Cryptocurrency Prediction Papers:

1. **[A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms]**

(Hamayel & Owda, 2021)

This study involves machine learning algorithms such as LSTM, bidirectional LSTM (bi-LSTM) and gated recurrent unit (GRU) to predict cryptocurrency prices. The data utilised was accessed from Marketwatch.com and involved the following five features: Date, Open, High, Low, and Close. The results are presented based on Root Mean Squared Error [RMSE] and Mean Absolute Percentage Error [MAPE]. Overall GRU performs more accurate price predictions than LSTM and bi-LSTM, with the lowest MAPE and RMSE results [Litcoin: MAPE 0.2116| RMSE: 0.825, Ethereum: MAPE 0.8267| RMSE: 26.59, Bitcoin: MAPE: 0.2454 | RMSE: 174.129]. However, all algorithms were noted as presenting excellent predictive results.

2. **[Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis]**

(Abraham et al., 2018)

This paper utilises a linear model that takes tweets and Google Trends as data inputs,, to accurately predict price changes to Bitcoin and Ethereum. The sentiment of tweets was captured using Natural Language Processing. However, this was found to be an unreliable indicator of whether cryptocurrency prices were falling, since the correlation was low. Conversely Google Trends and tweet volume were proven to be highly correlated with price. Therefore, Google Trends and tweet volume data were applied to multiple linear regression for predicting Bitcoin daily closing and opening price.

3. **[A Deep Learning-based Cryptocurrency Price Prediction Scheme for Financial Institutions]**

(Patel et al., 2020)

Patel et al. propose a new hybrid approach using an LSTM and GRU-based model to predict the prices of Litecoin and Monero cryptocurrencies. The data used was collected daily, with 1,279 data points for Litecoin and 1,851 data points for Montero, from Investing.com and involved five features (Price, Open, Close, High, Low and Volume). The data was prepared using min-max normalisation. The proposed hybrid approach is compared against an LSTM model based on MSE[Mean Squared Error], RMSE [Root Mean Squared Error], MAE [Mean Absolute Error] and MAPE [Mean Absolute Percentage Error]. Overall, Patel et al. 's (2020) LSTM and GRU-based hybrid model outperforms the LSTM model. However, the difficulty of cryptocurrency price prediction is stressed throughout the paper. This is largely attributed to the notoriously volatile nature of cryptocurrency that is often affected by external factors such as advancements in technology and public perception.

4. **[Stochastic Neural Networks for Cryptocurrency Price Prediction]**

(Jay et al., 2020)

The proposed stochastic neural network model is based on a random walk theory and induces layer-wise randomness into the observed feature activations of neural networks to simulate market volatility. This experiment aims to combat the erratic fluctuations in Cryptocurrency, by incorporating stochasticity into neural networks, creating stochastic MLP and LSTM models. The average relative improvement by using stochastic neural networks over regular neural networks is 1.56% in the Norm dataset at $\gamma = 0{:}1$, 1.73% at $\gamma = 0{:}12$ and 1.76% when it is set as a learnable parameter. The improvement is much more significant for the UNorm dataset where the average relative improvement is 3.91% at $\gamma = 0{:}1$, 4.52% at $\gamma = 0{:}12$ and 7.41% when it is set as a learnable parameter.

5. **Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system**

(M. et al., 2020)

M. et al. (2020), propose a price prediction implementation with two machine learning methods, linear regression (LR) and support vector machine (SVM), by using a time-series of daily Ethereum closing prices. The experimental evidence suggests that the SVM without features has a 10.62% higher accuracy [96.06%] than the LR method [85.46%]. The accuracy further increases 99% when additional features are incorporated into the SVM algorithm.

6. **Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators**

(Alonso-Monsalve et al., 2020)

This study focuses on four different network architectures - Convolutional neural networks, hybrid CNN-LSTM network, multi-layer perceptron and radial basis function neural network to predict whether six popular cryptocurrencies [Bitcoin, Dash, Ether, Litecoin, Monero and Ripple] will increase their value for the next minute. The Convolutional LSTM neural networks were shown to surpass the other architectures, while the CNN neural networks were also able to provide good results specially with Bitcoin, Ether and Litecoin. The data consisted of price series vs. USD., sampled at 1-min intervals from 1st July 2018 – 30th June 2019 (~1 year) and was sourced from Cryptocompare.

7. **Cryptocurrency forecasting with deep learning chaotic neural networks**

(Lahmiri & Bekiros, 2019)

The work of Lahmiri & Bekiros (2019) appears to be the first to make use of deep learning methods to forecast digital currency prices. The LSTM deep learning systems were found to be capable of learning the chaotic and self-similar patterns for the three major cryptocurrencies (Bitcoin, Digital Cash and Ripple) better as well as predicting their future dynamics more robustly versus the generalised regression neural network (GRNN). The RMSE confirms this with LSTM achieving significantly lower scores than the GRNN, [Bitcoin: 2.75×10^3 | Digital Cash: 19.2923 | Ripple: 0.0499].

8. **Evolvable fuzzy systems from data streams with missing values: With application to temporal pattern recognition and cryptocurrency prediction**

(Garcia et al., 2019)

The focus of this paper is on proposing an evolving granular fuzzy-rule-based model for temporal pattern recognition and time series prediction in an online nonstationary context, where values may be missing. This model consists of a modified rule structure that includes reduced-term consequent polynomials and is supplied by an incremental learning algorithm that simultaneously imputes missing data and updates model parameters and structure. The proposed model adaptation Evolving Fuzzy Granular Predictor [eFGP] provides prediction in a time-varying environment employing several IF-THEN rules. A Bitcoin dataset from bitcoincharts.com/charts is used to evaluate eFGP. The behaviour of eFGP was shown to be stable in different scenarios. It was compared against the evolving Granular Neural Network

[eGNN], the evolving Takagi-Sugeno [eTS] and the extended Takagi-Sugeno [xTS]. The eFGP is competitive with these other methods in terms of processing time.

Summary of Papers using TFT:

1. **Inter-Series Attention Model for COVID-19 Forecasting**

(Jin et al., 2021)

Jin et al. (2021) perform Covid-19 Forecasting in terms of newly confirmed cases, hospitalisations and deaths. The experimentation included a set of leading neural forecasters including TFT and ConvTrans.. However, the TFT performs poorly in this experimentation. This is largely attributed to the fact there is no employment of attention across multiple time series, which would dramatically improve the TFT's performance. The proposed neural forecasting model called Attention Crossing Time Series [ACTS] is shown to outperform the TFT in this instance. This model degenerates to an autoregression model similar to ConvTrans and TFT.

2. **A Data-Driven Forecasting Strategy to Predict Continuous Hourly Energy Demand in Smart Buildings**

(Mariano-Hernández et al., 2021)

The objective of this paper is to present an energy consumption forecasting strategy that allows hourly day-ahead predictions. The presented forecasting strategy is tested using real data from two buildings located in Valladolid, Spain. It is observed that in building 1 XGBoost and TCN performs the best based on the MAPE, while in building 2 CNN and TCN had the best MAPE results. In general, the assembled models obtained better results than the single models. The TFT is examined on the single model only.

Table 2. Performance of each model by building.

	Building 1				Building 2			
Model	R^2	RMSE (kWh)	MAE (kWh)	MAPE (%)	R^2	RMSE kWh)	MAE (kWh)	MAPE (%)
MLR	0.72	37.34	26.19	16.79	0.72	21.01	15.27	38.31
ANN	0.75	35.48	22.56	13.26	0.79	18.19	12.22	26.84
RF	0.83	29.45	16.2	9.22	0.86	15.16	9.15	19.68
XGBoost	0.85	27.23	15	8.83	0.87	14.12	8.22	17.92
LSTM	0.79	32.26	17.74	10.18	0.83	16.38	9.33	19.36
CNN	0.81	30.71	17.07	9.38	0.81	17.35	9.59	16.96
TCN	0.83	29.43	15.84	9.02	0.84	16.04	9.01	17.74
TFT	0.69	39.57	17.7	9.22	0.84	16.08	9.53	18.84

3. **Forecasting Market Prices using DL with Data Augmentation and Meta-learning: ARIMA still wins!**

(Shah & Shroff, 2021)

This study focuses on comparing the performance of deep-learning techniques for forecasting prices in financial markets. Shaf & Shroff (2021), apply gradient-based meta-learning to account for the non-stationarity of financial time series. Due to the chaotic and noisy nature of financial data, the fact that deep learning models are often trained on multiple time series, the feedback signals are also extremely noisy leading to poor optimisation. The synthetic data is deterministic, leading to noiseless signals that data deep learning methods perform better on. However, in this study, the autoregression model ARIMA model is shown to outperform the TFT with the inherently chaotic financial data. In general, ARIMA performs better for prediction when the regularities enabling prediction are themselves non-stationary and cannot, therefore, be exploited by machine-learning models. Since ARIMA relies purely on local signal computed every time it is applied. The local signal is much stronger when there is more noise in the longer signals. This is true for financial data since the inefficiencies that lead to potential trading signals are not similar over time and change continuously.

Table 1: Results on three datasets **SN** - Synthetic, **FR** - Forex, and **BN** - Banknifty for **NTR** - Normal Training, **DA** - Data Augmented Training and **M2L** - Meta Learning. **OS** - One step prediction, **TS** - Tens step prediction, **Average DL** - Average of performance of all five deep learning models

			ARIMA		Average DL		Best DL Model		Best DL
			RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	Model Name
SN	NTR	OS	0.5496	0.0387	0.3837	0.0127	**0.2313**	**0.0102**	NBEATS
		TS	0.8561	0.0527	0.8126	0.0454	**0.7591**	**0.0409**	NBEATS
	M2L	OS	-	-	0.3655	0.0262	0.2961	0.0219	NBEATS
		TS	-	-	0.7510	0.0530	0.7205	0.0514	LSTM
FR	NTR	OS	**0.1982**	**0.0027**	0.2962	0.0029	0.2734	0.0027	LSTM
		TS	**0.4754**	**0.0056**	0.5923	0.0058	0.5648	0.0055	LSTM
	DA	OS	-	-	0.2849	0.0028	0.2609	0.0027	Transformer
		TS	-	-	0.5980	0.0059	0.5680	0.0055	LSTM
	M2L	OS	-	-	0.3443	0.0038	0.3235	0.0036	MLP
		TS	-	-	0.6579	0.0069	0.6029	0.0064	LSTM
BN	NTR	OS	**435.18**	**0.0131**	816.63	0.0187	704.68	0.0158	NBEATS
		TS	**1196.09**	**0.0314**	1491.46	0.0337	1396.34	0.0312	LSTM
	DA	OS	-	-	858.85	0.0200	728.65	0.0165	MLP
		TS	-	-	1486.22	0.0339	1397.29	0.0313	LSTM
	M2L	OS	-	-	877.94	0.0238	784.62	0.0213	MLP
		TS	-	-	1674.17	0.0434	1562.72	0.0403	LSTM

Table 2: Rollout Testing with and without Data Augmentation for non-autoregressive models

			MLP		NBEATS		TFT	
			RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
SN		OS	0.2754	0.0127	0.2315	0.0102	0.4361	0.0225
		TS	0.9691	0.0522	0.9957	0.0529	0.8567	0.0479
FR	NTR	OS	0.2935	0.0029	0.3233	0.0033	0.3936	0.0038
		TS	0.6030	0.0059	0.7699	0.0079	0.7036	0.0068
	DA	OS	0.3093	0.0030	0.2984	0.0029	0.3005	0.0030
		TS	0.6877	0.0064	0.5982	0.0058	0.6784	0.0067
BN	NTR	OS	730.23	0.0164	707.82	0.0159	977.06	0.0233
		TS	1480.58	0.0309	1413.56	0.0318	1431.35	0.0339
	DA	OS	727.78	0.0165	851.85	0.0209	1006.77	0.0239
		TS	1462.83	0.0306	1551.39	0.0374	1583.22	0.0371

4. Extreme Precipitation Seasonal Forecast Using a Transformer Neural Network

(Civitarese et al., 2021)

Civitarese et al. (2021) present an approach for forecasting the quantiles of the maximum daily precipitation in each week up to six months ahead using the TFT model. To the Authors' knowledge, this is the first time the TFT has been used to forecast extreme weather. The experiments compare TFT predictions in two regions against climatology and a calibrated ECMWF SEAS5 ensemble forecast (S5). Hyperparameter optimization with random search was followed since it is the same scheme presented by Lim et. al (2020)[TFT authors]. The training data ranged from 1981-2010 (30 years), the validation data involved values from 2011-2014 (4 years) and the test data was from 2015-2019 (5 years). The TFT significantly outperformed S5 with a considerably lower q-risk for all points. Once again, the TFT had lower q-risks for most locations compared to climatology. The TFT was found to accurately raise the quantile level and also respond to changes that climatology cannot.

5. **Analysis and forecasting of rivers pH level using deep learning**

(Srivastava & Cano, 2021)

This paper investigates different deep learning approaches to analyse and forecast pH levels, which includes LSTM, GRU, RNN and TFT to determine which algorithm provides the best pH predictive forecast. In this instance, the novel Temporal Fusion Temporal (TFT) model outperformed the other deep learning methods. The model was highly accurate with an average RMSE score of .217 on the validation data. The experimental data was collected from Virginia Current Conditions Water Quality Data on 72 bodies of water. A batch size of 128 was utilised and Optuna was implemented to optimise the TFT.

6. **The Method of Constructing a Development Trajectory as the Basis of an Intelligent Module for Strategic Planning of the EPM System**

(Moskalenko & Fonta, 2021)

In this paper, Moskalenko & Fonta (2021) employ computational intelligence methods for solving strategic problems based on the analysis in strategic management and forecasting. It is proposed to use a neural network with the Temporal Fusion Transformer architecture for achieving forecasting insights. The application of TFT is suggested but not implemented within this paper.

7. **Trading with the Momentum Transformer: An Intelligent and Interpretable Architecture**

(Wood et al., 2021)

Wood et al. (2021) introduce the Momentum Transformer, which is an extension to the LSTM-based DMN. It is an attention-LSTM hybrid Decoder-Only Temporal Fusion Transformer style architecture. The experiments used a portfolio of 50 of the most liquid, continuous futures contracts over the period 1990-2020. The TFT-based architecture is shown to be the best performing and most interpretable. However, the results from the non-hybrid architectures such as the Transformer and Informer outperformed the TFT model during the SARS-CoV-2 market crash.

8. **Flight Demand Forecasting with Transformers**

(Wang et al., 2021)

This paper applies an attention-based neural network, TFT, to improve flight departure demand forecast accuracy for Pace. It highlights how TFT outperforms traditional methods by large margins with Aviation System Performance Metrics (ASPM) and System Wide Information Management (SWIM) as inputs. The TFT is compared against other candidate models: linear regression (LR) and autoregression (AR), seq2seq and seq2seq with attention. The MSE of the TFT is 53% better than the AR model and 31% better than the LR model. Even when a single data source (ASPM) is used, TFT still produces the predictions with the lowest error.