

RESEARCH ARTICLE

Using Sequence-to-Sequence Models for Carrier Frequency Offset Estimation of Short Messages and Chaotic Maps

CHRISTOPHER P. DAVEY¹, (Graduate Student Member, IEEE),
ISMAIL SHAKEEL^{1,2}, (Senior Member, IEEE), RAVINESH C. DEO¹, (Senior Member, IEEE),
SANCHO SALCEDO-SANZ³, AND JEFFREY SOAR⁴

¹School of Mathematics, Physics and Computing, University of Southern Queensland (USQ), Springfield, QLD 4300, Australia

²Information Sciences Division, Defence Science and Technology Group (DSTG), Contested Communications Branch, Canberra, SA 5111, Australia

³Department of Signal Processing and Communications, Universidad de Alcalá, 28805 Alcalá de Henares, Spain

⁴School of Business, University of Southern Queensland, Springfield, QLD 4300, Australia

Corresponding authors: Christopher P. Davey (christopher.davey@usq.edu.au) and Ravinesh C. Deo (ravinesh.deo@usq.edu.au)

This work was supported in part by the Department of Defence of the Commonwealth of Australia under Defence Science Partnership Scholarship under Agreement 10254.

ABSTRACT Deep Learning methods have produced good carrier frequency offset estimations for short message sequences in comparison with methods based on the Fast Fourier Transform. However, these performance gains were observed for short ranges of frequency offsets, sequences with predefined pilot symbols and periodic modulation schemes. Chaotic modulation has an advantage over periodic signals in offering security through the continuous changes produced by parameterising the chaotic map function. However, synchronisation of chaotic map parameters in coherent receivers is dependent on the carrier recovery of phase and frequency which dramatically reduces the demodulation performance under high noise levels. This article presents a stacked sequence-to-sequence neural network architecture for blind carrier frequency offset estimation of both periodic and chaotic modulation schemes. The results obtained demonstrate better performance than conventional methods in low SNR for the Additive White Gaussian Noise channel. While this technique operates without feature engineering, the results demonstrate that data augmentation produces a higher degree of accuracy for such models, indicating the benefit of integration with conventional signal pre-processing steps as part of the deep learning pipeline. The proposed neural network architecture is shown to perform carrier frequency offset estimation, not only for the selected periodic modulations, but also in the case of highly non-linear chaotic maps. This suggests the applicability of deep learning methods for synchronisation in waveforms that employ chaotic modulation schemes for secure communication and for applications where short and sporadic messaging are required (e.g., Internet of Things).

INDEX TERMS Chaotic communication, deep learning, fast fourier transforms, frequency synchronisation, carrier frequency offset estimation.

I. INTRODUCTION

The accuracy of Carrier Frequency Offset (CFO) estimation methods based on the Fast Fourier Transform (FFT) in single

The associate editor coordinating the review of this manuscript and approving it for publication was Juan Wang¹.

carrier communications is dependent on the sample length of the message, and on the Signal to Noise Ratio (SNR) [1]. Short sample message lengths are advantageous in low power Internet of Things (IoT) applications and pilot signals used for signal detection and synchronisation. Deep Learning (DL) methods have demonstrated to outperform FFT-based

methods under similar constraints [2], [3]. However, much of the experimentation to date has focused largely on phase amplitude modulation (PAM) or M -ary phase shift keying (M -PSK) modulations, and has not investigated the potential application to chaotic modulation techniques.

Chaotic modulations present a method for providing physical layer security, and are well suited to address the constraints placed on IoT applications [4]. Due to the continuously changing signal which results from parameterisation of the chaotic map sequence, chaotic modulations exhibit high autocorrelation for the same symbol and low cross-correlation between symbols [5]. This characteristic is advantageous for coherent detection, where each symbol is correlated with a potential mapping function at the receiver and is resilient to small levels of noise [5]. However to achieve demodulation the receiver is required to estimate the parameters for each chaotic map function which is known as sequence synchronisation [6]. Sequence synchronisation for chaotic maps is dependent on accurate estimation and removal of the CFO [6], [7]. For estimating frequency offsets in chaotic maps, autocorrelation methods are shown to be effective for fixed preambles [8], however these methods are difficult to implement for variable and non-repetitive sequences.

Given that deep neural networks can learn non-linear features, the estimation of CFO for randomised chaotic sequences is an application well suited to such methods. In this article we propose a data driven method for the estimation of the CFO in short sequences of BPSK, QPSK modulations, as well as for the Circular, Quadratic and Zadoff-Chu chaotic maps. The approach is applied to both fixed preamble and randomised sequences. The model performs an iterative estimation of the frequency offset using a sequence-to-sequence (Seq2Seq) block at each level. This approach is capable of more accurate CFO estimation for the M -PSK modulations in comparison with the FFT and Phase Locked Loop (PLL) approach. While brute force cross-correlation is more accurate without down-sampling at the matched filter (at the expense of execution time), the DL method is more accurate when compared with cross-correlation on the shorter down-sampled signal. The network can produce CFO estimates directly from the In-phase and Quadrature (IQ) values of the received signal, however data augmentation is shown to provide an advantage for the accuracy of the estimation.

A. BACKGROUND AND RELATED WORK

The use of the FFT is demonstrated to perform an approximation for the maximum-likelihood function of the parameters in a sinusoidal signal corrupted by Gaussian noise in [9]. The length of the FFT determines the accuracy of the measurement, and was found to be optimal at up to 4 times the length of the signal [9]. As the frequency step size of the FFT produces a coarse estimation, an interpolation is required to produce a finer estimate. In the case of [9] an iterative secant method is applied to the fine estimate of the frequency but

is indicated to produce a larger error in low SNR [9]. The threshold for the variance of the estimator in [9] is shown to be optimal above an SNR between 15 dB and 17 dB in [10] for corresponding sequence lengths between $N = 64$ to $N = 2048$.

Interpolation methods using points either side of the maximum value for the FFT are applied to calculate an adjustment term for the frequency estimate in [11] and [12] and improve on the method in [9]. These methods are shown to have a bias for short sequences and low SNR in [10] which proposes three and five point interpolation methods making use of the phase information in the FFT coefficients. Several methods of interpolation are compared in [13] which also makes use of three coefficients to demonstrate a method that approaches uniform error variance above 2 dB. An extended number of fourier coefficients weighted by an approximation of their mean square error are combined to estimate the frequency offset in [14], resulting in an estimator approaching the lower bound of variance close to 5 dB. However each of these methods share limitations in lower SNR and for short sequences. In addition the application of the FFT is applicable for periodic signals and are not appropriate for use with those chaotic modulations which do not exhibit distinctive peaks within the power spectrum.

DL approaches, in particular convolutional neural networks (CNN), are demonstrated to outperform FFT based methods on estimation of CFO for short random sequences in 1-bit ADC's at low SNR in [2]. The selection of DL models is able to extrapolate well over a wider range of SNR (between -20 and 40 dB), even though they are trained on a subset of the SNR (between 0 and 10 dB) [2]. The 1-bit quantization method reduces the amount of information available to the network for training [2] and for conventional methods it is known to require up to four times oversampling for the estimation of offset parameters [15]. In conventional methods, knowledge of modulation order M is applied to remove the modulation from the signal prior to the application of FFT estimation, however the generality of the 1-bit ADC in [2] did not motivate an exploration of the impact of the modulation on CFO estimation. As our method is applied after down-sampling at the matched filter output, the type of modulation is shown to have an influence on estimation accuracy for both FFT and DL approaches.

Further indication that DL can provide good frequency offset estimation for sinusoidal waveforms in low SNR is described in [3]. The network architecture was constrained specifically to the fully connected network (FCN) with the number of input nodes representing the length of the signal to be processed and being dependent on the range of the frequency offset, requiring larger dimensions for wider ranges of frequency [3]. FCN networks require a larger number of connections between layers as opposed to the CNN [16], hence consideration of CNN layers would provide flexibility for processing multiple signal lengths with a constant number of layer parameters. Although the choice of network architecture limited the range of frequency offset, it was shown

that the FFT and DL methods did decrease in accuracy under shorter signal lengths [3]. To address a wider frequency offset range, as well as several modulations, this article proposes the stacked network architecture, which incorporates CNN layers to extract features at each level rather than fully connected layers.

Short signals prevent the FFT from accurate spectral estimation due to the resulting coarse resolution, whereas a DL method for super-resolution estimation of the approximate spectrogram is proposed in [17]. A combination of both FCN (linear) and CNN layers are applied in the architecture, taking advantage of the ability of the CNN to accept multiple resolutions of input during training to learn translation invariant features [17]. A customised minimum distance loss is applied during the learning procedure and the model is shown to produce more accurate estimation than the periodogram and eigenvector (MUSIC) based estimators at a limited range of SNR [17]. The model is trained and tested on the complex sinusoid with amplitudes, frequency and phase selected from random normal distribution at different parameters [17]. A fixed output resolution is used to estimate the pseudo-spectrum of the signal which is then mapped onto a known frequency range [17], the resolution is dependent on the signal length and is fixed. Our proposed stacked model refines the peak frequency estimate at increasing resolutions for each stack in the network and estimates an error correction term to produce a high resolution estimate for the carrier frequency offset at the final layer.

The CNN is leveraged in the literature on the CFO estimation task, however as the signal varies over time, a recurrent neural network (RNN) may be applied to learn time dependent features over the signal. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) network models are trained to perform CFO estimation with the short training field (STF) of the IEEE 802.11ah preamble frame in [18]. Results demonstrate that the network performs well on the CFO estimation task in comparison with the conventional correlation method in low SNR [18]. The STF is a fixed pattern within the frame and is useful in simplifying the process of timing and CFO estimation [18]. It is designed to improve the resulting accuracy of the estimation method. In the proposed method, we experiment with both the fixed preamble as well as randomised sequences for several modulations and demonstrate that the DL approach can learn to estimate the CFO even where the modulation exhibits chaotic behaviour. In the proposed architecture, recurrent LSTM layers learn time dependencies resulting from features modelled by CNN layers and are organised in encoder-decoder blocks which share the hidden state for learnt time dependencies between them.

A common element in the cited literature is that the DL method is more accurate than conventional methods in low SNR and for short sequences. While the FCN layer is applied in [3] due to the constraints of the experiment, the CNN has advantages as an effective choice for feature extraction in the CFO estimation task [2], [17] and the use of the LSTM is

shown to be effective in [18]. It is clear a DL model can be constructed for a single modulation, the impact of estimating CFO for multiple modulations has not been investigated for such an approach. Spectral methods are optimal under the right conditions and would be useful to incorporate into the design of the network model as demonstrated in [17]. The chaotic map becomes deterministic when the state parameters are known. A recurrent network modelling approach may demonstrate the ability to learn implicit information from the signal, thereby aiding estimation of the CFO. A combination of RNN and CNN would enable a DL model to both extract translation invariant features as well as learn time dependent features. This article proposes a stacked architecture which estimates the probability of the peak frequency as well as an error correction term using sequence-to-sequence blocks comprised of CNN and LSTM units.

The rest of the paper has been structured in the following way: The next section describes the system model, as well as the conventional carrier offset estimation method. It also explains the proposed model architecture, as well as the data augmentation applied when training the model. Section III shows the experimental results obtained when the proposed DL approach is applied to a number of CFO estimation tasks. A discussion on these results is also provided in this section. Section IV closes the paper, by giving some final concluding remarks on the research carried out.

II. METHODS

When transmitted over a channel, the baseband signal $s(t)$ is subject to perturbations of timing t , phase θ and carrier frequency f_0 offsets, shown in Equation (1), where $a(t)$ represents the signal modulation after filtering, and $n(t)$ represents Additive White Gaussian Noise (AWGN).

$$s(t) = a(t)e^{j\theta} e^{j2\pi f_0 t} + n(t) \quad (1)$$

In this work the proposed model is trained on several modulations, which include Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK), as well as chaotic Circular, Quadratic and Zadoff-Chu maps. Frequency offsets for M -PSK modulations are estimated in two stages: first, a coarse estimate \hat{f}_1 is given by the position of maximum frequency of the coarse grained FFT (Equations (2)-(4)). The derivation for the use of the Discrete Fourier Transform (applied through the FFT) as an approximation for the maximum-likelihood estimator of the CFO is described in Rife and Boorstyn [9], in this article we apply the Matlab coarse frequency estimator [19] which is derived from the use of the FFT in [20]. The received signal $s(t)$ is first raised to the M th power $z(t) = s(t)^M$, then the FFT is calculated giving $S(k)$ (Equation (2)). The index k_m of the frequency, having the maximum absolute value for $S(k)$ (Equation (3)) is then divided by the modulation order M ($M = 2$ in BPSK and $M = 4$ in QPSK) and is scaled by the sampling frequency f_s over the length of the FFT N (Equation (4)). After the coarse estimate, a fine frequency adjustment \hat{f}_2 is estimated via a PLL implemented by the Matlab carrier synchronisation

function [21] derived in [22]. The difference in phase error estimates $\Delta\theta$ produced by the PLL are scaled to the frequency estimate via the sampling rate f_s and the down-sampling rate d , and the operation is averaged to estimate the adjustment for the frequency offset (Equation (5)). Finally, the frequency offset is estimated as the sum of the coarse frequency estimate and the fine frequency adjustment (Equation (6)). Improvement in accuracy can be gained by increasing the resolution of the FFT, results from [9] recommend a resolution up to four times the length of the original signal, depending on performance constraints. In our experiments the FFT resolution is set to $4\times$ the down-sampled received signal of 104 samples.

Two FFT interpolation methods are employed for comparison. Both methods adjust the index k_m through an estimate of the difference to the peak of the FFT, $\hat{\delta}$ and add it to the index as in Equation (10), the updated index, k_{adj} is then applied in estimating the frequency f_1 (replacing k_m with the adjusted index k_{adj}). The first interpolation method is described in [13] where the two values either side of the maximum index are used to estimate the difference from the peak of the FFT (Equation (7)), this method reduces the bias of the quadratic interpolation. The second method is proposed in [14] which incorporates all FFT coefficients (in $K < N/2 - 1$) and calculates an estimate for the adjustment $\hat{\delta}_k$ at each coefficient index k (Equation (8)). These estimates are aggregated through weighting each with an approximate of their mean square error term (Equation (9)) [14]. In the results section the first interpolation method is indicated on plots as 'Jacobsen' and the second 'Candan'. Both methods are suitable for use in multiple iterations, however in our comparison we generate results with only one application of each method.

$$S(k) = \sum_{n=0}^{N-1} z_i e^{j2\pi kn/N} \quad (2)$$

$$k_m = \operatorname{argmax} |S(k)| \quad (3)$$

$$\hat{f}_1 = \frac{f_s k_m}{N M} \quad (4)$$

$$\hat{f}_2 = \frac{1}{N} \sum_{i=1}^N \frac{f_s \Delta\theta}{d} \frac{\Delta\theta}{2\pi} \quad (5)$$

$$\hat{f}_0 = \hat{f}_1 + \hat{f}_2 \quad (6)$$

$$\hat{\delta} = -\operatorname{Re} \left[\frac{S_{k_m+1} - S_{k_m-1}}{2S_{k_m} - S_{k_m-1} - S_{k_m+1}} \right] \times \operatorname{Re} \quad (7)$$

$$\hat{\delta}_k = \frac{N}{\pi} \tan^{-1} \left(\tan\left(\frac{\pi k}{N}\right) \times \left\{ \frac{S_{k_m+k} e^{-j(\pi/N)k} - S_{k_m-k} e^{j(\pi/N)k}}{S_{k_m+k} e^{-j(\pi/N)k} + S_{k_m-k} e^{j(\pi/N)k} - \frac{2S_{k_m}}{\cos(k\pi/N)}} \right\} \right) \quad (8)$$

$$\hat{\delta} = \frac{\sum_{k=1}^K 1/\sin^2(\pi k/N) \hat{\delta}_k}{\sum_{k=1}^K 1/\sin^2(\pi k/N)}, K < N/2 - 1 \quad (9)$$

$$k_{adj} = k_m + \hat{\delta} \quad (10)$$

The cross-correlation method is applicable where a template such as a pilot signal is known. The template signal is rotated by frequency steps f_1, f_2, \dots, f_n between the range of the expected frequency offset (in our experiments ± 5 kHz). The complex cross-correlation between the received signal and the distorted template is calculated and the maximum cross-correlation is used to determine the index of the frequency estimate. In our randomised experiments, the DL model does not have any knowledge of the template used for the comparative method, whereas in the fixed preamble setting it is trained on a fixed sequence. Cross correlation is performed prior to down-sampling at $4\times$ sample length and post down-sampling at $2\times$ sample length for comparison. This method is computationally expensive and is most accurate on small frequency ranges and longer signal lengths.

A. DATA GENERATION

The data used in training and evaluation are divided into two experimental settings, the fixed preamble setting and the randomised sequence setting. In the fixed preamble setting, M -PSK sequences are generated by repeating a fixed message containing the 13 bit Barker code. For the chaotic maps, the initial conditions are predefined along with a fixed length for the recurrence relation within the map. Randomised sequences consist of random bits for the M -PSK messages and sliding windows of chaotic maps. Both types of sequences (fixed and random) are constructed where the bit sequence length is dependent on the number of bits per symbol and produce 2 samples per symbol resulting from matched filtering (up-sampled at $8\times$ and decimated at $4\times$ per sample respectively). After applying a root raised cosine matched filter at the transmitter and receiver, a 52-bit sequence for BPSK and 104-bit sequence for QPSK generate 104 samples. In the chaotic modulations 52 symbols are mapped to a resulting 104 symbols after matched filtering. All sequences are 104 samples in length.

Chaotic sequences cannot be randomised in the same manner as bit sequences, since they depend upon the initial conditions for each symbol and are parameterised depending on the mapping function. Given their reliance on successive feedback, a randomised chaotic sequence is generated by randomly selecting the number of feedback iterations from an initial condition and stepping the mapping function over the sequence length while storing the feedback signal to use as the initial conditions for the next sequence. The mapping functions for each of the chaotic maps are shown in Table 1, along with the feedback parameter and initial condition parameters. Figure 1 illustrates the IQ values for each of the corresponding map functions.

During the data generation process, no phase rotation is applied, and the frequency offset is selected from a random uniform distribution within the range ± 5 kHz with a sampling frequency $f_s = 1$ MHz. Noise is added for SNR, $E_s/N_0 = 0 \dots 9$ dB with the noise variance σ^2 being estimated from parameters E_s and N_0 in Equations (11)-(13), where E_s is the energy per channel symbol, N_0 the noise

TABLE 1. The set of chaotic map functions and their initial parameters used in generating sliding sequences.

Map Function	Parameters	Initial Value
Circular Map $x_{i+1} = x_i + b + \frac{a}{2\pi} \sin(2\pi x_i)$	$0 \leq a, b \leq 1$ $a = 0.5$ $b = 0.2$	$x_0 = 0.7$
Quadratic Map $x_{i+1} = b - ax_i^2$	$a = 4$ $b = 0.5$	$x_0 = 0.15$
Zadoff-Chu Map $x_i = \exp\left(-j \frac{\pi u i(i+1)}{N}\right)$	$0 < u < N$ $\gcd(N, u) = 1$	$N = 63$ $u = 1$ $x_0 = 0.7$

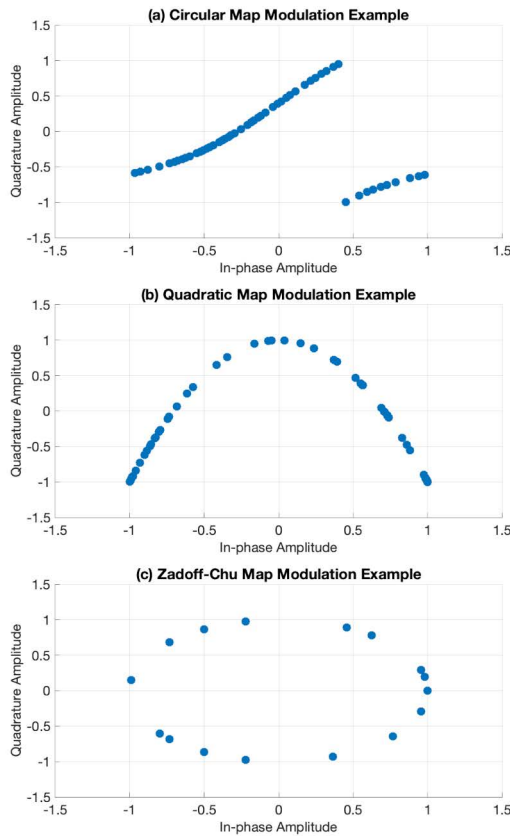


FIGURE 1. Example IQ plots of the chaotic map functions for a) Circular, b) Quadratic and c) Zadoff-Chu maps.

power spectral density, L the number of symbols, and n the bits per symbol. For training the network an offline dataset of 102400 sequences is generated for each modulation (502400 sequences) and each sequence is labelled with the corresponding signal random frequency offset that was applied to distort the signal.

$$E_s = \frac{\sum_{t=1}^L s(t)}{L/n} \tag{11}$$

$$N_0 = \frac{E_s}{E_s/N_0} \tag{12}$$

$$\sigma^2 = N_0/2 \tag{13}$$

B. NETWORK ARCHITECTURE

The intuition applied to the design of the network architecture was that the network should be capable of multiple stages of refinement in the task of frequency offset estimation. A stacked architecture was arrived at such that each stack would successively estimate a discrete set of steps for the frequency range where the step size decreases at each level in the stack. The final level then estimates the error between the coarse estimate of the previous layer and the target frequency. For comparison, the error adjustment layer is implemented with two approaches. The first applies a classification approach that is constrained within ± 100 Hz of the coarse estimate. The second approach applies a direct regression to provide a continuous error correction to compensate for broader variation of the error between the coarse estimate and target frequency offset.

Each stack consists of a subnetwork block which is responsible for learning features and performing estimation for that block. To perform feature extraction, as well as learn recurrence relationships, a sequence-to-sequence (Seq2Seq) network is defined within the feature extraction block. The Seq2Seq architecture follows the approach first defined in [23], however beam search is not applied during estimation and the inclusion of Convolutional layers differs from the original model. The block design includes a Convolutional (CNN) layer to extract input features, a bidirectional Long Short-Term Memory (LSTM) encoder, latent space implemented as a CNN layer, a bidirectional decoder LSTM layer followed by an output CNN layer. Classification is provided by a Dense block with a soft-max activation while regression is achieved with a tanh activation. Regularisation is provided by applying Batch Normalisation [24] following each CNN and intermediate Dense layer, and Layer Normalisation is applied after each LSTM layer. Max-pooling is applied to the output of intermediate CNN layers with Global Average Pooling applied prior to the Dense layer.

Aside from the estimation output, the hidden LSTM state is shared between encoder and decoder LSTM, and the hidden state of the decoder is forwarded to the encoder in the subsequent stack. The latent CNN state is also forwarded between network stacks and concatenated with the input features for the encoder in the subsequent stack. These skip connections enable multiple forward paths fusing latent features and sharing hidden recurrent state throughout the network and enable gradient flow during back-propagation [25]. Such connections are proposed to enable ensemble like behaviours in deep networks [26]. Figure 2 presents the schematic view of the sequence-to-sequence block as well as the dense estimator blocks for the network output and the interconnection between the blocks is illustrated in Figure 2. Three stacks were defined, with frequency bins of 100 and 50 Hz for both the classifier and regressor networks. A frequency adjustment of ± 100 Hz is applied for the final estimator of the classifier network, and a single continuous parameter applied in the final estimator of the regressor network. Table 2 lists the number of units for each layer type.

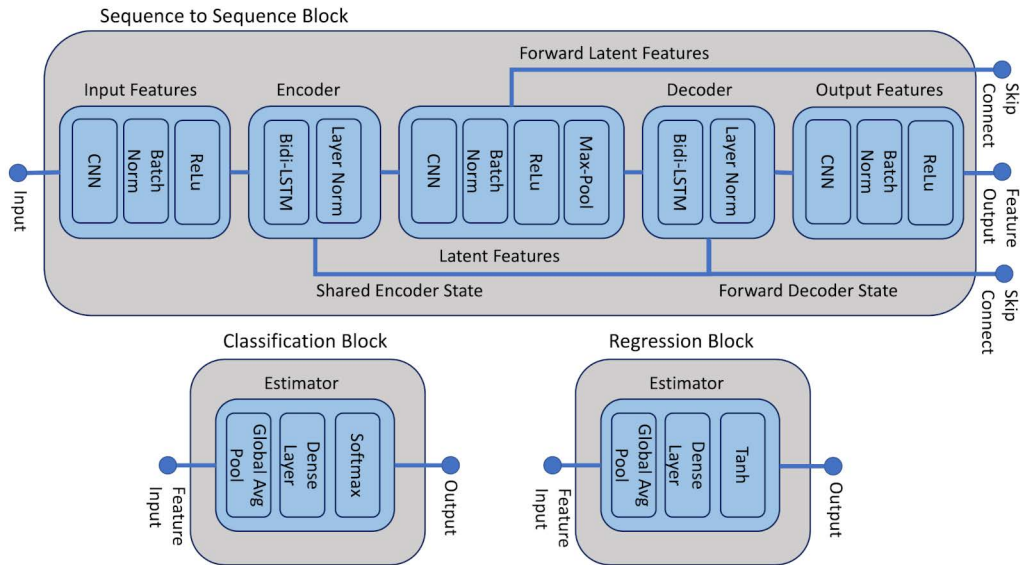


FIGURE 2. Sequence to Sequence blocks with CNN feature extraction are interconnected with paths for hidden recurrent state and latent CNN state. The final output of each block pools the output features of the sequence to sequence block to produce the estimate for either the frequency bin or the frequency error.

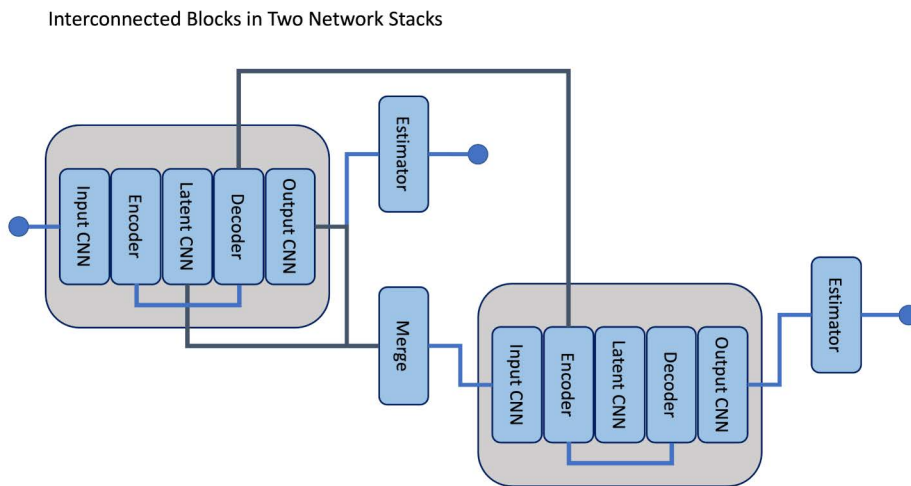


FIGURE 3. Interconnection between two sequence-to-sequence blocks shares the decoder hidden state with the encoder of the subsequent block and merges the latent CNN state with the CNN output via a concatenation.

TABLE 2. The set of chaotic map functions and their initial parameters used in generating sliding sequences.

Layer Type	Units/Filters
CNN	32
LSTM	128
Dense Classifier units (bin Hz) per stack	102 (step size 100 Hz) 202 (step size 50 Hz)
Dense Regressor	202 (step size 1 Hz) 1

During the network’s training, the data set is partitioned into 50% training, 20% validation and 30% test. A cyclical learning rate schedule [27] was applied which allowed the

learning rate to oscillate between 0.0001 and 0.001. Input data was scaled by dividing the input signal by the l^2 -norm and min-max normalising with parameters ± 1 . Target frequency is min-max normalised with parameters ± 5 kHz. Back-propagation is performed with Adam optimisation [28]. Cross-entropy loss is applied to the classification estimator and mean squared error loss is applied to the regression estimator. Each stack is trained iteratively, and the weights of each previous stack are frozen prior to training the subsequent stack. When training the final stack, the difference between the previous stack frequency estimate and the true target frequency is calculated and applied as the target after min-max normalisation (± 5 kHz).

TABLE 3. Data augmentation produced 17 features prior to the input for the network. Each of the features were derived from steps used in conventional synchronisation.

Operation	Description
$y_t = r_{t+1} - r_t$	Lag-1 Differences of IQ values of received signal r . The difference between lags are used in the fine tuning of coarse grained estimation in conventional methods.
$[r^2, r^4]$	Raised Powers of IQ values. In PAM and PSK modulations raising the signal to the order of the modulation results in a constant phase for the IQ coordinates.
$[\text{FFT}(r^2), \text{FFT}(r^4)]$	FFT for raised powers of received IQ values in r . A coarse grained FFT of the down-sampled signal is used to produce coarse estimates for the frequency.
$\hat{R}(k) = \frac{1}{N-k} \sum_{m=k}^{N-1} r(m)r^*(m-k)$	Autocorrelation of received signal r . The autocorrelation of the signal is used in conventional phase and frequency correction procedures.
$p = \begin{bmatrix} r , \angle r \\ p^2, p^4 \end{bmatrix}$	Raised Polar form p of received signal r . The raised polar form is used in conventional phase estimation algorithms.
$y'_t = r_{t+1}^* r_t$	Lag-1 Conjugate for received values in r . Conventional frequency estimation may employ the conjugate between two samples of a given interval (in this case lag 1).

The network models are trained under two experimental settings, fixed preambles and randomised sequences, with each setting producing separate models (eight individual models in total, four model variants in each setting). A third experiment explores the difference in training on a single modulation, as opposed to multiple modulations. In this task, two variants of the network model are independently trained on QPSK and Quadratic map modulations for each setting, resulting in eight individual models.

C. DATA AUGMENTATION

A comparison is made between models trained with and without data augmentation. For those networks that are trained without data augmentation, the complex signal is represented as a matrix with two columns for the in-phase and quadrature components. Those networks trained with data augmentation were supplied with 17 features derived from the treatment of the complex signal in conventional synchronisation algorithms, these are described in Table 3.

During evaluation, a separate feature importance analysis is undertaken by iteratively assigning uniform noise to each feature and calculating the difference in performance between the baseline model and the noisy input data.

III. RESULTS

The Mean Absolute Error (MAE), in Hz, produced by the Stacked Model and the FFT/PLL method for the CFO

TABLE 4. Comparison between STACKNet_C and STACKNet_R on fixed preamble sequences indicates a minor difference between model variants when data augmentation is applied. A slight improvement in MAE Hz does result from the regression model in comparison to the classification model.

Comparison	Mean Difference in MAE Hz
$\text{MAE}_{\text{STACKNet}_C 17F} - \text{MAE}_{\text{STACKNet}_C}$	-0.45
$\text{MAE}_{\text{STACKNet}_R 17F} - \text{MAE}_{\text{STACKNet}_R}$	0.99
$\text{MAE}_{\text{STACKNet}_C} - \text{MAE}_{\text{STACKNet}_R}$	-4.49
$\text{MAE}_{\text{STACKNet}_C 17F} - \text{MAE}_{\text{STACKNet}_R 17F}$	-5.94

estimation task is shown in Figure 4 for BPSK and QPSK modulations between 0 and 15 dB SNR in both experimental settings. Accuracy differs on each modulation for both the proposed and conventional methods, with the proposed method achieving higher accuracy on short sequences at 104 samples than the FFT/PLL method with 4x FFT resolution. Similarly the MAE, in Hz, for each chaotic map sequence is shown in Figure 5, where the panels on the left hand side show the proposed stacked network results for estimation using 104 samples and those on the right showing the effect of sample length on the brute force correlation method at 2x and 4x sample lengths (208 and 416 samples). The stacked network is more accurate than the cross-correlation with 2x upsampling, however the cross-correlation at 4x upsampling demonstrates much higher accuracy at the expense of execution timing. Like the BPSK and QPSK modulations, the kind of chaotic map influences the accuracy of the estimate.

Comparison is made between two configurations of the network architecture where error adjustment is implemented with either a classification layer (STACKNet_C) or as a regression layer (STACKNet_R). In addition, models are trained with and without data augmentation as indicated by the postfix 17F. In the fixed preamble setting, there is little difference between models that are trained with and without data augmentation, while the regression model achieves a lower MAE Hz on average than the classification model, indicated in Table 4. On randomised sequences, those models trained with data augmentation demonstrate slightly lower MAE (Hz) on most modulations and SNR. While the performance of the augmented classification and regression models (STACKNet_C17F and STACKNet_R17F) are similar, the regression model does appear to perform better on most modulations for randomised sequences, especially on QPSK and Quadratic modulations which exhibit higher MAE (Hz) for all models. Table 5 shows the mean improvement in MAE (Hz) between those models in the random setting.

In a separate experiment, the model architecture with data augmentation is trained on single modulations for QPSK and Quadratic maps. Figure 6 indicates a lower MAE Hz for the regression model with the exception of random QPSK where performance between the two variants are close. Figures 4 and 5, indicate that training on a single modulation produces results similar to training on multiple modulations and that performance is dependent on the type of modulation.

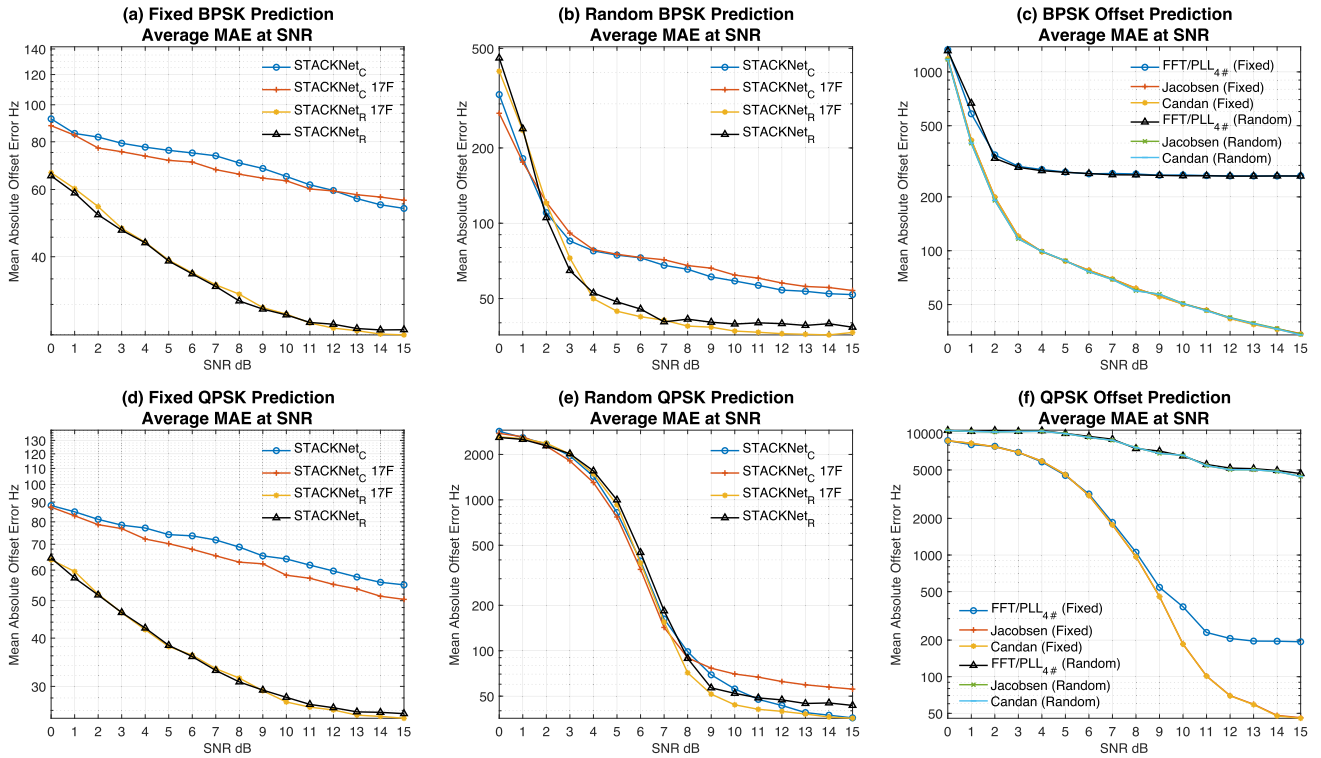


FIGURE 4. Comparison between stacked model configurations for classification (STACKNet_C) and regression (STACKNet_R) demonstrates better performance for CFO estimation on short BPSK and QPSK sequence lengths of 104 samples versus FFT/PLL methods.

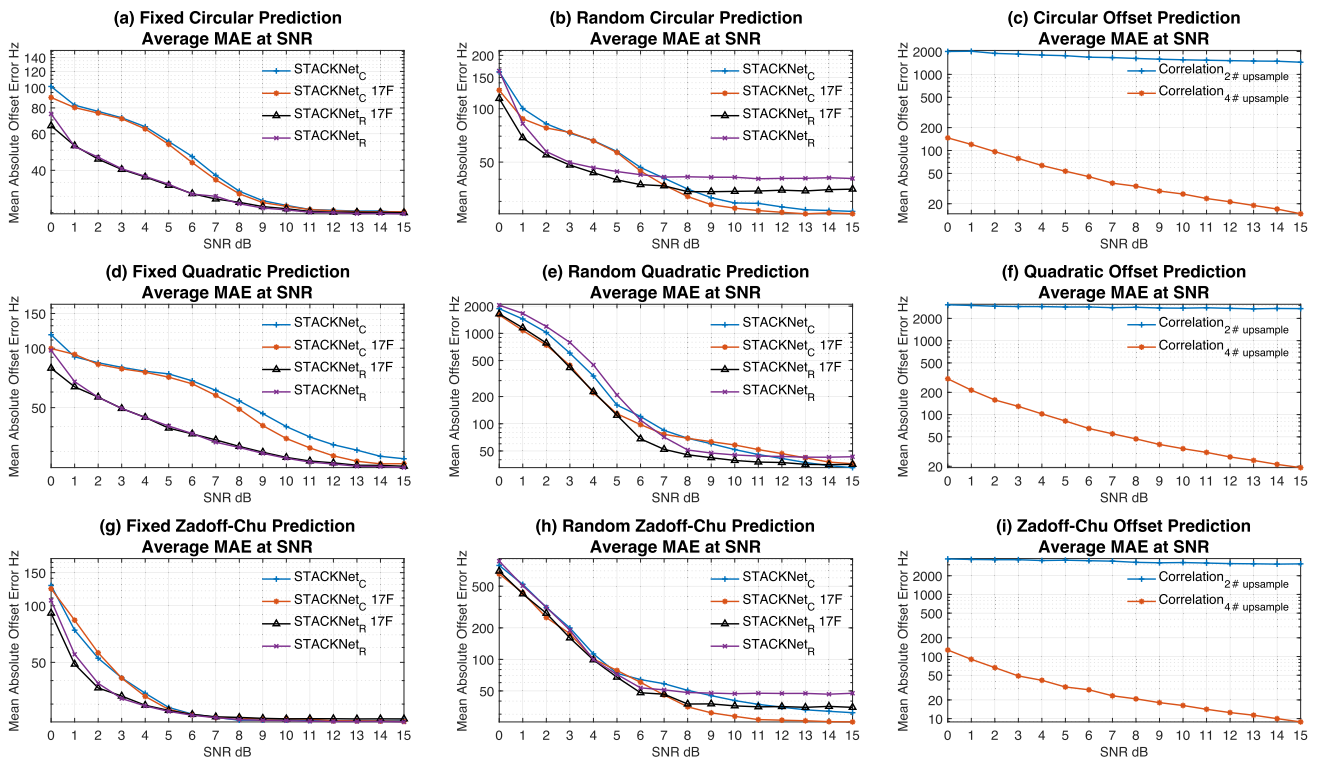


FIGURE 5. The stacked model demonstrates higher accuracy on the chaotic map than the cross-correlation with 2 × upsampling however does not perform as well as the 8 × upsampled cross-correlation.

Figure 7 displays a box plot for the execution timing of each method. The network is more complex than the

conventional FFT/PLL, and this is reflected in the timings, hence the trade-off between accuracy and complexity. It is

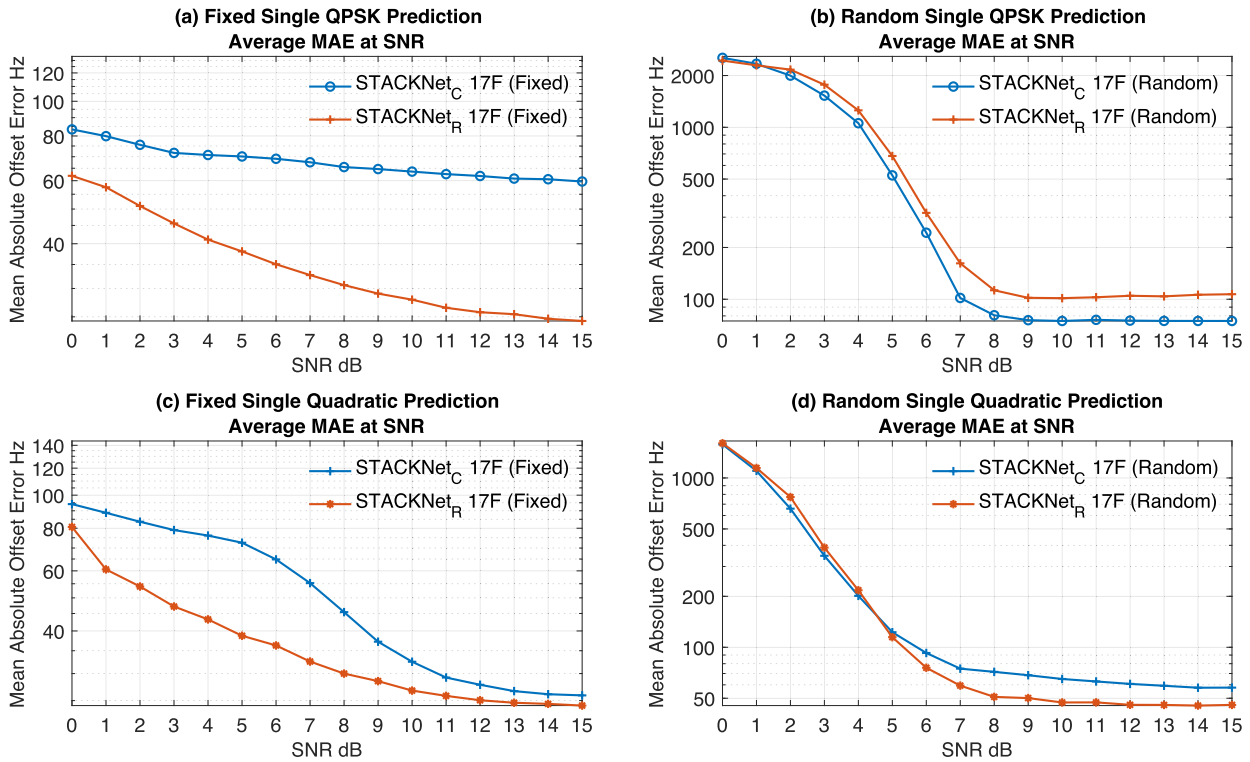


FIGURE 6. Models trained on a single modulation exhibit similar MAE to those trained on multiple modulations, indicating that training on multiple modulations does not appear to influence the performance of the model as much as the choice of modulation itself.

TABLE 5. Comparison between STACKNet_C and STACKNet_R on random sequences indicates an improvement in MAE Hz when data augmentation is applied and a small improvement in MAE Hz resulting from classification as opposed to regression.

Comparison	Mean Difference in MAE Hz
$MAE_{STACKNet_C 17F} - MAE_{STACKNet_C}$	25.7
$MAE_{STACKNet_R 17F} - MAE_{STACKNet_R}$	26.9
$MAE_{STACKNet_C} - MAE_{STACKNet_R}$	6.1
$MAE_{STACKNet_C 17F} - MAE_{STACKNet_R 17F}$	4.9

notable that it takes longer to process a single record on a DL model than it does to process a batch size of 100 records. This is due to the hardware environment being more suited to parallel execution, which will be an important consideration when integrating DL into other systems. Such an estimate may be taken as an average across windowed sequences for the received signal. The brute force cross-correlation method is much more expensive than the other two given the wide frequency range.

Those models constructed with data augmentation demonstrate an improvement over those learning from the unprocessed signal in the randomised setting. Both variants of the models (classification and regression) appear consistent in the influence of each of the features shown in Figure 8. One notable difference is that they disagree on the influence of the lagged difference for the conjugate of the signal where the imaginary value does not contribute as highly to the model

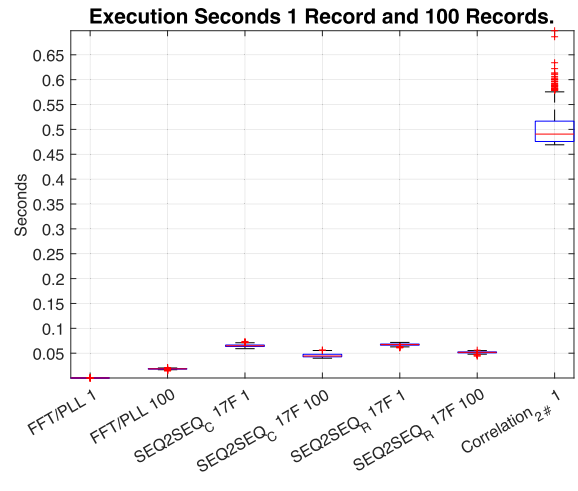


FIGURE 7. Execution speed of the simpler FFT/PLL method is faster in comparison to the deep network model which performs well on larger batches and is faster than the brute force cross-correlation method.

accuracy for the classification model STACKNet_C 17F as opposed to the regression model. Variables contributing the lowest scores include the signal raised to 4th power and the lag-1 difference of phase in the signal. Both models nominate the phase of the squared signal $\angle r^2$ as causing the highest MAE when the feature is replaced with Gaussian noise. The low resolution FFT (length of 104), appears to be influential to both models, however is not able to be used in isolation

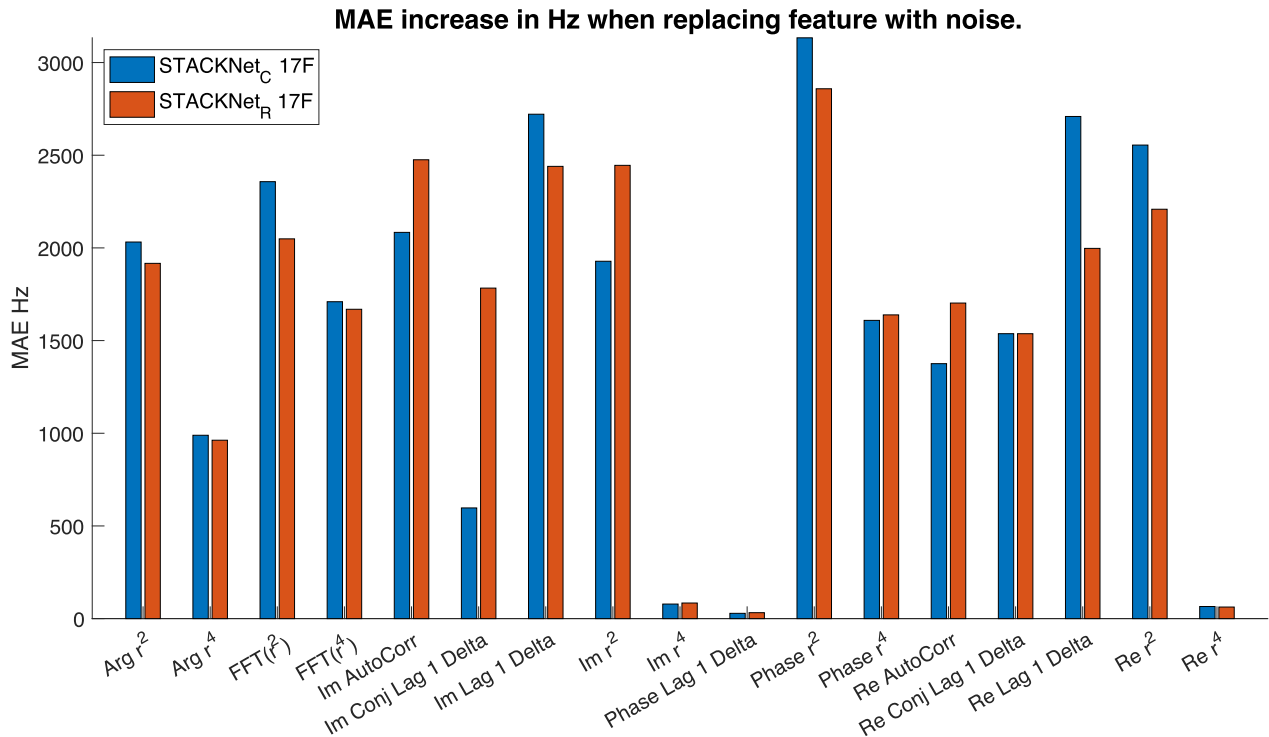


FIGURE 8. Feature importance indicated by the increase in MAE Hz when the feature is replaced with gaussian noise.

from the auto-correlation and squared polar form of the signal.

A. DISCUSSION

After training the proposed stacked network on the selected set of modulations, the model was able to produce more accurate CFO estimates than the FFT/PLL and the cross-correlation methods for short message sequences. On the other hand, the cross-correlation method required a longer message sequence to outperform the DL model. As shown in the related research, DL is capable of CFO estimation for short random sequences [2] and for noisy sinusoidal modulations [3], [18]. The stacked network models are also able to accept random sequences of several chaotic maps without reference to a template pilot sequence, indicating the ability of the trained network to estimate CFO without explicit knowledge of the feedback parameters for these types of signals. As such, this methodology is suitable for use with chaotic modulations and, given the ability to estimate frequency offset, it may be possible for such a method to estimate additional parameters required for chaotic synchronisation, such as the time dependent state variables of the chaotic map. Future research in this task may investigate the use of encoder-decoder networks in the estimation and tracking of multiple chaotic system parameters such as in [29].

Data augmentation was applied to the model, and in the randomised setting, demonstrated an improvement of approximately 20 Hz MAE over those models which did not make use of data augmentation. In the fixed preamble setting, data

augmentation did not demonstrate much influence over the performance of the model, this is indicative that the variation in message content is influential over the performance of the model, with a fixed preamble illustrating low variation (outside of the channel model) as opposed to randomised sequences. While DL is capable of representation learning without the requirement of manual feature engineering, it is also true that domain specific feature engineering does provide an advantage in the application of DL. Such an approach indicates that DL will be most useful where it can be incorporated into communications systems alongside conventional signal processing methods in a hybridised form.

In this study we applied simulations with an AWGN channel to generate the required data. The difficulty in the supervised learning approach is the requirement for off-line training, which requires a large volume of data especially when training across multiple signal modulations. The amount of data required increases with each supported modulation so as to ensure an equal sized population for each modulation in the training set. However this research has not investigated the potential for transfer learning [16] to enable the network to adapt to new modulations or channel models, which is a topic for future investigation.

Performance of the model is influenced by the modulation of the signal as shown in the results, hence the network model is learning features related to the modulation in the carrier offset estimation task. In an end-to-end learning setting, it may be possible to dynamically learn a suitable modulation to reduce receiver error as demonstrated in works such as [30]

and [31]. Future work will investigate methods of incorporating learnt CFO estimation which may jointly benefit from the modulations learnt at the transmitter, necessarily moving from an offline supervised learning problem to an online learning problem.

Execution timing demonstrated that the DL model is more efficient on batches of signal frames rather than on a single signal frame. This is also a result consistent with the benchmarking performed in [32]. This poses a design challenge for the practical application of DL models in communications systems, where batches of signal frames will be necessary to most efficiently make use of the DL architecture. Future work will be required to investigate the practical implementation challenges of integrating DL based CFO estimation within an end-to-end wireless communications system.

IV. CONCLUSION

In this article we have demonstrated the use of a stacked sequence-to-sequence encoder to perform carrier frequency offset estimation in multiple modulations, including for feedback dependent chaotic maps. The proposed architecture has been shown to outperform FFT/PLL and cross-correlation methods on short sequences, in both the fixed preamble setting and in the randomised setting without knowledge of the modulation, and in the randomised setting without a pilot template. However increasing the message sequence length did enable the cross-correlation method to outperform the DL model, at the expense of additional execution time. Data augmentation in the randomised setting, was shown to provide an increased accuracy for the CFO estimation (of approximately 20 Hz) and indicates that while DL models are capable of learning feature representations directly from raw IQ values, the use of appropriately chosen features is an avenue for enhancing the performance of the model. Iterative estimation was performed by separate stages of the stacked network architecture with an error correction performed at the final stack, thereby taking advantage of the composability of DL modules as a means of iteratively refining the CFO estimate. This work demonstrates the capability of DL techniques to estimate the carrier offset parameter for chaotic communications, and provides an incremental step towards the application of DL in short messaging systems and chaotic communication.

REFERENCES

- [1] W. G. Cowley, "Phase and frequency estimation for PSK packets: Bounds and algorithms," *IEEE Trans. Commun.*, vol. 44, no. 1, pp. 26–28, Jan. 1996.
- [2] R. M. Dreifuerst, R. W. Heath, Jr., M. N. Kulkarni, and J. Charlie, "Deep learning-based carrier frequency offset estimation with one-bit ADCs," in *Proc. IEEE 21st Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, May 2020, pp. 1–5.
- [3] H. R. Almayyali and Z. M. Hussain, "Deep learning versus spectral techniques for frequency estimation of single tones: Reduced complexity for software-defined radio and IoT sensor communications," *Sensors*, vol. 21, no. 8, p. 2729, Apr. 2021.
- [4] J. Zhang, T. Q. Duong, R. F. Woods, and A. J. Marshall, "Securing wireless communications of the Internet of Things from the physical layer, an overview," *Entropy*, vol. 19, no. 8, p. 420, 2017.
- [5] G. Kolumbán, M. P. Kennedy, and L. O. Chua, "The role of synchronization in digital communications using chaos. II. Chaotic modulation and chaotic synchronization," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 45, no. 11, pp. 1129–1140, Nov. 1998.
- [6] B. Jovic, *Synchronization Techniques for Chaotic Communication Systems*. Berlin, Germany: Springer, 2011.
- [7] A. J. Michaels, "Digital chaotic communications," Ph.d. dissertation, Georgia Inst. Technol., Atlanta, GA, USA, 2009.
- [8] J. Tao, L. Yang, and X. Han, "Enhanced carrier frequency offset estimation based on Zadoff–Chu sequences," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1862–1865, Oct. 2019.
- [9] D. C. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 5, pp. 591–598, Sep. 1974.
- [10] M. D. Macleod, "Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 141–148, Jan. 1998.
- [11] B. G. Quinn, "Estimating frequency by interpolation using Fourier coefficients," *IEEE Trans. Signal Process.*, vol. 42, no. 5, pp. 1264–1268, May 1994.
- [12] B. G. Quinn, "Estimation of frequency, amplitude, and phase from the DFT of a time series," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 814–817, Mar. 1997.
- [13] E. Jacobsen and P. Kootsookos, "Fast, accurate frequency estimators [DSP tips & Tricks]," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 123–125, May 2007.
- [14] U. Orguner and C. Candan, "A fine-resolution frequency estimator using an arbitrary number of DFT coefficients," *Signal Process.*, vol. 105, no. 10, pp. 17–21, Dec. 2014.
- [15] A. Host-Madsen and P. Handel, "Effects of sampling and quantization on single-tone frequency estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 650–662, Mar. 2000.
- [16] Y. Bengio, Y. LeCun, and G. E. Hinton, "Deep learning for AI," *Commun. ACM*, vol. 64, pp. 58–65, Jun. 2021.
- [17] G. Izacard, B. Bernstein, and C. Fernandez-Granda, "A learning-based framework for line-spectra super-resolution," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 3632–3636.
- [18] V. Ninkovic, A. Valka, D. Dumic, and D. Vukobratovic, "Deep learning-based packet detection and carrier frequency offset estimation in IEEE 802.11ah," *IEEE Access*, vol. 9, pp. 99853–99865, 2021.
- [19] Mathworks. *Commun.Coarse Frequency Compensator Compensate for Frequency Offset of PAM, PSK or QAM Signal*. Accessed: Sep. 29, 2022. [Online]. Available: <https://au.mathworks.com/help/comm/ref/Commun.coarsefrequencycompensator-system-object.html>
- [20] Y. Wang, K. Shi, and E. Serpedin, "Non-data-aided feedforward carrier frequency offset estimators for QAM constellations: A nonlinear least-squares approach," *EURASIP J. Adv. Signal Process.*, vol. 2004, no. 13, Dec. 2004, Art. no. 856139.
- [21] Mathworks. *Commun.Carriersynchronizer Compensate for Carrier Frequency Offset*. Accessed: Sep. 29, 2022. [Online]. Available: <https://au.mathworks.com/help/comm/ref/Commun.carriersynchronizer-system-object.html>
- [22] M. Rice, *Digital Communications: A Discrete-Time Approach*. Upper Saddle River, NJ, USA: Prentice-Hall, 2009.
- [23] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.
- [24] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [26] A. Veit, M. J. Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 550–558.
- [27] L. N. Smith, "Cyclical learning rates for training neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 464–472.
- [28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [29] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, "Data-driven discovery of coordinates and governing equations," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 45, pp. 22445–22451, Nov. 2019.

- [30] T. J. O'Shea, K. Karra, and T. C. Clancy, "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2016, pp. 223–228.
- [31] F. A. Aoudia and J. Hoydis, "End-to-End learning of communications systems without a channel model," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 298–303.
- [32] S. Bianco, R. Cadene, L. Celona, and P. Napoletano, "Benchmark analysis of representative deep neural network architectures," *IEEE Access*, vol. 6, pp. 64270–64277, 2018.



CHRISTOPHER P. DAVEY (Graduate Student Member, IEEE) received the master's degree in information technology from the Queensland University of Technology (QUT), Australia, in 2007, and the M.Sc. degree in mathematics and statistics from the University of Southern Queensland (USQ), Australia, in 2020, where he is currently pursuing the Ph.D. degree in deep learning for wireless communications. He has over a decade of professional experience in software development and systems integration. He has worked on "Artificial Intelligence for Decision-Making (AI4DM)" and "AI-Enabled Communicating Systems" research project funded by the Australian Government's Department of Defence.



ISMAIL SHAKEEL (Senior Member, IEEE) received the B.Eng. degree (Hons.) in electronic engineering from the University of South Australia, in 1997, the two master's degrees from the University of Canterbury (NZ) and Monash University, in 2001 and 2002, respectively, and the Ph.D. degree in telecommunications, in 2007. He joined the Defence Science and Technology Group (DSTG), in 2011. He is currently with the Information Sciences Division, DSTG, and also an Adjunct Professor with the University of Southern Queensland. Before joining DSTG, he has worked in both academia and industry and holds a patent and generated more than 40 technical reports and publications in the field of telecommunications. His current research interests include signal detection and classification techniques, artificial intelligence-enabled wireless communication, interference-resistant signalling, chaotic communication, and cooperative wireless communication.



RAVINESH C. DEO (Senior Member, IEEE) leads Advanced Data Analytics Research Laboratory as a Professor at the University of Southern Queensland (USQ), Australia. He is a Clarivate Highly Cited Researcher with publications ranking in top 1% by citations for field and publication year in the Web of Science citation index and is among scientists and social scientists who have demonstrated significant broad influence, reflected in the publication of multiple papers frequently cited by their peers. He leads cross-disciplinary research in deep learning and artificial intelligence having supervised more than 30 Ph.D./M.Sc. degrees. He has published more than 270 articles, 150 journals, and seven books with a cumulative citation that exceeds 11,100. He received the Employee Excellence Awards, the Elsevier Highly Cited Paper Awards, and the Publication Excellence and Teaching Commendations.



SANCHO SALCEDO-SANZ was born in Madrid, Spain, in 1974. He received the B.S. degree in physics from the Universidad Complutense de Madrid, Madrid, in 1998, the Ph.D. degree in telecommunications engineering from the Universidad Carlos III de Madrid, Madrid, in 2002, and the Ph.D. degree in physics from the Universidad Complutense de Madrid, in 2019. He spent one year at the School of Computer Science, University of Birmingham, U.K, as a Postdoctoral Research Fellow. He is currently a Full Professor with the Department of Signal Processing and Communications, Universidad de Alcalá, Spain. He has coauthored more than 220 international journal articles in the field of machine learning and soft-computing and its applications. His current research interests include soft-computing techniques, hybrid algorithms, and neural networks in different problems of science and technology.



JEFFREY SOAR is currently a Personal Chair in human-centered technology with the School of Business, University of Southern Queensland. He came to academic research from a long and distinguished career in industry including as the Chief Information Officer in government agencies in Australia and New Zealand. His research interests include AI, e-business, e-health, technology and development, and social and organizational change.

...