noise reduction according to Chaudhuri et al. are morphological operations. Morphological operations are designed to correct broken strokes, decompose connected strokes, remove random points, extract boundaries, etc. Its purpose is to remove noise from images that have poor quality of detail due to erroneous hand movements made when writing the input data [[7](#Dal08)].

##### Staff Line Removal

Staff lines are one of the most prominent features of sheet music. This represents a pitch the musician must play and the time and speed it has to be played. This staff line, however, creates an augmentation to the symbols this study aims to detect, therefore it has to be temporarily removed. One of the many approaches of staff line removal is based on morphological operations relying on relative ordering of pixel values. However, this is not determined as the best based on the study of Dalitz et al. They included in their study about the comparison of staff line removal algorithms is that there is no clear best algorithm to be used in removal of the staff lines [[8](#Win11)].

#### Normalization

Normalization aims to remove different instances of writing that may be caused by erroneous handwriting or image digitalization problems to obtain a uniform data set. Two of the four groups of approaches mentioned by Chaudhuri et al. are skew normalization and baseline extraction, as well as size normalization. Skew normalization and baseline extraction aims to correct the errors caused curving of the image. The size normalization aims to standardize the size of the character to be scanned. Size normalization may be used for both horizontal and vertical standardization. It is used for scaling and is helpful on adjusting the data to be processed so that the whole system would yield less errors in detecting the character [[7](#Dal08)].

#### Compression

As the word implies, compression aims to reduce the size of the of the size of the images. Large images are not suitable for Optical Character Recognition due to it needing a longer amount of time in the processing stages. Reduction of the data set’s size (while also preserving the shape information) will positively affect the speed of OCR. An approach for compression according to Chaudhuri et al. would be thresholding. Thresholding aims to reduce storage requirements by adding a threshold in binary images. Global thresholding assumes a threshold value basing on the estimated intensity difference of the characters and background of the binarized images. Local thresholding, on the other hand, assumes a threshold value basing on each pixel of a given local area [[7](#Dal08)].

### Segmentation

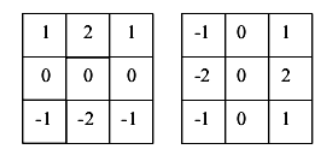
Segmentation is the process of the isolation of the individual character images from the refined image. This is considered as the area of the study having the most problem. [[8](#Win11)] this phase, the image sequence of characters is fragmented into sub-images of individual character. The pre-processed input is also fragmented into isolated characters, labeling each character with numbers. Numbering of images serves as the indicator of the number of characters in an image. After fragmentation, the characters are fit into a defined number of pixels. The size of the characters is then normalized due to the characters having varied sizes. The purpose of this is to prepare the image for feature extraction [[7](#Dal08)].

### Feature Extraction

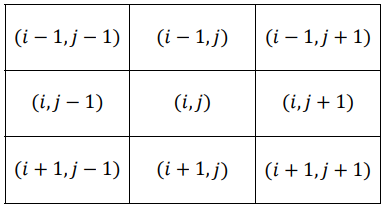
Feature extraction aims extract informative and non-redundant features from a data set. The redundant representative of pixels in an image is aimed to be reduced to a set of features, also called a feature vector. Feature selection would then determine a subset of initial features. This stage would then be better processed if the data is being represented by only relevant information as compared to the initial pre-processed data. One of the most difficult pattern recognition problems, the main goal of feature extraction is to obtain the vital features of symbols [[9](#Ful01)].

#### Gradient Feature Extraction

Magnitude and direction of the greatest change in intensity in a small neighborhood of each pixel is what is measured by Gradient Feature Extraction. Gradient Feature Vector is formed by combining the strength of a gradient per individual direction. Computing the horizontal and vertical components of the gradient is done by means of the Sobel operator, which is usually used in edge detection [[5](#Adi14)]. The following figures shown below are image representations for the Sobel technique [[6](#Cha17)] and neighborhood of pixel.



Horizontal Component Vertical Component



### Classification

#### Neural Networks

The design of neural networks works like a human brain. According to Vieria,et al., neural networks can be trained with examples using vectors, inputs and outputs of a system.

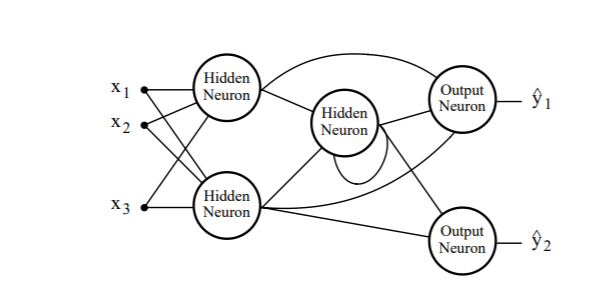


Figure 3. The general structure of an artificial neural network with 3 inputs and 2 outputs [[10](#Lar99)]

#### Fuzzy Neural Network

Neuro-Fuzzy networks are a combination of fuzzy logic and neural networks. The hybridization of these two method makes these two different approaches more efficient. Fuzzy logic is designed to handle non-Boolean information, which lie somewhere within 0 and 1. Therefore, fuzzy logic is an efficient way of processing information which data is good at explaining decisions. However, fuzzy logic’s weakness is not being able to automatically obtain the rules used to make said decisions. Neural networks, on the other hand, is designed to recognize certain patterns, depending on the rules it was given, however it does not explain very well how it can obtained decisions. The strengths of the two approaches are each the other’s weakness, therefore having to combine these two to make a hybrid system would eliminate previous boundaries that the approaches have individually [[11](#Nau97)].

There are 3 types of Neuro-Fuzzy systems, Cooperative Neuro-Fuzzy Systems, Concurrent Neuro-Fuzzy Systems, and Hybrid Neuro-Fuzzy Systems. Cooperative Neuro-Fuzzy Systems are used primarily on initial phases. Neural networks in this type use training data to divide fuzzy systems by sub-blocks. After this part, the neural network is not anymore used, and the fuzzy system is executed. Concurrent Neuro-Fuzzy Systems are a type of neuro-fuzzy system which is only classified as a neuro-fuzzy system loosely due to it having both approaches work together. Either the input is entered and pre-processed in the fuzzy system, and it is being processed and outputted in the neural network, or vice versa. [[12](#BUc95)]Hybrid Neuro-fuzzy, according to Nauck [[13](#Kru08)] “a fuzzy system that uses a learning algorithm based on gradients or inspired by the neural networks theory (heuristical learning strategies) to determine its parameters (fuzzy sets and fuzzy rules) through the patterns processing (input and output)”. Hybrid systems are composed of fuzzy logic, neural networks, genetic algorithms, and expert systems, proving their efficiency in facing a number of problems. Specific to every method are unique computational abilities that make them efficient for certain fields. However, these specific abilities also mean that a method is only able to do fairly limited tasks. [[9](#Ful01)].

Although there are a lot of different approaches, neuro-fuzzy system is the term used for approaches which display similar properties. A Neuro-Fuzzy system is a system that primarily uses a combination fuzzy logic and is being trained by a neural network-derived algorithm. A Neuro-Fuzzy system can be viewed as a neural network with three layers. The first layer is the input layer, the second layer represents an area where fuzzy rules are applied, and the third layer which represents the output of the fuzzy rules [[14](#Pal)].

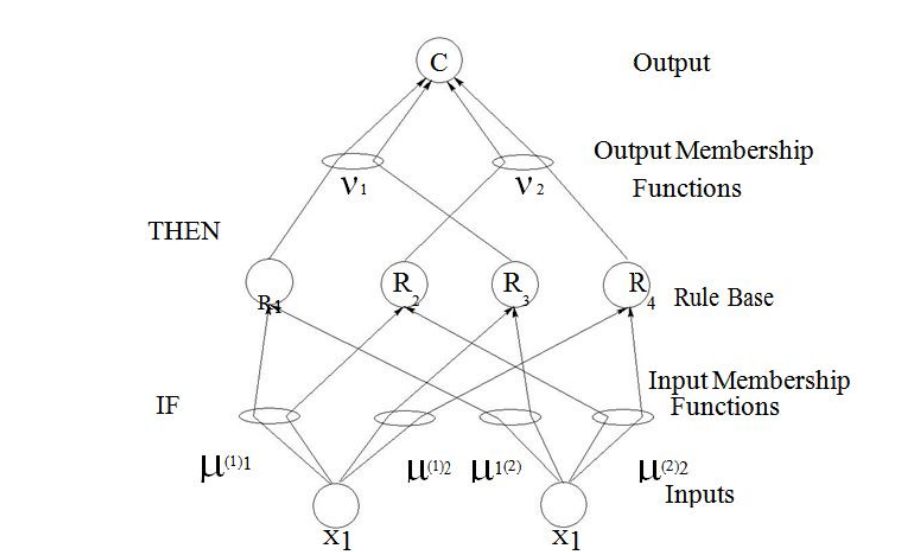


FIGURE 2a. Architecture of four rule fuzzy controller from neural networks point of view [[13](#Kru08)]

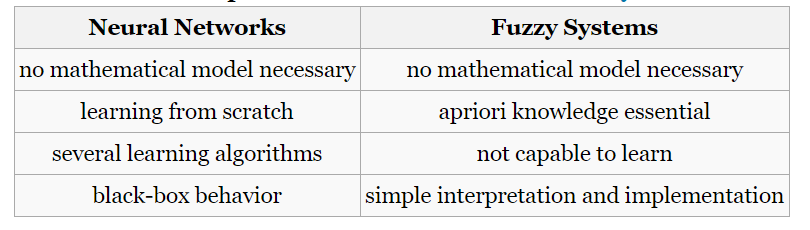
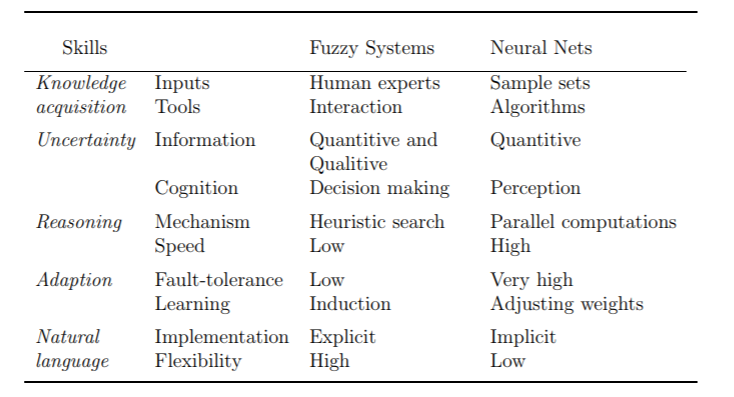
Table 2: Fuzzy System and Neural Networks Comparison table [[15](#LAZ65)].

TABLE 2a: Neural Networks and Fuzzy Systems Comparative Study. [[9](#Ful01)]



##### Fuzzy System

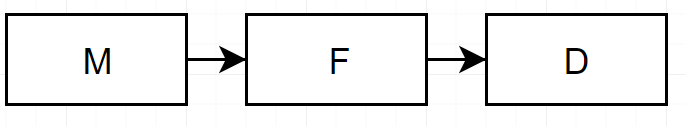
Mathematical calculus is used by fuzzy systems to manipulate knowledge with a degree of uncertainty. Using the Fuzzy set theory initiated by Zadeh, Lofti in 1965, [[10](#Lar99)] a Fuzzy System’s behavior could be described as:

**IF <premise> THEN <consequent>**

This system uses linguistic variables combined with symbolic terms each representing a fuzzy set which consists of three stages [[16](#Vei04)]:First, using fuzzification to map the values of the input to a certain level of compatibility to the respective fuzzy sets then processing of the rules so that the Fuzzy System will know the strength of each individual input, and finally, using defuzzification to transform the fuzzy value results into their numerical values.

According to the study of Vieria, et al., [[16](#Vei04)] fuzzy systems have the following advantages and disadvantages. Fuzzy systems have capacity of representing uncertainties with linguistic variables while having easy interpretation of results and addition of new rules make extension of base knowledge easier. Fuzzy Systems also have disadvantages such as only knowing how to answer what’s written on its rule base, changing the system topology would need the rule base altered, an expert is needed to determine the inference logical rules.

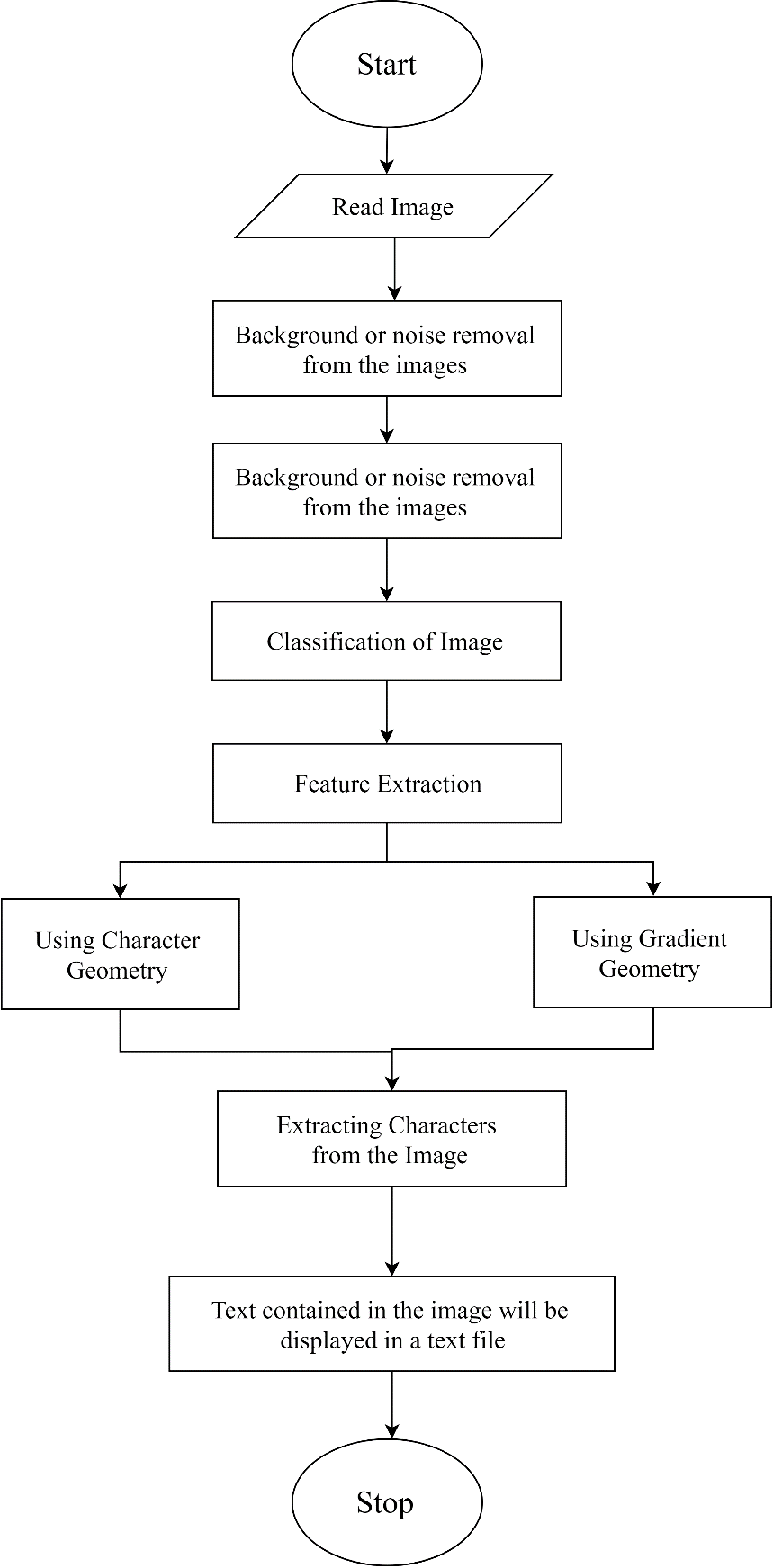
###### Neuro-Fuzzy for Image Processing and Pattern Recognition

The task of pattern recognition involves the computer transforming from the measurement space M to the feature space F, and finally, to the decision space D [[14](#Pal)] :

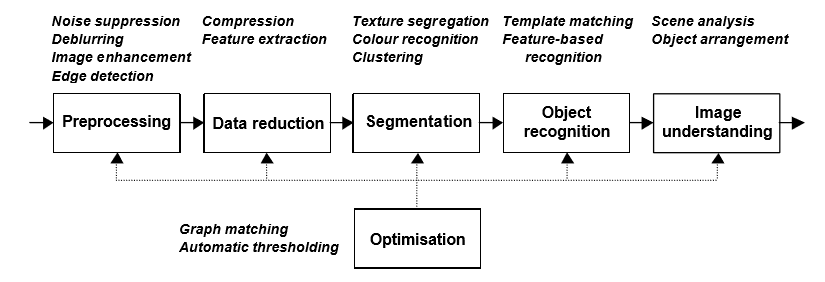
The measurement space involves image manipulations like noise reduction, filtering, enhancement, skeleton extraction, and contour extraction in order to extract features. This is commonly known as image processing. If the processed information is recognized and understood, then the complete system would then be called vision. [[2](#Gon08)]

## Theoretical Framework

Using a neural network in optical image recognition [[5](#Adi14)]



In this research, a photographed handwritten text is fed to the program. The objective of the program aims to identify all handwritten letters and convert it into its computerized form and put it in a .doc file. The image underwent 4 phases: Pre-processing, segmentation, feature extraction, and classification and recognition.

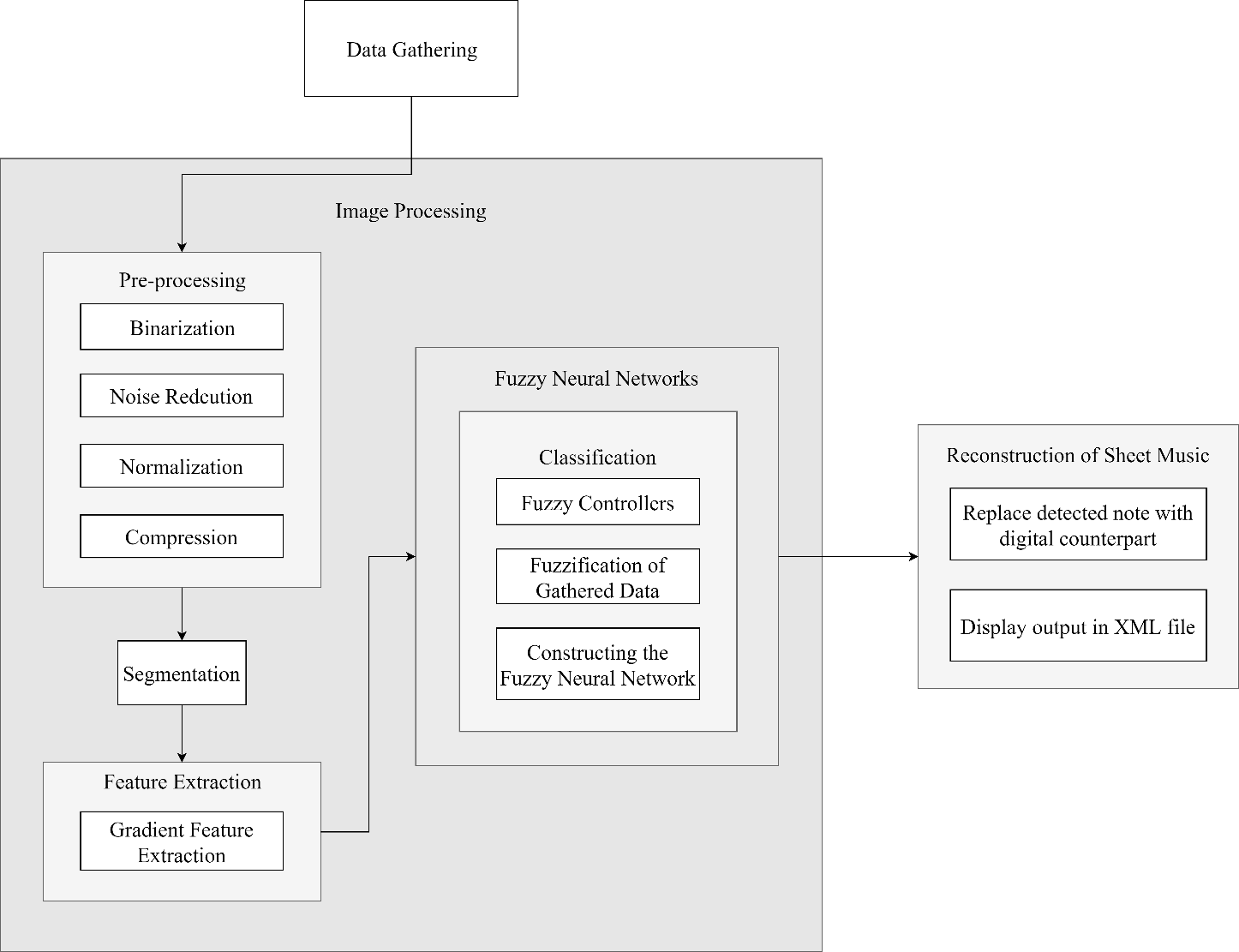


In this research, the researchers created a framework containing five different tasks: preprocessing for image enhancement and edge detection, data reduction for compression and feature extraction, segmentation for texture segregation and clustering, object recognition for feature-based recognition, and finally, image understanding for scene analysis and object arrangement. The researchers used this framework for image recognition using neural networks. [[17](#Egm01)]

# RESEARCH DESIGN AND METHOdology

3

## Conceptual Framework



## Methodology

In this study, the phases are as follows:

1. Data Gathering
2. Image Processing
   1. Pre-Processing
      1. Binarization
      2. Noise Reduction
      3. Normalization
      4. Compression
   2. Segmentation
   3. Feature Extraction
      1. Gradient Feature Extraction
   4. Classification
      1. Fuzzy Systems
         1. Fuzzy Controllers
         2. Fuzzification
         3. Constructing the Neuro Fuzzy Network
3. Reconstruction of Music Sheet
   1. Replace detected note with digital counterpart
   2. Display output in XML file

### Data Gathering

The data gathering for this study is done by having the researchers give participants papers containing musical symbols on staff lines, and papers for them to copy the symbols on. Half of the papers will contain a blank musical staff while the other half would be blank pieces of papers. The aim of this is to have a wide range of different handwritten symbols to make the system learn and determine what the handwritten symbols’ digital counterparts are.

### Classification by Image Processing

In this step, the data set will undergo many operations relating to correction for it to be refined and ready for classification.

#### Pre-processing

This first step of image processing would oversee the removal of additional unwanted symbols and disfigurements of the image to prepare it for segmentation.

##### Binarization

Binarization is the first step the researchers will take in pre-processing the data set. What this does is alter the image, having it convert a grayscale image to a binary image, enhancing its details and leaving symbols on paper distinct.

##### Noise Reduction

The next step in pre-processing is noise reduction. In this step, removal of unwanted markings occurs. This step specifically aims to separate the character and the staff through staff line removal. This makes the symbol much more readable than it was before.

##### Normalization

The third step in pre-processing is normalization. In here, the data set will be normalized in terms of distortion due to curvature of the image by skew normalization and baseline extraction. The data set will also undergo size normalization to standardize the dimensions of each detected symbol

##### Compression

The fourth and last step in pre-processing is compression. In this step, images will be compressed in terms of size of the image (in bytes) for it to be significantly efficient. This is to be done by thresholding.

### Segmentation

After pre-processing the data set, it would then undergo segmentation. Here, both the reference images and the data set will now be divided in sub-images. These segments are then fit into a defined number of pixels. This would cause the symbols to normalize due to it having varied sizes even after pre-processing.

### Feature Extraction

Next phase is feature extraction. This phase separates the background areas of the data set and the areas of the data set where there are significant features. The feature selection would now determine a subset of those significant features, obtaining only the vital features of the symbols. This is done by gradient feature extraction where the magnitude and direction with the greatest change in intensity in a small neighborhood is measured. A vector will then be formed by combining the strength of a gradient per individual direction.

### Fuzzy Systems

The rule base of the fuzzy system here will be interpreted as a neural network. The fuzzy sets fed to the system will be regarded as weights and neurons are modeled from the given input and output of the sets. This combined fuzzy logic and neural networks will be called a Hybrid Fuzzy Neural Network. In this step, the classified data will undergo the Fuzzy System.

#### Fuzzy Controllers

Linguistic terms of the fuzzy rules will be defined to express membership functions.

#### Fuzzification of Gathered Data

From the gathered data, the researchers will use a Fuzzy Inference System such as that of FisPro to create the fuzzy rules needed to tune and modify the Neuro-Fuzzy Network.

#### Constructing the Neuro-Fuzzy Network

The proponents of the research will follow the steps done by Babuska R. [[17](#Bab)] With two approaches, gathered data from classified images and prior knowledge will be used to create the neuro-fuzzy system If-then fuzzy rules derived from fuzzy sets will create expert knowledge to be used by the neural network. A layer of the neural network will be the rule base and will be fine-tuned by fuzzy logic.

### Reconstruction the musical notation symbol

After successfully classifying/determining a symbol, the program will then replace the handwritten symbol with its digital counterpart that was taken from the database with digital symbols on where the symbol is originally located in the staff and placed on an XML file. However, if the origin data set of the symbol does not have a staff, it would just be placed on an XML file without a staff.

# References

|  |  |
| --- | --- |
| 1 | Novotny, Jiri and Pokorny, Jaroslav. *Introduction to Optical Music Recognition: Overview and Practical Challenges*. Prague, 2015. |
| 2 | Gonzales, Rafael and Woods, Richard. *Digital Image Processing*. Pearson Education, Inc., Upper Saddle, 2008. |
| 3 | Zeng, Andy. *Optical Music Recognition*. Princeton University, Princeton, 2014. |
| 4 | Rebelo, A. and Capela, G. *Optical Recognition of Music Symbols*. Springer-Verlag, 2009. |
| 5 | Adi, C. *Hand Written Character Recognition Using Neural Network*. 2014. |
| 6 | Chaudhuri, A., Mandaviya, K, Badelia, P., and K Ghosh, S. *Optical Character Recognition Systems for Different Languages with Soft Computing*. Springer International Publishing AG, Basel, Switzerland, 2017. |
| 7 | Dalitz, Christoph, Droettboom, Michael, Czerwinski, Bastian, and Fujinaga, Ichiro. *A Comparative Study of Staff Removal Algorithms*. IEEE, 2008. |
| 8 | Win, H.P.P., Khine, P.T T. , Tun, K.N.N. *Bilingual OCR System for Myanmar and English*. International Journal of Engineering and Computer Science, 2011. |
| 9 | Barve, S. *Optical Character Recognition Using Artificial Neural Network*. International Journal of Advanced Research in Computer Engineering & Technology, 2012. |
| 10 | Fuller, Robert. *Neuro Fuzzy Methods*. Abo Akademi University, Turku, 2001. |
| 11 | Veira, Jose, Fernando, Dias, and Alexandre, Mota. *Neuro-Fuzzy Systems: A Survey*. Portugal, 2004. |
| 12 | Nauck, Detlef, Klawonn, Frank, and Kruse, Rudolf. *Foundations on Neuro-Fuzzy Systems*. John Wiley & Sons, Inc., New York, 1997. |
| 13 | BUckley, James K., Youchi, and Hayashi. *Neural networks for fuzzy systems, Fuzzy Sets and Systems*. Elsevier B.V., Amsterdam, 1995. |
| 14 | Krus, R. *Fuzzy Neural Network*. Institute for Information and Communication Systems, Germany, , 2008. |
| 15 | Pal, S. and Ghosh, A. *Neuro-fuzzy Computing for Image Processing and Pattern Recognition.* Taylor and Francis Ltd., England. |
| 16 | L. A., Zadeh. *Fuzzy Sets; Information and Control*. California, 1965. |
| 17 | Larsen, Jan. *Introduction to Artificial Neural Networks*. Technical University of Denmark, Denmark, 1999. |
| 18 | Rebelo, Ana, Capela, G., and Cardoso, Jaime S. *Optical recognition of music symbols*. Springer-Verlag , Berlin, 2009. |
| 19 | Pacha, Alexander and Eidenberger, Horst. *Towards Self-Learning Optical Music Recognition.* Interactive Media Systems, Vienna, 2017. |
| 20 | Savic, Bojovic. *Training of hidden Markov models for cursive handwritten word recognition*. Proceedings of the International Conference on Pattern Recognition, Washington, DC, 1973. |