Programa de Doctorado en Ingeniería Matemática, Estadística e Investigación Operativa

por la

Universidad Complutense de Madrid

y la

Universidad Politécnica de Madrid

Advanced data-driven modal decomposition techniques for medical imaging applications

Tesis Doctoral

Nourelhouda Groun

Directores

Soledad Le Clainche

Jesús Garicano-Mena

Año

2024

UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingenieros de Telecomunicación

Doctoral Degree in Mathematical Engineering, Statistics and Operational Research

Advanced data-driven modal decomposition techniques for medical imaging applications

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Nourelhouda Groun
Master's degree in functional analysis and theory of linear operators

Under the supervision of:
Dr. Soledad Le Clainche (Supervisor)
Dr. Jesús Garicano-Mena (Supervisor)

Madrid, 2024

Title: Advanced data-driven modal decomposition techniques for medical imaging applications

Author: Nourelhouda Groun

Doctoral Programme: Mathematical Engineering, Statistics and Operational Research (IMEIO)

Thesis Supervision:

Dr. Soledad Le Clainche, Escuela Técnica Superior de Ingeniería Aeronáutica y del Espacio (Supervisor)

Dr. Jesús Garicano-Mena, Escuela Técnica Superior de Ingeniería Aeronáutica y del Espacio (Supervisor)

External Reviewers:

Thesis Defense Committee:

Thesis Defense Date:

This thesis has been supported by the grant TED2021-129774B-C21, funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU"/PRTR. The thesis has also been supported by 041 Bis/PG/Espagne/2020-2021 funded by the Algerian Ministry of Higher Education and Scientific Research. Finally, we extend our gratitude to the Algerian government for the provided doctoral scholarship.

Esta tesis ha contado con el apoyo de la subvención TED2021-129774B-C21, financiadas por MCIN/AEI/10.13039/501100011033 y por la Unión Europea "NextGenerationEU"/PRTR. La tesis también ha sido apoyada por 041 Bis/PG/Espagne/2020-2021 financiada por el Ministerio de Educación Superior e Investigación Científica de Argelia. Finalmente, extendemos nuestro agradecimiento al gobierno argelino por la beca de doctorado proporcionada.

Knowledge is light and ignorance is darkness.

Acknowledgement

These past three years have been an amazing journey for me. I consider pursuing my PhD in Universidad Politécnica de Madrid (UPM) is one of the best experiences of my life. This journey have been a roller coaster ride, as it might seem scary but in fact it is full of excitement and good memories. Nevertheless, these past three year would not be the same without the help and support of many people.

First, I would like to express my gratitude for my supervisors Dr. Soledad Le Clainche and Dr. Jesus Garicano Mena, for their continuous support and never-ending guidance. They have trusted me every step of the way and never doubted me, so thank you for your patience and for making me believe in my abilities. I would also like to thank my tutor Dr. Eusebio valero for accepting me in his research group and for his trust and support. During all my time as a PhD student, I have felt proud to work with these three outstanding researchers. They have helped me learn and overcome all the challenges I have faced, and guided me to produce high-quality researches. Not only they are professionals and respected experts in their field, but also very kind and warm-hearted people.

I would like to thank next our NUMATH group. Since the beginning they welcomed us, took the time to listen to our concerns and obstacles and tried their best to help us, by giving helpful advices and instructions.

Through these years I had the chance to meet many amazing colleagues and friends. My office mates Andres, Kenza, Kheir-eddine, Himpu, Pablo, Pol, Rodrigo, Fernando, Alfredo, Aurelio, Saumitra and Ja Qing, thank you all for the fun conversations and cheerful time in the office. Also Ivan, Jonathan, David, Stefano and Vector, having lunch with all these nice people was my favorite time of the day. All these people have been a marvelous support through the tough times of my PhD. I appreciate each and everyone of them.

I thank all my friends who have shared great moment with me through these past three years and in my life in general. Sara, Kenza and Chahinez, thank you for being there for me in every step. I think meeting such friends is a great blessing, I consider my self extremely lucky and I'm grateful for everything you have done for me so far.

Finally, I would like to give my greatest appreciation to my family. I first show my most sincere to my parents, my mother Bahria and my father Salem, for their love, caring and unlimited support. They have tried their best, their whole life, to provide me a good life, and encourage me to pursue my studies and career. I also thank my twin sister and soulmate Fatima for standing by my side my whole life. My older brothers Houcine and Houssam for supporting and protecting me. And of course my younger, little brother, my sweet angel

Abdelkoudouss, thank you for always checking on me and missing me every single day. I would also like to thank a friend, which I consider family, Hanene, I want to say thank you for being my best friend all these years, for supporting and believing in me. Thank you all for everything.

Most importantly, thank you Allah for all these blessings.

Abstract

Data-driven modal decomposition techniques are methods with sound mathematical foundation, used to analyze and extract useful information from complex datasets. These techniques rely heavily on numerical linear algebra concepts. Consider the singular value decomposition (SVD), which is considered the most important matrix factorization of the computational era. This method is representative of most of the data-driven methods. Hence, this basis will allow us to interpret most of these modal decomposition techniques as matrix factorization approaches (illustrated in Fig. (1)) as follows:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{C}^*, \tag{1}$$

where $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ is the data matrix. The matrix $\boldsymbol{A} \in \mathbb{C}^{n\times r}$ describes the spatial aspect of the dataset, $\boldsymbol{C} \in \mathbb{C}^{m\times r}$ is related to the temporal aspect of the dataset, $\boldsymbol{B} \in \mathbb{C}^{r\times r}$ is a diagonal matrix (with $1 \leq r \leq min(m, n)$), whose role is to gauge the importance of the spatial and temporal components of the data matrix, whereas * denotes the complex-conjugate.

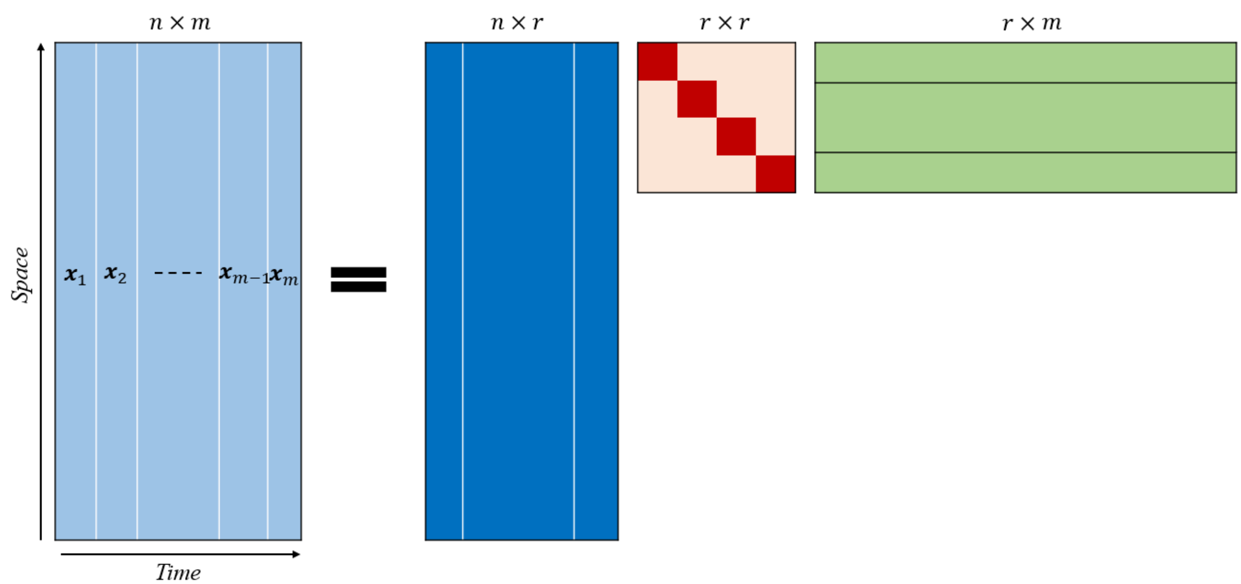


Figure 1: Diagram illustrates the matrix factorization procedure.

Among the plethora of modal decomposition techniques, in this contribution, we focus on one of the most well-known data-driven techniques, the dynamic mode decomposition (DMD). This technique was originally introduced in the field of fluid mechanics, but extended its application to cover plentiful other fields (biology, healthcare, sports ...). Following its great success in these different fields, many researchers have worked on developing extensions of this technique, which can be more robust, more accurate and cover a wider range of applications. We introduce one of these extensions, named the higher order dynamic mode decomposition (HODMD), which is widely used in fluid dynamics applications. The HODMD method excels at identifying and extracting dominant patterns representing the complex dynamics of fluid flows.

In this contribution, we shift the domain of applications of the HODMD technique from fluid dynamics to the medical field. More specifically, we consider the analysis of data obtained from common use medical imaging techniques. The investigation of this technique on medical data has led not only to the demonstration of the robustness of the HODMD applications, but also to the discovery of several, new applications.

First, the HODMD technique is used to analyze echocardiography video loops taken from mice in healthy conditions and mice afflicted with different cardiac diseases (diabetic cardiomyopathy, obesity, TAC hypertrophy or myocardial infarction). The outcome of this first application proves, once again, the efficiency of the HODMD algorithm in pattern identification and feature extraction, as the technique was able to capture and represent the dominant features related to the different cardiac conditions.

Second, the HODMD algorithm is used for the analysis of sequences of Magnetic Resonance Imaging (MRI). In this application the HODMD is used first as a reduced order model (ROM) to generate new, extended databases. Furthermore, the HODMD is combined with a novel interpolation technique with the aim of recovering missing information and the reconstruction of corrupted images.

Keywords: Higher Order Dynamic Mode Decomposition, Data analysis, Feature extraction, Medical imaging.

Resumen

Las técnicas de descomposición modal basadas en datos son métodos con una base matemática sólida, que se utilizan para analizar y extraer información útil de conjuntos de datos complejos. Estas técnicas se basan en gran medida en conceptos de álgebra lineal numérica. Considere la Singular Value Decomposition (SVD), que se considera la factorización de matrices más importante de la era computacional. Este método es representativo de la mayoría de los métodos basados en datos. Por lo tanto, esta base nos permitirá interpretar la mayoría de estas técnicas de descomposición modal como enfoques de factorización de matrices (ilustrada en la Fig. (2)) de la siguiente manera:

$$\boldsymbol{X} = \boldsymbol{ABC}^*, \tag{2}$$

donde $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ es la matriz de datos. La matriz $\boldsymbol{A} \in \mathbb{C}^{n\times r}$ describe el aspecto espacial del conjunto de datos, $\boldsymbol{C} \in \mathbb{C}^{m\times r}$ está relacionado con el aspecto temporal del conjunto de datos, $\boldsymbol{B} \in \mathbb{C}^{r\times r}$ es una matriz diagonal (con $1 \leq r \leq min(m, n)$) cuyo papel es medir la importancia de los componentes espaciales y temporales de la matriz de datos, mientras que * denota el conjugado complejo.

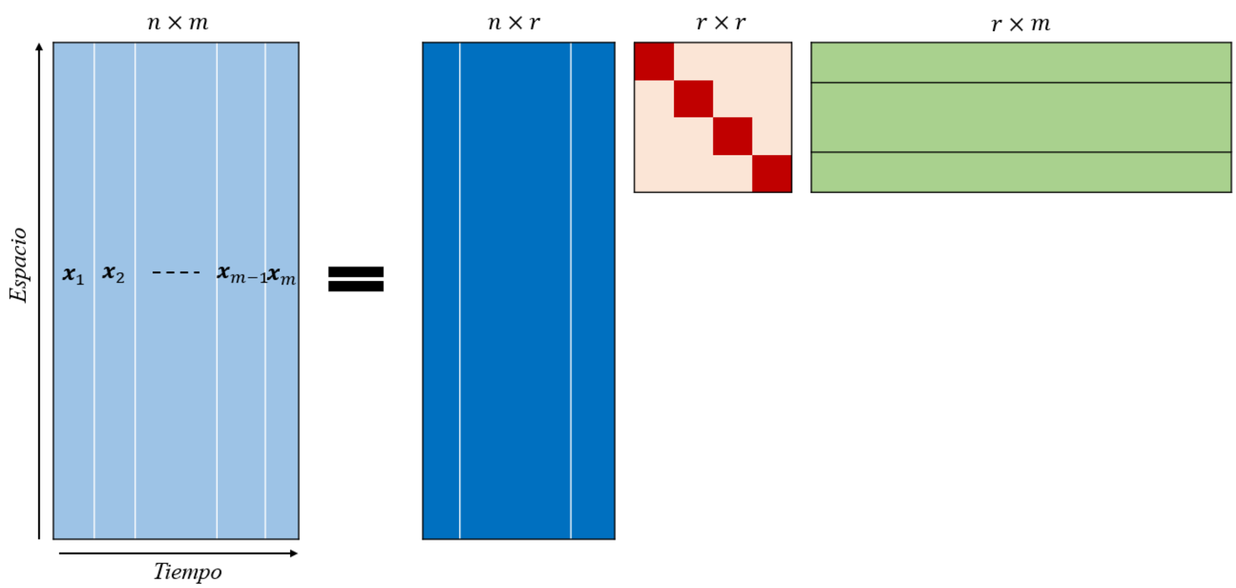


Figure 2: El diagrama ilustra el procedimiento de factorización matricial.

Sin embargo, entre la plétora de técnicas de descomposición modal, en esta contribución, nos centramos en una de las técnicas basadas en datos más conocidas, la Dynamic Mode Decomposition (DMD). Esta técnica se introdujo originalmente en el campo de la mecánica de fluidos, pero extendió su aplicación a muchos otros campos (biología, salud, deportes...). Tras su gran éxito en estos diferentes campos, muchos investigadores han trabajado en el desarrollo de extensiones de esta técnica, que pueden ser más robustas, más precisas y cubrir una gama más amplia de aplicaciones. Presentamos una de estas extensiones, Higher Order Dynamic Mode Decomposition (HODMD). Esta técnica, es ampliamente utilizada en aplicaciones de dinámica de fluidos. El método HODMD sobresale en la identificación y extracción de patrones dominantes que representan la dinámica compleja de los flujos de fluidos.

En esta contribución, cambiamos el dominio de aplicaciones de la técnica HODMD de la dinámica de fluidos al campo médico. Más específicamente, consideramos el análisis de datos obtenidos de técnicas de imágenes médicas de uso común. La investigación de esta técnica en datos médicos ha llevado no solo a la demostración de la robustez de las aplicaciones HODMD, sino también al descubrimiento de varias aplicaciones nuevas.

Primero, la técnica HODMD se usa para analizar bucles de video de ecocardiografía tomados de ratones en condiciones saludables y ratones afectados por diferentes enfermedades cardíacas (miocardiopatía diabética, obesidad, hipertrofia TAC o infarto de miocardio). El resultado de esta primera aplicación demuestra, una vez más, la eficiencia del algoritmo HODMD en la identificación de patrones y extracción de características, ya que la técnica fue capaz de capturar y representar las características dominantes relacionadas con las diferentes condiciones cardíacas.

Segundo, el algoritmo HODMD se utiliza para el análisis de secuencias de Imágenes por Resonancia Magnética (MRI). En esta aplicación, el HODMD se usa primero como un Reduced Order Model (ROM) para generar nuevas bases de datos ampliadas. Además, el HODMD se combina con una novedosa técnica de interpolación con el objetivo de recuperar información faltante y reconstruir imágenes corruptas.

Palabras clave: Higher Order Dynamic Mode Decomposition, Análisis de datos, Extracción de características, Imágenes médicas.

Abstrait

Les techniques de décomposition modale basées sur les données sont des méthodes reposant sur des bases mathématiques solides, utilisées pour analyser et extraire des informations utiles à partir d'ensembles de données complexes. Ces techniques s'appuient fortement sur les concepts d'algèbre linéaire numérique. Considérons la décomposition en valeurs singulières (ou SVD, de l'anglais singular value decomposition), qui est considérée comme la factorisation matricielle la plus importante de l'ère informatique. Cette méthode est représentative de la plupart des méthodes basées sur les données. Par conséquent, cette base nous permettra d'interpréter la plupart de ces techniques de décomposition modale comme des approches de factorisation matricielle (illustrées dans la Fig. (3)) comme suit:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{C}^{*}, \tag{3}$$

où $\boldsymbol{X} \in \mathbb{R}^{n \times m}$ est la matrice de données. La matrice $\boldsymbol{A} \in \mathbb{C}^{n \times r}$ décrit l'aspect spatial de l'ensemble de données, $\boldsymbol{C} \in \mathbb{C}^{m \times r}$ est lié à l'aspect temporel de l'ensemble de données, $\boldsymbol{B} \in \mathbb{C}^{r \times r}$ est un matrice diagonale (avec $1 \leq r \leq min(m, n)$) dont le rôle est de jauger l'importance de les composantes spatiales et temporelles de la matrice de données, alors que * désigne le complexe-conjugué.

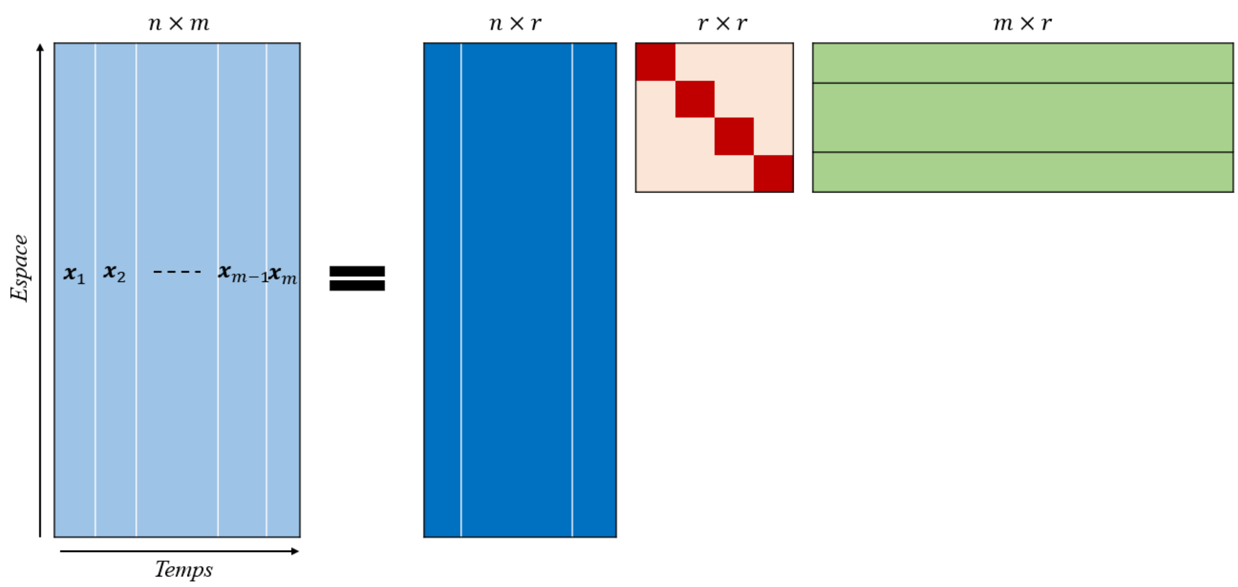


Figure 3: Le diagramme illustre la procédure de factorisation matricielle.

Parmi la pléthore de techniques de décomposition modale, dans cette contribution, nous nous concentrons sur l'une des techniques basées sur les données les plus connues, la décomposition en mode dynamique (DMD). Cette technique a été introduite à l'origine dans le domaine de la mécanique des fluides, mais a étendu son application à de nombreux autres domaines (biologie, santé, sport...). Suite à son grand succès dans ces différents domaines, de nombreux chercheurs

ont travaillé au développement d'extensions de cette technique, qui peuvent être plus robustes, plus précises et couvrir un plus large éventail d'applications. Nous introduisons l'une de ces extensions, appelée décomposition en mode dynamique d'ordre supérieur (HODMD), qui est largement utilisée dans les applications de dynamique des fluides. La méthode HODMD excelle dans identifier et extraire des modèles dominants représentant la dynamique complexe des écoulements de fluides.

Dans cette contribution, nous déplaçons le domaine d'application de la technique HODMD de la dynamique des fluides vers le domaine médical. Plus spécifiquement, nous considérons l'analyse des données obtenues à partir de techniques d'imagerie médicale d'usage courant. L'investigation de cette technique sur des données médicales a conduit non seulement à la démonstration de la robustesse des applications HODMD, mais également à la découverte de plusieurs nouvelles applications.

Premièrement, la technique HODMD est utilisée pour analyser des boucles vidéo d'échocardiographie réalisées sur des souris en bonne santé et des souris atteintes de différentes maladies cardiaques (cardiomyopathie diabétique, obésité, hypertrophie TAC ou infarctus du myocarde). Le résultat de cette première application prouve, une fois de plus, l'efficacité de l'algorithme HODMD dans l'identification de modèles et l'extraction de caractéristiques, car la technique a pu capturer et représenter les caractéristiques dominantes liées aux différentes conditions cardiaques.

Deuxièmement, l'algorithme HODMD est utilisé pour l'analyse de séquences d'Imagerie par Résonance Magnétique (IRM). Dans cette application, le HODMD est d'abord utilisé comme modèle d'ordre réduit (ROM) pour générer de nouvelles bases de données étendues. De plus, le HODMD est combiné à une nouvelle technique d'interpolation dans le but de récupérer les informations manquantes et de reconstruire les images corrompues.

Mots clés: Décomposition en mode dynamique d'ordre supérieur, Analyse de données, Extraction de caractéristiques, Imagerie médicale.

Contents

List of Figures

List of Tables

Abbreviations

DMD Dynamic Mode Decomposition

HODMD Higher Order Dynamic Mode Decomposition

SVD Singular Value Decomposition

HOSVD Higher Order Singular Value Decomposition

POD Proper Orthogonal Decomposition

PCA Principal Component Analysis

ROM Reduced Order Model

RRMSE Relative Root Mean Squared Error

CNN Convolutional Neural Network

SVM Support Vector Machine

LR Learning Rate

ReLU Rectified Linear Unit

MRI Magnetic Resonance Imaging

LAX Long Axis

SAX Short Axis

H Healthy

DC Diabetic Cardiomyopathy

MI Myocardial Infarction

Ob Obesity

TAC Transverse Aortic Constriction (TAC) Hypertension

Nomenclature

$\boldsymbol{V}$ Data matrix (Snapshot matrix)

T Time interval

t_k Temporal term at time k

$\boldsymbol{v}_k$ Snapshot collected at time t_k

K Number of snapshots

Δt Sampling time (time between snapshots)

$\boldsymbol{T}$ Data in tensor form (snapshot tensor)

$\boldsymbol{R}$ Koopman matrix

M Number of modes

$\boldsymbol{u}_m$ DMD modes

a_m Amplitudes of the DMD modes

ω_m Frequencies of the DMD modes

δ_m Growth rates of the DMD modes

ε_1 SVD threshold

ε_2 DMD threshold

ε_{HOSVD} HOSVD threshold

d Delay parameter of HODMD

$\boldsymbol{Y}$ Original MRI slice

$\boldsymbol{Y}^{approx}$ New reconstructed MRI slice

$\|.\|_F$ Frobenius norm

I Number of MRI sequences (slices)

$\mathrm{N_x}$ Number of pixels in the X axis

$\mathrm{N_y}$ Number of pixels in the Y axis

x_1 Position of pixel in X axis

x_2 Position of pixel in Y axis

Part I

General introduction

Summary This part provides a general overview of the research described in this thesis. This overview presents the research's background, including a profound description of state of the art, the interest and the goal of the research. Finally, we introduce the scientific dissemination produced during this investigation process and a short description of the structure of the thesis.

Chapter 1

Introduction

Dynamical systems provide one of the most complete and well-connected fields of mathematics, bridging diverse topics from linear algebra and differential equations to numerical analysis and geometry. These systems provide a mathematical framework that describes the world around us, whether turbulent fluids, climate science, disease modeling, neuroscience, and nearly every other system that evolves in time. Hence, the ability to understand, control, and predict the behavior of these systems contributes to make our lives much easier. However, handling these complex systems can be very challenging, as we face several obstacles such as large scale, high-fidelity mathematical models for the analysis and design of complex systems, which can be very time consuming and computationally expensive. Moreover, the increase in measurement data from complex systems, where the amount of data we are obtaining from experiments, simulations and historical records can be extremely large. In order to address some of these challenges, we can always resort to data-driven techniques. These algorithms, which rely on data, also involve the use of statistical analysis, applied mathematics, optimization and other computational methods to analyse large datasets and extract patterns, correlations, and useful information and insights.

The main aim of this work revolves around the development of hybrid reduced order models (ROMs) [2] based on physical principles. These ROMs are established on the use of different data-driven modal decomposition tools to extract information about the physics of the data understudy. In addition, the knowledge obtained from these ROMs can be combined with machine learning approaches to create advanced, robust and efficient novel tools, as well as enhancing the existing deep learning algorithms.

In this contribution, the leading modal decomposition tool used is a technique that has proven to be one of the most sophisticated data-driven techniques, the Higher Order Dynamic Mode Decomposition (HODMD). Nevertheless two other modal decomposition techniques, named the singular value decomposition (SVD), and the higher order singular value decomposition (HOSVD), are also employed.

In this research, the proposed techniques are used to inspect two different medical imaging datasets: i) echocardiography images taken from several mice, either in healthy conditions or afflicted by different cardiac diseases. ii) The second dataset is cardiac cine magnetic resonance imaging (MRI), taken from mice, either in healthy conditions or diagnosed with SFSR4 Hypertrophy.

Using these physics-aware ROMs it is possible to achieve several applications: i) first, we exploit the HODMD advantageous properties in dynamics identification, as well as the use of the HOSVD for noise cleaning, in order to investigate relevant frequencies and coherent patterns for each disease in both datasets (echocardiography and MRI). ii) Second, Using the analysis of the MRI datasets, we provide a three-dimensional reconstruction of the heart and build a reduced order model of the heart dynamics. Finally, the HODMD is combined with a singular value decomposition (SVD) based interpolation technique, to reconstruct corrupted or missing MRI images.

1.1 State of the art

When it comes to the history of data science, many may think that it was founded in the early 70s following the establishment of the term data science by Bill Cleveland in 1974. However, data science was in fact introduced a decade earlier, when the mathematician John Tukey wrote an article in 1962 titled "The Future of Data Analysis" [3], where he surprised the statistics profession by explaining why he thought such research was too narrowly focused and the research scope of statistics needed to be dramatically enlarged and redirected. Furthermore, he mentions the possibility of the existence of a science that learns from data. As a consequence, and starting from this point, researchers from different fields began to show interest in this unrecognized science.

Through the years, data science has rapidly evolved because of the different discoveries it can provide from raw data in numerous fields such as marketing ([4], [5]), cyber-security ([6], [7], [8]), sports ([9], [10], [11]), health care ([12], [13], [14]) etc. . . . These data based methods are being more and more approachable because of the fact that they learn from data and do not require the knowledge of the underlying equations. Note that these statements do not imply that engineering or mathematical physics are being replaced by data science, on the contrary, it is enhancing it for our century.

When working with data science, researchers usually -but not necessarily- deal with large data matrices, whether it is a statistical data (gender, age, wight, hight...) or data produced by complex systems [15]. For example, an experiment or a simulation will result a data matrix with each column includes all of the measurements at a given time; the intensity of pixels of a grayscale image can be sorted in matrix, or a number of images can be reshaped into column vectors and arranged in a matrix where each column will represent a frame of a video.

When it comes to the medical field, medical imaging [16] occupies a major part of medical data. However, in order to make an accurate diagnosis, a considerable number of images need to be examined but only a few will be expected to show abnormalities. Images can be taken from different measurements, at different points in time from the same modality or obtained from different modalities such as computerized tomography (CT), magnetic resonance (MR), angiography or ultrasound. Nevertheless, analyzing this large amount of images in difficult for medical practitioners and possibly leading to late or wrong diagnoses, which summons the intervention of different data science techniques.

1.1.1 Data science for medical imaging

The involvement of data science in medical imaging has been active for decades, especially the use of neural networks, as they are very efficient and cover a wide range of applications. Cios *et al.* [17] used in their study two types of backpropagational neural networks in order to distinguish between normal, myocardial infarctions of the anterior ventricular septum (AVS), and hypertrophic cardiomyopathy (HCM) in a first data set, and between normal and myocardial infarctions of the posterior wall (PW) in a second set using 2D echocardiographic images. Gross *et al.* [18] were able to train a neural network (NN) to chose one or more diagnoses from a list of 12 possible diagnoses provided by pediatric radiologist after the observation of 77 neonatal chest radiographs. The NN was tested with 103 chest radiographs from a separate group of neonates and reached to approximately 80% agreement between the NN and the radiologists' predictions (see also [19–21]).

Chin-Tu *et al.* [22] on the other hand tried to benefit the medical field by developing a Constraint Satisfaction Neural Network (CSNN) to solve the problem of medical image segmentation. The image segmentation process was achieved by using the CSNN with fixed structure and weights on medical images obtained from CT, MRI and PET (see also [23–26]).

Fujita *et al.* [27] introduced a computerized system that can aid in the radiologist's diagnosis in the detection and classification of coronary artery diseases in myocardial SPECT bull's-eye images. Even though their approach did not do better than an experienced radiologist, it was able to outperform a radiology resident. With another image processing technique Crooks and Fallone [28] proposed an algorithm for edge detection and edge enhancement for medical imaging. The algorithm, which was called Histogram Shifting (HS), focused on two parameters: a scaling factor and the kernel size taken over the image in a stepwise fashion. The algorithm gave satisfying results when applied to different types of medical images such as radiotherapy simulator, nuclear medicine bone scan and MRI scan of the head (see also [29, 30]). Another paper was proposed by Hunter *et al.* [31], where in their work they focused on the detection of the center of the Left Ventricle (LV) cavity and it's boundaries in short axis mid-papillary muscle echocardiographic images. In order to cover a similar application to the previous work, Boudraa *et al.* [32] introduced an automated outlining technique of the left ventricular contour and its bounded area in gated isotopic ventriculography, where their main interest was determining an important parameter for measuring cardiac function, called the ejection fraction (EF). Their method was tested on 50 patients and got positive feedback from a team of trained clinicians (see also [33, 34]).

Van Leemput *et al.* [35] approached the medical field through tissue classification of magnetic resonance (MR) images of the brain. The method, which was an extension of their previous work [36], aimed to ameliorate the previous technique by incorporating contextual information during classification. The validation of their approach was performed on both simulated data and on real images (see also [37–40]).

Lelieveldt *et al.* [41] were able to establish a time-continuous segmentation of cardiac image sequences. They tested the approach on short-axis cardiac MRI with total of 1200 image frames from 25 subjects (15 normal subjects and 10 myocardial infarction patients) and four-chamber echocardiographic image sequences from 129 patients. Although their method performed slightly more accurately for MRI than for echocardiograms, it still generates a nearly perfect time-continuous segmentation results, which are consistent with cardiac dynamics (see also [42–45]).

Another work dedicated for the medical field was introduced by Bona *et al.* [46], with a data-driven comparison and registration of three-dimensional images called the "Volume-Matcher 3D" project. The authors used neural networks to register a source volume (SV) with a target volume (TV) when nonlinear deformations occur, to find a law that could model a growing intracranial lesion inside the brain. The method gave promising results when tested on different kinds of medical imaging.(see also [47–49]).

Xulei Qin *et al.* [50] proposed an extraction technique to automatically detect the cardiac myofiber orientations from high frequency ultrasound images. This method was tested on both phantom and pig hearts and showed satisfying results in both cases. Berikol *et al.* [51] also have contributed to the medical field by using support vector machine (SVM) to diagnose Acute coronary syndrome (ACS) with the purpose of helping the physicians with the decision of discharging or hospitalizing the patients. The data they used included sex, age, past history, ECG, CK-MB and troponin-I level, and echocardiographic data. The performance of their approach was compared with results obtained from three other methods (artificial neural network, Naïve Bayes, and Logistic Regression) and it shows SVM achieving the highest predicting accuracy among the other methods (99.13%). Zhang *et al.* [52] presented a fully automated computer vision pipeline for echocardiogram interpretation in clinical practice. Starting from automatically determining echocardiographic views, followed by the segmentation of cardiac chambers across 5 common views. Next, they proceed with measurements of cardiac structure and function, where they used the results obtained from the previous step to generated two automated measures of left ventricular (LV) function (ejection fraction, longitudinal strain). Finally, they build a model to detect hypertrophic cardiomyopathy, pulmonary arterial hypertension and cardiac amyloidosis. Moreover, they tested their approach on a dataset consists of more than 14000 echocardiograms.

Till this day data science is still finding its way to adjust with any changes that occurs in the medical field; e.g. since the beginning of Covid-19, data science has contributed in the detection of this disease in different ways using raw chest X-ray images (see [53–62]).

1.1.2 3D Reconstructions

Beside the already mentioned applications of data science in the medical field, three-dimensional (3D) reconstructions of the heart is quite a useful practice in medicine. In 1991, and starting from an early stage, Kuwahara and Eiho [63] took the challenge to build a 3D reconstruction of the heart using gated MRI method, which was able to provide several sets of cross-sectional images of a left ventricle and the whole heart during a cardiac cycle. The researchers used three pairs of magnetic resonance (MR) images, each pair taken across a different axis, to reconstruct a 3D image of the left ventricle. Salustri and Roelandt [64] took a step further by developing their own image acquisition technique in 1994, where they introduced a rotational imaging probe [65, 66] in order to achieve the 3D reconstruction of the heart. Using their acquisition technique, the researchers obtained cross-sectional images, which they re-sample and converted from polar to Cartesian coordinates. Next, an interpolation technique was used to fill the space between individual cross sections, followed by an enhancement step allowed them to display the three dimensional reconstruction of the heart. Miquel *et al.* [67] used a series of two-dimensional (2D) MR images (5 mm thick slices, 2–3 mm apart) to build the 3D reconstruction of the heart, using commercially available software package, visualization tool kit (Vtk) [68]. The authors used a semi-auto automated technique in every slice to segment the cardiac chambers and vessels independently, then 3D surface was extracted from the segmented dataset providing the 3D reconstruction, using the early mentioned software. S. Jacobs *et al.* [69] created an anatomical correct 3D rapid prototyping model (RPT) for patients with complex heart disease. Patient's datasets were obtained by segmenting computer tomography (CT) and magnetic resonance imaging (MRI) images, which provides the target volume and structures at risk. The proposed approach successfully facilitates surgical procedures due to better planning and improved orientation. Banerjee *et al.* [70] introduced an automated pipeline for generating patient-specific 3D heart models from cine magnetic resonance (MR) slices. Following the automatic selection of the MR images, the developed technique perform a deep learning-based segmentation step to extract the heart contours. Next, the contours are aligned in 3D space while correcting any possible misalignment, producing the 3D reconstruction. The technique was evaluated on 20 healthy subjects, showing an average reduction of misalignment artefacts, which contributes to the efficient realization of precision medicine, enabling the enhanced interpretability of clinical data (see also [71–73]).

1.1.3 Information recovery

Generally, 3D reconstructions of the heart require cardiac cine MRI datasets, but the resolution of these images can be unsatisfying. In some occasions MR images can be adversely affected by numerous imaging artifacts, which can cause losing information and corrupted images (e.g. inter-slice motion artifacts can be caused by cardiac and respiratory motion, together with fast flowing blood). Similarly, local loss of signal in particular slices can be caused by adjacent tissues with different properties or implants [74, 75]. In order to solve these concerns, researchers have introduced various and different tools. Interpolation in particular has been the base of several information recovery techniques. Grevera *et al.* [76] provided an objective comparison between 8 slice interpolation techniques, using datasets from different medical applications, different body parts, different modalities, and different patients. Leng *et al.* [77] proposed a registration-based image interpolation technique. The proposed method is divided into two steps, a registration process, which is carried out to construct a corresponding transformation between the slices, followed by interpolating intensity values along the matching lines to construct additional images between each two neighboring slices. The authors tested their method on different types of images including two series of medical images: 1) an MR brain sequence, where they used their method to produce 3 additional inter slices between four sub-sequences of slices (with 40 slices per sub-sequence). 2) Computed Tomography (CT) chest sequence, where they removed all the odd slices and reconstructed them using interpolation between the even slices. Ehrhardt *et al.* [78] derived an interpolation scheme from the optical flow equation, mainly to generate images at predefined phases of the cardiac cycle in cine MRI sequences of patients with myocardial infarction. The introduced method consisted of two main steps: first, a non-linear registration algorithm is used to determine the optical flow between the temporal images. Next, the calculated optical flow field is is used to generate the interpolated images at the desired time. The results obtained were compared with previous techniques and it showed that the presented method outperformed several interpolation techniques. Horváth *et al.* [79] introduced a high order slice interpolation method, which employs both object and intensity interpolation to interpolate a whole stack of images, solving the problem of combining higher order interpolations of structure motion and intensity. The proposed technique was tested on the human spinal cord along the neck captured with a slice-wise inversion recovery MR sequence, where they used a leave-one-slice-out test to evaluate how well the left-out slices can be interpolated. Lin *et al.* [80] proposed a new algorithm for slice interpolation in MRI called the decomposition-reconstruction (D-R) method. As its name indicates, the method first decomposes the information contained in MRI and then reconstruct them based on the rule of organ consistency. The inter-slice interpolation algorithm was applied to a 3 neighboring MRI of human throat and its performance was compared to three other interpolation techniques (Linear interpolation, Spline interpolation and Cubic interpolation). The results showed that the D-R method is more suitable for MRI interpolation (see also [81, 82]).

1.2 Interests and objectives

In this work we try to address some of the challenges identified in section 1.1 using different modal decomposition based ROMs. In particular, we introduce novel approaches for the wide, complicated field of medicine. Several data-driven modal decomposition tools are employed, including the SVD and HOSVD. Nevertheless, the fluid dynamic tool, the higher order dynamic mode decomposition (HODMD) [83] is the technique employed the most. This technique, which was developed for the analysis of complex data modeling non-linear dynamical systems, has sound mathematical foundation, but can be formulated as a fully data-driven technique. The HODMD technique has proved its efficiency in identifying complex dynamics with outstanding denoising properties, as well as providing robust and accurate results even when dealing with limited amounts of data. The remarkable performance of the HODMD algorithm has been proven in different applications, such as dealing with turbulent flows [84], for the study of cross-flow instabilities [85] and even for bio-inspired propulsion and reduced order modeling [86] (see also [87–90]).

In our work, we have explore the use of the HODMD and two other modal decomposition techniques to build hybrid ROMs, capable of achieving a variety of applications: i) the investigation of the relevant frequencies and the identification of the features and patterns related to several cardiac conditions in two different medical datasets. ii) Building a reduced order model capable of providing new databases, allowing us to achieve a 3D reconstruction of the heart lasting for more than one cardiac cycle. iii) Finally, the HODMD technique is paired with an SVD based interpolation technique, producing an information recovery method. This novel technique is used to reconstruct all the temporal snapshots of a missing slice.

Applying our proposed methods to both datasets gave the following results: i) for the first dataset, the proposed method needed only the maximum of 100 snapshots to analyze the signal and capture two lines of frequencies: an upper branch of frequency representing the heart rate and a lower branch of frequency representing the respiratory rate. Furthermore, the HODMD algorithm provided the dominant features for each cardiac disease through the constructed DMD modes. The resulted modes can be divided into two sets, modes belong to the upper branch of frequencies and exhibit the heart area in the echoradiography image, and modes that belong to the lower branch of frequencies and exhibit the area of the lungs in the echocardiography image. All the datasets from each mentioned cardiac disease were analyzed to demonstrate the robustness of the proposed method, and as it was anticipated, the algorithm still captures the two lines of frequencies representing the heart rate and respiratory rate, and provides similar modes that represent the dominant patterns for each cardiac disease. ii) As for the second dataset, similarly to the previous case, regardless the very limited amount of data (20 snapshots), the HODMD was able to capture the relevant frequencies representing the heart rate for each slice. Furthermore, the HODMD is used as a ROM, which permits us to extend the original database that is formed by a small number of snapshots (20 snapshot per slice) to generate new databases consists of 100 snapshots per slice, providing a total number of 1000 snapshot per dataset. These new generated databases are used to build a 3D reconstruction of the heart covering additional cardiac cycles. Finally, a novel technique, combining the HODMD algorithm with the interpolation technique is used to recover the information of a whole slice, using the neighboring two slices. The new technique was able to reconstruct all the snapshots of a corrupted slice.

1.3 Scientific Dissemination

The present thesis has made various publications, including first author and collaboration publications. In particular, three first-authored publications, two in the journal of Computers in Biology and Medicine (CIBM) and one in the Results in Engineering journal, and a second-authored publication in the journal of Physics of Fluids (PoF), all peer-reviewed Q1 journals that meets the requirement of Universidad Politécnica de Madrid.

1.3.1 Accepted journal papers

- Groun, N., Villalba-Orero, M., Lara-Pezzi, E., Valero, E., Garicano-Mena, J. and Le Clainche, S., 2022. Higher order dynamic mode decomposition: From fluid dynamics to heart disease analysis. Computers in Biology and Medicine, 144, p.105384.
- Groun, N., Villalba-Orero, M., Lara-Pezzi, E., Valero, E., Garicano-Mena, J. and Le Clainche, S., 2022. A novel data-driven method for the analysis and reconstruction of cardiac cine MRI. Computers in Biology and Medicine, 151, p.106317.
- Groun, N., Begiashvili, B., Valero, E., Garicano-Mena, J., Le Clainche, S. (2023). Higher Order Dynamic Mode Decomposition Beyond Aerospace Engineering. Results in Engineering, 101471.
- Begiashvili, B., Groun, N., Garicano-Mena, J., Le Clainche, S. and Valero, E., 2023. Data-driven modal decomposition methods as feature detection techniques for flow problems: A critical assessment. Physics of Fluids, 35(4).

1.3.2 Conference papers

- Groun, N., Begiashvili, B., Valero, E., Garicano-Mena, J. and Le Clainche, S., 2022, May. Data-Driven Methods Beyond Aerospace Field. In International Symposium on Unmanned Systems and The Defense Industry (pp. 105-110). Cham: Springer International Publishing.

1.3.3 Publication under review and work on progress

- Groun, N., Villalba-Orero, M., Casado-Mart, L., Lara-Pezzi,E., Valero, E., Garicano-Mena, J., and Le Clainche, S. . 2023. Data-Driven Modal Decomposition techniques: Effective Tool For Image Classification Improvement. Artificial Intelligence in Medicine (Under review).
- Patent request: Método implementado por ordenador de aumento de la eficiencia en la clasificación de imágenes (under review).

1.3.4 Participation in conferences

- ISUDEF 2022 (International Symposium on Unmanned Systems and the Defense Industry), 30/05/2022 - 01/06/2022, Online.
- Modelling the Cardiac Function - MCF2022 iHEART Conference, 30/09/2022 - 02/10/2022, Italy-Cetraro.
- Spanish Fluid Mechanics Conference (SFMC2023), 02-05/07/2023, Barcelona - Spain.

1.4 Outline of the Thesis

In the following, a short description of the structure of this thesis is given.

This first part (I) has provided a general introduction to our work, accompanied by the state of the art, the interests and objectives of the research and finally, the scientific dissemination.

Part II contains two chapters.
Chapter 2 introduces a number of preliminary concepts. These preliminary concepts are leveraged to better describe the methodology employed in this work, which is included in chapter 3.

Part III contains three chapters: chapter 4, and it gives a detailed description of the datasets used in this research. Meanwhile chapters 5 and 6 provides integrated analysis of the results.

Part IV, consisting only of chapter 7, includes the reached conclusions 7.1 and possibilities of future works 7.2.

Part V, which is the last part, assembles the appendices.

Part II

Methodology

Summary This part contain two chapters. Chapter 2 provides a general overview of the main techniques leveraged in this thesis. This chapter deals with the topics of modal decomposition techniques and interpolation basics. The precursory view of these techniques helps to better structure the tools used in the thesis, namely the HODMD and the SVD based interpolation technique, introduced in chapter 3.

Chapter 2

Preliminaries

2.1 Modal decomposition techniques

Data-driven modal decomposition methods mainly focus on the analysis and interpretation of data. The data can be statistical data (gender, age, wight, height...) or data produced by complex systems. Moreover, time-resolved data are considered one of the richest information sources. This type of data considers taking measurements of experiments at a uniform sampling rate, resulting in large data matrices. Alternatively, the data can also be images (snapshots) acquired in different, equidistant time instants. In this last case, the data can be organized in one of these two following forms:

- In matrix form (Fig. (2.1)) as:

$$\boldsymbol{V}_1^K = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_K], \tag{2.1}$$

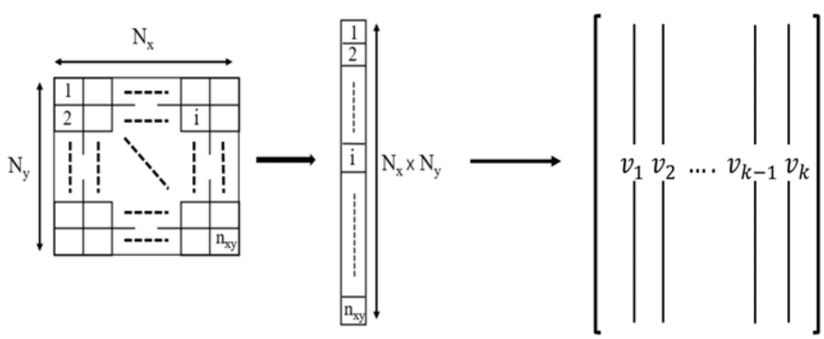


Figure 2.1: Arrangement of data in matrix from.

where $\boldsymbol{v}_k$ is a reshaped snapshot collected at time t_k, with $k = 1, \ldots, K$. In particular, each column is a vector that contains the pixels of the image analyzed. Hence $\boldsymbol{V}_1^K \in \mathbb{R}^{J\times K}$, with $J = \mathrm{N_x}$ (number of pixels in X) $\times \mathrm{N_y}$ (number of pixels in Y) and K is the total number of snapshots.

- In tensor form (Fig. (2.2)) as:

$$\boldsymbol{T}_{i_1,i_2,k} \quad \text{for} \quad i_1 = 1,\dots,I_1; i_2 = 1,\dots,I_2 \quad \text{and} \quad k = 1,\dots,K, \tag{2.2}$$

  where i_1 and i_2 the indexes representing the position of each pixel in the plane of the image, and K is again the total number of snapshots.

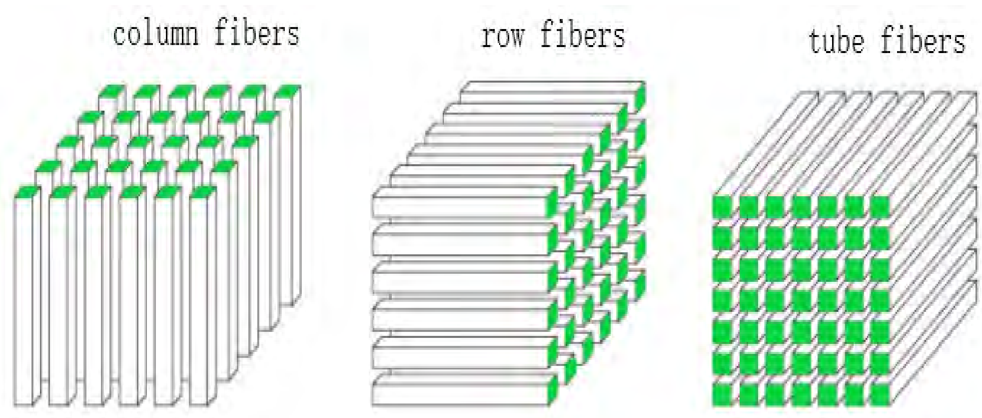


Figure 2.2: Visualization of a third order tensor and its fibers.

Before diving into the specifics of the different tools, we recall that most data-driven modal analysis techniques can be interpreted as matrix factorization strategies ([91, 92]) as presented in Eq. (1) and Fig. (1), making these techniques highly related. Hence, in the following we will introduce some of these tools, stressing the connection between them.

2.1.1 Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD) is a matrix decomposition tool and it is considered as one of the most important algorithms in the past decades. This tool provides a foundation for most of the data-driven methods that are currently being used. The SVD method was discovered by Eugenio Beltrami and Camille Jordan [93] over a 100 years ago, even thought other researchers like James Joseph Sylvester, Erhard Schmidt and Hermann Weyl also rederived the SVD using their own methods [93]. However, the researchers that turned the field of numerical linear algebra upside down with their method are Golub and Reinsch when they developed the SVD algorithm in 1970 [94].

Mathematically, the SVD allows us to represent a snapshot matrix (as in eq. (2.1)) as a product of three other matrix factor as follows:

$$\boldsymbol{V}_1^K = \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{T}^T = \sum_{j=1}^{r=min(J,K)} \sigma_j \boldsymbol{w}_j \boldsymbol{t}_j^T, \tag{2.3}$$

where $\boldsymbol{W} \in \mathbb{C}^{J\times J}$ and $\boldsymbol{T} \in \mathbb{C}^{K\times K}$ are real, orthonormal matrices. The columns of $\boldsymbol{W}$ (noted $\boldsymbol{w}_j$) are the left singular vectors of $\boldsymbol{V}_1^K$ (codify the spatial aspects of the data), and the columns of $\boldsymbol{T}$ (noted $\boldsymbol{t}_j$) are the right singular vectors of $\boldsymbol{V}_1^K$ (codify the temporal aspects of the data). The matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{J\times K}$ is a matrix with real, non-negative entries on the diagonal and zeros off the diagonal. The elements of $\boldsymbol{\Sigma}$ (noted σ_j) are the singular values, with $\sigma_1 \geq \cdots \geq \sigma_j \geq \cdots \geq \sigma_r \geq 0$.

Furthermore, the SVD guaranties an optimal, low rank representation of the dataset using a rank of approximation $r' \leq r$, allowing eq. (2.3) to be be rewritten as

$$\boldsymbol{V}_1^K \simeq \sum_{j=1}^{r'} \sigma_j \boldsymbol{w}_j \boldsymbol{t}_j^T. \tag{2.4}$$

Note that the parameter r' can be chosen either directly (manually by specifying an integer value) or indirectly specifying a given tolerance ε_{SVD} as

$$\frac{\|\boldsymbol{V}_1^K - \sum_{j=1}^{r'} \sigma_j \boldsymbol{w}_j \boldsymbol{t}_j^T\|_F}{\sqrt{\sum_{j=1}^{r} \sigma_j^2}} \leq \varepsilon_{SVD}, \tag{2.5}$$

where $\|.\|_F$ is the Frobenius norm.

2.1.2 Higher order singular value decomposition (HOSVD)

Through the years, there have been many attempts to extend the application of the SVD to higher than two order tensors. However, some of the most known extensions (e.g., *the canonical decomposition*) face a major challenge: the determination of the *rank*. And this is because the ranks of the various fibers of a *tensor* do not coincide in general if the order of the *tensor* is higher than 2. Hence, in order to address this obstacle, other sophisticated techniques have been developed. The *higher order singular value decomposition* ([94],[93]) is one of these techniques. This tool is a robust extension of the SVD, originally introduced by Tucker [95] in 1966; this technique resurfaced and became widespread in the early 2000's thanks to Lathauwer *et al.* [96].

Similarly to SVD, HOSVD has many versions; in this contribution we focus on *economy* HOSVD (illustrated in Fig. (2.3)).

For a given Nth order *tensor* $\boldsymbol{T}$ of size $I_1 \times I_2 \times \cdots \times I_N$, with $I_1 > 1, I_2 > 1, \ldots, I_N > 1$ *economy* HOSVD decomposes the *tensor* $\boldsymbol{T}$ as follows:

$$\boldsymbol{T}_{i_1,i_2,\ldots,i_N} = \sum_{n_1=1}^{r'_1} \sum_{n_2=1}^{r'_2} \cdots \sum_{n_N=1}^{r'_N} \boldsymbol{\Sigma}_{n_1 n_2 \ldots n_N} \boldsymbol{W}^1_{i_1 n_1} \boldsymbol{W}^2_{i_2 n_2} \ldots \boldsymbol{W}^N_{i_N n_N}, \tag{2.6}$$

such that, $i_1 = 1, 2 \ldots, I_1$; $i_2 = 1, 2, \ldots, I_2$; ... ; $i_n = 1, 2, \ldots I_N$. Denoting $r_1, r_2, \ldots\ r_N$ the ranks of the fibers of the tensor along the different dimensions of the *tensor*, thus $r'_1, \ldots r'_N$

are such that $r'_1 \geq max\{r_1, I_1\}, \ldots, r'_N \geq max\{r_N, I_N\}$. The elements of the matrices $\boldsymbol{W}^1, \boldsymbol{W}^2, \ldots, \boldsymbol{W}^N$ of sizes $I_1 \times r'_1$; $I_2 \times r'_2$; …; $I_N \times r'_N$, respectively, are called *mode matrices* and $\boldsymbol{\Sigma}_{n_1 n_2 \ldots n_N}$ are the components of a Nth order tensor $\boldsymbol{\Sigma}$, of size $r'_1 \times r'_2 \times \cdots \times r'_N$ known as the *core tensor*. Furthermore, since the right side of the expression (2.6) is a *tensor product* of the core tensor and the *mode matrices*, it could be written as:

$$\boldsymbol{T} = \mathbf{tprod}(\boldsymbol{\Sigma}, \boldsymbol{W}^1, \boldsymbol{W}^2, \ldots, \boldsymbol{W}^N), \tag{2.7}$$

where we employ here the notation **tprod** used in library [97].

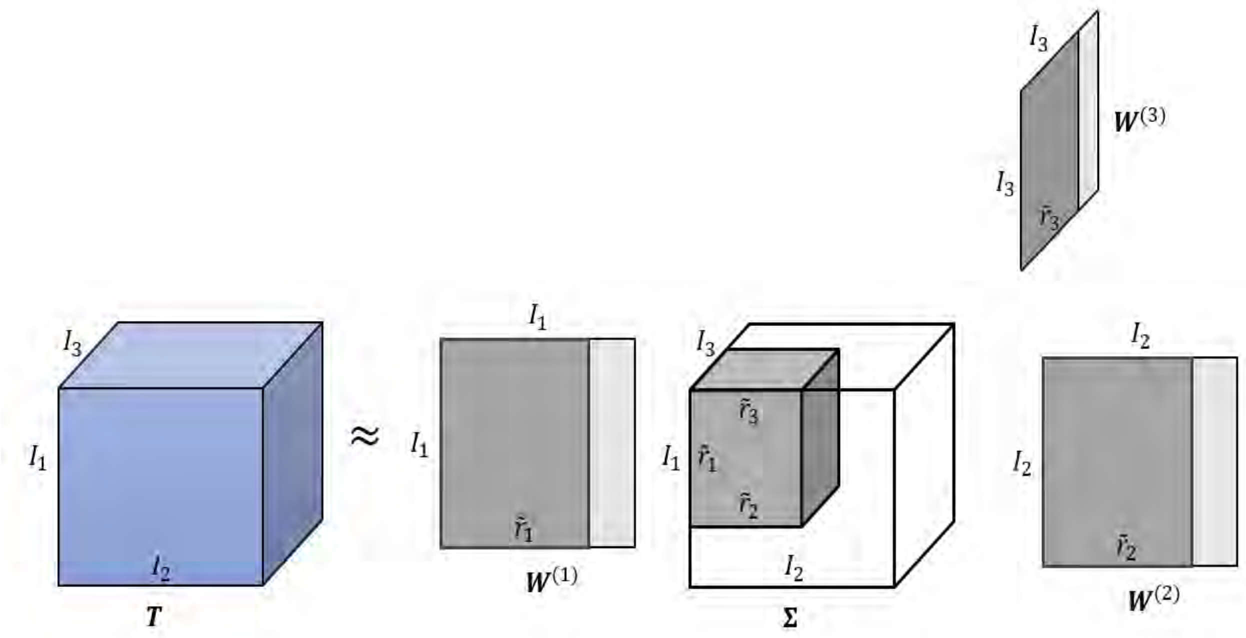


Figure 2.3: Visualization of truncated-HOSVD for a third-order tensor $\boldsymbol{T}$.

In more mathematical details, *economy* HOSVD is calculated as follows:

i) The different dimensions of the tensor $\boldsymbol{T}$ are unfolded into the matrices $\boldsymbol{D}^1, \ldots, \boldsymbol{D}^N$ (whose columns are the fibers of $\boldsymbol{T}$ along the nth dimension of the tensor).
ii) Economy SVD is applied to these matrices $\boldsymbol{D}^n$ (whose rank is r_n) as:

$$\boldsymbol{D}^n = \boldsymbol{W}^n \boldsymbol{\Sigma}^n (\boldsymbol{T}^n)^T \text{ for } n = 1, \ldots, N. \tag{2.8}$$

iii) Once the matrices $\boldsymbol{W}^n$ -which represent the mode matrix along the nth dimension- have been calculated, the core tensor can be computed using **tprod** as

$$\boldsymbol{\Sigma} = \mathbf{tprod}(\boldsymbol{T}, (\boldsymbol{W}^1)^T, (\boldsymbol{W}^2)^T, \ldots, (\boldsymbol{W}^N)^T). \tag{2.9}$$

Note that the the elements of the diagonal matrix $\mathbf{\Sigma}^n$ are the HOSVD singular values along the *nth* dimension of the tensor and they will be denoted as

$$\sigma_1^n \geq \cdots \geq \sigma_{r_n-1}^n \geq \sigma_{r_n}^n > 0. \tag{2.10}$$

iv) In order to complete the HOSVD decomposition, we consider the truncation values $\tilde{r_1}, \tilde{r_2}, \ldots, \tilde{r_N}$, such that $\tilde{r_1} \leq r_1 \leq \tilde{r_2} \leq r_2 \leq \cdots \leq \tilde{r_N} \leq r_N$. Hence, the Forbenius norm of the error of the approximation is given as

$$\|\boldsymbol{T} - \boldsymbol{T}^{approx}\|_F \leq \sqrt{\sum_{n=1}^{N} [(\sigma_{\tilde{r}_n+1}^n)^2 + \cdots + (\sigma_{r_n}^n)^2]}. \tag{2.11}$$

Now, similarly to the SVD, we impose the following in order to make sure that the relative RMS error of the truncated approximation is smaller than a given threshold ε_{HOSVD} using:

$$\mathbf{RRMSE} = \frac{\|\boldsymbol{T} - \boldsymbol{T}^{approx}\|}{\|T\|} = \sqrt{\sum_{n=1}^{N} \frac{(\sigma_{\tilde{r}_n+1}^n)^2 + \cdots + (\sigma_{r_n}^n)^2}{(\sigma_1^n)^2 + \cdots + (\sigma_{r_n}^n)^2}} \leq \varepsilon_{HOSVD}. \tag{2.12}$$

Further improvement to this approximation can be achieved by equi-distributing the upper bound in the left hand side of eq. (2.12) along the various dimensions of the tensor. This can be accomplished by choosing $\tilde{r_1}, \ldots, \tilde{r_n}$ as the smallest indices such that

$$\sqrt{\frac{(\sigma_{\tilde{r}_n+1}^n)^2 + \cdots + (\sigma_{r_n}^n)^2}{(\sigma_1^n)^2 + \cdots + (\sigma_{r_n}^n)^2}} \leq \frac{\varepsilon_{HOSVD}}{\sqrt{N}}, \quad \text{for } n = 1, \ldots, N, \tag{2.13}$$

which in cases where the singular values decay very fast starting from $\sigma_{\tilde{r}_n+1}^n$, can be rewritten

$$\frac{\sigma_{\tilde{r}_n+1}^n}{\sigma_1^n} \leq \frac{\varepsilon_{HOSVD}}{\sqrt{N}}. \tag{2.14}$$

2.1.3 Principal Component Analysis (PCA)

Principal component analysis (PCA) [98, 99] is one of the central applications of the SVD. PCA pre-processes the data by mean subtraction and setting the variance to unity before performing the SVD. This technique results in a coordinate system transformation, which is determined by principal components (PCs).

Consistently with PCA literature, measurements are generally arranged into row vectors. The PCA algorithm analyses the data matrix and tries to find "components" that capture the maximal variance within the data. Hence, applying PCA on a data matrix $\boldsymbol{V} \in \mathbb{R}^{n \times p}$ yields:

$$\boldsymbol{V} = \boldsymbol{W}\boldsymbol{C}^T,$$

where $\boldsymbol{C} \in \mathbb{R}^{r \times p}$ contains the top r principal components and $\boldsymbol{W} \in \mathbb{R}^{n \times r}$ contains the loading weights (see also Fig. 2.4).

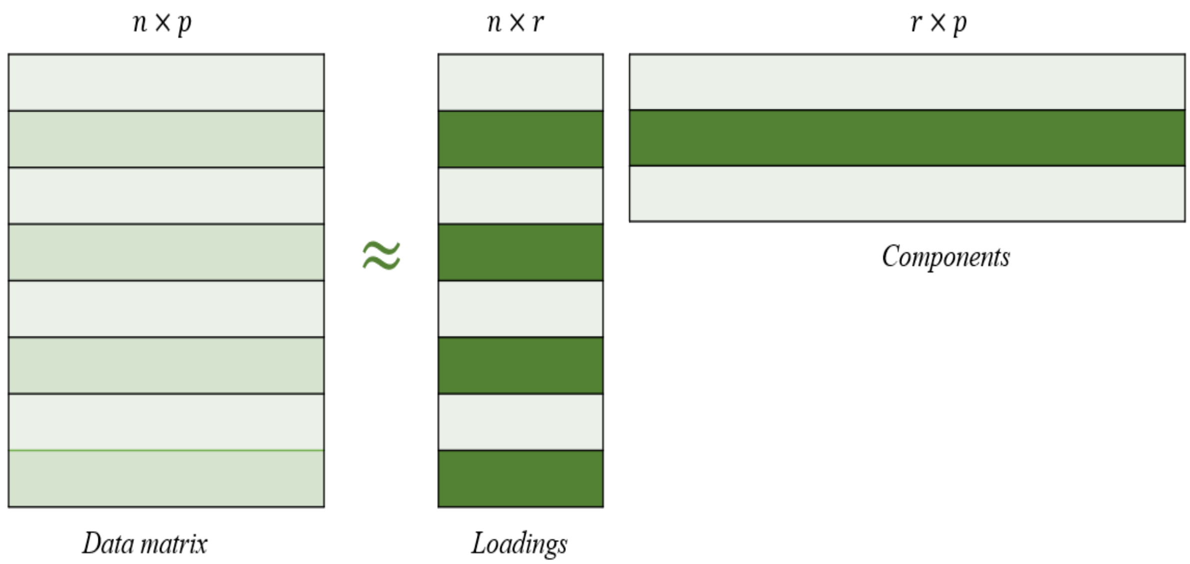


Figure 2.4: Diagram for Principal Component Analysis (PCA).

Mathematically, and as mentioned earlier, in this case the data matrix $\boldsymbol{V}$ is organized in rows instead of columns:

$$\boldsymbol{V} = \begin{bmatrix} — & \boldsymbol{v_1} & — \\ — & \boldsymbol{v_2} & — \\ — & \vdots & — \\ — & \boldsymbol{v_n} & — \end{bmatrix},$$

where each row holds the information of a certain experiment.

i) The first step in the PCA algorithm is to compute the average row (row-wise mean):

$$\overline{\boldsymbol{v}} = \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{v}_j,$$

and build an average matrix:

$$\overline{\boldsymbol{V}} = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \\ \vdots \\ \mathbf{1} \end{bmatrix} \begin{bmatrix} — & \overline{\boldsymbol{v}} & — \end{bmatrix}.$$

ii) The second step is to subtract the mean to obtain what we call the mean centered data:

$$\boldsymbol{\mathcal{B}} = \boldsymbol{V} - \overline{\boldsymbol{V}}.$$

iii) Next, we compute the cross-correlation matrix of the the matrix $\boldsymbol{\mathcal{B}}$:

$$\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{B}}^T \boldsymbol{\mathcal{B}}.$$

Hence, computing the eigen-decomposition of the cross-correlation matrix $\boldsymbol{\mathcal{C}}$, which yields:

$$\boldsymbol{\mathcal{C}} \boldsymbol{\Lambda} = \boldsymbol{\Lambda} \boldsymbol{\mathcal{D}},$$

will allow us to obtain the principal components $\boldsymbol{\mathcal{T}}$ as:

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{B}} \boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda}$ are the eigenvectors of the cross-correlation matrix.

2.1.4 Proper Orthogonal Decomposition (POD)

Proper Orthogonal Decomposition (POD) is a powerful technique commonly used for dimensionality reduction. The POD was originally developed by Pearson in 1901 [100] and is often formulated by taking the SVD of the data matrix as shown in eq. (2.3). This method provides a hierarchical decomposition of data into an orthogonal basis of spatially correlated modes.

The method of *snapshots* was later introduced by Sirovich in 1987 [101] and it considers the correlation matrix $(\boldsymbol{V}_1^K)^T \boldsymbol{V}_1^K$. The eigendecomposition of this matrix is related to the SVD:

$$(\boldsymbol{V}_1^K)^T \boldsymbol{V}_1^K = \boldsymbol{T} \boldsymbol{\Sigma} \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{\Sigma} \boldsymbol{T}^T = \boldsymbol{T} \boldsymbol{\Sigma}^2 \boldsymbol{T}^T, \tag{2.15}$$

hence

$$(\boldsymbol{V}_1^K)^T \boldsymbol{V}_1^K T = \boldsymbol{T} \boldsymbol{\Sigma}^2, \tag{2.16}$$

Once $\boldsymbol{T}$ and $\boldsymbol{\Sigma}$ are computed, the columns of $\boldsymbol{W}$ (i.e., POD modes) can be reconstructed by

$$\boldsymbol{W} = \boldsymbol{V}_1^K \boldsymbol{T} \boldsymbol{\Sigma}^{-1} \tag{2.17}$$

It is habitual to subtract the mean of $\boldsymbol{V}_1^K$ before computing POD modes, in which case POD is equivalent to PCA. Keeping in mind that mean-subtraction does not have an effect on the basic calculations, however, it can affect the interpretation of the results [102], especially in cases where capturing the mean flow is essential [103].

2.1.5 Dynamic Mode Decomposition (DMD)

Dynamic mode decomposition (DMD) is a data driven technique originated in the fluid dynamics field. This tool was developed by Peter Schmid [91], generally used to obtain linear reduced order models for high dimensional complex systems, as well as to extract dominant patterns from experimental data. Additionally, the DMD creates a decomposition in space and time in which DMD modes describe dominant spatial structure (as seen in Fig.(2.5)).

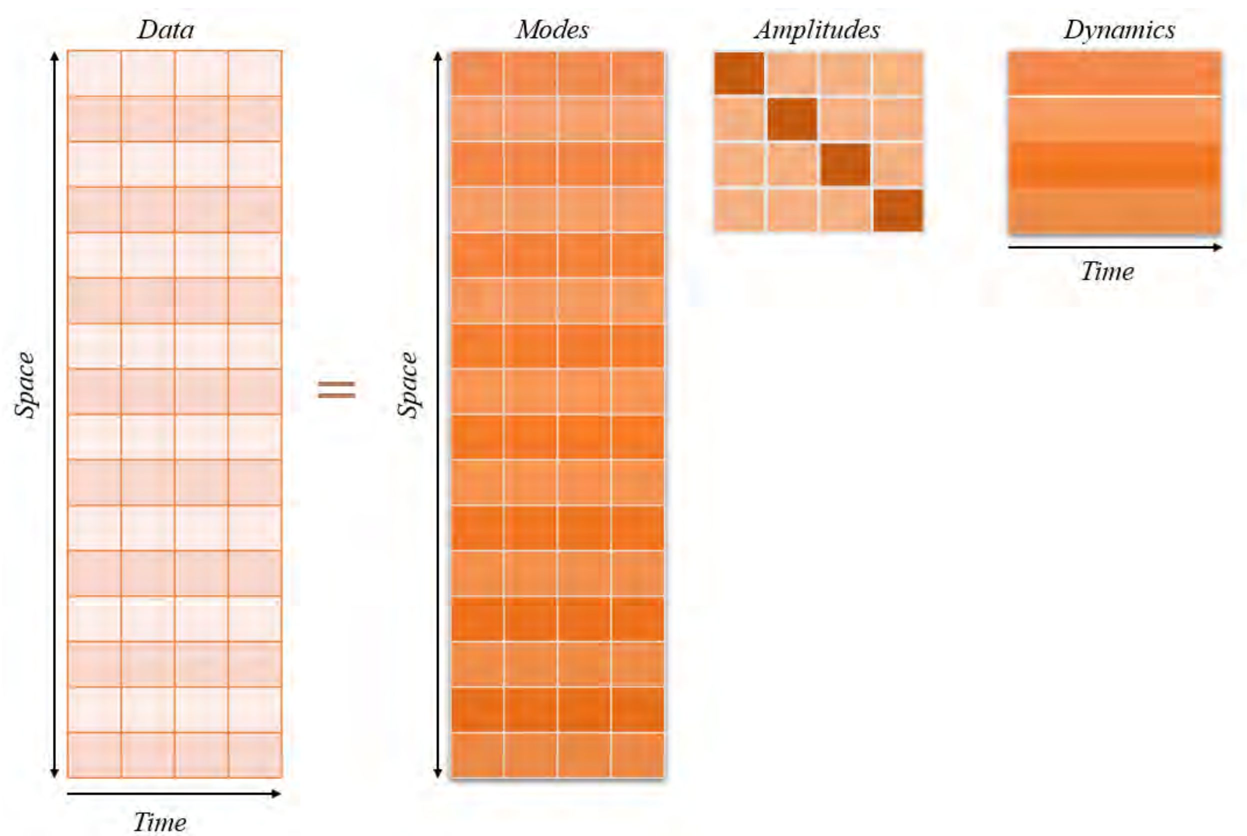


Figure 2.5: Illustration of the dynamic mode decomposition.

Mathematically, while working with DMD the data used in the process is acquired from a dynamical system:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}, t; \mu), \tag{2.18}$$

where $\boldsymbol{f}$ is the dynamics, $\boldsymbol{x}$ is the state of the dynamical system and μ contains parameters of the system and t is of course the time.

i) In order to apply DMD we take measurements of the state $\boldsymbol{x}$ at different time instants t_k such that $k = 1, 2, \ldots, K$, with a uniform sampling time Δt. These measurements are called *snapshots*, which will be arranged into two large matrices:

$$\boldsymbol{V}_1^{K-1} = \begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_{K-1} \\ | & | & & | \end{bmatrix}, \qquad \boldsymbol{V}_2^{K} = \begin{bmatrix} | & | & & | \\ \boldsymbol{v}_2 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_{K} \\ | & | & & | \end{bmatrix}.$$

The DMD operates under the assumption that a linear relationship exists between $\boldsymbol{V}_1^{K-1}$ and $\boldsymbol{V}_2^{K}$ as:

$$\boldsymbol{V}_2^{K} = \boldsymbol{A}\boldsymbol{V}_1^{K-1}. \tag{2.19}$$

Hence, the DMD revolves about finding the best linear fit operator $\boldsymbol{A}$ such that Eq. (2.19) is realized.

ii) The linear operator $\boldsymbol{A} \in \mathbb{R}^{n_p \times n_p}$ is unknown and very large ($n_p >> 1$). As a consequence, the SVD factorization of the data matrix $\boldsymbol{V}_1^{K-1}$ is employed to build a reduced representation of $\boldsymbol{A}$.

Nevertheless, for a rank of approximation denoted by r, the SVD factorization of $\boldsymbol{V}_1^{K-1}$ yield:

$$\boldsymbol{V}_1^{K-1} \approx \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{T}^*,$$

where $\boldsymbol{W} \in \mathbb{C}^{J\times r}$ (with the left singular vectors of $\boldsymbol{W}$ are POD modes), $\boldsymbol{\Sigma} \in \mathbb{C}^{r\times r}$ and $\boldsymbol{T} \in \mathbb{C}^{K-1\times r}$.

Employing the pseudo-inverse of $\boldsymbol{V}_1^{K-1}$ we obtain the matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{V}_2^{K}\boldsymbol{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^*,$$

which permits us to compute a low-dimensional linear model $\tilde{\boldsymbol{A}}$ of the dynamical system by projecting the full matrix $\boldsymbol{A}$ onto POD modes:

$$\tilde{\boldsymbol{A}} = \boldsymbol{W}^*\boldsymbol{A}\boldsymbol{W} = \boldsymbol{W}^*\boldsymbol{V}_2^{K}\boldsymbol{T}\boldsymbol{\Sigma}^{\boldsymbol{-1}}.$$

iii) Next, we calculate the eigendecomposition of $\tilde{\boldsymbol{A}}$:

$$\tilde{\boldsymbol{A}}\boldsymbol{\Psi} = \boldsymbol{\Psi}\boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda}$ is a matrix containing the eigenvalues on its diagonal and $\boldsymbol{\Psi}$ has the eigenvectors as it's columns.
iv) Finally, using the eigenvalues and eigenvectors of the matrix $\tilde{\boldsymbol{A}}$ we reconstruct the eigendecomposition of the matrix $\boldsymbol{A}$ such that:

- The eigenvalues of $\boldsymbol{A}$ are given by $\boldsymbol{\Lambda}$.
- The eigenvectors of $\boldsymbol{A}$ are given by the columns of $\quad \boldsymbol{\Phi} = \boldsymbol{V}_2^{K}\boldsymbol{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi} \in \mathbb{C}^{n_p\times n_p}$.

2.1.6 The Koopman operator

The *Koopman* operator is a linear, infinite-dimensional operator that was introduced by B. Koopman in 1931 [104]. This operator represents the action of a nonlinear dynamical system on the Hilbert space of measurement functions of the state.

Mathematically, for a considered the dynamical system as introduced in Eq. (2.18) and a measurement function of the state $\boldsymbol{x}$, $g : \mathcal{D} \longrightarrow \mathbb{C}$ (with $\mathcal{D}$ is an n-dimensional manifold), the action of Koopman operator $\boldsymbol{K}$ is equal to the composition of the function g with the the nonlinear evolution:

$$\boldsymbol{K}g = g \circ \boldsymbol{f} \Rightarrow \boldsymbol{K}g(\boldsymbol{x}_k) = g(\boldsymbol{f}(\boldsymbol{x}_k)) = g(\boldsymbol{x}_{k+1}).$$

That is, the Koopman operator advances the measurements one time step into the future and remeasures the system at that next time step.

This operator is strongly connected to DMD, as it was proven by Rowley and Mezić *et al.* [105] that under certain conditions (such as the special case of linear systems where the observable is the full flow state), the DMD provides a finite-dimensional approximation to eigenvalues and eigenvectors of the infinite-dimensional Koopman operator.

In more details, we first consider a set of p observables:

$$g_n : \mathcal{D} \longrightarrow \mathbb{C} \text{ with } n = 1 \dots p,$$

where we denote $\mathbf{g} = [g_1\, g_2\, \dots\, g_p]^T$ the column vector of observables. Now, by considering a set of initial conditions $\{x_1\, x_2\, \dots\, x_{k-1}\}$ for the considered dynamical system, we can construct data matrices $\boldsymbol{Y}$ and $\boldsymbol{Y'}$, such that, $y_k = \mathbf{g}(\boldsymbol{x}_k)$ gives the columns of the matrix $\boldsymbol{Y}$ and the columns of $\boldsymbol{Y'}$ are obtained by evolving (2.18) forward in time and viewing the output vector through the observables, denoted by $\boldsymbol{y'} = \mathbf{g}(\boldsymbol{f}(\boldsymbol{x}_k))$. Hence, the desired Koopman approximation is given by the resulting DMD algorithm on the data of observables.

2.2 Interpolation Basics

Mathematically, in particular in numerical analysis, interpolation is a type of estimation, which permits the constructing of new data points based on the range of a discrete set of known data points. Interpolation tools are useful methods in data science, as they are employed to determine or estimate the hypothetical values for an unknown variable based on the observation of other data points obtained from experiments.

Linear interpolation provide one of the most basic and general interpolation formula. It is believed that linear interpolation goes back to the last three centuries B.C, and even a description of linear interpolation can be found in the ancient Chinese mathematical text called "The Nine Chapters on the Mathematical Art" [106]. The main formula of linear interpolation is given as follows:

$$y = y_1 + \frac{x - x_1}{x_2 - x_1}(y_2 - y_1),$$

where (x_1, y_1) and (x_2, y_2) are two known data points and (x, y) represents the data point to be estimated.
Various further interpolation techniques have been introduced over the years, such as bilinear interpolation [107], Kriging [108, 109], bicubic interpolation [110] etc ....
In the following we give a general overview of the interpolation technique used in this contribution, named spline interpolation.

2.2.1 Spline interpolation

Spline interpolation [111, 112] is a type of interpolation where the interpolant is a special form of piece-wise polynomial. In particular, this tool fits low-degree polynomials to small subsets of the values instead of fitting one single, high-degree polynomial to all of the values at once. Spline interpolation is considered on of the most useful interpolation techniques, as its interpolation error can be made small even when using low-degree polynomials [113].

The mathematical representation of this tool is as follows:

We first introduce a piece-wise polynomial (defined by multiple sub-polynomials, where each sub-polynomial applies to a different sub-interval in the domain), such that, for an interval $[a, b]$, which can be segmented into sub-intervals $[x_i, x_{i+1}]$, where $i = 0, \dots, n-1$, $x_0 = a$ and $x_n = b$, We define the piece-wise polynomial order $p(x)$ on $[a, b]$ as

$$p(x) = p_i(x), \;\; x_{i-1} \leq x \leq x_i, \;\; i = 1, \dots, n.$$

Each polynomial $p_i(x)$ is a function defined on $[x_{i-1}, x_i]$, with $i = 1 \dots, n$.

Hence, for giving data points $x_0, \dots, x_n$ and for $p \in C[a, b]$, the spline interpolation $s(x)$ at these points is defined as

$$s(x) = p(x_{i-1})\frac{x - x_i}{x_{i-1} - x_i} + p(x_i)\frac{x - x_{i-1}}{x_i - x_{i-1}}, \;\; x_{i-1} \leq x \leq x_i, \;\; i = 1, \dots, n. \tag{2.20}$$

Note that the points $x_0, \dots, x_n$ are called the knots of the spline.

Recall also that splines allows to enforce elementwise continuity on function and arbitrary $k-th$ order derivatives. Furthermore, the spline functions are found to be finite dimensional in nature, which is the primary reason for their utility in computations and representation.

Splines are one of the most used techniques and cover numerous applications in many fields. Applications such as image interpolation and digital filtering [114], improving ocean tide predictions [115], optimizing pharmaceutical formulations [116], recovering dynamic information contained in discrete data [117], are only a few of the wide range of applications this tool covers (see also [118–122]).

Chapter 3

Methodology

In this section we introduce the techniques used in this thesis. Starting with the higher order dynamic mode decomposition (HODMD) algorithm, which is the method that has been the focus of our work. Next, the information recovery technique, which combines the HODMD with an interpolation technique, is reported in details.

3.1 Higher order dynamic mode decomposition (HODMD)

Higher order dynamic mode decomposition (HODMD) is an extension of the well-known DMD method (2010) [91]. HODMD was developed by S. Le Clainche and José M. Vega (2017) [123] for the analysis of complex data modeling non-linear dynamical systems.
Similarly to DMD, HODMD decomposes spatio-temporal data into a number of modes; each mode has its associated frequency, growth rate and amplitude. Furthermore, the HODMD method excels at identifying complex dynamics while presenting robust noise rejection properties, as it has been proved in several occasions (e.g. [85, 86, 124]).

Since this algorithm is a data driven method, the data will be arranged, as mentioned above, either in matrix form as in eq. (2.1) or in tensor form as in eq. (2.2).

Note that the standard DMD described in 2.1.5, rely on following *Koopman assumption*:

$$\boldsymbol{v}_{k+1} = \boldsymbol{R}\boldsymbol{v}_k, \quad \text{for} \quad k = 1, ..., K-1, \tag{3.1}$$

where $\boldsymbol{v}_k$ are the spatio-temporal data organized in K equispaced J-dimensional *snapshots* and $\boldsymbol{R}$ is the *Koopman* matrix.
Meanwhile, the HODMD relies on a *higher order Koopman assumption*, which relates each snapshot with the d previous delayed snapshots using several *Koopman* matrices as follows:

$$\boldsymbol{v}_{k+d} = \boldsymbol{R}_1\boldsymbol{v}_k + \boldsymbol{R}_2\boldsymbol{v}_{k+1} + \cdots + \boldsymbol{R}_d\boldsymbol{v}_{k+d-1} \quad \text{for} \quad k = 1, \ldots, K-d. \tag{3.2}$$

The HODMD algorithm represents the spatio-temporal data $\boldsymbol{v}_k$ as an expansion of M modes $\boldsymbol{u}_m$, each mode has its own amplitude a_m, frequency ω_m and growth rate δ_m as follows:

$$\boldsymbol{v}(t) \simeq \sum_{n=1}^{M} a_m \boldsymbol{u}_m e^{(\delta_m + i\omega_m)(t-t_1)} \quad \text{for} \quad t_1 \leq t \leq t_1 + T. \tag{3.3}$$

The HODMD algorithm can be summarized in the following steps:

i) First, the data undergoes a dimensionality reduction via the application of the *truncated* SVD to the full snapshot matrix eq. (2.1) as

$$\boldsymbol{V}_1^K \simeq \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{T}^T, \tag{3.4}$$

where the number of modes retained N, is defined as $\sigma_{N+1}/\sigma_1 \leq \varepsilon_1$, where $\sigma_1, \ldots, \sigma_N$ are the singular values and the threshold ε_1 is selected according to the level of noise in the data. The previous equation can be written as:

$$\boldsymbol{V}_1^K \simeq \boldsymbol{W}\widehat{\boldsymbol{V}}_1^K, \quad \text{where} \quad \widehat{\boldsymbol{V}}_1^K = \boldsymbol{\Sigma}\boldsymbol{T}^T, \tag{3.5}$$

and $\widehat{\boldsymbol{V}}_1^K$ is called *the reduced snapshot matrix*.

ii) Next, the higher order Koopman assumption defined in eq. (3.2) is used to the reduced snapshot matrix as

$$\widehat{\boldsymbol{V}}_{d+1}^K \simeq \widehat{\boldsymbol{R}}_1\widehat{\boldsymbol{V}}_1^{K-d} + \widehat{\boldsymbol{R}}_2\widehat{\boldsymbol{V}}_2^{K-d+1} + \cdots + \widehat{\boldsymbol{R}}_d\widehat{\boldsymbol{V}}_d^{K-1}, \tag{3.6}$$

where $\widehat{\boldsymbol{R}}_k = \boldsymbol{W}^T\boldsymbol{R}_k\boldsymbol{W}$.

This equation can be represented using the reduced snapshot matrix and the modified Koopman matrix $\tilde{\boldsymbol{R}}$ as follows :

$$\widetilde{\boldsymbol{V}}_2^{K-d+1} = \widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{V}}_1^{K-d}, \tag{3.7}$$

where

$$\tilde{\boldsymbol{V}}_1^{K-d} = \begin{bmatrix} \widehat{\boldsymbol{V}}_1^{K-d} \\ \widehat{\boldsymbol{V}}_2^{K-d+1} \\ \ldots \\ \widehat{\boldsymbol{V}}_d^{K-1} \end{bmatrix}, \quad \tilde{\boldsymbol{V}}_2^{K-d+1} = \begin{bmatrix} \widehat{\boldsymbol{V}}_2^{K-d+1} \\ \ldots \\ \widehat{\boldsymbol{V}}_d^{K-1} \\ \widehat{\boldsymbol{V}}_{d+1}^{K} \end{bmatrix}, \quad \tilde{\boldsymbol{R}} = \begin{bmatrix} 0 & \boldsymbol{I} & 0 & \ldots & 0 & 0 \\ 0 & 0 & \boldsymbol{I} & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \boldsymbol{I} & 0 \\ \widehat{\boldsymbol{R}}_1 & \widehat{\boldsymbol{R}}_2 & \widehat{\boldsymbol{R}}_3 & \ldots & \widehat{\boldsymbol{R}}_{d-1} & \widehat{\boldsymbol{R}}_d \end{bmatrix}. \tag{3.8}$$

Note that eq. (3.6) suggests that, as illustrated in Fig. (3.1), it is possible to compare HODMD with the sliding window process [125], usually carried out in power spectral density analysis. This allows HODMD to calculate several temporal modes associated with a

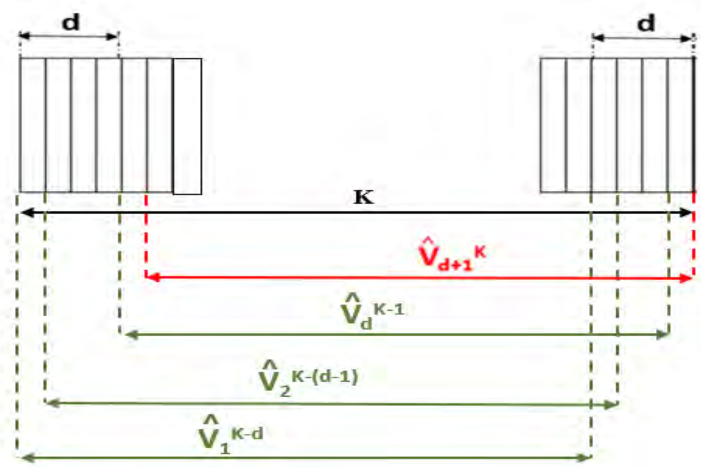


Figure 3.1: *Reduced snapshot matrix* $\widehat{V}_1^{d+1}$ and the sliding window process defined in eq. (3.6).

single spatial mode, ameliorating the performance of HODMD compared to the standard DMD.

A second dimensionality reduction is carried out to the matrix containing the reduced snapshots using truncated SVD and the tolerance ε_1 as $\tilde{\sigma}_{N'+1}/\tilde{\sigma}_1 < \varepsilon_1$, where N' is the number of retained SVD modes and $\tilde{\sigma}_i$ are the singular values. This truncation yields

$$\tilde{\boldsymbol{V}}_1^{K-d+1} \simeq \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{T}}^T \simeq \tilde{\boldsymbol{U}}\overline{\boldsymbol{T}}_1^{K-d+1}, \text{with} \quad \overline{\boldsymbol{T}}_1^{K-d+1} = \tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{T}}^T. \tag{3.9}$$

This step is completed through pre-multiplying eq. (3.7) by $\tilde{\boldsymbol{U}}^T$, and invoking eq. (3.9) gives:

$$\overline{\boldsymbol{T}}_2^{K-d+1} = \bar{\boldsymbol{R}}\overline{\boldsymbol{T}}_1^{K-d}, \tag{3.10}$$

such that $\overline{\boldsymbol{R}} \in N' \times N'$ is the new *Koopman matrix* defined as $\overline{\boldsymbol{R}} = \tilde{\boldsymbol{U}}^T\tilde{\boldsymbol{R}}\tilde{\boldsymbol{U}}$. However, the matrix $\overline{\boldsymbol{R}}$ is not computing with this expression (as it would be typically very memory demanding), instead we use the methodology presented in the next step.

iii) The following step is computing the DMD modes, frequencies and growth rates. In order to do so, $\overline{\boldsymbol{R}}$ must be computed first, which is simply done by applying, once again, the SVD on the matrix $\overline{\boldsymbol{T}}_1^{K-d}$

$$\overline{\boldsymbol{T}}_1^{K-d} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T, \tag{3.11}$$

then we substitute eq. (3.11) in eq. (3.10) and multiply the result by $\boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{U}^T$ to obtain:

$$\overline{\boldsymbol{R}} = \overline{\boldsymbol{T}}_2^{K-d+1}\boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{U}^T. \tag{3.12}$$

Once the matrix $\overline{\boldsymbol{R}}$ has been calculated, the reduced DMD expansion for the reduced snapshots (eq. (3.5)) can be computed as follows:

$$\hat{\boldsymbol{v}}_k = \sum_{m=1}^{M} \hat{a}_m\hat{\boldsymbol{u}}_m e^{(\delta_m+i\omega_m)t_k}, \quad \text{for} \quad k = 1, \ldots, K. \tag{3.13}$$

In particular, the eigendecomposition of the matrix $\overline{\boldsymbol{R}}$

$$\overline{\boldsymbol{R}}\hat{\boldsymbol{V}}_{\boldsymbol{k}} = \lambda_k \hat{\boldsymbol{V}}_{\boldsymbol{k}},$$

provides on the one hand, the reduced DMD modes $\hat{\boldsymbol{u}}_m$, which are calculated by keeping the first M elements of the vector $\hat{\boldsymbol{q}}_m = \tilde{\boldsymbol{U}}\bar{\boldsymbol{q}}_m$, where $\bar{\boldsymbol{q}}_m$ represents the eigenvectors of $\overline{\boldsymbol{R}}$. In the other hand, the associated eigenvalues $\boldsymbol{\mu}_m$ provides the frequencies ω_m and growth rates δ_m by the following expression:

$$\delta_m + i\omega_m = log(\boldsymbol{\mu}_m)/\Delta t. \tag{3.14}$$

iv) Finally, the amplitudes a_m are computed via least-square fitting of eq. (3.13). The DMD modes are ordered with respect to decreasing value of a_m. It is possible to determine the number of M modes to retrain in the DMD expansion eq. (3.3) using a different tolerance ε_2 such as $a_{M+1}/a_1 \leq \varepsilon_2$, and finally computing the DMD expansion for the original *snapshots*.

3.2 Multidimensional higher order dynamic mode decomposition

Multidimensional iterative HODMD is an extension of HODMD presented in [83], used for the analysis and pattern identification in complex, highly noisy, experimental data. For this algorithm, and unlike standard HODMD where the data is organized in a snapshot matrix, the data is organized in tensor form as presented in eq. (2.2) and employs the HOSVD instead of the standard SVD. The HOSVD algorithm applies SVD along each spatial direction in order to better clean the data. Meanwhile, the iterative algorithm is basically obtaining the expansion eq. (3.3) using HODMD and then the multidimensional HODMD is applied iteratively over the reconstructed data until the number of HOSVD modes is the same between two consecutive iterations (see [83] for more details).

As mentioned, the data matrix is substituted by a multidimensional snapshot *tensor* as in eq. (2.2) and the expansion presented in eq. (3.3) in this situation is defined as follows:

$$\boldsymbol{T}_{j_1 j_2 k} \simeq \sum_{m=1}^{M} a_m \boldsymbol{U}_{j_1 j_2 m} e^{(\delta_m + i\omega_m)(k-1)\Delta t}, \quad \text{for} \quad k = 1, \dots, K, \tag{3.15}$$

where $j_1 = 1, \dots, J_1; j_2 = 1, \dots, J_2; j_3 = 1, \dots, J_3$ (J_k is the size of the vector $\boldsymbol{T}_{j_1 j_2 k}$).

This algorithm has two main steps. First one is a dimensionality reduction, but instead of using SVD, the HOSVD algorithm is applied to *the snapshot tensor*, which yields the following decomposition:

$$\boldsymbol{T}_{j_1 j_2 k} \simeq \mathbf{tprod}(\boldsymbol{\Sigma}, \boldsymbol{W}^{(1)}, \boldsymbol{W}^{(2)}, \mathsf{T}) \simeq \sum_{p_1=1}^{P_1} \sum_{p_2=1}^{P_2} \sum_{n=1}^{N} \boldsymbol{\Sigma}_{p_1 p_2 n} \boldsymbol{W}^{(1)}_{j_1 p_1} \boldsymbol{W}^{(2)}_{j_2 p_2} \mathsf{T}_{kn}, \tag{3.16}$$

where $\boldsymbol{\Sigma}$ is the *core tensor* and the columns of $\boldsymbol{W}^{(1)}, \boldsymbol{W}^{(2)}$ are called spatial modes. The columns of $\mathbf{T}_{kn}$ are the temporal modes of the decomposition. The process of obtaining these three sets of modes (by applying SVD to the three matrices, whose columns are the associated fibers of the tensor), results three sets of non-zero singular values $\sigma_{p_1}^{(1)}, \sigma_{p_2}^{(2)}$ and $\sigma_n^{(t)}$.

Hence, eq. (3.16) can be rewritten as

$$\boldsymbol{T}_{j_1 j_2 k} \simeq \underbrace{\sum_{n=1}^{N} \sum_{p1=1}^{P_1} \sum_{p_2=1}^{P_2} \boldsymbol{\Sigma}_{p_1 p_2 n} \boldsymbol{W}_{j_1 p_1}^{(1)} \boldsymbol{W}_{j_2 p_2}^{(2)} \frac{1}{\sigma_n^t}}_{\widehat{\boldsymbol{\Sigma}}_{j_1 j_2 n}} \underbrace{\sigma_n^t \mathbf{T}_{kn}}_{\widehat{\mathbf{T}}_{kn}}, \tag{3.17}$$

which allows to introduce both the spatial modes $\widehat{\boldsymbol{\Sigma}}_{j_1 j_2 n}$ and the rescaled temporal modes $\widehat{\mathbf{T}}_{kn}$ such that

$$\widehat{\boldsymbol{\Sigma}}_{j_1 j_2 n} = \sum_{p1=1}^{P_1} \sum_{p_2=1}^{P_2} \boldsymbol{\Sigma}_{p_1 p_2 n} \boldsymbol{W}_{j_1 p_1}^{(1)} \boldsymbol{W}_{j_2 p_2}^{(2)} / \sigma_n^t \text{ and } \widehat{\mathbf{T}}_{kn} = \sigma_n^t \mathbf{T}_{kn}. \tag{3.18}$$

We determine the number of modes to retain N from each one of the spatial modes, using the tolerance ε_{HOSVD} (as in eq. (2.14)). Finally, steps from ii) to iv) of the standard HODMD introduced in 3.1 is applied to the temporal modes $\mathbf{T}$.

3.3 SVD-based interpolation technique

The interpolation technique used in this work (that we shall evaluate in chapter 6.3) is inspired by an HOSVD based image resolution enhancement method, which was developed by Rövid *et al.* [126]. In particular, color channels (if it is a RGB image) of a considered image $f(x) = (x_1, x_2, x_3)$ (where x_1 and x_2 represent the horizontal and vertical coordinates of the pixel, respectively and x_3 is related to the color components), are stored in a $(m \times n \times 3)$ tensor $\boldsymbol{\mathcal{B}}$ (where m and n correspond to the width and height of the image). The tensor $\boldsymbol{\mathcal{B}}$ goes under a factorization process using the HOSVD algorithm, which yields:

$$\boldsymbol{\mathcal{B}} = \textbf{tprod}(\boldsymbol{\mathcal{D}}, \boldsymbol{U}^{\boldsymbol{(1)}}, \boldsymbol{U}^{\boldsymbol{(2)}}, \boldsymbol{U}^{\boldsymbol{(3)}}). \tag{3.19}$$

The main concept is to show the effectiveness of image scaling in the HOSVD based domain, where additional values are interpolated between each neighbouring pixel pair of the columns of the matrices $U^{(n)}$. When all the columns are updated, the new image can be reconstructed using eq. (3.19).

In this contribution, we follow a similar framework of this method to introduce a new information recovery technique. However, in our case, two dimensional arrays are employed instead of tensors and the HOSVD is substituted by the SVD algorithm, as explained in the following (see also Fig. (3.2)).

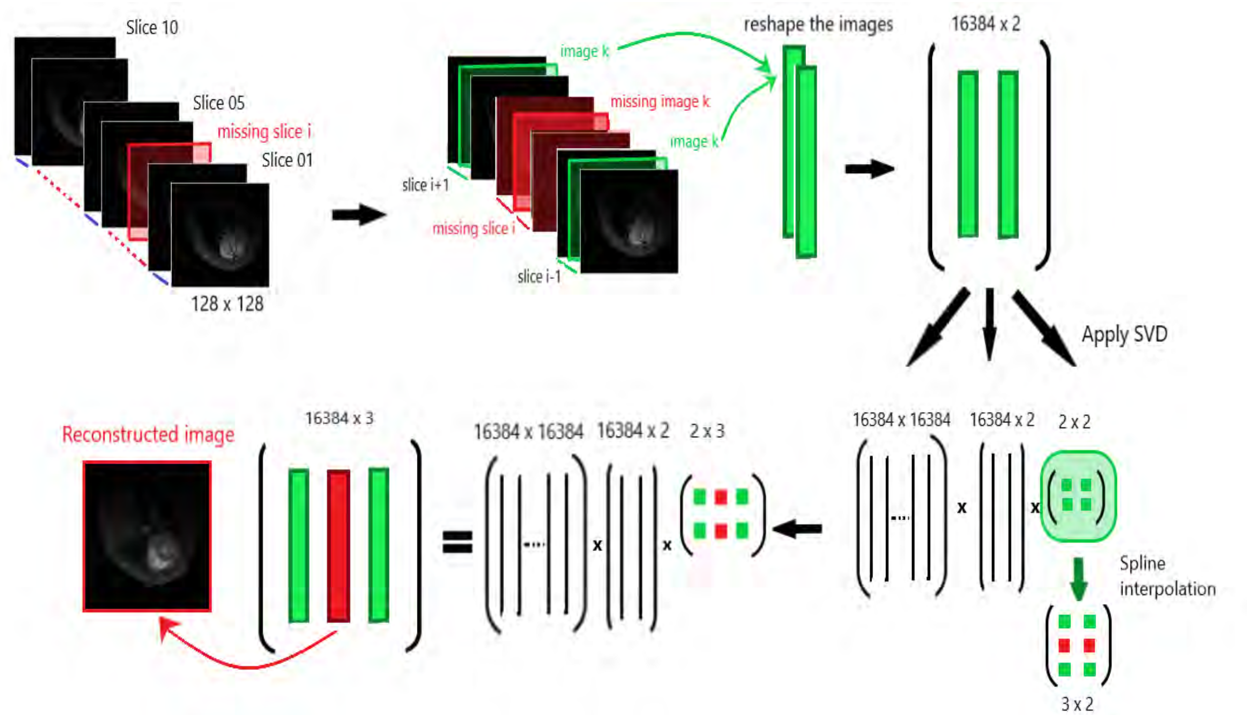


Figure 3.2: Schematic diagram for the interpolation technique pipeline.

To reconstruct the missing information of an image f_k (for $k = 1, ..., K$, with K is the number of snapshots) of a slice i, we consider a $J \times 2$ matrix denoted $\boldsymbol{M}$. The matrix contains in columns the pixels of the kth image of slices $i-1$ and $i+1$, respectively, with $J = N_x \times N_y$. First, we apply the SVD the matrix $\boldsymbol{M}$ as follows:

$$\boldsymbol{M} = \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{T}^T, \tag{3.20}$$

where $\boldsymbol{W} \in \mathbb{R}^{J\times J}$ and $\boldsymbol{T} \in \mathbb{R}^{2\times 2}$ are unitary matrices and J is number of SVD modes. The matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{J\times 2}$ contains real, non-negative singular values on the diagonal and zeros off the diagonal.

To generate the missing image, the matrix $\boldsymbol{T}$ will be updated, where the number of columns will remain the same and the number of lines will be extended to 3. Lets note $\tilde{\boldsymbol{T}}$ the updated matrix, the lines $\tilde{\boldsymbol{T}}_s$ with $1 \leq s \leq 3$, will be determined as follows: $\tilde{\boldsymbol{T}}_1 = \boldsymbol{T}_1$, $\tilde{\boldsymbol{T}}_3 = \boldsymbol{T}_2$, meanwhile, spline interpolation introduced in 2.2.1 is used to determine the elements of $\tilde{\boldsymbol{T}}_2$. When all the lines are determined, eq. (3.20) is used to obtain a new matrix containing 3 columns, representing the pixels of the two previous images in addition to the pixels of the reconstructed missing image (the pipeline of the method is illustrated in Fig. (3.2)).
This procedure is repeated for each time step t_k in order to reconstruct all the snapshots of the missing slice.

The reconstruction error incurred by the interpolation process can be assessed using the relative root mean square error (RRMSE), which is defined as follows:

$$RRMSE = \frac{\|\boldsymbol{Y}^{approx} - \boldsymbol{Y}\|_F}{\|\boldsymbol{Y}\|_F}, \tag{3.21}$$

where $\boldsymbol{Y}^{approx}$ is the new reconstructed ith slice, $\boldsymbol{Y}$ is the ith original slice and $\|.\|_F$ is the Frobenius norm for matrices and tensors.

Part III

Material and Results

Summary In this part we provide first a detailed description of the different datasets used in this work. Next, the main results obtained are profoundly elaborated.
In particular, three chapters are included. Chapter 4 explain in details the acquisition of the echocardiography and MRI datasets investigated in this contribution. The description includes the type of cardiac diseases investigated, how the diseases are modeled and the size of the different datasets. Next, the results obtained from the analysis of the different medical datasets are split into two chapters: i) chapter 5 covers the results obtained from analysing the echocardiography datasets, which involves the identification and extraction of features ii) Chapter 6 offers the results corresponding to the MRI datasets, which include 3D reconstructions of the heart, enlarging databases using HODMD-based reduced order model and finally, MR images reconstruction using the information recovery technique.

Chapter 4

Data bases

In the following, both the datasets (echocardiogrpahy and magnetic resonance imaging (MRI)) used in this contribution are explained in details, including the cardiac diseases investigated, the generation of images and the size of the different databases.

4.1 Echocardiography datasets

The data includes several sets of medical images, specifically, echocardiography taken from different mice. The data is analyzed to investigate the following cardiac diseases:

- Myocardial infarction, also known as a heart attack. It mainly occurs when one or more areas of the heart muscle don't get enough oxygen, which usually happens when blood flow to the heart muscle is blocked by a blood clot (thrombus), causing the death of the heart muscles and leaving a permanent damage.
- Obesity, it is a complex disease involving an excessive amount of body fat. It increases the risk of other diseases and health problems such as diabetes and high blood pressure. Furthermore, people with obesity are more likely to develop a rapid and irregular heart rate, which can lead to strokes.
- Cardiac hypertrophy (TAC hypertrophy) is the abnormal enlargement, or thickening of the heart muscle, resulting from increases in cardiomyocyte size and changes in other heart muscle components. Thickened heart muscle, as well as the abnormal structure of heart cells, can cause changes in the heart's electrical system, resulting in fast or irregular heartbeats.
- Diabetic cardiomyopathy is defined by the existence of abnormal myocardial structure in people with diabetes, it can lead to inability of the heart to circulate blood through the body effectively causing heart failure.

4.1.1 Mouse model diseases

Hearts from healthy C57BL/6 10-weeks-old mice were used as control (CTL). Diabetes was induced in mice by injecting streptozotocin (STZ, 50mg/kg, 0.05mol/L in citrate buffer, pH 4.5, Sigma, St. Louis, USA) i.p for five consecutive days [127] and images were assessed 16 weeks post-induction. Obesity was induced feeding mice with high fat diet [45 kcal% (24 g%) palm oil-based fat, 35 kcal% (41 g%) carbohydrate, 20 kcal% (24 g%) protein; based on OpenSource Diets No. D12451, Research Diet Services, Wijk bij Duurstede, The Netherlands] and images were assessed 88 weeks post-induction. Cardiac hypertrophy images were obtained from mutant mice previously described (SFSR4 KO, [128]) and images were assessed 24 weeks of age. Another model of hypertrophy was induced using the aorta constriction surgery [129] and images were assessed 4 weeks post-induction. Lastly, a model of myocardial infarction was used from mice subjected to left anterior descending coronary artery ligation ([130], [129]) and images were assessed 4 weeks post-induction.

The total number of datasets in this study is 260, equally split into five cardiac conditions, including healthy conditions and the four previously mentioned diseases.

4.1.2 Echocardiography images acquisition

Transthoracic echocardiography was performed under isoflurane anaesthesia by an expert operator using a high-frequency ultrasound system (Vevo 2100, Visualsonics Inc, Canada) with a 40-MHz linear probe. Isoflurane was administered in 100% oxygen and the dose was adjusted to maintain podal reflex (light anaesthesia plane). Mice were placed in supine position using a heating platform and warmed ultrasound gel was used to preserve normothermia. A base apex electrocardiogram was used for continuous monitoring. Bidimensional (2D) parasternal standard long and short axis views (LAX and SAX, respectively shown in Fig. (4.1)) of the left ventricle (LV) were obtained as previously described [127].

Offline, 260 LAX and SAX video loops from different mouse models of cardiac diseases (diabetes, obesity, hypertrophy and infarction) and from healthy mice were exported as DICOM format, including at least 3 cardiac cycles.

4.2 Cardiac cine MRI datasets

Three MRI datasets are employed in this research. Each dataset consists of 10 MRI sequences (slices). In particular, 20 images (snapshots) of $128 \times 128 \times 1$ resolution were acquired for each slice, resulting a total number of 200 snapshot per dataset. The datasets are taken from three mice, one with a healthy heart and two were diagnosed with a cardiac disease named SFSR4 hypertrophy.

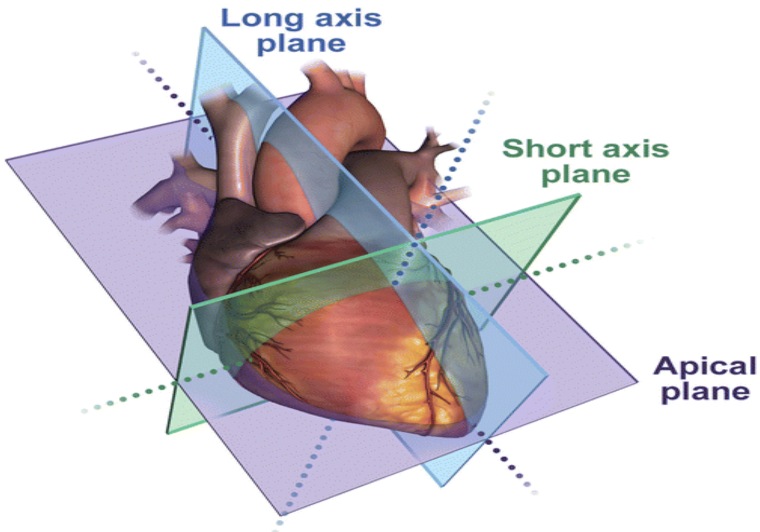


Figure 4.1: Imaging planes of the heart: the long axis (LAX) and the short axis (SAX). Original source of the image [1].

4.2.1 MRI acquisition

Hearts from healthy 10 months-old mice (and SFSR4 wild-type mice [128]), were used. All the data have been obtained from previous studies performed in accordance with protocols approved by the CNIC's Institutional Animal Care and Research Advisory Committee of the Ethics Committee of the Regional Government of Madrid (PROEX177/17). Mice were anesthetized by inhalation of 3-4% isoflurane for induction and 1.5-2% for maintenance), administered in 95% oxygen using a nasal mask. Ophthalmic gel was placed in their eyes to prevent retinal drying. Normothermia was maintained with forced air warming the procedure. The core body temperature, electrocardiogram, heart rate and respiratory rate were continuously monitored with a CMR-compatible system for rodents.

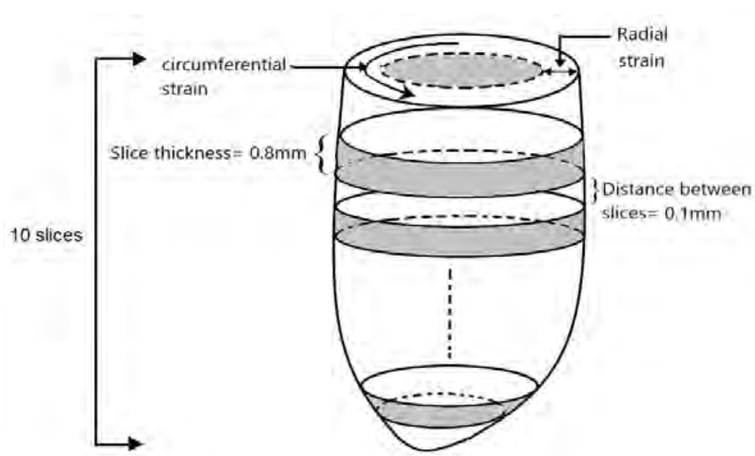


Figure 4.2: Sketch of a heart illustrating the acquisition of data using cine MRI.

In vivo cardiac images were acquired with a 7 by using 7-T Agilent/Varian scanner (Agilent, Santa Clara, CA, USA) equipped with a DD2 console, with an actively shielded 115/60 gradient. A surface coil was used for MR signal transmission and reception. MRI experiments were conducted by applying an *ECG-triggered fast gradient echo cine* sequence with the following imaging parameters: TR/TE values around 125/1.25 msec (they are minimum values depending on heart rate and number of frames per heart cycle); field of view, 30 × 30 mm; acquisition matrix, 128 × 128; flip angle 15°; 4 averages; 20 cardiac phases; 10 slices; slice thickness, 0.8 mm and gap, 0.1 mm.

Chapter 5

Results for echocardiography datasets

In this chapter, the results obtained from application of the HODMD algorithm to the echocardiography datasets are explained in details.

A total number of 260 datasets are analysed using the HODMD algorithm, however, for the sake of brevity, the results of only two datasets from each cardiac disease are presented in this document [1].

5.1 HODMD for feature extraction

In this section, echocardiography datasets were analyzed using HODMD (presented in 3.2). As stated before (see section 4.1), the datasets consist of video loops taken from both LAX and SAX view. Each video covers at least three cardiac cycles, but not surpassing 300 frames per video.

All the datasets went through the same procedure represented in Fig. (5.1) in order to be analyzed by HODMD. The first step is extracting all the frames from the videos, cropping and removing some parts of the image to eliminate any disturbances that might affect the signal, leaving only the part displaying the heart area in the medical image (as shown in Fig. (5.2)). Each one of these modified frames will be converted into a grayscale image, and then sorted in a *tensor*, which is the HODMD input. The HODMD technique will denoise the data, capture the relevant frequencies and provide the DMD modes, which represent the main patterns related to the different cardiac conditions. In each case a different number of modes are identified, however in this document only two modes of each dataset are displayed.

[1]The material discussed in this section has been published as full paper titled "Higher Order Dynamic Mode Decomposition: from Fluid Dynamics to Heart Disease Analysis" [131] in the peer-reviewed, JCR indexed journal Computers in Biology and Medicine (CIBIM).

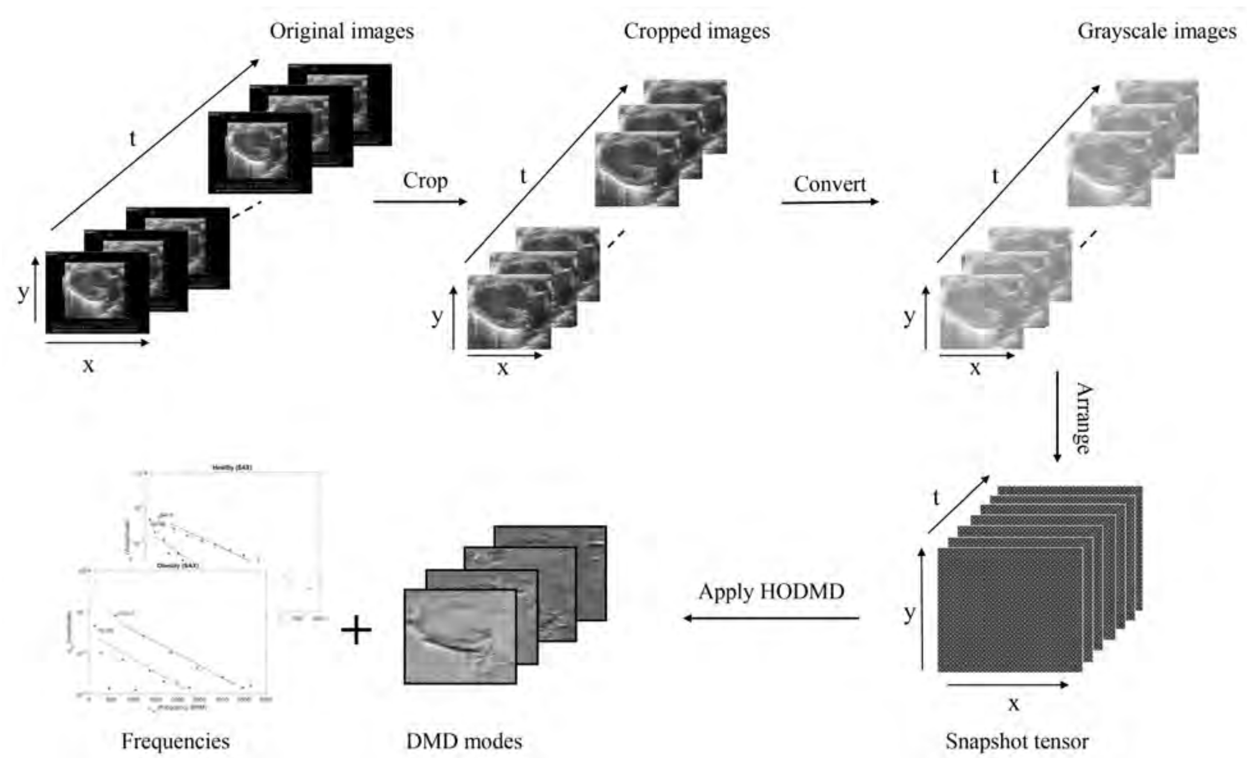


Figure 5.1: Schematic diagram for the higher order dynamic mode decomposition (HODMD) analysis pipeline.

The next step is selecting the following HODMD parameters:

1. The number of snapshots K, where the goal is to capture several full cardiac cycles.
2. Two different thresholds: ε_1 for dimentionality reduction, which should be larger than the noise level, and ε_2 for amplitude truncation, such that, decreasing its value increases the number of the identified frequencies.
3. The time step Δt, which is the temporal distance between snapshots, and the sampled interval T.
4. Finally, the index d, which we increase or decrease until we minimize the relative RRMS error and capture two accurate lines of frequencies.

After some calibration, the main parameters are chosen as follows:

1. The number of snapshots K: in the cases of, healthy LAX data, diabetic cardiomyopathy (both LAX and SAX) and obesity (SAX), the number of snapshots was set to $K = 200$, and for the rest of the data sets the number of snapshots was set to $K = 100$. The choice of this parameter is highly related to the noise level in the dataset. Hence, in cases where we have noisy data, the analysis of longer sequences allows to segregate the cardiac from the respiratory frequencies.

2. The two thresholds were fixed to be $\varepsilon_1 = \varepsilon_2 = 5 \times 10^{-4}$ for all the datasets.

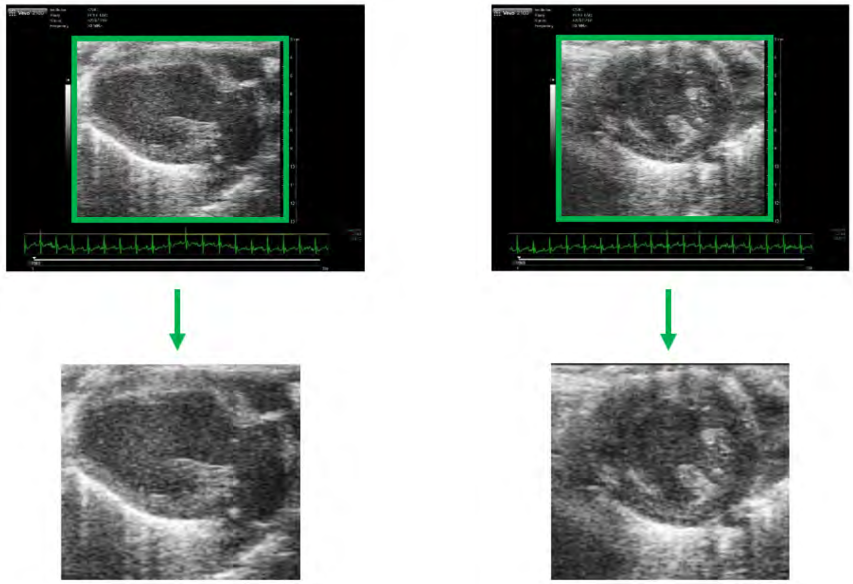


Figure 5.2: Echocardiography frame at time $t = 0$, pre and post cropping.

3. The time step or the time between the snapshots Δt is estimated to be 4 milliseconds ($4 \times 10^{-3} s$), which is given by the ultrasound scanner after being set by specialists to guarantee the production of proper, representative images. Regarding the time interval, $T = [0, K \times \Delta t]$ is scaled with the time step Δt when the number of snapshots is set to 100, and with double the time step ($2\Delta t$) when the number of snapshots is set to 200.

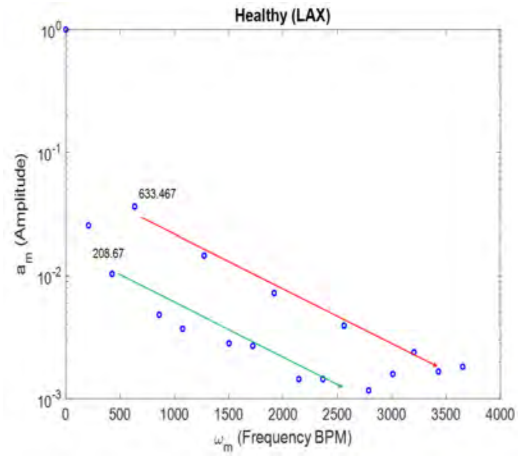
4. Finally, the index d is chosen between ($30 \sim 35$) when $K = 100$, and between ($60 \sim 70$) when $K = 200$, in good agreement with the calibration process described in [123].

5.1.1 Healthy data

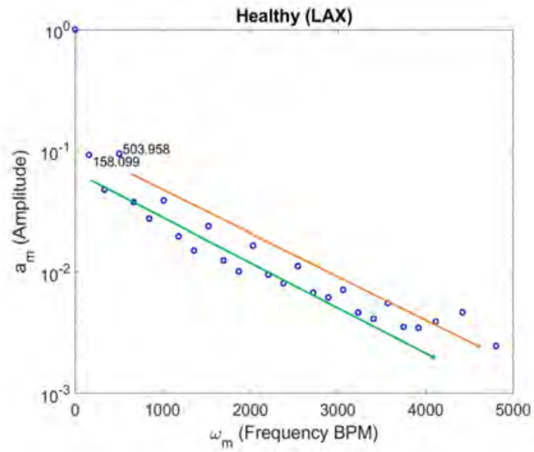
The first dataset analyzed by the HODMD algorithm is the data taken from the healthy subjects. As it is shown in fig. (5.3) and Fig. (5.4), the HODMD was able to capture and segregate two signals, which can be seen represented in two different lines of frequencies for both LAX and SAX views. These signals, which are periodic and regular, were confirmed to be related to the heart rate and the respiratory rate. Particularly, the upper branch is representing the frequency of the signal caused by heart and the lower branch is representing the frequency of the signal caused by the lungs. Second, as anticipated, the HODMD algorithm delivered a set of DMD modes representing the main patterns related to this cardiac condition. As can be observed in Fig. (5.3) and Fig. (5.4), the obtained DMD modes display more details and clear patterns, which may not be as clear in the original image. Furthermore, in this dataset in particular, an interesting observation we encountered is the fact that the DMD modes can be divided into two sets: modes that belong to the upper branch and present their highest intensity in the area representing the heart, where we have a clear view of the left ventricle in the long axis view, and modes that belong to the lower branch and present a higher intensity in the area of the image representing the lungs.

The heart and respiratory rates captured in the analysis of these datasets are as follows: for the first dataset, the dominant frequency in the upper branch is 633 beats per minute (BPM) for the LAX view and 644 BPM for the SAX view. For the respiratory rate, 208 breaths per minute and 203 breaths per minute were captured for LAX view and SAX view, respectively. In the second dataset, 503 BPM and 514 BPM were captured as dominant frequencies in the upper branch for LAX view and SAX view, respectively. Meanwhile, for the respiratory rate, the dominant frequencies are 158 breaths per minute for the LAX and 210 breaths per minute for the SAX. It is also worth to mention that the changes in the frequencies between the long axis view and the short axis view are normal, anticipated and do not contrast with the efficiency of the algorithm, as several factors effect both the heart rate and respiratory rate, including gender, age and also the dose anesthesia (the values of all the frequencies from the upper and lower branches can be found in Table 1 in Appendix V).

01- Healthy heart results (LAX)

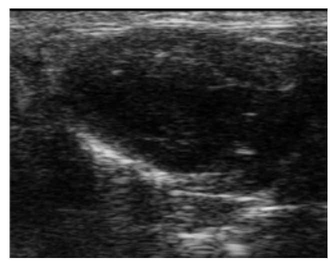


(a) Frequency obtained from the healthy heart (first dataset)



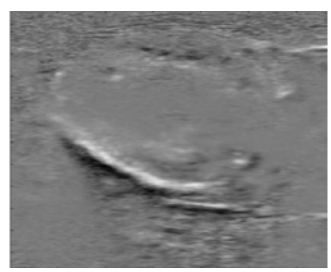
(b) Frequency obtained from the healthy heart (second dataset)

(c) Original image

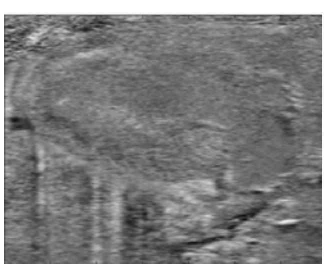


(d) Original image

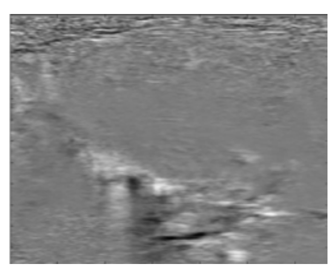
(e) Upper branch mode



(f) Upper branch mode



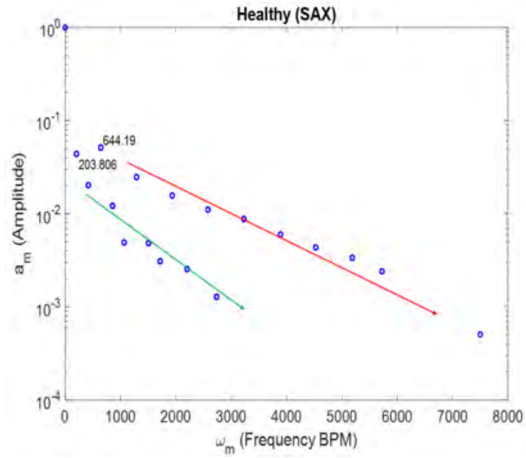
(g) Lower branch mode



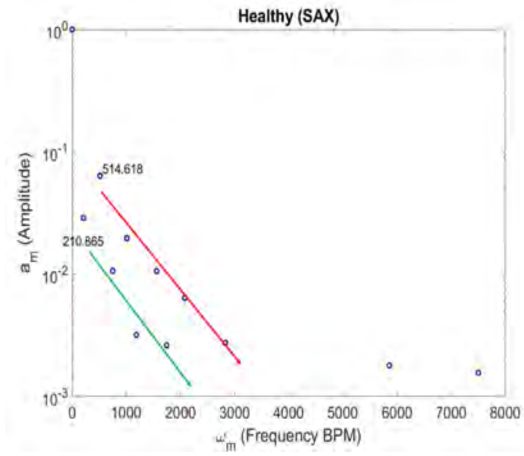
(h) Lower branch mode

Figure 5.3: The frequencies captured from analyzing the data of the healthy hearts (LAX) and the dominant modes related to each branch of frequency. The values of the frequencies from the upper and lower branches can be found in Table 1 in Appendix V.

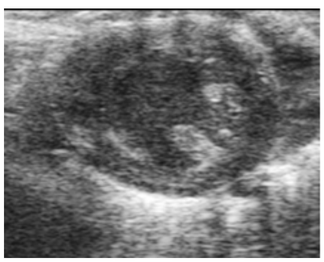
02- Healthy heart results (SAX)



(a) Frequency obtained from the healthy heart (first dataset)



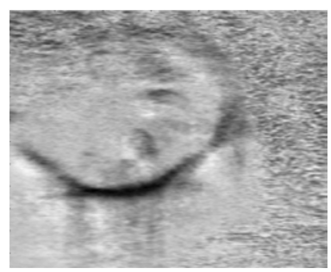
(b) Frequency obtained from the healthy heart (second dataset)



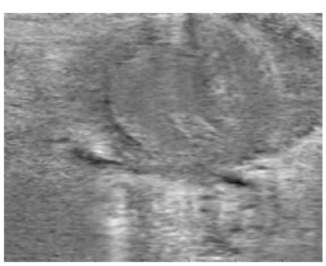
(c) Original image

(d) Original image

(e) Upper branch mode



(f) Upper branch mode



(g) Lower branch mode

(h) Lower branch mode

Figure 5.4: The frequencies captured from analyzing the data of the healthy hearts (SAX) and the dominant modes related to each branch of frequency. The values of the frequencies from the upper and lower branches can be found in Table 1 in Appendix V.

5.1.2 Unhealthy data

The analysis of the datasets obtained from the unhealthy subjects (diagnosed with the cardiac diseases introduced in 4.1) gave the following results:

1- Captured frequencies:

As can be seen in (Fig. 5.5, Fig. 5.7, Fig. 5.11) for the LAX data and in (Fig. 5.6, Fig. 5.9, Fig. 5.12) for SAX data, we still capture the two distinct lines of frequencies in each case. The frequencies might have some changes when compared to the healthy data but it is not necessarily related to the disease, as it can be caused by the anesthesia that all the mice go through, to try to keep the heart rate stable. We note also that there is a difference between the frequencies obtained from the LAX view and the ones obtained from the SAX view because in the case of short axis view the mice are taking a higher doze of anesthesia. Hence, in general, the captured frequencies, which still regular and periodic, fluctuate between 400 to 700 in the heart rate and between 150 to 250 for the respiratory rate. These changes are considered normal and anticipated following several factors including type of the disease, gender, age and as mentioned earlier, the dose of anesthesia (all the frequencies can be found in Tables 2, 3, 4 and 5 in appendix A V).

2- Extracted features:

Our main focus of this analysis is the patterns represented in the DMD modes, since one of key outputs of the HODMD algorithm are the extracted dominant patterns. Hence, the obtained modes can be seen representing -once again- clearer features and detailed patterns better than the original echocardiography images. These modes are considered strong representers of the shape of the heart in particular unhealthy conditions and display certain features related to a specific disease.

The modes obtained from analyzing the pathological heart datasets can be sorted into three main categories: myocardial infarction models (obtained from analyzing the myocardial infarction datasets), hypertension models (obtained from analyzing the obesity and TAC hypertrophy datasets), and diabetic cardiomyopathy models (obtained by analyzing diabetic cardiomyopathy datasets).

Case 1: Myocardial infarction model

In the case of the myocardial infarction, the resulted modes, for the LAX data in particular, show very similar abnormalities. As can be seen in Fig. (5.5) we can clearly see the change in the shape of the heart, as we can notice that the left ventricle does not show the normal shape of a heart. The modes instead show a perspicuous circle called aneurysm, which is an outward bulging, likened to a bubble or balloon, caused by a localized, abnormal, weak spot on a blood vessel wall. Furthermore, the posterior wall is inconspicuous, and it is undetectable because the myocardial thickness is lost due to the infarction. In the short axis view, we can not clearly see the aneurysm but in both cases the noticeable matter is how the interior wall is thinner and almost invisible (see Fig. (5.6)).

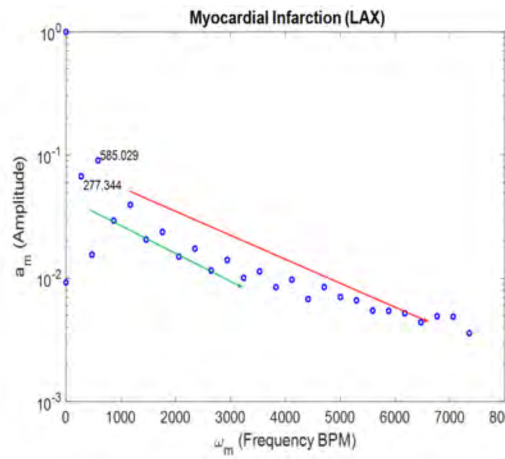
Case 2: Hypertrophy models

The diseases included in this case are obesity and TAC hypertrophy. As seen in Fig. (5.8), all the modes show hypertrophy and they share the same abnormalities. The mutual and most obvious feature is the increased wall thickness compared to the healthy case, which is evidenced in the circular shape of the heart and in the left ventricle, in particular. However, we can not make the same comments on the SAX view data, where similarly to the previous cases, this view does not show clear vision of the heart (see Fig. (5.10)).

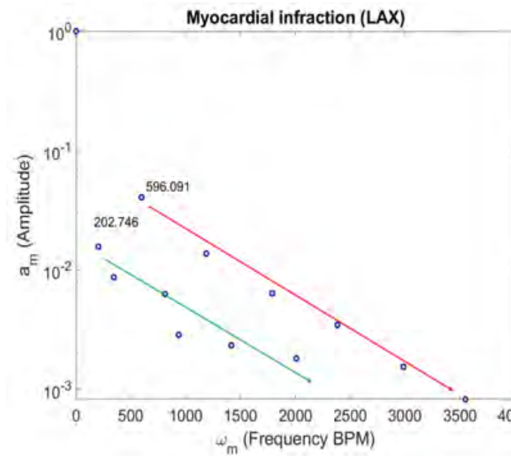
Case 3: Diabetic Cardiomyopathy

Finally, for the diabetic cardiomyopathy dataset, by observing Fig. (5.11) no abnormalities were captured, although the echocardiography was taken from mice with diabetic cardiomyopathy, the echocardiography images do not capture any changes in the shape of the heart. Nevertheless, this does not overlay the fact that the DMD modes are still offering an obvious enhancement of the features related to these datasets. We can notice the same enhancement in the SAX data, where as seen in Fig. (5.12), the DMD modes display detailed patterns and more characteristics, which can not be easily seen in the original images.

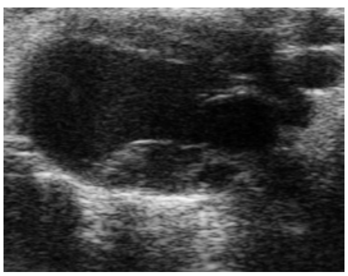
1- Myocardial infarction results (LAX)



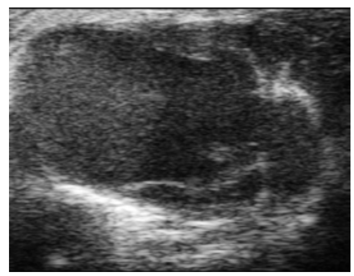
(a) Frequency obtained from the heart with myocardial infarction (first dataset)



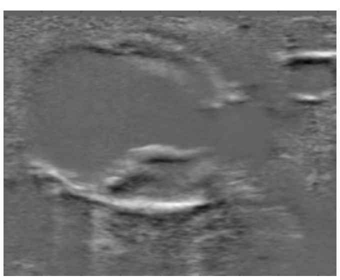
(b) Frequency obtained from the heart with myocardial infarction (second dataset)



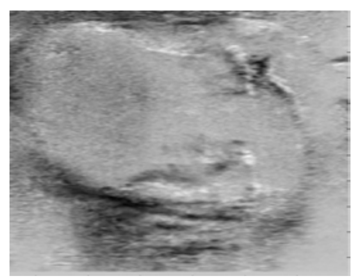
(c) Original image



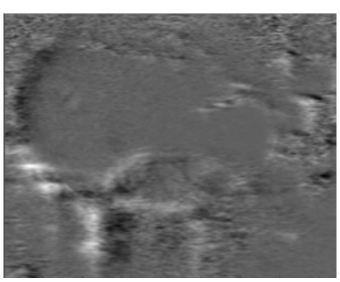
(d) Original image



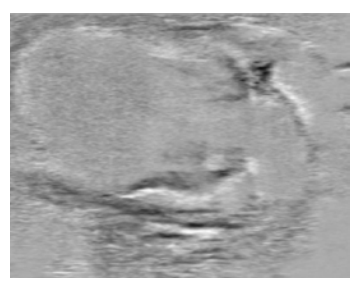
(e) Upper branch mode



(f) Upper branch mode



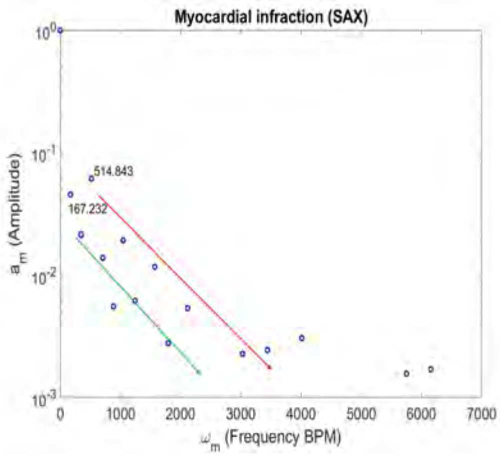
(g) Lower branch mode



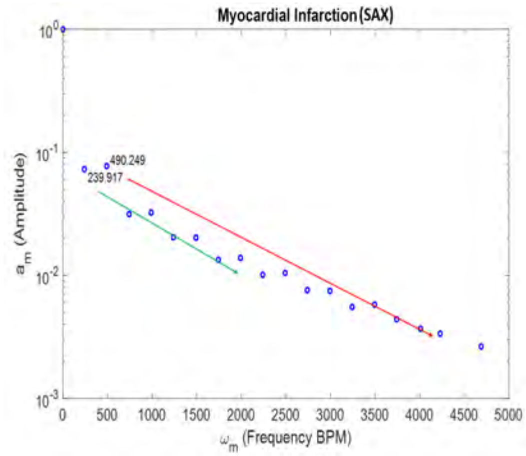
(h) Lower branch mode

Figure 5.5: The frequencies captured from analyzing the data of the hearts diagnosed with myocardial infarction and the dominant modes related to each branch of frequency. The values of the frequencies from the upper and lower branches can be found in Table 2 in Appendix V.

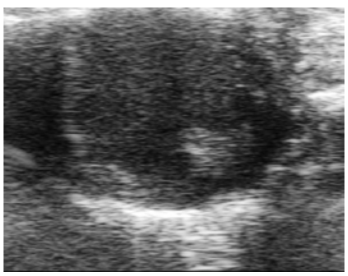
2- Myocardial infarction results (SAX)



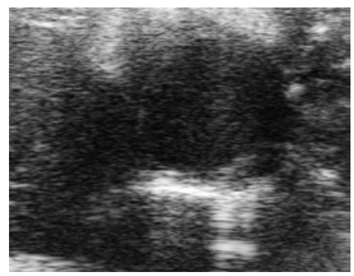
(a) Frequency obtained from the heart with myocardial infarction (first dataset)



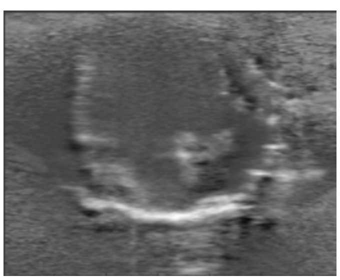
(b) Frequency obtained from the heart with myocardial infarction (second dataset)



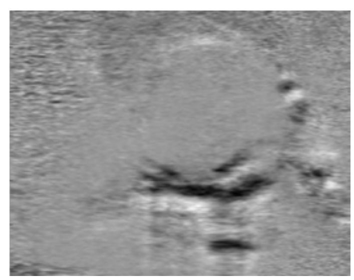
(c) Original image



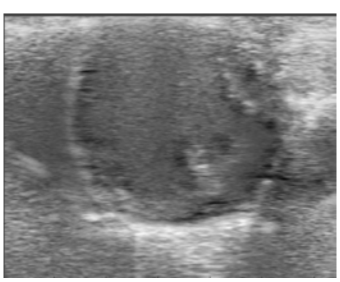
(d) Original image



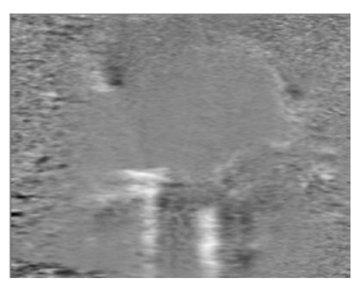
(e) Upper branch mode



(f) Upper branch mode



(g) Lower branch mode

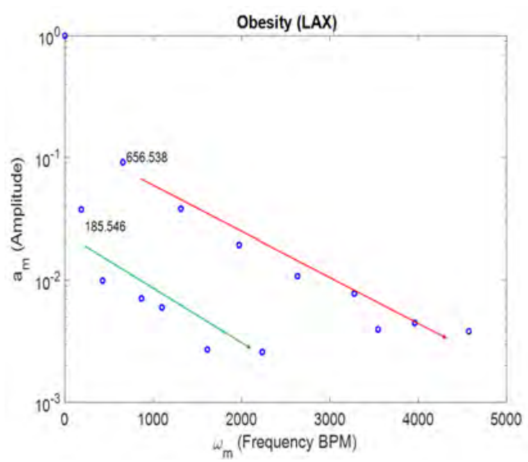


(h) Lower branch mode

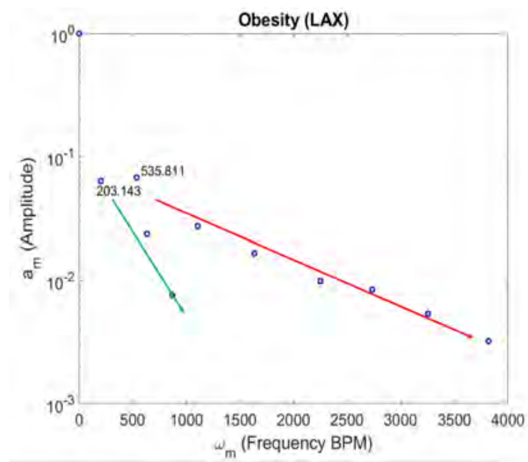
Figure 5.6: The frequencies captured from analyzing the data of the hearts diagnosed with myocardial infarction and the dominant modes related to each branch of frequency. The values of the frequencies from the upper and lower branches can be found in Table 2 in Appendix V.

1- Hypertension models results for LAX data

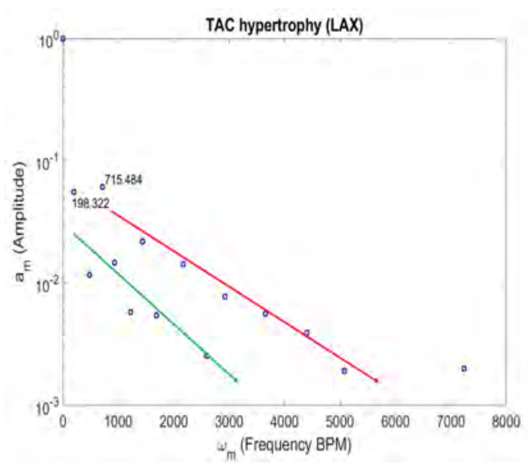
1.1- Captured frequencies:



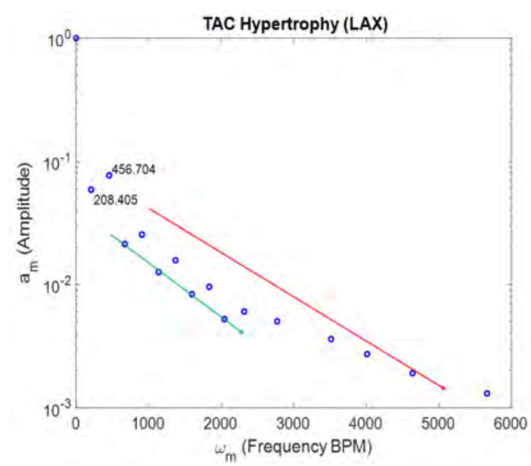
(a) Frequency of the heart with obesity (First dateset)



(b) Frequency of the heart with obesity (Second dataset)



(c) Frequency of the heart with TAC hypertrophy (First dateset)



(d) Frequency of the heart with TAC hypertrophy (Second dataset)

Figure 5.7: The frequencies captured from analyzing the data of the hypertension models (obesity and TAC hypertrophy). The values of the frequencies from the upper and lower branches can be found in Tables 3 and 4 in Appendix V .

1.2- Extracted patterns:

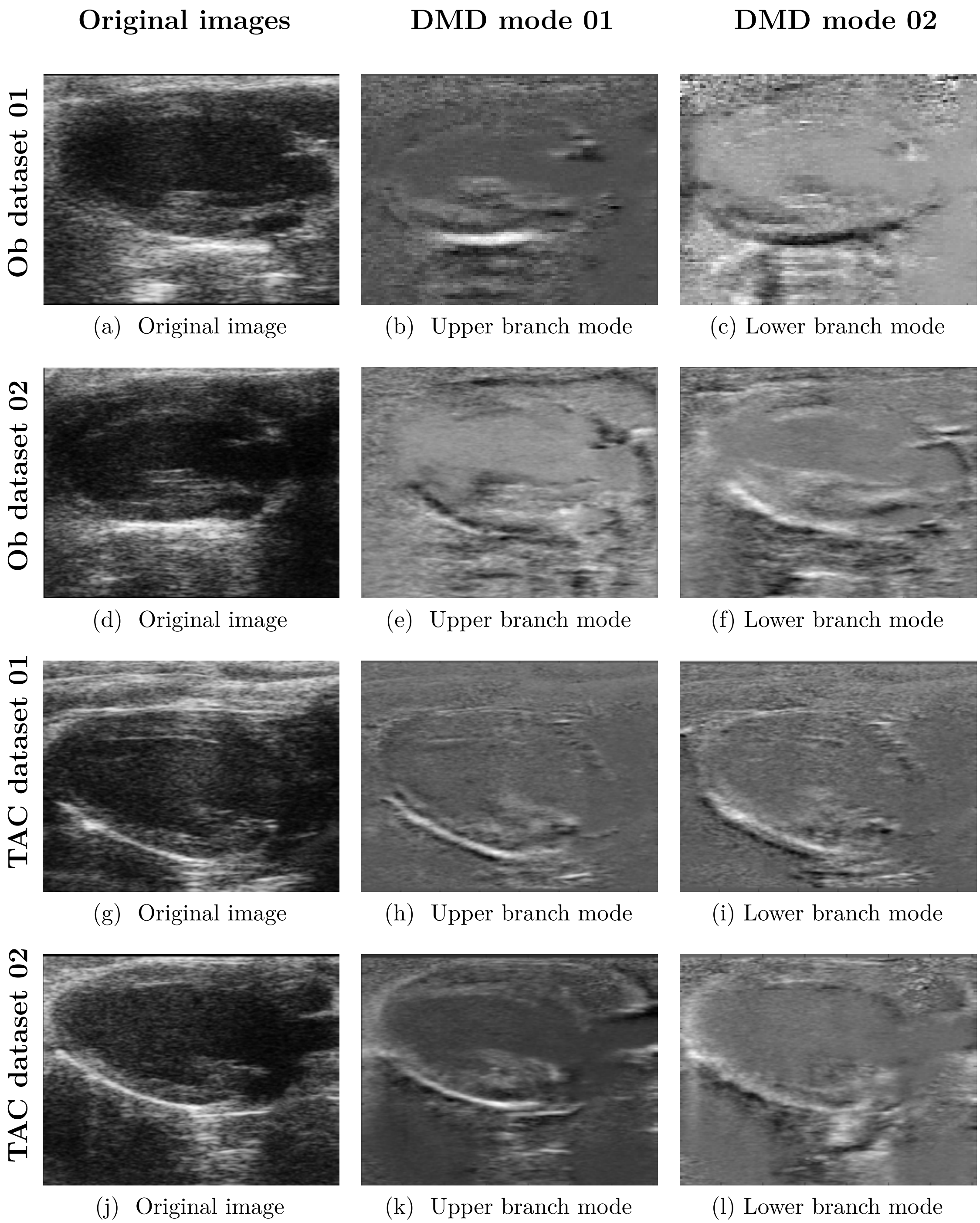
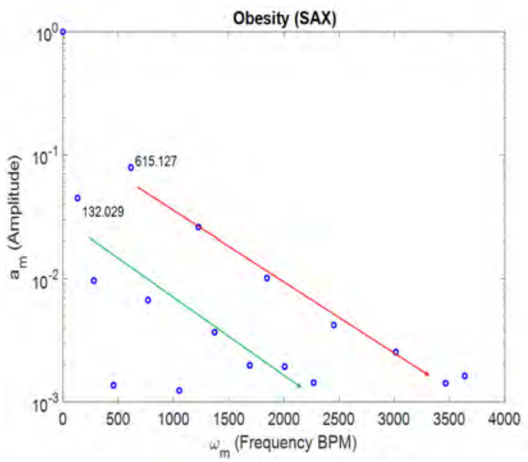


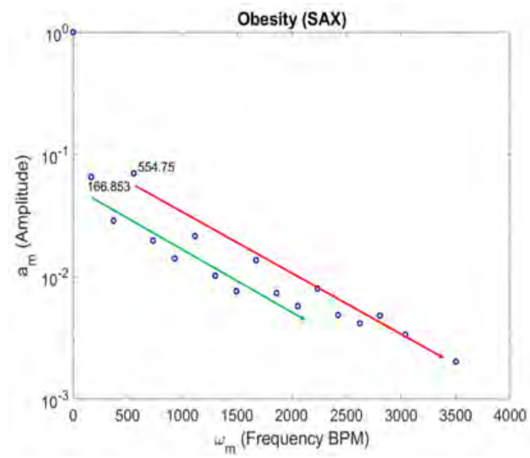
Figure 5.8: The DMD modes obtained from analyzing the LAX datasets of the hypertrophic hearts: obesity (Ob) and TAC hypertension (TAC).

2- Hypertension models results for SAX data

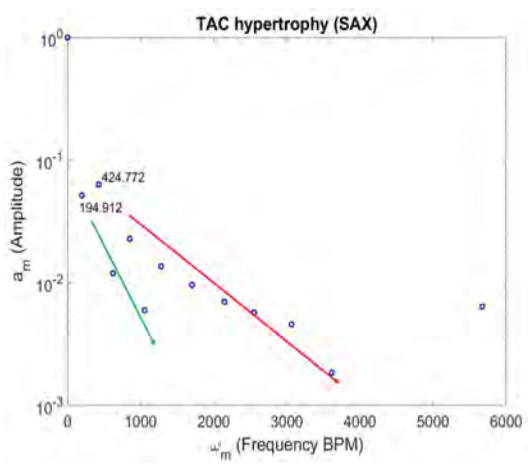
2.1- Captured frequencies:



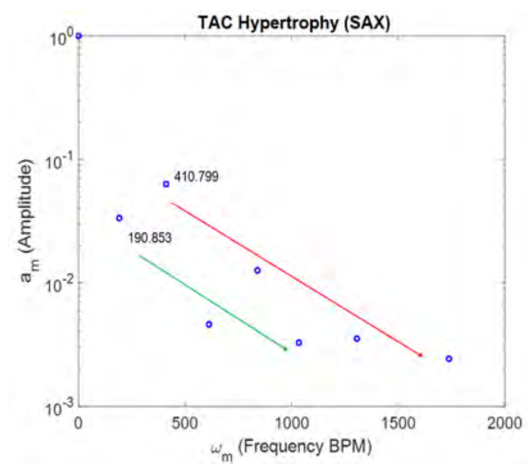
(a) Frequency of the heart with obesity (First dateset)



(b) Frequency of the heart with obesity (Second dataset)



(c) Frequency of the heart with TAC hypertrophy (First dateset)



(d) Frequency of the heart with TAC hypertrophy (Second dataset)

Figure 5.9: The frequencies captured from analyzing the data of the hypertension models (obesity and TAC hypertrophy). The values of the frequencies from the upper and lower branches can be found in Tables 3 and 4 in Appendix V.

2.2- Extracted patterns:

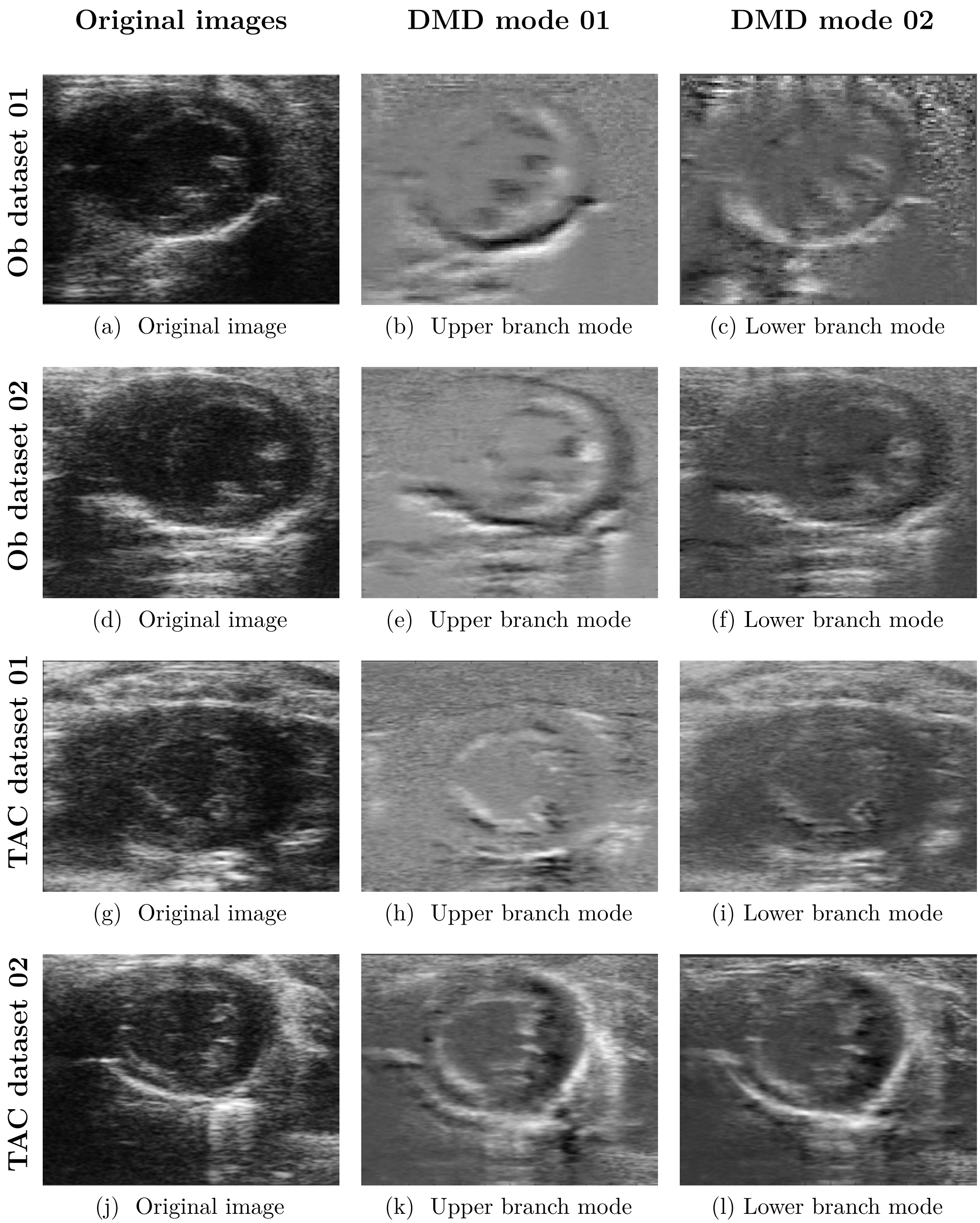
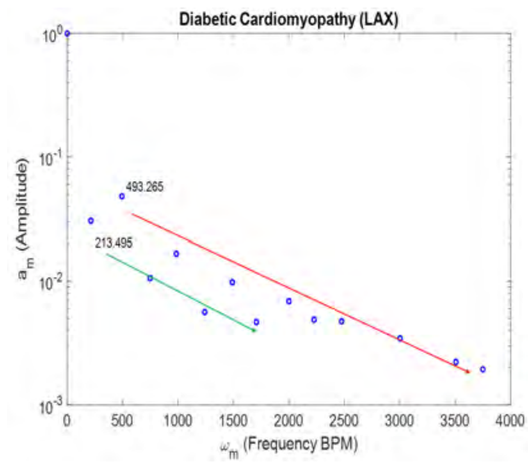
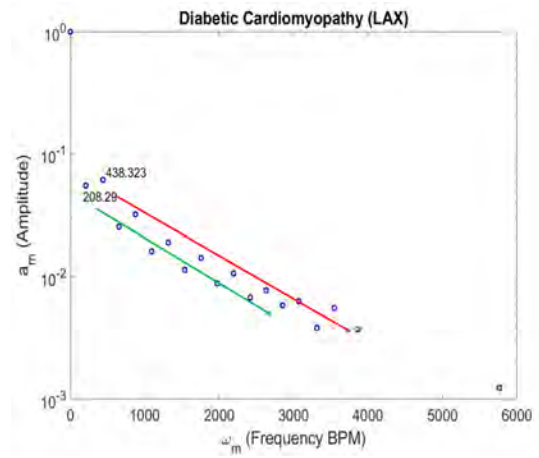


Figure 5.10: The DMD modes obtained from analyzing the SAX datasets of the hypertrophic hearts: obesity (Ob) and TAC hypertension (TAC).

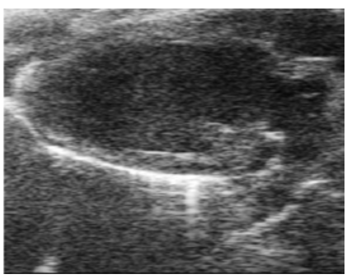
1- Diabetic cardiomyopathy results (LAX)



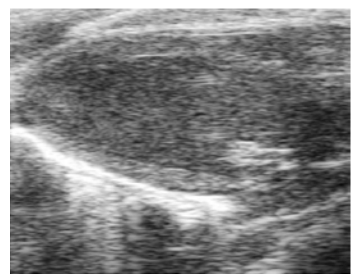
(a) Frequency of the heart with diabetic cardiomyopathy (First dateset)



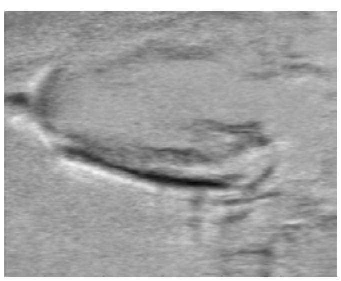
(b) Frequency of the heart with diabetic cardiomyopathy (second dateset)



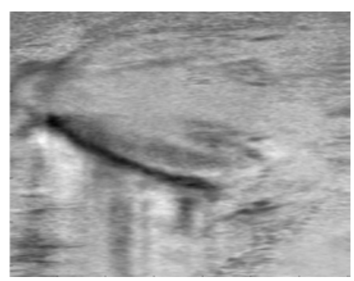
(c) Original image



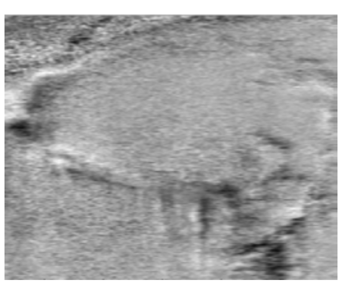
(d) Original image



(e) Upper branch mode



(f) Upper branch mode

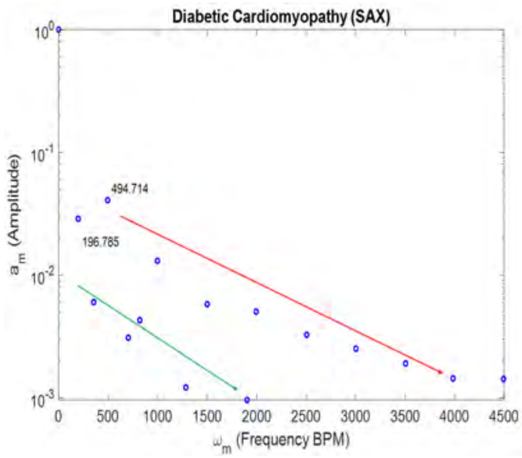


(g) Lower branch mode

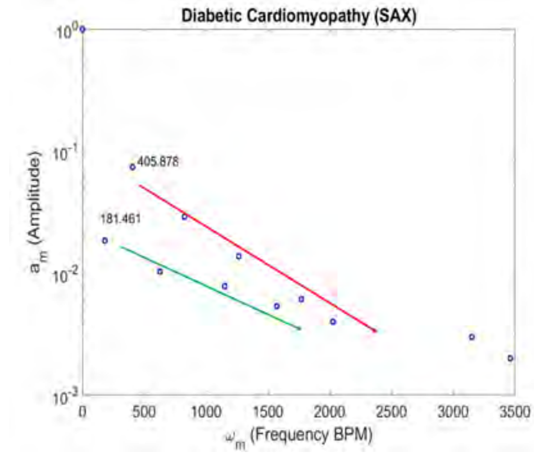
(h) Lower branch mode

Figure 5.11: The frequencies captured from analyzing the data of the hearts diagnosed with diabetic cardiomyopathy and the DMD modes related to each branch of frequencies. The values of the frequencies from the upper and lower branches can be found in Table 5 in Appendix V.

2- Diabetic cardiomyopathy results (SAX)

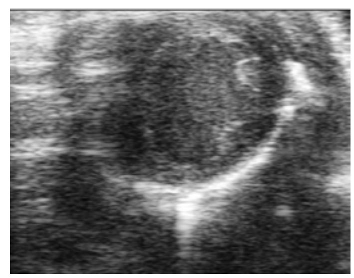


(a) Frequency of the heart with diabetic cardiomyopathy (first dateset)

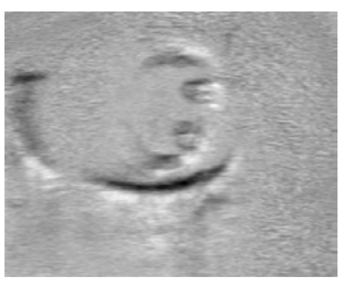


(b) Frequency of the heart with diabetic cardiomyopathy (second dateset)

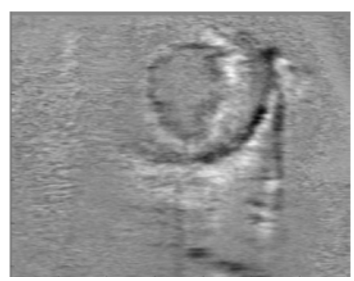
(c) Original image



(d) Original image

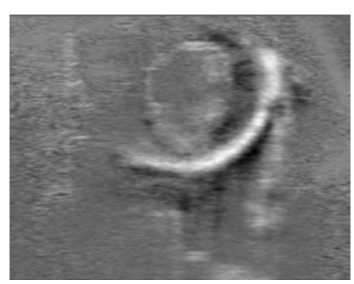


(e) Upper branch mode



(f) Upper branch mode

(g) Lower branch mode



(h) Lower branch mode

Figure 5.12: The frequencies captured from analyzing the data of the hearts diagnosed with diabetic cardiomyopathy and the DMD modes related to each branch of frequencies. The values of the frequencies from the upper and lower branches can be found in Table 5 in Appendix V.

Figures (5.13 and 5.15) shows the original images of some of the additional datasets and their extracted features (DMD modes) in Figures (5.14 and 5.16).

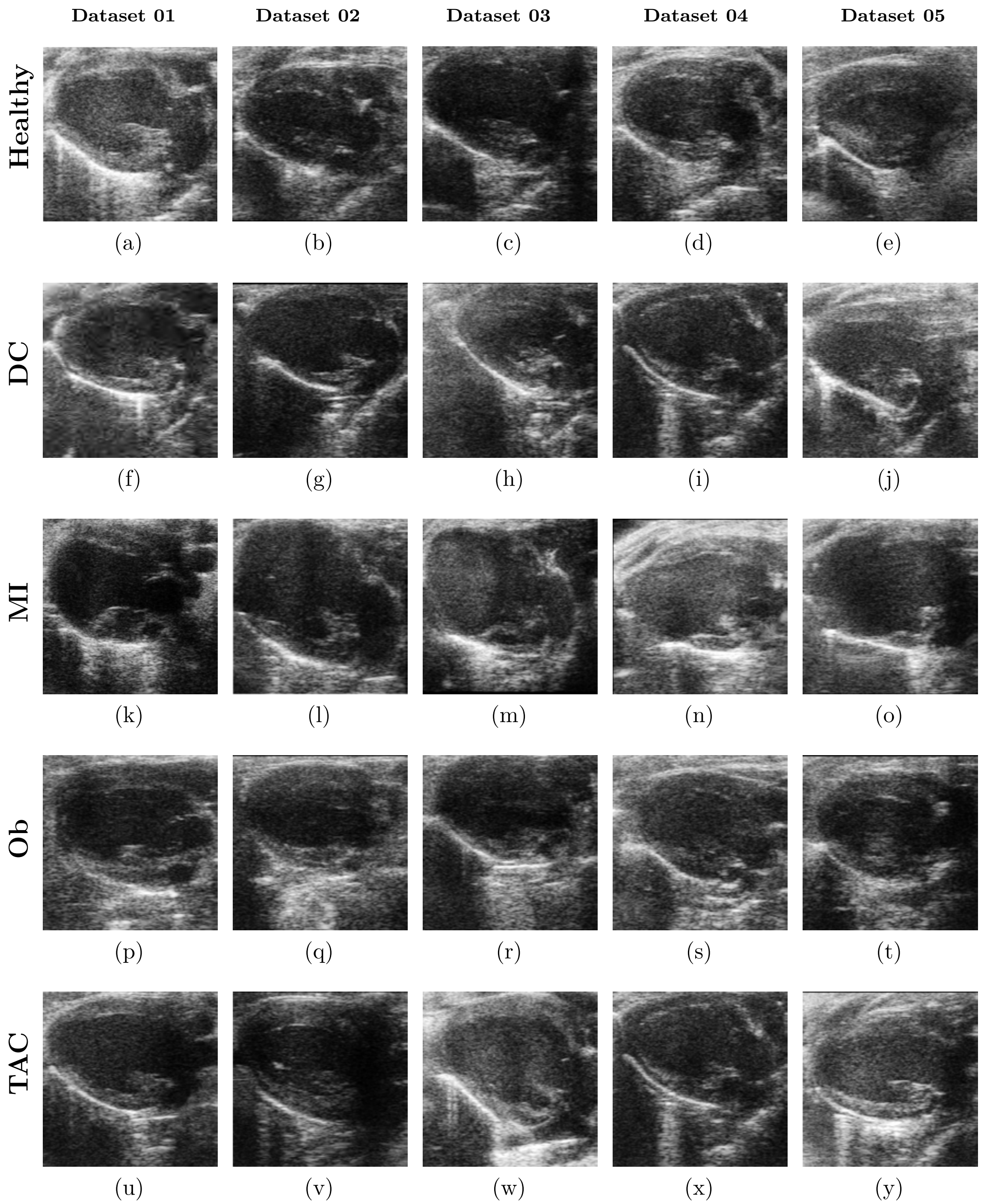


Figure 5.13: Original echocardiography images (LAX) pre-processing. Dc: diabetic cardiomyopathy, MI: myocardial infarction, Ob: obesity, TAC: TAC hypertrophy

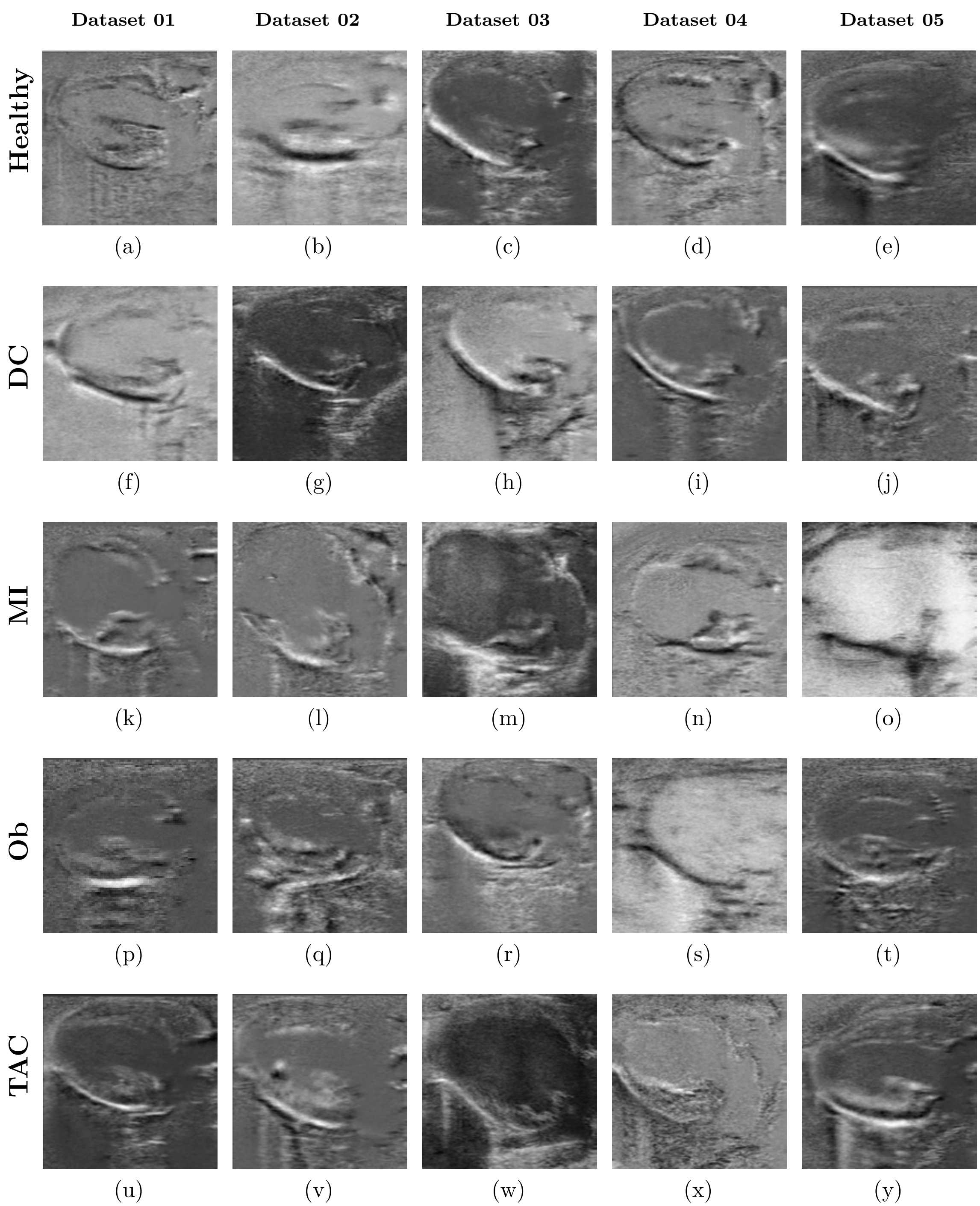


Figure 5.14: Echocardiography images (LAX) post-processing (DMD modes obtained using HODMD as pre-processing step). Dc: diabetic cardiomyopathy, MI: myocardial infarction, Ob: obesity, TAC: TAC hypertrophy.

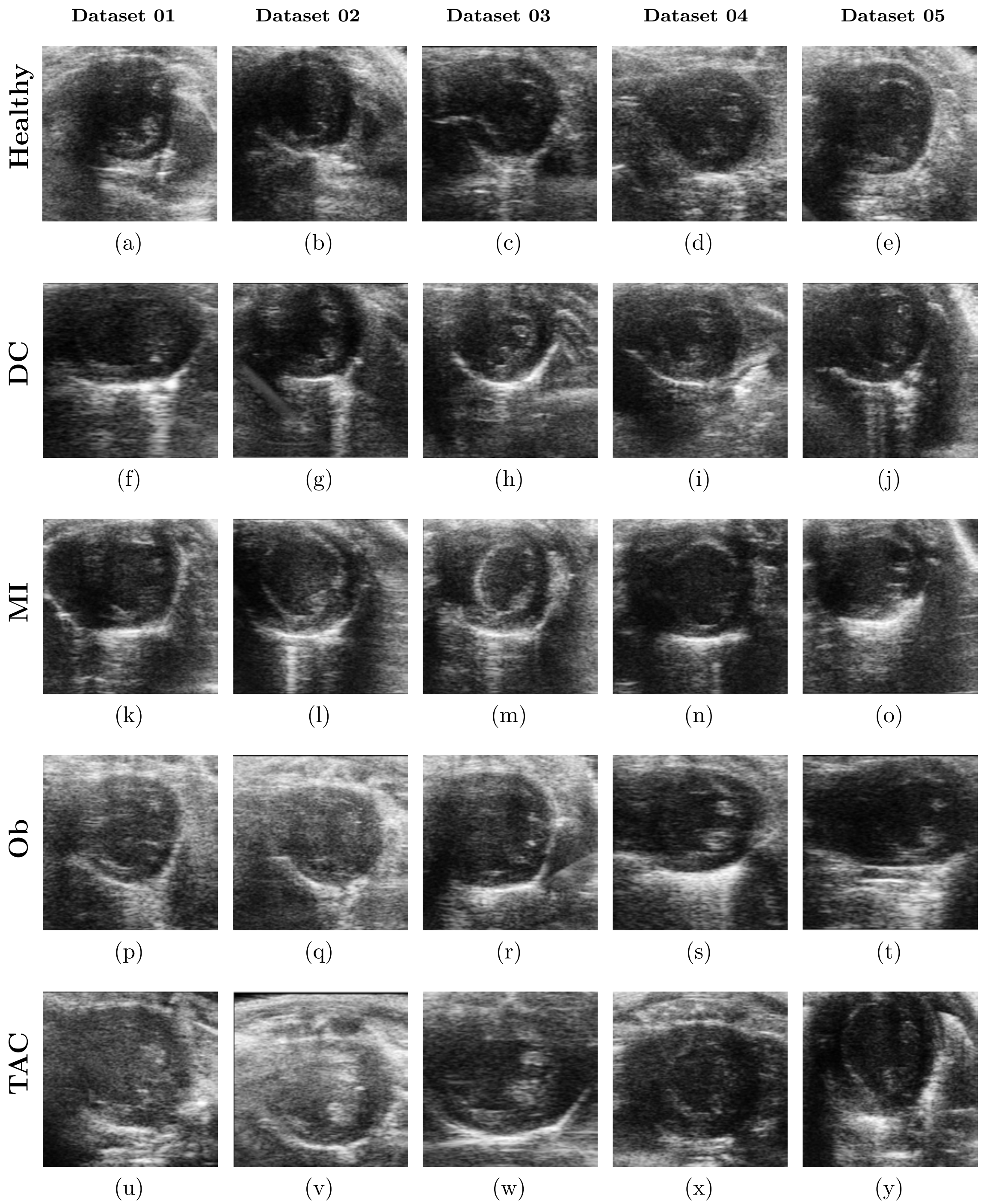


Figure 5.15: Original echocardiography images (SAX) pre-processing. Dc: diabetic cardiomyopathy, MI: myocardial infarction, Ob: obesity, TAC: TAC hypertrophy

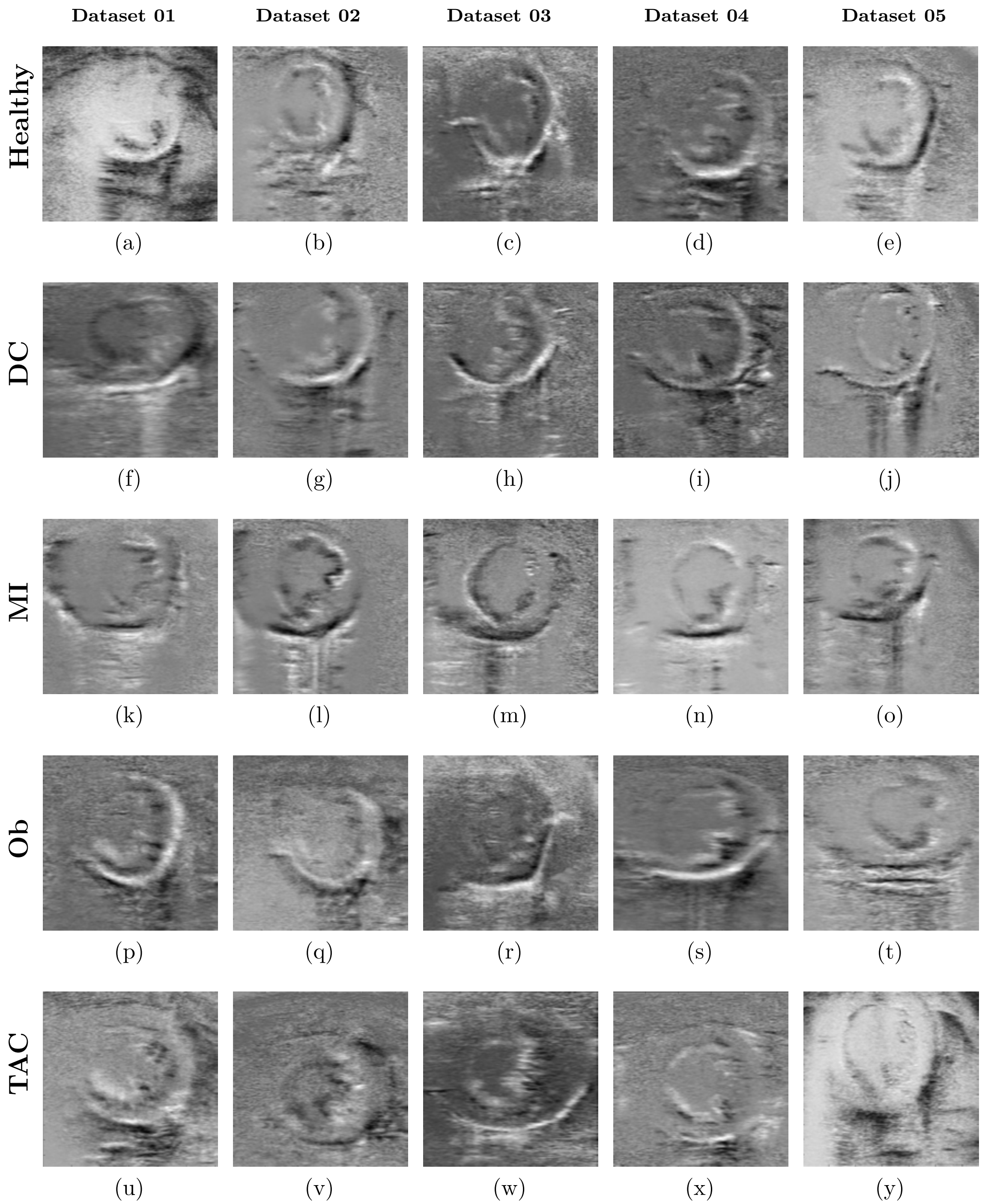


Figure 5.16: Echocardiography images (SAX) post-processing (DMD modes obtained using HODMD as pre-processing step). Dc: diabetic cardiomyopathy, MI: myocardial infarction, Ob: obesity, TAC: TAC hypertrophy

5.2 Conclusions

In this chapter, we have explored the application of the main modal decomposition tool employed. The HODMD algorithm is first used analyse different sets of echocardiography sequences. In particular, echocardiography video loops taken from LAX and SAX views are analyzed. Employing and combining the HODMD algorithm with other tools gave the following main conclusions:

- The capability of the HODMD algorithm to capture and segregate two distinct signals in each dataset. These two signals represent both the heart rate and respiratory rate, which can vary from a dataset to another following several factors, such as the cardiac condition analyzed, dose of anesthesia, as well as gender and age of the mice.
- The HODMD technique, with the help of the HOSVD, was also able to identify and extract main patterns and features related to the different cardiac conditions analyzed. The dominant features were represented in sets of DMD modes, which exhibit the form of the heart in the different cardiac conditions investigated.

The different results obtained from the analysis of the echocardiography databases demonstrate the advantages and benefits of employing different modal decomposition techniques to structure novel approaches capable of covering several applications while delivering competent results.

Chapter 6

Results for MRI datasets

In this chapter we present the results obtained from the analysis of the MRI datasets. In particular, three MRI datasets were analyzed, including one dataset taken from a healthy subject and two taken from mice with Cardiac hypertrophy [1].

Each MRI dataset analyzed is this chapter consist of 10 sequences, where each sequence represents a different slice of the heart. All the sequences consist of 20 snapshots (frames), with 128×128 resolution (more details can be found in section 4.2.1).

6.1 MRI analysis and 3D reconstructions

All the MRI datasets analyzed were first processed as follows: the snapshots of each sequence were imported and organized in an individual tensor as in eq. (2.2). Next, all the tensors were analyzed using the HODMD algorithm with the following parameters:

- The number of snapshots $K = 20$ representing a complete cardiac cycle, as mentioned before.
- The tolerances $\varepsilon_1 = \varepsilon_2 = 5 \times 10^{-4}$, which are related to the noise level and the amplitude truncation, respectively.
- The time step Δt is estimated to be $\Delta t = 8 \times 10^{-3}$.
- Finally, the index d is set to $d = 2$ for all the slices (following the calibration tips presented in Ref. [87]).

Applying HODMD to all the tensors provides the reconstructed slices, which will be joined in one tensor, presented as follows:

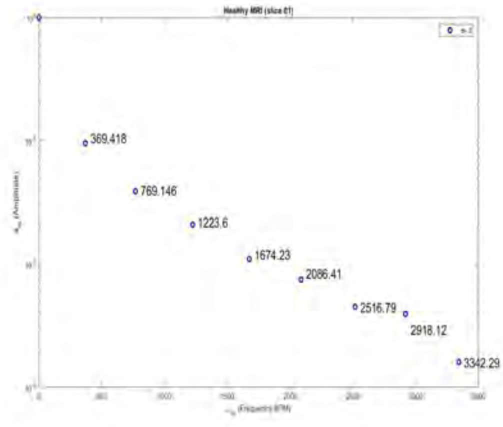
$$\boldsymbol{X}_{x_1,x_2,i,k} \text{ for } x_1 = 1,\ldots,N_x; x_2 = 1,\ldots,N_y; i = 1,\ldots,I \text{ and } k = 1,\ldots,K, \tag{6.1}$$

[1]The results presented in this section (6) have been used to produce a second paper titled "A novel data-driven method for the analysis and reconstruction of cardiac cine MRI" ([132]), which is published in the peer-reviewed, JCR indexed journal Computers in Biology and Medicine (CIBM).

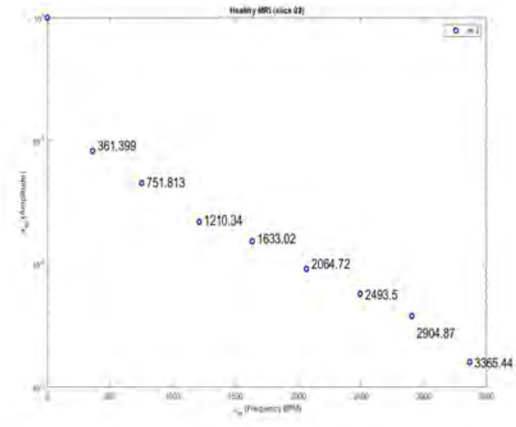
where x_1 and x_2 represent the position of each pixel in the plane containing the image (as defined before), I is the number of slices and k correspond to the snapshot number. This last tensor will be used next to demonstrate the rest of the applications i.e. the 3D visualisation of the heart (as we shall present in 6.1.1 and 6.1.2), HODMD-based reduced order model (ROM) (section 6.2) and the information recovery tool (section 6.3).

6.1.1 Analysis of the healthy data

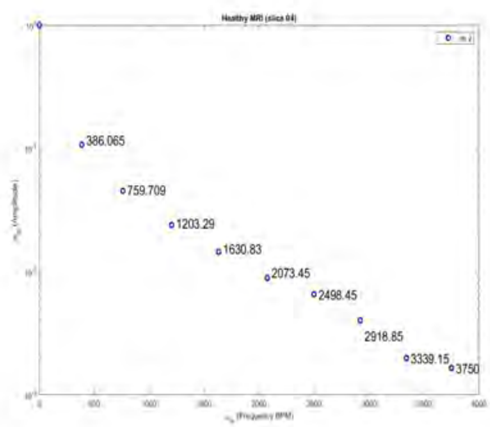
As explained above, HODMD is applied first to each one of the slices separately, and then to all the slices simultaneously (joined in one tensor as presented in eq. (6.1)). As expected, the solution is periodic (as shown in Figures (6.1) and (6.2a)). While analysing the slices individually and simultaneously, the method was able to capture a leading frequency (the frequency related to the heart) confined between 350 and 370 beats per minute (BPM) for all the slices. The differences found between the captured values are mainly connected to the presence of noise. It is worth to mention that the number of snapshots used for this analysis is only 20 snapshots, where using larger number of snapshots will imply obtaining more accurate results, hence, the value of the leading frequency will be similar in all the tests carried out (the values of the captured frequencies for all the slices can be found in Table 6 in appendix V).



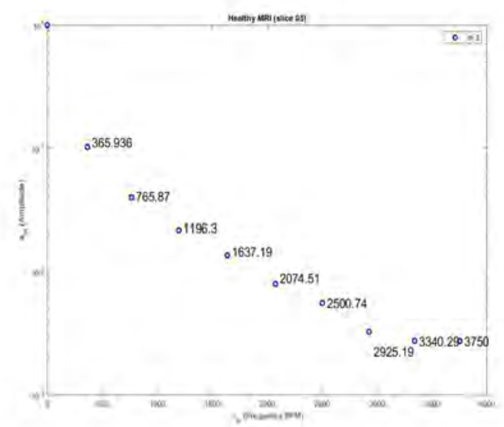
(a)



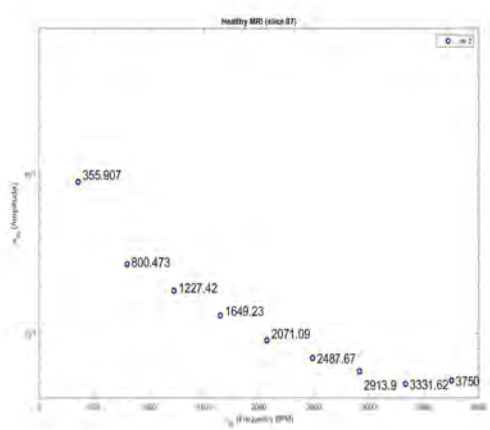
(b)



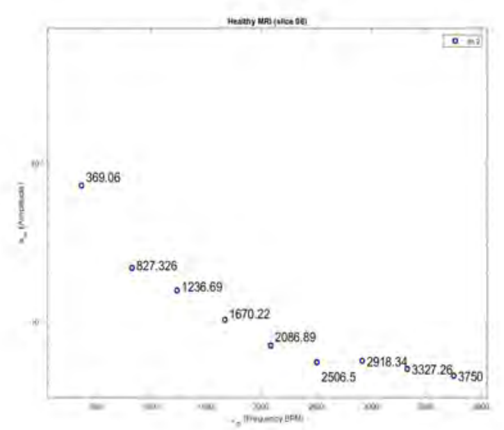
(c)



(d)



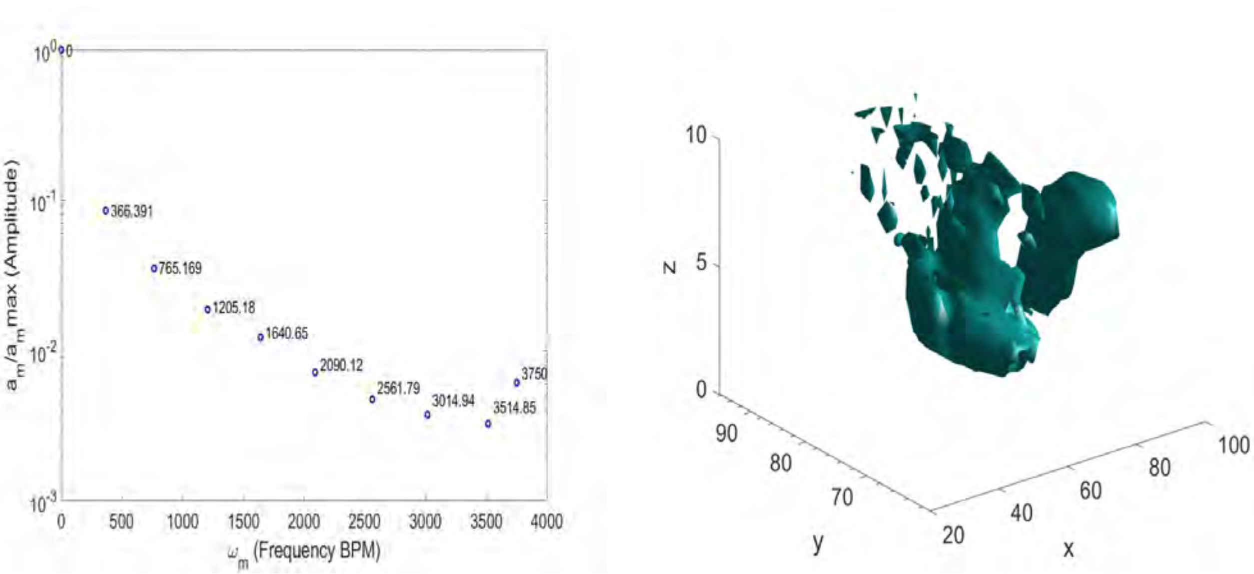
(e)



(f)

Figure 6.1: Frequencies captured by the application of HODMD to the individual MRI sequences for slice 01 (a), slice 02 (b), slice 04 (c), slice 05 (d), slice 06 (e) and slice 08 (f)

HODMD is used next to provide the 3D reconstruction of the heart, as presented in Fig. (6.2b). Analysing the data tensor presented in eq. (6.1), the method identifies the leading frequency as ~ 370 BPM, in good agreement with the results obtained when the algorithm was applied in the original core database (individual slices). Applying the HODMD algorithm to the tensor containing all the slices permits the use of the HODMD extrapolation properties to fill in any gaps between the individual slices. As a consequence, a 3D visualization of the full heart can be carried out. In our case, the 3D representation (shown in Fig. (6.2b)) is carried out using the MATLAB® [133] visualization tool "*isosurface*".



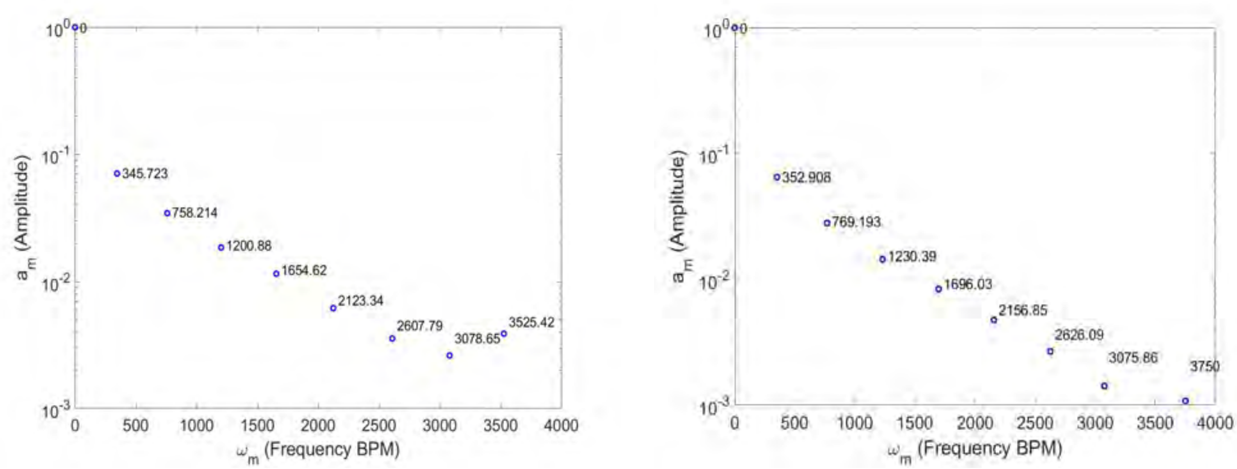
(a) (b)

Figure 6.2: (a): Frequencies identified by HODMD applied to the tensor containing all the slices and (b): 3D reconstruction of the healthy heart.

6.1.2 Analysis of the hypertension datasets

The following results are obtained from analyzing two MRI datasets of the hypertrophic hearts. Similarly to the previous case, the HODMD is used first to analyze all the slices separately, then the tensor containing all the reconstructed slices. Figure (6.3) shows the frequencies captured when analyzing all the slice simultaneity. As shown in the figure, the HODMD technique still captures a regular, periodic signal representing the heart rate, where the leading frequency is still in the range of 350-370 BPM despite of the very limited amount of data, which -ones again- demonstrates the efficiency of the HODMD tool when dealing with limited data.
The frequencies obtained from analyzing all the slices can be found in Tables 7 and 8 in appendix V.

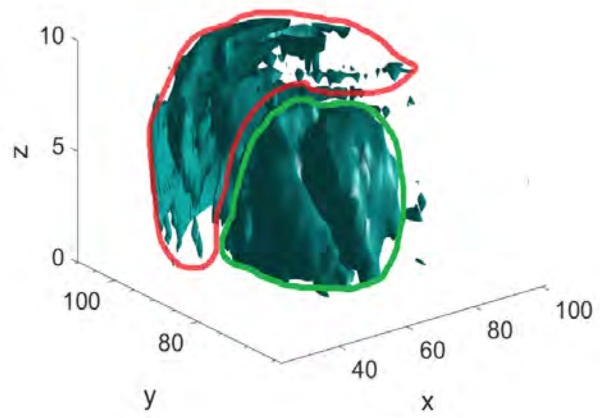
Regarding the 3D reconstruction of these two datasets, a noticeable difference between the healthy and the unhealthy hearts is observed. In particular, as shown in Fig. (6.4), we can notice the enlargement of the hearts diagnosed with hypertrophy when compared with the regular thin shape of the healthy one, as well as the stiffness of the hypertrophic hearts compared with the smooth and regular contraction and relaxation movement of the healthy heart. These observed features are highly related to this disease in particular, which motivate the investigation of other pathologies using 3D reconstructions.



(a) (b)

Figure 6.3: Frequencies identified by HODMD when applied to the tensors containing the reconstructed MRI slices of the first (a) and second (b) hypertrophic hearts.

Nevertheless, as seen in Fig. (6.4), these datasets show the heart (marked in green) as well as its surrounding thoracic cage (marked in red), which may not give the best 3D visualisation of the heart. Hence, a segmentation stage, which enhances the 3D reconstruction, is considered as future work.



(a) (b)

Figure 6.4: 3D reconstruction of the unhealthy heart, (a) 3D layout of the first hypertrophic heart and (b) 3D layout of the second hypertrophic heart.

6.2 HODMD as a reduced order model

In this application we work on improving the 3D reconstruction presented in the previous section. In particular, the HODMD algorithm is used as a reduced order model (ROM) to generate new databases covering more cardiac cycles and providing a 3D reconstruction, varying for longer periods of time.

In order to accomplish this goal, we simply apply the HODMD algorithm to the tensor containing all the slices, but with a change in the temporal term. In more details, we will not apply the DMD expansion defined in Eq. (3.3) to its original temporal term $[t_1, t_K]$, but to a larger one $[t_{K+1}, t_R]$, with $R > K$, exceeding the original number of time steps. As a consequence, the new database generated will consist of 100 snapshots for each slice instead of 20 snapshots. In this way, we will be able to represent the evolution of the 3D reconstruction, with additional reconstructed snapshots covering more than one cardiac cycle.

A comparison between the original snapshots and the new, generated ones for the dataset of the healthy heart is shown in Fig. (6.5). Similar comparison concerning the hypertension datasets can be found in Fig. (1) and Fig. (2) in the appendix V.

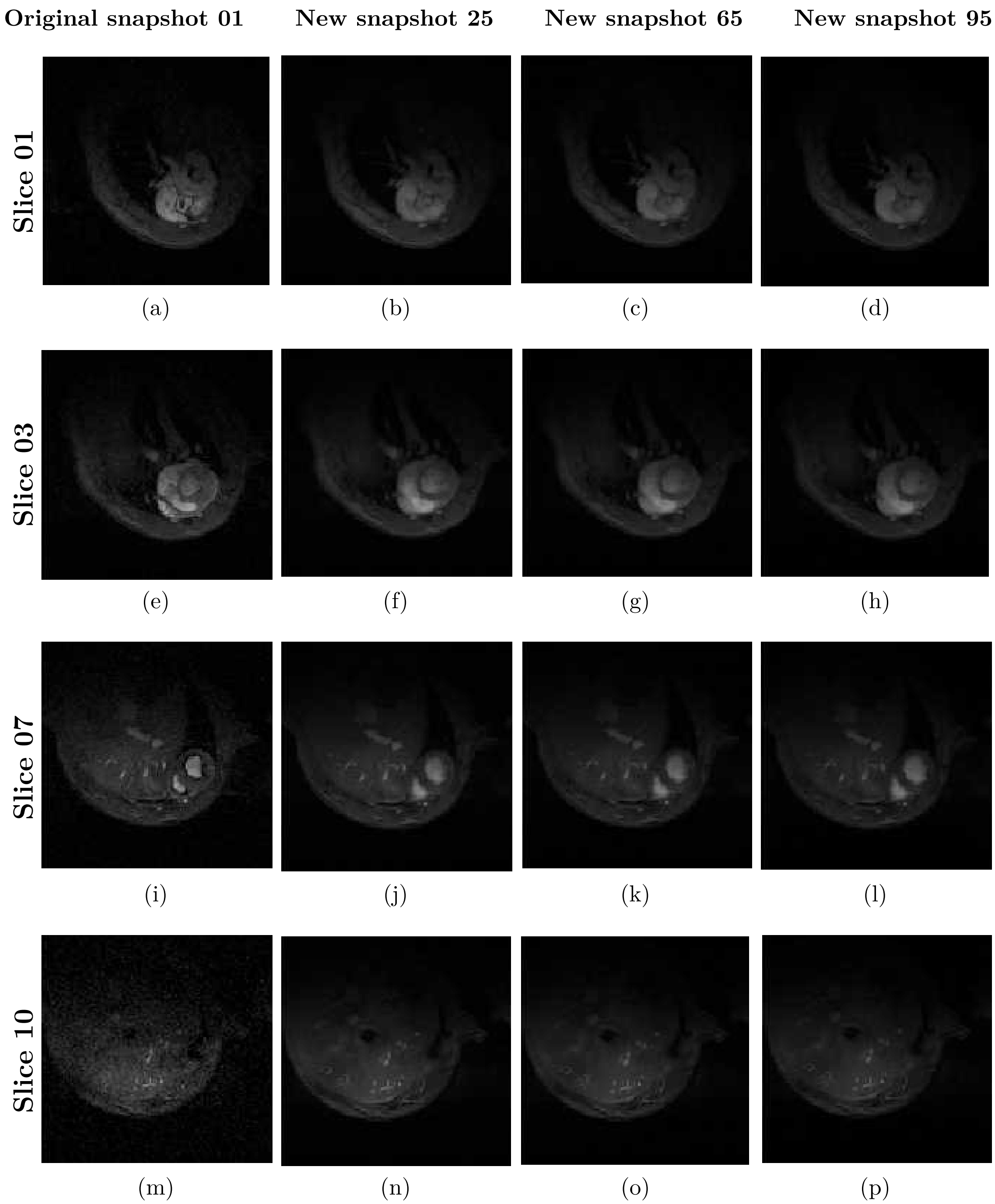


Figure 6.5: Different reconstructed snapshots using a ROM based on HODMD.

6.3 HODMD for information recovery

In the following, HODMD is combined with the interpolation technique introduced in section 3.3. This new method is applied to the tensor containing all the reconstructed slices in order to recover the information of one missing slice (as illustrated previously in Fig. (3.2)). In this contribution, the $7th$ slice has been completely removed, and the two neighboring slices (slice 06 and slice 08) are used to reconstruct all the information of the missing slice.

In more details, for $k = 1, \ldots, 20$, the k-*th* snapshot (frame) from missing slice 07 can be reconstructed by the following procedure:

- First, the k-*th* frames from the neighboring slices 06 and 08 are reshaped into vectors and arranged into a matrix $\boldsymbol{M} \in \mathbb{R}^{16384\times 2}$.
- Next, the matrix $\boldsymbol{M}$ is factorized using SVD as

  $$\boldsymbol{M} = \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{T}^T,$$

  where $\boldsymbol{W} \in \mathbb{R}^{16384\times 16384}$ and $\boldsymbol{T} \in \mathbb{R}^{2\times 2}$ represent the left and right singular vectors, respectively. Meanwhile $\boldsymbol{\Sigma} \in \mathbb{R}^{16384\times 2}$ contains the singular values.
- The information recovery technique is applied next to the matrix of the right singular vectors $\boldsymbol{T}$, where *spline* interpolation (introduced in 2.2.1) is performed between the points of the two lines present in the matrix $\boldsymbol{T} \in \mathbb{R}^{2\times 2}$, providing a new matrix $\tilde{\boldsymbol{T}} \in \mathbb{R}^{3\times 2}$. The new interpolated line is related to the information of the snapshot to be reconstructed.
- Finally, the reconstructed data of the missing snapshot (frame) is retrieved as

  $$\tilde{\boldsymbol{M}} = \boldsymbol{W}\boldsymbol{\Sigma}\tilde{\boldsymbol{T}}^T \in \mathbb{R}^{16384\times 3},$$

  where the first and third columns of $\tilde{\boldsymbol{M}}$ correspond to the k-*th* snapshot of slices 06 and 08, meanwhile the second column correspond to the k-*th* snapshot of slice 07.

A comparison between the snapshots of the original $7th$ and the same snapshots from the new interpolated slice is shown in Fig. (6.6). As seen the images present a qualitative similar shapes and intensities. The noise is the main difference found between the two slices; hence the reconstructed images are clean.

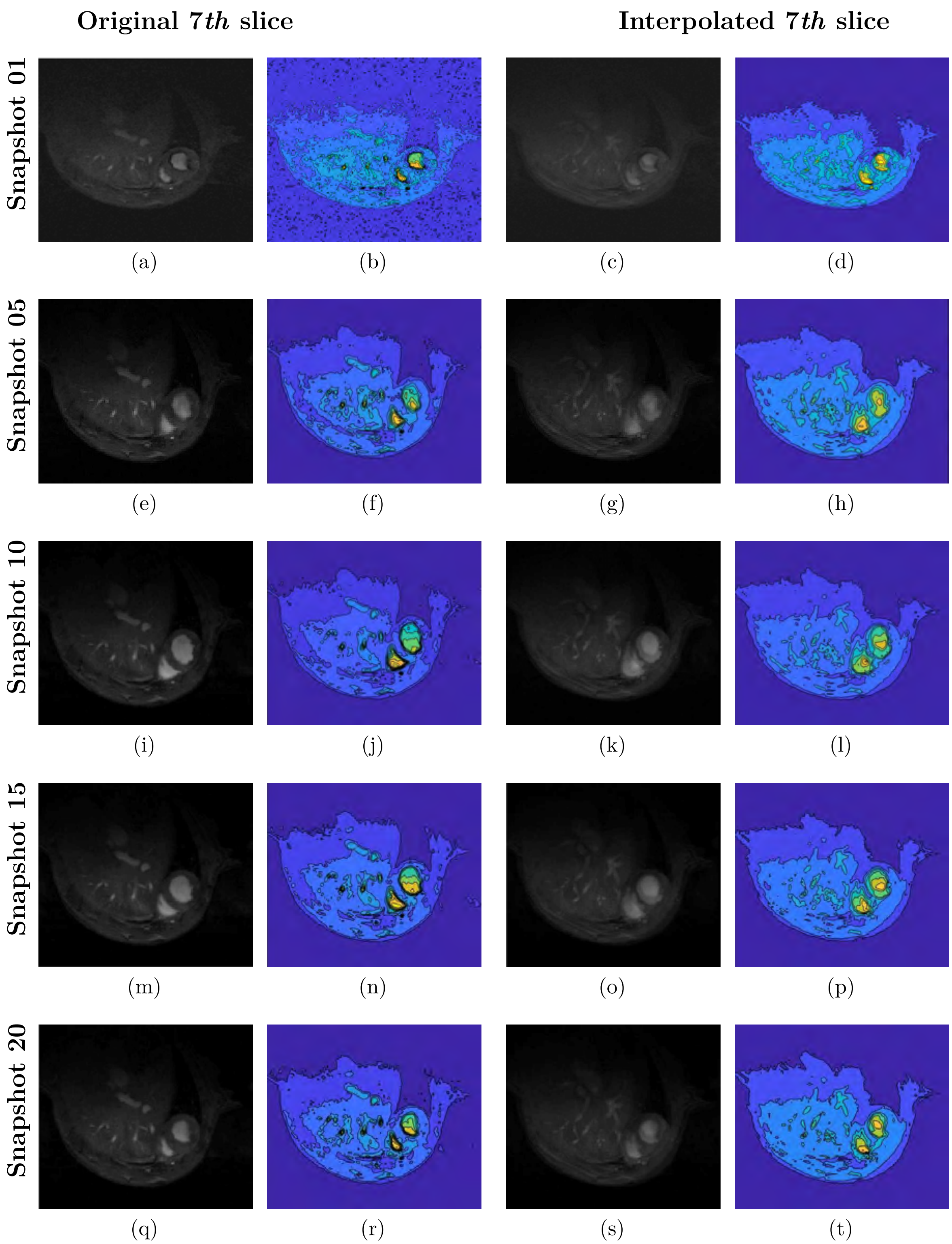
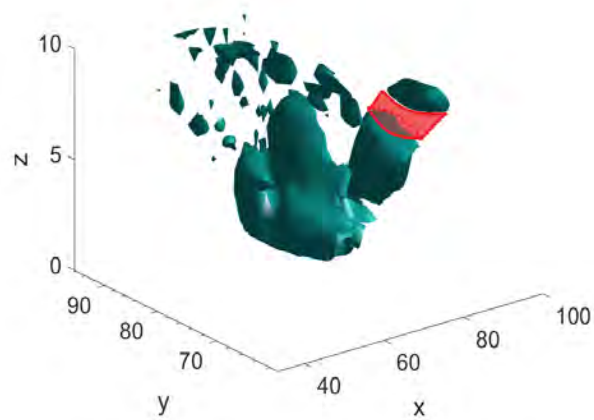
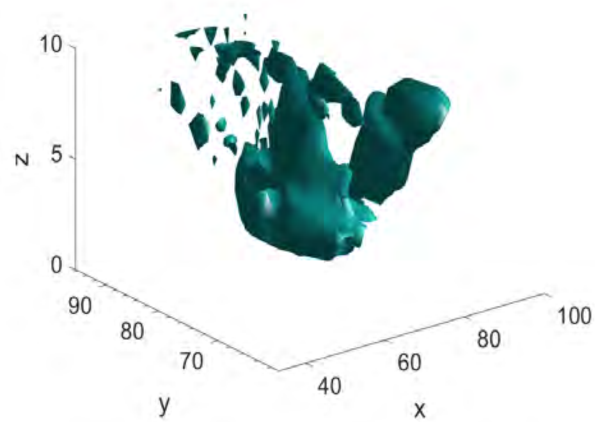


Figure 6.6: A comparison between snapshots 01, 05, 10, 15 and 20 taken from original slice 07 and the interpolated slice 07. first and second columns correspond to the original snapshot, and third and fourth columns to the reconstruction. The second and fourth columns present the intensity of the images (included for clearer comparison).

As a final step on this section, the 3D reconstruction of the heart is used to evaluate the SVD based interpolation tool. Figure (6.7) shows the 3D reconstruction of the heart with the missing slice (Fig. (6.7a)), as well as the 3D reconstruction with the new interpolated slice (Fig. (6.7b)). The reconstruction error between the 3D layouts of the original data and the one with the recovered slice was calculated using the RRMSE (defined in eq. (3.21)) and estimated to be 0.09. Meanwhile the error between the original slice and the interpolated one is 0.2.
It is worth to mention that, alongside using *spline* interpolation, several other interpolation techniques have been tested. In particular, *kriging*, *linear*, *cubic*, *pchip* and *makima* delivered similar reconstructions to the ones presented in this section, with the same reconstruction error. Meanwhile, *nearest*, *next*, *previous* showed poor performance, increasing the error to 0.12. Hence, it is highly recommended to explore using other type of interpolation techniques (e.g. machine learning based methods) in future works to reduce this error, but this remains as an open topic.



(a)



(b)

Figure 6.7: 3D reconstruction of the heart before and after the missing information recovery step. (a) 3D visualization of the heart after removing the information of the $7th$ slice marked in red. (b) 3D visualization of the heart with the interpolated $7th$ slice.

Similar test has been carried out in slices 02, 05 and 09, where the RRMSE for the 3D reconstruction was always smaller than 0.09, showing the robustness of this methodology. Similarly to the previous test, the error between the original slices and the interpolated ones is always set to 0.2 (see Table 6.1). The main source if this error is the noise and surrounding areas of the heart. Hence, cleaning the original data and performing a segmentation phase to exclude the surrounding regions can significantly reduce the reconstruction error (these proposals are considered as future works).

Reconstructed slice	RRMS Error:	
	Between slices	Between 3D reconstructions
Slice 02	0.2	0.08
Slice 05	0.2	0.07
Slice 09	0.2	0.08

Table 6.1: Estimated RRMS error. The second column shows the error between the original and the reconstructed slices. The third column shows the error between the 3D reconstruction with the original slice and 3D reconstruction with the reconstructed slice.

Additional results concerning the other tested slices can be found in Figures (3) and (4) in the Appendix V.

6.4 Conclusions

For this section the second medical dataset is investigated. In particular, we once again explore different applications of the modal decomposition techniques. but with MRI datasets (10 sequences per dataset representing 10 different slices of the heart). This MRI databases correspond to two different cardiac conditions (healthy and hypertension). The analysis of this databases gave the following results:

- First, each MRI sequence is analyzed separately using the HODMD algorithm, which captures and provides the signals representing the heart rates. The leading frequency captured is always between 350-370 BPM, which demonstrates the robustness of the HODMD technique and its efficiency when dealing with limited data.
- The MRI sequences were first analyzed individually then simultaneously, which allows us to visualize our database in a 3D format displaying the different hearts studied.
- Another application investigated in this section is employing the HODMD tool as a reduced order model. Particularly, with some changes in the HODMD parameters, the original MRI sequences, which had 20 frames per sequence (covering one cardiac cycle), are extended to include 100 frame per sequence. The new generated images expanded the databases to cover approximately 3 full cardiac cycles.
- The 3D visualisation proposed in the previous application is used to evaluate an SVD-based interpolation tool introduced for the reconstruction of corrupted images. In more details, a full MRI sequence (representing a specific slice of the heart) is fully removed, where the proposed algorithm employs the two neighboring MRI sequences to reconstruct the missing one. The technique shows promising results with an error of 0.09 for the full 3D reconstruction.

Once again, we see the modal decomposition tools covering a wide range of applications, showing a promising future for these techniques in the medical field.

Part IV

Conclusions and future works

Summary This last part includes the general dissertation of the thesis with a statement of the conclusions reached while pursuing the PhD. It also discusses a number of possible avenues to continue and improve the results of the research.

Chapter 7

Conclusions and future works

7.1 Conclusions

This contribution represents the research work conducted by the author over three academic years. The major purpose of this research has been to introduce hybrid ROMs based on physical principles. These ROMs employ data-driven modal decomposition techniques and machine learning tools to develop novel and efficient approaches. The main tool utilized is the fluid dynamic tool, Higher Order Dynamic Mode Decomposition (HODMD). This algorithm has shown numerous of times its abilities, robustness and efficiency in several application in the fluid mechanics industry. As a consequence, we decided to exploit the properties of this algorithm as well as other modal decomposition tools to cover many complex applications in medicine.

This thesis includes eight chapters. Starting from a general introduction (chapter 1), which covers our research background, state of the art, the main objectives of the research and the resultant scientific dissemination. Chapters 2 and 3, present the preliminaries and the fundamental methodologies employed in the investigation, respectively. The three following chapters include a thorough description of the datasets used (chapter 4) and a detailed elaboration on the main results obtained from the analysis of the echocardiography datasets (chapter 5) and the MRI databases (chapter 6). In the following (chapter 7), we summarize the main conclusions and report and report several future works. We introduce last a final part (chapter V) gathering the appendices.

The main applications covered in this investigation have been summarized as follows:

- The first application has been explored is the analysis of echocardiography video loops. The echocardiography sequences, which were taken with respect to a long axis view (LAX) and a short axis view (SAX), are analyzed using the HODMD algorithm, in order to capture relevant frequencies and obtain useful information and insights. Furthermore, this analysis was aimed at another essential goal, which is the identification and extraction of dominant features and patterns related to different cardiac conditions.

- For the second application, a different dataset was investigated. The HODMD was used this time to analyze magnetic resonance imaging (MRI) sequences. The analysis included the identification of relevant frequencies, as well as building 3D reconstructions displaying the movement of the heart.
- The following application exploits the properties of the HODMD algorithm as a reduced order model (ROM). In particular, the ROM has been used to extend the original MRI sequences and generate new databases covering additional cardiac cycles.
- The last application introduced a novel technique for information recovery. In more details, the new tool combined the HODMD algorithm with an SVD based interpolation technique to recover the missing information of corrupted MR images.

Main research outcomes obtained from these applications are:

- The analysis of the echocardiography datasets using the HODMD algorithm was done first. The proposed technique was able to identify and segregate two different signals. The signals, which are regular and periodic, represent the heart rate and respiratory rate. These captured rates were always varying between 400 to 700 beats per minute and 150 to 250 breaths per minute, which are considered ordinary for mice. The frequencies captured as well as the number of modes identified in each case is different, but it is normal and expected as it can be affected by several factors such as the noise level in the dataset, dose of anesthesia, gender, age, cardiac conditions etc ....
- The HODMD technique successfully identified and extracted sets of DMD modes representing the dominant features and patterns related to the different cardiac conditions. These DMD modes display the shape of the heart in its healthy conditions, as well as its shape when afflicted by different cardiac diseases. Furthermore, based on the patterns extracted, the DMD modes obtained from the unhealthy hearts can be sorted into three main categories: (i) modes representing the patterns that characterize the myocardial infarction (such as the myocardial wall deformation), obtained from analysing the myocardial infarction datasets, (ii) modes representing the patterns that characterize different types of hypertrophy (circular shape of the hear), obtained from analyzing obesity and TAC hypertrophy datasets and (iii) diabetic cardiomyopathy models, obtained by analyzing diabetic cardiomyopathy datasets, which did not show any abnormalities or any certain feature that might represent this specific cardiac disease. All these presented results have demonstrated that the proposed method has a remarkable ability in capturing spatio-temporal patterns from different, limited and noisy data.
- Similarly to the previous dataset, analyzing the MRI datasets has also confirmed the efficiency of the HODMD algorithm when dealing with limited data, where analyzing each slice, which contains only 20 snapshots, HODMD was able to capture the heart rate for each slice. In all cases, the leading frequency has always been between 350-370, proving the robustness of the algorithm. Moreover, by applying HODMD to all the MRI sequences of each dataset individually then all the slices simultaneously, a 3D visualisation of the heart is provided.

- An HODMD-based reduced order model (ROM) of the heart is next developed. As mentioned earlier, the original dataset consists of 20 snapshots per sequence, however, using the HODMD technique as a ROM, allowed us to generate a new database consisting of 100 snapshots for each sequence. Furthermore, these additional generated snapshots were used to extend the evolution of a 3D reconstruction of the heart, providing a 3D visualisation of the heart, covering more than one cardiac cycle.
- The SVD-based interpolation algorithm introduced was used to reconstruct corrupted images and recover missing information from several different slices. The algorithm presented was capable of satisfactorily reconstructing a completely missing slice from the information of neighboring slices. The technique was tested on 4 different slices, giving similar results with an error never exceeding 0.09 for the full 3D reconstruction.

All the applications and results presented in this contribution highlight the wide range of applications of the different modal decomposition algorithms. These technique have proven to be very powerful and useful in building efficient ROMs, not only for fluid mechanics and aerospace engineering problems, but also for complex applications in the medical field. Hence, this shows the great potentials of the proposed methods to extend further and cover more applications in other industries.

The data has been treated and analyzed using codes implemented in MATLAB.

All the data have been obtained from previous studies performed in accordance with protocols approved by the CNIC's Institutional Animal Care and Research Advisory Committee of the Ethics Committee of the Regional Government of Madrid (PROEX177/17).

7.2 Future works

The presented thesis has made important progresses in the application of data-driven modal decomposition techniques in the medical field. However, while investigating this field, we have faced several difficulties and limitations, which would be addressed in future works. The followup activities mainly include two points:

- The 3D visualisation of the heart is potentially useful in many applications in the medical field. Even though the HODMD permits these 3D reconstructions, the nature of the datasets does not consistently provide the best visualisation of the heart. As a consequence, a segmentation process should be considered to eliminate the surrounding thoracic cage and extract only the heart for a better visualisation of the shape of the different phases of the cardiac cycles investigated.
- The third and last point concerns the information recovery technique. As stated in the third section of chapter 6, several interpolation techniques were tested to reduce the reconstruction error. However, the results attained could be improved. We shall consider both more developed techniques (machine learning based) as well as segmentation stages

for a better reconstruction of the recovered images.

References

[1] Carol Mitchell, Peter S Rahko, Lori A Blauwet, Barry Canaday, Joshua A Finstuen, Michael C Foster, Kenneth Horton, Kofo O Ogunyankin, Richard A Palma, and Eric J Velazquez. Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the american society of echocardiography. *Journal of the American Society of Echocardiography*, 32(1):1–64, 2019. Cited on page(s) (document), 4.1.

[2] Wilhelmus HA Schilders, Henk A Van der Vorst, and Joost Rommes. *Model order reduction: theory, research aspects and applications*, volume 13. Springer, 2008. Cited on page(s) 1.

[3] John W Tukey. The future of data analysis. *The Annals of Mathematical Statistics*, 33(1):1–67, 1962. Cited on page(s) 1.1.

[4] Daniel M Ringel and Bernd Skiera. Visualizing asymmetric competition among more than 1,000 products using big search data. *Marketing Science*, 35(3):511–534, 2016. Cited on page(s) 1.1.

[5] Bruno JD Jacobs, Bas Donkers, and Dennis Fok. Model-based purchase predictions for large assortments. *Marketing Science*, 35(3):389–404, 2016. Cited on page(s) 1.1.

[6] Farhad Foroughi and Peter Luksch. Data science methodology for cybersecurity projects. *arXiv preprint arXiv:1803.04219*, 2018. Cited on page(s) 1.1.

[7] Iqbal H Sarker, ASM Kayes, Shahriar Badsha, Hamed Alqahtani, Paul Watters, and Alex Ng. Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data*, 7(1):1–29, 2020. Cited on page(s) 1.1.

[8] Bhavani Thuraisingham, Murat Kantarcioglu, Kevin Hamlen, Latifur Khan, Tim Finin, Anupam Joshi, Tim Oates, and Elisa Bertino. A data driven approach for the science of cyber security: Challenges and directions. In *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, pages 1–10. IEEE, 2016. Cited on page(s) 1.1.

[9] Carson K Leung and Kyle W Joseph. Sports data mining: predicting results for the college football games. *Procedia Computer Science*, 35:710–719, 2014. Cited on page(s) 1.1.

[10] Adam Maszczyk, Artur Gołaś, Przemysław Pietraszewski, Robert Roczniok, Adam Zając, and Arkadiusz Stanula. Application of neural and regression models in sports results prediction. *Procedia-Social and Behavioral Sciences*, 117:482–487, 2014. Cited on page(s) 1.1.

[11] John McCullagh et al. Data mining in sport: A neural network approach. *International Journal of Sports Science and Engineering*, 4(3):131–138, 2010. Cited on page(s) 1.1.

[12] Elaine A Ryan and Michael J Farquharson. Breast tissue classification using x-ray scattering measurements and multivariate data analysis. *Physics in Medicine & Biology*, 52(22):6679, 2007. Cited on page(s) 1.1.

[13] Weiping Zhang, Jingzhi Yang, Yanling Fang, Huanyu Chen, Yihua Mao, and Mohit Kumar. Analytical fuzzy approach to biological data analysis. *Saudi Journal of Biological Sciences*, 24(3):563–573, 2017. Cited on page(s) 1.1.

[14] Jyoti Islam and Yanqing Zhang. Brain mri analysis for alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Informatics*, 5(2):1–14, 2018. Cited on page(s) 1.1.

[15] Julio M Ottino. Complex systems. *American Institute of Chemical Engineers. AIChE Journal*, 49(2):292, 2003. Cited on page(s) 1.1.

[16] Xiaofeng Zhang, Nadine Smith, and Andrew Webb. Medical imaging. In David Dagan Feng, editor, *Biomedical Information Technology*, Biomedical Engineering, pages 3–27. Academic Press, Burlington, 2008. Cited on page(s) 1.1.

[17] K.J. Cios, K. Chen, and R.A. Langenderfer. Use of neural networks in detecting cardiac diseases from echocardiographic images. *IEEE Engineering in Medicine and Biology Magazine*, 9(3):58–60, 1990. Cited on page(s) 1.1.1.

[18] GW Gross, JM Boone, V Greco-Hunt, and B Greenberg. Neural networks in radiologic diagnosis. ii. interpretation of neonatal chest radiographs. *Investigative Radiology*, 25(9):1017—1023, September 1990. Cited on page(s) 1.1.1.

[19] Ronald H. Silverman and Andrew S. Noetzel. Image processing and pattern recognition in ultrasonograms by backpropagation. *Neural Networks*, 3(5):593–603, 1990. Cited on page(s) 1.1.1.

[20] AS Miller, BH Blott, and TK Hames. Neural networks for electrical impedance tomography image characterisa tion. *Clinical Physics and Physiological Measurement*, 13(A):119, 1992. Cited on page(s) .

[21] E. I. Mohamed, C. Maiolo, R. Linder, S. J. Pöppl, and A. De Lorenzo. Artificial neural network analysis: a novel application for predicting site-specific bone mineral density. *Acta Diabetologica*, 40(S1):s19–s22, October 2003. Cited on page(s) 1.1.1.

[22] C.-T. Chen, E.C.-K. Tsao, and W.-C. Lin. Medical image segmentation by a constraint satisfaction neural network. *IEEE Transactions on Nuclear Science*, 38(2):678–686, 1991. Cited on page(s) 1.1.1.

[23] J. Alirezaie, M.E. Jernigan, and C. Nahmias. Automatic segmentation of cerebral mr images using artificial neural networks. *IEEE Transactions on Nuclear Science*, 45(4):2174–2182, 1998. Cited on page(s) 1.1.1.

[24] Jzau-Sheng Lin. Segmentation of medical images through a penalized fuzzy hopfield netw ork with moments preservation. *Journal of The Chinese Institute of Engineers*, 23(5):633–643, 2000. Cited on page(s) .

[25] Ian Middleton and Robert I. Damper. Segmentation of magnetic resonance images using a combination of neural networks and active contour models. *Medical Engineering & Physics*, 26(1):71–86, January 2004. Cited on page(s) .

[26] S.R. Kannan, S. Ramathilagam, R. Devi, and A. Sathya. Robust kernel fcm in segmentation of breast medical images. *Expert Systems with Applications*, 38(4):4382–4389, 2011. Cited on page(s) 1.1.1.

[27] Hiroshi Fujita, Tetsuro Katafuchi, Toshiisa Uehara, and Tsunehiko Nishimura. Application of artificial neural network to computer-aided diagnosis of coronary artery disease in myocardial spect bull's-eye images. *Journal of Nuclear Medicine*, 33(2):272–276, 1992. Cited on page(s) 1.1.1.

[28] I Crooks and BG Fallone. A novel algorithm for the edge detection and edge enhancement of medical images. *Medical Physics*, 20(4):993–998, 1993. Cited on page(s) 1.1.1.

[29] Chuan-Yu Chang. Two-layer competitive based Hopfield neural network for medical image edge detection. *Optical Engineering*, 39(3):695, March 2000. Cited on page(s) 1.1.1.

[30] H Tang, E.X Wu, Q.Y Ma, D Gallagher, G.M Perera, and T Zhuang. Mri brain image segmentation by multi-resolution edge detection and region selection. *Computerized Medical Imaging and Graphics*, 24(6):349–357, 2000. Cited on page(s) 1.1.1.

[31] I.A. Hunter, J.J. Soraghan, J. Christie, and T.S. Durrani. Detection of echocardiographic left ventricle boundaries using neural networks. In *Proceedings of Computers in Cardiology Conference*, pages 201–204, 1993. Cited on page(s) 1.1.1.

[32] A. Boudraa, M. Arzi, J. Sau, J. Champier, S. Hadj-Moussa, J.-E. Besson, D. Sappey-Marinier, R. Itti, and J.-J. Mallet. Automated detection of the left ventricular region in gated nuclear cardiac imaging. *IEEE Transactions on Biomedical Engineering*, 43(4):430–436, 1996. Cited on page(s) 1.1.1.

[33] J. Weng, A. Singh, and M.Y. Chiu. Learning-based ventricle detection from cardiac mr and ct images. *IEEE Transactions on Medical Imaging*, 16(4):378–391, 1997. Cited on page(s) 1.1.1.

[34] Gordon D Waiter, Fergus I McKiddie, Thomas W Redpath, Scott I.K Semple, and Roger J Trent. Determination of normal regional left ventricular function from cine-mr images using a semi-automated edge detection method. *Magnetic Resonance Imaging*, 17(1):99–107, 1999. Cited on page(s) 1.1.1.

[35] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of mr images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908, 1999. Cited on page(s) 1.1.1.

[36] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based bias field correction of mr images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):885–896, 1999. Cited on page(s) 1.1.1.

[37] Elaine A Ryan and Michael J Farquharson. Breast tissue classification using x-ray scattering measurements and multivariate data analysis. *Physics in Medicine and Biology*, 52(22):6679–6696, November 2007. Cited on page(s) 1.1.1.

[38] H.S. Choi, D.R. Haynor, and Y. Kim. Partial volume tissue classification of multichannel magnetic resonance images-a mixel model. *IEEE Transactions on Medical Imaging*, 10(3):395–407, 1991. Cited on page(s) .

[39] John S DaPonte and Porter Sherman. Classification of ultrasonic image texture by statistical discriminant analysis and neural networks. *Computerized Medical Imaging and Graphics*, 15(1):3–9, 1991. Cited on page(s) .

[40] Y.M. Kadah, A.A. Farag, J.M. Zurada, A.M. Badawi, and A.-B.M. Youssef. Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images. *IEEE Transactions on Medical Imaging*, 15(4):466–478, 1996. Cited on page(s) 1.1.1.

[41] Boudewijn P.F. Lelieveldt, Rob J. van der Geest, Johan H.C. Reiber, Johan G. Bosch, Steven C. Mitchell, and Milan Sonka. Time-continuous segmentation of cardiac image sequences using active appearance motion models. In Michael F. Insana and Richard M. Leahy, editors, *Information Processing in Medical Imaging*, pages 446–452, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg. Cited on page(s) 1.1.1.

[42] David T. Gering. Automatic segmentation of cardiac mri. In Randy E. Ellis and Terry M. Peters, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003*, pages 524–532, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. Cited on page(s) 1.1.1.

[43] Amin Katouzian, Ashwin Prakash, and Elisa Konofagou. A new automated technique for left- and right-ventricular segmentation in magnetic resonance imaging. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3074–3077, 2006. Cited on page(s) .

[44] Su Huang, Jimin Liu, Looi Chow Lee, Sudhakar K Venkatesh, Lynette Li San Teo, Christopher Au, and Wieslaw L Nowinski. An image-based comprehensive approach for automatic segmentation of left ventricle from cardiac short axis cine mr images. *Journal of Digital Imaging*, 24(4):598–608, 2011. Cited on page(s) .

[45] Tao Liu, Yun Tian, Shifeng Zhao, Xiaoying Huang, and Qingjun Wang. Residual convolutional neural network for cardiac image segmentation and heart disease diagnosis. *IEEE Access*, 8:82153–82161, 2020. Cited on page(s) 1.1.1.

[46] Sergio Di Bona and Ovidio Salvetti. Neural method for three-dimensional image matching. *Journal of Electronic Imaging*, 11(4):497–506, 2002. Cited on page(s) 1.1.1.

[47] George K. Matsopoulos, Nikolaos A. Mouravliansky, Pantelis A. Asvestas, Konstantinos K. Delibasis, and Vassilis Kouloulias. Thoracic non-rigid registration combining self-organizing maps and radial basis functions. *Medical Image Analysis*, 9(3):237–254, 2005. Cited on page(s) 1.1.1.

[48] J. Zhang, Y. Ge, S.H. Ong, C.K. Chui, S.H. Teoh, and C.H. Yan. Rapid surface registration of 3D volumes using a neural network approach. *Image and Vision Computing*, 26(2):201–210, February 2008. Cited on page(s) .

[49] Lifeng Shang, Jian Cheng Lv, and Zhang Yi. Rigid medical image registration using pca neural network. *Neurocomputing*, 69(13):1717–1722, 2006. Blind Source Separation and Independent Component Analysis. Cited on page(s) 1.1.1.

[50] Xulei Qin, Zhibin Cong, Rong Jiang, Ming Shen, Mary B Wagner, Paul Kirshbom, and Baowei Fei. Extracting cardiac myofiber orientations from high frequency ultrasound images. In *Medical Imaging 2013: Ultrasonic Imaging, Tomography, and Therapy*, volume 8675, page 867507. International Society for Optics and Photonics, 2013. Cited on page(s) 1.1.1.

[51] Göksu Bozdereli Berikol, Oktay Yildiz, and İ Türkay Özcan. Diagnosis of acute coronary syndrome with a support vector machine. *Journal of Medical Systems*, 40(4):84, 2016. Cited on page(s) 1.1.1.

[52] Jeffrey Zhang, Sravani Gajjala, Pulkit Agrawal, Geoffrey H Tison, Laura A Hallock, Lauren Beussink-Nelson, Mats H Lassen, Eugene Fan, Mandar A Aras, ChaRandle Jordan, et al. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. *Circulation*, 138(16):1623–1635, 2018. Cited on page(s) 1.1.1.

[53] Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, and U. Rajendra Acharya. Automated detection of covid-19 cases using deep neural networks with x-ray images. *Computers in Biology and Medicine*, 121:103792, 2020. Cited on page(s) 1.1.1.

[54] Ezz El-Din Hemdan, Marwa A Shouman, and Mohamed Esmail Karar. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. *arXiv preprint arXiv:2003.11055*, 2020. Cited on page(s) .

[55] Govardhan Jain, Deepti Mittal, Daksh Thakur, and Madhup K. Mittal. A deep learning approach to detect covid-19 coronavirus with x-ray images. *Biocybernetics and Biomedical Engineering*, 40(4):1391–1405, 2020. Cited on page(s) .

[56] Asmaa Abbas, Mohammed M Abdelsamea, and Mohamed Medhat Gaber. Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network. *Applied Intelligence*, 51(2):854–864, 2021. Cited on page(s) .

[57] Parnian Afshar, Shahin Heidarian, Farnoosh Naderkhani, Anastasia Oikonomou, Konstantinos N. Plataniotis, and Arash Mohammadi. Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. *Pattern Recognition Letters*, 138:638–643, 2020. Cited on page(s) .

[58] Asif Iqbal Khan, Junaid Latief Shah, and Mohammad Mudasir Bhat. Coronet: A deep neural network for detection and diagnosis of covid-19 from chest x-ray images. *Computer Methods and Programs in Biomedicine*, 196:105581, 2020. Cited on page(s) .

[59] K Shankar and Eswaran Perumal. A novel hand-crafted with deep learning features based fusion model for covid-19 diagnosis and classification using chest x-ray images. *Complex & Intelligent Systems*, 7(3):1277–1293, 2021. Cited on page(s) .

[60] Tej Bahadur Chandra, Kesari Verma, Bikesh Kumar Singh, Deepak Jain, and Satyabhuwan Singh Netam. Coronavirus disease (covid-19) detection in chest x-ray images using majority voting based classifier ensemble. *Expert Systems with Applications*, 165:113909, 2021. Cited on page(s) .

[61] Rachna Jain, Meenu Gupta, Soham Taneja, and D Jude Hemanth. Deep learning based detection and analysis of covid-19 on chest x-ray images. *Applied Intelligence*, 51(3):1690–1700, 2021. Cited on page(s) .

[62] Aras M. Ismael and Abdulkadir Şengür. Deep learning approaches for covid-19 detection based on chest x-ray images. *Expert Systems with Applications*, 164:114054, 2021. Cited on page(s) 1.1.1.

[63] Michiyoshi Kuwahara and Shigeru Eiho. 3-d heart image reconstructed from mri data. *Computerized Medical Imaging and Graphics*, 15(4):241–246, 1991. Cited on page(s) 1.1.2.

[64] Alessandro Salustri and Jos Roelandt. Three dimensional reconstruction of the heart with rotational acquisition: methods and clinical applications. *British Heart Journal*, 73(5 Suppl 2):10, 1995. Cited on page(s) 1.1.2.

[65] J Roelandt, A Salustri, W Vletter, Y Nosir, and N Bmining. Precordial multiplane echocardiography for dynamic anyplane, paraplane and three-dimensional imaging of the heart. *Quantitative 3-D Echocardiography*, page 9, 1994. Cited on page(s) 1.1.2.

[66] Jos RTC Roelandt, J Folkert, Wim B Vletter, Meindert A Taams, Leo Bekkering, H Glastra, Kie K Djoa, and Frank Weber. Ultrasonic dynamic three-dimensional visualization of the heart with a multiplane transesophageal imaging transducer. *Journal of the American Society of Echocardiography*, 7(3):217–229, 1994. Cited on page(s) 1.1.2.

[67] ME Miquel, DLG Hill, EJ Baker, SA Qureshi, RDB Simon, SF Keevil, and RS Razavi. Three-and four-dimensional reconstruction of intra-cardiac anatomy from two-dimensional magnetic resonance images. *The International Journal of Cardiovascular Imaging*, 19(3):239–254, 2003. Cited on page(s) 1.1.2.

[68] Will Schroeder, Kenneth M Martin, and William E Lorensen. *The visualization toolkit an object-oriented approach to 3D graphics*. Prentice-Hall, Inc., 1998. Cited on page(s) 1.1.2.

[69] Stephan Jacobs, Ronny Grunert, Friedrich W Mohr, and Volkmar Falk. 3d-imaging of cardiac structures using 3d heart models for planning in heart surgery: a preliminary study. *Interactive Cardiovascular and Thoracic Surgery*, 7(1):6–9, 2008. Cited on page(s) 1.1.2.

[70] Abhirup Banerjee, Julià Camps, Ernesto Zacur, Christopher M Andrews, Yoram Rudy, Robin P Choudhury, Blanca Rodriguez, and Vicente Grau. A completely automated pipeline for 3d reconstruction of human heart from 2d cine magnetic resonance slices. *Philosophical Transactions of the Royal Society A*, 379(2212):20200257, 2021. Cited on page(s) 1.1.2.

[71] Alessandro Salustri and Jos RTC Roelandt. Ultrasonic three-dimensional reconstruction of the heart. *Ultrasound in Medicine & Biology*, 21(3):281–293, 1995. Cited on page(s) 1.1.2.

[72] Chandrajit Bajaj and Samrat Goswami. Multi-component heart reconstruction from volumetric imaging. In *Proceedings of the 2008 ACM symposium on Solid and Physical Modeling*, pages 193–202, 2008. Cited on page(s) .

[73] Aqeel Al-Surmi, Rahmita Wirza, Mohd Zamrin Dimon, Ramlan Mahmod, and Fatimah Khalid. Three dimensional reconstruction of human heart surface from single image-view under different illumination conditions. *American Journal of Applied Sciences*, 10(7):669, 2013. Cited on page(s) 1.1.2.

[74] Pedro F Ferreira, Peter D Gatehouse, Raad H Mohiaddin, and David N Firmin. Cardiovascular magnetic resonance artefacts. *Journal of Cardiovascular Magnetic Resonance*, 15(1):1–39, 2013. Cited on page(s) 1.1.3.

[75] AWM Van der Graaf, P Bhagirath, S Ghoerbien, and MJW Götte. Cardiac magnetic resonance imaging: artefacts for clinicians. *Netherlands Heart Journal*, 22(12):542–549, 2014. Cited on page(s) 1.1.3.

[76] G.J. Grevera and J.K. Udupa. An objective comparison of 3-d image interpolation methods. *IEEE Transactions on Medical Imaging*, 17(4):642–652, 1998. Cited on page(s) 1.1.3.

[77] Juelin Leng, Guoliang Xu, and Yongjie Zhang. Medical image interpolation based on multi-resolution registration. *Computers and Mathematics with Applications*, 66(1):1–18, 2013. Cited on page(s) 1.1.3.

[78] Jan Ehrhardt, Dennis Säring, and Heinz Handels. Optical flow based interpolation of temporal image sequences. In *Medical Imaging 2006: Image Processing*, volume 6144, page 61442K. International Society for Optics and Photonics, 2006. Cited on page(s) 1.1.3.

[79] Antal Horváth, Simon Pezold, Matthias Weigel, Katrin Parmar, and Philippe Cattin. High order slice interpolation for medical images. In Sotirios A. Tsaftaris, Ali Gooya, Alejandro F. Frangi, and Jerry L. Prince, editors, *Simulation and Synthesis in Medical Imaging*, pages 69–78, Cham, 2017. Springer International Publishing. Cited on page(s) 1.1.3.

[80] Qinghua Lin, Qin Zhang, and Li Tongbin. Slice interpolation in mri using a decomposition-reconstruction method. In *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*, pages 678–681, 2017. Cited on page(s) 1.1.3.

[81] David H. Frakes, Lakshmi P. Dasi, Kerem Pekkan, Hiroumi D. Kitajima, Kartik Sundareswaran, Ajit P. Yoganathan, and Mark J. T. Smith. A new method for registration-based medical image interpolation. *IEEE Transactions on Medical Imaging*, 27(3):370–377, 2008. Cited on page(s) 1.1.3.

[82] G.P. Penney, J.A. Schnabel, D. Rueckert, M.A. Viergever, and W.J. Niessen. Registration-based interpolation. *IEEE Transactions on Medical Imaging*, 23(7):922–926, 2004. Cited on page(s) 1.1.3.

[83] Soledad Le Clainche, José M Vega, and Julio Soria. Higher order dynamic mode decomposition of noisy experimental data: The flow structure of a zero-net-mass-flux jet. *Experimental Thermal and Fluid Science*, 88:336–353, 2017. Cited on page(s) 1.2, 3.2.

[84] S Le Clainche, Daulet Izbassarov, M Rosti, Luca Brandt, and Outi Tammisola. Coherent structures in the turbulent channel flow of an elastoviscoplastic fluid. *Journal of Fluid Mechanics*, 888, 2020. Cited on page(s) 1.2.

[85] Soledad Le Clainche, Zhong-Hua Han, and Esteban Ferrer. An alternative method to study cross-flow instabilities based on high order dynamic mode decomposition. *Physics of Fluids*, 31(9):094101, 2019. Cited on page(s) 1.2, 3.1.

[86] Soledad Le Clainche. Prediction of the optimal vortex in synthetic jets. *Energies*, 12(9):1635, 2019. Cited on page(s) 1.2, 3.1.

[87] Jose Manuel Vega and Soledad Le Clainche. *Higher order dynamic mode decomposition and its applications*. Academic Press, 2020. Cited on page(s) 1.2, 6.1.

[88] Nourelhouda Groun, Beka Begiashvili, Eusebio Valero, Jesús Garicano-Mena, and Soledad Le Clainche. Data-driven methods beyond aerospace field. In *International Symposium on Unmanned Systems and The Defense Industry*, pages 105–110. Springer, 2022. Cited on page(s) .

[89] N Groun, B Begiashvili, E Valero, J Garicano-Mena, and S Le Clainche. Higher order dynamic mode decomposition beyond aerospace engineering. *Results in Engineering*, page 101471, 2023. Cited on page(s) .

[90] Jiaqing Kou, Soledad Le Clainche, and Weiwei Zhang. A reduced-order model for compressible flows with buffeting condition using higher order dynamic mode decomposition with a mode selection criterion. *Physics of Fluids*, 30(1), 2018. Cited on page(s) 1.2.

[91] Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, August 2010. Cited on page(s) 2.1, 2.1.5, 3.1.

[92] Beka Begiashvili, Nourel Groun, Jesús Garicano-Mena, Soledad Le Clainche, and Eusebio Valero. Data-driven modal decomposition methods as feature detection techniques for flow problems: A critical assessment. *Physics of Fluids*, 35(4), 2023. Cited on page(s) 2.1.

[93] G. W. Stewart. On the Early History of the Singular Value Decomposition. *SIAM Review*, 35(4):551–566, December 1993. Cited on page(s) 2.1.1, 2.1.2.

[94] G H Golub and Reinsch. Singular value decomposition and least squares solutions. page 18, 1971. Cited on page(s) 2.1.1, 2.1.2.

[95] Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, September 1966. Cited on page(s) 2.1.2.

[96] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A Multilinear Singular Value Decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, January 2000. Cited on page(s) 2.1.2.

[97] The MathWorks Inc. Tensor product toolbox version: 9.9 (r2020b), 2009. Cited on page(s) 2.1.2.

[98] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417, 1933. Cited on page(s) 2.1.3.

[99] Ian Jolliffe. Principal component analysis, brian everitt, 2005. Cited on page(s) 2.1.3.

[100] Karl Pearson. LIII. *On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, November 1901. Cited on page(s) 2.1.4.

[101] Lawrence Sirovich. Turbulence and the dynamics of coherent structures. i. coherent structures. *Quarterly of Applied Mathematics*, 45(3):561–571, 1987. Cited on page(s) 2.1.4.

[102] Anindya Chatterjee. An introduction to the proper orthogonal decomposition. *Current Science*, pages 808–817, 2000. Cited on page(s) 2.1.4.

[103] Scott B Leask and Vincent G McDonell. On the physical interpretation of proper orthogonal decomposition and dynamic mode decomposition for liquid injection. *arXiv preprint arXiv:1909.07576*, 2019. Cited on page(s) 2.1.4.

[104] Bernard O Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences of the United States of America*, 17(5):315, 1931. Cited on page(s) 2.1.6.

[105] Clarence W Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S Henningson. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115–127, 2009. Cited on page(s) 2.1.6.

[106] Joseph Needham. *Science and Civilisation in China: Volume 3, Mathematics and the Sciences of the Heavens and the Earth*, volume 3. Cambridge University Press, 1959. Cited on page(s) 2.2.

[107] A Calderón. Intermediate spaces and interpolation. *Studia Mathematica*, (1):31–34, 1963. Cited on page(s) 2.2.

[108] Noel Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015. Cited on page(s) 2.2.

[109] WJ Sacks and TJ Welch. Mitchell, and hp wynn,"design and analysis of computer experiments,". *Statistical Science*, 4(4):409–453, 1989. Cited on page(s) 2.2.

[110] William S Russell. Polynomial interpolation schemes for internal derivative distributions on structured grids. *Applied Numerical Mathematics*, 17(2):129–171, 1995. Cited on page(s) 2.2.

[111] Kevin Johnston, Jay M Ver Hoef, Konstantin Krivoruchko, and Neil Lucas. *Using ArcGIS geostatistical analyst*, volume 380. Esri Redlands, 2001. Cited on page(s) 2.2.1.

[112] Sky McKinley and Megan Levine. Cubic spline interpolation. *College of the Redwoods*, 45(1):1049–1060, 1998. Cited on page(s) 2.2.1.

[113] Charles A Hall and W Weston Meyer. Optimal error bounds for cubic spline interpolation. *Journal of Approximation Theory*, 16(2):105–122, 1976. Cited on page(s) 2.2.1.

[114] Hsieh Hou and H Andrews. Cubic splines for image interpolation and digital filtering. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(6):508–517, 1978. Cited on page(s) 2.2.1.

[115] Christian Le Provost, Florent Lyard, and Jean-Marc Molines. Improving ocean tide predictions by using additional semidiurnal constituents from spline interpolation in the frequency domain. *Geophysical Research Letters*, 18(5):845–848, 1991. Cited on page(s) 2.2.1.

[116] K Takayama, Y Obata, M Morishita, and T Nagai. Multivariate spline interpolation as a novel method to optimize pharmaceutical formulations. *Die Pharmazie-An International Journal of Pharmaceutical Sciences*, 59(5):392–395, 2004. Cited on page(s) 2.2.1.

[117] YX Liu and Ronald Rousseau. Splines can recover dynamic information contained in discrete data. In *Proceedings of ISSI*, pages 1022–1024, 2011. Cited on page(s) 2.2.1.

[118] Young Kinh-Nhue Truong and Muhammad Sarfraz. *Topics in splines and applications*. BoD–Books on Demand, 2018. Cited on page(s) 2.2.1.

[119] John Tetteh, Sian Howells, Ed Metcalfe, and Takahiro Suzuki. Optimisation of radial basis function neural networks using biharmonic spline interpolation. *Chemometrics and Intelligent Laboratory Systems*, 41(1):17–29, 1998. Cited on page(s) .

[120] Wei-Chen Wu, Tsun-Hsien Wang, and Ching-Te Chiu. Edge curve scaling and smoothing with cubic spline interpolation for image up-scaling. *Journal of Signal Processing Systems*, 78:95–113, 2015. Cited on page(s) .

[121] L Pedroso, A Arco, I Figueiras, JV Araújo dos Santos, JLM Fernandes, and H Lopes. Application of cubic spline interpolation with optimal spatial sampling for damage identification. *Structural Control and Health Monitoring*, 29(1):e2856, 2022. Cited on page(s) .

[122] Igor V Yuyukin. Spline interpolation of navigational isolines. *Vestnik Gosudarstvennogo universiteta morskogo i rechnogo flota imeni admirala SO Makarova*, 11:1026–1036, 2019. Cited on page(s) 2.2.1.

[123] Soledad Le Clainche and José M Vega. Higher order dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 16(2):882–925, 2017. Cited on page(s) 3.1, 5.1.

[124] Soledad Le Clainche. An introduction to some methods for soft computing in fluid dynamics. In Francisco Martínez Álvarez, Alicia Troncoso Lora, José António Sáez Muñoz, Héctor Quintián, and Emilio Corchado, editors, *14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019)*, pages 557–566, Cham, 2020. Springer International Publishing. Cited on page(s) 3.1.

[125] Peter Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967. Cited on page(s) 3.1.

[126] András Rövid, Imre J Rudas, Szabolcs Sergyán, and László Szeidl. Hosvd based image processing techniques. In *Proc. of the 10th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, pages 297–302, 2011. Cited on page(s) 3.3.

[127] María Villalba-Orero, Marina M López-Olañeta, Esther González-López, Laura Padrón-Barthe, Jesús M Gómez-Salinero, Jaime García-Prieto, Timothy Wai, Pablo García-Pavía, Borja Ibáñez, Luis J Jiménez-Borreguero, et al. Lung ultrasound as a translational approach for non-invasive assessment of heart failure with reduced or preserved ejection fraction in mice. *Cardiovascular Research*, 113(10):1113–1123, 2017. Cited on page(s) 4.1.1, 4.1.2.

[128] Javier Larrasa-Alonso, María Villalba, Carlos Martí-Gómez, Paula Ortiz-Sánchez, Marina López-Olañeta, Maria Ascension Rey-Martin, Fatima Sanchez-Cabo, François McNicoll, Michaela Müller-McNicoll, Pablo Garcia-Pavia, et al. The srsf4-gas5-glucocorticoid

receptor axis regulates ventricular hypertrophy. *Circulation Research*, 2021. Cited on page(s) 4.1.1, 4.2.1.

[129] Enrique Lara-Pezzi, Philippe Menasché, Jean-Hugues Trouvin, Lina Badimón, John PA Ioannidis, Joseph C Wu, Joseph A Hill, Walter J Koch, Albert F De Felice, Peter de Waele, et al. Guidelines for translational research in heart failure. *Journal of Cardiovascular Translational Research*, 8(1):3–22, 2015. Cited on page(s) 4.1.1.

[130] Maria Villalba-Orero, Marina López-Olañeta, Pablo García-Pavía, and Enrique Lara-Pezzi. Systolic dysfunction in infarcted mice does not necessarily lead to heart failure: Need to refine preclinical models. *Journal of Cardiovascular Translational Research*, 10(5):499–501, 2017. Cited on page(s) 4.1.1.

[131] Nourelhouda Groun, María Villalba-Orero, Enrique Lara-Pezzi, Eusebio Valero, Jesús Garicano-Mena, and Soledad Le Clainche. Higher order dynamic mode decomposition: From fluid dynamics to heart disease analysis. *Computers in Biology and Medicine*, 144:105384, 2022. Cited on page(s) 1.

[132] Nourelhouda Groun, María Villalba-Orero, Enrique Lara-Pezzi, Eusebio Valero, Jesús Garicano-Mena, and Soledad Le Clainche. A novel data-driven method for the analysis and reconstruction of cardiac cine MRI. *Computers in Biology and Medicine*, 151:106317, 2022. Cited on page(s) 1.

[133] MATLAB. *9.9.0.1538559 (R2020b) Update 3*. The MathWorks Inc., Natick, Massachusetts, 2020. Cited on page(s) 6.1.1.

Part V

Appendices

Summary This section provides additional results for the two datasets analyzed. In particular, the section includes tables containing all captured frequencies, graphs simplifying the different procedures taken through this analysis, as well as additional experiments that have not been included in the main structure of the document.

In more details, in appendix A V, all the frequencies captured from the HODMD analysis are presented. Tables from 1 to 5 include the frequencies of all the DMD modes obtained from the two datasets analyzed for each one of the cardiac conditions.
In appendix B V we present all the frequencies obtained from the analysis of the three investigated MRI datasets. Tables 6, 7 and 8 display the frequencies captured in all the slices for the healthy dataset and the two hypertension datasets, respectively.
In appendix C V the results obtained from applying the HODMD-based reduced order model introduced in 6.2 to the hypertension datasets are presented. In particular, Fig. 1 and Fig. 2 show a comparison between the original snapshots and the new generated snapshots for several MRI slices.
Finally, in appendix D V we include two figures 3 and 4. The figures represent additional results related to the application of the information recovery technique presented in section 3.3 to two other slices other than the one presented in the main structure of the document.

Appendix A

In this section all the frequencies captured from the HODMD analysis of the echocardiography datasets are presented. Tables from 1 to 5 include the frequencies of all the DMD modes obtained from the two datasets analyzed for each one of the cardiac conditions.

1- Healthy datasets

Frequencies	Healthy dataset 01		Healthy dataset 02	
	LAX	SAX	LAX	SAX
ω_1	0	0	0	0
ω_2	208.6	203.8	158.09	210.8
ω_3	425.4	419.0	334.8	514.6
ω_4	633.4	644.1	503.9	750.3
ω_5	859.2	856.2	667.6	1011.5
ω_6	1074.2	1063.9	846.6	1185.25
ω_7	1273.2	1289.2	1010.8	1560.8
ω_8	1502.4	1505.0	1180.3	1747.05
ω_9	1720.6	1716.4	1358.01	2079.3
ω_{10}	1916.9	1934.4	1521.8	2828.5
ω_{11}	2144.6	2203.4	1691.8	5855.9
ω_{12}	2365.7	2582.6	1867.7	
ω_{13}	2560.7	2735.8	2030.3	
ω_{14}	2786.9	3230.1	2204.5	
ω_{15}	3009.0	3889.9	2377.7	
ω_{16}	3203.3	4526.7	2541.08	
ω_{17}	3428.3	5186.9	2717.7	
ω_{18}	3652.0	5725.5	2890.19	
ω_{19}		7499.9	3057.05	

Table 1: Frequencies of the healthy datasets.

2- Myocardial infarction datasets

Frequencies	MI dataset 01		MI dataset 02	
	LAX	SAX	LAX	SAX
ω_1	0	0	0	0
ω_2	202.7	167.2	277.3	239.9
ω_3	343.7	344.3	473.6	490.2
ω_4	596.09	514.8	585.0	741.9
ω_5	811. 6	704.8	870.4	989.2
ω_6	935.5	882.8	1173.0	1239.5
ω_7	1186.1	1043.6	1460.8	1490.6
ω_8	1416	1240.5	1759.1	1741.6
ω_9	1787.9	1568.8	2057.4	1990.7
ω_{10}	2008.3	1792.1	2350.3	2239.5
ω_{11}	2382.4	2116.6	2648.2	2490.6
ω_{12}	2984.1	3025.8	2940.3	2739.0
ω_{13}	3549.8	3438.5	3238.9	2993.5
ω_{14}		4009.07	3528.7	3241.4
ω_{15}			3528.7	3241.4
ω_{16}			3828.4	3493.7
ω_{17}			4117.9	3741.8
ω_{18}			4414.8	4008.5
ω_{19}			4707.8	4228.9
ω_{20}			5003.1	4688.1
ω_{21}			5296.6	
ω_{22}			5591.7	

Table 2: Frequencies of the myocardial infarction (MI) datasets.

3- Obesity datasets

Frequencies	Ob dataset 01		Ob dataset 02	
	LAX	SAX	LAX	SAX
ω_1	0	0	0	0
ω_2	185.5	132.0	203.1	166.8
ω_3	425.9	279.6	535.8	369.6
ω_4	656.5	457.7	633.2	554.7
ω_5	867.5	615.1	868.7	731.5
ω_6	1099.5	770.0	1105.9	929.8
ω_7	1313.2	1052.8	1634.9	1114.9
ω_8	1612.2	1226.1	2249.8	1299.5
ω_9	1972.8	1371.9	2732.8	1494.07
ω_{10}	2234.4	1690.5	3253.7	1673.61
ω_{11}	2632.0	1846.8	3817.4	1860.7
ω_{12}	3275.6	2006.9		2055.6
ω_{13}	3545.7	2269.2		2235.4
ω_{14}	3959.5	2451.2		2423.1
ω_{15}	4572.2	3012.9		2805.1
ω_{16}		3464.0		3500.9
ω_{17}		3637.3		

Table 3: Frequencies of the obesity (Ob) datasets.

4- TAC hypertension datasets

Frequencies	TAC dataset 01		TAC dataset 02	
	LAX	SAX	LAX	SAX
ω_1	0	0	0	0
ω_2	198.3	194.9	208.4	190.8
ω_3	480.4	424.7	456.7	410.7
ω_4	715.4	621.02	671.9	612.0
ω_5	933.04	849.4	908.9	839.0
ω_6	1219.6	1055.8	1138.5	1033.7
ω_7	1436.6	1278.1	1371.0	1306.1
ω_8	1687.5	1701.9	1598.5	1737.9
ω_9	2169.9	2146.2	1835.3	
ω_{10}	2591.1	2550.38	2045.0	
ω_{11}	2921.8	3061.9	2317.5	
ω_{12}	3652.9	3612.9	2772.3	
ω_{13}	4398.5		3514.4	
ω_{14}	5076.6		4011.9	
ω_{15}	72.4		4637.4	
ω_{16}			5665.3	

Table 4: Frequencies of the TAC hypertension (TAC) datasets.

5- Diabetic cardiomyopathy datasets

Frequencies	DC dataset 01		DC dataset 02	
	LAX	SAX	LAX	SAX
ω_1	0	0	0	0
ω_2	213.4	196.7	208.2	181.4
ω_3	493.2	353.8	438.3	405.8
ω_4	747.4	494.7	652.41	626.9
ω_5	984.3	703.2	875.02	824.8
ω_6	1240.6	819.7	1091.4	1149.07
ω_7	1489.9	997.7	1315.9	1264.3
ω_8	1707.0	1284.6	1536.2	1571.5
ω_9	1999.6	1500.8	1755.05	1769.8
ω_{10}	2227.1	1902.9	1973.5	2026.1
ω_{11}	2475.5	1995.4	2192.34	
ω_{12}	3003.2	2505.0	2416.6	
ω_{13}	3506.5	3003.3	2629.1	
ω_{14}	3750	3505.1	2851.9	
ω_{15}		3984.0	3086.87	
ω_{16}		4493.7	3312.9	

Table 5: Frequencies of the diabetic cardiomyopathy (DC) datasets.

Appendix B

In the following, additional results related to the analysis of the MRI datasets are presented. In more details, this section presents all the frequencies obtained from the analysis of the three investigated MRI datasets. Tables 6, 7 and 8 display the frequencies captured in all the slices for the healthy dataset and the two hypertension datasets, respectively.

1- Healthy MRI dataset

Frequencies	MRI sequences									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
ω_1	0	0	0	0	0	0	0	0	0	0
ω_2	369	361	376	386	365	360	355	369	373	342
ω_3	769	751	756	759	765	768	800	827	861	867
ω_4	1223	1210	1217	1203	1196	1200	1227	1236	1287	1294
ω_5	1674	1633	1637	1630	1637	1641	1649	1670	1635	1657
ω_6	2086	2064	2074	2073	2074	2071	2071	2086	2074	2085
ω_7	2516	2493	2503	2498	2500	2492	2487	2506	2503	2514
ω_8	2918	2904	2922	2918	2925	2923	2913	2918	2933	2938
ω_9	3342	3365	3331	3339	3340	3336	3331	3327	3353	3352
ω_{10}				3750	3750	3750	3750	3750	3750	3750

Table 6: Frequency of the healthy MRI dataset.

2- First hypertension MRI dataset

Frequencies	MRI sequences									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
ω_1	0	0	0	0	0	0	0	0	0	0
ω_2	343	353	339	341	353	356	356	345	404	497
ω_3	770	763	740	758	761	777	811	838	872	868
ω_4	1217	1226	1188	1170	1194	1217	1246	1263	1296	1259
ω_5	1640	1646	1629	1625	1633	1657	1667	1674	1696	1691
ω_6	2074	2093	2068	2066	2071	2085	2096		2084	2125
ω_7	2500	2516	2509	2498	2516	2520	2526	2573	2532	2578
ω_8	2933	2935	2937	2941	2931	2923	2911	3028	2935	3104
ω_9	3345	3341	3352	3338	3333	3335		3428	3335	3566
ω_{10}	3750	3750	3750	3750		3750	3750	3750	3750	3750

Table 7: Frequency of the first hypertension MRI dataset.

3- Second hypertension MRI dataset

Frequencies	MRI sequences									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
ω_1	0	0	0	0	0	0	0	0	0	0
ω_2	384	351	332	302	321	332	365	386	402	461
ω_3	770	791	752	734	744	766	818	861	882	943
ω_4	1236	1213	1208	1190	1213	1221	1256	1266	1294	1382
ω_5	1656	1638	1647	1636	1646	1653	1676	1680	1680	1714
ω_6	2092	2087	2065	2065	2086	2065	2102	2111	2092	2140
ω_7	2518	2518	2482	2493	2504	2484	2521	2539	2537	2518
ω_8	2940	2923	2916	2920	2931	2920	2943	2963	2944	2981
ω_9	3350	3338	3343	3338	3333	3328	3339	3364	3341	3416
ω_{10}		3750	3750	3750	3750	3750	3750	3750	3750	3750

Table 8: Frequency of the Second hypertension MRI dataset.

Appendix C

In this section the results obtained from applying the HODMD-based reduced order model introduced in 6.2 to the hypertension datasets are presented. In particular, Fig. (1) and Fig. (2) show a comparison between the original snapshots and the new generated snapshots for several MRI slices of the first and second hypertension datasets.

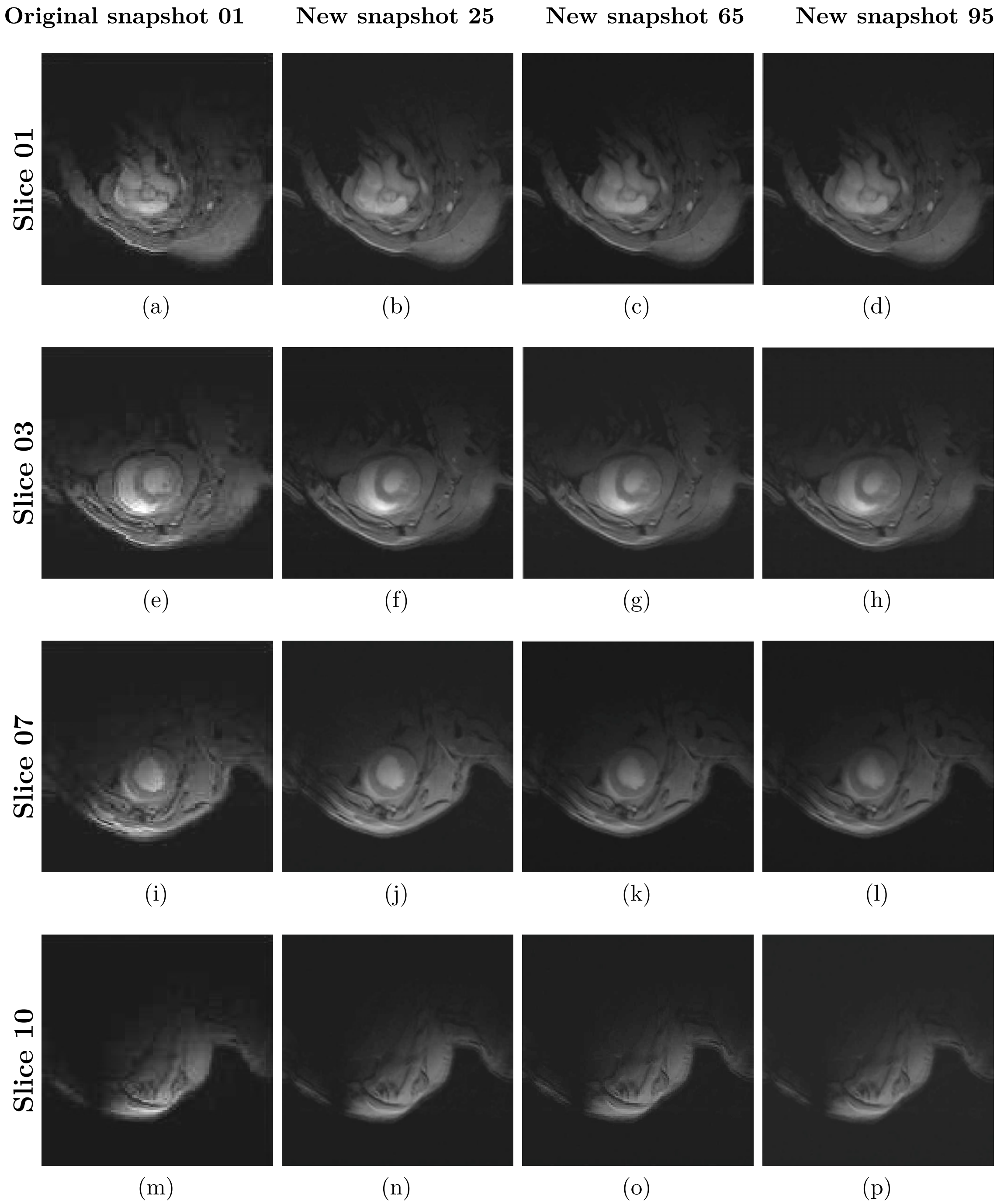


Figure 1: Different reconstructed snapshots obtained from applying the HODMD-based ROM on the first hypertension dataset.

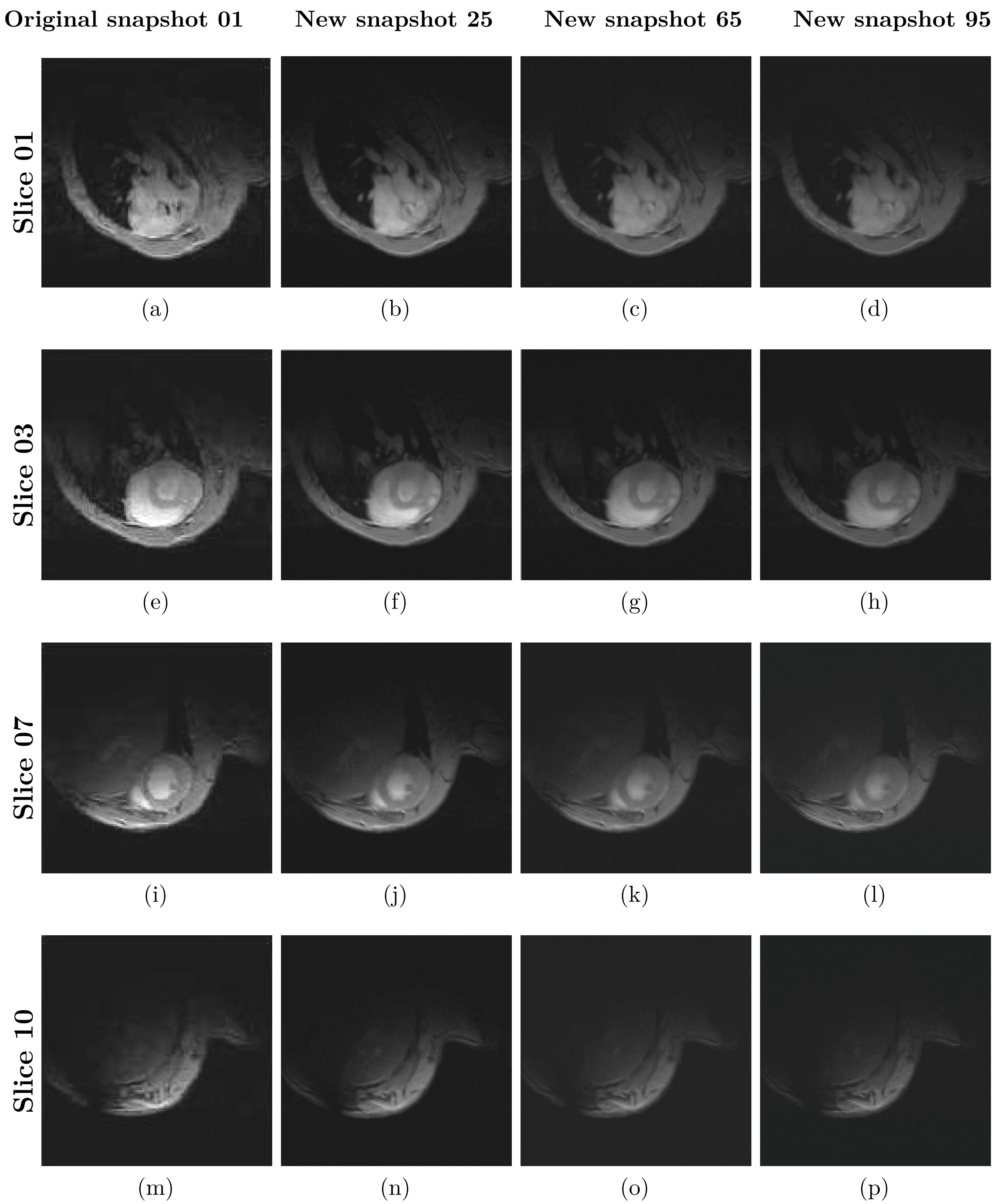


Figure 2: Different reconstructed snapshots obtained from applying the HODMD-based ROM on the second hypertension dataset.

Appendix D

Finally, this section includes two figures (3) and (4). The figures represent additional results related to the application of the information recovery technique presented in 3.3 (see Fig. (3.2)) to slices 05 and 09 of the healthy MRI sequence. Figure (3) exhibit a comparison between the original snapshots 1, 5, 10, 15 and 20 of slice 5 and snapshots 1, 5, 10, 15 and 20 of the new interpolated slice. Meanwhile, Fig. (4) exhibit a comparison between the original snapshots 1, 5, 10, 15 and 20 of slice 9 and snapshots 1, 5, 10, 15 and 20 of the new interpolated slice.

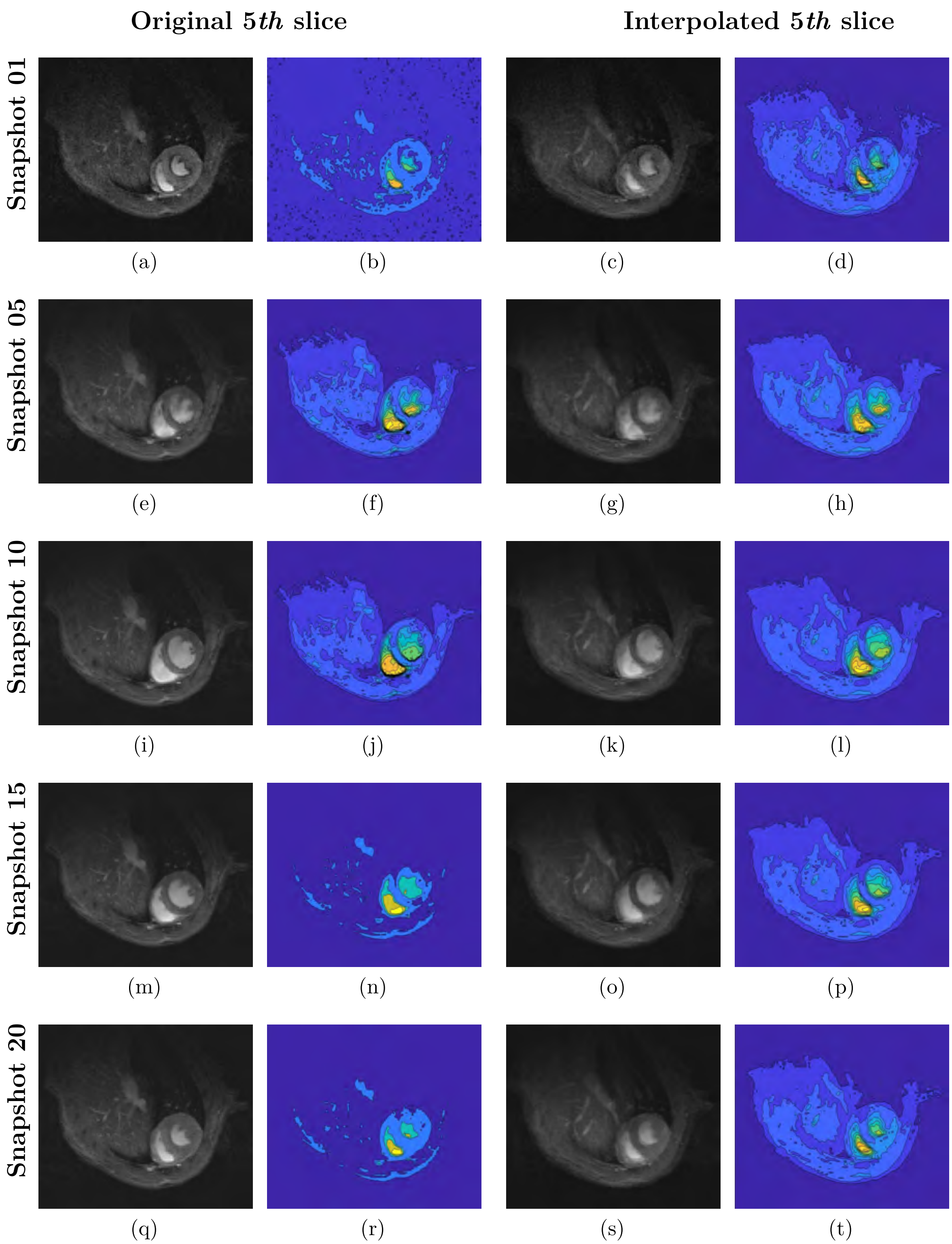


Figure 3: A comparison between snapshots 01, 05, 10, 15 and 20 taken from original slice 05 and the interpolated slice 05. first and second columns correspond to the original snapshot, and third and fourth columns to the reconstruction. The second and fourth columns present the intensity of the images (included for clearer comparison).

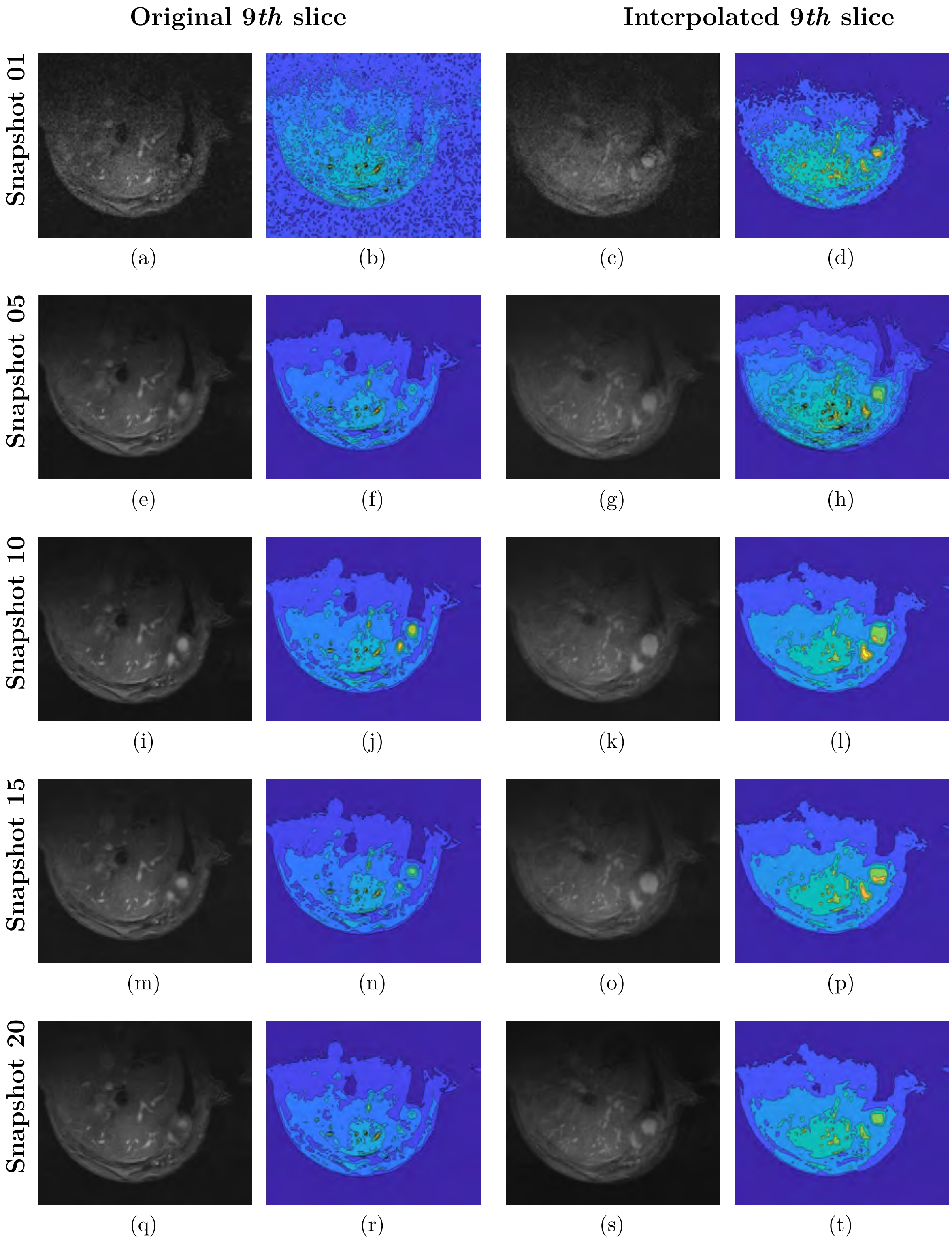


Figure 4: A comparison between snapshots 01, 05, 10, 15 and 20 taken from original slice 09 and the interpolated slice 09. first and second columns correspond to the original snapshot, and third and fourth columns to the reconstruction. The second and fourth columns present the intensity of the images (included for clearer comparison).