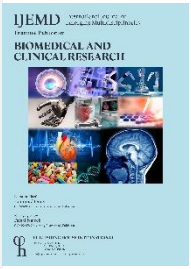
IJEMD-BMCR, 1 (2) (2023), 1 – 17 https://doi.org/10.54938/ijemdbmcr.2023.01.2.244

International Journal of Emerging Multidisciplinaries:
Biomedical and Clinical Research

Research Paper
Journal Homepage: www.ojs.ijemd.com
ISSN (print): 2957-8620 ISSN (online): 2960-0731



Optimizing MRI Data Processing by exploiting GPU Acceleration for Efficient Image Analysis and Reconstruction

Irfan Ullah[1*] and Hammad Omer[1]

[1] *Department of Electrical and Computer Engineering, COMSATS University Islamabad, Islamabad, Pakistan*
**Corresponding author*

Abstract

Magnetic Resonance Imaging (MRI) is a non-invasive medical imaging modality, offering detailed anatomical insights without ionizing radiation The advent of Graphics Processing Units (GPUs) has stimulated a paradigm shift, propelling MRI techniques to new frontiers. Through parallel processing capabilities, GPUs have expedited real-time imaging, complex image reconstruction, noise reduction, and intricate data analysis. This paper provides an extensive survey on the GPUs based MRI techniques and its synergistic impact on core methodologies such as Diffusion Tensor Imaging (DTI) and Functional MRI (fMRI). Through parallel processing and frameworks like CUDA and OpenCL, GPUs have overcome computational hurdles in MRI data processing. Challenges like memory constraints and data transfer bottlenecks are addressed through hybrid CPU-GPU strategies and algorithmic enhancements. The integration of GPUs yields faster scans, enhanced image quality, and real-time insights, benefiting patient care and accelerating medical research. Considering the ethical issues regarding patient data privacy and algorithmic fairness, GPUs' potential in MRI research and development is evident. This paper concludes by looking ahead about the future of GPU based MRI, urging further exploration to uncover new possibilities and shape a transformative path for the future of medical imaging.

Keywords: DTI; DWI; GPU; MRF; MRI.

Email addresses: irfan_ullah@comsats.edu.pk (I. Khattak), hammad.omer@comsats.edu.pk (H. Omer)

1. Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive medical imaging technique that uses strong magnetic fields and radio waves to generate detailed and high-resolution images of the internal structures of the human body [1]. It provides valuable information about the anatomy, function, and pathology of various tissues and organs without utilizing ionized radiation, making it a safer option compared to techniques like X-rays or CT scans. MRI has become an essential tool in medical imaging due to its ability to provide multi-dimensional images with exceptional soft tissue contrast and spatial resolution.

MRI technology excels at differentiating between various soft tissues, such as muscles, organs, and nerves. This ability of MRI is particularly useful for diagnosing the brain, spinal cord, joints, and abdominal organs [1]. MRI is non-invasive and does not carry the risks associated with exposure to radiation. This makes it a safer option, especially for pediatric and pregnant patients [2]. MRI can acquire images in multiple planes (axial, sagittal, coronal), allowing physicians to view structures from different angles [1]. This is crucial for accurate diagnosis and surgical planning. Functional MRI (fMRI) can assess brain activity by measuring changes in blood flow and oxygenation [3]. It is widely used in neuroscience to map brain sections responsible for specific functions. Cardiac MRI can provide detailed images of the heart's anatomy and function, aiding in the diagnosis of various heart conditions, including congenital defects and heart muscle diseases [4]. MR images are used to visualize tumors, assess their size and location, and determine if they have spread to nearby tissues [5]. It helps in cancer staging and treatment planning [5]. MRI is essential for evaluating musculoskeletal conditions such as joint injuries, ligament tears, and bone fractures [6]. Techniques like MR angiography (MRA) [7] can visualize blood vessels without the need for invasive procedures. This is valuable for assessing conditions like aneurysms and vascular malformations. Techniques like MR angiography (MRA) can visualize blood vessels without the need for invasive procedures. This is valuable for assessing conditions like aneurysms and vascular malformations [7].

MRI data acquisition commences by applying strong magnetic fields and Radio Frequency (RF) pulses that excite the hydrogen nuclei in the human body. This will result in altering the energy state of the hydrogen nuclei [1]. After removing magnetic field hydrogen nuclei will return to their actual state by releasing absorbed energy in the form of RF signal. The resulting signals are captured by receiver coils and transformed into raw *k*-space data, representing spatial frequency information. Afterwards, the raw *k*-space data undergoes certain preprocessing steps i.e., noise reduction and correction, filtering, Fourier transforms etc.to be converted to MR image [1].

The raw *k*-space data is transformed into image space through a Fast Fourier Transform (FFT). However, due to the limited speed of conventional CPUs, this step can be time-consuming for large datasets. Here, Graphics Processing Units (GPUs) play a pivotal role [8]. GPU’s parallel processing capabilities accelerate the FFT, allowing for rapid conversion of *k*-space data into MR images [9]. Moreover, recent advancements in MRI techniques, such as parallel imaging and Compressed Sensing [10-11], often involve complex iterative reconstruction algorithms that benefit immensely from GPU acceleration [10-11]. Additionally, deep learning methods for image denoising, super-resolution, and reconstruction have further integrated GPUs to expedite training and inference processes [11].

GPUs accelerate the reconstruction process [12-13], making it feasible to reconstruct images on-the-fly during data acquisition, enabling immediate visualization and analysis. This real-time capability opens doors to interventions, adjustments, and decisions guided by live imaging feedback. However, the pivotal

question remains: Why is GPU acceleration crucial for MRI techniques? This paper aims to address this question by delving into the challenges of MRI data processing, showcasing recent research that highlights the role of GPUs, and highlighting their transformative impact on expediting MRI image generation, enabling timely clinical diagnoses, and advancing the frontiers of MRI research.

MRI exploits the magnetic properties of hydrogen nuclei (protons), that is the majority part of human body due to the prevalence of water molecules [1]. At the core of an MRI system lies the magnet that produces strong magnetic field B_0 (Figure 1(a)). Normally protons are randomly aligned (Figure 1(b)) in the human body tissues but when the human body is exposed to a strong magnetic field, protons (in body tissues) align with direction of the magnetic field (Figure1(c)). Afterwards, a short duration of Radio Frequency (RF) energy is applied by RF coils mounted inside the bore of the MRI machine [1], as a result protons absorb energy and temporarily perturbed from their native energy orbit. Intelligent placement of these coils around the patient's body optimizes the signal-to-noise ratio and enhances the quality of acquired images [1]. Once the RF pulse ends, the protons gradually return to their original alignment along the magnetic field. While returning, proton releases the absorbed energy in the form of RF signals. Emitted RF signals are then detected and received by the RF receiver coils in MRI machine (Figure 1(a)). These signals are amplified for further processing.

Spatial information is encoded into the MRI signal by using magnetic field gradients (Figure 1(a)). These gradients vary the magnetic field strength across the imaging area, causing protons at different positions to precess at slightly different frequencies [1]. By altering the gradients during data acquisition, the MRI machine can distinguish the location of the protons and create spatial information. During the imaging process, the MRI machine applies a series of RF pulses and gradient variations. The resulting signals are collected for various spatial locations within the imaging slice or volume. These signals represent the raw MRI data, which contains information about tissue properties and their spatial distribution [1].

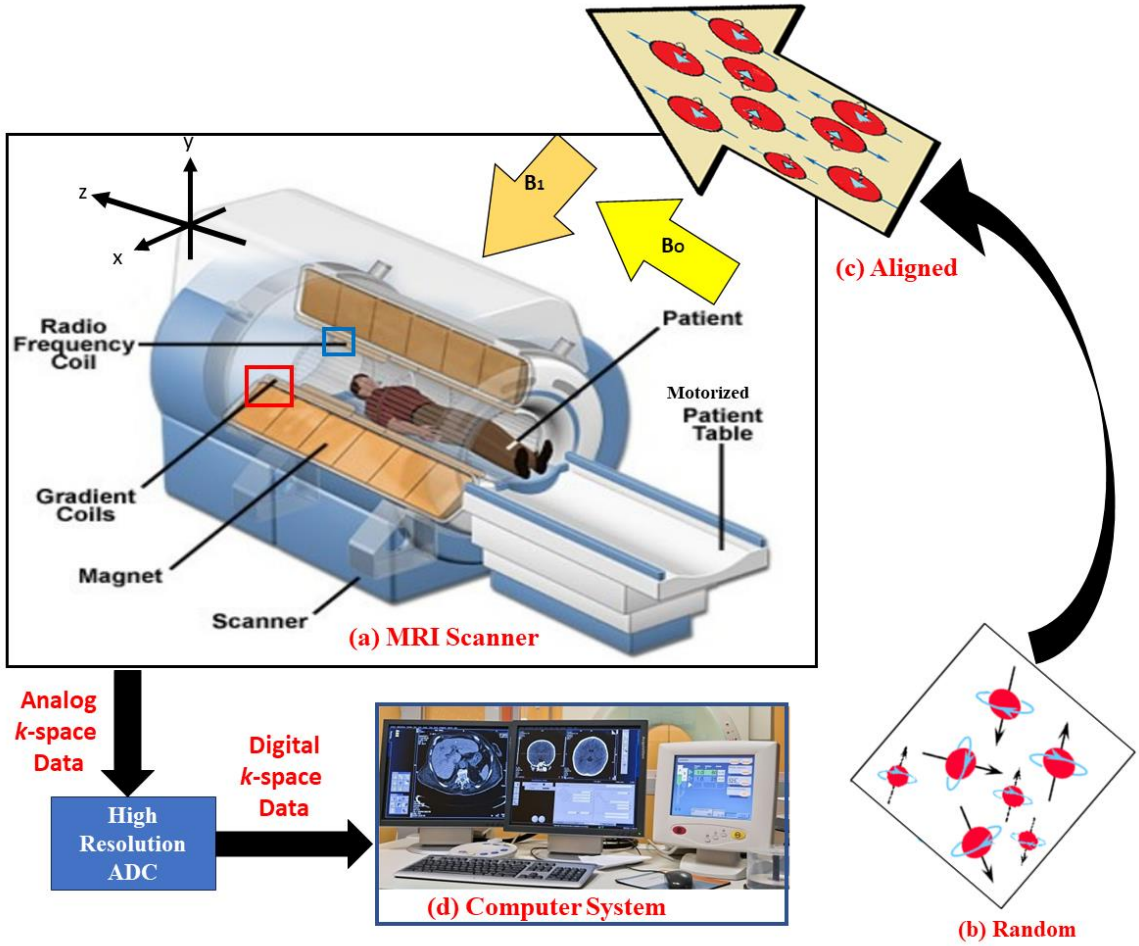


Figure 1: Schematic Representation of an MRI Scanner and its Working. The figure illustrates the essential components of an MRI scanner, including the main magnet, gradient coils, RF coils, computer system, patient table, and user interface. The working block diagram showcases the interplay of these components, highlighting their roles in generating high-resolution images of internal anatomical structures [1].

2. Data Acquisition, Processing, and Computational Challenges in MRI Imaging:

The collected raw data is digitized using high resolution Analog to Digital Converters (ADC) and amplified then transferred to a sophisticated computer system (Figure 1(d)) equipped with specialized software for further processing [1]. Furthermore, the computer system consists of a display and control interface. The MRI technologist/radiologist interacts with the system through a user-friendly display and control interface. This interface allows for the configuration of imaging parameters, real-time monitoring of scans, and adjustments as needed (Figure 1(d)).

The MR image reconstruction process involves advanced mathematical techniques, mainly a Fourier transformation, which converts the frequency-domain data into spatial information. The resulting images represent the distribution of different tissue types, and their contrast is determined by the relaxation times (T1 and T2) of the tissues. Different tissues have varying T1 and T2 relaxation times, leading to differences in signal intensity and contrast in the final images [1]. These contrasts allow radiologists to distinguish between various tissues and identify abnormalities.

The computation complexity of post-processing from raw *k*-space data to generate MR images involves a series of moderate-level operations, including filtering, contrast enhancement, and potentially image registration. These operations generally require linear algebra and basic image processing techniques, resulting in a computational workload that scales approximately linearly with the size of the raw MR data. The computational complexity in post processing exponentially increases when reducing the MRI scan time. The conventional approach to acquiring MRI data involves acquiring a full set of *k*-space data points [1], which can be time-consuming, especially for high-resolution or dynamic imaging. To address this under-sampling is employed to reduce scan time by acquiring only a subset of *k*-space data [14]. The reduced data acquisition requires specialized reconstruction algorithms to fill in the missing *k*-space data in frequency domain and some advanced image-based reconstruction techniques for processing data in image domain to generate high-quality artifact free MR images [14]. This process can involve complex optimization and iterative algorithms. Parallel imaging methods, like SENSE (Sensitivity Encoding) and GRAPPA (Generalized Auto-calibrating Partially Parallel Acquisitions), utilize multiple receiver coils to simultaneously acquire data, effectively reducing the amount of data that needs to be acquired. This under-sampling, while reducing scan time, can introduce artifacts in the reconstructed images due to missing data points [14]. Obtaining clinically feasible MR image from under-sampled *k*-space data has a severe impact on post-processing. One of the most transformative impacts of accelerated MRI data acquisition is the ability to perform real-time or near-real-time imaging. For instance, in dynamic imaging studies like cardiac cine-MRI or real-time functional MRI (fMRI), rapid image generation is crucial to capture dynamic physiological processes [3][7].

In a study by Knoll et al. [15], advanced image reconstruction methods based on parallel imaging and compressed sensing were applied to enhance MRI image quality. The reconstruction process involved solving complex optimization problems, contributing to increased computation time. Similarly, in research by Hammernik et al. [16], deep learning techniques were employed for image denoising and reconstruction, which introduced additional computational demands due to the training and deployment of neural networks. To address these computational challenges, GPUs have emerged as a pivotal technology.

3. GPU Technology and Parallel Computing:

GPUs, originally designed for rendering graphics, offer thousands of cores optimized for parallel processing, allowing them to perform numerous tasks simultaneously [17-19]. GPU architecture excels at accelerating matrix operations and data-intensive tasks, making them ideal for speeding up the iterative reconstruction algorithms and neural network models in MRI image processing [12-16]. GPU architecture is particularly advantageous, where operations can be divided into smaller and data independent modules (program) [12-16].

The memory hierarchy within GPUs, comprising global memory, shared memory, and registers, optimizes data access and minimizes memory-related bottlenecks (Figure 2). This design is especially relevant for handling the vast datasets involved in MRI. Additionally, the Single Instruction Multiple Data (SIMD) architecture of GPUs allows a single instruction to operate on multiple data elements simultaneously [12-16]. This architecture is well-suited for MRI post-processing tasks like convolution and filtering.

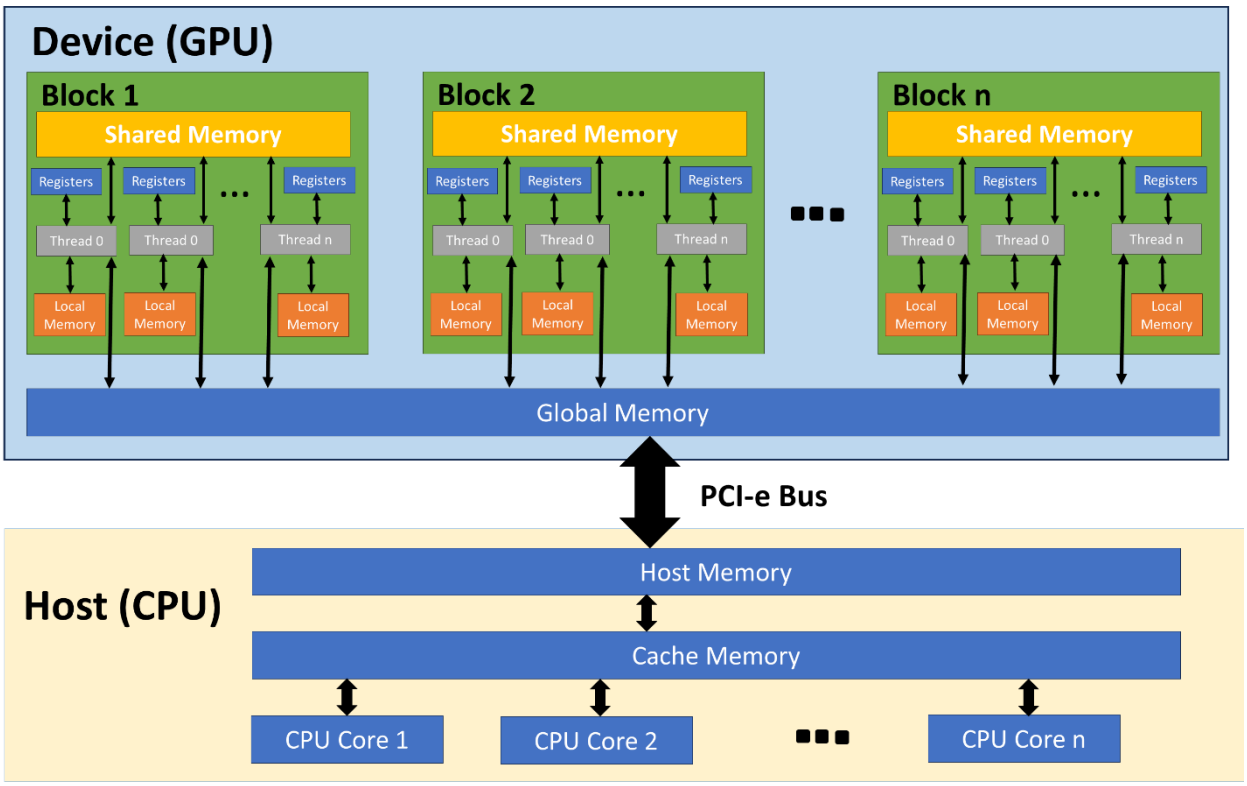


Figure 2:A block diagram of a CPU-GPU computing model. The GPU's architecture includes register, local, shared, and global memories, while the CPU operates with cache and host memory, enhancing data access speeds at different levels. Communication between CPU and GPU is supported by high-speed PCI-e Bus that optimizes performance for diverse applications.

Efficient data transfer mechanisms between the CPU and GPU memory via PCI-e data bus shown in Figure 2 ensure seamless communication, which is crucial for iterative MRI reconstruction algorithms. Furthermore, GPUs support the execution of complex algorithms employed in MRI, such as parallel imaging techniques like SENSE (Sensitivity Encoding) [12] and Compressed Sensing (CS). These algorithms involve complex calculations that can be executed efficiently through GPU parallelism.

The complicated and data-intensive nature of MRI algorithms presents an enormous computational challenge in post-processing raw MRI data to obtain clinically feasible MR images. As the demand for higher resolution, faster acquisitions, and advanced reconstruction methods increases, traditional CPU-based processing is not able to meet the required demand. The increasing complexity of imaging protocols, coupled with the need for real-time or near-real-time analysis, exacerbates the situation. CPUs are versatile for general-purpose computing; their fewer cores and general-purpose architecture may result in slower processing speeds and reduced efficiency for specialized workloads. Ultimately, GPUs offer a significant performance advantage in medical imaging due to their ability to handle the specific demands of MRI data processing. To further elaborate the dominance of GPU based computation over CPU based computation let assume two matrices.

$$\text{Matrix A (1000x1000):} \quad A \in \mathbb{R}^{1000 \times 1000} \text{----------(i)}$$

$$\text{Matrix B (1000x1000):} \quad B \in \mathbb{R}^{1000 \times 1000} \text{----------(ii)}$$

Simple multiplication operation is required to perform on both matrices and resultant data will be stored in matrix C as shown in equation (iii). Where C is the resulting matrix of dimensions 1000x1000:

$$C = A * B \text{----------(iii)}$$

The algorithm for performing matrix multiplication using CPU is shown in table 1. Algorithm involves three nested loops iterating over the indices i, j, and k ranging from 1 to 1000. For each pair of indices i and j, the algorithm calculates the element C[i][j] by performing a summation of products between elements from matrices A and B. The resulting element C[i][j] accumulates these products over the innermost loop, showcasing a cubic time complexity of $O(n^3)$, where n represents the size of the matrices.

Table 1:CPU-Based Computation Algorithm for matrices multiplication requires nested loop.

```
for i = 1 to 1000:
   for j = 1 to 1000:
      C[i][j] = 0
      for k = 1 to 1000:
         C[i][j] += A[i][k] * B[k][j]
```

The parallelized version of the matrix multiplication algorithm is shown in table 2. Each element in the resulting matrix C can be computed independently by a separate GPU thread. Each thread, identified by indices i and j, calculates the element C[i][j] by summing the products of corresponding elements from matrices A and B, where k ranges from 1 to 1000. This parallel execution significantly reduces the time complexity to $O(n^2)$ by utilizing the GPU's multiple cores and simultaneous arithmetic operations.

Table 2: GPU-Based Computation Algorithm for matrices multiplication

$\boldsymbol{A}$ as an $n \times 1000$ Matrix
$\boldsymbol{B}$ as a $1000 \times m$ Matrix
Resulting Matrix $\boldsymbol{C}$ will be $n \times m$ Matrix.
$\boldsymbol{C[i][j]}$ represents the element in the $\boldsymbol{i^{th}}$ row and $\boldsymbol{j^{th}}$ column of matrix $\boldsymbol{C}$.
The dot product of the $\boldsymbol{i^{th}}$ row of matrix $\boldsymbol{A}$ and the $\boldsymbol{j^{th}}$ column of matrix $\boldsymbol{B}$ *is required to compute* $\boldsymbol{C[i][j]}$.
Thread i, j computes: C[i][j] = Σ (A[i][k] * B[k][j]) for k = 1 to 1000
Where:
$\boldsymbol{A_i} = \{\boldsymbol{A[i][1], A[i][2], \ldots A[i][1000]}\}$ represents the $\boldsymbol{i^{th}}$ row of Matrix $\boldsymbol{A}$.
$\boldsymbol{B_i} = \{\boldsymbol{B[1][j], B[2][j], \ldots B[1000][j]}\}$ represents the $\boldsymbol{j^{th}}$ column of Matrix $\boldsymbol{B}$.
This dot product involves multiplying corresponding elements of $\boldsymbol{A_i}$ and $\boldsymbol{B_i}$ than summing up these products from k = 1 to 1000, the resulting sum is the value of $\boldsymbol{C[i][j]}$.
This calculation is repeated for each i and j in the resulting matrix $\boldsymbol{C}$, and each computation of $\boldsymbol{C[i][j]}$ is independent of the others. Also, there are no data dependencies making it amenable to compute it in parallel fashion with multiple threads.

The Compute Unified Device Architecture (CUDA) and Open Computing Language (OpenCL) are two parallel computing frameworks [17-20]. Both have played important roles in advancing GPU programming for medical imaging applications, including MRI. CUDA and OpenCL were designed for multi-platform use, provide programmers with powerful tools to use the parallel processing capabilities of GPUs for computational tasks.

CUDA, specifically designed for NVIDIA GPUs, offers an extensive set of libraries, APIs, and development tools [18]. Its architecture enables direct access to GPU hardware, allowing programmers to design and optimize algorithms for specific tasks in medical imaging. CUDA's close integration with the GPU architecture facilitates efficient data movement and computation, making it particularly effective for accelerating iterative algorithms, image reconstruction, and deep learning techniques applied in MRI data processing.

OpenCL [20] is a more versatile framework than CUDA that supports GPUs from different vendors, enabling cross-platform development. Its flexibility makes it a compelling choice for medical imaging software developers seeking to deploy applications on diverse hardware. OpenCL abstracts the underlying hardware, allowing developers to target both GPUs and other accelerators like CPUs or FPGAs [20]. Although it may not provide the same level of optimization as CUDA for specific GPU architectures, OpenCL's broad compatibility makes it a suitable choice for larger-scale applications where hardware heterogeneity is a consideration.

4. GPU Acceleration in MR Imaging:

Parallelization of data acquisition has emerged as a promising strategy for accelerating the raw data collection process in MRI. While GPUs are not directly involved in the data acquisition itself, they play a pivotal role in expediting subsequent data processing stages. To illustrate, consider the study conducted by Wu et al. [21] which focuses on real-time MRI reconstruction using GPU acceleration. In this research, the authors present a method that utilizes the parallel computing power of GPUs to accelerate the reconstruction of dynamic MRI images. By distributing the computational workload across multiple GPU cores, they achieve remarkable reductions in processing time, allowing for near-instantaneous image generation during data acquisition. This demonstrates how GPUs can significantly enhance the efficiency of MRI data processing, paving the way for real-time imaging applications and more timely clinical diagnoses [47][48].

Conventional MRI scans mostly generate images after post-processing but real-time visualization during scans can provide immediate feedback and enable adjustments for optimal data collection. GPU-based techniques, such as those demonstrated in the study by Uecker et al. [22], have achieved significant acceleration in the iterative image reconstruction process, allowing for on-the-fly visualization of dynamic processes within the body. This dynamic imaging capability not only enhances clinical decision-making but also accelerates advanced research in areas like functional MRI. By using GPUs' parallel processing prowess, real-time imaging techniques in MRI are poised to transform the way we observe and understand physiological changes in real-time.

Image reconstruction in MRI has made remarkable strides in acceleration through the integration of GPUs, enabling the enhancement of both iterative and non-iterative methods. In a significant contribution by

Hammad et al. [12], titled "QR-decomposition based SENSE reconstruction using parallel architecture," the authors introduce a novel approach that deployed GPUs for parallel imaging reconstruction. By utilizing the computational power of GPUs, their QR-decomposition-based SENSE method accelerates the reconstruction process, yielding significant reductions in processing time. QR decomposition-based SENSE reconstruction is an innovative technique and provides enough evidence to prove that GPUs have the potential to revolutionize MRI image reconstruction, driving towards more efficient and timely clinical diagnoses. Furthermore, in the context of non-iterative parallel imaging reconstruction, Lustig et al. [23] illustrate the potential of GPUs in their work on " Compressed Sensing Parallel Imaging MRI: Scalable Parallel Implementation and Clinically Feasible Runtime." In their work, GPUs play an active role in expediting image reconstruction with their parallel processing capabilities. These studies jointly evident the transformative impact of GPUs on MRI image reconstruction, ultimately propelling advancements in both clinical practice and research domains.

Magnetic Resonance Fingerprinting (MRF) [24-25] is an innovative MRI technique that simultaneously acquires and analyzes multiple tissue parameters, such as relaxation times and proton density, in a single rapid scan, providing comprehensive tissue characterization and enabling quantitative imaging. MRF utilizes specialized pulse sequences and advanced reconstruction algorithms to generate unique "fingerprints" for different tissues, enhancing diagnostic capabilities and expanding applications in research and clinical settings [24]. Despite MRF potential benefits in the field of MRI, MRF faces several challenges when implemented on MRI scanners. One key challenge is the increased demand for computational resources due to the complex pulse sequences and iterative reconstruction methods used in MRF. This can lead to longer reconstruction times and may require specialized hardware, such as GPUs, to accelerate processing. The study by Omer et al. [25], titled "GPU accelerated grouped magnetic resonance fingerprinting using clustering techniques," showcases another facet of GPU utilization in MRI. The researchers demonstrate the effectiveness of GPUs in rapidly reconstructing MRF data through the application of clustering techniques. By exploiting the parallelism in MRF using GPUs, the researchers achieved accelerated data processing, contributing to the efficiency and speed of a cutting-edge imaging technique [25-26]. This example further highlights the versatility of GPUs in MRI, from accelerating image reconstruction to enhancing the efficiency of advanced imaging methodologies.

Denoising and artifact reduction are the other critical challenges in MRI that directly impact image quality and diagnostic accuracy. In recent years, GPU has emerged as a powerful tool in addressing these challenges by enabling advanced processing techniques. Utilizing the parallel processing capabilities of GPUs, researchers have developed innovative algorithms for denoising and artifact reduction that substantially enhance image clarity and fidelity. For instance, the study by Li et al. [27], " An automatic restoration framework based on GPU-accelerated collateral filtering in brain MR images," employs GPU-accelerated deep learning neural networks to effectively denoise and restore images acquired under noisy conditions. Similarly, the work by Z. Wang et al [28]., "One-Dimensional Deep Low-Rank and Sparse Network for Accelerated MRI," highlights the important role of GPUs in the implementation of the proposed deep learning network. The complex computations involved in the proposed deep learning network, such as convolutional operations and matrix manipulations, are distributed across the numerous cores of GPUs. This parallelization significantly accelerates the training and inference processes, enabling rapid

reconstruction of MRI images from under-sampled data. Ultimately contributing to the production of high-quality MRI images that can lead to more accurate diagnoses and improved patient care.

5. Advanced Imaging Techniques:

Diffusion Tensor Imaging (DTI) [29] has become a vital tool in studying the microstructural organization of biological tissues, particularly in the brain's white matter. However, the computational demands of processing and analyzing DTI data can be substantial, making efficient techniques is essential for getting maximum benefit. GPU acceleration has emerged as a novel solution in this context. GPUs' parallel processing architecture aligns seamlessly with the inherent parallelism in DTI computations, such as diffusion tensor estimation and tractography. This synergy enables researchers and clinicians to expedite these processes, ultimately reducing processing times from hours to minutes.

DTI tractography, which reconstructs the pathways of nerve fibers in the brain, is a computationally intensive process. Researchers like Garyfallidis et al. [30] developed tools like DiPy, which utilize parallel processing to enhance tractography performance. With GPUs, DiPy can achieve real-time or near-real-time tractography, enabling researchers to visualize complex brain connectivity patterns quickly and efficiently.

Diffusion Tensor Estimation (DTE) [31] is the process of estimating diffusion tensors from raw MRI data. DTE is a highly computation intensive process due to the complex nature of the underlying mathematical operations and the large volume of data involved in the estimation. The tensor eigenproblem is vital in various fields and has gained recent interest from both mathematical and application-specific communities [32]. A method known as the shifted symmetric higher-order power method (SS-HOPM), an extension of the matrix power method for symmetric tensors, was introduced by Kolda and Mayo [33]. Research study conducted by Yaniv et al. [32] concentrates on effectively implementing the SS-HOPM algorithm, particularly using GPU acceleration for scenarios involving numerous small tensor eigenproblems. The primary motivation of this research is the identification of nerve fibers in brain tissue using diffusion-weighted MRI data. This involves processing data from countless tiny cubic millimeter-sized units called voxels [32]. Each voxel requires the solution to a small tensor eigenvalue problem to determine the count and orientations of nerve fiber bundles within it. Due to the independent nature of voxel computations, parallelism is feasible, and hence, significant improvement is achieved by the author after implementing the algorithm on the GPU based parallel framework [32]. However, reduction in the processing time is achieved at the cost of adding another layer of complexity to the computation.

Functional Magnetic Resonance Imaging (fMRI) [3] is another MRI based neuro imaging technique just like DTI but it primarily measures changes in blood oxygenation levels associated with brain activity. It detects brain regions that are activated during various cognitive tasks or resting states [3]. The analysis of fMRI data involves preprocessing for the removal of artifacts, statistical analysis to identify activated regions, and possibly connectivity analysis to study how brain regions interact. While its insights into cognitive processes and brain functions are valuable, the computational complexity of fMRI analysis presents challenges. The complexity arises from handling large datasets, performing statistical tests, and dealing with the inherent noise in the signal. fMRI data acquisition generates large datasets with high spatial and temporal resolutions, demanding intensive storage and processing resources. Preprocessing involves tasks like motion correction, spatial normalization, and noise reduction, which require sophisticated

algorithms and significant computation [33]. Statistical analysis is often performed using complex regression models, and subsequent corrections for multiple comparisons contribute to the computational load. Moreover, tasks like functional connectivity analysis require complex mathematical calculations. These complexities emphasize the significance of computational tools, including GPUs, in accelerating fMRI analyses [33]. With optimized algorithms and hardware acceleration, the computational bottlenecks in fMRI processing can be addressed, advancing our understanding of brain function and disorders.

Dynamic Contrast-Enhanced MRI (DCE-MRI) [34] is a technique used to assess tissue perfusion and microvascular properties by tracking the uptake and distribution of contrast agents over time. From a computational complexity perspective, DCE-MRI involves acquiring a sequence of images over multiple time points, often resulting in large datasets [35]. The analysis of DCE-MRI data encompasses several stages, including preprocessing, pharmacokinetic modeling, and quantitative parameter estimation. Computational complexity arises from tasks such as motion correction, image registration, iterative optimization for model fitting, and mathematical operations to derive perfusion-related parameters [35]. GPU-based perfusion analysis enables researchers and clinicians to efficiently process and interpret large datasets, contributing to improved understanding of tissue perfusion dynamics, aiding in clinical decision-making, and advancing medical research [36].

The usability of Machine Learning (ML) techniques [37-39] in MRI has revolutionized the field by providing advanced tools for data analysis, interpretation, and decision-making. ML promotes computational algorithms that enable computers to learn and make predictions or decisions based on training data [5]. These algorithms can automatically extract complex patterns, relationships, and features from complex MRI datasets that might not be easily detectable to naked eye [37-39]. In recent past, different MRI based studies were conducted to predict the impact of different activities such as sports, memorization etc. on health of brain, to improve the one's lifestyle [40-42]. All of those techniques are based on segmentation and feature extraction. Segmentation and feature extraction are critical steps in almost all MRI analysis/techniques (including techniques that involve ML), allowing researchers and clinicians to delineate region of interest and extract meaningful quantitative information from MR images [5][40]. With the increasing complexity of MRI datasets and the demand for near-real-time processing, GPUs have emerged as valuable tools to accelerate these tasks. GPU-based segmentation methods, such as region resizing, thresholding, and clustering, enable rapid identification and separation of anatomical or pathological regions within MR images [25]. Feature extraction, which involves quantifying characteristics like intensity, shape, or texture, are benefited by parallel and multicore architecture of GPUs [43]. Hence. GPUs efficiently process large datasets and get specific information in significantly less time. This acceleration is particularly relevant in applications such as tumor detection, brain tissue classification, and functional region localization. GPU-based segmentation and feature extraction enhance the accuracy, efficiency, and scalability of MRI analysis, paving the way for more advanced clinical diagnostics and research insights [37][43].

6. Challenges and Future Directions:

A major limitation of GPU-based MRI processing is the constraint due to the limited memory capacity of GPUs [17] and can be quite challenging especially when handling large MRI datasets. Addressing memory limitations and data transfer bottlenecks is crucial when implementing GPU-based MRI processing. GPUs offer exceptional parallel processing capabilities, if utilized appropriately will overcome the limited memory capacities. Efficient memory management and data movement strategies are essential to get maximum GPU performance [18]. Techniques such as memory optimization, data compression, and smart memory allocation can reduce memory constraints [17-19]. One such technique (based on clustering [25]) to avoid memory limitation of GPU was successfully implemented for memory intensive MRF algorithm and get significant improvement in term of computational complexity and speedup.

Another primary concern is the overhead incurred during data transfer between the CPU and GPU [12-13]. This data movement involves memory copying and communication latency, which can impact overall performance. To mitigate this, techniques like data prefetching, overlap of computation and data transfer, and utilizing high-speed interconnects can improve overall efficiency [18-20]. Moreover, employing GPU-resident memory, such as High Bandwidth Memory (HBM), can improve data transfer bottlenecks and enhance processing speed [44].

Coordinating tasks and ensuring data integrity between the CPU and GPU introduces synchronization complexities, particularly in scenarios where frequent data exchanges are required. Moreover, achieving optimal load balancing to fully utilize both processing units while avoiding resource underutilization can be difficult.

A Collaboration between hardware experts, software developers, and domain specialists is vital to address challenges mentioned in previous paragraphs, Optimized algorithms, parallelization, and GPU-specific programming frameworks like CUDA or OpenCL can help maximize GPU performance while minimizing memory and data transfer overheads. Furthermore, advances in GPU architecture, like larger memory capacities and faster interconnects, continue to aid in overcoming these limitations and pushing the boundaries of GPU-based MRI processing [17].

A hybrid approach that combines the strengths of both CPUs and GPUs to process raw MRI data can address some of the major limitations of CPU- GPU based MRI processing. In hybrid approach, CPU's handle complex control flow while GPUs execute computationally intensive operations in parallel. If the workload is properly distributed, it will result in efficient utilization of hardware resources and overall improvement of large-scale data processing with reduced memory bottlenecks. Techniques like task parallelism and data partitioning are essential in optimizing the workload distribution [18].

The incorporation of GPUs for accelerating MRI data analysis presents a compelling opportunity for advancements in medical research. However, this technological progress raises notable ethical and privacy concerns. One such concern is patient data security, as the fast-processing capabilities of GPUs require robust storage and transmission mechanisms, therefore preventing unauthorized access to sensitive patient data is a bit challenging. Maintaining patient privacy is equally critical because MRI data often contains sensitive health information. Implementing effective de-identification techniques and adhering to

regulatory frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) are vital to safeguard patient anonymity [45].

Another significant concern is the potential for algorithmic bias within GPU-accelerated MRI analysis [46]. Since algorithms hold a important role in interpreting a patient's health condition, any biases embedded within them could yield conflicting outcomes that might misguide the entire treatment procedure. Ensuring fairness and mitigating bias requires periodically updating and rigorous testing of algorithms. Furthermore, Patients need to be fully aware about the privacy concerns when their data is processed (as it requires storage for some time) using GPU framework and agree to it. Overall, addressing ethical and privacy concerns is vital to get full potential of GPU-accelerated MRI data analysis while maintaining individual rights and data integrity.

Accelerated MRI techniques, often made possible using GPU-based processing. They have profound clinical applications and implications that impact patient care, diagnosis, and medical research. These techniques offer the potential to significantly reduce scan times while maintaining/improving image quality, leading to several important benefits:

Patient Comfort and Compliance: MRI scans can be uncomfortable and claustrophobic for some patients, especially those who find it challenging to remain still for extended periods. Accelerated MRI techniques can shorten scan durations, making the experience more tolerable for patients and reducing the need for sedation.

Pediatric Imaging: Children often have difficulty staying still during MRI scans, which can lead to motion artifacts. Accelerated MRI techniques enable faster scans, reducing the likelihood of motion-related artifacts and the need for repeated scans in pediatric patients.

Emergency Cases: In emergency situations, quick diagnosis is crucial. Accelerated MRI allows for faster image acquisition, which can be essential for patients with acute conditions such as traumatic injuries, strokes, or internal bleeding.

High Patient Throughput: In busy clinical settings, such as hospitals and imaging centers, accelerated MRI techniques can increase patient throughput by reducing the time required for each scan. This is particularly valuable for institutions that have high demand for MRI services.

Real-Time Imaging: Accelerated MRI can enable real-time imaging of dynamic processes within the body, such as cardiac motion and joint movements. This capability is useful for guiding procedures, interventions, and surgeries.

Functional Imaging: Techniques like functional MRI (fMRI) can benefit from accelerated imaging, allowing researchers and clinicians to capture rapid changes in brain activity during tasks or stimulation.

Monitoring Disease Progression: For patients with chronic conditions that require regular monitoring, accelerated MRI techniques can make the scanning process more efficient and less burdensome, facilitating longitudinal studies of disease progression.

Improved Image Quality: In some cases, accelerated MRI techniques can also improve image quality by reducing motion artifacts and susceptibility effects. This is especially important for obtaining clear images in challenging areas of the body, such as the lungs or abdomen.

Clinical Research: Accelerated MRI can enable researchers to conduct more studies within a given timeframe, leading to faster advancements in medical research. Clinical trials involving MRI can benefit from reduced scan times and improved patient comfort.

Diffusion Imaging: DTI and related techniques can benefit from accelerated MRI, allowing for more comprehensive and higher-resolution mapping of white matter tracts in the brain.

Multi-Contrast Imaging: In cases where multiple types of contrasts are needed for accurate diagnosis, accelerated MRI can make it more feasible to acquire multiple sets of images in a single scan session.

Cost Efficiency: Shorter scan times can reduce operational costs by optimizing resource utilization and increasing the number of scans that can be performed in a day.

It's important to note that while accelerated MRI techniques offer significant advantages, they also come with considerations such as potential trade-offs in image quality and the need for specialized sequences and equipment. Therefore, the implementation of these techniques requires careful validation and optimization to ensure that clinical diagnoses are accurate and reliable.

Overall, accelerated MRI techniques have the potential to revolutionize the field of medical imaging by providing faster, more efficient, and more patient-friendly imaging solutions for a wide range of clinical applications.

7. Conclusion

The synergic combination of MRI and GPUs has revolutionized medical imaging. GPU acceleration has significantly enhanced MRI techniques, enabling real-time imaging, advanced image reconstruction, denoising, and analysis. Fundamental techniques like DTI and fMRI have seen remarkable improvements through GPU processing. The parallel computing power of GPUs, combined with frameworks like CUDA and OpenCL, has addressed computational challenges, accelerating MRI data processing.

Despite challenges like memory constraints and data transfer bottlenecks, the hybrid approach of CPU-GPU collaboration, memory optimization, and algorithmic refinement has improved overall efficiency. Ethical considerations include patient data security and algorithmic bias, requiring robust data protection and fairness measures.

Accelerated MRI techniques have led to faster scans, improved image quality, and real-time insights, benefiting patient care and medical research. The encouraging outcomes achieved thus far inspire the research community to further explore, innovate, and refine the combination of MRI with GPU based parallel frameworks. The incorporation of GPUs with MRI is shaping the future of MR imaging, driving advancements and promising new horizons in diagnosis, treatment, and understanding of human health.

References

[1] McRobbie, D.W., Moore, E.A., Graves, M.J. & Prince, M.R. *MRI from Picture to Proton*, second ed., Cambridge University Press. ISBN: 978052186572, 2006.

[2] Schwenzer, N.F. et al. *Non-invasive assessment and quantification of liver steatosis by ultrasound, computed tomography, and magnetic resonance*. Journal of hepatology, **51**(3), 433-445 (2009).

[3] Matthews, P.M. & Jezzard, P. *Functional magnetic resonance imaging*. Journal of Neurology, Neurosurgery & Psychiatry, **75**(1), 6-12 (2004).

[4] Li, L., Ding, W., Huang, L., Zhuang, X. & Grau, V. *Multi-modality cardiac image computing: A survey*. Medical Image Analysis, 102869 (2023).

[5] Irmak, E. *Multi-classification of brain tumor MRI images using deep convolutional neural network with fully optimized framework*. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, **45**(3), 1015-1036 (2021).

[6] Kogan, F., Broski, S.M., Yoon, D. & Gold, G.E. *Applications of PET-MRI in musculoskeletal disease*. Journal of Magnetic Resonance Imaging, **48**(1), 27-47 (2018).

[7] Özsarlak, Ö., Van Goethem, J.W., Maes, M. & Parizel, P.M. *MR angiography of the intracranial vessels: technical aspects and clinical applications.* Neuroradiology, **46,** 955-972 (2004).

[8] Kowalik, G.T. *Rapid online reconstruction of non-Cartesian magnetic resonance images using commodity graphics cards* (Doctoral dissertation, UCL (University College London)) (2016).

[9] Yang, Z. & Jacob, M. *Efficient NUFFT algorithm for non-Cartesian MRI reconstruction*. Proc. IEEE International Symposium on Biomedical Imaging (ISBI), 117-20 (2009).

[10] Sorensen, T.S., Atkinson, D., Schaeffter, T. & Hansen, M.S. *Real-time reconstruction of sensitivity encoded radial magnetic resonance imaging using a graphics processing unit. IEEE transactions on medical imaging*, **28**(12), 1974-1985 (2009).

[11] Bustin, A., Fuin, N., Botnar, R.M. & Prieto, C., *From compressed-sensing to artificial intelligence-based cardiac MRI reconstruction. Frontiers in cardiovascular medicine*, **7,** 17 (2020).

[12] Ullah, I. et al. *QR-decomposition based SENSE reconstruction using parallel architecture*. *Computers in biology and medicine*, **95,** pp.1-12 (2018).

[13] Qazi, S.A., Tariq, F., Ullah, I. & Omer, H. *Parallel implementation of L+ S signal recovery in dynamic MRI. Magnetic Resonance Materials in Physics, Biology and Medicine*, **34,** 297-307 (2021).

[14] Ullah, I., Inam, O., Aslam, I. & Omer, H. Accelerating parallel magnetic resonance imaging using p-thresholding based compressed-sensing. *Applied Magnetic Resonance*, **50**(1-3), 243-261 (2019).

[15] Knoll, F. et al. Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues. *IEEE signal processing magazine*, **37**(1), 128-140 (2020).

[16] Hammernik, K. et al. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic resonance in medicine*, **79**(6), 3055-3071 (2018).

[17] Cheng, J., Grossman, M. & McKercher, T., *Professional CUDA c programming*. John Wiley & Sons, 2014.

[18] Guide, D., Cuda c programming guide. *NVIDIA, July 29*, 31 (2013).

[19] Zeller, C. Cuda c/c++ basics. *NVIDIA Coporation* (2011).

[20] Sorensen, T. & Donaldson, A.F. April. The hitchhiker's guide to cross-platform opencl application development. In *Proceedings of the 4th International Workshop on OpenCL* 1-12 (2016).
[21] Kamimura, H.A. et al. Real-time passive acoustic mapping using sparse matrix multiplication. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, **68**(1), 164-177 (2020).
[22] Uecker, M. et al. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magnetic resonance in medicine*, **71**(3), 990-1001 (2014).
[23] Murphy, M. et al. Fast $\ell_1 $-SPIRiT compressed sensing parallel imaging MRI: scalable parallel implementation and clinically feasible runtime. *IEEE transactions on medical imaging*, **31**(6), 1250-1262 (2012).
[24] Ma, D. et al. Magnetic resonance fingerprinting. *Nature*, **495**(7440), 187-192 (2013).
[25] Ullah, I., Hassan, A.M., Saad, R.M. & Omer, H., GPU accelerated grouped magnetic resonance fingerprinting using clustering techniques. *Magnetic Resonance Imaging*, **97,** 13-23 (2023).
[26] Hassan, A.M., Saad, R.M., Ullah, I. & Omer, H. GPU Accelerated Grouped Magnetic Resonance Fingerprinting using Clustering Techniques. In *Proc. Intl. Soc. Mag. Reson. Med* **29,** 1558 (2021).
[27] Chang, H.H. & Li, C.Y. An automatic restoration framework based on GPU-accelerated collateral filtering in brain MR images. *BMC Medical Imaging*, **19**(1), 1-13 (2019).
[28] Wang, Z. et al. One-dimensional deep low-rank and sparse network for accelerated MRI. *IEEE Transactions on Medical Imaging*, **42**(1), 79-90 (2022).
[29] Le Bihan, D. et al. Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, **13**(4), 534-546 (2001).
[30] Garyfallidis, E. et al. Dipy, a library for the analysis of diffusion MRI data. *Frontiers in neuroinformatics*, **8,** 8 (2014).
[31] Koay, C.G., Chang, L.C., Carew, J.D., Pierpaoli, C. & Basser, P.J. A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. *Journal of magnetic resonance*, **182**(1), 115-125 (2006).
[32] Kolda, T.G. & Mayo, J.R. Shifted power method for computing tensor eigenpairs. *SIAM Journal on Matrix Analysis and Applications*, **32**(4), 1095-1124 (2011).
[33] Cohen, J.D. et al. Computational approaches to fMRI analysis. *Nature neuroscience*, **20**(3), 304-313 (2017).
[34] Sourbron, S.P. & Buckley, D.L. Classic models for dynamic contrast-enhanced MRI. *NMR in Biomedicine*, **26**(8), 1004-1027 (2013).
[35] Quan, T.M. & Jeong, W.K. Compressed sensing reconstruction of dynamic contrast enhanced MRI using GPU-accelerated convolutional sparse coding. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* IEEE 518-521 (2016).
[36] Hachaj, T. and Ogiela, M.R. Visualization of perfusion abnormalities with GPU-based volume rendering. *Computers & Graphics*, **36**(3), 163-169 (2012).
[37] Chato, L. and Latifi, S. Machine learning and deep learning techniques to predict overall survival of brain tumor patients using MRI images. In *2017 IEEE 17th international conference on bioinformatics and bioengineering (BIBE)* IEEE 9-14 (2017).
[38] De Filippis, R. et al. Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: a systematic review. *Neuropsychiatric disease and treatment*,1605-1627 (2019).

[39] Reig, B., Heacock, L., Geras, K.J. & Moy, L. Machine learning in breast MRI. *Journal of Magnetic Resonance Imaging*, **52**(4), 998-1018 (2020).
[40] Rahman, M., Aribisala, B., Ullah, I. & Omer, H. Association between scripture memorization and brain atrophy using magnetic resonance imaging. *Acta Neurobiologiae Experimentalis*, **80**(1), 90-97 (2020).
[41] Churchill, N. et al. Brain structure and function associated with a history of sport concussion: a multi-modal magnetic resonance imaging study. *Journal of neurotrauma*, **34**(4), 765-771 (2017).
[42] Zhang, H. et al. GPU-accelerated GLRLM algorithm for feature extraction of MRI. *Scientific reports*, **9**(1), 10883 (2019).
[43] Zheng, T., Nellans, D., Zulfiqar, A., Stephenson, M. & Keckler, S.W. Towards high performance paged memory for GPUs. In *2016 IEEE International Symposium on High Performance Computer Architecture (HPCA)* IEEE 345-357 (2016).
[44] Edemekong, P.F., Annamaraju, P. & Haydel, M.J. Health insurance portability and accountability act (2018).
[45] Aja-Fernández, S. et al. Validation of deep learning techniques for quality augmentation in diffusion MRI for clinical studies. *NeuroImage: Clinical*, ***39*****,** 103483 (2023).
[46] Stone, S.S. et al. Accelerating advanced MRI reconstructions on GPUs. In Proceedings of the 5th conference on Computing frontiers 261-272 (2008).
[47] Wang, H., Peng, H., Chang, Y. & Liang, D. A survey of GPU-based acceleration techniques in MRI reconstructions. Quantitative imaging in medicine and surgery, **8**(2), 196 (2018).