CHAPTER 2

BACKGROUND OF METHODOLOGIES

Chapter Highlights

- ***Magnetic resonance spectroscopy (MRS)***
- ***Thresholding Methods***
- ***Wavelet Transform***
- ***Machine Learning***
- ***Deep Learning***

This chapter provides an overview of concepts and methods used in studies performed and presented throughout the thesis, followed by an account of usage and applications of these concepts and methods in the area of MR spectroscopy addressing different issues in acquired spectra for a better understanding, in terms of both qualitative as well as quantitative measures.

2.1 Magnetic Resonance Spectroscopy

Magnetic resonance spectroscopy (MRS) is a promising non-invasive and non-ionizing methodology of studying and quantifying the biochemical information of tissue metabolites from the anatomical region of interest[1]. This method can aid in early diagnosis of different types of cancer pathology, study treatment progression or therapy. Because of these attributes, this valuable tool is also sometimes referred to as "virtual biopsy"[2]. An in-vivo MR spectrum in the time domain can be expressed as

$$\mathrm{y(t)} = \sum_{k=1}^{K} a_k e^{(j\phi_k)} e^{(-d_k t + 2\pi f_k t)} + B(t) + w(t) + \epsilon(t) \qquad (2.1)$$

where K denotes the number of metabolites, a_k represents the amplitude, ϕ_k signifies the phase shift, d_k is the damping correction, and f_k denotes frequency shifts due to the field inhomogeneity of the k^{th} metabolite; $B(t)$ denotes the baseline due to macromolecules/lipid contamination, $w(t)$ marks the water resonance and, $\epsilon(t)$ is the white noise with standard deviation σ.

MRS finds its application in both research as well as clinical setup, and an increasing number of research publications in the last two decades ensures its near acceptance as an important clinical diagnostic tool. Post-acquisition of signals, the standard workflow of an in-vivo MRS study, follows three steps: pre-processing, spectral analysis, and quantitation. Pre-processing of an acquired MR spectrum is fundamental for precise analysis and subsequent quantitation of the concerned metabolites, as along with the chemical information, a significant amount of irrelevant information like noise, spikes, spurious echoes are also present. [3]

An acquired *in-vivo* time-domain MRS spectrum, also called free induction decay (FID), is a discrete, complex-valued signal. The real and imaginary part are called absorption and dispersion spectra. Mathematically, it can be represented as a parametric linear combination model given by:

$$S(t) = S_X(t) + S_M(t) + \eta \quad (2.2)$$

where,

$$S_X(t) = \sum_{k=1}^{K} A_k . B_k . \exp(j\varphi_k) \exp(-d_k t + 2\pi j f_k t) \quad (2.2.1)$$

$$S_M(t) = \sum_{m=1}^{M} A_m . \exp(j\varphi_m) . F_g(f_c, t) \quad (2.2.2)$$

Here, $S_X(t)$ is the summed metabolite spectra and $S_M(t)$ is the summed macromolecular contribution. $A_k,\ B_k, \varphi_k, d_k, f_k$ are amplitude, metabolite basis profile, phase-shift, damping, and frequency shift of k^{th} metabolite component respectively. A_m, is the amplitude of the m^{th} macromolecular component, F_g is the gaussian basis for macromolecular components, and η denotes gaussian noise of mean zero and standard deviation σ, and other artifacts. t is the discrete time samples.

For an *in-vivo* MR spectrum, there are two major obstacles because of strong metabolite-macromolecule overlap and high noise content: (1) this inverse problem of fitting is ill-conditioned, and (2) efficient reconstruction of small metabolite peaks embedded in noise in frequency spectrum.

2.2. Sparsity in Signal Processing

Sparsity is a key notion in signal processing that refers to the property of a signal or representation of a signal in which a significant number of its coefficients or elements are zero or unimportant. A sparse signal, in other words, is one that may be represented by a minimal number of non-zero coefficients or elements. It is important in signal processing because it enables for efficient signal encoding, compression, and processing. It is commonly used in signal processing techniques and algorithms like compressed sensing, sparse coding, and sparse recovery. The fundamental idea behind sparsity is that many natural or real-world signals have concentrated energy or information embedded in a few large components, whereas the remainder have relatively lesser energy or are considered noise-like. Unnecessary or redundant information can be deleted or efficiently compressed by representing and processing signals in a sparse domain, resulting in more efficient data representation and perhaps reduced computational complexity. Sparse signal representations can be achieved by employing different sparsity-promoting bases or dictionaries. [4-5]

In the context of norm-based sparsity, the norm employed is often a measure of the amplitude or energy of the signal. The L1 norm (also known as the Manhattan norm or the absolute norm) and the L2 norm (also known as the Euclidean norm or the least squares norm) are two commonly used norms. The L1 norm is the sum of the absolute values of the signal's coefficients or constituents. Because it supports sparse representation, it is frequently employed as a metric of sparsity. A signal with a lower L1 norm has a lower number of meaningful non-

zero coefficients, indicating a higher level of sparsity. The L2 norm, on the other hand, calculates the square root of the sum of the squares of the signal's coefficients or elements. It represents the signal's magnitude or energy. Even though it does not directly measure sparsity, it can nevertheless be relevant in the context of norm-based sparsity analysis. [6-11]

The sparsity exhibited by a signal or representation in a certain transform domain is referred to as transform-based sparsity. It entails encoding a signal in a transform domain in which a significant fraction of its coefficients or elements become zero or insignificant resulting in a sparse representation. Transform-based sparsity is frequently achieved by employing transforms such as the Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT), or other sparsifying transforms such as the Total Variation (TV) transform or the Sparse Representation Transform (e.g., dictionary-based sparse representations).

For instance, the DFT can offer a sparse representation for frequency-domain signals. However, the DWT is well-known for its ability to capture localised or sparse features in the time-frequency plane. Sparsity based on transforms is commonly used in signal processing applications. Transform coding approaches, such as JPEG and MPEG, use the sparsity of signals in the transform domain to obtain high compression ratios in image and video compression.

The suitable transform is chosen based on the characteristics of the signal and the task at hand. The capacity to represent signals sparsely in transform domains has revolutionized signal processing, allowing for more efficient and accurate algorithms in disciplines such as image processing, audio and speech processing, communication systems, medical imaging, and many more.

2.3. Wavelet transform

It is a powerful multi-scale signal analysis tool which decomposes a signal into its components of varied resolution by scaling and shifting a localized-support mother wavelet function. The selection of a wavelet basis to decompose a signal in terms of wavelet and scaling function depends upon the nature of signal under investigation. Replacing the infinitely oscillating basis of Fourier transform (FT) with locally oscillating real wavelet basis for Discrete wavelet transform (DWT) provides optimal and sparser representation of signals containing singularities. But real wavelets suffer from oscillations, shift variance, aliasing, and directionality, noting that FT does not suffer from these shortcomings. In context with noise reduction or feature extraction from complex signals, wavelets have several advantages. With its ability to simultaneously capture time-frequency information, it works better when sharp changes/spikes, non-uniform noise background present as expected with MRS data and can be adaptive to large variations as well smooth changes. With multi-resolution approach, we can decompose the data into different scales and resolutions, and be able to preserve and isolate key signal components from noisy data using thresholding and other non-linear methods making it more robust approach to denoise the data faithfully.

2.3.1. Double Density Dual Tree Discrete Wavelet Transform (DDDTDWT)

The DDD-DWT [12] is a modified version of the DTDWT [13] that overcomes some of its drawbacks, such as shift variance and aliasing artefacts. It improves the DT-DWT by increasing the degree of oversampling, which improves the transform's temporal and frequency localization features. This is accomplished by including an additional level of sample at each decomposition level. Like the DT-DWT, the DDD-DWT employs two sets of wavelet filters for each level of decomposition. These filters are designed to have complimentary frequency responses, which eliminates aliasing problems and increases the transform's directional selectivity. The fundamental difference between the DDD-DWT and the DT-DWT is that it

samples the signal at twice the rate. Oversampling allows for a denser representation of the signal at each level, resulting in improved localization in both the time and frequency domains.

The DDD-DWT has shown promise in a variety of image processing applications, including denoising, edge detection, and feature extraction. It improves upon the standard DWT and DT-DWT in terms of maintaining fine details and properly localising signal characteristics.

2.3.2. Rational-Dilation Wavelet Transform (RADWT)

This wavelet transform belongs to a family of over-complete, rational dilation-based (i.e., non-dyadic) transforms[14]. The term "rational dilation wavelet transform" refers to a wavelet transform variation that employs rational dilation factors rather than integer or fractional dilation factors. The dilation factor in the standard wavelet transform specifies the scale at which the wavelet function is analysed. By allowing for non-integer scale factors, rational dilation factors may give a more flexible and fine-grained analysis. Unlike most of the existing over-complete WTs, where over-completeness is attained only by increasing the temporal sampling in frequency bands, overcomplete RADWTs increase sampling both in time and frequency bands providing different redundancy factors and better resolution of the signal. The implementation of this transform is based on an FFT-based filter bank which provides greater design flexibilities and can be utilized to generate several wavelet attributes that is difficult to realize with FIR filter-bank-based transforms. An array of Q-factors, frequency resolution and redundancy factors are achievable with this family of WT. This could result in better time-frequency localization qualities or other desirable aspects in certain applications. Because of these properties, this approach of wavelet decomposition has been extensive used over MRS signals.

2.3.3. Dual Tree Complex Wavelet Transform (DTCWT)

Complex wavelet transform proposes Fourier like complex basis representation of scaling and wavelet function. DTCWT takes real and imaginary wavelet functions as individual orthonormal bases forming two filterbank trees (dual tree approach) and capable of overcoming the above issues for real as well as complex signals with only 2x redundancy for 1D signals. The detailed information about DTCWT implementation can be obtained from [15]. The scaled coefficients capture local information and noise from different parts of spectra which may not be considered globally, and training over individual scale-coefficients may help in reducing variance. Also, a level thresholding over the coefficients reduces the noise and provide a sparse representation of data to reduce computational complexities as well as faithful reconstruction of denoised spectra. The DT-CWT performs better than the real-valued wavelet transforms in various ways, including enhanced shift invariance, directionality, and better representation of signals with complex structures or features. Image processing, image analysis, denoising, feature extraction, texture analysis, and object detection are some of the applications. Since MRS signals are complex-valued signals, in the present study, DTCWT was chosen for the wavelet decomposition and optimal presentation of time-frequency response coefficients as features for Machine Learning (ML)/ Deep Learning (DL) models in this thesis.

2.4. Wavelet thresholding

Wavelet thresholding methods are a type of signal and image denoising, compression, and sparse signal representation approach. These approaches take advantage of the sparsity property of signals in wavelet transform domains to remove noise or reduce the number of coefficients while keeping significant features.

Wavelet thresholding works by applying a thresholding operation on the wavelet coefficients. The magnitude of each coefficient is compared to a predefined threshold value during the thresholding procedure. If the magnitude is less than the threshold, the coefficient is set to zero,

removing it from the representation. If the magnitude is greater than the threshold, the coefficient is kept.

There are various common thresholding methods, which include:

a) Hard Thresholding: In hard thresholding, every coefficient whose magnitude is less than the threshold is set to zero, while the rest coefficients stay intact.
b) Soft Thresholding: Soft thresholding is like hard thresholding in the sense that the coefficients below the threshold are decreased towards zero by a predetermined amount rather than being set to zero. This mild shrinkage preserves vital features while also decreasing noise.
c) Minimax: It is a wavelet thresholding approach for determining the best threshold value for denoising signals. This criterion tries to choose a threshold that achieves the optimum overall denoising performance in terms of balancing noise reduction and signal retention. The aim to handle uncertainty in the statistical features of the signal and noise motivates the minimax thresholding criterion. It offers a conservative approach to denoising by considering the most adverse conditions for preserving signal features.
d) SureShrink: It is a threshold selection strategy aimed at reducing the Stein Unbiased Risk Estimate (SURE). SURE is a statistical measure that calculates the mean squared error between the original and thresholded signals. SureShrink dynamically adjusts the threshold based on the wavelet coefficients' local properties to achieve optimal results.

These thresholding approaches can be used over wavelet coefficients produced from different wavelet transforms, such as the discrete wavelet transform (DWT), stationary wavelet transform (SWT), and dual tree complex wavelet transform (DT-CWT). In wavelet thresholding, the threshold value must be carefully chosen. It can be determined manually using

previous knowledge or adaptively using statistical approaches or information criteria such as SURE etc.

Overall, wavelet thresholding methods are useful for denoising and compressing signals by utilising the sparsity in wavelet transform domains. They provide a good blend of noise reduction and signal retention, making them popular in a variety of signal processing applications.

2.5. Machine learning

2.5.1. Support Vector Regressor

A support vector regressor (SVR) is a support vector machine (SVM)-based regression algorithm [16] that finds a hyperplane in a high-dimensional feature space that approximates the relationships between input variables and continuous goal values. It seeks to reduce the difference between projected and actual target values while increasing the margin or distance between the hyperplane and the training data points.

It is very effective for dealing with non-linear regression problems and can handle high-dimensional data by utilising different kernel functions. The primary idea behind SVR is to strike a balance between achieving a small error on training data and keeping strong generalisation performance on unknown data. It achieves this by providing a regularisation parameter that governs the trade-off between model complexity and the amount of error tolerated in training data.

2.5.2. Random Forest

A Random Forest regressor is a machine learning method that is used for regression tasks. It is a member of the ensemble technique family, which combines numerous independent models to make predictions. Random Forest algorithm [17] is well-known for its versatility and durability, making it attractive in a variety of disciplines.

The Random Forest regression algorithm builds an ensemble of decision trees. Each tree is trained using a different random subset of the training data as well as a separate random subset of the input features. This randomization aids in reducing overfitting and improving the model's generalisation ability.

During training, each Random Forest decision tree is constructed by recursively splitting the data based on distinct feature splits. The divides are chosen to minimise the volatility of the target variable. The target variable is continually valued in regression tasks, and the predictions from each tree are combined to generate the final prediction.

Random Forest regressors have a number of advantages. They can handle both numerical and categorical data, as well as complex interactions between input and target variables. They are also less sensitive to data outliers and noise than other regression algorithms. However, Random Forest regressors can be computationally expensive, and optimal performance may necessitate careful adjustment of hyperparameters such as the number of trees and the maximum depth of each tree.

2.3.3. XGBoost

XGBoost (Extreme Gradient Boosting) [18] is a powerful and versatile algorithm that excels in predictive tasks, particularly when it comes to handling structured data and large datasets. It is a well-known machine learning technique that excels in both regression and classification tasks. It is a gradient boosting framework implementation, which is a strategy for combining numerous weak prediction models to generate a stronger overall model.

The core idea underlying XGBoost is to build an ensemble of decision trees iteratively, with each consecutive tree striving to repair the faults produced by prior trees. This iterative technique is concerned with minimising a certain loss function, such as mean squared error in regression or log loss in classification.

XGBoost includes a number of critical aspects that add to its effectiveness: (a) Gradient-based optimisation: XGBoost optimises the objective function by calculating the loss function's gradient and Hessian. This data is utilised to drive the development of each tree, resulting in more accurate forecasts. (b) Regularisation: To prevent overfitting, XGBoost applies regularisation techniques. It contains options for adjusting the complexity of individual trees, such as maximum depth, minimum child weight, and column subsampling. (c) Missing value handling: XGBoost can handle missing values in data automatically during training, removing the need for explicit data imputation. (d) Parallel processing is supported by XGBoost, allowing for efficient processing and scalability over multiple cores or distributed systems. This makes it appropriate for huge datasets.

Because of its excellent predicted accuracy and versatility, XGBoost has gained popularity and is widely employed in different data science competitions and real-world applications. It is available in a variety of programming languages, including Python, R, Java, and others, making it suitable for a variety of contexts. It is normally used after preprocessing your data, specifying the proper hyperparameters, and training the model on your training data. Following training, you can use the model to generate predictions on new, previously unknown data.

2.6. Deep Learning

2.6.1. Convolutional Neural Network (CNN)

A Convolutional Neural Network is a deep learning technique that is commonly used to analyse visual data such as photos and videos. CNNs have excelled in tasks such as image

classification, object identification, and image segmentation, whereas 1D CNNs have been used for signal analysis, peak estimation, noise and artifact removal.

CNNs are based on the concept of convolution, which is a mathematical operation that combines input data with a collection of learnable filters to generate a feature map. CNNs use convolutional layers to learn hierarchical representations of input data automatically. In the earliest layers, the network gradually learns to detect low-level features (such as edges and textures) and gradually integrates them to detect increasingly complex patterns and objects in deeper layers [19].

Some of the most important components of a standard CNN architecture are:

1. Convolutional layer: It is made up of several learnable filters that convolve with the input data to extract features. Each filter slides over the input, executing element-wise multiplication and summation operations to generate a feature map.
2. Pooling layers: Pooling layers reduce the size of feature maps by down-sampling their spatial dimensions while maintaining the most important characteristics. Max pooling (selecting the maximum value within a pool) and average pooling (getting the average value inside a pool) are two common pooling processes.
3. Non-linear activation functions, such as the ReLU (Rectified Linear Unit), are commonly used to incorporate non-linearities into the network, allowing it to describe complicated interactions between features.
4. Fully connected layers: The feature maps are flattened and fed into fully connected layers after numerous convolutional and pooling layers. These layers are in charge of generating final predictions based on the learned features.

5. Dropout: It is a regularisation technique that is often employed in CNNs. During training, it randomly sets a fraction of the connections between neurons to zero, which helps to prevent overfitting and enhances generalisation.

Using a labelled dataset, a CNN is trained by optimising its weights. To compute and update the gradients, stochastic gradient descent (SGD) or other optimisation methods are often used, together with backpropagation.

One of the primary advantages of CNNs is their ability to efficiently capture local spatial dependencies via shared weight parameters and weight sharing, making them well-suited for analysing photos and other grid-like data. Furthermore, innovations such as pre-trained CNN models (such as VGG, ResNet, and Inception) have aided transfer learning, in which pre-trained models are used as a starting point for comparable tasks, enabling effective learning with less data.

2.6.2. Long Short-Term Memory (LSTM)

LSTM architecture [20] is a form of recurrent neural network (RNN) [21] architecture developed to solve the vanishing gradient problem and capture long-term dependencies in sequential input. LSTMs are very useful for time series analysis, natural language processing, speech recognition, and other sequential data challenges. This problem is addressed by LSTMs, which include a memory cell as well as three gating mechanisms that control the flow of information through the network: the input gate, forget gate, and output gate.

The LSTM design facilitates the capture of long-term dependencies by allowing for the efficient flow of information. LSTMs can process sequences of arbitrary length while conserving useful information over time by selectively updating and keeping information via gating techniques. The parameters of the LSTM network are learned during training by optimising a loss function using gradient descent and backpropagation through time (BPTT).

Using the chain rule, the gradients are efficiently transferred through time, allowing the LSTM to learn from sequential input.

Language modelling, machine translation, sentiment analysis, speech recognition, and other applications have seen success using LSTMs.

2.6.3. U-Net

The U-Net is a convolutional neural network (CNN) architecture that was introduced in 2015 for biomedical image segmentation. It is called "U-Net" because of its U-shaped architecture, which consists of an encoder path and a decoder path.

The U-Net architecture was created primarily for accurate image segmentation, such as segmenting organs or structures in medical imaging. It is commonly employed in medical image analysis, but it has also found uses in satellite image analysis, cell segmentation, signal processing and other domains.

A high-level overview of the U-Net architecture are as follows:

I. The encoder path collects context and extracts high-level characteristics from the input image. It is made up of several convolutional and pooling layers. With each pooling procedure, the spatial dimensions are reduced while the number of channels (feature maps) is increased.

II. The decoder path uses transposed convolutions to upsample the feature maps to the original input resolution (also known as deconvolutions or upsampling). Each up-sampling procedure improves spatial dimensions while decreasing channel count. To connect the matching feature maps from the encoder path to the decoder path, skip connections are used. These skip connections aid in the recovery of detailed information that has been lost during the encoding process.

III. Expansion Path: The feature maps from the decoder path are concatenated with the equivalent feature maps from the encoder path in the expansion path. This information fusion enables the network to use both high-level context and specific spatial information.

IV. The final layer of the U-Net is commonly a 1x1 convolution, which reduces the number of channels to the necessary number of segmentation classes. To build the segmentation map, this layer uses pixel-wise categorization.

The U-Net design is well-known for its ability to capture fine features while preserving global context. Its symmetric architecture with skip connections aids in spatial information preservation and allows for exact segmentation.

Since its introduction, the U-Net architecture has undergone several modifications and improvements, including changes to the encoder and decoder paths, the incorporation of attention mechanisms, and the use of different loss functions, to improve its performance for specific tasks and datasets. Some of the modifications in U-Net architecture incorporated have used one or combination of the following methods:

a) **Residual connection:** It allows the network to learn residual mappings rather than the desired underlying mapping. These blocks are made up of a shortcut link, also known as a skip connection, that allows the network to bypass one or more layers and more efficiently disseminate gradients, preventing them from disappearing as the network goes deeper. The skip connection connects a block's original input to its output, resulting in a residual connection. The output of a residual block is the mathematical sum of the input and output of the block's convolutional layers. This connection ensures that the gradient can flow straight from the block's output to its input, allowing the network to learn residual functions that capture the changes needed to transform.

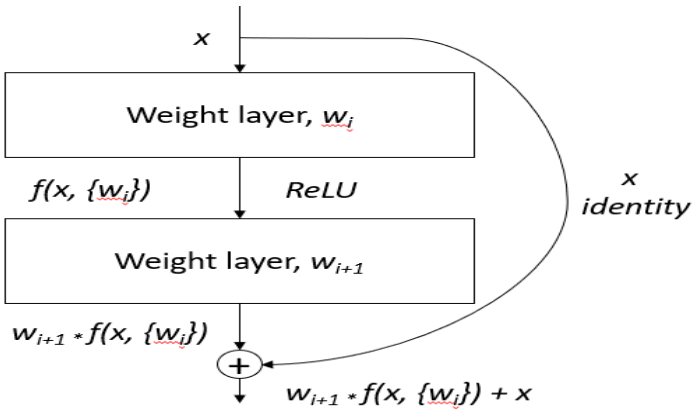

Figure 2.1: A schematic of a residual connection implemented in a Deep Learning (DL) model. The activation function (*f*) used is ReLU, w_i is the weight at i^{th} *layer*, *x* is the input.

b) **Attention mechanism**: It is a component in neural network topologies that allows the network to focus on certain areas of the incoming data while processing it. In natural language processing (NLP), picture analysis, and sequence-to-sequence tasks, attention mechanisms have received a lot of attention. Traditionally, each input element in neural networks is handled equally throughout processing. Attention mechanisms, on the other hand, enable the network to dynamically allocate computational resources and selectively attend to distinct parts of the input data. The value or relevance of various input elements can be used to guide this selective attention. The attention network encodes a sequence of input items, such as words in a phrase, pixels in an image, or peaks in a spectrum. Depending on the job, this encoding step may employ recurrent neural networks (RNNs), convolutional neural networks (CNNs), or other forms of neural networks.

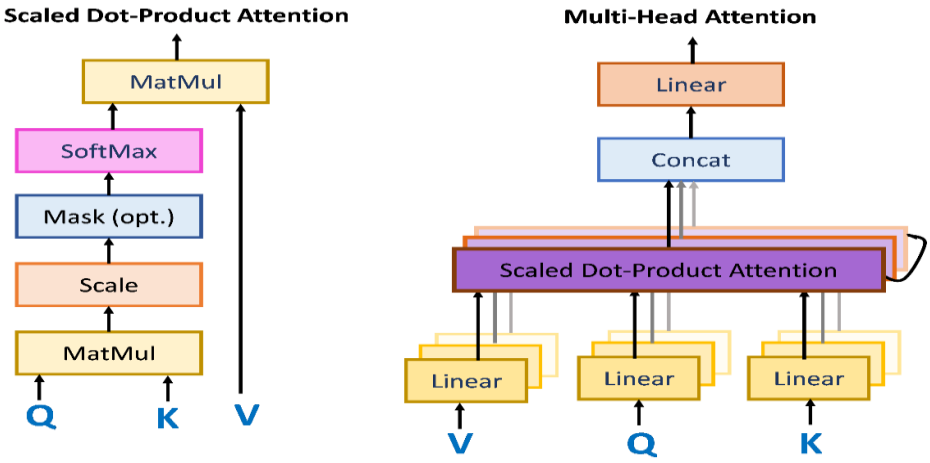

Figure 2.2: Attention modules used in a DL network. Left: Scaled Dot-Product Attention; Right: Multi-Head Attention module.

c) **Inception module:** The utilisation of "Inception modules," which are miniature convolutional networks with several filter sizes (1x1, 3x3, 5x5) running in parallel, is the core innovation of the Inception network. The network can collect both local and global characteristics at different scales by mixing filters of different sizes. Furthermore, the Inception network employs other strategies to reduce computational complexity, such as dimensionality reduction via 1x1 convolutions and the inclusion of auxiliary classifiers at intermediary layers to address the vanishing gradient problem and the high computational cost of deeper networks during training. Overall, the Inception architecture achieves a good mix of accuracy and processing efficiency, making it appropriate for a wide variety of computer vision tasks. Since its inception, several versions of the Inception network have been constructed, each one improving on the preceding one in some way.

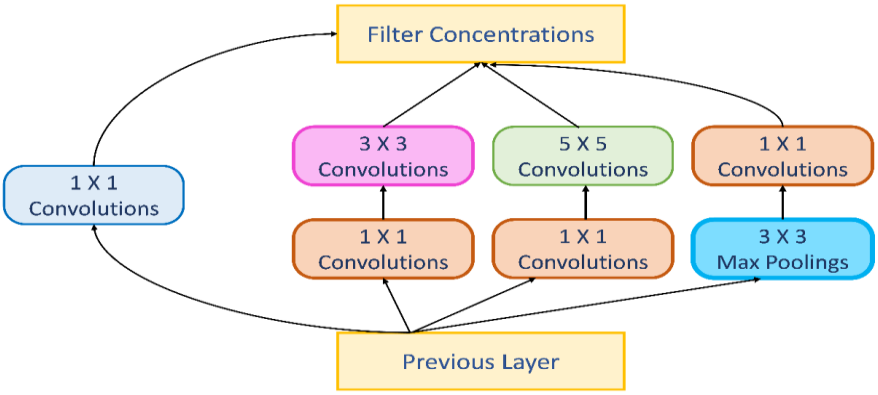

Figure 2.3: A standard architecture of inception module.

2.7. Review of previous works in MRS

'Decomposing a signal by transformation, applying a threshold on the decomposed coefficients to remove unwanted points, and reconstructing the signal with remaining coefficients' are the standard steps for signal denoising since a long time. Transforming a signal into different spaces e.g., wavelets present a variety of unique feature coefficients at different levels of transformation/decomposition depending upon the signal to noise contributions. Traditional denoising methods have used either time domain or frequency domain-based analysis. Fourier transform (FT) and its variants have been used as global transformation functions for signal analysis, but are regarded inadequate in many cases related to singularities or non-stationary signals. The short time-frequency transform (STFT) method, incorporating the short-term window function, also does not meet the requirements. The introduction of wavelet transform (WT) as an alternative has opened a new paradigm in the field of signal processing. The seminal works of Daubechies [4] in filter design and that of Mallat [5] in the development of efficient algorithms for discrete WTs created a platform for the current advances in wavelet-based signal processing [6]. Several thresholding criteria has been designed to utilise this information to achieve better denoising outcomes. Thresholding a wavelet coefficient for denoising was first proposed by Donohue, which used a fixed noise threshold for the decomposition levels, also

termed as the universal threshold [19]. Stein's Unbiased Risk Estimate (SURE) threshold, Minimax threshold [21] are some widely used and better performing fixed threshold estimates. Based on wavelet shrinkage [17], [22–27]; wavelet coefficient modelling [28], and modulus maxima [29]; many denoising methods have been developed with performances better than the conventional filtering operation [30]. Further, considering different properties of wavelet decomposition coefficients, different approaches of thresholding criteria have been proposed [31–35]. The idea of adaptive thresholding [22], [36] has evolved from the limitations of fixed thresholding methods like a) same noise threshold for each coefficient of a decomposition level without considering its bias and b) inability to discriminate between the coefficients when magnitudes of signal and noise coefficients are close. The above-mentioned issues are practically visible when analysing a real experimental signal. Srivastava et al. [37] proposed a denoising method for adaptive noise threshold selection depending on the decomposition levels selected with application on cw-ESR spectra and presented promising results.

An MR spectrum acquired in a clinical setting at low TE has a significant contribution of macromolecular compounds (MMs) spread gradually along the metabolite peak spectrum as baseline. Since, in recent studies, it has been confirmed that MMs are important to diagnostics and related to aging, characterization of MMs also requires equal attention as metabolite peaks. various prospective as well as retrospective methods have been proposed in the past studies. Among prospective methods, (i) taking long-TE time acquisitions to suppress the contribution of MM in spectra [6, 7], and (ii) method based on T_1 and T_2 differences by using inversion recovery (IR) methods, either single IR or double IR, to obtain metabolite-nulled or MM-nulled spectra [8-10] have been prominent. Both the approaches provide good suppression but has various limitations in terms of peak information loss, SNR reduction, weighting issues (T_2-weighting for long TE, and T_1-weighting for IR-methods) [1]. Among retrospective methods, different time- and frequency-domain fitting-based approaches have been used. Either using

metabolite-nulled MM spectra for parameterization or mathematical modelling of MM spectra using a set of gaussian, Lorentzian or Voigt model functions has been the two main approaches [15-21]. Hankel-Lanczos singular value decomposition (HLSVD) based methods, followed by Advanced Method for Accurate, Robust and Efficient Spectral fitting (AMARES) were among the first methods used MRS domain [11-14, 17]. For MM parametrization, HLSVD doesn't take prior knowledge of MM components, whereas AMARES is highly user intensive in terms of creating a knowledge base of peaks. In recent years, machine learning and deep learning-based approaches for metabolite fitting and spectral mining have been used, but in almost every study, MMs were discarded along with other baseline artifacts and noise.

Quantitation can be approached in many ways, of which, the curve-fitting approaches to map individual components has been an established strategy. Different parametric and non-parametric model approaches have been developed to perform analysis and obtain spectral parameters from data, either in time-domain or frequency domain [10-18]. Similarly, machine/deep learning approaches have been implemented in a variety of tasks [19-20], and in recent years, MRS domain. Das et al. [21], Hatami et at. [22] proposed a Random Forest (RF) regression model and CNN network respectively for metabolite concentration estimation. Gurbani et at. [23] proposed an unsupervised convolutional encoder-model decoder approach to accelerate spectral fitting of whole brain MRSI spectra. Kyathanahally et al. [24] proposed a convolutional neural network (CNN) for ghosting artifact detection and removal. Gurbani et al. [25] developed CNN architectures for overall quality assessment of MR spectra based on artifacts present and spectral fitting. Lee et al. [26] designed a CNN network for spectral fitting and metabolite peak estimation. For correcting frequency and phase of acquired MR spectra, Ma et al. [27] came up for a convolutional neural network-based approach. Similarly, to address the issues of noise degradation of MR spectra, Lei et al. [28] used a stacked auto-encoder (SAE)

model. Hu et al. [29] and Dandil et al. [30] proposed LSTM based designs to address the issues of noise, and grading of brain tumors, respectively.

Table 2.1: A review of previous ML/DL based application in MRS analysis

Applications (Dataset)	Network architecture	Important issues	Ref.
Estimation of parameters of a model-based analysis of MRS data for metabolite quantification. Synthetic: 1 million* In-vivo: 287	Random Forest Regression approach	-Training model accounted for spectral features such as MM baseline, LW, SNR variation with metabolite conc.	[1] Das et al. (2018)
Estimation of spectral parameters addressing MRS quantification problem Synthetic: 5×10^5 *	CNN	- CReLU has been used. - more realistic data generation including phase variation due to ECC, residual water, or non-ideal lineshapes is expected to examine the robustness of quantification with CNN.	[2] Hatami et al. (2018)
Rapid spectral fitting of whole brain data (MRSI) In-vivo: 102,005 from volumetric EPSI scans of 4 healthy and 6 newly diagnosed GB patients	Convolutional encoder-model decoder (CEMD) – unsupervised learning-based task	-Autoencoder in ML correlates to curve-fitting. - Encoders reduces data into lower dimensional space which may not be readily interpretable (in place, wavelet can be used).	[3] Gurbani et al. (2019)
Detection and removal of ghosting artifacts in MR Spectroscopy Synthetic: two sets of 30000 spectra In-vivo: 13 patients	Different DL approaches-FCNN, CNN, SWWAE-residual network	- data converted to time-frequency spectrogram before processing - different CNNs were tested to enhance accuracy	[4] Kyathanahally et al. (2018)
Identification and filtering of artifacts from MRSI spectra	CNN	- Bayesian optimization technique used to tune network architecture	[5] Gurbani et al. (2018)

Intact metabolite spectrum mining for robust quantification using CNN model Synthetic: 50000 In-vivo: 8 healthy patients	CNN	Gaussian model functions used for detailed MM-baseline simulation. - residual water and other artifacts have not been considered in the data simulation. - generic CNN model used. Better performing architecture is likely.	[6] Lee et al. (2019)
Applications (Dataset)	**Network architecture**	**Important issues**	**Ref.**
MRS frequency and phase correction using CNN model Synthetic: 41000 MEGA_PRESS transients In-vivo: 33	CNN	- separate model path for frequency and phase correction. - dataset specific to the MEGA-PRESS spectra	[7] Ma et al. (2021)
Denoising MRS spectra using Deep learning model Synthetic: 192 In-vivo 5 patients	Stacked auto-encoder (patch- based training)	- represent noisy data in a sparse feature vector using SAE without apriori information.	[8] Lei et al. (2021)
Denoising repeatedly sampled in-vivo SVS data Synthetic: 50 virtual persons In-vivo: 8 healthy, 1 Parkinson's patient	Causal-LSTM		[9] Hu et al. (2021)
Automatic grading of brain tumors using LSTM NN on MRS signals.	LSTM	-	[10] Dandil et. al. (2020)
Target metabolite isolation and big data driven measurement uncertainty estimation Synthetic: 100000 (rat brain)	CNN	- assessment of measurement uncertainty is heuristic. For general application, a theory-oriented, formal approach is required. - approach is computation cost intensive, and demand large data storage.	[11] Lee et al. (2020)

Wavelet scattering CN for quantification Synthetic: 10000	Wavelet scattering CN	- features extracted using scattering transform and fed into neural network for quantification. - amplitude of simulated metabolites did not imitate in-vivo metabolite concentration.	[12] Shamaei et al. (2021)

* ISMRM MRS Fitting Challenge 2016 dataset