# APPLYING MACHINE LEARNING (THE K-MEANS ALGORITHM) TO CLUSTERING AND ANALYZING SYNOVIAL FLUID CONTENTS AMONG DIFFERENT AGES AND GENDERS IN HEALTHY AND OSTEOARTHRITIS PATIENTS

by

BADER EID ALABKARY

A dissertation submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN ELECTRICAL AND COMPUTER ENGINEERING

2024

Oakland University
Rochester, Michigan

Doctoral Advisory Committee:

Mohamed A. Zohdy**,** Ph.D., Chair.
Hoda Soliman Abdel-Aty-Zohdy**,** Ph.D.
Suzan Kamel-ElSayed**,** Ph.D.
Christopher Kobus**,** Ph.D.

*Many thanks are due to*
*ALLAH (God)*

*To my lovely wife Dr. Galelah Alawami and sons*
*Ahmed, Ali, Adam, and Ameer*

*To my parents*
*especially my mother*
*And to brothers and sisters*

# ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Professor /Dr. Mohamed A. Zohdy, for his invaluable support and guidance throughout my research, as well as for giving me the opportunity to explore and contribute to my field of interest.

I am also sincerely thankful to my doctoral advisory committee: Professor /Dr. Hoda Soliman Abdel-Aty-Zohdy, Dr. Suzan Kamel-ElSayed, and Dr. Christopher Kobus. I greatly appreciate their time, effort, and willingness to serve on my committee, as well as their valuable feedback and guidance, which significantly improved the quality of my work.

I owe my deepest thanks to my wife Dr. Galelah Alawami and my sons Ahmed, Ali, Adam, and Ameer for their unwavering support and encouragement throughout this journey, helping me to accomplish my goals.

I would like to extend my heartfelt thanks to my mother and sisters for their constant encouragement and for helping me stay focused during my studies.

Finally, I am immensely grateful to my dear friends Abdulghani Alabkari, Ahmad Almuhaimeed, Salamah Alaqool, Ali Sonbul, Ahmad Safar, Jafar Abdrabulrasul, Baqer Aljabr, Talal Omar, as well as to Qatif General Hospital, Hani AlKhawajah from King Fahad University Hospital, for their generous support and encouragement throughout my dissertation work.

Bader Eid Alabkary

# ABSTRACT

APPLYING MACHINE LEARNING (THE K-MEANS ALGORITHM) TO CLUSTERING AND ANALYZING SYNOVIAL FLUID CONTENTS AMONG DIFFERENT AGES AND GENDERS IN HEALTHY AND OSTEOARTHRITIS PATIENTS

by

Bader Eid Alabkary

Adviser: Mohamed A. Zohdy, Ph.D.

Machine learning, a subset of AI, has made a significant impact on the medical field by improving the speed and accuracy of test results. Among the many discrete ML tools, k-means is a type of data clustering that uses unsupervised ML to divide unclassified data into different groups with similar variances. This dissertation applied the k-means clustering algorithm to analyze synovial fluid compositions of healthy people and osteoarthritis (OA) patients, focusing on four components: hyaluronic acid (HA), chondroitin sulfate (C6S, C4S), and the C6S ratio. The main objective was to identify distinct patterns and clusters within these datasets based on age and gender. Data was extracted from two previously published research studies. The first dataset comprised 187 healthy participants, with ages ranging from 10 to 90 years. The second dataset consisted of 133 OA participants with ages ranging from 55 to 90 years. Applying ML algorithms, specifically k-means clustering, the MATLAB program was used for data analysis. The findings showed the k-means clustering successfully highlighted age- and gender-related synovial fluid concentration patterns. In addition, for both healthy and OA

groups, younger people had higher levels of synovial fluid components, which decreased with age. In healthy people, HA levels were high among younger people but decreased with age. In the OA group, HA levels increased in older patients. These findings confirmed the potential of synovial fluid concentration in diagnosing joint health. These findings also asserted the utility of ML techniques, such as k-means clustering, in medical data analysis.

TABLE OF CONTENTS

Table of Contents—Continued

Table of Contents—Continued

Table of Contents—Continued

Table of Contents—Continued

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES—Continued

# CHAPTER ONE

# INTRODUCTION

## 1.1 Background knowledge

The modern world has seen exponential growth of technologies with almost unimaginable capabilities. Many machines today perform tasks that have replaced human labor. The quality of such work has also improved over the years [1]. Artificial intelligence (AI) is one of the latest inventions that has revolutionized the world. It was first introduced by John McCarthy in 1956, who asserted that AI could reproduce human intelligence through computers [2]. McCarthy further argued AI may be considered as the science and engineering of producing intelligent computers [3].

Machine learning, or ML, is also a type of artificial intelligence. Al Turing introduced the first application of machine learning in the 1950s when a computer was programmed to learn and become artificially intelligent with the aim of accelerating efficacy and reducing public transportation risks [4]. During the 1980s and 1990s, there was a rise in interest in AI. Particularly, fuzzy expert systems, Bayesian networks, artificial neural networks, and hybrid intelligent systems found applications in numerous healthcare settings [5].

It is important to mention here that the medicine being practiced today was an outcome of experimentation and careful data analysis. Therefore, the capability of AI to process huge amounts of training data can be very useful for data analysis [1]. That is the reason medical research has benefitted so much through the introduction of AI, as this field relies heavily on data analysis techniques to explain patterns within patient datasets

[6]. Since examining big data can be very difficult, AI offers a viable option in data analysis, with its machine-like speed and human-like understanding. In other words, AI can comprehend almost all available data [1]. Thus, it makes sense that healthcare applications, as compared to other sectors, made the most investment in research about AI in 2016 [5].

To date, medical science has utilized ML at higher levels to get better predictive capabilities [6]. Machine learning is a type of AI that is different from its standard statistical model. ML not only interprets data but also makes reproducible and precise predictions [7]. Furthermore, ML applies specific tactics to navigate the complex relations between several variables to anticipate a certain result. ML can easily process convoluted variable relations as compared to typical statistical procedures [6]. Additionally, ML technologies have helped doctors as their analysts and assistants. ML has helped to identify medical trends and establish disease prediction models [4]. ML algorithms have also helped in analyzing Electronic Health Record (EHR) data, healthcare claims, and other areas such as hospital readmission and the occurrence of chronic illnesses to anticipate the probability of particular health outcomes. ML can be highly beneficial in helping doctors identify high-risk patients and take preventive measures to avoid severe outcomes for patients [8]. Moreover, ML can assist in diagnosing and treating as it can learn to interpret medical images, including X-rays and CT scans, helping make accurate diagnoses [8].

In healthcare settings, ML technology has led to important discoveries as well. For example, AI has aided critical improvements in joint diseases such as rheumatoid arthritis (RA), osteoarthritis (OA), and Anterior Cruciate Ligament (ACL) [9-10]. ML has

enabled doctors to comprehend pain in new joints caused by the above-mentioned diseases is due to stiffening joint fluids (called synovial fluids). In the context of humans and animals, synovial fluid markers are typically referenced in patients who have joint diseases such as RA and OA. Using AI tools, it is easy to find the total joint biomarker by examining the volume of joint fluid [11]. This total volume of synovial fluid can be measured using several radiology tests and CT scans [11]. At times, a basic blood test and X-ray may suffice. Moreover, rigorous tests such as MRI, synovial fluid test, and ultrasound can be performed on the affected area [11]. It is also important to note standard viscometers are employed by labs to perform tests which tend to be bulky and expensive. These may make mistakes and are not 100% foolproof because the fluid is manually drawn from the joint and is incompatible with the actual medical environment [12-13.] ML presents itself as the solution, especially in the case of RA and OA through the analysis of synovial fluid. Combined with multi-molecule synovial fluid analysis, ML can be used to accurately measure the knee's biomarker plural [14].

ML can also be useful for discovering hidden patterns and biomarkers in synovial fluids. A k-means clustering ML algorithm can be utilized to garner complicated patterns from the data of patients. On top of that, ML does not need any specialized programming, as it inherently understands various algorithmic and statistical models [4]. K-means may be considered one of the important ML procedures due to its established partitional clustering algorithm. This is all the more popular due to its simple, straightforward implementation and efficacy [15]. Synovial fluid research powered by ML k-means Clustering and elbow methods is critical in decoding complicated patterns efficiently. While the k-means algorithm offers information about the categorization of synovial fluid

content on the basis of identical retention levels, the Elbow approach offers a more refined process to determine standard cluster configurations.

## 1.2 Purpose of The Study

A viscous material called synovial fluid is found in the chambers of synovial joints. Its primary purpose is to reduce friction between the articular cartilages of synovial joints when in motion [16]. The characteristics and makeup of synovial fluid vary greatly with age, potentially affecting the health and operation of joints. The quality and content of modified hyaluronan acid in synovial fluid may be connected to the age-related increase in cartilage breakdown in non-osteoarthritic joints [17]. Understanding how aging affects knee synovial fluid is essential for knee-related disease diagnosis, treatment, and prevention. There are many approaches available for laboratory analysis of synovial fluid data, especially with the advancements in computer methods such as machine learning. Generally speaking, ML offers faster data representation than biological data analysis, which allows it to provide engineering solutions and act as an important standard [18]. Apart from age, gender is another factor that affects the amount of synovial fluid in healthy people. Due to increasing incidences of joint diseases in elderly people, especially in older women, it is essential to recognize and distinguish alterations in chondroitin sulfate CS isomers related to joint pathology from normal, age-related phenomena [19].

To analyze the data of factors such as these, an important method in ML is to group or classify data into a set of clusters or categories. Clustering aims to discover finite natural groupings with hidden or latent structures from a set of finite unlabeled data [20]. K-means is a well-known partitional clustering algorithm. Its simplicity, easy

implementation, efficiency, and empirical success have contributed to its popularity. Many practical problems can be solved using k-means [20]. With the growing use of ML and AI across different domains, there are many extant studies exploring ML approaches in the health and engineering fields. Those focused on synovial fluid primarily use statistical methods to analyze data. However, there is limited data that integrates biological results with the engineering field. There are fewer still studies using ML (k-mean algorithm) to interpret synovial fluid statistics. Therefore, it is necessary to conduct further studies on the efficiency of k-mean clustering for analysis of synovial fluid data.

Thus, the main purpose of this study is to evaluate the effectiveness and performance of ML (k-means clustering algorithm) in analyzing synovial fluid compositions—specifically hyaluronic acid (HA), chondroitin sulfate (CS), chondroitin 6-sulfate (C6S), and chondroitin 4-sulfate (C4S)—across different ages and genders in both osteoarthritis (OA) patients and healthy patients. Also, this study aims to provide a comprehensive comparison of different ML (k-means clustering) algorithms as applied to datasets derived from the synovial fluid of OA patients containing HA, CS, C6S, and C4S. The ultimate goal is to gain a better understanding of the physiological differences in synovial fluid composition among people of different ages, which could lead to improved diagnoses, treatment, and management of joint-related conditions.

## 1.3 Significance of The Study

This study is important in the light of how critical AI has become today in the health and engineering fields. The researchers hope the results provide essential objective information regarding the investigation of the application of ML techniques, specifically the k-means clustering algorithm and elbow method to cluster and analyze synovial fluid

contents among people of different ages and genders. The study will be beneficial for health providers by offering the possibility of improvements in the diagnoses and management of joint-related conditions. Hospitals can promote their diagnostic accuracy and treatment selection for patients with joint diseases by using ML techniques to analyze synovial fluid contents. It is also significant for professionals and engineers who work in biomedical and computational fields as they can design and implement innovative technological solutions that integrate ML algorithms with medical devices for joint disease. Finally, the details emergent from this study provide insights for future researchers who may explore newer applications of ML techniques in healthcare.

## 1.4 The Research Questions

The following research questions will be addressed in the study:

**RQ1: How does the k-means clustering algorithm expose distinct clusters within datasets that contain information found in the compositions of synovial fluid (HA, C6S, C4S, and C6S: C4S) in patients of all ages and genders who have osteoarthritis (OA) and healthy people?**

**RQ2: What are the characteristics of the clusters identified based on the concentrations of hyaluronic acid (HA), chondroitin sulfate (CS), chondroitin 6-sulfate (C6S), and chondroitin 4-sulfate (C4S) in the synovial fluid of osteoarthritis patients (OA) and healthy people?**

The purpose of these research questions is to investigate specific aspects related to the clustering and analysis of synovial fluid contents among different age groups and genders, with a focus on HA, CS, C6S, and C4S for healthy people and OA patients. The first step would be to investigate the effectiveness of integrating the database from

medical research with computer engineering. Thereafter, the researcher will investigate how applying different approaches to extant databases may contribute to methodological advancements in the engineering field. Third, an evaluation will be conducted to check the suitability of k-means clustering algorithm and elbow method for segregating synovial fluid data into meaningful clusters, as well as how age- and gender-related differences impact the distribution of synovial fluid contents, particularly HA, CS, C6S, and C4S. Finally, this study aims to determine the optimal number of clusters using the elbow method, assess the stability and reliability of clustering results across different age groups, and then compare the results of k-means clustering among synovial fluid HA, CS, C6S, and C4S for a healthy group as well as OA patients.

### 1.4.1 Hypotheses:

H1. The k-means clustering algorithm will identify distinct patterns in synovial fluid contents that differentiate OA patients from healthy individuals.

H2. Age and gender will significantly influence synovial fluid contents.

H3. Machine learning (k-means clustering) algorithms will exceed traditional statistical methods in identifying patterns for OA diagnosis.

## 1.5 Definition of Terms

**Machine Learning:** ML is a type of AI that utilizes computing algorithms that are flexible and are established with experience. Supervised and unsupervised algorithms are the primary forms of ML [2]. While supervised learning incorporates the training of an algorithm for forming relationships with the help of existing labeled data, unsupervised algorithms can generate complex new connections within unlabeled datasets [8].

**Clustering:** Cluster analysis is a method that categorizes data by examining the underlying structures of items within a dataset and their connections [21].

**K-means:** Unsupervised machine learning method k-means data clustering divides unlabeled data into a fixed number of unique groups with uniform variances. These entities, which are defined by common traits, are called clusters [22].

**Elbow method:** The optimal number of centroids (k) in the k-means clustering technique is determined using the elbow technique. This approach facilitates the identification of the optimal number of clusters [23].

**Synovial fluid**: The viscous material found in the cavities of synovial joints is called synovial fluid. Reducing friction between the articular cartilages of synovial joints when in motion is the main purpose of synovial fluid [16].

**Chondroitin sulfate (CS):** Synovial fluid contains chondroitin sulfate a glycosaminoglycan (GAG), which is a major component of the extracellular matrix in connective tissues. It is an extended sequence of repeating sugar molecules [24].

**Chondroitin-6-sulfate (C6S):** A subtype of chondroitin sulfate, CS6 is found in cartilage, connective tissues, and synovial fluid, among other tissues. The structure of each isomer of chondroitin sulfate, such as C4S (chondroitin-4-sulfate) and C6S, varies due to changes in the location and number of sulfate groups attached to the sugar units [24].

**Hyaluronic acid (HA):** HA reduces inflammation and preserves the natural cartilaginous matrix, which helps to maintain increased fluid viscosity and the integrity of joints [25].

## 1.6 Overview of Dissertation Structure

**In chapter One,** I present an introduction to ML in health research and the importance of synovial fluid analysis using machine learning (k-means clustering).

**Chapter Two** consists of a review of relevant literature on topics such as Synovial Fluid Analysis: Clinical Significance and Challenges (part 1), and Machine Learning Applications in Biomedical Research and Clustering Techniques in Synovial Fluid Studies (part 2).

**Chapter Three** delineates the methodological approach used for this study. I pay particular attention to the rationale behind using a k-means and elbow clustering approach to examine and analyze synovial fluid contents (C6S, C4S, C6S:C4S, and HA) among OA and healthy groups of different ages and genders. In addition, chapter three explains the k-means design, study participants, methods for data collection, data storage, and data analysis, as well as the strategies for validating the database. In this chapter, I also address the research ethics of this study.

**In chapter Four,** the collected data is analyzed in detail. The result of k-means and elbow clustering on synovial fluid contents (C6S, C4S, C6S:C4S, and HA) for OA and healthy groups of different ages and genders are also shown.

**Finally, in chapter Five,** I summarize the research study and discuss the data and results. I also note the implications and limitations of this study, as well as recommendations for future research. I conclude the dissertation with a comprehensive list of references and appendices.

# CHAPTER TWO

# LITERATURE REVIEW

## 2.1 Introduction

This chapter provides a detailed overview of synovial fluid analysis and the application of machine learning in the medical field and comprises two sections. The first part, Synovial Fluid Analysis: Clinical Significance and Challenges, is further divided into seven parts for ease of understanding. These are A: Properties of synovial fluids; B: Previous research focused on synovial fluids and joint health; C: Previous studies on chondroitin sulfates and hyaluronic acid in synovial fluid, D: Finding out how thick synovial fluids are; F: How the make-up of synovial fluids changes with age; and G: Researching synovial fluid with people who have an anterior cruciate ligament (ACL). The second part focuses on Machine Learning, and it is divided into three sub-parts, A: Machine learning applications in medical research; B: Clustering techniques for medical data analysis; and C: Using K-means clustering in synovial fluid studies.

## 2.2 Synovial Fluid Analysis: Clinical Significance and Challenges

Synovial joints are the most common type of joints in the human body [26]. These joints are found in different areas of the human body, such as the knees, waist, ankles, shoulders, and hips. Synovium tissue lines all human joints, except the weight-bearing ones [27]. These joints are provided with articular cartilage, a thin layer of hyaline cartilage covering the entire articulating surface of each bone, which prevents friction between the bones [28]. In contrast to cartilaginous joints, the articular cartilages of synovial joints are not physically linked. Instead, they function as a Teflon-like covering, facilitating frictionless movement between the bones while protecting the underlying

bone tissue [28]. Inside the articular capsule, there is a delicate synovial membrane that produces synovial fluid (synovia = "a thick fluid") [28], which is a viscous and slippery fluid [28]. Found in joints, it is referred to as synovial fluid due to its similarity to egg whites [27]. In addition, synovial fluid may be described as the collection of fluid that remains constricted to particular joint spaces. It is an ultrafiltrate of blood plasma, and comprises mainly of hyaluronan, lubricin, proteinase, collagenases, and prostaglandins. Such fluid is physiologic, working as a joint space lubricator of articular cartilage. It also diffuses requisite nutrients to its neighboring structures such as cartilage, meniscus, and labrum [29]. Figure 2. 1. and Figure 2.2 illustrates the distribution of synovial fluid in the body.

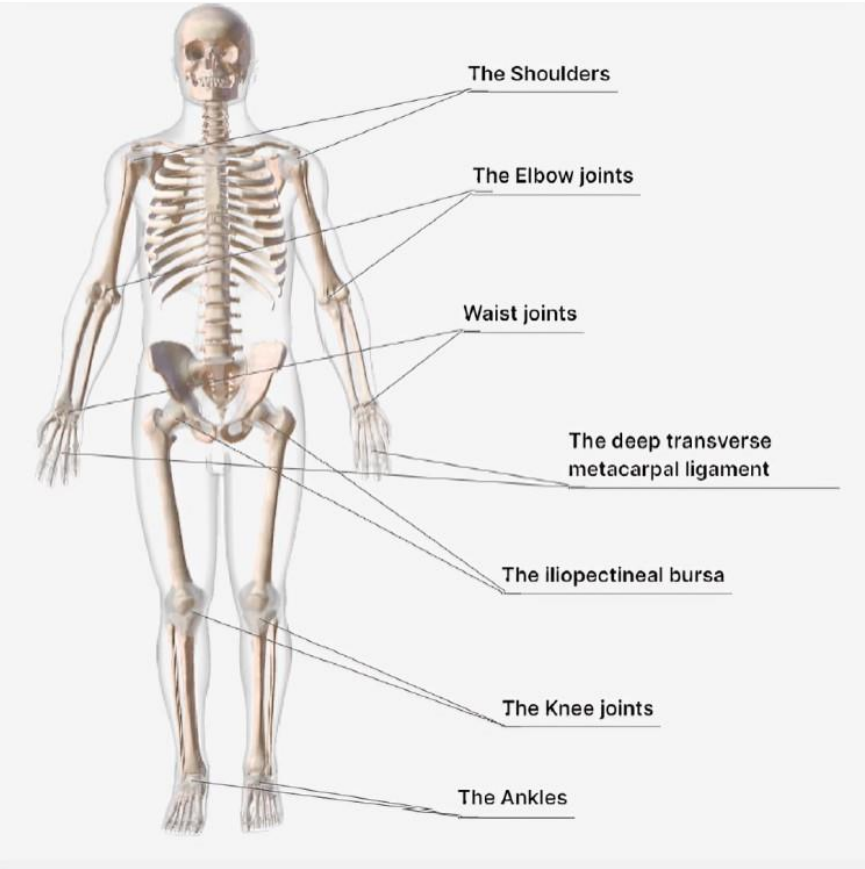


Figure 2. 1 Major Synovial Fluids in human body.

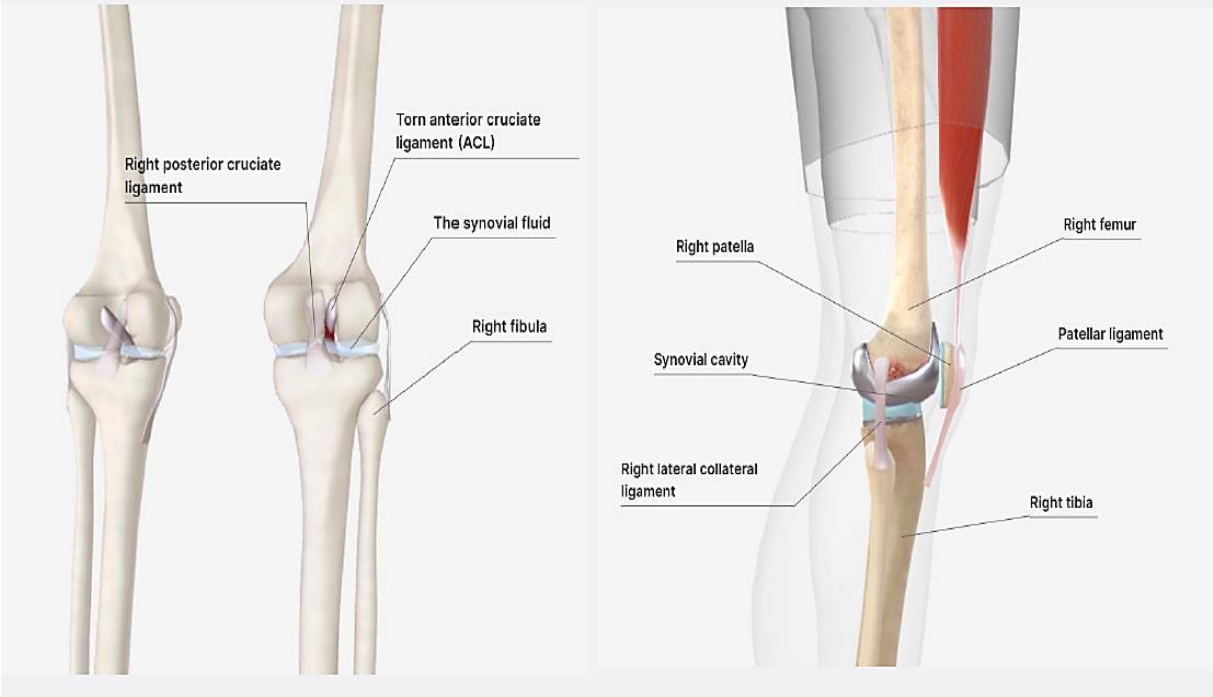


Figure 2.2. illustrates the distribution of synovial fluid in the body.

The main role of synovial fluid is to minimize friction between the articular cartilages of synovial joints during movement. The viscosity of this lubricating solution is produced by the presence of high molar mass hyaluronan (HA) in the fluid. Inflammation and oxidative stress increase the natural breakdown of hyaluronan, leading to many joint-related illnesses [38]. Since the synovial fluid surrounds joints and lubricates them, its analysis detects the source of joint pain and swelling. The working of the knee joint system is tested through its synovial fluid properties [30]. Synovial fluid presence in the joints was first investigated by the Greek Hippocratic text, “Of the Places of Man”. A Roman physician [31] also provides an earlier reference to synovial fluid in his text, “On the functions of parts of the human body.” [31]

### 2.2.1 Properties of Synovial Fluids

The fluid that lines joints is called synovial fluid; it is extremely viscous, resembling plasma [16,29, 32]. One of its main functions is to facilitate joint lubrication. Synovial fluid lubrication effectively protects the articular cartilage against degradation in a healthy joint. However, in an affected joint, the synovial fluid decreases in viscosity, resulting in reduced effectiveness of lubrication [32]. Diseases may be distinguished by quantifying the viscosity of synovial fluid, with rheumatoid arthritis having a higher viscosity than osteoarthritis, and both having higher viscosity than normal synovial fluid. [32]. The multiple characteristics of synovial fluid are outlined in Table 2.1. The physical characteristics include: 1 - Volume: Joints typically hold a little volume of fluid, while the knee joint typically holds a maximum of 4 ml of fluid [27]. 2- Color: In its normal state, synovial fluid appears colorless to pale yellow [33]. As noted previously, the term "synovia" refers to its resemblance to the white part of eggs [26]. The color of the fluid can become a deeper yellow in the presence of noninflammatory and inflammatory effusions and may even exhibit a greenish tinge in cases of bacterial infection [33]. 3 –The viscosity of synovial fluid results from the polymerization of hyaluronic acid, which plays a crucial role in adequately lubricating the joints. In conditions such as arthritis, both the production of hyaluronate and its ability to polymerize can be affected, leading to a decrease in the fluid's viscosity. Various methods are available to measure the viscosity of synovial fluid, with the simplest one involving the observation of the fluid's ability to form a string when drawn from a syringe. A string measuring 4 to 6 centimeters is considered normal. Another method to assess the amount of hyaluronate polymerization is the Ropes or mucin clot test [33, 27]. 4- Clotting: When

fibrinogen is present, synovial fluid may clot. Fibrinogen may have entered the synovial capsule as a consequence of a painful tap or injury to the synovial membrane [27].

Table 2.1 Characteristics of synovial fluid.

| ***Characteristics*** | **Normal Synovial Fluid Values^2** |
|---|---|
| Volume | < 3.5 ml |
| Color | Colorless to pale yellow |
| Clarity | Clear |
| Viscosity | Able to form a string 4-6 cm long |
| Leukocyte count | < 200 cells/ µL |
| Neutrophils | <25% of the differential |
| Crystals | None present |
| Glucose: plasma difference | <10 mg/dL lower than the blood glucose |
| Total protein | <3 g/dL |

Synovial fluid is a thick liquid consisting of plasma transudate and different chemicals that cells produce in joint tissue. The main functions of synovial fluid are to lubricate the synovial joints and supply nutrition to the intraarticular tissues [34]. A proteoglycan is a protein core with glycosaminoglycan (GAG) side chains attached to it. It is an important part of synovial fluid and the extracellular matrix of articular cartilage [34]. GAGs are polysaccharides composed of disaccharide units that are joined together by covalent bonds in a linear fashion. Their disaccharide repeating unit consists of a

hexosamine (containing D-glucosamine and D-galactosamine) and a uranic acid (D-glucuronic acid and Liduronic acid) [48]. The GAG consists of four main categories: hyaluronic acid (HA), chondroitin sulfate (CS), fucosylated chondroitin sulfate (FCS), and heparin/heparan sulfate, dermatan sulfate, and keratan sulfate [35]. All connective tissues have chondroitin sulfate (CS), a naturally occurring GAG, particularly in the extracellular matrix (ECM) of articular cartilage [19]. The synovial fluid consistently releases this chemical compound, and its levels and sulfation pattern reveal the joint tissue's metabolism [25]. Chondroitin sulfate (CS) is a GAG composed of a repetitive sequence of disaccharides consisting of N-acetylgalactosamin GalNAc and Glycosaminoglycans GlcA [35]. It possesses a smaller chain length compared to HA, consisting of 20-100 repeating units [35]. It is mostly found in the extracellular matrix of tissues and plasma membranes. This polymer has substantial variation in both its length and structure, which varies depending on the specific placements of the sulfates [35]. The largest amount of chondroitin 6-sulfate (C6S) seen in synovial fluid appears to come from articular cartilage. Chondroitin 4-sulfate (C4S) is found extensively not only in the cartilage of joints but also in the synovium, ligaments, and menisci [19]. Several GAG polysaccharides, including C6S (also known as chondroitin sulfate C, or CSC), and C4S (also known as chondroitin sulfate A, or CSA) [19].

Many joint illnesses alter the concentration of GAG in synovial fluid; for example, HA and total GAG levels are lower in patients with arthritis than in people enjoying normal knee health [42]. Pauilo et al. indicated the relative quantities of chondroitin 4- and 6-sulfates in cartilages from different human bones. The cartilages found in the joints and vertebral bodies consist mostly of chondroitin 6-sulfate, whereas

the cartilages involved in development and located under the joints have almost equal quantities of chondroitin 4-sulfate and chondroitin 6-sulfate. There are variations in the relative abundance of chondroitin sulfate in the cartilage of different animals as they age. Research has determined that chondroitin 4-sulfate is the primary glycosaminoglycan found in young cartilage, whereas elderly cartilage has a higher amount of chondroitin 6-sulfate [36]. Hyaluronic acid (HA) is a polysaccharide composed of repeating units of disaccharides, namely N-acetyl-D-glucosamine (GalNAc) and D-glucuronic acid (GlcA) [35]. Typically, it consists of 100 to 20,000 repeating units and has a molecular weight ranging from 105 to 108 Da. This sets it apart from other GAGs, which are often smaller in size [35]. Hyaluronic acid (HA) supports both stationary and moving lubrication of the joint [35].

In their book, Mundt & Shanahan [27] explained that Synovial fluid contains other chemical components such as proteins, glucose, uric acid, lactic acid, lactate dehydrogenase (LD), and rheumatoid factor (RF). The synovial fluid includes all proteins found in plasma, except for high-molecular-weight proteins such as fibrinogen and beta-2 macroglobulin, which are found in small amounts. The glucose levels in synovial fluid are usually lower than such levels in the blood, but decreased levels can indicate infectious joint disorders. Uric acid levels range from 6 to 8 mg/dL are useful for determining gout. When its lactic acid levels surpass 25 mg/dL —even reaching up to 1,000 mg/dL — this can be a sign of septic arthritis, although it is not commonly tested. Increased levels of lactate dehydrogenase (LD) may be related to rheumatoid arthritis (RA), infectious arthritis, and arthritis. Patients with RA mostly have the rheumatoid factor (RF) in their blood and synovial fluid.

### 2.2.2 Research Focused on Synovial Fluids and Joint Health

The issue of knee health problems is critical to specialists and medical researchers. There is much ongoing investigation in knee health being conducted by researchers who work in the biomedical field. They focus on electrical devices that measure and determine the level of knee health, especially synovial fluids and joint health. Human beings are exposed to varying loads on their joints while moving, especially their knee joints. Such loads on joints contribute to tissue damage due to the development of degenerated medical conditions and cartilage damage [37]. These conditions are known as osteoarthritis (OA). In their study, Spain & Cheneler [37] focused on the acoustic monitoring of joint health. The researchers identified classification methods that demonstrate the severity of knee damage. Lawrence and Kellgren's system identified radiographic evidence of OA concentration and scaled it from 0 to 4. However, the information obtained through radiographic evidence was not enough to define cartilage damage since this method had low inter-observer precision. The researchers also viewed other methods that quantify OA development effectively, including radiation-free methods and non-invasive ones. They defined the clinical settings and developed different therapies that would be effective in limiting and reducing OA. Variation of knee sounds during loading conditions and movements are an indicator of structural changes in knee joints and are linked with OA. The researchers also viewed the vibroarthrography and acoustic emission from knee joints. Monitoring the knee joints for OA enabled them to demonstrate synovial fluid needs for joints.

In their research paper, Prekasan & Saju [31] explained that synovial fluid is composed of hyaluronic acid and is identical to blood plasma with 2% protein content. It

is utilized for the natural lubrication of elbow and knee joints. It contains surface-active phospholipids, proteoglycan 4, and hyaluronic acid. Chondrocytes secreted in these lubricants are synoviocytes and articular cartilage. Synovial spaces in the joints are responsible for the concentration of synovial fluid. Lubricating system deficiencies contribute to the erosion of articular cartilage which creates arthritic conditions. The molecular concentration of hyaluronic acid is 1-4 mg/ml, proteoglycan is 0.05-0.35 mg/ml, and active phosphor lipids are 0.1 mg/ml. Synovial fluid interacts with the joint surfaces in relative motion to reduce friction effects during joint movements.

On their part, Seidman & Limaiem [29] asserted that during synovial fluid gross assessment, doctors examine viscosity level, clarity, volume, and color of the fluid. This helps identify symptoms that cause infection in the joints. During chemical testing of synovial fluid, experts examine the concentration of uric acid in joints. This is done via glucose testing, protein testing, and lactose quantity tests that may explain the reason for joint infection and other health diseases in patients. The microscopic testing of synovial fluid is supported by crystal examination of infection-causing elements. During microscopic testing, doctors examine bacterial culture that have accumulated in the joints by testing individual immune systems (WBCs). Gram stain is also calculated and is used to identify the causes of infections, factors resulting in synovial fluid changes in the body as well as treatment required.

On the other hand, More et al. [26], reviewed the rheological properties of synovial fluid and explained OA remedy options. They discussed the lubricative properties of synovial fluid and related it with an electrolytic solution that assists the joints in proper movement. They identified the interaction between the different

components of synovial fluid and connected this to the complex rheological properties of the fluid. Their analysis of inflammatory rheumatoid and osteoarthritis diseases demonstrated the significant efficiency of synovial fluid and its lubrication capabilities. They argued that viscos supplementation cases replenish synovial fluid concentration and affect its molecular weight as well as rheological properties. They also revised the rheological properties and polymetric properties of synovial fluid and noted the challenges related to its clinical effectiveness and viscos supplements.

Furthermore, Marian et al. [38] explored the mechanism of synovial fluid for joint longevity. It has already been noted multiple times in this paper that the fundamental function of synovial fluid is to lubricate the joints. The constituents of synovial fluid include active phospholipids, hyaluronic acid, and surface-active protein. Marian et al. [38] also identified its other protein components such as gamma globulin seric protein and human serum protein which assist the synovial fluid in lubricating joints. They identified the bio tribological system of synovial fluid and reviewed the current knowledge of lubrication. The prospective usage of synovial fluid enabled the researchers to identify its ability to combat osteoarthritic. The major focus of their research was to identify the role of protein on prosthetic lubrication, physical variable effects in prosthetic lubrication, and optimal material combination for prosthetics. They tested the wear mechanism of synovial fluid to identify prosthetic lubrication characteristics. Synovial fluid was studied under bio tribology conditions to identify certain prosthetic components; the superior lubrication properties of the fluid was also discussed in their study. They noted the fluid lubricated knee and elbow joints and prevented uric acid accumulation in them. The prosthetic testing of synovial fluids

demonstrated the exact chemical composition of different components and their lubrication mechanism for the development of an effective lubricative mechanism.

Evaluating pathological changes in the fluid to develop a numerical understanding, Liao et al. [39] investigated the roughness of cartilage surface to understand synovial fluid characteristics. The significant focus of the researchers was osteoarthritic conditions of synovial fluid and cartilage contact gap. Initially, they evaluated the cartilage surface topography in healthy as well as osteoarthritic conditions. They developed connections between the healthy and sick stages through a numerical approach with consideration of horizontal and vertical roughness of cartilage surface. Calibration against existing experimental data about the synovial fluid provided constitutive equations for the viscosity of synovial fluid. The results showed cartilage surface reduced its gap permeability by around 30-60%. While testing the gap permeability for discrete cartilage surface sizes, results showed the gap permeability of different cartilage surfaces decreasing and becoming more significant. Gap permeability resulted from early fluid ultrafiltration into joint tissues. The gap permeability of the cartilage surface was highly affected by pathological synovial fluid. OA cartilage surface increased the size of gap permeability up to a few hundred times. The pressure gradient remained always less than 106 Pa/m which resulted in fluid ultrafiltration in the tissues. The function of articular joints was highly affected by a change in OA conditions and cartilage contact gaps.

### 2.2.3 Previous Studies on Chondroitin Sulfates and Hyaluronic Acid in Synovial Fluid

There is much extant research on chondroitin sulfates and hyaluronic acid in synovial fluids. Uesaka et al., [19] confirmed the significance of chondroitin sulfate isomers in the synovial fluid of osteoarthritis patients. Their study aimed to investigate the

correlation between age, gender, and the severity of OA in knee OA patients and a control group of healthy individuals. The sample for this study was the concentrations of chondroitin 6-sulfate (C6S) and chondroitin 4-sulfate (C4S) in synovial fluid from 133 patients and 27 healthy individuals. The results indicated that the control group had lower C6S and C4S levels and a smaller C6 ratio than those with grade 1 OA. Grade 3 and 4 patients had reduced C6S concentrations and ratios. The gender characteristics showed significant differences, with females showing lower values than men in both the control and OA groups. In addition, a multiple regression analysis revealed a negative connection between age and all three biochemical variables in the control group. However, a significant opposite correlation was observed between the range of osteoarthritis and the concentrations of C6S and the C6 ratio. On the other hand, there was a less negative correlation between the severity of osteoarthritis and the concentration of C4S.

The study by Nakayama et al, [40] investigated the impact of age and sex on chondroitin sulfates (CS) in normal synovial fluid. A total of 82 healthy individuals aged 20 to 79 years provided samples of synovial fluid. The results indicated the levels of CS and HA, as well as the ratio of C6S to C4S, changed depending on age. Their values ranged between 20 and 30. Statistically, the concentration of C6S and the ratio of C6S to C4S were significantly lower in those aged 60-70 compared to those aged 20-30. Women had significantly lower concentrations of CS and a lower C6S:C4S ratio compared to males. Age showed a high positive correlation with the C6S concentration and the C6S:C4S ratio, a weak positive correlation with the C4S concentration, and a moderate positive correlation with the HA concentration. Sex exhibited a modest link with the levels of C6S and C4S, as well as the C6S:C4S ratio. However, there was little correlation between sex and the

concentration of HA.

On the other hand, the purpose of the study by Andersson et al, [34] was to confirm the accuracy and reliability of a high-performance liquid chromatography (HPLC) method for measuring CS and HA levels in synovial fluid. Samples were collected from 25 patients with osteoarthritis and 13 patients with knee injuries. Synovial fluid (SF-control) and aggrecan were treated with chondroitin and labeled with fluorophores. The researchers used quantitative HPLC analysis and mass spectrometry to evaluate the N-glycan profiles. The results showed unsaturated uranic acid and sulfated-N-acetyl galactosamine (ΔUA-GalNAc4S and ΔUA-GalNAc6S) accounted for 95% of the overall CS-signal in the SF-control sample. The coefficient of variation for HA and CS variants in SF-control ranged from 3% to 12% and from 11% to 19%, respectively. The biofluid stability test showed recoveries between 81% and 140%. A total of 61 distinct N-glycans were identified in the synovial fluid samples, but no significant variations were observed across the three patient groups. The CS-profile was similar to pure aggrecan from equivalent samples, and the contribution of aggrecan to the N-glycan profile was minimal.

Moreover**,** Fonsi et al, [41] argued HA is commonly administered by intra-articular (IA) injections to augment the viscosity of synovial fluid. This therapy is routinely employed to alleviate symptoms associated with OA. In this study, the researchers introduce HCS, a novel blend of substances that combine HA and chondroitin sulfate (CS), two important members of the GAG family. These compounds are key constituents of the joint. HA contributes to the thickness and stickiness of the synovial fluid, while CS enhances the flexibility and resilience of the cartilage. OA joints have decreased levels of HA and CS, which are linked to the gradual deterioration and depletion of cartilage. The

study aimed to assess the pharmacokinetic (PK) characteristics of both HA and CS following intra-articular (IA) treatment in a validated animal model of osteoarthritis OA. The effectiveness of an IA injection of mono-iodoacetate (MIA) in inducing symptoms of OA in the knee of rats was validated by motion impairment assessments and histological studies. Subsequently, the pharmacokinetic parameters of HA and CS were assessed following intravenous injection. Each active component was individually analyzed: HA was tagged with tritium (3H-HA) and CS was labeled with carbon 14. The text "(14C-CS)" remained unchanged. The ultimate radio-labeled solution replicated the refrigerated HCS formulation.

On the other hand, the purpose of M.P. Brown's [42] study was to examine how exercise and osteochondral (OC) injury affected the concentration and chain length of CS and HA in synovial fluid of horses. Serum and synovial fluid were collected from three groups of horses: (1) healthy horses after 8 weeks of rest, (2) the same horses after undergoing 9 months of treadmill training, and (3) horses with osteochondral (OC) damage resulting from racing. Group 3 horses were also used to gather articular cartilage. The concentrations and lengths of CS and HA were measured using gel chromatography and fluorophore-assisted carbohydrate electrophoresis. The results showed the peak chain length of synovial fluid CS in the OC injury group had a considerable increase (18.7 kDa) compared to rested horses (11.6 kDa), whereas exercise resulted in an intermediate chain length (15.6 kDa). The cartilage and serum obtained from the group with OC injuries had unusually elongated CS chains, similar to those observed in the synovial fluid of these horses. The overall surface area of the head and neck was notably smaller in the group with OC damage compared to the group that was at rest. Both the group with OC damage and

the group that exercised had notable reductions in the length of synovial fluid HA chains, in comparison to the group that rested. The length of the synovial fluid CS chain was extended by physical training and an osteochondral injury. Engaging in physical activity led to a little rise, whereas an OC injury resulted in a significant increase. Unlike in the case of CS, the length of the synovial fluid HA chain was reduced as a result of OC damage, and to a lesser degree due to exercise. An examination of the length of chains in synovial fluid, CS, and HA may offer a valuable method for assessing the health of joints.

### 2.2.4 Measuring Viscosity of Synovial Fluids

In order to monitor the health of fluid at joints in human body beings, measurement is very important. The particular reason for synovial fluid measurement is to expostulate explanations for joint pain and dysfunction. Assessing synovial fluid can help doctors reduce or analyze one of the many causes of joint pain. There are many traditional methods to measure synovial fluid such as blood tests, pathology, X-ray, MRI, ultrasound, and CT [43]. However, certain medical studies have propounded discrete new methods to measure synovial fluid.

Wang & Li [44] designed an online viscosity measurement method that is based on piezoelectric torsional transducer and electrochemical impedance. The test of the ring array of piezoelectric patches share mode and actuated the diagnostic system according to torsional resonance. They developed quantitative EMI-based measurement equivalence that assists online viscosity measurements, and derived viscous damping through fluid contact and tracking piezoelectric impedance of synovial fluid. They also tested the viscosity of glycerol-water, which is a concentrated solution of various components. The results of viscosity measurements agree well with those given by standard laboratory

viscometers. The device was highly sensitive toward viscous materials and had high-quality features to measure viscosity of any material.

On their part, Krishnan et al. [43] stated that by measuring the variation in density of synovial fluid, arthritis can be diagnosed. In their research, the authors used electrical bio-impedance, a non-invasive way to measure density of synovial fluid. This was done by calculating the opposition a tissue gives when an electric signal is applied to it. The values were thus recorded as drops in voltage in the area surrounding the joint tissue; it was by nature a resistance and capacitive reactance. The resultant numbers would differ between a normal person and someone suffering from arthritis since a fixed frequency and voltage is applied to the joints. Certain hardware was necessary to execute signal generator circuit compatible with variable frequency and voltage. Since the procedure was non-invasive, voltage drop was calculated by placing four electrodes that recorded the signals and responses. The results showed resistance and reactance varied from a normal person to an arthritis-afflicted patient when they were administered the same fixed frequency and amount of voltage. When such input voltage or frequency was changed, the corresponding measured voltage drop as calculated in bio-impedance also changed.

Again, Iqbal et al. [45] researched automated detection of human knee joint synovial fluid. They analyzed the magnetic resonance image and developed deep learning-based automated detections with transfer learning. Deep learning is beneficial in understanding analytical tools that help diagnostic teams evaluate synovial fluid concentration in knees more closely. Analytical tools in radiology and medicine provide labeled diseases in detail and reduce ensuing medical field challenges. Transfer learning

makes the knowledge process easier, focusing on the needs of relevant diseases, locating accurate solutions with discrete resolution parameters. In the study under reference, transfer learning started from the pre-training stage to move large-scale databases for certain tasks. The researchers tested the synovial fluid concentration in human knees through a deep learning process using magnetic images. They proposed a neutral architecture that specializes in automated detection of fluid through resonance images. This research covered the training, evaluation, and development stages of the proposed model. Independent database structures were utilized to diagnose the fluid level. The experimental model of testing was highly sensitive and specific to synovial fluid evaluation. It was specifically designed to bring accuracy to the experiments and make significant perceptions about testing the fluid level. The proposed structure provided a feasible way for expediting and automating synovial fluid analysis. The annotated database of the research comprised small tunes that comprehensively transferred the understanding of synovial fluid with magnetic resonance images. These images detailed the disease's causes and made the treatment process easier.

In another extant study, Jun Fu et al. [46] performed the viscosity test of synovial fluid to diagnose periprosthetic joint infection (PJI). The main purpose of this research was to identify an accurate testing methodology that provides authentic results for PJI testing. The researchers examined 45 patients suffering from PJI (n 1⁄4 15), revision for aseptic failure (n 1⁄4 15), and primary arthroplasty (n 1⁄4 15). The researchers argued synovial fluid viscosity varies in patients with PJI. They identified the sensitivity of viscosity difference to develop an accurate testing framework, utilizing the musculoskeletal infection criteria to identify viscosity levels of synovial fluid. They also

measured the plasma D-dimer level, c-reactive protein, and erythrocyte sedimentation rate of each patient. The results indicated synovial fluid viscosity level was 0.0011, which was significantly lower in patients with PJI as opposed to those with no such infection. Synovial fluid viscosity level exceeded CRP, ESR, and plasma D-dimer, with a sensitivity of 93.33% and specificity of 66.67%.

### 2.2.5 Age-Related Changes in Synovial Fluid Composition

As the human body ages, it affects changes in the level of growth in varied stages and different parts of the body. The term aging denotes the various physiological transformations bodies go through from adulthood till death. The bones in the body are remodeled throughout life as older bone tissues are replaced by newer ones. This systematized remodeling ensures a balance between bone resorption and formation which, in turn, maintains skeletal integrity. Like most things, such balance also alters as people age [47]. It needs to be enumerated here that bones are not in direct contact with each other within joints. Indeed, the articular cartilage that lines joints have a cushioning effect. Synovial fluid plays a crucial role in maintaining joint health and functionality by acting as a lubricant and shock absorber. In a joint, bones are cushioned by articular cartilage, synovial membranes, and synovial fluid [49]. The joints are also protected by synovial membranes while the synovial fluid acts as the lubricant to ensure seamless movement of the joints. However, the synovial fluid reduces as people age and the cartilages thin out — resulting in stiffness around the joints. Adding to the woe, ligaments become shorter with age and lose flexibility which also causes the joints to feel stiff [48].

The changes in synovial fluid associated with aging can impact its viscosity and

elasticity. In younger individuals, synovial fluid is more viscous, providing effective lubrication and joint cushioning. However, as individuals age, the viscosity of synovial fluid tends to decrease, leading to reduced lubrication and possibly contributing to joint degeneration. The elasticity of synovial fluid may also decrease with age, limiting its ability to absorb and distribute mechanical stresses within the joints, particularly the knee [50]. Furthermore, the aging process may result in stiffer and less flexible joints, with a potential decrease in joint fluid. The cartilage may begin to wear away, leading to the rubbing of bones against each other. Additionally, minerals may be deposited in and around certain joints, a condition known as calcification [51].

It may be reasonably expected that when these shifts occur, they result often in biological function decline, causing corresponding mental changes such as psychological and behavioral ones. Sometimes such transformation is evident, while at other times, these may not be clearly perceivable. As bones age, their mineral content also starts depleting. This makes people susceptible to osteoporosis, wherein bones lose density, become more fragile and, thus, likely to suffer from fractures. As human beings age, bone resorption rate by osteoclast cells becomes greater than the speed at which bone forms. Since the multinucleated cells comprise mitochondria and lysosomes that aid in bone resorption, bones weaken as the process reverses [47].

With all these references to synovial fluid in this paper, it is necessary at this point to expound on the crucial role this plays in joint health. It has already been noted that this fluid lubricates joints to ensure smooth movement. Further, healthy joints have greater quantities of high molar mass HA molecules within the synovial fluid which ensures perfect viscosity in its role as a lubricant. As people grow older, the size of HA molecules

declines, causing its function ability as a cushioning support and lubricant to lessen [30,49].

Uesaka et al. [25] investigated sex differences and age-related changes in HA and chondroitin sulfate (CS) isomer in regular synovial fluid. This research yielded the result that older people and gender profoundly affected the knee joint tissues, resulting in osteoarthritis. They determined HA and CS isomer levels in normal synovial fluid to test the influence of age and sex on its concentration. They tested a sample population comprising healthy subjects of different ages that informed data on the influence of age on knee joints lubricants' concentration. The 187 healthy volunteers tested ranged in age between 14 and 89 years. The results of the study showed that most synovial fluid components had a negative correlation with the age factor. There were no sex-related differences in synovial fluid concentration. However, CS isomer in the normal synovial fluid was significantly affected by such differences. The affected concentration also varied among women and men.

Temple-Wong et al. [52], in their study, analyzed size distribution and hyaluronan concentration in human knee synovial fluid. They associated synovial fluid variations with cartilage degeneration and age factors. The potential criteria identified in cartilage wear for normal knee joints was an alternative to synovial fluid quantity and lubricant content. They made specific correlations between synovial fluid and age factor. They analyzed the synovial fluid of donors (age range 23–91 years) who were not suffering from osteoarthritis. The researchers also investigated healthy communities to identify molecular weight of synovial fluid hyaluronan and protein components. Concentration was checked for both left and right knees of the people sampled. Bland-Altman and T-

test were implemented to test the synovial fluid concentration. Regression analysis was then performed to define the relationship between each biochemical component. The results highlighted that at older ages, cartilage wear is related to the quality of synovial fluid and hyaluronan content of joints. The study authors identified the concentration of all biochemical components of synovial fluid. Healthy joints contain a protein concentration range of 2.5–7 MDa (–9.4 %/decade), 1–2.5 MDa (–11.3 %/decade), 0.5–1 MDa (–12.5 %/decade), and 0.03– 0.5 MDa (–13.0 %/decade). They concluded that hyaluronan concentration was more strongly related to age than joint grade.

2.2.6. Studies On Synovial Fluid of Patients with Anterior Cruciate Ligament (ACL) Problems

Keiko Amano et al. [53], investigated the connection between posttraumatic osteoarthritis and anterior cruciate ligament (ACL) tears. They considered both biomechanical factors and changes in biochemical profiles within the knee joint after injury and ACL reconstruction (ACLR), as potential contributors to joint degeneration. Their study aimed to assess how the inflammatory response after injury, indicated by concentrations of cytokines, metalloproteinases, and cartilage degradation markers in the synovial fluid, may influence articular cartilage degeneration over a three-year period following ACLR. The researchers utilized T1ρ and T2 quantitative magnetic resonance imaging to evaluate cartilage degeneration at multiple time points — before surgery, as well as at 6 months, 1 year, 2 years, and 3 years post-surgery. Through Pearson correlation coefficients and K-means clustering, they identified intricate relationships between inflammation and cartilage turnover, which were also influenced by factors such as timing after injury and patient-specific factors. Notably, at the time of ACLR surgery,

the profiles of inflammatory cytokines, degradative enzymes, and cartilage breakdown products in the synovial fluid showed potential as predictors of abnormal cartilage tissue integrity (indicated by increased T1ρ and T2 values) during the first three years after surgery. The study comprised 26 participants from a longitudinal cohort study who underwent ACLR at an average of 8.5 weeks after injury (ranging from 4 to 19 weeks). At the time of surgery, synovial fluid was aspirated and analyzed for biomarkers using immunoassays. The T1ρ and T2 values of the articular cartilage were examined through magnetic resonance scans to evaluate cartilage condition over the specified time frame.

On the other hand, Kim Wang et al. [54], conducted a study focusing on immune cell profiles in synovial fluid following ACL and meniscus injuries. The research involved 29 patients, with 14 having meniscus injuries and 15 with ACL injuries. All participants had a normal contralateral knee. During the ACL and/or meniscus surgery, synovial fluid was extracted from both the uninjured and injured knees. The collected synovial fluid cells were processed, washed, and then labeled with a combination of fluorescent antibodies targeting cell surface proteins. Polychromatic flow cytometry was used to analyze the immune cells present in the synovial fluid. The results of the study revealed fundamental information, indicating that ACL and meniscus injuries create a microenvironment within the synovial fluid that is rich in immune cells, mainly comprising various T cells with multiple T helper phenotypes. Further investigations exploring the correlation between immune cells and joint degeneration could potentially lead to a deeper understanding of the underlying pathophysiology of posttraumatic osteoarthritis (PTOA) following joint injuries.

On their part, Kiapour et al. [55], conducted a study to investigate the potential chondroprotective effects of bridge-enhanced ACL repair by comparing the protein contents in synovial fluid. The researchers hypothesized post-surgical changes in the synovial fluid proteome would differ between untreated and repaired knees, and these changes would be related to the extent of cartilage damage. They performed the study on 30 adolescent Yucatan mini-pigs, wherein unilateral ACL transection (ACLT) was performed, and the pigs were randomly divided into two groups: one receiving no further treatment (ACLT, n = 14), and the other undergoing bridge-enhanced ACL repair (BEAR, n = 16). Using an advanced isotopically labeled high resolution LC MS/MS-based proteomics approach, the protein profiles of synovial fluid were analyzed at 6- and 12-months post-ACL transaction in both untreated and repaired porcine knees. At 6 months, there were no significant differences in cartilage lesion area or the quantitative abundance levels of secreted proteins between the ACLT and BEAR groups. However, by 12 months, the ACLT group showed greater cartilage damage compared to the BEAR group ($p = 0.005$). This difference in cartilage damage was accompanied by variations in the abundance levels of certain secreted proteins, with higher levels of Vitamin K-dependent protein C ($p = 0.001$), and lower levels of Apolipoprotein A4 ($p = 0.021$) and cartilage intermediate layer protein 1 ($p = 0.049$) in the ACLT group compared to the BEAR group. The study also revealed group differences in the secreted proteins that underwent significant changes in abundance between 6 and 12 months in both ACLT and BEAR knees. These findings suggest knees treated with bridge-enhanced ACL repair may maintain an environment that better protects the extracellular matrix, a function not observed in the ACLT knees.

Moving from swine to human beings, Brophy et al. [56] conducted a study to explore the association between patient- and injury-specific factors and the proteomics of synovial fluid in knees with ACL tears. They collected synovial fluid samples from 105 patients (38 male and 67 female) who had experienced an acute traumatic ACL tear. Various patient-specific factors, including age, sex, body mass index, time from injury, presence or absence of concomitant meniscal tears, and the location of any accompanying bone bruises (if present), were recorded. The protein concentration of the synovial fluid was measured, and samples were subjected to multi-affinity high-abundance protein depletion for benchmarking. Using an advanced isotropically labeled high-resolution nano-liquid chromatography with tandem mass spectrometry-based proteomic approach, the researchers determined the protein profile of the synovial fluid for each patient and injury factor. The findings revealed the proteomics of synovial fluid in ACL tears were associated with patient sex, injury pattern, and the location of bone bruises, but not with patient age, body mass index, or time from injury. Specifically, knees with an isolated ACL tear displayed higher levels of glutathione peroxidase 1 (GPX1) and plastin 3 compared to knees with both ACL and meniscal tears. Additionally, a bone bruise on the lateral femoral condyle was linked to elevated levels of leptin and glucose-6-phosphate dehydrogenase (G6PD), while a bone bruise on the lateral tibial plateau was associated with decreased GPX1 levels. Male patients had higher matrix metalloproteinase 9 and lower G6PD levels than female patients. In summary, the results indicated that patient sex, injury pattern, and bone bruise location significantly influenced the proteomic profile of the synovial fluid resulting from ACL tears.

Matthew T. Kingery et al, [57] aimed to examine the changes in the concentration of specific proinflammatory and anti-inflammatory cytokines in synovial fluid during the period between an ACL injury and surgical reconstruction. To do this, they collected synovial fluid samples from patients with an acute ACL injury during their initial office visit within two weeks of the traumatic event. Additional synovial fluid samples were obtained just before ACL reconstruction surgery. These samples were processed with a protease inhibitor, and the concentrations of 10 cytokines of interest were measured using a multiplex magnetic bead immunoassay. The main focus was on assessing the alterations in cytokine concentrations between the time points of the acute injury and the surgery. The study comprised 20 patients with an average age of 30.2 ± 8.3 years. The synovial fluid samples for acute injury were collected at an average of 6.6 ± 3.8 days after the injury, while the samples from the surgery were collected at an average of 31.6 ± 15.6 days after the initial injury samples. By employing linear mixed-effects models to account for the impact of concurrent meniscal injuries and individual patient variations, the researchers found statistically significant increases in the concentrations of RANTES and bFGF and statistically significant decreases in the concentrations of IL-6, MCP-1, MIP-1b, TIMP-1, IL-1Ra, and VEGF between the two time points. The study demonstrated the ongoing changes in the intra-articular microenvironment during the initial inflammatory response in the acute post-injury period. Specifically, six synovial fluid cytokines showed significant decreases, while two cytokines displayed significant increases between the first clinical presentation shortly after the injury and the time of surgery, which was approximately one month later.

## 2.3 Machine Learning

Machine learning (ML) is a discipline in which computers generate probabilistic statistical models from data and employ these models to predict and analyze information. It is often referred to as statistical machine learning [58]. According to Herbert A. Simon, "learning" may be defined as the ability of a system to improve its performance by going through a process [58]. In addition, Mahesh [59] defined ML as the scientific investigation of algorithms and statistical models that computer systems employ to execute a particular task without direct programming. Learning algorithms are utilized in numerous applications that we utilize on a daily basis. Arthur Samuel [59] defines machine learning as a discipline that authorizes computers to acquire information and improve their performance without explicit programming. Samuel was famous for creating software that demonstrated exceptional performance in the game of Checkers. Machine learning is utilized to produce predictions and interpret data, especially new and unfamiliar information. The prediction of data enhances computer intelligence and improves skills, while data analysis enables people to gain new knowledge and discover unexpected conclusions [58]. The purpose of machine learning is to extract knowledge from data [59]. The main goal of machine learning is to analyze the different models to be learned and the methods to acquire them, to enable the model to make accurate predictions and evaluate the data effectively [58].

H. Li stated in his book [58], machine learning is characterized by the following: (1) It is dependent on computers and networks; (2) Machine learning is a field of research that focuses on analyzing data to make predictions and decisions; (3) The primary objective of machine learning is to analyze and predict data; (4) Machine learning is

method-centered, as it involves the development and application of models for the purpose of prediction and analysis; (5) Machine learning is a combination of computer science, optimization, computation, information, statistics, and probability. Its theoretical system and methodology have been progressively established in conjunction with its advancement. Machine learning algorithms may be categorized into four primary groups: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [58, 59, 60] as explained in Figure 2.1.

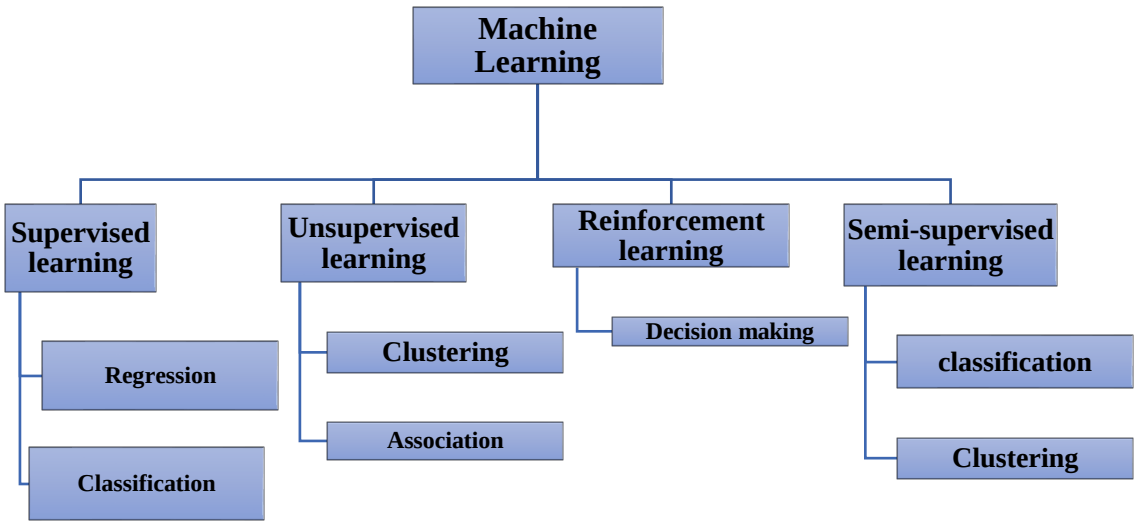


Figure 2.1. Types of machine learning.

There are three basic ML algorithms [59]. The first is supervised learning, a type of ML comprising algorithms that need external support. Herein, researchers can learn about functions which track the path of input-output pairs of data. A set of training examples are culled for labeled data, and supervised learning methods infer functions from such

data. While the input data is further subdivided into train and test data, it is only the former that requires prediction or classification. The train data reveals patterns that feed into the algorithms; these are then applied to the test dataset to further throw up predictions or classifications [59, 60].

The second type of ML is unsupervised learning which, on the other hand, requires no interference from human beings and is used to understand certain features of the data. The algorithm reads extant features to classify new data and is primarily used for clustering and feature reduction of data [59, 60].

The third type of ML is reinforcement learning wherein the algorithm allows a machine to internally find the optimal behavior for best efficiency within particular contexts or situations. This is thus an environment-driven method [60]. It works on a rewards or penalties system — data obtained is used primarily by environmental activists to check factors related to their field. While there is some demand for reinforcement learning in AI models, it is not of much use for solving simpler problems [60]. There is a fourth type of ML that falls in between the supervised and unsupervised learning methods. Termed semi-supervised learning, this algorithm uses both labeled and unlabeled data [59, 60]. Since labeled data is not that abundantly available in the world and unlabeled data is more prevalent, semi-supervised learning is a good support for predicting outcomes using the latter. It is beneficial in applications such as machine translation, detecting fraud, data labeling as well as text classifications [59, 60].

### 2.3.1. Machine learning applications in medical research

Researchers have recently used advanced techniques such as machine learning and artificial intelligence to analyze and examine health-related data. ML applications in

healthcare are helping medical practices develop diagnostic and treatment capabilities [61, 62]. ML has a significant impact on healthcare by enhancing the speed and accuracy of medical practices [62]. Many researchers use medical data for illness detection and pattern identification [61]. A number of studies have investigated how ML impacts the medical field. In their study, Sidey-Gibbons & Sidey-Gibbons [63] examined the use of ML in cancer diagnosis by developing and evaluating three predictive models using descriptions of cell nuclei from breast mass samples. , They emphasized Regularized General Linear Models (GLMs), Support Vector Machines (SVMs) with a radial basis function kernel, and Single-layer Artificial Neural Networks. The dataset, consisting of 683 samples, was divided into training (456 samples) and validation (227 samples) sets. The models were trained on the training data and tested on the validation data, with the SVM algorithm achieving the highest accuracy (0.96) and area under the curve (0.97). The study also assessed the performance of voting ensembles, which marginally improved accuracy, sensitivity, and specificity. The study provides a detailed, practical guide using the R statistical programming environment, making it accessible for clinicians and researchers to develop and evaluate machine learning algorithms in healthcare.

On their part Bianchi et al. [64] looked at the application of ML methodologies for the development of temporomandibular joint (TMJ) osteoarthritis (OA) diagnosis. TMJ-OA is the second-most prevalent musculoskeletal disease, surpassed only by determined low back pain. Early detection of TMJ problems is necessary to prevent morphological damage, given that these illnesses affect 5-12% of the population and result in $4 billion in annual healthcare costs. The study examined the diagnosis validity

of four ML models — logistic regression, random forest, lightGBM, and XGBoost. The researchers evaluated these models using 52 clinical, biochemical, and high-quality CBCT indicators from patients with TMJ OA and a control group. Significant factors that identify early stages of TMJ OA include important traits and interactions, headaches, range of mouth opening, different biomarkers, and imaging features. The XGBoost + LightGBM model showed an accuracy of 0.823, an AUC of 0.870, and an F1-score of 0.823, indicating positive possibilities for further investigation into personalized treatment approaches to enhance joint health.

Many existing studies have explored applications of ML algorithms in orthopedic injuries — including ACL tears — with promising results, such as improved accuracy in diagnosis, personalized treatment plans, and enhanced predictive modeling of patient outcomes. That has helped lay the groundwork for this research to build upon and expand our understanding of ACL patient care through machine learning. Kunze KN et al.'s [65] study aimed to develop ML algorithms to predict whether patients undergoing ACLR would achieve the minimal clinically important difference (MCID) on the International Knee Documentation Committee (IKDC) score at a minimum 2-year follow up. The researchers retrospectively analyzed data from an ACLR registry involving 442 patients. Thirty-six variables were considered, while the study population was divided for training and testing sets. Six ML algorithms were employed and evaluated for performance using various metrics. The results identified specific preoperative and intraoperative factors, such as body mass index, medial collateral ligament examination grade, femoral tunnel fixation method, prior contralateral knee surgery, and preoperative knee extension, as significant predictors of achieving MCID. The elastic-net penalized logistic regression

(ENPLR) algorithm exhibited the best predictive ability, emphasizing the utility of machine learning in assessing patient outcomes after ACLR.

The study by D.-S. Chen et al, [66] introduced the novel approach of integrating unsupervised ML and supervised ML-derived radiomics in the context of ACL ruptures. The research comprised 68 patients, with demographic and radiomics features recorded and used as input for ML algorithms. The unsupervised algorithm classified patients into five groups revealing, among others, significant differences in ACL rupture incidences and left knee involvements. Utilizing a supervised ML algorithm, a radiomics model was constructed, and through feature selection methods like t-test and least absolute shrinkage and selection operator (LASSO), 43 radiomics features were identified. The combination of LASSO and random forest classifier demonstrated the highest sensitivity, specificity, accuracy, and area under the curve (AUC), highlighting its efficacy in predicting ACL rupture. The study validated the clinical application of unsupervised ML in ACL rupture and identified seven radiomics features as potential predictors for this condition, suggesting the promising utility of radiomics in predicting ACL injuries.

Research by R. K. Martin et al, [67] aimed to enhance the prediction of ACLR revision by employing machine learning analysis on the Norwegian Knee Ligament Register (NKLR). Examining data of 24,935 patients with a mean follow-up of 8.1 years, the study developed four ML models, including Cox Lasso, survival random forest, generalized additive model, and gradient-boosted regression. All models demonstrated moderate concordance and were well-calibrated, with the Cox Lasso model requiring only five variables for accurate outcome prediction. A novel in-clinic calculator, the Revision Risk Calculator, was created to estimate individualized risk levels for ACL

revision, ranging from near 0% for low-risk patients to 20% for high-risk patients within 5 years. The study concluded that ML analysis of a national knee ligament registry can moderately accurately predict the risk of ACL reconstruction revision, offering the potential for a practical in-clinic tool for personalized risk stratification based on minimal input variables.

Corban et. al [68] conducted a systematic review to investigate the current and potential applications of artificial intelligence (AI) in managing ACL injuries within orthopedic surgery. As technological innovation becomes integral to the field, the study explored the future role of AI, leveraging ML algorithms with adaptive learning and problem-solving capabilities to enhance accuracy. Nineteen publications were included in this systematic review, categorizing AI applications into prediction, diagnosis, intraoperative utilization, and postoperative care and rehabilitation. The applications spanned image interpretation, automated chart review, physical examination assistance through optical tracking, generation of predictive models, and optimization of postoperative care. While the review reflected a growing interest in AI among orthopedic surgeons and some studies showcasing comparable or superior outcomes, it acknowledged the existence of challenges that needed to be addressed before widespread adoption of this technology in orthopedic practice.

Uozumi, Y et.al [69] conducted a study to introduce an automated segmentation approach for assessing the anteromedial (AM) and posterolateral (PL) regions of bone tunnels after double bundle ACL reconstruction [4]. The procedure, which typically involves creating two bone tunnels, is crucial for knee injuries. Utilizing K-means clustering, the study evaluated six patients (mean age 27 ± 7, four males/two females).

The proposed method successfully segmented the bone tunnels for all patients, indicating its efficacy in delineating the regions after double bundle ACL reconstruction. In summary, the study suggested the introduced segmentation method was suitable for analyzing bone tunnels for double bundle ACL reconstruction. The aforementioned studies have explored applications of ML algorithms in orthopedic injuries including ACL tears. The aim of this study is to understand optimal clustering methodologies for ACLR patient data, focusing on knee measurements and their association with demographic factors such as age, weight, height and activities.

### 2.3.2 Clustering Techniques for Medical Data Analysis

An important and effective method for organizing and analyzing data is clustering. In this study, I will use clustering as a methodology to analyze the data, which will be explained in Chapter 3. To that end, some extant studies that used clustering methods to examine medical data are enumerated here.

The purpose of Khan's [70] study was to determine brain cancers from MRI images using a graphical user interface (GUI) developed in MATLAB. The program used GUI to apply various combinations of segmentation, filters, and other techniques to process the images in order to achieve the best possible results. First, the researcher applied the Prewitt horizontal edge-emphasizing filter to the image. The next step in tumor detection was to identify "watershed pixels." The most important element of this project was the integration of all the MATLAB applications with the "MATLAB Guide" GUI. By applying the aforementioned filter combinations and using more image processing techniques, Khan [56] was able to effectively identify brain cancers in early stages. For a long time now, scientists have evinced interest in the development of biomedical image

processing. However, several issues have developed in this field, especially regarding MRI imaging. A major concern with the biological image segmentation approach for MRI images is its significant variability based on a particular application. For example, the segmentation software used for MRI imaging has different requirements than CT scan imaging segmentation. In addition, every image — even if it originates from the same imaging technique such as MRI — possesses unique features. A different set of MRI images may yield discrete results when interpreted by the same segmentation algorithm. To detect the tumor and determine its location with more precision and less computational time, Khan [70] developed another method for segmenting MRI brain images in this study. They focused on a new hybrid clustering approach that aimed to decrease computing time. Additionally, it proposed a binarization method for precisely calculating the area in terms of square millimeters (mm^2) using typography and digital image units. After determining the tumor's location and computing the CPU time, the researcher compared the simulation results with existing techniques and proposed shaft algorithm. This approach outperformed existing current algorithms along with a significant reduction in computing time.

On the other hand, Islam & Atiku [71] argued current clustering techniques are a significant method of data mining. The Internet of Medical Things (IoMT) is a developing field that influences knowledge extraction and networking, playing an important role in biomedical research, especially in remote monitoring and healthcare services. During the COVID-19 pandemic, its psychological impact on human reaction has become a critical field of study. Clustering methods have proven effective in identifying patterns and structures in both labeled and unlabeled datasets. The authors of

this study reviewed certain methods in medical IoMT for healthcare and used clustering processes. They used five selection methods: Minimum Redundancy — Maximum Relevance (mRMR), Random Forest, Normalized Mutual Information Feature Selection (NMIFS), F-Test, and Chi-Square, as well as five clustering algorithms — hierarchical clustering, DBSCAN, k-means, shrinkage clustering, and fuzzy c-means clustering.

Lötsch & Ultsch [72] noted that in biomedical research, clustering on projected data is a common practice, with Principal Component Analysis (PCA) often being the projection method. PCA focuses on data differences, while the clustering goal is to identify data focusing, which can sometimes lead to conflicting results. This study examined how PCA and other projection methods such as independent component analysis (ICA), Isomap, Multidimensional Scaling (MDS), and t-distributed Stochastic Neighborhood Embedding (t-SNE) perform when combined with common clustering algorithms such as k-means, k-medoids, and hierarchical clustering techniques. Thereafter this study evaluated these elements by using both artificial and real biomedical datasets. It also employed different assessments through Voronoi tessellation and standard cluster quality measures. The findings showed no single combination consistently captured the correct classifications in the projections and clusters, underscoring the importance of different inspections. Neighborhood-based approaches such as t-SNE or multiple learning techniques such as isomap consistently outperformed or matched PCA. The study suggested different comparisons of multiple projection and clustering methods instead of using PCA as the default projection method before clustering. A proposed visualization technique used Voronoi tessellation to evaluate the

effectiveness of different combinations, helping identify the best approach for a given dataset.

The purpose of a study by Takeda et al., [73] was to investigate patterns of joint destruction in knee radiographs of patients with rheumatoid arthritis (RA) who underwent total knee arthroplasty (TKA) over a 16-year period. They applied software to automatically measure 831 preoperative knee x-rays to determine medial and lateral joint spaces, spur period, and femoro-tibial angles. Subsequently, non-hierarchical clustering analysis was conducted based on these parameters. The analysis showed similarities in the radiographic features and distribution of different clusters over time. Results indicated that all measured parameters significantly increased between 2006 and 2021 except for the lateral spur area. The knee radiographs were placed in three clusters: Cluster 1 (conventional RA type), Cluster 2 (osteoarthritis type), and Cluster 3 (less destructive type). The results indicated a significant decrease in the proportion of patients in Cluster 1, but Clusters 2 and 3 showed significant increases. Also, the Cluster 3 patients had higher DAS28-CRP scores compared to those in Clusters 1 and 2. These findings indicate in recent decades, radiographs of RA patients undergoing TKA are increasingly showing features typical of OA, suggesting a shift in the disease pattern.

### 2.3.3 Using K-Mean Clustering in Synovial Fluid Studies

A perusal of extant literature shows multiple studies that used k-mean clustering to analyze synovial fluid. In their research, heard et al., [74] assessed whether specific Matrix Metalloproteinases (MMPs) and Tissue Inhibitors of MMPs (TIMPs) in synovial fluid could help distinguish between normal and early OA conditions. Utilizing multiplexing technology, the researchers examined the protein expression levels of 8

MMPs and 4 TIMPs in synovial fluid collected from patients with normal knee joints, early OA, advanced OA, and RA. Significant distinctions in MMP and TIMP expressions were noted between normal and arthritic synovial fluids, except for MMP 12. The Principal Component Analysis (PCA) technique was applied to identify influential MMPs in distinguishing between these groups. Furthermore, k-means clustering was utilized to confirm the identified clusters based on PCA results, revealing three distinct clusters: normal joints clustering with early OA, and separate clusters for advanced OA and RA. These findings suggest unique MMP and TIMP expression profiles exist in normal, advanced OA, and RA synovial fluids, enabling differentiation between these conditions. However, these profiles were not effective in diagnosing early OA. Nevertheless, high-throughput multiplex technology for MMPs and TIMPs may hold promise for assessing disease severity and treatment response monitoring.

On the other hand, the study by Heard et al., [75] explored whether cytokine/chemokine profiles in synovial fluid and sera could effectively differentiate between mild/moderate OA, normal individuals, and severe OA. Utilizing multiplex technology, the researchers quantified expression levels of 42 cytokines in the synovial fluid of individuals diagnosed with severe OA (n=20) and mild/moderate OA (n=12), as well as normal controls (n=34). They also examined the same 42 cytokines in serum samples from patients with severe OA (n=26), mild/moderate OA (n=74), and normal individuals (n=100). Comparisons between the treatment groups were followed by PCA and k-means clustering of the significantly different cytokines/chemokines to reveal groupings based on physician diagnosis. The findings showed differences in cytokine/chemokine levels in synovial fluid samples among control, mild/moderate OA,

and severe OA cases, as well as differences in serum samples between normal and mild/moderate OA, and control and severe OA. However, no distinctions were observed between mild/moderate and severe OA serum samples. The study suggested comparing cytokine/chemokine expression levels in synovial fluid and serum could potentially serve as a diagnostic platform for early OA detection. The k-means analysis validated visual groupings identified by PCA, particularly distinguishing between normal and mild/moderate OA serum samples with 96% accuracy for normal, and 93% for mild/moderate OA samples. This high-throughput and cost-effective assay may offer clinicians a valuable diagnostic tool to complement existing clinical and imaging methods for OA diagnosis.

## 2.4. Summary of the Chapter

In this chapter, I provided a detailed of previse research that done in synovial fluid and machine learning. This chapter was dividing two part the first part was about Synovial Fluid Analysis: Clinical Significance and Challenges, is further divided into seven parts for ease of understanding, Properties of synovial fluids, Previous research focused on synovial fluids and joint health; Previous studies on chondroitin sulfates and hyaluronic acid in synovial fluid and researching synovial fluid with people who have an anterior cruciate ligament (ACL). The second part was about Machine Learning, and it is divided into three sub-parts, Machine learning applications in medical research; Clustering techniques for medical data analysis; and Using K-means clustering in synovial fluid studies. The results of this review were that there many research investigating the use of machine learning in the medical field. However, there was

limited data of using k-mean cluster in synovial fluid research. This dissertation will investigate the gap of literature.

# CHAPTER THREE

# METHODOLOGY

## 3.1 The Objectives of the Study

The purpose of this study is to evaluate the effectiveness and performance of machine learning (k-means clustering algorithm) in analyzing synovial fluid compositions—specifically hyaluronic acid (HA), chondroitin sulfate (CS), chondroitin 6-sulfate (C6S), and chondroitin 4-sulfate (C4S)—across different ages and genders in both osteoarthritis (OA) patients and healthy patients. Also, this study aims to provide a comprehensive comparison of different machine learning (k-means clustering) algorithms as applied to datasets derived from the synovial fluid of OA patients containing HA, CS, C6S, and C4S. This chapter provides a description of the methodology of this study, the rationale for clustering (k-means and elbow) methods. This chapter also describes the data collection and analyses and the methods I use to validate the findings.

## 3.2 Machine Learning Techniques

### 3.2.1 Clustering

Clustering may be considered a data mining technique that is employed to examine variable data with multiple lots. Clustering is the procedure of gathering discrete data into clusters [23]. In the context of data management, it is important to categorize the data into clusters or sets of categories [23]. Clustering applications have diverse uses in different fields such as engineering, social science, and computer science [63]. Clustering algorithms vary according to diverse requirements; however, they are roughly categorized into two groups: hierarchical clustering, and partitional clustering [66].

Hierarchical clustering can store data in two types of mode, viz. agglomerative and divisive modes. Some examples include MONA, DIANA, CURE, and BIRTH. On the other hand, partial clustering does not employ a hierarchical structure and creates a partition of the data from some clusters only. Examples include k-means and k-medoids algorithms [63].

Clustering analysis is very useful as it helps to extract important information from data and may be used to identify interesting patterns in various digital data sets. Bioinformatics, pattern recognition, data mining, economics, and image processing are some of the fields where clustering is used [76]. The procedure of clustering involves grouping data based on their similarities within a certain dataset [77]

The main objective of cluster analysis is to partition a group, which is known as a cluster, into a homogeneous subgroup through selected similarity measures. In each subgroup, the similarity of the objects is comparatively higher than the objects that come from other subgroups [78]. It is important to mention that clustering is very useful for identifying statistical patterns in data. Continuous and discrete variables typically represent those patterns.

There are several clustering techniques that may be employed, with k-means being the oldest method. It continues to be widely used. Herein, the number of clusters represented by k is fixed, whereas the starting centroids are selected on a random basis. After the iteration of the algorithm, each data point is assigned to the closest centroid. The process is repeated until the convergence criteria are satisfied. Regardless of the widespread use of k-means, it does have certain drawbacks that necessitate an exploration of the techniques it employs. This would help in enhancing the final outcomes of the

analysis [76]. In this research, I will delineate the utilization of clustering methods through k-means.

### 3.2.2 K-means Clustering Methods

The method I use in this dissertation is k-means clustering. K-means is a type of data clustering that uses unsupervised machine learning. It can divide the unlabeled data into different groups that have the same variances. The resultant groups are clusters that have certain similarities [79]. K-means is also considered the most simple and common method of clustering that can effectively manage a lot of datasets. It was developed by McQueen in 1967 [80]. K-means has special applications in data mining and pattern recognition fields because it implements a straightforward clustering approach based on the Euclidean distance function. The main objective of this technique is to reduce the performance index of cluster and squares' sum errors. It also lessens the error criterion which is important for finding out the optimal number of clusters (k), which satisfy a fixed standard. Its functioning objectives include speed, efficiency, and conciseness [81]. K-means encompasses vector quantization techniques that can divide bigger datasets into smaller ones. Its algorithm is intended to divide the data points into (k) clusters and categorize them to produce Voronoi cells. The primary objective here is to make sure the squared distance between data points in a single cluster is minimized [80].

K-means uses a gradient method that is based on the objective function to reach a peak value. The computation process involves a gradient-based trend-seeking function; the process converges into a local minimum when the starting cluster centroids are not properly selected. This leads to ineffective clustering [81].

There are certain merits and demerits of using k-means as discussed by Ravel & Jani [76], asserted the clustering technique is very fast. It is useful for handling large datasets as its time complexity is *O(nkl)*. Here, *n* refers to the pattern numbers, *k* is the number of the clusters, and *l* refers to the iterations number. The method also uses Euclidean distance which makes it efficient for numbers with meaningful geometric and statistical results. On the other hand, although the starting value for *k* is crucial, there is no accurate procedure that can be used to select such a value. As a result, different clusters will have different k values. Moreover, starting cluster centroids are also important and if they are far from the main clusters of data, the algorithm can start producing unlimited iterations which leads to inaccurate clustering. Furthermore, the k-means technique is not suitable for clustering datasets that contain substantial noise.

Syakur et al [23] argued k-means is a clustering technique that partitions data set into a cluster called k. K-means is essentially a distance-based algorithm that subdivides the data into numerous clusters that have numerical attributes. I have enumerated below the steps that are usually followed:

1. Determine the maximum number of desired iterations and the optimal number of clusters, k.

2. Carry out the k midpoint cluster initialization process and then use the centroid count feature given by the following equation:

$$Ci = \frac{1}{M}\sum_{j=1}^{M} \text{Xj}$$

1

Equation 1 is done as much as p dimensions from i = 1 to i = p

3. Connect every data point from an observation to the nearest cluster. The following formula can be used to calculate Euclidean distance spacing measurements:

$$d = \sqrt{(\text{x1-x2})^2 + (\text{y1-y2})^2} \qquad 2$$

4. Rearranging the distribution of the data among groups by calculating the difference between the data and each category's centroid.

$$aij \begin{cases} 1 & d = min\{D(Xi,\ Ci) \\ 0 & otherwise \end{cases} \qquad 3$$

5. Determine the cluster midpoint's location once more. After comparing the shortest path from data xi to group k, calculate the membership value of point $xi$ to the centers of group $ci$, and assign c1 as the group's center to 1. This method uses distance and data membership value within the group to determine the objective function. There is an equation that can be used to determine the goal function as specified by MacQueen (1967): Let k be the dimensionless number of groups and n be the total amount of data. The membership value of data point xi in group c1 is represented by the value ai1, where and is a binary value between 0 and 1.

If the data point belongs to group c1, ai1 = 1; otherwise, ai1 = 0.

$$J = \sum_{i=1}^{n} \sum_{l=1}^{k} aicD(Xi, C1)^2 \qquad 4$$

6. Return to step 3 if there is a divergence from the cluster midpoint position or if the number of iterations is greater than the maximum number allowed. If not, give the result of the clustering.

Li [58], suggested k-means is a clustering algorithm that uses sample sets division. This means the clustering technique will divide the sample set into k subsets which will later form k classes. In other words, it will divide n samples into k classes. Furthermore, the distance between the class center and each sample is the smallest and every sample belongs to only one class, which means k-means clustering is a hard type of clustering.

A given dataset is classified according to the formation of clusters. The main objective is to define the k-centers. There is one center for each cluster. It is important to place these centers cautiously because discrete locations can yield variable outcomes [59]. Also, the functional aims of the k-means clustering can be applied to specific non-numeric datasets as well.

The primary advantages of adopting the k-means clustering approach are brevity, efficiency, and promptness. The following equation can be used to describe the procedure of clustering. Defines the total within-cluster variation as the sum of the squared distances between items and the corresponding centroid [81]:

$$W(Ck) = \sum_{Xi \epsilon Ck} (Xi - \mu k)^2 \quad 5$$

$W(Ck)$ represents the within-cluster total, $xi$ is a data point for a cluster, $Ck$ indicates a cluster for each data point, and $\mu k$ defines the mean value of the points that is assigned to

the cluster $Ck$. The sum of total within-clusters of the sum of squares measures compactness (TW) as follows [81]:

$$\mathrm{TW} = \sum_{K=1}^{K} W(C_k) = \sum_{k=1}^{k} \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad 6$$

I have used an unsupervised K-means clustering algorithm in this dissertation because the data lacks labels and is able to learn natural patterns within it, among several other reasons. Because the make-up of synovial fluid can change slightly with age, gender, and health status (OA vs. healthy), K-means can find the main group or cluster that may show the relationship or trend in the data. The dataset involves multi-dimensional information combined into one record, such as HA concentration, C6S, C4S, and C6S:C4S, among others. In this respect, K-means shall find groupings useful in high-dimensional data in exploring biological data with many demographic categories. Finally, the unsupervised nature of K-means makes the generation of hypotheses possible through patterns that are outlined and perhaps not foreseen. It enables the establishment of key findings for further research, such as the identification of specific biomarkers involved in the progression of OA across varying demographic groups.

### 3.2.3 Optimally Determine Number of Clusters (Elbow Method)

I used Optimally determine number of clusters (elbow method) because its important steps for find the number of k. The elbow method is a procedure that is used to attain the standard number of centroids (k) in k-means clustering approach [84]. In other words, the number of clusters can be determined by the elbow method. Here, certain

numbers of clusters need to be picked. The average of one or more clusters does not offer appropriate data modeling marginally. For graphing purposes, the variance percentage of the clusters can be plotted against the optimum number of clusters. The initial clusters will penetrate various amounts of information. However, at a specific point, the marginal number of clusters that are inserted will decline substantially and will offer an angle on the graph [81]. Additionally, the proper value of k can be referred to as the clusters' number, which is known as the elbow criterion. The main premises comprise starting at the value of k=2 and increasing it by one point and finding the cluster and the expenditure that comes with its training [81]. In order to locate the proper k-value, there is a need to continuously iterate from k=1 to k=n. Moreover, for every k value, there is a need to calculate the within-cluster sum of squares value known as WCSS [82].

In this approach for finding the appropriate k-value, we iterate continuously from k=1 to k=n. At each k value, we compute the within-cluster sum of squares (WCSS) value [82].

$$\mathrm{WSS} = \sum_{i=1}^{m} (Xi - Ci)^2$$

7

The Elbow method and k-means may be applied together to plot the relation between cluster names and the corresponding error. Furthermore, when the k value increases, the error will be reduced. As a result, the graph will show a gradual decrease until it is stabilized. For example, when we switch from k=2 to k=3, the error may moderately decrease. However, from k=3 to k =4, a significant drop will be formed at an elbow where k=3 and in must Elbow results K= 3. Thus, the optimum value of the cluster is calculated to be k=3 which uses the elbow and k-means method. This signifies the best

point of clustering. The elbow method can be represented by the sum of squared error [83]. The following equation shows the above relationship:

$$SSE = \sum_{k-1Xi}^{k} \sum_{\epsilon Sk} ||x_i - c_k \Big\|_2^2 \quad 8$$

3.2.4 K-Means ++

Some of the drawbacks of k-means are offset by the usage of k-means++. This is a regular k-means algorithm which adopts a more robust methos to initialize cluster centroids. This is significant because the success of k-means algorithm depends heavily on how well the centroids (also called mean points) are initialized [84]. When allowed to perform multiple trials, k-means++ yields substantially faster and more accurate results [85] than other k-means. When it is run, say, 2k times more, researchers may argue that the result has a constant approximation at least one time. Further, if k-means++ finds a number of new centers each time, its performance improves much more, as evidenced by experiments [85].

In this study, I have chosen to employ k-means clustering and the optimally determined number of clusters (elbow method) study since this is the most appropriate approach for my topic—it accorded me an opportunity to gather rich and deep information from different sources. However, there are fewer sources about k-means++ that may affect the results of this study.

## 3.3 Research Questions

The following research questions will be addressed in the study:

**RQ1: How does the k-means clustering algorithm expose distinct clusters within datasets that contain information found in the compositions of synovial fluid (HA, C6S, C4S, and C6S:C4S) in patients of all ages and genders who have osteoarthritis (OA) and are healthy?**

**RQ2: What are the characteristics of the clusters identified based on the concentrations of hyaluronic acid (HA), chondroitin sulfate (CS), chondroitin 6-sulfate (C6S), and chondroitin 4-sulfate (C4S) in the synovial fluid in osteoarthritis patients (OA) and healthy people?**

### 3.3.1 Hypotheses

1. The k-means clustering algorithm will identify distinct patterns in synovial fluid contents that characterize OA patients from healthy individuals.
2. Age and gender will significantly influence synovial fluid contents.
3. Machine learning (k-means clustering) algorithms will exceed traditional statistical methods in identifying patterns for OA diagnosis.

## 3.4 Data Collection Procedures

The data for this study has been taken from two previously published research. The first study is titled, "Age-related changes and sex differences in chondroitin sulfate isomers and hyaluronic acid in normal synovial fluid," by authors Uesaka, Miyazaki, and Ito [25]; it was published in 2004. The purpose of their study was to determine if gender and age had any effect on the HA and CS isomers found in normal synovial fluid. In healthy people of different ages, the researchers examined the amounts of HA, C6S, and C4S. Samples of synovial fluid were collected from 187 participants, ranging in age from 14 to 89 years. The results indicated significant negative correlation with age in C6S,

C4S, HA concentrations, and C6S:C4S. The concentration of HA did not show any differences related to gender. However, in most age groups, women exhibited significantly lower levels of C6S and C4S, as well as a lower C6S:C4S ratio, compared to men.

The second research is titled “Significance of chondroitin sulfate isomers in the synovial fluid of osteoarthritis patients,” by Uesaka, Nakayama, Yoshihara, and Ito [19], and It was published in 2002. This study for the osteoarthritis patients for the same authors for the first data. Researchers explored the relationship between types of CS— C6S, C4S, C6S:C4S—and HA in synovial fluid, as well as the extent of osteoarthritis (OA) in knee joints. The participants of their study comprised 133 patients with knee OA, and 27 healthy people. The findings indicated women had lower values than men in both groups. Age and sex showed a moderate inverse relationship with types of CS and HA. Specifically, sex was linked to C6S levels and C6S:C4S. However, among OA patients, age and sex had little correlation with the variables.

For the current research, I applied k-means for synovial fluid of healthy knee joints using Shinji et al.’s dataset [25]. I used k-means for synovial fluid of OA in knee joints on the data found in Uesaka et al.’s published article [19]

3.4.1 Criteria for Data Selection

I have used the following criteria to select the published data for machine learning (k-means clustering):

1- I found there is a lack of use of any ML k-means clustering algorithm in synovial fluid composition in knee joint research, so further medical studies on the efficiency of such clustering are important.

2- The data I have used from the two extant studies [19,25] were focused on age-related changes and sex differences in synovial fluid composition. Since the purpose of this dissertation is to analyze synovial fluid components among different ages and genders by using ML (k-means clustering), I deemed these available datasets suitable.

3- These datasets include synovial fluid samples from a broad population range, including people of different ages and genders. This broad coverage implies the k-means clustering will have enough data variability to identify significant patterns in the clustering analysis.

4- The data from the two studies [19,25] were already organized in a way that could be directly used for ML without extensive additional preprocessing. The measurement of synovial fluid components (HA, CS, C6S, C4S) was suitable for k-means clustering techniques.

5- The k-means clustering method will be able to identify natural groupings or clusters within the dataset that may not have been evident from the correlation analysis methods used in either of the studies [19,25].

### 3.4.2 Study Participants

The participants (healthy group) from [25] for my study were 187 volunteers—99 male and 88 female. Their ages ranged from 14 to 89 years. The total number of people in each group was as follows: 10 in the age group of 10 to 19 years, 29 in the age group of 20 to 29 years, 21 in the age range of 30 to 39 years, 15 in the age group of 40 to 49 years, 26 aged between 50 to 59 years, 21 aged 60 to 69 years, 36 aged 70 to 79 years, and 29 in the age of 80 to 89 years. The participants were in good health condition without any history of knee diseases, injuries, obesity, or abnormalities in lower part

alignment. All participants volunteered for the study and signed the consent form, following the ethical guidelines the research ethics committee put in Nippon Medical School and Kitamurayama Public Hospital in Tokyo, Japan.

The participants of my OA group from [19], were 133 patients with OA of the knee: 37 male and 96 female. Their ages ranged from 50 to 90 years. The total number of people in each OA age group was as follows: 27 aged 50 to 59 years, 59 in the age range of 60 to 69 years, 34 in the age range of 70 to 79 years, and 13 people aged 80 to 89 years. All participants volunteered for the study and had signed the consent form, following the ethical guidelines that the research ethics committee put in Nippon Medical School and Kitamurayama Public Hospital in Tokyo, Japan.

3.4.3 Sample Withdrawal

Both Uesaka et al [19] and Uesaka et al [25], explained that after the knee joint was thoroughly cleaned and disinfected, the synovial fluid was aseptically collected using a 21-gauge needle through the lateral infrapatellar route. To remove cells and joint debris, the fluid samples were collected in sterile plastic tubes and centrifuged at 10,000 g for 15 minutes at 4°C. Until the biochemical tests were carried out, the resultant supernatants were kept at -80°C. A volume of 0.2 milliliters of synovial fluid was used for each experiment. Thereafter, the necessary amount of HA, CS, C6S, and C4S were extracted. The same methods were used for OA patients. The summary of biochemical results showing synovial fluid concentrations of CS and HA in healthy groups of different ages and genders is provided here Table 3.1. The summary of biochemical results highlighting synovial fluid concentrations of CS and HA in the OA group across different ages and

genders is shown here in Table 3.2 This data used for this study to apply the K means approach combined with the optimally determine number of clusters (elbow method).

Table 3.1. Data characteristics for healthy group from Uesaka et al.'s study (2004), [25].

| Age | Man | Women | Total | Volume (ml) | C6S (nmol/ml) | C4S (nmol/ml) | C6S: C4S | HA (mg/ml) |
|---|---|---|---|---|---|---|---|---|
| **(10-19)** | 6 | 4 | 10 | 0.5 ± 0.3 | 136.7 ± 47.8 | 22.7 ± 7.9 | 6.2 ± 1.5 | 3.8 ± 0.7 |
| **(20 - 29)** | 19 | 10 | 29 | 0.6 ± 0.3 | 144.1 ± 66.2 | 21.0 ± 7.2 | 6.7 ± 1.2 | 3.4 ± 0.6 |
| **(30 - 39)** | 14 | 7 | 21 | 0.5 ± 0.3 | 120.0 ± 31.2 | 19.8 ± 6.3 | 6.2 ± 0.8 | 3.4 ± 0.8 |
| **(40 - 49)** | 9 | 6 | 15 | 0.5 ± 0.3 | 105.8 ± 30.5 | 18.8 ± 4.8 | 5.7 ± 1.1 | 3.2 ± 0.5 |
| **(50 - 59)** | 18 | 8 | 26 | 0.5 ± 0.3 | 101.1 ± 33.4 | 19.1 ± 5.4 | 5.3 ± 1.2 | 3.2 ± 0.6 |
| **(60 - 69)** | 11 | 10 | 21 | 0.4 ± 0.3 | 89.3 ± 32.3 | 18.4 ± 4.9 | 4.8 ± 1.0 | 2.9 ± 0.5 |
| **(70 - 79)** | 7 | 21 | 36 | 0.6 ± 0.3 | 72.4 ± 28.4 | 15.7 ± 33.3 | 4.6 ± 1.4 | 2.5 ± 0.6 |
| **(80 - 89)** | 15 | 22 | 29 | 0.6 ± 0.3 | 62.2 ± 26.5 | 14.5 ± 3.9 | 4.2 ± 1.1 | 2.1 ± 0.7 |
| **Total** | 99 | 88 | 187 | 0.5 ± 0.3 | 99.3±47.7 | 18.2±5.8 | 5.3±1.4 | 2.9±0.8 |

Table 3.2 Data characteristics for OA group from Uesaka et al.'s study (2002), [19]

| Age | Man | Women | Total | Volume (ml) | C6S (nmol/ml) | C4S (nmol/ml) | C6S: C4S | HA (mg/ml) |
|---|---|---|---|---|---|---|---|---|
| **(50 - 59)** | 8 | 19 | 27 | 9.9 ± 6.3 | 119.1 ± 66.1 | 22.8 ± 8.3 | 5.1 ± 1.2 | 1.8 ± 0.6 |
| **(60 - 69)** | 19 | 40 | 59 | 11.1 ± 9.1 | 72.8 ± 42.4 | 17.2 ± 6.7 | 4.1 ± 1.0 | 1.9 ± 0.5 |
| **(70 - 79)** | 7 | 27 | 34 | 9.2 ± 9.9 | 56.2 ± 22.2 | 16.8 ± 5.5 | 3.3 ± 0.8 | 1.8 ± 0.7 |
| **(80 - 89)** | 3 | 10 | 13 | 8.7 ± 9.5 | 44.9 ± 13.3 | 17.0 ± 3.9 | 2.6 ± 0.4 | 2.1 ± 0.5 |
| **Total** | 37 | 96 | 133 | 10.1 ± 8.8 | 75.2±48.7 | 18.2±6.9 | 4.0±1.2 | 1.9±0.6 |

### 3.4.4 Explanation Terms

**Synovial fluid**: The viscous material found in the cavities of synovial joints is called synovial fluid. Reducing friction between the articular cartilages of synovial joints when in motion is the main purpose of synovial fluid [16].

**Chondroitin sulfate (CS):** Synovial fluid contains chondroitin sulfate a glycosaminoglycan (GAG), which is a major component of the extracellular matrix in connective tissues. It is an extended sequence of repeating sugar molecules [24].

**Chondroitin-6-sulfate (C6S):** A subtype of chondroitin sulfate, CS6 is found in cartilage, connective tissues, and synovial fluid, among other tissues. The structure of each isomer of chondroitin sulfate, such as C4S (chondroitin-4-sulfate) and C6S, varies due to changes in the location and number of sulfate groups attached to the sugar units [24].

**Hyaluronic acid (HA):** HA reduces inflammation and preserves the natural cartilaginous matrix, which helps to maintain increased fluid viscosity and the integrity of joints [25].

## 3.5 Data Processing

I used MATLAB, a powerful language designed for technical computing, to process the data. It combines programming, calculation, and visualization in a user-friendly environment where issues and their fixes are presented using familiar mathematical symbols [86]. MATLAB has multiple typical functions including development algorithms, clustering, math and computation, data analysis, and engineering graphics, and application development [86]. I applied k-means clustering, an unsupervised ML algorithm within MATLAB, to manage the data through the coding process.

Many ML estimators within the MATLAB require scaling datasets to standardize independent variables' range. However, such scaling often prevents the estimator from correctly gleaning information from other features [81]. It may so happen that the estimator makes mistakes when individual features fail to resemble usual data distribution patterns, Gaussian processes with a zero mean and no variance across units. When working with datasets, the distribution shape and data centralization are ignored when each feature's mean value is taken off. In this, standard deviation representation looks like this:

$$Z = \frac{\aleph - \mu}{\sigma} \qquad 9$$

In the above specification, the normalized or standardized value is represented by z, whereas x is the data point's raw value, μ stands for the population mean, while the dataset's standard divisor for the population is σ [81].

I created a MATLAB code for the process. The flowchart in Figure 3.1 shows the processing of a dataset using the elbow and k-means clustering approaches to identify optimal clustering and yield results. The database for synovial fluid characteristics C6S, C4S, C6S:C4S, and HA for the healthy group were recorded to analyze and compare the difference between ages and genders. The dataset was processed in five steps. The first step was to input the datasets by extracting information from both, including both demographic information (age and gender) and synovial fluid components such as C6S and C4S, C6S ratio, and HA. These variables were used in the clustering analysis to identify patterns and differences among participants based on age and gender. The second step comprised data preprocessing through scaling and standardization, which ensured consistency and comparability across all parameters. The third step used the elbow method to find the best number of clusters. This method involved adding up all the squared errors (SSE) for all possible cluster numbers to find the point where adding more clusters stopped making a big difference in the SSE. The fourth step involved running the k-means clustering algorithm and using the dataset to determine the range of cluster numbers, as well as calculating the SSE to evaluate the compactness of each cluster. The fifth step was the optimal result, followed by output analysis, visualization, and presentation in an organized and meaningful way. The process used was based on the MATLAB code I created for the database of synovial fluid characteristics C6S, C4S, C6S:C4S, and HA for the OA group.

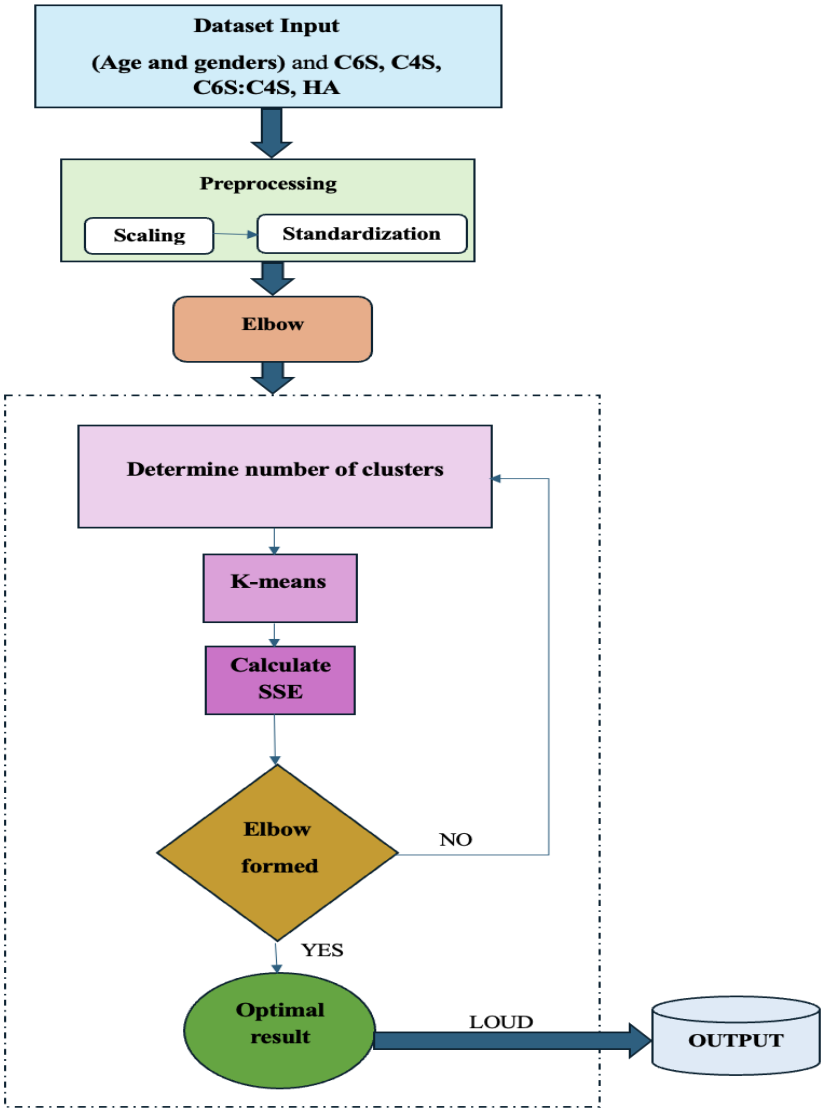


Figure 3.1 Process flowchart for the proposed system.

## 3.6 Strategies to Validate Findings

Validity indicates how accurate a study is, and how well the results may be generalized or trusted. Validity is a significant strength of a study, determined by the accuracy of the findings from the perspectives of both researcher and participants [87]. Validity guarantees reliability and credibility [87]. In this study, which applies k-means clustering and the elbow method to analyze synovial fluid components among different ages and genders, I used both internal validity (how the study's design, data, and methods support its conclusions) and external validity (generalization of findings). I also conducted the peer debriefing, which provided an extra check of the results of this research. I shared the study process with my committee members at Oakland University.

## 3.7 Research Ethics

It is important to follow ethical guidelines to ensure data is dealt with appropriately and the integrity of research is maintained. I adhered to procedures that comply with the rules and guidelines of Oakland University. I obtained permission from my adviser to use the published data. I cited the sources of my data. I ensured my research protects the privacy and confidentiality of participants whose data may be included. The authors of my data sources had secured informed consent from all participants of their research. I took care to avoid data misrepresentation or manipulation. All the data I used, and the results I generated from k-means clustering, are accurate to the best of my knowledge. I employed machine learning techniques—such as the k-means clustering— responsibly, with a full understanding of their limitations and potential biases.

## 3.8 Summary of Chapter Three

In this chapter, I have explained the methodology used in this research—the clustering approach, k-means clustering algorithm, and elbow methods. The data source for this study has been taken from two previously published research, one for the healthy group and the other for the OA group. I also delineated details about study participants and the process of sample withdrawal. In Chapter Four, I report the results of this study.

# CHAPTER FOUR

# RESULTS OF STUDY

The purpose of this study was to evaluate the effectiveness and performance of machine learning (k-means clustering algorithm) in analyzing synovial fluid compositions—specifically HA, CS, C6S, and C4S—across different ages and genders in both OA patients and healthy people. Furthermore, this study aimed to provide a comprehensive comparison of different machine learning (k-means clustering) algorithms as applied to datasets derived from the synovial fluid of OA patients containing HA, CS, C6S, and C4S. In this chapter, I present the results of the study in two parts for ease of understanding. In part one, I provide detailed description of the k-means results for the healthy groups. In second part two, I furnish a detailed description of the k-means results for the OA group.

## 4.1 Results of K-Means Clustering for Healthy Group

To interpret the results of the k-means clustering algorithm from Uesaka et al.'s [25] data. The heatmaps were selected as one of the multiple ways to visualize the clustering results. The first result is the heatmap for the normal group, as shown in Figure 4.1. This shows the correlation matrix for healthy group data with various measurements. The heatmaps provide details of relationships between different synovial fluid compositions across various age groups and genders. Each cell determines the correlation coefficient between two variables on the axis, which ranges from -1 to 1. A coefficient close to -1 shows a negative correlation between variables, which means one variable increases as the other decreases. A coefficient close to 1 indicates a strong positive

correlation, meaning one variable increases as the other increases. When the coefficient is zero, there is no correlation between variables.

In addition, the color yellow indicates a strong positive correlation between variables, while blue indicates negative coloration between variables. Figure 4.1 shows a weak negative correlation between age and other variables. For example, the correlations between age and HA, C4S, and C6S: C4S had very weak negative coefficients at around -0.96. Between age and C6S coefficients, this was around -0.98. This result indicates there is no significant relationship between these variables in the healthy group.

Furthermore, women had a very weak negative correlation with HA, C6S, C4S, and C6s: C4S. The coefficients at -0.93 to -0.76 indicate there was an almost weak relationship between these variables in the healthy group. Women also had a very weak negative correlation with HA, C6S, C4S, and C6s: C4S. The coefficients at -0.93 to 0.76 implies an almost weak relationship between these variables for the healthy group. However, a weak positive correlation was revealed between men with HA—with a coefficient of 0.12—and men with C4S, with a coefficient of 0.09. Additionally, the correlation coefficients for men with C6S and men with C6S: C4S were 0.22 and 0.28 respectively, indicating a positive correlation. The coefficient of age with synovial fluid volume was 0.16, which means a weak positive correlation. The coefficient of men with synovial fluid volume was 0.22, a weak positive correlation. However, a strong positive correlation was revealed between women with volume with a coefficient of 0.63. These results indicate as people age, the levels of synovial fluid components tend to decrease significantly. In addition to age, gender appears to be another important factor influencing synovial fluid composition, as the results indicate women had negative

correlations with the concentrations of C6S, C4S, and HA but the men had positive correlations.

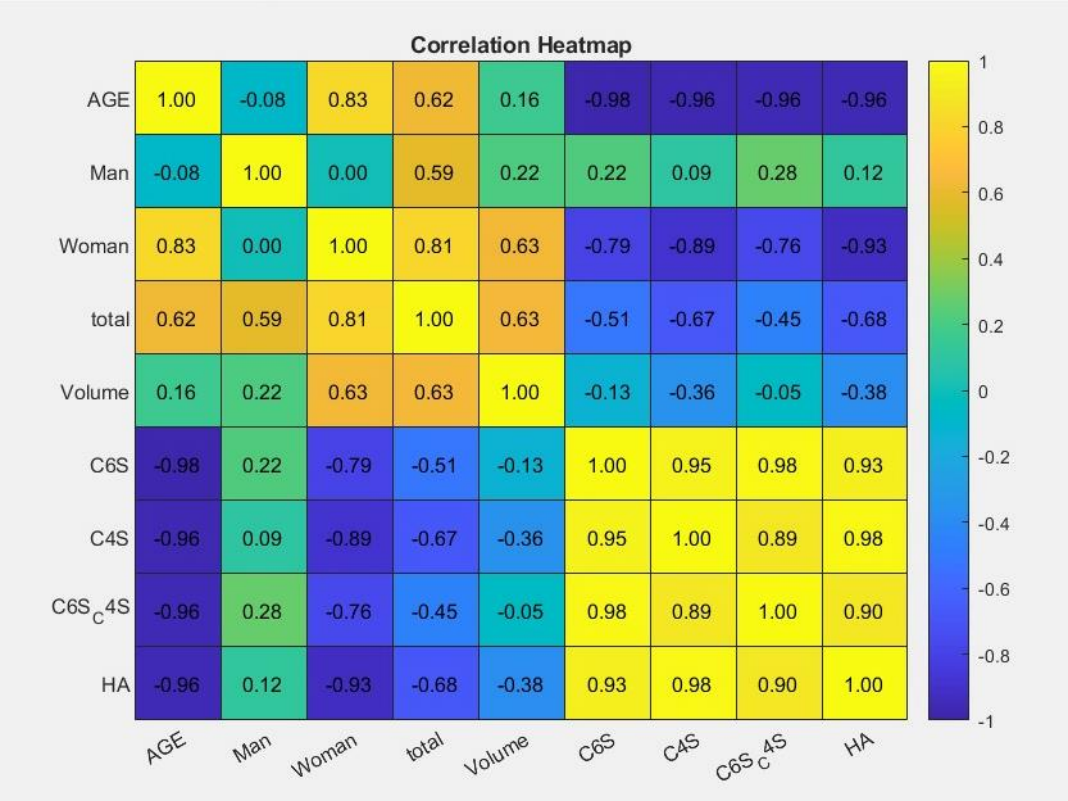


Figure 4.1. Heatmap for healthy group.

The second way I visualized k-means clustering was through scatter plots. To start the k-means procedure, it is important to specify the number of k, which represents the number of clusters in the data. This input is critically important to the quality of clusters, particularly in datasets with more than three variables [77]. In this study, the elbow method determined the optimal number of clusters (k) for k-means clustering. Figure 4.2. indicates the elbow graph and shows the within-cluster sum of squares (WCSS) for different numbers of clusters, from 1 to 10, before and after scaling. The

point at which the curve starts to stabilize (the elbow point) is where the convergence criterion is reached [77].

K values in Figure 5 before scaling:

1- k=1: WCSS appears to be around 60.

2- k=4: WCSS appears to be around 10.

3- K=7 WCSS appears to be around 1 or 2.

After scaling:

1- k=1: WCSS also appears to be around 60.

2- k=4: WCSS appears to be around 10, similar to the unscaled case.

3- k=7: WCSS appears to be around 1 or 2, similar to the unscaled case.

In this plot, the optimal number of clusters (elbow) before scaling appeared to be around K = 3 but after scaling, the elbow appeared to be around K = 4. As the curve flattened beyond K = 4, the WCSS continued to decrease the elbow metric until it reached zero (60).

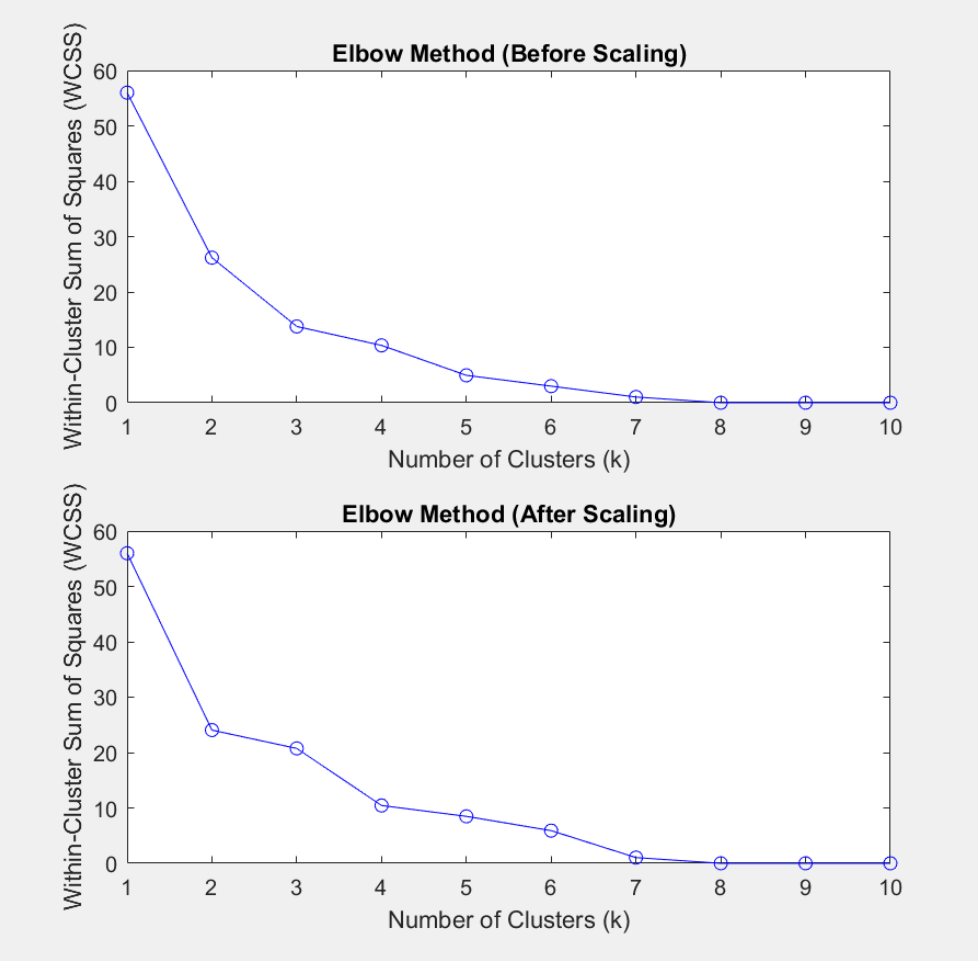


Figure 4.2. The Elbow Method.

## 4.2 Results of K-Means Clustering for Healthy Group with Age Factor, CS and HA in Synovial Fluid

### 4.2.1 Results for Variables Age and C6S

The scatter plot in Figure 4.3. represents the results of k-means clustering for the healthy group based on two variables: age on the x-axis, and C6S concentration (nmol/ml) on the y-axis. The age range was 10 to 90 years, with C6S concentration (nmol/ml) ranging from 0 to 300. The red, black, and blue points indicate the identification of three distinct clusters with centroids marked by black Xs. Each color represents a different cluster for the k-means algorithm. The red cluster comprises

younger people (aged roughly between 10 and 50 years) with higher C6S levels ranging from 120 to 160 nmol/ml. The black cluster includes older people (aged approximately 80 to 90 years) with the lowest C6S levels ranging between 50 and 80 nmol/ml. The blue cluster comprises middle-aged people (aged 60 to 70 years) with moderate C6S levels ranging from 80 to 100 nmol/ml. The large 'X' markers represent the clusters' centroids, which are the average of all points within each cluster. These centroids are the points that minimize the total distance from all points within the cluster.

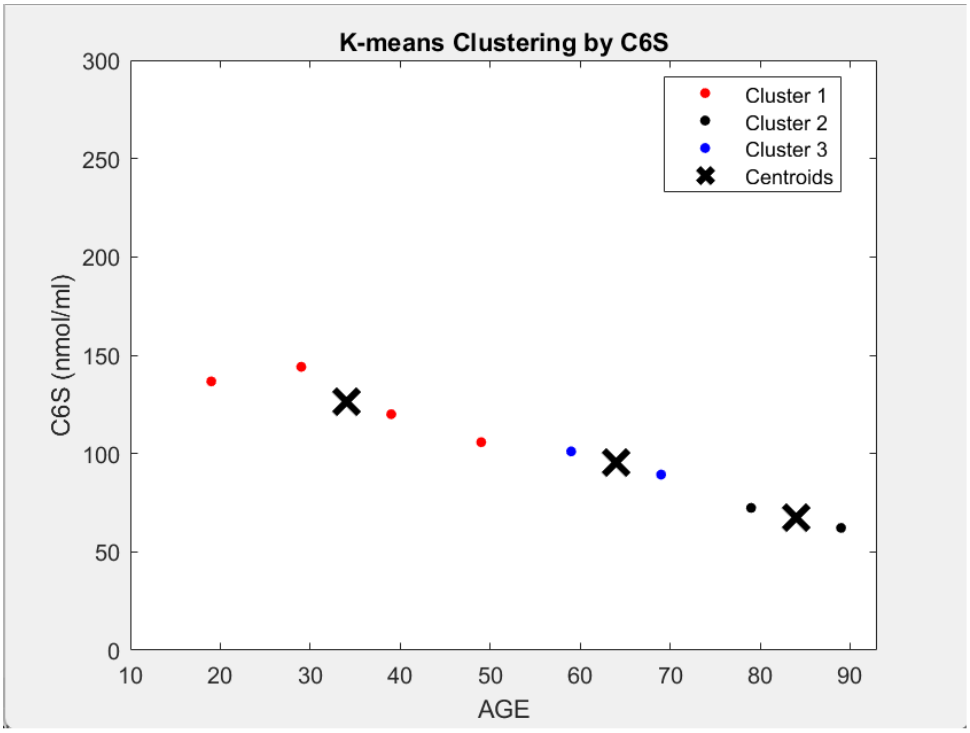


Figure 4.3 Result of k-means cluster for variables age and C6S.

4.2.2 Results for Variables Age and C4S

The scatter plot in Figure 4.4 represents the results of k-means clustering for the healthy group based on two variables: age on the x-axis and C4S concentration (nmol/ml) on the y-axis. The age range was 10 to 90 years, with C4S concentration (nmol/ml) ranging from 5 to 45. The red, black, and blue points indicate the identification of three distinct clusters with centroids marked by black X. Each color represents a different cluster for the k-means algorithm. The red cluster comprises younger people (aged roughly between 10 and 50 years) who had higher C4S levels, ranging from 15 to 25 nmol/ml. The black cluster includes older people (aged approximately 80 to 90 years) with the lowest C4S levels ranging between 10 and 15 nmol/ml. The blue cluster comprises middle-aged people aged between 60 to 70 years, who had moderate C4S levels ranging from 15 to 20 nmol/ml.

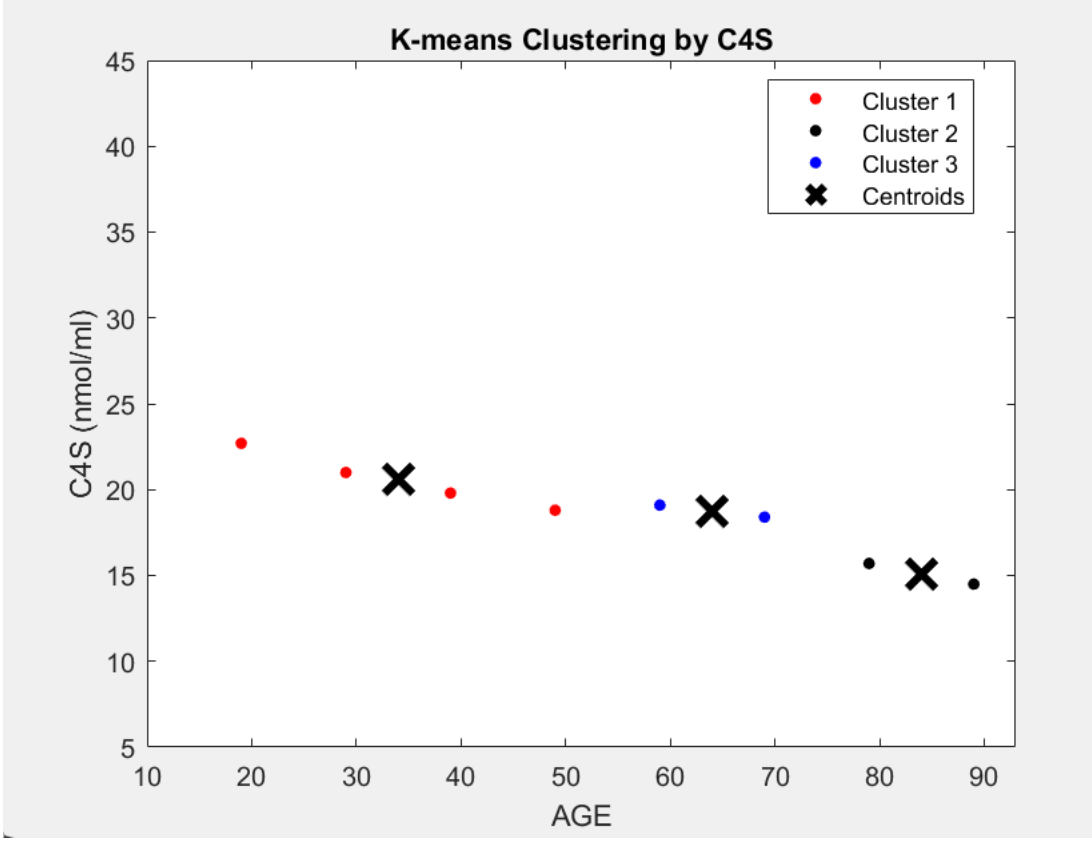


Figure 4.4. Result of k-means cluster for variables age and C4S.

4.2.3 Results for Variables Age and C6S:C4S

Figure 4.5. shows the results of k-means clustering on data with two features: age on the x-axis and C6S:C4S concentration (nmol/ml) on the y-axis. The age range was 10 to 90 years, with C6S:C4S concentration (nmol/ml) ranging from 2 to 10. The first cluster is red and constitutes younger people (aged roughly between 10 and 50 years) with higher C4S levels ranging from 5 to 7 nmol/ml. The second cluster is black, including older people (aged approximately 80 to 90 years) with the lowest C6S:C4S levels, between 4 and 5 nmol/ml. The blue cluster comprises middle-aged people aged between 60 and 70 years, who had moderate C6S:C4S levels, ranging from 5 to 6 nmol/ml. The large 'X'

markers represent the clusters' centroids, which are the average of all points within each cluster. These centroids are the points that minimize the total distance from all points within the cluster.

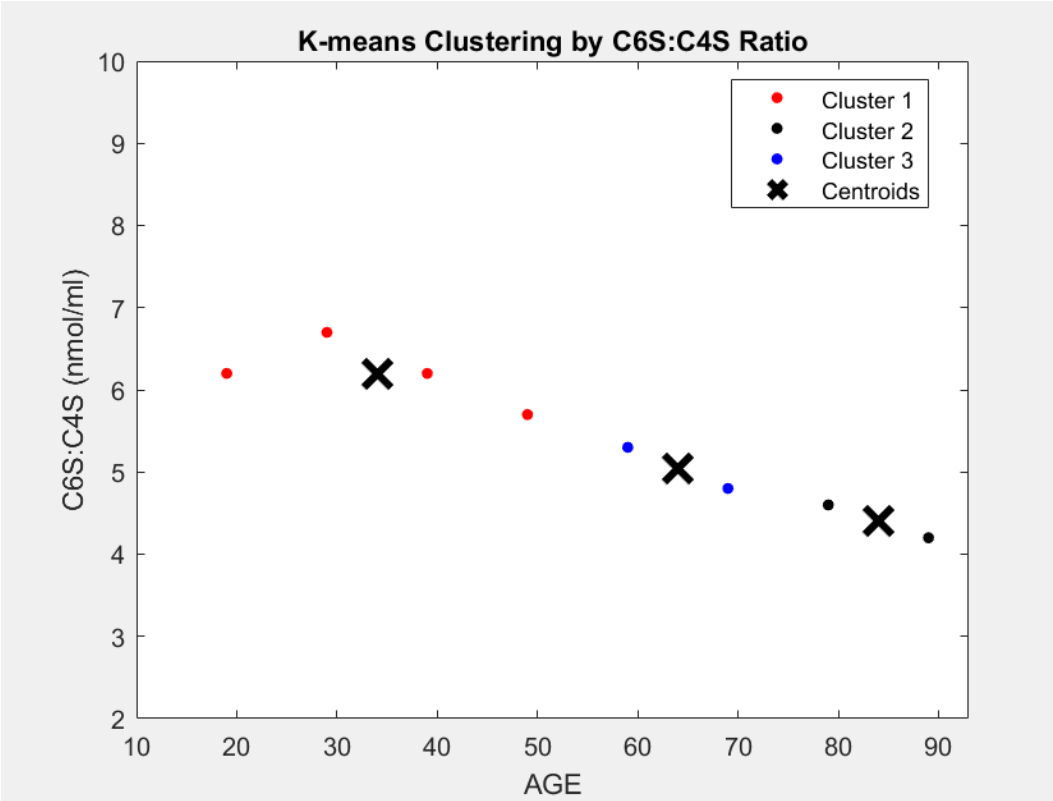


Figure 4.5. Result of k-means cluster for variables age and C6S:C4S

4.2.4 Results for Variable Age and HA

Figure 4.6. presents the k-means clustering results for two features: age on the x-axis and HA concentration 1 to 5 (nmol/ml) on the y-axis. The first cluster is red and comprises younger people (aged roughly between 10 and 50 years) with higher HA levels ranging from 3 to 4 nmol/ml. The second cluster is black, including older people (aged approximately 80 to 90 years) with the lowest HA levels between 2 to 2.5 nmol/ml. The

blue cluster consists of middle-aged people who were between 60 and 70 years old, who had moderate HA levels ranging from 3 to 3.5 nmol/ml.

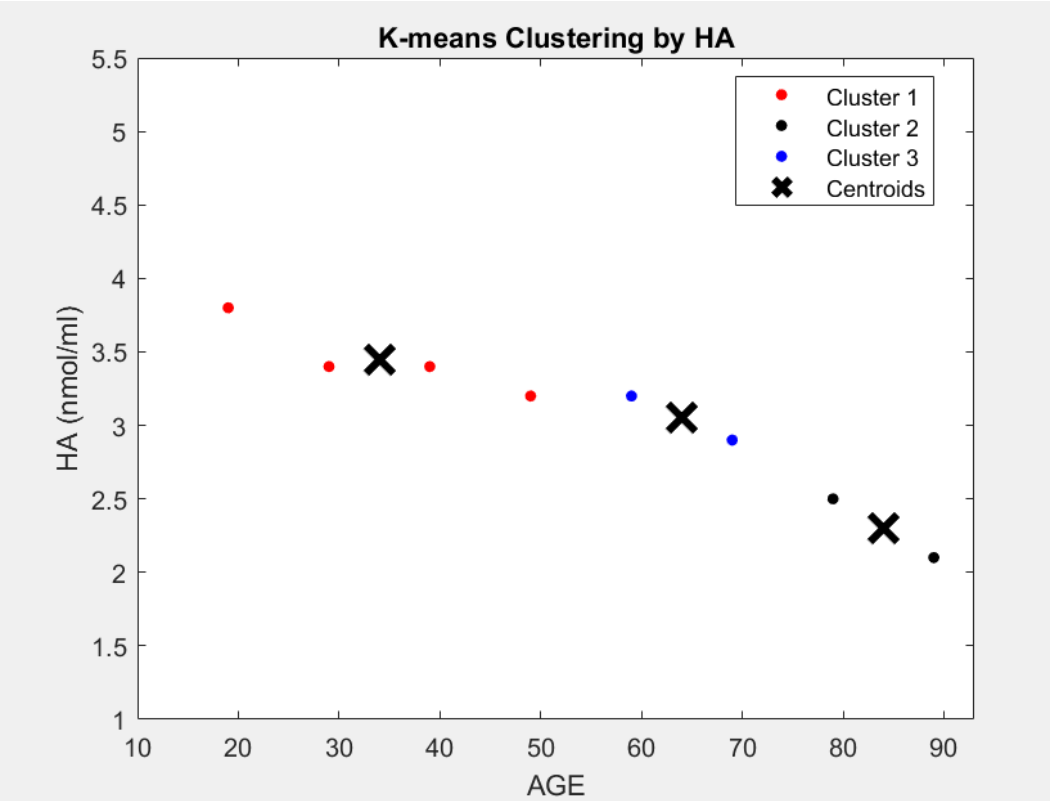


Figure 4.6. Result of k-means cluster for variables age and HA

In conclusion, the k-means clustering analysis indicates clear related patterns of C6S, C4S, C6S:C4S, and HA in synovial fluid. The results show younger people (10 to 50 years of age) consistently demonstrate higher levels of synovial fluid concentrations, while older people (80 to 90 years old) have the lowest concentrations. However, people in the middle age group (60 to 70 years) are in a moderate range. These results reveal C6S, C4S, C6S:C4S, and HA in synovial fluid decrease as age increases. These clusters

indicate a negative correlation between age and changes in synovial fluid composition for the healthy group.

## 4.3 Results for K-Means Cluster for Men in Healthy Group with CS And HA in Synovial Fluid

### 4.3.1 Results for Men with C6S

The scatter plot in Figure 4.7. illustrates the results of k-means clustering for the healthy group based on two variables: men on the x-axis and C6S concentration (nmol/ml) on the y-axis. The men's age span is 10 to 90 years, with C6S concentration (nmol/ml) ranging from 0 to 300. There are three clusters—red, black, and blue points—with centroids marked by black Xs. Each color represents a different cluster for the k-means algorithm. The red cluster has men aged between 30 and 60 years with C6S levels ranging from 100 to 150 nmol/ml. The black cluster includes older men (approximately 80 to 90 years old) with the lowest C6S levels between 50 and 70 nmol/ml. The blue cluster comprises men aged between 20 to 60 years with moderate C6S levels ranging from 100 to 150 nmol/ml. The centroid of the red and blue cluster is pointed around age 50 with a C6S of around 120 nmol/ml, but the centroid of the black cluster is pointed around age 85 with a C6S of around 70 nmol/ml.

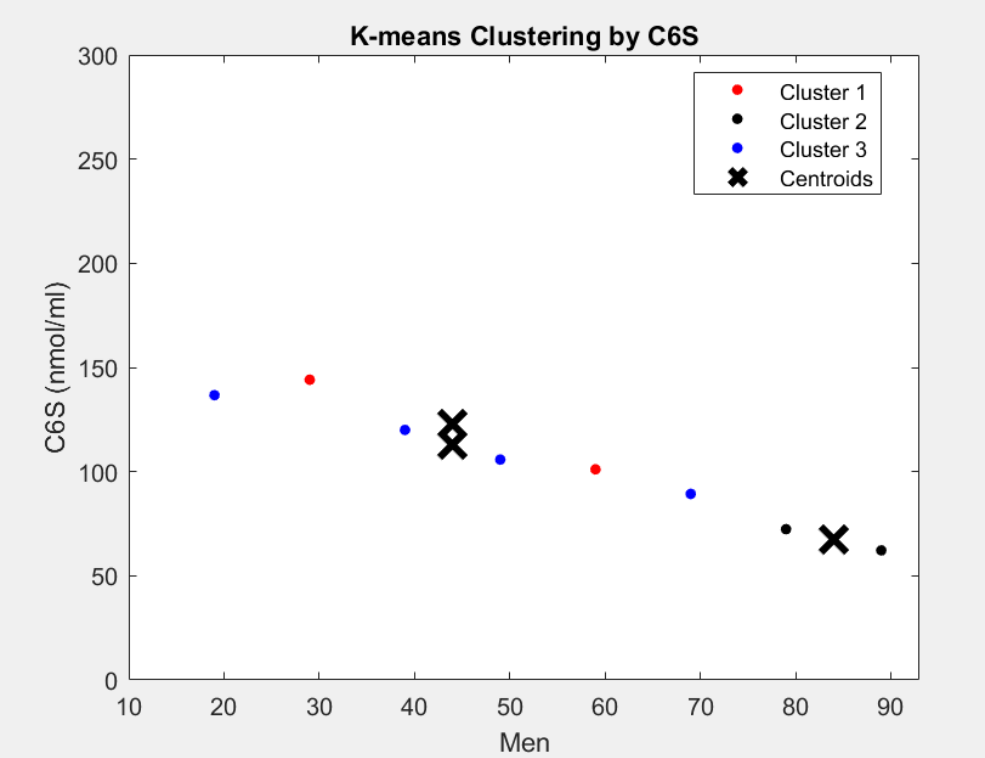


Figure 4.7. Result of k-means cluster for men with C6S.

4.3.2 Results for Men with C4S

The healthy group's k-means clustering results are shown in Figure 4.8. as a scatter plot. The x-axis shows age of the men, and the y-axis conveys the C4S concentration (nmol/ml). The age range is 10 to 90 years, with C4S concentration (nmol/ml) ranging from 5 to 45. There are three clusters: red, black, and blue. Each color represents a different cluster for the k-means algorithm. The red cluster comprises men aged roughly between 25 and 60 years with higher C4S levels, ranging from 19 to 22 nmol/ml. The black cluster includes older men, aged approximately 80 to 90 years with the lowest C4S levels between 10 and 15 nmol/ml. The blue cluster has men aged between 10 and 70 years who have moderate C4S levels ranging from 15 to 23 nmol/ml. The centroid for red and blue clusters are around age 45 with a concentration of

approximately 21 nmol/ml, but the centroid for black cluster is around age 85 with a concentration of approximately 15 nmol/ml.

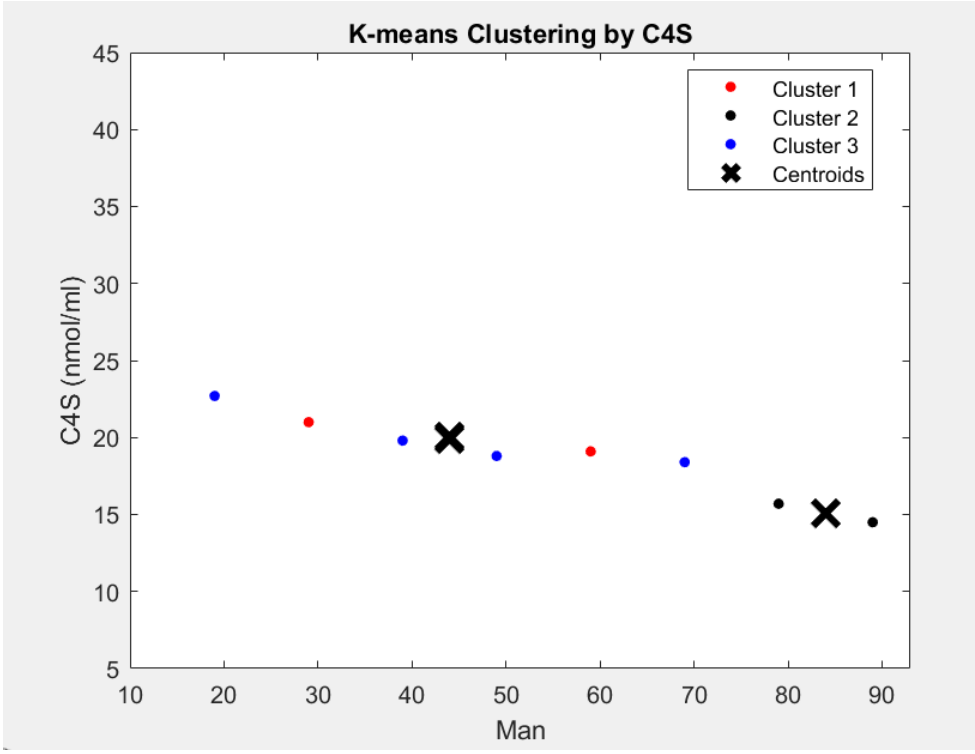


Figure 4.8. Result of k-means cluster for men with C4S

4.3.3 Results for Men with C6S:C4S

Figure 4.9. highlights the results of k-means clustering on data with two features: men's age on the x-axis and C6S:C4S concentration (nmol/ml) on the y-axis. The age range is 10 to 90 years, with C6S:C4S concentration (nmol/ml) ranging from 2 to 10. The first cluster is red and comprises men aged roughly between 30 and 60 years, who had higher C4S levels ranging from 5 to 6 nmol/ml. The second cluster is black, including older men aged approximately 80 to 90 years with the lowest C6S:C4S levels between 4 and 5 nmol/ml. The blue cluster comprises men aged between 10 and 70 years with

moderate C6S:C4S levels ranging from 5 to 7 nmol/ml. The centroid for the clusters red and blue at age 40 and a ratio of about 6.5, but the centroid for the black cluster is around age 85 with a concentration of approximately 15 nmol/ml.

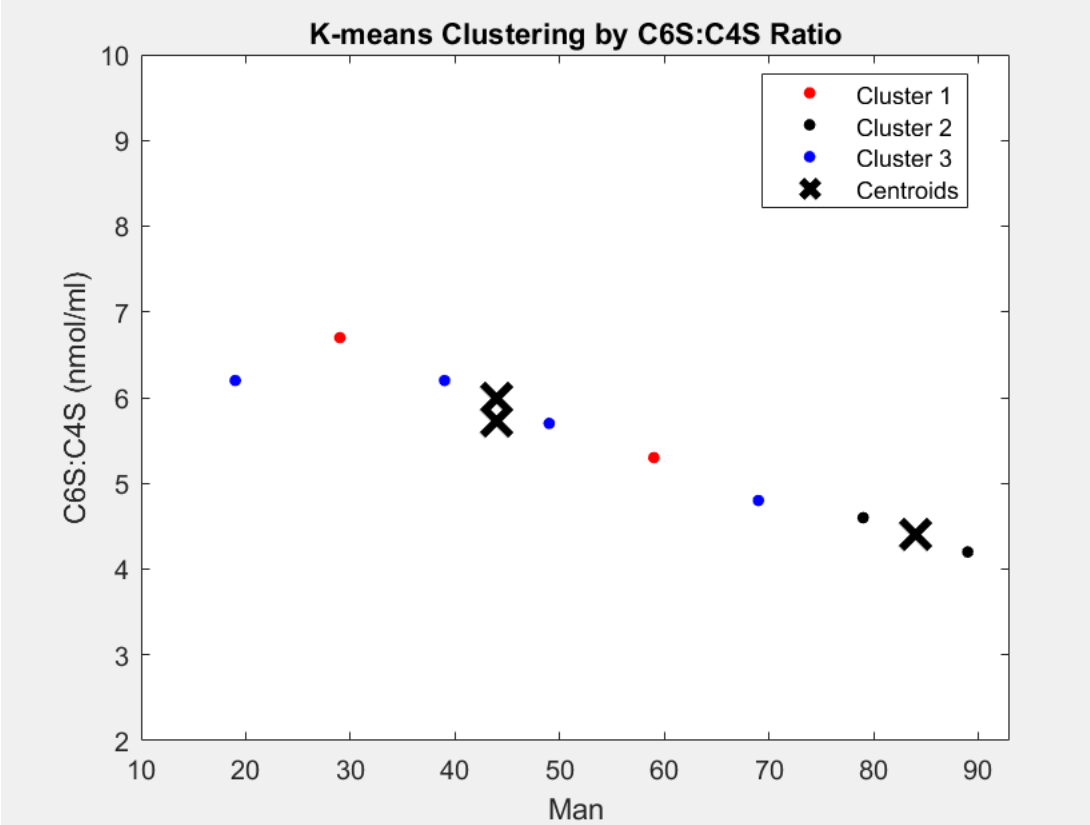


Figure 4.9. Result of k-means cluster for men with C6S:C4S.

4.3.4 Results for Men with HA

The scatter plot in Figure 4.10. reveals the k-means clustering results for two features: men's age on the x-axis and HA concentration of 1 to 5.5 (nmol/ml) on the y-axis. The first cluster is red and contains men aged roughly between 30 and 60 years with HA levels ranging from 3 to 3.5 nmol/ml. The second cluster is black, including older men aged approximately 80 to 90 years; with the lowest HA levels between 2 and 2.5

nmol/ml. The blue cluster comprises men aged between 10 to 70 years, who had HA levels ranging from 2.9 to 3.8 nmol/ml. The large 'X' markers represent the clusters' centroids, which are the average of all points within each cluster. The centroid for the clusters red and blue is at age 40 and a ratio of about 3.4, but the centroid for the black cluster is around age 85 with a concentration of approximately 2.2 nmol/ml.

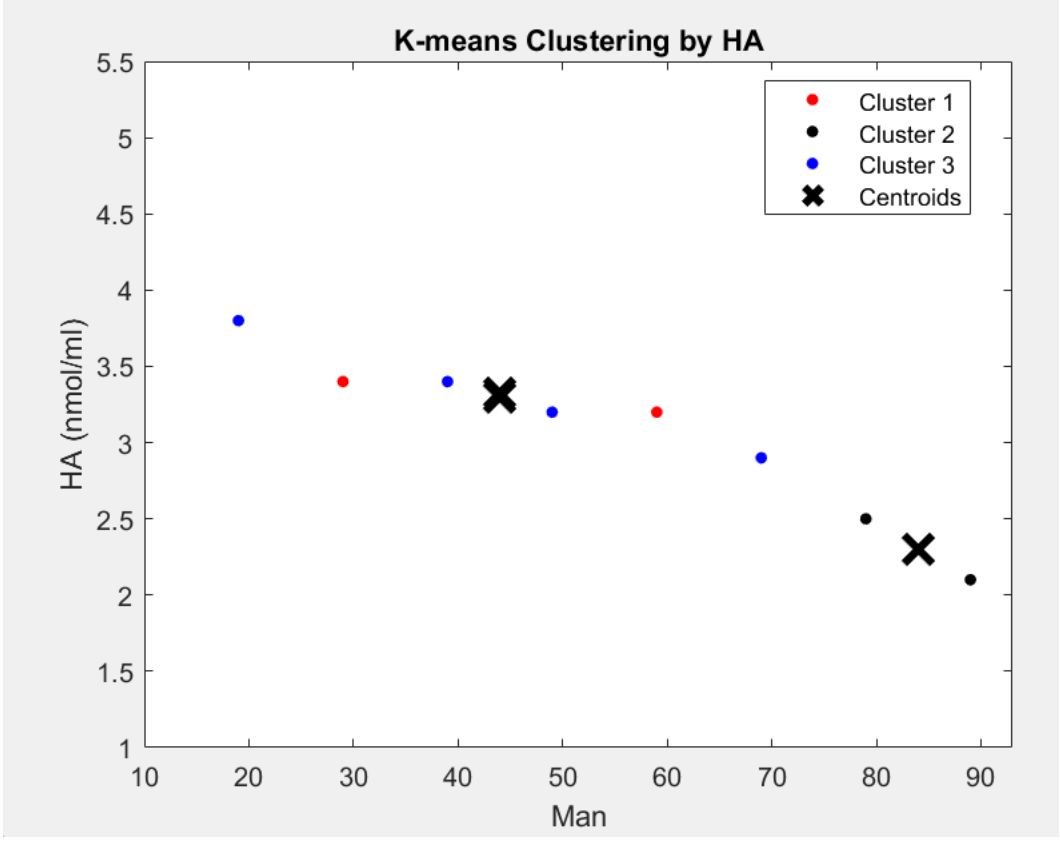


Figure 4.10. Result of k-means cluster for men with HA.

The k-means clustering analysis for men was conducted based on age and the amounts of hyaluronic acid (HA) and chondroitin sulfate (C6S, C4S, and the ratio of C6S to C4S) in synovial fluid. The results indicate men aged 10 to 70 years have relatively

higher concentrations of HA and CS in synovial fluid. However, the men aged 50 to 90 years have the lowest relative concentrations of HA and CS in synovial fluid. The men aged between 30 and 60 years have moderate biomarker levels.

## 4.4 Results of K-Means Cluster for Women in Healthy Group with HA in Synovial Fluid

### 4.4.1 Results for Women with C6S

The scatter plot in Figure 4.11. shows the outcomes of k-means clustering for the healthy group. It is based on two variables: the number of women (x-axis) and the concentration of C6S (nmol/ml) (y-axis). The women's age range is 10 to 90 years, with C6S concentrations (nmol/ml) ranging from 0 to 300. There are three clusters, red, black, and blue points, with centroids marked by black Xs. Each color represents a different cluster for the k-means algorithm. The red cluster comprises women aged between 10 and 40 years with C6S levels ranging from 120 to 150 nmol/ml. The blue cluster includes older women (approximately 80 to 90 years old) with the lowest C6S levels between 60 and 70 nmol/ml. The black cluster comprises women aged between 50 to 70 who had moderate C6S levels ranging from 80 to 120 nmol/ml. The centroid of the red cluster is pointed around age 30 with a C6S of around 120 nmol/ml, but the centroid of the blue cluster is pointed around age 85 with a C6S of around 70 nmol/ml. The centroid of the black cluster is pointed around age 60 with a C6S of around 100 nmol/ml.

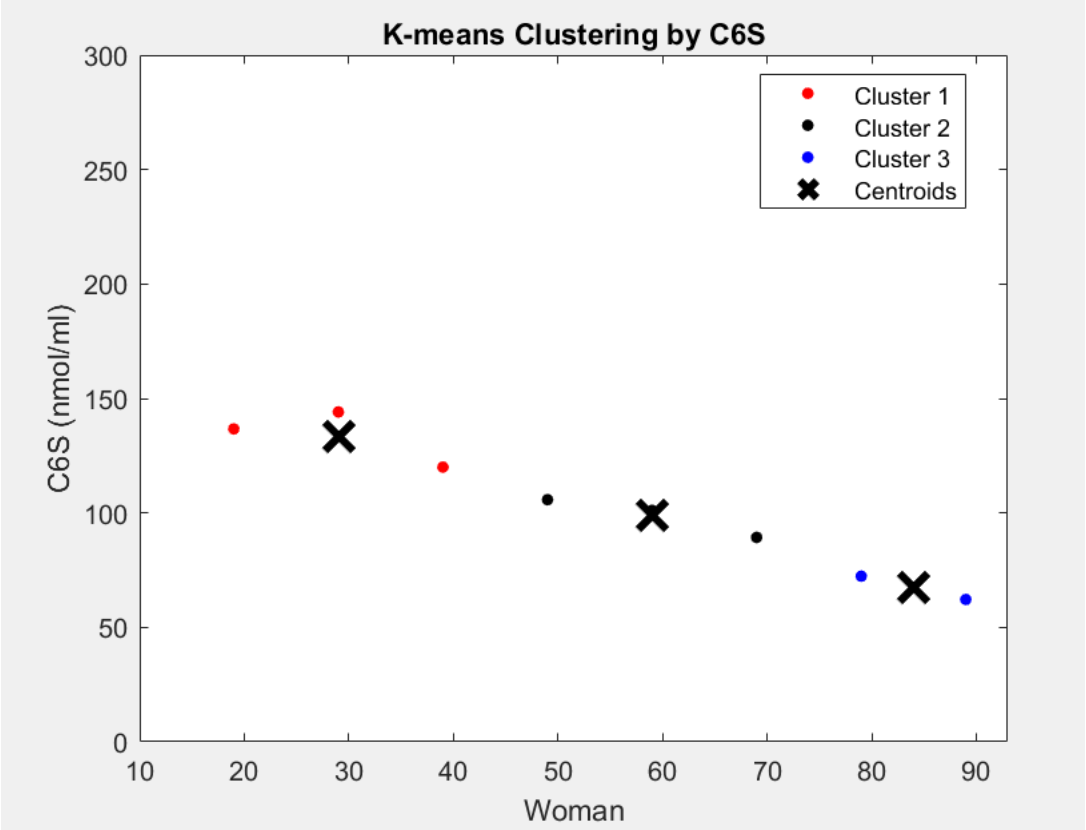


Figure 4.11. Result of k-means cluster for women with C6S.

4.4.2 Results for Women with C4S

The healthy group's k-means clustering results are shown in Figure 4.12. as a scatter plot. The x-axis shows the age of the women, and the y-axis comprises the C4S concentration (nmol/ml). The age range is 10 to 90 years, with C4S concentration (nmol/ml) ranging from 5 to 45. There are three clusters: red, black, and blue. Each color represents a different cluster for the k-means algorithm. The red cluster is for women aged roughly between 10 and 40 years with higher C4S levels ranging from 18 to 22 nmol/ml. The centroid of the red cluster is pointed around age 30 with a C4S of around 22 nmol/ml. The blue cluster includes older women aged between 80 to 90 years with the lowest C4S levels between 10 and 15 nmol/ml. The centroid of the blue cluster is pointed

around age 85 with a C4S of around 15 nmol/ml. The black cluster comprises women aged between 50 and 70 years who had moderate C4S levels ranging from 18 to 19 nmol/ml. The centroid of the black cluster is pointed around age 60 with a C4S of around 19 nmol/ml.

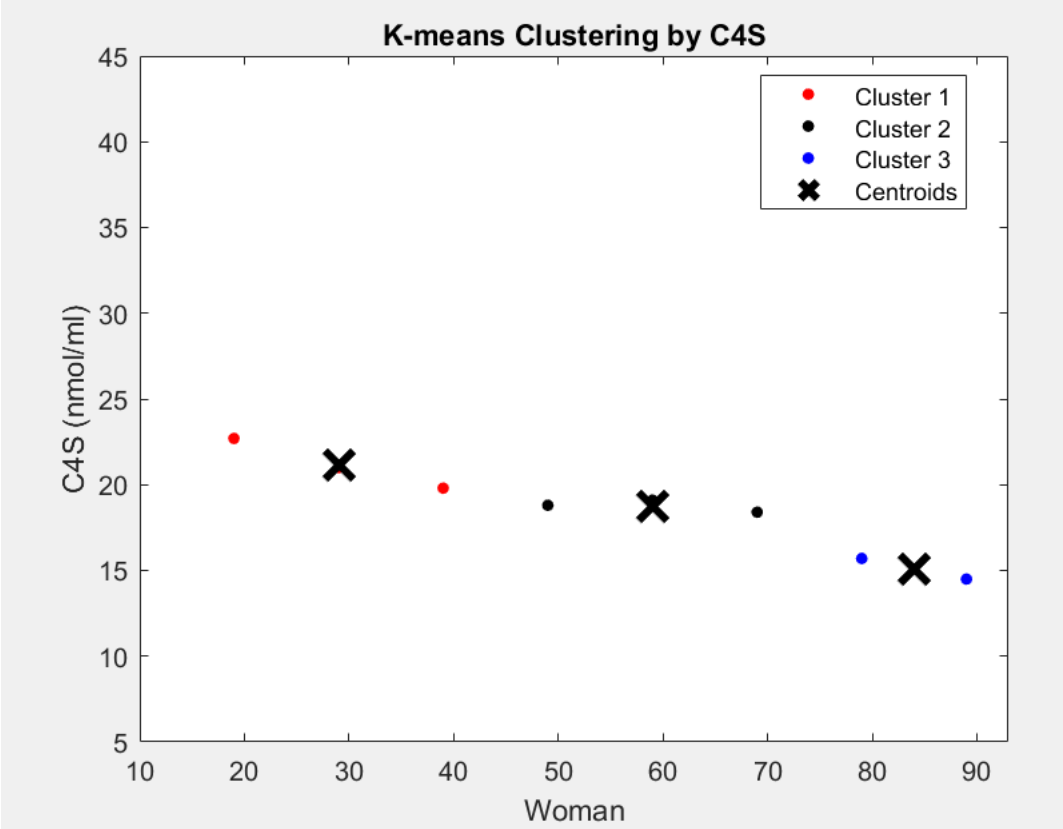


Figure 4.12. Result of k-means cluster for women with C4S.

4.4.3 Results for Women with C6S:C4S

Figure 4.13. show presents the results of k-means clustering on data with two features: women's age on the x-axis and C6S:C4S concentration (nmol/ml) on the y-axis. The age range was 10 to 90 years, with C6S:C4S concentration (nmol/ml) ranging from 2 to 10. The first cluster is red and contains women aged roughly between 10 and 40 years

who had higher C4S levels ranging from 6 to 7nmol/ml. The second black cluster comprises women aged between 50 and 70 years, with moderate C6S:C4S levels ranging from 4.5 to 6 nmol/ml. The third cluster is blue, including older women aged approximately 80 to 90 years; with the lowest C6S:C4S levels between 4 and 5 nmol/ml. The large 'X' markers represent the clusters' centroids, which are the average of all points within each cluster. The centroid of the red cluster is pointed around age 30 with a C6S:C4S of around 6 nmol/ml, but the centroid of the blue cluster is pointed around age 85 with a C6S:C4S of around 4.5 nmol/ml. The centroid of the black cluster is pointed around age 60 with a C6S:C4S of around 5.5 nmol/ml.

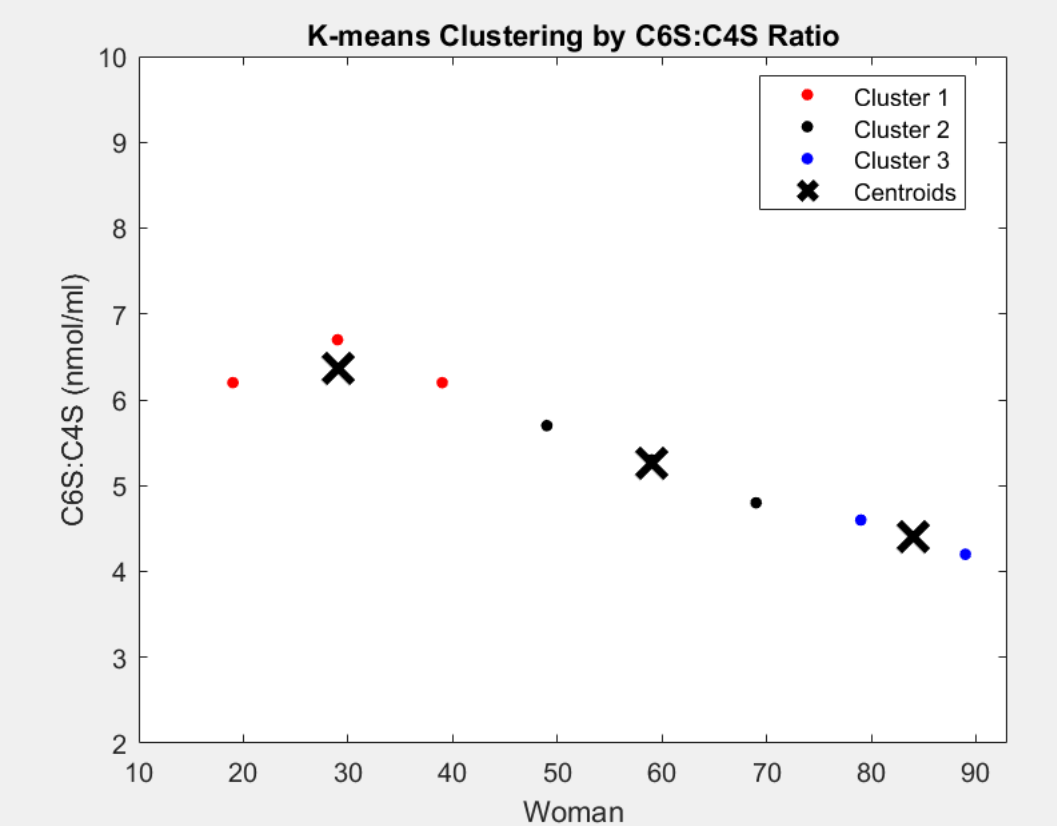


Figure 4.13. Result of k-means cluster for women with C6S:C4S.

4.4.4 Results for Women with HA

The scatter plot in Figure 4.14. shows the k-means clustering results for two variables: women’s age on the x-axis and HA concentration of 1 to 5.5 (nmol/ml) on the y-axis. The first cluster is red and contains data of women aged roughly between 10 and 40 years with HA levels ranging from 3 to 4 nmol/ml. The second is a black cluster comprising women aged 50 to 70 years who had HA levels ranging from 2.5 to 3.5 nmol/ml. The third cluster is blue, including older women aged approximately 80 to 90 years with the lowest HA levels between 2 and 2.5 nmol/ml. The large 'X' markers represent the clusters' centroids, which are the average of all points within each cluster. The centroid for the red cluster is at age 30 and has a ratio of about 3.4, but the centroid for the blue cluster is around age 85 with a concentration of approximately 2.2 nmol/ml. The centroid for the black cluster is at age 60 and has a ratio of about 3.2.

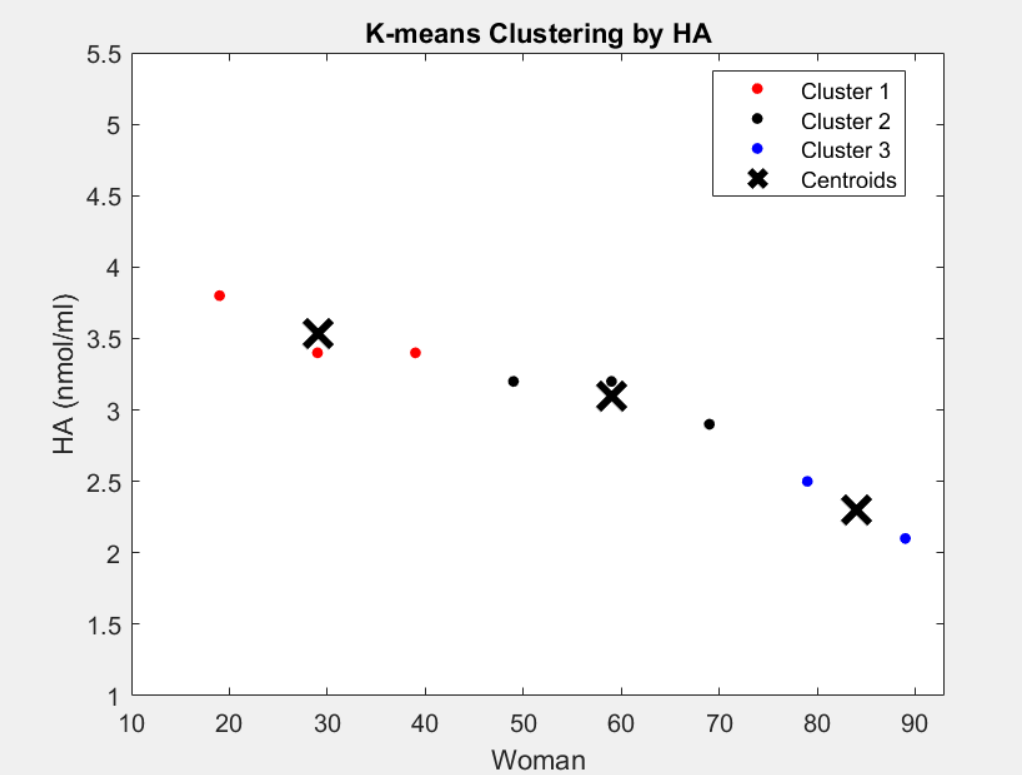


Figure 4.14 Result of k-means cluster for women with HA.

The k-means clustering analysis for women, based on their age and the amounts of HA and CS (C6S, C4S, and the ratio of C6S to C4S), indicates three different groups. The results for the younger women (aged 10 to 40 years) shows higher levels of these components, whereas older women (80 to 90 years old) have the lowest concentrations. Women in the middle age range (50 to 70 years) have moderate levels of C6S, C4S, and HA.

## 4.5 Results of K-Means Clustering for Osteoarthritis (OA) Group

One method to interpret the results of the k-means clustering algorithm for Uesaka et al.'s [19]data is using to visualize the clustering results. The first result is the heatmap for the OA group, as shown in Figure 18. This shows the correlation matrix for

OA group data with multiple measurements. This heatmap explains in great detail the relationships between synovial fluid compositions (C6S, C4S, C6S: C4S, and HA) for different age groups and gender. Each cell determines the correlation coefficient between two variables on the axis, which ranges from -1 to 1. A coefficient close to -1 shows a negative correlation between variables, which means one variable increases as the other decreases. A coefficient close to 1 indicates a strong positive correlation, meaning one variable increases as the other increases. Where the coefficient is zero, there is no correlation between the variables. In addition, the color yellow indicates a strong positive correlation between variables, while blue indicates negative coloration between variables.

Figure 4.15. shows the results for variable age with C6S:C4S. There was a very weak negative correlation between these variables with coefficients around -0.1 that are revealed by the dark blue color. Also, the correlation between age and C6S coefficients was around -0.95, a very weak negative correlation, while the correlation between age and C4S coefficients was around -0.79, which was a weak negative correlation. These results indicate there is no significant relationship between these variables for the OA group. However, the relationship between age and HA had a strong positive correlation coefficient of around 0.73, which was a significant relationship between these two variables. Furthermore, women had a negative correlation with HA and C4S. The coefficients at -0.48 in HA and -0.24, indicate an almost negative relationship between these variables in the OA group. This means there is no significant relationship between these variables. Women had a week positive correlation with C6S:C4S coefficients at 0.33, and for women with C6S, the coefficients were at 0.09. This implies a significant

relationship between these variables. However, a weak positive correlation was revealed between men and C6S, with a coefficient of 0.23.

Men with C6S:C4S had a coefficient of 0.45, a positive correlation. Additionally, the correlation coefficients for men with HA, and men with C4S, were -0.31 and -0.08, indicating a negative correlation. However, a weak positive correlation was revealed between men and C6S, with a coefficient of 0.23. Men and C6S:C4S with a coefficient of 0.45 had a positive correlation. Additionally, the correlation coefficients for men with HA, and men with C4S, were -0.31 and -0.08, indicating a negative correlation. The results for the men are similar to those of the women, conveying a significant relationship for men with C6S and C6S:C4S, while is no significant relationship for men with HA and C4S in OA group. The coefficient of age with synovial fluid volume in the OA group was -0.68, implying a weak negative correlation. The coefficient of men with synovial fluid volume was 0.97, which means a very strong positive correlation. Furthermore, a strong positive correlation was revealed between women with volume, at a coefficient of 0.85.

These results indicate that as people age, the levels of synovial fluid components tend to decrease significantly. In addition to age, gender appears to be another important factor influencing synovial fluid composition, as the results show women and men had negative correlations with the concentrations of C4S and HA, and positive correlations with the concentrations of C6S and C6S:C4S in the OA group.

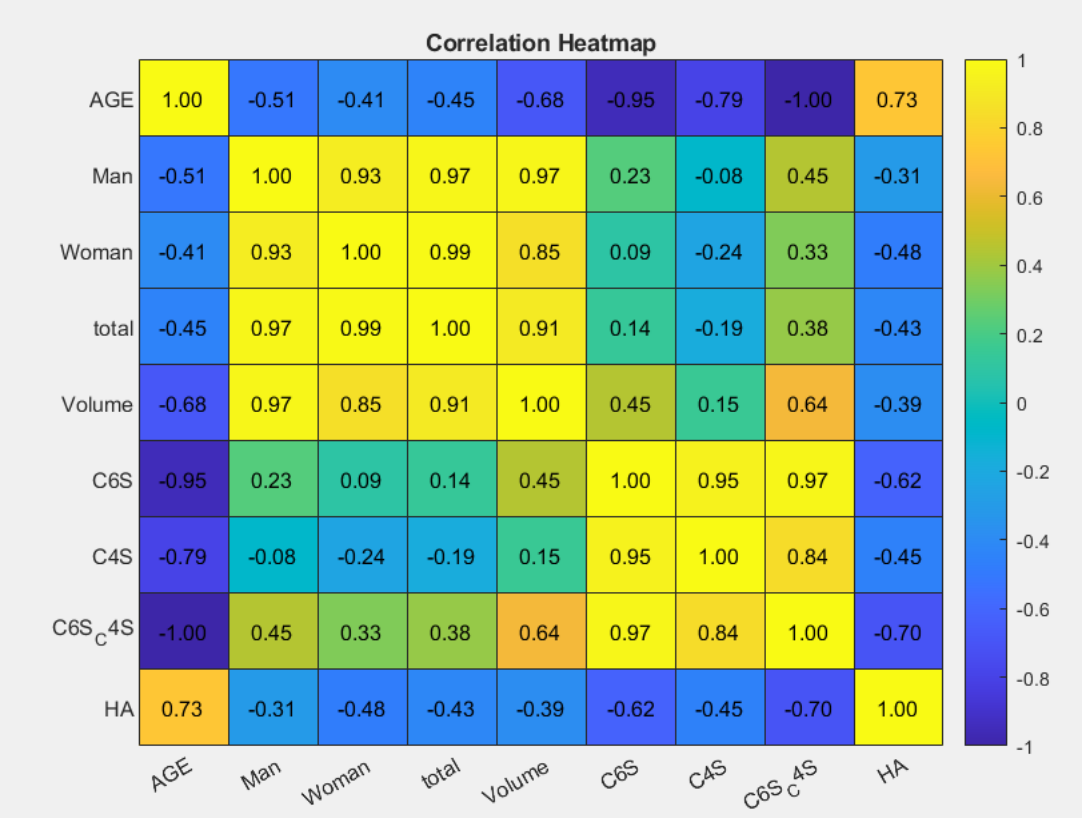


Figure 4.15. heatmap for Osteoarthritis (OA) group.

The second way to interpret the results of k-means is through scatter plots. To start the k-means procedure, it is important to specify the number of k, which represents the number of clusters in the data. The graph in Figure 4.16. shows the elbow method for determining the optimal number of k clusters by examining the Within-Cluster Sum of Squares (WCSS). The top plot represents the OA data before scaling, and the bottom plot represents the data after scaling. In both plots, the WCSS starts with a high of 1 cluster, then drops clearly as the number of clusters increases. In both plots, the sharp decline in WCSS from 1 to 3 clusters, followed by a leveling off, confirms K= 3 is the optimal number.

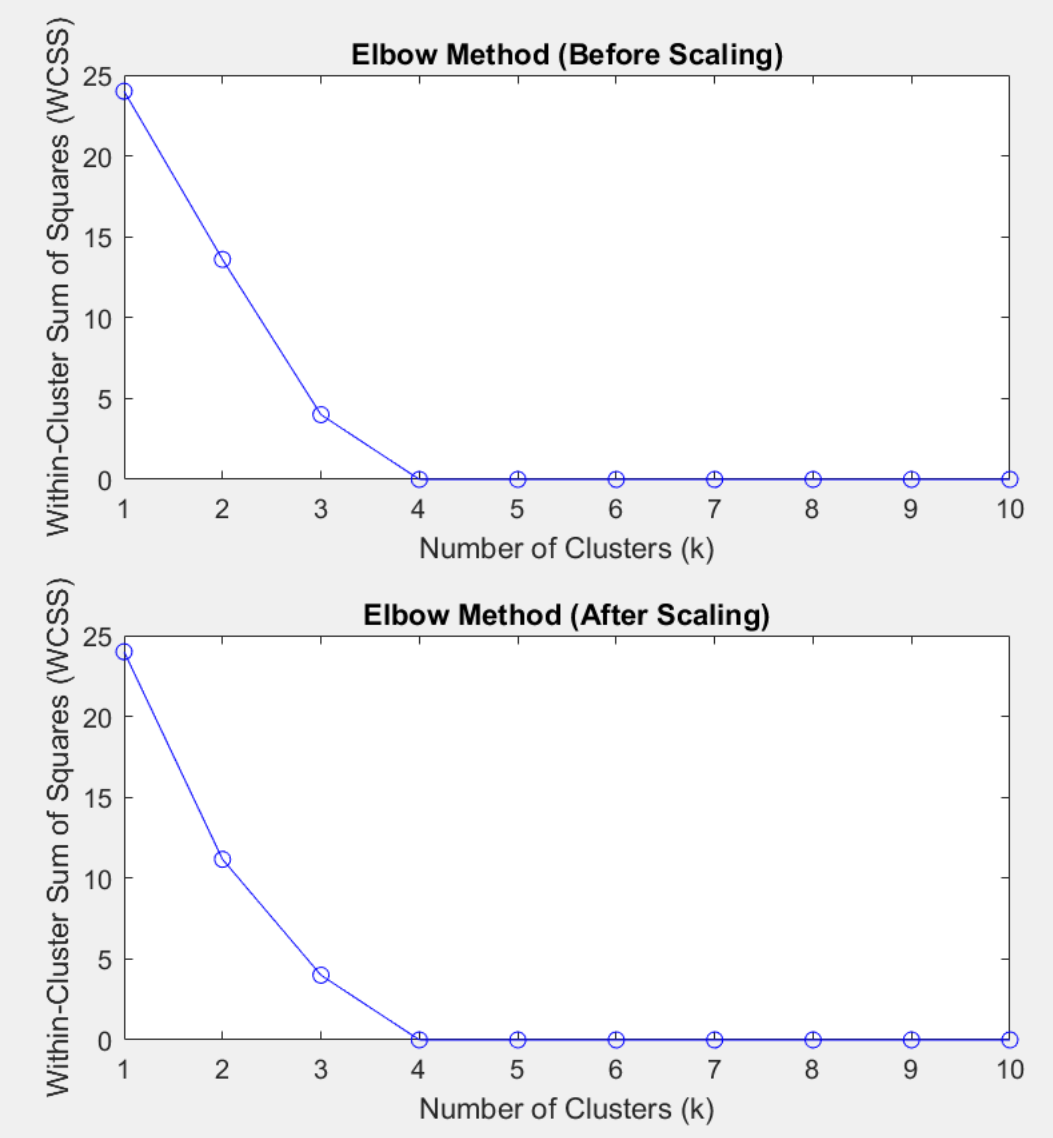


Figure 4.16 Elbow method for OA group.

## 4.6 Results of K-Means Cluster for Variables Age and CS, HA in Synovial Fluid in Osteoarthritis (OA) Group

### 4.6.1 Results for Age and C6S

The scatter plot in Figure 4.17. represents the results of k-means clustering for the OA group based on two variables: age on the x-axis and C6S concentration (nmol/ml) on the y-axis. The age range is 55 to 90 years, with C6S concentration (nmol/ml) ranging from 0 to 300. The red, black, and blue points indicate the identification of three distinct clusters with centroids marked by black Xs. Each color represents a different cluster for the k-means algorithm. The red cluster comprises older people (aged roughly between 80 and 90 years) who had higher C6S levels ranging from 50 to 70 nmol/ml. The black cluster includes middle-aged people (67 years old), with C6S levels of 70 nmol/ml. The blue cluster comprises younger people who are between 55 and 60 years old, with higher C6S levels ranging from 100 to 120 nmol/ml. The large 'X' markers represent the clusters' centroids, which are the average of all points within each cluster. These centroids are the points that minimize the total distance from all points within the cluster.

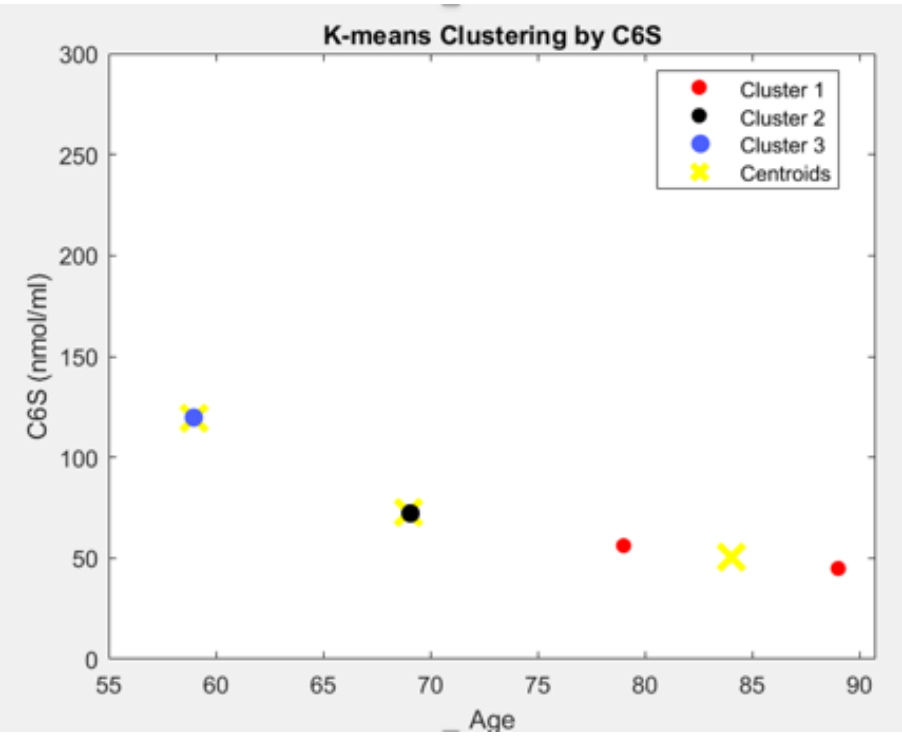


Figure 4.17**.** Result of k-means cluster for age with C6S.

4.6.2 Results for Age with C4S

The scatter plot in Figure 4.18. shows the results of k-means clustering for the healthy group based on two variables: age on the x-axis and C4S concentration (nmol/ml) on the y-axis. The age range is 10 to 90 years, with C4S concentration (nmol/ml) ranging from 5 to 45. Each color represents a different cluster for the k-means algorithm. The red cluster has older people (aged roughly between 80 and 90 years) who had higher C4S levels ranging from 15 to 25 nmol/ml. The black cluster includes middle-aged people (aged approximately 69 years old) with C4S levels of 17 nmol/ml. The blue cluster comprises younger people (between 55 and 60 years old) with higher C4S levels ranging from 23 nmol/ml. The large 'X' markers represent the clusters' centroids on the same value of clusters in blue and black clusters but the centroids for the red cluster at age 85.

These centroids are the points that minimize the total distance from all points within the cluster.

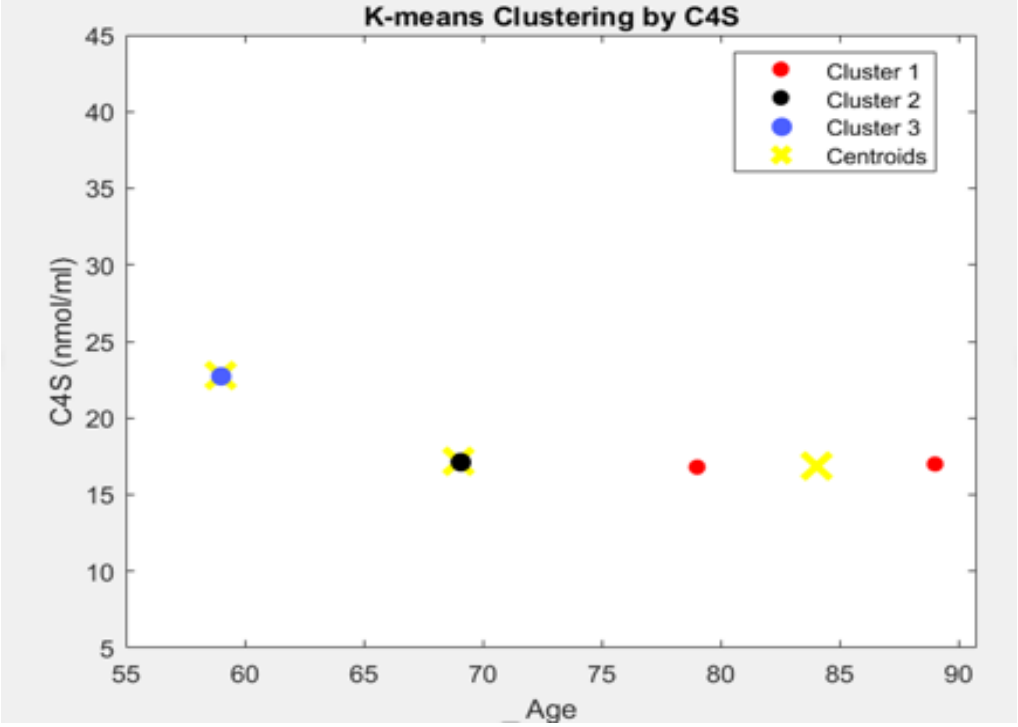


Figure 4.18. Result of k-means cluster for age with C4S.

4.6.3 Results for Age with C6S:C4S

Figure 4.19. shows the results of k-means clustering for data with two features: age on the x-axis and C6S:C4S concentration (nmol/ml) on the y-axis. The age range is 55 to 90 years, with C6S:C4S concentration (nmol/ml) ranging from 2 to 10. The first cluster is red and contains older people (aged roughly between 80 and 90 years) who had the lowest C6S:C4S levels ranging from 2 to 2.9 nmol/ml. The second cluster is black, including middle-aged people aged (approximately 69 years old) with the C6S:C4S level at 4 nmol/ml. The blue cluster comprises younger aged people (between 55 and 60 years old) with the highest C6S:C4S levels ranging from 5 to 5.5 nmol/ml. The large 'X'

markers represent the clusters' centroids, which are the average of all points within each cluster. These centroids are the points that minimize the total distance from all points within the cluster.

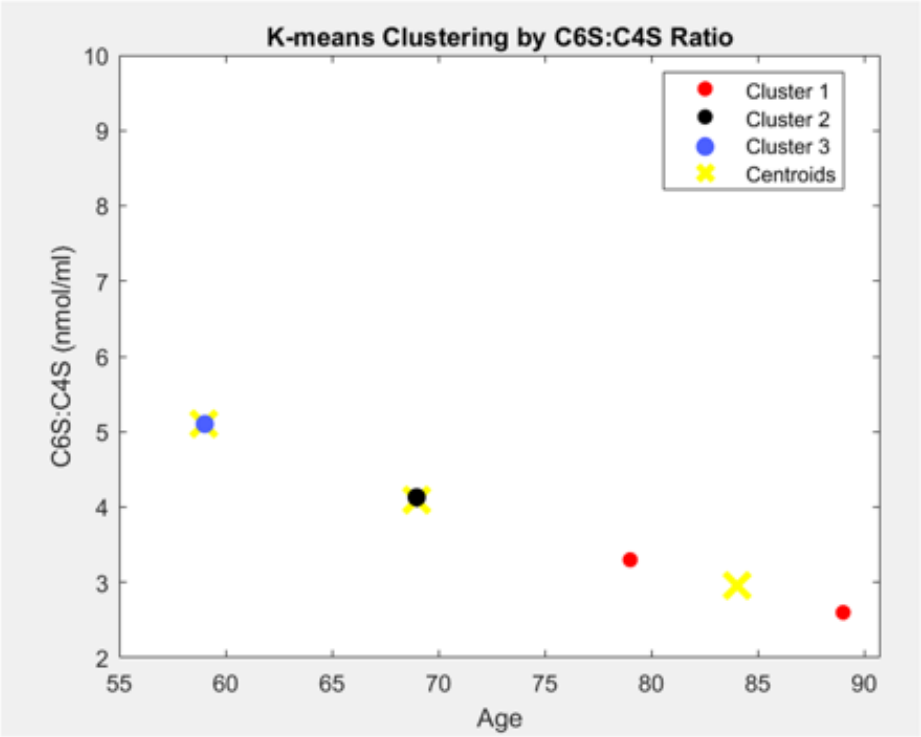


Figure 4.19. Result of k-means cluster for age with C6S:C4S.

4.6.4 Results for Age with HA

Figure 4.20. shows the k-means clustering results for two features: age on the x-axis and HA concentration 1 to 5 (nmol/ml) on the y-axis. The first cluster is red and encompasses older people (aged roughly between 78 and 90 years) with higher HA levels ranging from 1.8 to 2.5 nmol/ml. The second cluster is black, including middle-aged people (approximately 69 years old); with HA levels at 1.9 nmol/ml. The blue cluster comprises younger aged people who were between 55 and 60 years old with moderate

HA levels ranging from 1.9 nmol/ml. The large 'X' markers represent the clusters' centroids, which are the average of all points within each cluster. These centroids are the points that minimize the total distance from all points within the cluster.

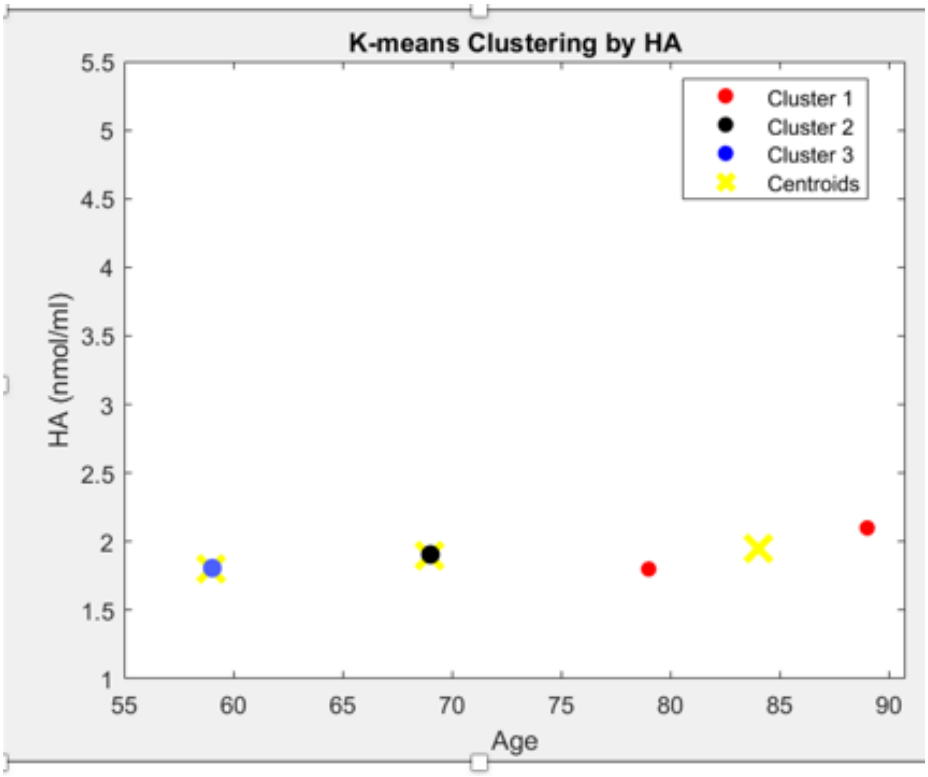


Figure 4.20. Result of k-means cluster for age with HA.

The results of k-means clustering analysis for the OA group, based on age and comparing concentrations of C6S, C4S, C6S:C4S ratio, and HA in synovial fluid, indicates younger people aged 55 to 60 years have the highest levels of CS components, except with HA. The results for the older age group (80 to 90 years old) indicate they have the lowest levels of CS components, except HA. For people in the middle-age group, CS components and HA show middle levels for the OA group.

## 4.7 Results of K-Means Clustering for Men in OA Group with CS and HA in Synovial Fluid

### 4.7.1 Results for Men with C6S

The scatter plot in Figure 4.21. illustrates the results of k-means clustering for the OA group based on two variables: men on the x-axis and C6S concentration (nmol/ml) on the y-axis. The men's age range is 55 to 90 years, with C6S concentration (nmol/ml) ranging from 0 to 300. There are three clusters, red, black, and blue points, with centroids marked by black Xs. Each color represents a different cluster for the k-means algorithm. The red cluster contains men aged between 65 and 70 years, with C6S levels ranging from 56 to 70 nmol/ml. The black cluster includes older men (approximately 80 to 90 years old) with the lowest C6S levels between 45 and 50 nmol/ml. The blue cluster comprises men aged between 55 to 60 years, who had the highest C6S levels ranging from 100 to 120 nmol/ml. The large 'X' markers represent the clusters' centroids on the same value of clusters in blue and black clusters but the centroids for the red cluster at age 85. These centroids are the points that minimize the total distance from all points within the cluster.

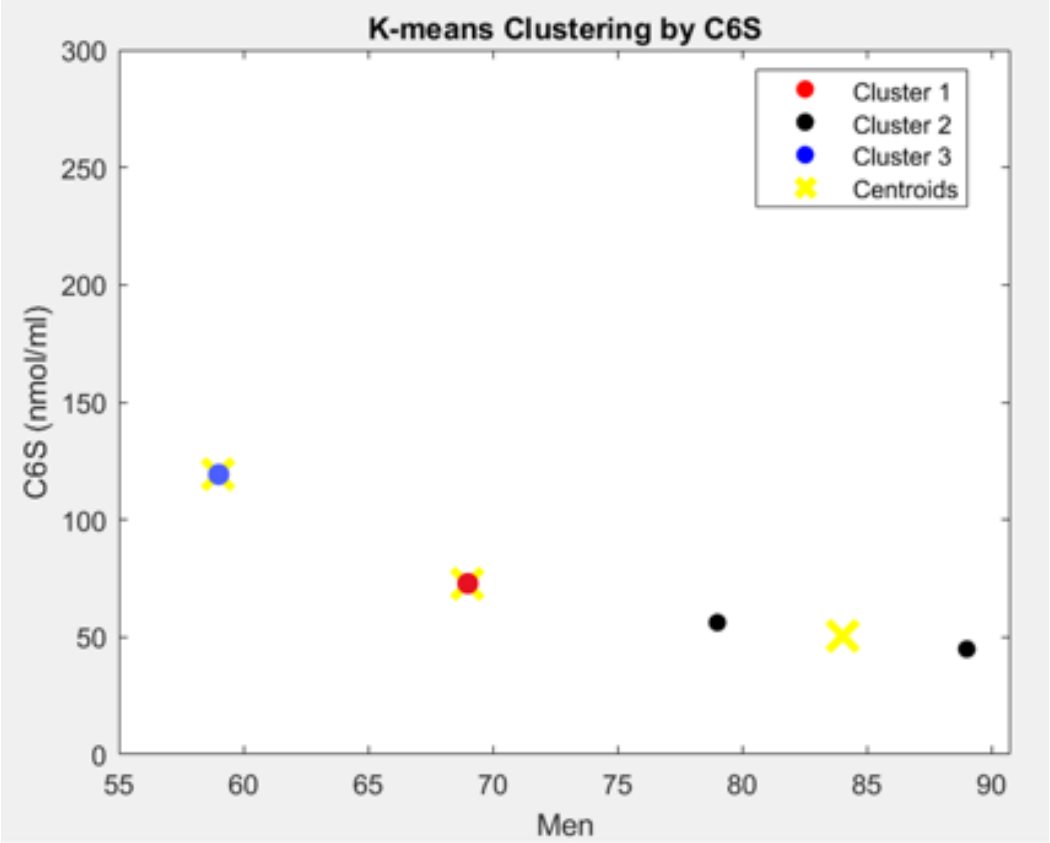


Figure 4.21. Result of k-means cluster for men with C6S.

4.7.2 Results for Men with C4S

The scatter plot in Figure 4.22. represents the results of k-means clustering for the OA group based on two variables: men age on the x-axis and C4S concentration (nmol/ml) on the y-axis. The men age range is 55 to 90 years, with C4S concentrations (nmol/ml) ranging from 5 to 45. Each color represents a different cluster for the k-means algorithm. The black cluster contains older men (aged roughly between 80 and 90 years) who had C4S levels ranging from 16 to 17 nmol/ml. The red cluster includes middle-aged men (approximately 69 years old) with C4S levels of 17 nmol/ml. The blue cluster comprises younger men who were between 55 and 60 years old and had higher C4S

levels ranging from 23 nmol/ml. The large 'X' markers represent the clusters' centroids on the same value of clusters in blue and black clusters but the centroids for the red cluster at age 85. These centroids are the points that minimize the total distance from all points within the cluster.

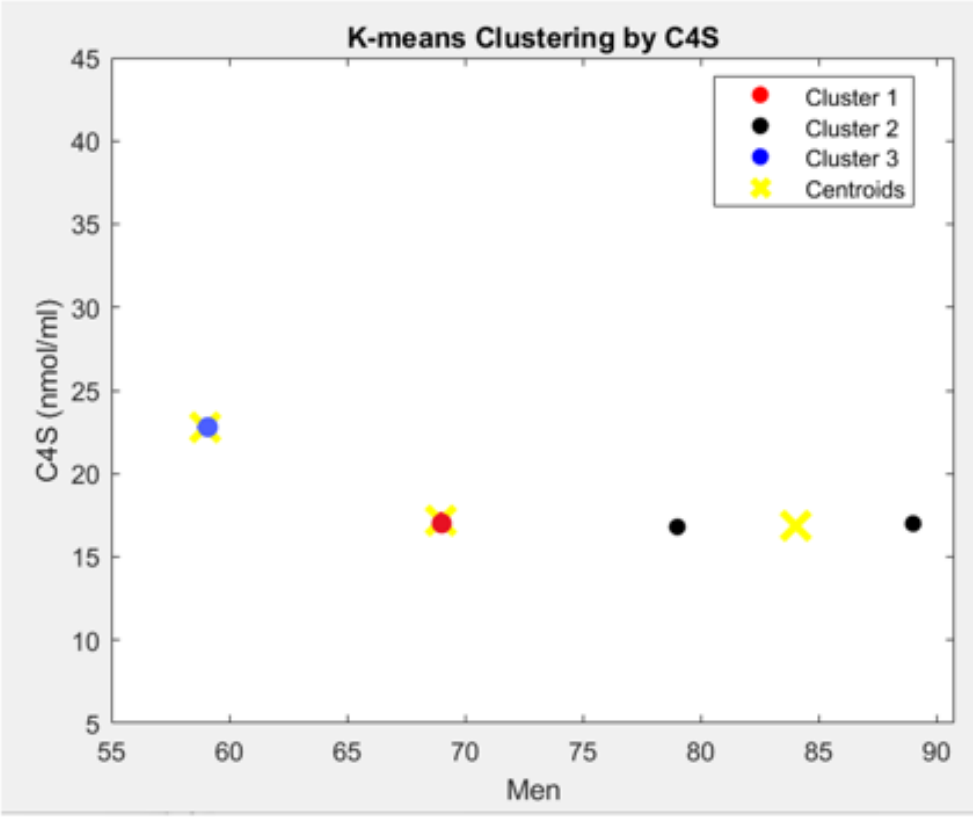


Figure 4.22. Result of k-means cluster for men with C4S.

4.7.3 Results for Men with C6S:C4S

The scatter plot in Figure 4.23. shows the results of k-means clustering on data with two features: men on the x-axis and C6S: C4S concentration (nmol/ml) on the y-axis. The men age range is 55 to 90 years, with C6S: C4S concentrations (nmol/ml)

ranging from 2 to 10. The first cluster is red and contains middle-aged men (68 years) who had middle C6S:C4S levels ranging from 4 to 4.5 nmol/ml. The second cluster is black, including older men (approximately 80 to 90 years old); with the lowest C6S: C4S level 2 to 3 nmol/ml. The blue cluster comprises younger men aged between 55 and 60 years who had the highest C6S: C4S levels ranging from 5 to 5.5 nmol/ml. The large 'X' markers represent the clusters' centroids, which are the average of all points within each cluster. These centroids are the points that minimize the total distance from all points within the cluster.

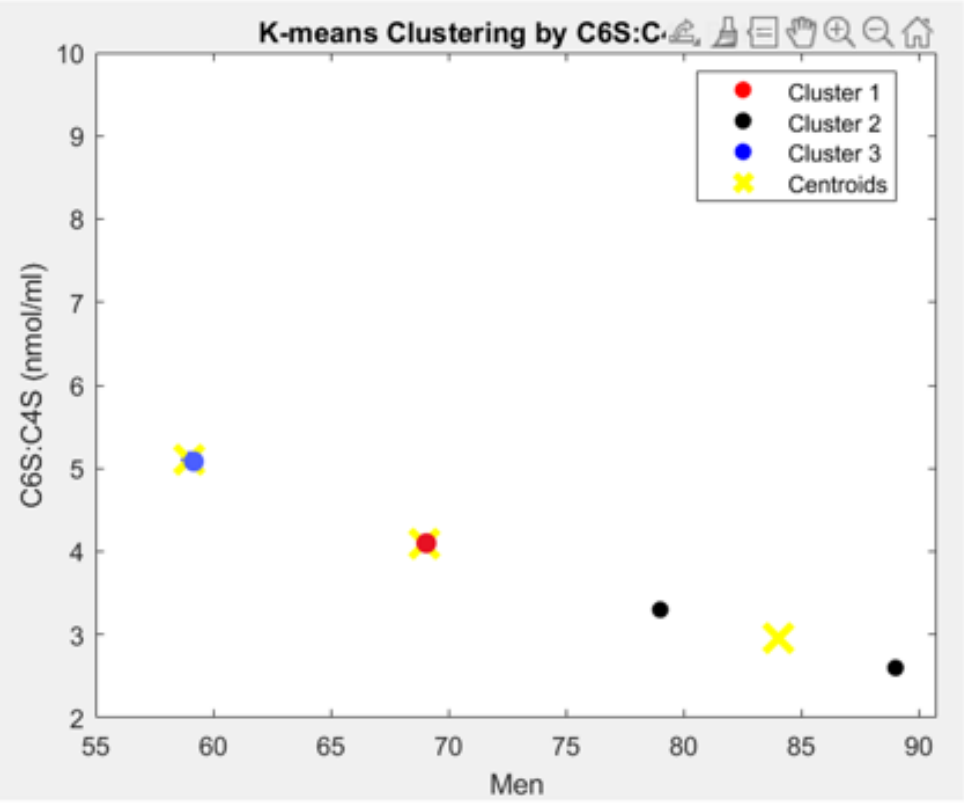


Figure 4.23. Result of k-means cluster for men with C6S:C4S.

4.7.4 Results for Men with HA

Figure 4.24. shows the k-means clustering results for two features: men on the x-axis and HA concentration 1 to 5 (nmol/ml) on the y-axis. The first cluster is red, including middle-aged (approximately 69 years old) with HA levels of 1.9 nmol/ml. The second cluster is black and contains men in the oldest age range (roughly 80 to 90 years) who had higher HA levels ranging from 1.8 to 2.5 nmol/ml. The third is the blue cluster, comprising younger men (aged between 55 to 60 years) who had moderate HA levels ranging from 1.7 nmol/ml. The large 'X' markers represent the clusters' centroids, which are the average of all points within each cluster. These centroids are the points that minimize the total distance from all points within the cluster.

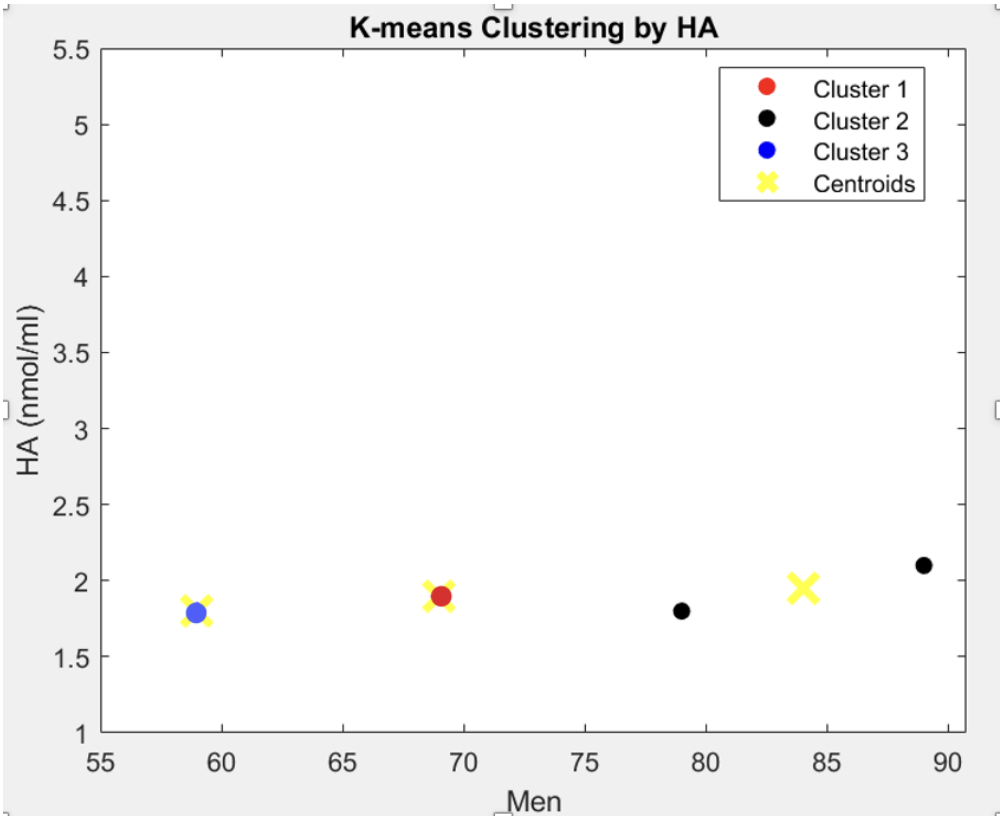


Figure 4.24. Result of k-means cluster for men with HA.

The results of the k-means clustering for men with osteoarthritis (OA), based on age and synovial fluid compositions (C6S, C4S, C6S:C4S ratio, and HA), indicates youngest men (aged between 55 to 60 years) have the highest levels of CS. while these levels decline in middle-aged men (65 to 70 years old) and reach their lowest in older men, (aged around 80 to 90 years). However, HA levels increase with age, with older men (around 90 years old) having the highest HA levels.

## 4.8 Results of K-Means Clustering for Women in OA Group with CS And HA in Synovial Fluid

### 4.8.1 Results for Women with C6S

The scatter plot in Figure 4.25. illustrates the results of k-means clustering for the OA group based on two variables: women on the x-axis and C6S concentration (nmol/ml) on the y-axis. The women's age range is 55 to 90 years, with C6S concentration (nmol/ml) ranging from 0 to 300. There are three clusters, red, black, and blue points, with centroids marked by black Xs. Each color represents a different cluster for the k-means algorithm. The red cluster contains women aged between 65 and 80 years, with C6S levels ranging from 56 to 70 nmol/ml. The black cluster includes older women (approximately 89 years old) with the lowest C6S levels at between 45 and 50 nmol/ml. The blue cluster comprises women aged between 55 to 60 years who had the highest C6S levels ranging from 100 to 120 nmol/ml. The large 'X' markers represent the clusters' centroids on the same value of clusters in blue and black clusters but the centroids for the red cluster are on middle. These centroids are the points that minimize the total distance from all points within the cluster.

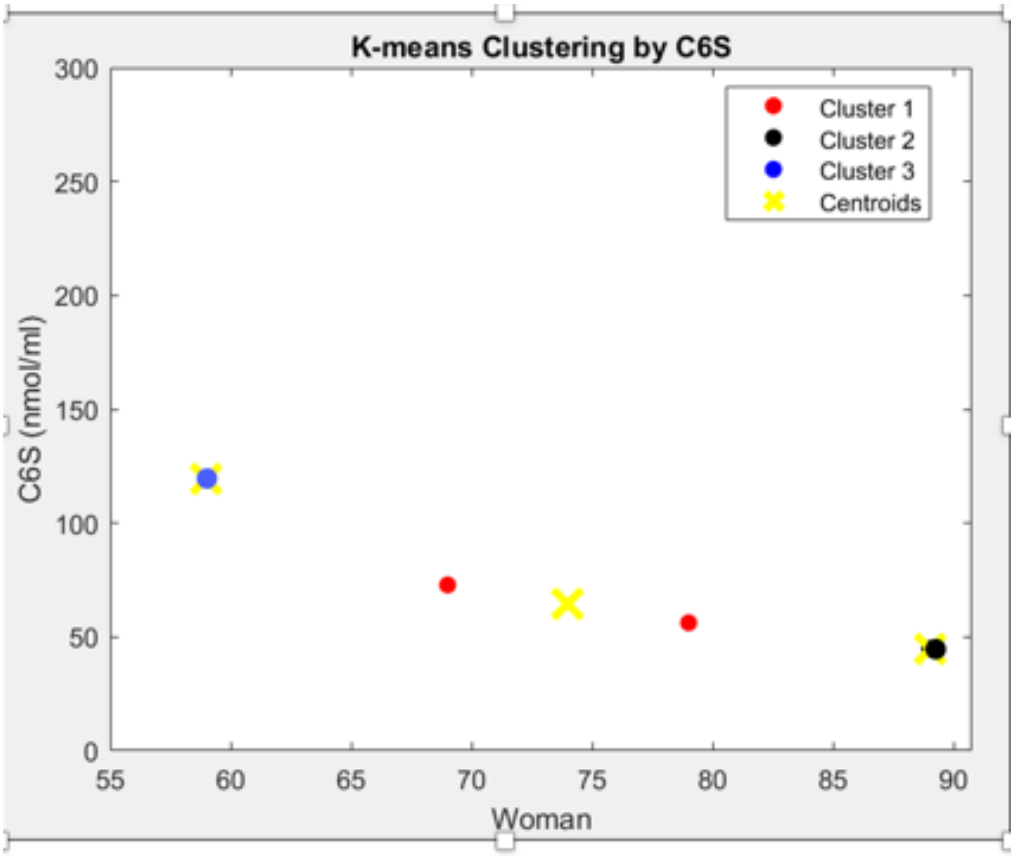


Figure 4.25. Result of k-means cluster for women with C6S.

4.8.2 Results for Women with C4S

The scatter plot in Figure 4.26. represents the results of k-means clustering for the OA group based on two variables: women's age on the x-axis and C4S concentration (nmol/ml) on the y-axis. The women's age range is 55 to 90 years, with C4S concentrations (nmol/ml) ranging from 5 to 45. Each color represents a different cluster for the k-means algorithm. The black cluster denotes older women (aged 90 years) who had C4S levels ranging from 16 to 17 nmol/ml. The red cluster includes middle-aged women (aged approximately 69 to 80 years) with C4S levels of 17 nmol/ml. The blue

cluster comprises younger women (between 55 and 60 years of age) with higher C4S levels ranging from 23 nmol/ml. The large 'X' markers represent the clusters' centroids on the same value of clusters in blue and black clusters but the centroids for the red cluster are in the middle. These centroids are the points that minimize the total distance from all points within the cluster.

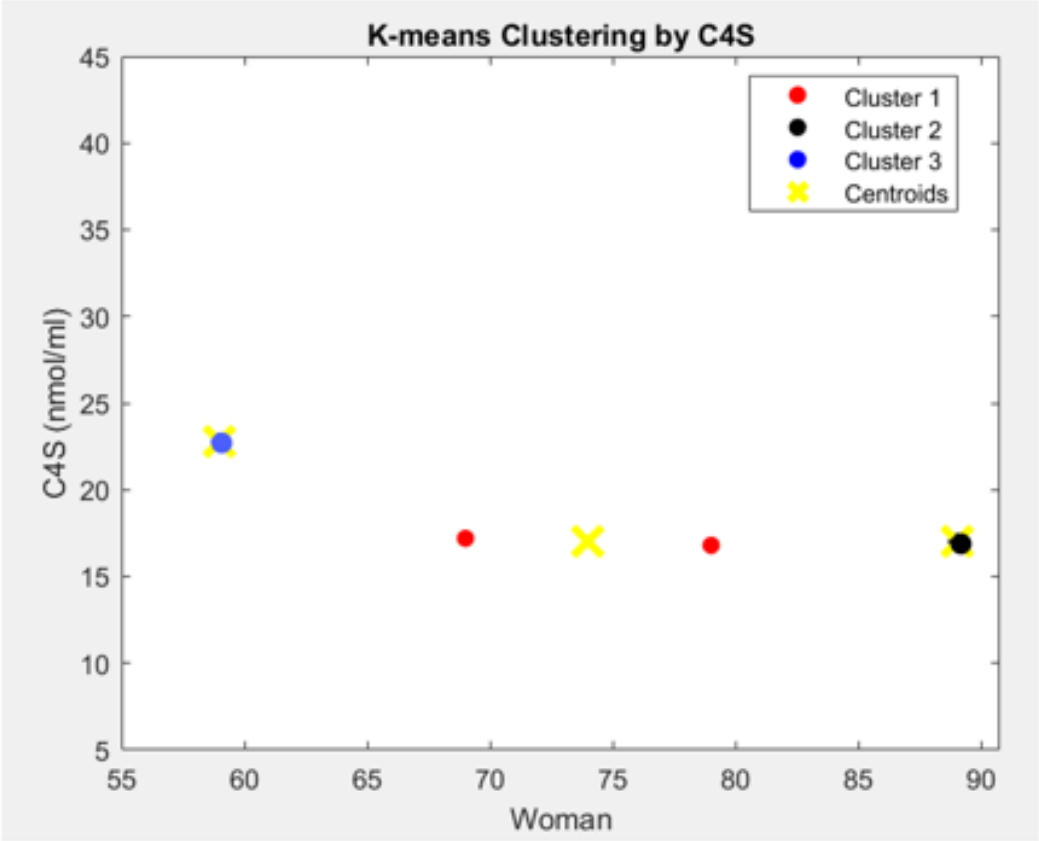


Figure 4.26. Result of k-means cluster for women with C4S

4.8.3 Results for Women with C6S:C4S

The scatter plot in Figure 4.27. shows the results of k-means clustering on data with two features: women on the x-axis and C6S:C4S concentration (nmol/ml) on the y-axis. The women's age range is 55 to 90 years, with C6S:C4S concentrations (nmol/ml) ranging from 2 to 10. The first cluster is red and contains middle-aged women (aged 68 to 80 years) who had middle C6S:C4S levels, ranging from 3 to 4.5 nmol/ml. The second cluster is black, including older women approximately 90 years old; with the lowest C6S:C4S level at 2.5 nmol/ml. The blue cluster comprises younger women aged between 55 and 60 years who had the highest C6S:C4S levels at 5 nmol/ml. The large 'X' markers represent the clusters' centroids, which are the average of all points within each cluster. These centroids are the points that minimize the total distance from all points within the cluster.

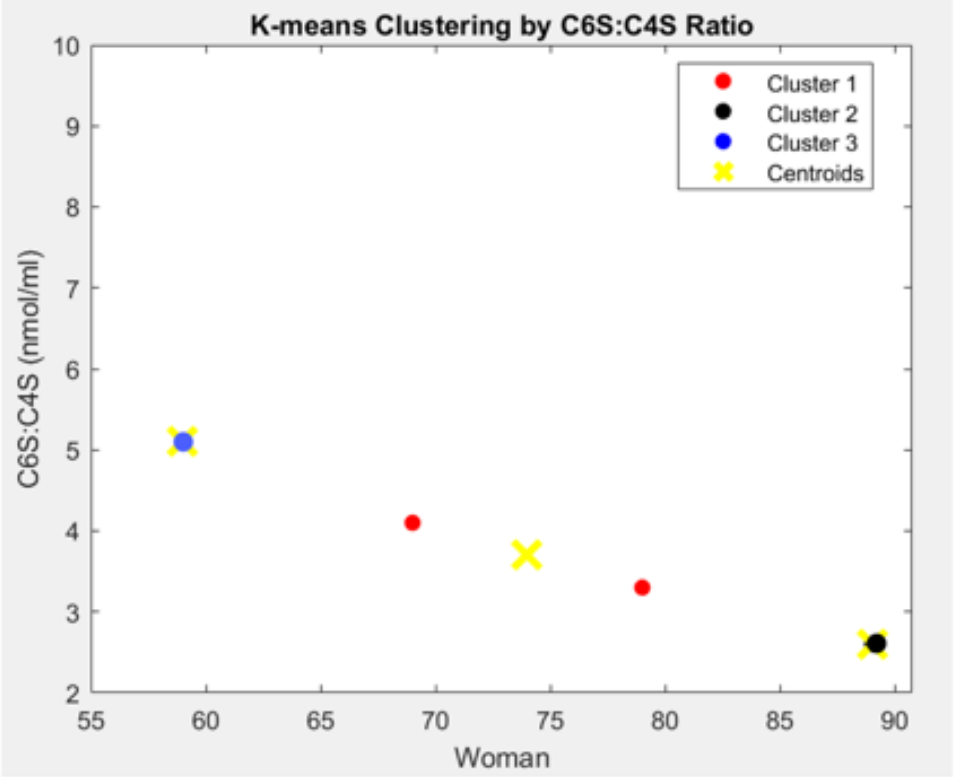


Figure 4.27. Result of k-means cluster for women with C6S:C4S

4.8.4 Results for Women with HA

Figure 4.28. shows the k-means clustering results for two variables: women on the x-axis and HA concentration 1 to 5.5 (nmol/ml) on the y-axis. The first cluster is red, including middle-aged women aged approximately 69 to 80 years; with HA levels of 1.7 to 1.9 nmol/ml. The second cluster is black and contains women in the oldest age (90 years) who had higher HA levels ranging from 1.8 to 2.5 nmol/ml. The third is a blue cluster comprising younger women (aged between 55 and 60 years) who had moderate HA levels ranging from 1.7 nmol/ml. The large 'X' markers represent the clusters'

centroids, which are the average of all points within each cluster. These centroids are the points that minimize the total distance from all points within the cluster.

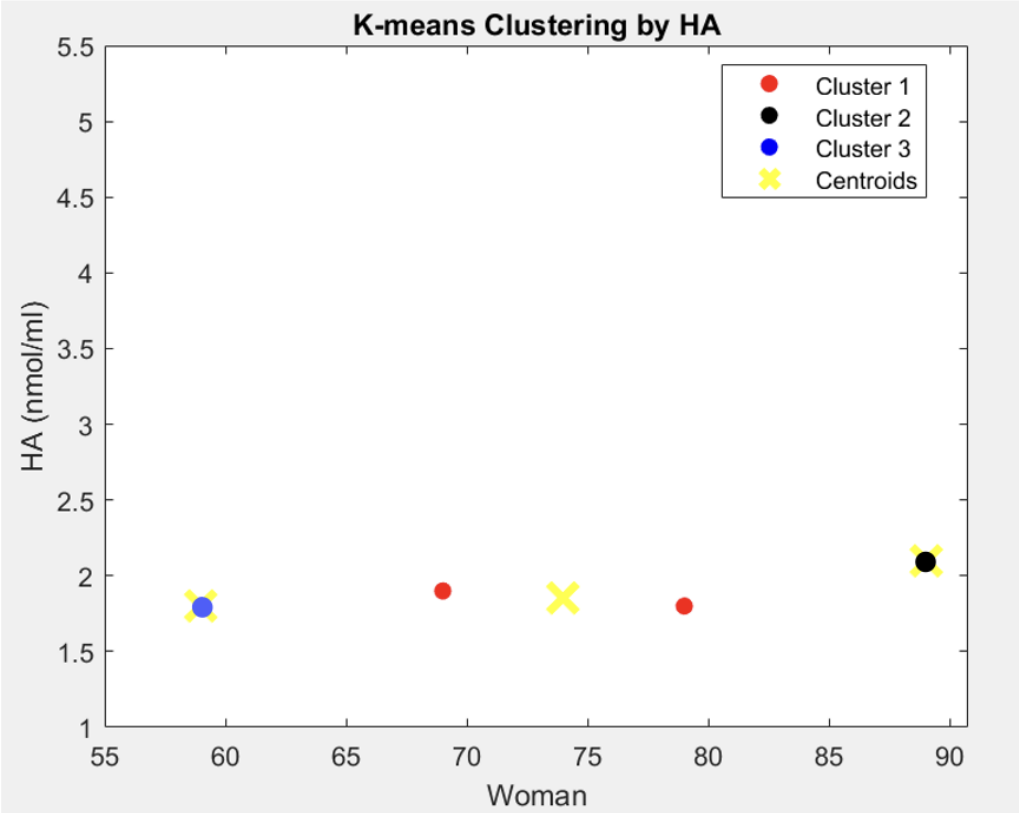


Figure 4.28. Result of k-means cluster for women with HA

The results of the k-means clustering for women with osteoarthritis (OA) based on age and synovial fluid compositions (C6S, C4S, C6S:C4S ratio, and HA) is similar to that of the men; this indicates youngest women (aged 55 to 60 years) have the highest levels of CS. These levels decline for middle-aged women (in the age group of 65 to 70 years) and reach their lowest in older women (who are around 80 to 90 years old). However, the HA levels increase with age, with older women (aged around 90 years) having the highest HA levels.

## 4.9 Comparison of K-Means Clustering Results for Healthy Group and OA Group

1- In the healthy group, the younger people consistently showed higher levels of C6S, C4S, and HA, but as age increased, these concentrations decreased, with older people revealing the lowest levels.

2- In the OA group, the younger people also had the highest levels of C6S, C4S, and C6S:C4S while the oldest people showed the lowest levels, except for HA, which increased with age.

3- In the healthy group, those in middle-age category always had moderate levels of C6S, C4S, and HA.

4- In the OA group, middle-aged people had moderate levels of C6S, C4S, and C6S:C4S, but had lower levels compared to their counterparts in the healthy group.

5- In the healthy group, younger people had higher levels of HA and older people, the lowest. However, in the OA group, HA levels increased with age, with the highest levels observed in the oldest OA patients.

6- In the healthy group, women had a negative correlation with C6S, C4S, C6S:C4S, and HA, but the men had a positive correlation with these variables.

7- In the OA group, men and women had a positive collaboration with C6S and C6S:C4S, but a negative correlation with HA and C4S.

8- Both men and women in the healthy group had similar patterns of k-means clustering, with younger people having higher levels of CS and HA than the older people.

9- Both men and women in the OA group had similar patterns of k-means clustering, with younger people having higher levels of CS than older people. Also, both men and women

in the OA group had similar patterns of k-means clustering with HA; the levels increased in older OA patients.

10- The healthy group clusters indicated a more balanced distribution of people across age ranges because the data for this age group was from 10 to 90 years.

11- The OA group clusters were more concentrated, particularly in the younger and older age groups, except for HA, because the age of OA patents in this dataset ranged from 55 to 90 years.

## 4.10 Summary of Chapter

This chapter focused on the k-means clustering analysis of the two data: healthy group data and OA data. In the first part of this chapter, I enumerated the results of the k-means cluster for the healthy group, while the second part of the chapter explained the results of k-means clustering for the OA group. Moving forward, in Chapter Five, I will provide certain important responses to the research questions of this study. I will also discuss the findings, implications of study, its limitations, and offer suggestions for future research.

# CHAPTER FIVE

# DISSECTION AND CONCLUSIONS

In this chapter, the results of k-means clustering analysis are discussed as pertaining to the research questions posed at the beginning of the study in chapter three. To iterate, the purpose of this dissertation was to investigate how the k-means algorithm could expose distinct clusters based on the concentrations of hyaluronic acid (HA), chondroitin sulfate (C6S, C4S), and the C6S:C4S in the synovial fluid of both osteoarthritis (OA) patients and healthy people. By analyzing these variables, the study aimed to identify key patterns and differences in synovial fluid composition across different age groups and genders. Additionally, the study explored the characteristics of the clusters identified by the algorithm, providing insight into how these components are distributed in healthy and OA-affected populations. In the following pages, I examine my findings and compare these with relevant extant study results. concerning the existing research literature. I also enumerate the limitations of this study, its implications and, finally, my suggestions for future research.

## 5.1 Findings for The First Question

RQ1: How does the k-means clustering algorithm expose distinct clusters within datasets that contain information found in the compositions of synovial fluid (HA, C6S, C4S, and C6S:C4S) in patients of all ages and genders who have osteoarthritis (OA) and are healthy?

The k-means clustering algorithm effectively identified distinct clusters based on the compositions of synovial fluid in both OA patients and healthy people by grouping

data points such as age, gender with synovial fluid components HA, C6S, C4S, and the C6S:C4S that share similar characteristics. This finding is consistent with Li's [58] idea that the main objective of ML is to analyze the different models to be learned and the methods to acquire them, to enable the model to make accurate predictions and evaluate data effectively. Xie [15] also argued clustering methods are effective in organizing and analyzing data. In their research, Kumar & Vashistha [80] offered results that are similar to mine regarding the effectiveness of k-means clustering algorithm in identifying distinct clusters; k-means is considered the most simple and common method of clustering that can effectively manage a lot of datasets.

In my study, the algorithm presented three clear clusters and centroid points in each cluster within the datasets for both healthy and OA groups, delineated primarily by age and gender with the concentrations of HA, C6S, C4S, and the C6S: C4S in synovial fluid. These results are compatible with Omar et al.'s [81] study which stated k-means method uses a gradient method that is based on the objective function for reaching a peak value. The computation process involves the gradient-based trend-seeking function, while the process might converge into a local minimum when the starting cluster centroids are not properly selected. This may lead to ineffective clustering. In this dissertation, I found that for the healthy group, the k-means algorithm identified three clear clusters for people aged 10 to 90 years. However, in the OA group, k-means clustering indicated sharper in three clusters as the age ranged 55 to 60 years. and they like the healthy group, younger, middle, and older that show how k-means detect early disease effects.

## 5.2 Findings for the Second Question

RQ2: What are the characteristics of the clusters identified based on the concentrations of hyaluronic acid (HA), chondroitin sulfate (CS), chondroitin 6-sulfate (C6S), and chondroitin 4-sulfate (C4S) in the synovial fluid in osteoarthritis patients (OA) and healthy people?

Characteristics of clusters in the healthy group showed younger people had higher levels of C6S, C4S, and HA; these results indicate better joint health and cartilage function for younger people. Nakayama et al. [40] argued younger healthy people's values of C6S, C4S, and HA were highest, and showed a tendency to decrease with age. In the study I have used for my first dataset, , Uesaka et al. [25], noted "Results indicate that CS, C6S in particular, from cartilaginous tissues into the articular cavity is more active in younger individuals in whom the cartilage mass is large and metabolic activity on the surface of the articular cartilage is high" (p. 473). This aligns with my study: middle-aged people had moderate biomarker levels, which showed early signs of knee joint aging. Thus, these findings are compatible with the results of Uesaka et al. [25] and Nakayama et al. [40].

Applying k-means clustering, my study indicated older, healthy people had lower levels of synovial fluid components, indicating a normal age-related decline in cartilage integrity and knee joint function. This finding is compatible with that of Gul et al. [88], who argued how change in synovial joints affects older individuals' functional abilities. As well, Balazs [89] reported value of HA in normal synovial fluid decreased with age. On their part, Nakayama et al. [40]. observed the concentration of C6S, C4S, C6S:C4S, and HA at ages 60 to 70 years was significantly lower than those aged 20 to 30 years.

Platt et al. [90] found in an experiment with animals that as they aged, the C6S:C4S ratio for the proteoglycan that was freshly generated inside cartilaginous tissues declined. This evidence supports my result too. In general, the results of my study are similar to those of Uesaka et al. [25] who used statistical results. Both confirmed ages have a negative correlation with C6S, C4S, HA concentrations, and the C6S:C4S ratio. The k-means clustering analysis for women, based on their age and the amounts of HA and chondroitin sulfate (C6S, C4S, and the ratio of C6S to C4S), indicates women had lower values than men across most age groups. The results are also similar with Uesaka et al, [25] and Nakayama et al, [40] for healthy people.

On the other hand, characteristics of clusters in the OA group in my study showed young OA patients also had high levels of C6S, C4S, and C6S:C4S, implying early-stage OA with relatively less joint degradation. Also, middle-aged OA patients had moderate levels of C6S, C4S, and C6S:C4S, while older OA patients had low levels of C6S, C4S, and C6S:C4S, reflecting the progression of joint degradation. A study by Lewis et al. [91] discovered 45 OA patients (mean age: 70 years) had much lower C6S levels and C6S:C4S compared to 13 control participants (mean age: 34 years); he found there was no difference in C4S concentrations between the two groups. Sharif et al. (86) found OA patients (mean age: 68.7 years) showed higher levels of C4S concentrations and lower levels of C6S:C4S. In contrast, my study reveals differences of C4S, C6S, C6S:C4S concentration levels between the OA patients and also between healthy people. My study applied data from Uesaka et al.’s [19] research and has many similar findings. There was a difference between the genders' values. All values or women, the numbers were significantly lower than men in the OA group. My results also indicated men and women

had a positive correlation with C6S concentration and the C6S:C4S, but a negative correlation with C4S concentration. This is compatible with results revealed by Uesaka et al. [19].

However, Uesaka et al. [19] didn't get results for HA levels in OA patients. The findings of my study showed HA levels were higher in older people. Comparing the two groups, HA decreased with age for the healthy group but increased in OA patients. The CS concentrations and HA patterns in synovial fluid differed with age and gender for both OA patients and healthy people.

## 5.3 Implications of the Study

This research is important because it explains how to use machine learning k-means clustering to analyze synovial fluid data for healthy people and OA patients. The results of this study may have a significant impact on the fields of engineering and biomedicine. In this section, I highlight the implications drawn from my research.

The first implication is for the field of engineering when engineers use MATLAB programs to analyze ML. I used the MATLAB application of k-means clustering algorithm to explicate the effectiveness of ML tools in engineering, especially for the analysis of biomedical data. Also, the study shows how unsupervised ML approaches such as k-means can extract critical patterns from biomedical data. Moreover, the integration of ML with traditional biomedical data analysis offers engineers multiple opportunities, including using MATLAB to automate medical or engineering data analysis, which reduces the processing time. Additionally, ML technology can design predictive models that predict the progress of diseases and identify high-risk patients,

such as OA patients. Furthermore, engineers can use ML capabilities to develop software for clinicians to help them get faster results.

The second implication of the study results is for the field of biomedicine. I have explained the potential of ML (K-means clustering) in developing biomarker research by identifying distinct patterns in synovial fluid components such as C6S, C4S, and HA. The findings of this study highlight the difference between healthy people and OA patients. These findings can provide more targeted interventions based on a patient's biomarker profile, which supports the development of treatment strategies. Also, the age-related biomarker patterns suggest tracking OA progression and treatment differently.

The third implication of the study results is for researchers; demonstrating the effectiveness of k-means clustering may help future researchers use these approaches to expand other areas of investigation that require pattern recognition and data clustering. Also, the findings of my study promote collaboration between different disciplines, such as ML, engineering, and biomedical research. Finally, by applying ML algorithms to medical problems, future studies can delve deeper to find insights that are not clear through traditional statistical methods.

## 5.4. Limitations of Study

Despite the many advantages enumerated in this study, the k-means approach—like any other method—is not without certain limitations, which affect the results. The first limitation is it is hard to generalize results from one case to a population. In my study, the participant size was small, especially the second dataset for the OA group. A larger and more diverse sample could provide a broader understanding of synovial fluid composition across different populations. The second limitation is that my study focused

on a limited set of biomarkers—HA, C6S, C4S, and the C6S:C4S—while other compositions in synovial fluid, such as glycosaminoglycans, were not examined. The third limitation pertains to the usage of previously published data, which may introduce variability in data collection methods, samples, and analysis methods that were beyond the scope of my control. Further, using previous data limited my control over certain variables, such as the history of the patients or OA severity.

## 5.5 Conclusions

This research provides a comprehensive comparison of different ML (k-means clustering) algorithms as applied to datasets derived from healthy people and OA patients. The study aimed to find optimal clustering methodologies for OA and healthy patients' data, focusing on synovial fluid components C6S, C4S, C6S: C4S, and HA and their association with demographic factors such as age and gender. I gathered data from two previously published research studies: in the first data, the participants were 187 healthy people ranging in age from 10 to 90 years. The second data was for 133 OA patients whose age ranged from 55 to 90. I used ML algorithms (k-means clustering) for MATLAB programs. The study's findings highlight the differences in clustering results among k-means algorithms. The use of k-means clustering proved effective in identifying distinct clusters in synovial fluid compositions; these results are similar to the biological results found in existing research studies. The findings indicate that in the healthy group, younger people exhibited higher levels of synovial fluid components C6S, C4S, C6S: C4S, and HA, while older people showed a significant decline in these components. However, in the OA group, while C6S and C4S levels similarly decreased with age, HA levels increased. Also, in both cases, women had lower values than men.

## 5.6 Future Work

Future research should address several points to build on the findings of this study. First, researchers need to explore the application of advanced ML algorithms such as k-means ++ to better capture the complex relationships between biomarkers and disease stages. Second, scope of study should be expanded to include a broader range of synovial fluid components, which could give a more comprehensive understanding of the mechanisms underlying OA and joint health. Third, a study with fresh data may discover new patterns. I worked with data published in two studies in 2002 and 2004. Overall, my study contributes to the growing field of biomedical and engineering research that supports the integration of machine learning into the diagnosis and management of joint health and disease.

# REFERENCES

Artificial Intelligence and Machine Learning


[1] S. Momtazmanesh, A. Nowroozi, and N. Rezaei, "Artificial Intelligence in Rheumatoid Arthritis: Current Status and Future Perspectives: A State-of-the-Art Review," Rheumatol. Ther., vol. 9, no. 5, pp. 1249-1304, Oct. 2022. doi: 10.1007/s40744-022-00475-4.

[2] X. G. Han and W. Tian, "Artificial Intelligence in Orthopedic Surgery: Current State and Future Perspective," Chin Med J (Engl), vol. 132, no. 21, pp. 2521-2523, Nov. 5, 2019. doi:10.1097/CM9.0000000000000479.

[3] A. Amisha, P. Malik, M. Pathania, and V. K. Rathaur, "Overview of artificial intelligence in medicine," J. Family Med. Prim. Care, vol. 8, no. 7, pp. 2328–2331, 2019. [Online]. Available: https://doi.org/10.4103/jfmpc.jfmpc_440_19

[4] H. Habehh and S. Gohel, "Machine Learning in Healthcare," Curr. Genomics, vol. 22, no. 4, pp. 291-300, Dec. 2021. Doi: 10.2174/1389202922666210705124359.

[5] A. Amisha, P. Malik, M. Pathania, and V. K. Rathaur, "Overview of artificial intelligence in medicine," J. Family Med. Prim. Care, vol. 8, no. 7, pp. 2328–2331, 2019. [Online]. Available: https://doi.org/10.4103/jfmpc.jfmpc_440_19

[6] Martin RK, Solvejg W, Ayoosh P, Andreas P, Håvard V, et al. "Predicting Anterior Cruciate Ligament Reconstruction Revision: A Machine Learning Analysis Utilizing the Norw."

[7] K. N. Kunze, E. M. Polce, A. S. Ranawat et al., "Application of Machine Learning Algorithms to Predict Clinically Meaningful Improvement After Arthroscopic Anterior Cruciate Ligament Reconstruction," Orthopedic Journal of Sports Medicine, vol. 9, no. 10, 2021. Doi: 10.1177/23259671211046575.

[8] Q. An, S. Rahman, J. Zhou, and J.J. Kang, "A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges," Sensors, vol. 23, p. 4178, 2023. [Online]. Available: https://doi.org/10.3390/s23094178

[9] J. Taborri et al., "A Machine-Learning Approach to Measure the Anterior Cruciate Ligament Injury Risk in Female Basketball Players," Sensors, vol. 21, no. 9, p. 3141, 2021. Doi: 10.3390/s21093141.

[14] Binvignat, M. V. Pedoia, A.J. Butte, et al., "Use of machine learning in osteoarthritis research: a systematic literature review," RMD Open, vol. 8, p. e001998, 2022. Doi: 10.1136/rmdopen-2021-001998.

[15] Q. Y. Xie, "K-centers dynamic clustering algorithms and applications," Ph.D. dissertation, University of Cincinnati, 2013.

[18] Kording, K. P., Benjamin, A. S., Farhoodi, R., et al., "The Roles of Machine Learning in Biomedical Science." In: National Academy of Engineering. Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2017 Symposium. Washington (DC): National Academies Press (US); 2018 Jan 22.

[20] Xie, Q.Y., 2013. *K-centers dynamic clustering algorithms and applications*. University of Cincinnati.

[21] Mantas Lukauskas & Tomas Ruzgas, 2022. "A New Clustering Method Based on the Inversion Formula," Mathematics, MDPI, vol. 10(15), pages 1-16, July.

[22] I. Logunova, "K-Means Clustering in Machine Learning," Serokell, Jan. 9, 2023. [Online]. Available: https://serokell.io/blog/k-means-clustering-in-machine-learning.

[23] M.A. Syakur, B.K. Khotimah, E.M.S. Rochman, and B.D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best C," 2018. [Online]. Available: https://doi.org/10.1088/1742-6596/1116/4/042022.

[45] G. Iqbal, N. Shahzad, N. Rafiq, G. Mustafa, and J. Ma, “Deep learning based automated detection of human knee joint’s synovial fluid from magnetic resonance images with transfer learning,” IET Image Processing, May 2020, doi: 10.1049/iet-ipr.2019.1646.

[58] H. Li, Machine Learning Methods, L. Lin and H. Zeng, Trans., 2nd ed. Singapore: Tsinghua Univ. Press, 2024. ISBN: 978-981-99-3916-9. [Online]. Available: https://doi.org/10.1007/978-981-99-3917-6

[59] B. Mahesh, "Machine Learning Algorithms - A Review," International Journal of Science and Research (IJSR), vol. 9, no. 6, pp. 381-386, June 2020. DOI: 10.21275/ART20203995.

[60] H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Computer Science, vol. 2, no. 3, p. 160, 2021. DOI: 10.1007/s42979-021-00592-5.

[61] Q. An, S. Rahman, J. Zhou, and J. J. Kang, "A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges," Sensors, vol. 23, no. 9, p. 4178, 2023. Doi: 10.3390/s23094178.

[62] M. Javaid, A. Haleem, R. Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications," International Journal of Intelligent Networks, vol. 3, pp. 58-73, 2022. Doi: 10.1016/j.ijin.2022.05.002.

[63] J. Sidey-Gibbons and C. Sidey-Gibbons, "Machine learning in medicine: a practical introduction," BMC Medical Research Methodology, vol. 19, p. 64, 2019. Doi: 10.1186/s12874-019-0681-4.

[64] J. Bianchi, A. Ruellas, J. Gonçalves, B. Paniagua, J. Prieto, M. Styner, T. Li, H. Zhu, J. Sugai, W. Giannobile, E. Benavides, F. Soki, M. Atabey, L. Ashman, D. Walker, S. M. Reza Soroushmehr, K. Najarian, and L. Cevidanes, "Osteoarthritis of the Temporomandibular Joint can be diagnosed earlier using biomarkers and machine learning," Scientific Reports, vol. 10, p. 10661, 2020. Doi: 10.1038/s41598-020-64942-0.

[65] K. N. Kunze, E. M. Polce, A. S. Ranawat et al., "Application of Machine Learning Algorithms to Predict Clinically Meaningful Improvement After Arthroscopic Anterior Cruciate Ligament Reconstruction," Orthopedic Journal of Sports Medicine, vol. 9, no. 10, 2021. Doi: 10.1177/23259671211046575.

[66] D.-S. Chen et al., "A Novel Application of Unsupervised Machine Learning and Supervised Machine Learning-Derived Radiomics in Anterior Cruciate Ligament Rupture," Risk Management and Healthcare Policy, vol. 14, pp. 2657-2664, 2021. Doi: 10.2147/RMHP.S312330.

[67] K. Martin et al., "Predicting Anterior Cruciate Ligament Reconstruction Revision: A Machine Learning Analysis Utilizing the Norwegian Knee Ligament Register," The Journal of Bone and Joint Surgery, vol. 104, no. 2, pp. 145-153, Jan. 19, 2022. Doi: 10.2106/JBJS.21.00113.

[68] Corban et al., "Artificial Intelligence in the Management of Anterior Cruciate Ligament Injuries," Orthopedic Journal of Sports Medicine, vol. 9, no. 7, 2021. Doi: 10.1177/23259671211018418.

[69] Y. Uozumi, K. Nagamune, D. Araki, Y. Hoshino, R. Kuroda, and M. Kurosaka, "An Automatic Segmentation of Bone Tunnels after Anterior Cruciate Ligament Reconstruction in MDCT Image Using K-means Clustering," Procedia Computer Science, vol. 22, pp. 590-598, 2013. Doi: 10.1016/j.procs.2013.09.139.

[70] S. Khan, "Design and implementation of efficient brain tumor segmentation with hybrid fuzzy K-means clustering technique," 2024.

[71] A. Islam, S. Seth, T. Bhadra, S. Mallik, A. Roy, A. Li, and M. Sarkar, "Feature selection, clustering and IoMT on biomedical engineering for COVID-19 pandemic: A comprehensive review," Journal of Data Science and Intelligent Systems, vol. 10, 2023. Doi: 10.47852/bonviewJDSIS3202916.

[72] J. Lötsch & A. Ultsch, "Comparative assessment of projection and clustering method combinations in the analysis of biomedical data," 2023. Doi: 10.21203/rs.3.rs-2658032/v1.

[76] Ravel, U.R. and Jani, C. (2016) Implementing & Improvisation of K-Means Clustering Algorithm. International Journal of Computer Science and Mobile Computing, 5, 191-203.

[77] T. S. M. Duarte, "Learning Analytics of K-Means Clustering: A Pilot Study," M.S. thesis, Faculty of Economics, Univ. of Porto, Porto, Portugal, 2018. [Online]. Available: https://repositorio-aberto.up.pt/handle/10216/116970.

[78] E. Backer and A. Jain, "A clustering performance measure based on fuzzy set decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 3, no. 1, pp. 66–75, Jan. 1981.

[79] I. Logunova, "K-Means Clustering in Machine Learning," Serokell, Jan. 9, 2023. [Online].
Available: https://serokell.io/blog/k-means-clustering-in-machine-learning.

[80] Kumar, J. and Vashistha, R. (2017) Estimation of Inter-Centroid Distance Quality in Data Clustering Problem Using Hybridized K-Means Algorithm. 2017 2nd International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, 22-24 February 2017, 1-7. https://doi.org/10.1109/ICECCT.2017.8117896

[81] T. Omar, A. K. Alzahrani, and M. Zohdy, “Clustering approach for analyzing the student’s efficiency and performance based on data,” Journal of Data Analysis and Information Processing, vol. 08, no. 03, pp. 171–182, Jan. 2020, doi: 10.4236/jdaip.2020.83010.

[82] MacQueen, J. B. (1965). On the asymptotic behavior of K-means (No. WMSI WORKING PAPER-89). CALIFORNIA UNIV LOS ANGELES WESTERN MANAGEMENT SCIENCE INST.

[83] Jamadi, N. A., Siraj, M. M., Din, M. M., Mammy, H. K., & Ithnin, N. (2018). Privacy Preserving Data Mining Based on Geometrical Data Transformation Method (GDTM) and K-Means Clustering Algorithm. International Journal of Innovative Computing, 8(2).

[84] “ML: K-means++ algorithm,” GeeksforGeeks, https://www.geeksforgeeks.org/ml-k-means-algorithm/ (accessed Nov. 1, 2023).

[85] D. Arthur and S. Vassilvitskii, "*k-means++: The Advantages of Careful Seeding*," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, New Orleans, LA, USA, 2007, pp. 1027-1035.

[86] S. Ackerman and T. Whittaker, "What is MATLAB?" University of Wisconsin-Madison, 2000. [Online]. Available: https://cimss.ssec.wisc.edu/wxwise/class/aos340/spr00/whatismatlab.htm. [Accessed: Sep. 17, 2024].

Synovial Fluid and Medical Research

[11] G. H. Krishnan, A. Nanda, and A. Natarajan R, “Synovial Fluid Density Measurement for Diagnosis of Arthritis,” Biomedical & Pharmacology Journal, vol. 7, no. 1, pp. 221–224, Jun. 2014, Doi: 10.13005/bpj/476.

[12] G. Wang and F. Li, “An Online Viscosity Measurement Method Based on the Electromechanical Impedance of a Piezoelectric Torsional Transducer,” IEEE Sensors Journal, vol. 18, no. 21, pp. 8781–8788, Nov. 2018, Doi: 10.1109/jsen.2018.2867102.

[13] J. Fu, M. Ni, W. Chai, X. Li, L. Hao, and J. Chen, “Synovial Fluid Viscosity Test is Promising for the Diagnosis of Periprosthetic Joint Infection,” The Journal of Arthroplasty, vol. 34, no. 6, pp. 1197–1200, Jun. 2019, Doi: 10.1016/j.arth.2019.02.009.

[16] T. M. Tamer, "Hyaluronan and synovial joint: function, distribution and healing," Interdisciplinary Toxicology, vol. 6, no. 3, pp. 111-125, Sep. 2013, Doi: 10.2478/intox-2013-0019.

[17] M. M. Temple-Wong et al., “Hyaluronan concentration and size distribution in human knee synovial fluid: variations with age and cartilage degeneration,” Arthritis Research & Therapy, vol. 18, no. 1, Jan. 2016, Doi: 10.1186/s13075-016-0922-4.

[19] S. Uesaka, Y. Nakayama, K. Yoshihara, and H. Ito, "Significance of chondroitin sulfate isomers in the synovial fluid of osteoarthritis patients," Department of Orthopedic Surgery, Nippon Medical School, Tokyo, Japan.

[24] Vincent, H. K., Percival, S. S., Conrad, B. P., Seay, A. N., Montero, C., & Vincent, K. R., "Hyaluronic Acid (HA) Viscosupplementation on Synovial Fluid Inflammation in Knee Osteoarthritis: A Pilot Study," The Open Orthopedics Journal, vol. 7, pp. 378–384, 2013. Doi: 10.2174/1874325001307010378.

[25] S. Uesaka, K. Miyazaki, and H. Ito, “Age-related changes and sex differences in chondroitin sulfate isomers and hyaluronic acid in normal synovial fluid,” Modern

Rheumatology, vol. 14, no. 6, pp. 470–475, Dec. 2004, Doi: 10.1007/s10165-004-0351-0.

[26] S. More, A. Kotiya, A. Kotia, S. K. Ghosh, L. A. Spyrou, and I. E. Sarris, "Rheological properties of synovial fluid due to viscosupplements: A review for osteoarthritis remedy," Computer Methods and Programs in Biomedicine, vol. 196, p. 105644, Nov. 2020, Doi: 10.1016/j.cmpb.2020.105644.

[27] L. A. Mundt and K. Shanahan, Graff's Textbook of Routine Urinalysis and Body Fluids, 2nd ed. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2011.

[28] "Synovial joints," Physiopedia. https://www.physio-pedia.com/Synovial_Joints#:~:text=Synovial%20joints%20are%20the%20most,presence%20of%20a%20joint%20cavity.

[29] Seidman AJ, Limaiem F., "Synovial Fluid Analysis." 2022 May 8. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–. PMID: 30725799.

[30] Walker, W., "Effects of Ageing on Joints," Apr. 11, 2022.

[31] D. Prekasan and K. K. Saju, "Review of the Tribological Characteristics of Synovial Fluid," Procedia Technology, vol. 25, pp. 1170–1174, Jan. 2016, Doi: 10.1016/j.protcy.2016.08.235.

[32] B. P. Conrad, "The effects of glucosamine and chondroitin on the viscosity of synovial fluid in patients with osteoarthritis," M.S. thesis, Dept. Eng., Univ. of Florida, Gainesville, FL, 2001.

[33] S. K. Strasinger and M. S. Di Lorenzo, "Urinalysis and Body Fluids," 6th edition, Philadelphia, PA, USA: F.A. Davis Company, 2014.

[34] E. Andersson, E. Tykesson, L. S. Lohmander, N. G. Karlsson, C. Jin, E. Mirgorodskaya, P. Swärd, and A. Struglics, "Quantification of chondroitin sulfate, hyaluronic acid and N-glycans in synovial fluid - A technical performance study," Osteoarthritis and Cartilage Open, vol. 5, no. 3, p. 100380, Jun. 2023, Doi: 10.1016/j.ocarto.2023.100380.

[35] M. M. Abdallah, N. Fernández, A. A. Matias, and M. d. R. Bronze, "Hyaluronic acid and Chondroitin sulfate from marine and terrestrial sources: Extraction and purification methods," *Carbohydrate Polymers.*

[36] P. A. S. Mourão, S. Rozenfeld, J. Laredo, and C. P. Dietrich, "The distribution of chondroitin sulfates in articular and growth cartilages of human bones,"

*Departamento de Bioquímica e Farmacologia, Escola Paulista de Medicina*, São Paulo, Brazil, Aug. 1975.

[37] Spain L, Cheneler D (2020) Acoustic monitoring of joint health". InData Acquisition-Recent Advances and Applications in Biomedical Engineering Intech Open.

[38] M. Marian, R. Shah, B. Gashi, S. Zhang, K. Bhavnani, S. Wartzack, and A. Rosenkranz, "Exploring the lubrication mechanisms of synovial fluids for joint longevity - A perspective," Colloids Surf B Biointerfaces, vol. 206, p. 111926, Oct. 2021. doi: 10.1016/j.colsurfb.2021.111926. Epub Jun. 16, 2021. PMID: 34153619.

[39] J. Liao, S. Miramini, X. Liu, and L. Zhang, "Computational study on synovial fluid flow behaviour in cartilage contact gap under osteoarthritic condition," Computers in Biology and Medicine, vol. 123, p. 103915, Aug. 2020, Doi: 10.1016/j.compbiomed.2020.103915.

[40] Y. Nakayama, T. Narita, A. Mori, S. Uesaka, K. Miyazaki, and H. Ito, "The Effects of Age and Sex on Chondroitin Sulfates in Normal Synovial Fluid," *Arthritis & Rheumatism*, vol. 46, no. 8, pp. 2105–2108, Aug. 2002, Doi: 10.1002/art.10424.

[41] Fonsi, M., El Amrani, A.-I., Gervais, F., and Vincent, P., "Intra-Articular Hyaluronic Acid and Chondroitin Sulfate: Pharmacokinetic Investigation in Osteoarthritic Rat Models," Current Therapeutic Research, vol. 92, 2020, Art. no. 100573, Doi: 10.1016/j.curtheres.2019.100573.

[42] M.P. Brown, D.V.M., M.Sc., T.N. Trumble, D.V.M., Ph.D., A.H.K. Plaas, Ph.D., M. Romano, B.S., J. Hernandez, D.V.M., Ph.D., K.A. Merritt, B.S., "Exercise and injury increase chondroitin sulfate chain length and decrease hyaluronan chain length in synovial fluid," *Osteoarthritis and Cartilage*, vol. 15, no. 8, pp. 937-944, June 2007, Doi: 10.1016/j.joca.2007.04.005.

[43] G. H. Krishnan, A. Nanda, and A. Natarajan R, "Synovial Fluid Density Measurement for Diagnosis of Arthritis," Biomedical & Pharmacology Journal, vol. 7, no. 1, pp. 221–224, Jun. 2014, Doi: 10.13005/bpj/476.

[44] G. Wang and F. Li, "An Online Viscosity Measurement Method Based on the Electromechanical Impedance of a Piezoelectric Torsional Transducer," IEEE Sensors Journal, vol. 18, no. 21, pp. 8781–8788, Nov. 2018, Doi: 10.1109/jsen.2018.2867102. [46] J. Fu, M. Ni, W. Chai, X. Li, L. Hao, and J. Chen, "Synovial Fluid Viscosity Test is Promising for the Diagnosis of Periprosthetic Joint Infection," The Journal of Arthroplasty, vol. 34, no. 6, pp. 1197–1200, Jun. 2019, Doi: 10.1016/j.arth.2019.02.009.

[47] J. Liao, S. Miramini, X. Liu, and L. Zhang, “Computational study on synovial fluid flow behavior in cartilage contact gap under osteoarthritic condition,” Computers in Biology and Medicine, vol. 123, p. 103915, Aug. 2020, Doi: 10.1016/j.compbiomed.2020.103915.

[48] Abdullzaher, et al, Effects Ageing on Bone. (2020),

[49] Department of Health & Human Services, "Joints," Better Health Channel. [Online]. Available: https://www.betterhealth.vic.gov.au./health/conditionsandtreatments/joints.

[50] “Effects of ageing on joints,” Physiopedia. https://www.physio-pedia.com/Effects_of_Ageing_on_Joints.

[51] “Aging changes in the bones - muscles - joints: MedlinePlus Medical Encyclopedia.” https://medlineplus.gov/ency/article/004015.htm

[52] M. Temple-Wong et al., “Hyaluronan concentration and size distribution in human knee synovial fluid: variations with age and cartilage degeneration,” Arthritis Research & Therapy, vol. 18, no. 1, Jan. 2016, Doi: 10.1186/s13075-016-0922-4.

[53] K. Amano, J. L. Huebner, T. V. Stabler, M. Tanaka, C. E. McCulloch, I. Lobach, N. E. Lane, V. B. Kraus, C. B. Ma, and X. Li, "Synovial Fluid Profile at the Time of Anterior Cruciate Ligament Reconstruction and Its Association With Cartilage Matrix Composition 3 Years After Surgery," The American Journal of Sports Medicine, vol. 46, no. 4, pp. 890-899, 2018.

[54] S.Y. Kim-Wang, A.G. Holt, A.M. McGowan, et al., "Immune cell profiles in synovial fluid after anterior cruciate ligament and meniscus injuries," Arthritis Res Ther, vol. 23, p. 280, 2021. https://doi.org/10.1186/s13075-021-02661-1.

[55] A.M. Kiapour, J.T. Sieker, B.L. Proffen, T.T. Lam, B.C. Fleming, and M.M. Murray, "Synovial fluid proteome changes in ACL injury-induced posttraumatic osteoarthritis: Proteomics analysis of porcine knee synovial fluid," PloS one, vol. 14, no. 3, p. e0212662, 2019. https://doi.org/10.1371/journal.pone.0212662.

[56] R.H. Brophy, L. Cai, Q. Zhang, R.R. Townsend, and M.F. Rai, "Proteomic Profile Analysis of Synovial Fluid in Patients with Anterior Cruciate Ligament Tears," The American Journal of Sports Medicine, vol. 50, no. 11, pp. 2935-2943, 2022. Doi: 10.1177/03635465221112652.

[57] M.T. Kingery, U. Anil, E.J. Berlinberg, A.J. Clair, L. Kenny, and E.J. Strauss, "Changes in the Synovial Fluid Cytokine Profile of the Knee Between an Acute Anterior Cruciate Ligament Injury and Surgical Reconstruction," The American

Journal of Sports Medicine, vol. 50, no. 2, pp. 451-460, 2022. https://doi.org/10.1177/03635465211062264.

[73] R. Takeda, T. Matsumoto, Y. Omata, H. Inui, S. Taketomi, Y. Nagase, T. Nishikawa, H. Oka, and S. Tanaka, "Changes in knee joint destruction patterns among patients with rheumatoid arthritis undergoing total knee arthroplasty in recent decades," *Clinical Rheumatology*, vol. 42, 2023. Doi: 10.1007/s10067-023-06620.

[74] Heard, B.J., Martin, L., Rattner, J.B. et al. Matrix metalloproteinase protein expression profiles cannot distinguish between normal and early osteoarthritic synovial fluid. BMC Musculoskeletal Discord 13, 126 (2012). https://doi.org/10.1186/1471-2474-13-126.

[75] B. J. Heard, M. J. Fritzler, J. P. Wiley, J. McAllister, L. Martin, H. El-Gabalawy, D. A. Hart, C. B. Frank, and R. Krawetz, "Intraarticular and Systemic Inflammatory Profiles May Identify Patients with Osteoarthritis," The Journal of Rheumatology, vol. 40, no. 8, pp. 1379-1387, Aug. 2013. [Online]. Available: https://doi.org/10.3899/jrheum.121204.

[87] L. Leung, "Validity, reliability, and generalizability in qualitative research," J. Family Med. Prim. Care, vol. 4, no. 3, pp. 324-327, Jul.-Sep. 2015, Doi: 10.4103/2249-4863.161306.

[88] R. Gul, I. Nazir, M. I. Amirzada, F. Jahan, F. Naseer, and T. A. Baig, "Aging and Synovial Joint Function: Changes in Structure and Implications for Mobility," *Intech Open*, submitted Nov. 6, 2023, reviewed Nov. 10, 2023, published Mar. 27, 2024. DOI: 10.5772/intechopen.1003866.

[89] E. A. Balazs, "The physical properties of synovial fluid and the special role of hyaluronic acid," in *Disorders of the Knee*, A. J. Helfet, Ed. Philadelphia, PA: Lippincott, 1982, pp. 61–74.

[90] D. Platt, J. L. E. Bird, and M. T. Bayliss, "Ageing of equine articular cartilage: structure and composition of aggrecan and decor in," *Equine Vet. J.*, vol. 30, pp. 43–52, 1998.

[91] S. Lewis, M. Crossman, J. Flannelly, *et al.*, "Chondroitin sulphation patterns in synovial fluid in osteoarthritis subsets," *Ann. Rheum. Dis.*, vol. 58, pp. 441–445, 1999.

# PUBLICATIONS

1- B. Alabkary and M. Zohdy, "Review of Synovial Fluid Properties and Measurement," Orthopedics and Rheumatology Open Access Journal, vol. 20, no. 2, pp. 556033, 2022. Doi: 10.19080/OROAJ.2022.20.556033.

2- B. Alabkary, G. Alawami, C. Kobus, S. Kamel-ElSayed, H. S. Abdel-Aty Zohdy, and M. Zohdy, "Comparison of Machine Learning Algorithms (K-means) Using Data of Anterior Cruciate Ligament Reconstruction (ACLR) Patients," Engineering Technology Open Access Journal, vol. 6, no. 1, pp. 555676, 2024. Doi: 10.19080/ETOAJ.2024.06.555676.