JOURNAL OF INTERNATIONAL CRISIS AND RISK COMMUNICATION RESEARCH
ISSN: 2576-0017
2025, VOL 8, NO 2

Downside Risk of the South African Mining Index: Leveraging the Power of Long Short-Term Memory and Explainable AI

Katleho Makatjane[1], Claris Shoko[2] and Ntebogang Moroke[3]

[1,2]Department of Statistics, University of Botswana
[3]Department of Statistics and Operations Research, North-West University

Abstract

With a primary focus on downside risk (i.e., losses), this study uses a combination of deep learning and probability distributions to the South African mining index. We model the downside risk by combining probability distribution, SHAP, and long-short-term memory. In contrast, Expected shortfall (ES) and value at risk (VaR) are used in the downside risk assessment. To illustrate the varied levels of model performance across various distributions (normal, student t, skew-normal, generalised hyperbolic, and Laplace), these risk metrics are backtested at 90%, 95%, and 99% confidence intervals. The findings show that normalcy assumptions are inadequate for accounting for high losses. In terms of forecasting power, the generalised hyperbolic distribution performs better than all the other distributions, especially over longer forecasting horizons and greater confidence levels. Findings from this study further show that the Russian-Ukraine war, ZAR/USD exchange rate and COVID-19 have the highest impact on the downside risk.

Keywords: *Value at Risk, Expected Shortfall, Explainable AI, Long-Short-Term-Memory, SHAP*

1 Introduction

With the advancement of economic globalisation and deregulation, there has been a discernible escalation in price volatility, thereby amplifying the overall risk associated with revenue (Kakade, K., Jain, I., and Mishra, A. K. 2022). The Value-at-Risk (VaR) has gained recognition as a significant indicator for evaluating market risk, and it is commonly utilised by financial risk managers and institutions across the world. The Value-at-Risk offers a quantitative representation of the downside risk of an asset over a specific timeframe, during which it is invested with a particular statistical confidence level. This metric according to Makatjane, Moroke and Munapo (2021) serves various purposes, such as setting position limits, measuring risk-adjusted returns, and assessing models. With the advancement of economic globalisation and deregulation, there has been a discernible escalation in price volatility, thereby amplifying the overall risk associated with revenue (Kakade et al., 2022). The Value-at-Risk (VaR) has gained recognition as a significant indicator for evaluating market risk, and it is commonly utilised by financial risk managers and institutions across the world. The Value-at-Risk offers a quantitative representation of the downside risk of an asset over a specific timeframe, during which it is invested with a particular statistical confidence level. This metric according to Makatjane et al (2021) serves various purposes, such as setting position limits, measuring risk-adjusted returns, and assessing models. The Basel Committee's internal model approach, which enables institutions to establish their risk models for internal risk management, has resulted in an augmentation in the diversity of methods for calculating risk. The primary classification of VaR estimation models encompasses parametric and nonparametric models. Nonparametric

models, such as historical simulation, bootstrap simulation, and simulation by default, do not rely on a distribution of historical data and, therefore, have no underlying assumptions or parameters. Because of its simplicity, the non-parametric approach supports a wide range of positions, including speculative ones, and permits flexibility in their estimation. However, they are less reliable than parametric approaches and perform much worse in practice because they tend to overestimate risk for exceptionally volatile (or stable) periods. To estimate an asset's risk, the parametric approach to VaR uses an estimate of volatility and expected return in addition to examining the distribution of return on assets over time. Recent research has employed several parametric methods to estimate the VaR, including the use of complex volatility models, density function approaches, and the consideration of higher-order conditional moments.

Financial markets on the other hand allow investors and traders (those who seek to make a profit from short-term price movements) to make capital gains if they make the right choices. However, these movements are highly nonlinear and it is difficult to consistently make the right decisions. Due to the rapid development of machine learning and deep learning in recent years, considerable progress has been made in stock market forecasting (Çelik, İcan, and Bulut 2023). In the field of financial time series analysis, different methods have been investigated by different researchers; for instance, Bao, Yue, and Rao (2017) focused on open-high-low-close (OHLC) bars and technical indicators by combining wavelet transformations and stacked autoencoders with long-short-term memory (LSTM) networks. By contrast, Sirignano and Cont (2021) found that the assimilation of technical indicators with intraday stock data enhances network stacking performance, and further evidence that dilated convolutional networks outperform LSTM in several metrics was provided by Borovykh, Bohte, and Oosterlee (2017), while Dixon (2018) also contributed by showing that recurrent neural networks perform better on limit order book data than feedforward networks with lag characteristics. It is interesting to note that the machine learning literature, as demonstrated by the works of Hochreiter (1997) and Bayer (2015), appears to diverge from the statistical modelling literature, which is represented by the works of Box (2013) and Hamilton (2020). This discrepancy points to different viewpoints or approaches used by various research paradigms to handle the intricacies of financial time series data.

Building upon this approach, we employ Long Short-Term Memory networks to estimate Value-at-Risk and Expected Shortfall across varying market and economic conditions. To enhance the transparency and interpretability of our model, we integrate Explainable Artificial Intelligence techniques, specifically Shapley Additive Explanations. These methods enable the identification of key determinants influencing risk estimates, thereby fostering trust and confidence in decision-making processes. By emphasising feature importance, Shapley Additive Explanations improve model comprehensibility, equipping financial analysts and policymakers with deeper insights into risk dynamics. Additionally, we undertake a comparative evaluation of the normal, Laplace, skew-normal, Student’s t, and generalised hyperbolic distributions, integrating these with Long Short-Term Memory network predictions to refine both in-sample and out-of-sample forecasting of Value at Risk and Expected Shortfall across multiple time horizons and confidence levels. The incorporation of these distributions accommodates crucial data properties, including heavy tails, skewness, and extreme fluctuations, which are characteristic of financial time series. By synthesising explainability techniques with a comprehensive range of probability distributions, we enhance the precision, robustness, and practical applicability of risk estimation. To the best of our knowledge, this study is the first to propose such a comprehensive and rigorous methodological framework for improving the accuracy and interpretability of deep learning-based downside risk estimation in

the South African context. No prior research has provided such a systematic mapping and robust application to enhance both the predictive performance and explainability of advanced risk modelling techniques.

Research Highlights and Key Contributions

This study is one of the first to apply the combination of deep learning with probability distributions to the South African mining index. Moreover, we focus on the downside risk (i.e., the losses). This way, we aim to capture the possibility of achieving losses beyond the expected or average returns. For instance, if a stock has an expected return of 8%, but there is a possibility it could return 5% due to unfavourable market conditions, the difference (i.e., 8% - 5% = 3%) represents the downside risk (loss), hence, the application of our approach to these specific energy market has not been thoroughly explored in the existing literature. The highlights and key findings of this study are summarised in Table 1

Table 1: Performance and Contributions

Characteristic	Highlights	Findings	Contributions
Risk Estimation Method	Utilises LSTM networks for VaR and ES estimation	Effective in estimating downside risk in fluctuating market conditions	Pioneering LSTM-based downside risk prediction in South African markets
Model Interpretability	Incorporates Explainable AI (SHAP) for model transparency	Provides clear insights into key risk determinants	Enhances model transparency, aiding informed decision-making
Comparative Risk Estimation	Evaluates multiple probability distributions	Better fit for financial time series, capturing heavy tails and skew	Refines risk estimation by incorporating diverse distributions
Focus Area	South African mining sector	Accurately models downside risk for the sector	Tailored approach to sector-specific risk analysis
Methodological Framework	Bridges statistical modelling and machine learning	Provides a comprehensive risk estimation methodology	Combines traditional statistical methods with modern machine learning
Multi-Horizon Forecasting	Estimates VaR and ES over multiple time horizons	Provides a more complete view of risk across different periods	Expands the scope of risk assessment with multi-horizon forecasting
Innovative Approach	The first study to combine LSTM, probability distributions, and SHAP	Integrates deep learning with probability distributions for robust risk modelling	Integrates advanced techniques for a more comprehensive risk framework

Methods and Procedures

The data used is the five business day financial time series/Johannesburg stock exchange (FTSE/JSE) Mining Index. For the period of January 2010 to 28 February 2025 it was obtained from https://za.investing.com/indices/ftse-jse-mining. (accessed on 3 March 2025).

LSTM Architecture

The hidden layer's memory blocks comprise the LSTM's structure. Cells and input and output gates make up these blocks. The forget gate (f_t) can be used to overcome the LSTM's drawbacks. The forget gate resets the cell variable leading to the forgetting of the stored input, c_t, whereas the input and output gates manage the reading of inputs from the feature vector, x_t, and writing of output, h_t, respectively. The gates regulate the action of the memory block, whereas the forget gate weighs the information inside the cells, such that any time previous information becomes unimportant for some cells, it will reset the state of the different cells. Forget cells also enable prediction by making cells forget their previous state, thereby restricting biases in prediction. The computation operation within an LSTM block is as follows: Input values can only be conserved in the cell state if the input gate allows them. The input value of i_t and the expected value of the memory cells, $\tilde{C}_t$, at step t, is calculated as follows

$$i_t = \sigma\left(\omega_i\left[h_{t-1}, x_t\right] + b_t\right) \tag{1}$$

$$\tilde{C}_t = \tanh\left(\omega_c\left[h_{t-1}, x_t\right] + b_t\right) \tag{2}$$

Where $\omega_c\left[h_{t-1}, x_t\right]$ and b represent the weight matrices and the bias, respectively. The forget gate controls the weight of the state cell unit, and the value of the forget gate is computed as

$$f_t = \sigma\left(\omega_f\left[h_{t-1}, x_t\right] + b_f\right). \tag{3}$$

This process updates the new state of the memory cell as follows

$$\tilde{C}_t = i_t\tilde{C}_t + f_t\tilde{C}_{t-1} \tag{4}$$

and with this new state of the memory cell, the output of the gate is computed as

$$O_t = \sigma\left(\omega_x\left[h_{t-1}, xX_t\right] + b_0\right). \tag{5}$$

Hence, the final output value of the cell can then be explained as

$$h_t = O_t * \tanh\left(c_t\right) \tag{6}$$

where, i, f, o and c are input, forget, output gate and the cell activation vectors, respectively. All the features of the LSTM network architecture can be trained using the sigmoid ϕ and tanh functions. Figure 1 presents the standard structure of the LSTM.

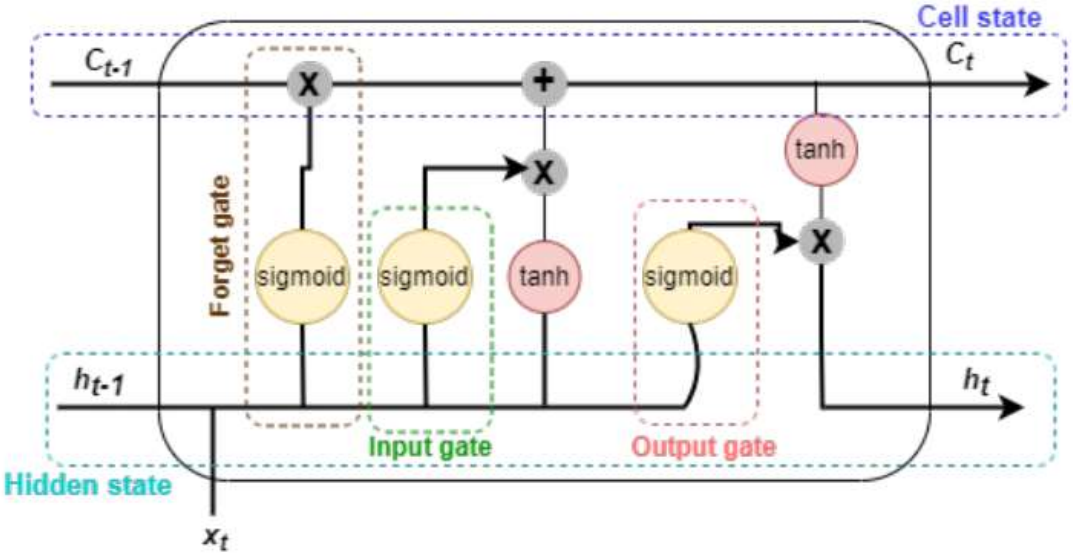

Figure 1: Structure of LSTM (Youness and Adam, 2023)

Shapley Additive Explanations (SHAP): Feature importance

In this study, we used Shapley Additive exPlanations (SHAP), an Explainable Artificial Intelligence (XAI) technique that explains the prediction process by providing information on how to explain the contribution of each feature to the final prediction from the fitted LSTM model. The SHAP is a method used to improve interpretability of artificial intelligence methods, and was proposed by Lundberg and Lee (2017). The idea behind the SHAP is to average the impact of a feature over all possible combinations of features in a linear model. According to Youness and Aalah (2023), the Sharpley value ϕ_j is defined as follows

$$\phi_j(f,x)=\sum_{S\subseteq\{x_1,\cdots,x_p\}\setminus\{x_j\}}\frac{|S|!(p-|S|-1)!}{p!}\left(f_x(S\cup\{j\})-f_x(S)\right) \tag{7}$$

Where p is the number of all features, S is a subset of features used in the model, and j is the j^{th} feature. f_x is the prediction function of a single input x, defined as

$$f_x(S)=\int f(x_1,\cdots,x_p)\,dP_{x\notin S}-E_X(f(X)) \tag{8}$$

The function performs several integrations for each of the features that do not contain S. A linear g model is finally fitted to the features and their impacts, represented as follows:

$$f(x)=g(x')=\phi_0+\sum_{i=1}^{p}\phi_i x_i \tag{9}$$

where $f(x)$ represents the original model, $g(x)$ is the explanatory model, x' is the simplified input such that $x=h_x(x')$, it has several omitted features, and $\emptyset_0=f(h_x(0))$; it represents the model output with all simplified inputs. Sharpley values are the only set of values that satisfy the following three properties: local accuracy, missingness, and consistency.

Risk Measures

Market risk is the chance that changes in market factors would result in a decline in the capital value. For financial institutions, one of the most important rules is to evaluate their exposure to market risk. Peihani (2016) states that estimating losses that might transpire when stock returns or asset prices in the portfolio are dropping is one way to measure this risk. This is the goal of the risk-measures methodology. These approaches have become a vital tool for financial institutions in managing market risk since the Basel Committee on Banking Supervision at the Bank for International Settlements mandates that a financial institution meet the capital requirements of basic risk estimates (Abad and Benito, 2013). In this study, we compare the two risk measures, Value-at-Risk and Expected Shortfall.

LSTM-Value-at-Risk and LSTM-ES

We predict volatility for various time horizons after training the LSTM model. The Value-at-Risk (VaR) and expected shortfall (ES) are computed using the forecasted vitality and several distributions, including normal, student t, Laplace, and skew-normal, among others. After that, the VaR and ES are computed over a range of time horizons, including one, two, and three months. For 90%, 95%, and 99% confidence intervals, this is carried out. Next, we use the backtesting procedure to assess how well risk measures VaR and ES perform. To evaluate the model's correctness, two-sided tests are typically used. In this study, we follow a similar approach proposed by Du et al (2019) and we model the market returns as

$$\begin{aligned} r_t &= \mu_t + z_t \sigma_t \\ \mu_t &= f_1\left(\Omega_{t-1}; \theta\right) \\ \sigma_t^2 &= f_2\left(\Omega_{t-1}; \theta\right) \end{aligned} \quad . \tag{10}$$

where r_t represents the log return at time t, μ_t denotes the corresponding conditional expected mean of the log return, and σ_t signifies its conditional volatility. The term z_t is an innovation random variable with a mean of zero and a standard deviation of one. We define Ω_{t-1} it as the set of information available at time t and θ as the set of parameters. Under this structure, the conditional mean μ_t and conditional volatility σ_t are estimated using two nonlinear functions, $f_1\left(\Omega_{t-1}; \theta\right)$ $f_2\left(\Omega_{t-1}; \theta\right)$ which are approximated by two separately trained LSTM architectures. With the estimated $\hat{\mu}_t$ and $\hat{\sigma}_t$ we then estimate the VaR at a time t at level as

$$VaR_\alpha^t = \hat{\mu}_t + q_\alpha \times \hat{\sigma}_t \tag{11}$$

where q_α is the α^{th} quantile of the assumed innovation distribution. Value-at-Risk is frequently seen as a sufficient risk indicator, although it does not account for subadditivity or other elements of market risk. Unlike VaR, ES evaluates the riskiness of an instrument by taking into account both the size and probability of losses over a specific threshold (Basel, 2019). The expected size of return that is above VaR, or for a probability level, is provided by the Expected Shortfall. More specifically, ES can be expressed as follows at a significance level of α.

$$ES_t^\alpha = E\left[r_t \mid r_t \le VaR_t^\alpha \right] \tag{12}$$

although ES and VaR are both widely used in finance for managing and estimating market risk, ES is thought to be a more prudent and dependable risk measure than VaR because it takes into

account the severity of losses that go beyond the VaR level. However, calculating ES requires estimating the entire tail of the distribution rather than just one percentile, which complicates the procedure.

Prediction interval evaluation metrics

Prediction intervals (PIs) are the key instruments utilised in probabilistic prediction, and these are a 188 set of potential values within which the predicted result exists based on a specified degree of confidence. They convey the intrinsic uncertainty inherent in the prediction model due to variability in data and the complexity of the systems being investigated. Prediction intervals can measure this uncertainty for better decision and risk management systems. In this study, we used the Mean Prediction Interval Width (MPIW).

Mean Prediction Interval Width (MPIW) is a metric used to evaluate the performance of prediction intervals, measuring the average width of the intervals, with smaller widths indicating better predictive accuracy. The MPIW is calculated as follows (Pearce et al., 2018)

$$MPIW = \frac{1}{n}\sum\left(U_i - L_i\right) \tag{13}$$

Where n is the number of prediction intervals, U_i is the upper limit of the i^{th} prediction interval, L_i is the lower limit of the i^{th} prediction interval. A good prediction interval method should aim to have a high probability of containing the true value (PICP) while also maintaining a narrow MPIW, indicating a more precise prediction.

Empirical analysis

This section provides an empirical analysis of the study, which used time series data to evaluate the LSTM architecture. Figure 1 presents the closing stock price trends. Figures 1(a) and 1(b) illustrate the non-normality of the data, while the kernel density plot in Figure 1(b) suggests a slight rightward skew in the distribution of closing stock prices. Moreover, Figure 1(a) highlights seasonal patterns accompanied by distinct positive and negative fluctuations. According to Jacobo and Marengo (2020), these instances of volatility clustering are attributable to major events such as the COVID-19 pandemic and the 2015–2016 Shanghai stock exchange crash. As depicted in Figure 1(a), this financial market exhibits the highest concentration of stock return losses, which may offer valuable insights when accounting for conditional heteroscedasticity.

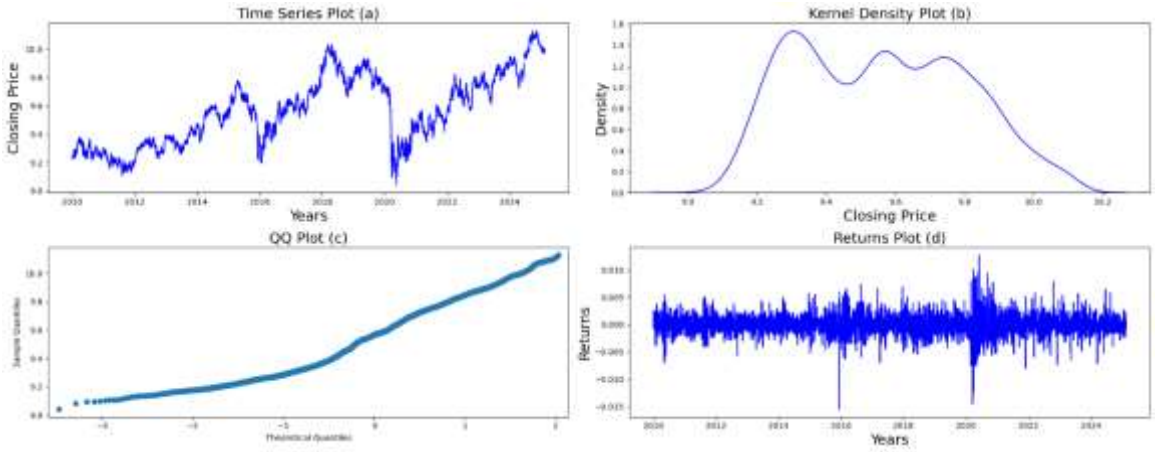

Figure 2 Time series plots

Moreover, to support the results reported in Figure 2(b) and Figure 2(c), the Jarque-Bera test statistic and Kolmogorov-Smirnov in Table 2 both reject the null hypothesis of normality implying that the returns on the FTSE/JSE Mining index understudy are not normally distributed. This is also confirmed by high Kurtosis indicating leptokurtic returns while the negative skewness indicates asymmetric returns. Meanwhile, the mean of the returns is positive, indicating a slightly positive average performance. Finally, the standard deviation, which measures the volatility or dispersion of the returns around the mean, has a value of 0.010100 which indicates a certain level of variability in the returns. Chinhamu et al. (2015a), using the same FTSE/JSE mining index in this study, observed that the mean was positive, indicating that the overall mining index was slightly increasing. The leptokurtic behaviour of the returns is indicated by a large kurtosis in their study, also indicating that the series has tails that are significantly heavier than those of the normal distribution.

Table 2 Descriptive statistics and normality tests

	Mean	**st dev**	**skewness**	**kurtosis**	**JB-Test**	**KS-Test**
Returns	0.0002443	0.010100	-0.5138089	7.63198	8376.572(0.001)	0.483628(0.001)

Short, Medium and Long-Term In Sample Forecasting

Based on the results reported in Table 3, the LSTM shows significant improvements in error metrics up to 300 epochs, after which performance deteriorates. Initially, as the number of epochs increases, the error metrics decrease, while coverage and accuracy improve, suggesting better generalisation. At 300 epochs, the model achieves its best performance with an MSE of 0.0038, MAE of 0.0443, and coverage of 71.19%, indicating an optimal balance between bias and variance. However, after 300 epochs, error metrics worsen, with MSE spiking to 0.1079 at 350 epochs, likely due to overfitting, as the model starts memorising noise instead of learning patterns. However, these results are in contrast with the one found by Makatjane, Shoko and Sigauke (2025) where the results of these authors showed that the LSTM takes longer to train but is faster to predict due to cellular memory caching implementation details beyond the scope of their work.

Table 3: The Forecasting performance of the LSTM

Architecture	**# epochs**	**MSE**	**MFE**	**MAE**	**MASE**	**WAPE**	**Coverage**	**Time**
LSTM	100	0.0299	0.1636	0.1636	21.5783	1.7056	0%	0.0979
	150	0.0061	0.0656	0.0656	8.6559	0.6842	50.00%	0.1179
	200	0.0039	0.0471	0.0476	6.2834	0.4967	65.23%	0.0899
	250	0.0041	0.0476	0.0478	6.3095	0.4987	67.28%	0.0905
	300	0.0038	0.0437	0.0443	5.8465	0.4621	71.19%	0.0886
	350	0.1079	0.2098	0.2098	27.6746	2.1874	29.01%	0.0905
	400	0.1066	0.2013	0.2013	26.5520	2.0987	45.88%	0.0881

Feature Importance and prediction of downside risk in the FTSE/JSE-Mining Index

To accurately forecast the downside risk, in high-frequency time series, Shoko, C., Moroke, N. D., Makatjane, K. (2024) advise that modelling this kind of time series data requires the inclusion of lagged features to allow forecasters to react dynamically to the changing conditions. Nonetheless, we include features such as the Ukraine and Russia war market

sentiments, the South African Rand against the United States dollar (ZAR/USD) exchange rate, the COVID-19 era, the 2015/2016 Shanghai Stock exchange crash market sentiments, interest and inflation rates of South Africa in this study. The SHAP summary plot in Figure 3 provides insights into the relative importance of features in predicting the outcomes for the FTSE/JSE mining index under study, with different economic and geopolitical factors influencing stock price predictions. The Russia-Ukraine war, the ZAR/USD exchange rate, and COVID-19 have the highest impact on the output of the model, suggesting that these events strongly affect market trends. The 2015/2016 Shanghai Stock Exchange crash, interest rate, and inflation also contributed, though their influence is relatively lower. Jabeur, Mefteh-Wali and Viviani. (2024) in their study also illustrate that the utilisation of XGBoost along with the SHAP approach had provided a significant boost in increasing the gold price forecasting performance.

Figure 3: A feature Importance showing Features that highly affect the Model.

To predict downside risk, we incorporate a range of distributions, to allow for a comprehensive assessment of risk by capturing varying degrees of tail heaviness and distributional properties. Our approach differs from that of Chinhamu et al. (2015b) because we use the LSTM architecture instead of the generalised Pareto distribution. Moreover, we mitigate the influence of extreme values through the application of the cubic spline method, which provides a more refined and smoothed representation of the data. Given in Table 4 to Table 6 are the VaR and ES estimates at different confidence levels, and different time horizons, with various distributions. At a quantile level of 90%, the estimated Value-at-Risk and Expected Shortfall across various distributions and time horizons are as follows: for a 5-day horizon, the normal distribution estimates a VaR of -3.2679 and an ES of -4.2807, suggesting that there is a 90% confidence that the market value will not decrease by more than 3.2679% nor experience losses greater than 4.2807%. The t-distribution gives a VaR of -0.1975 and an ES of 0.8492, indicating that the market value will potentially increase by 0.8492% in extreme scenarios, while the Laplace distribution estimates a VaR of -0.9595 and an ES of -29.8395, showing the possibility of extreme losses of up to 29.8395% at the 90% confidence level. On behalf of the skew-normal distribution, the VaR is -3.4447 and the ES is -102.0343, indicating the potential for severe losses. Finally, the generalised hyperbolic distribution gives a VaR of -0.2207 and an ES of -24.1733, suggesting a maximum loss of 24.1733% in adverse conditions. The same interpretation can be done for other forecasting horizons. Finally, these findings suggest that the FTSE/JSE mining index under study is highly susceptible to extreme market shocks.

Table 4 90% Confidence level in-sample Forecasting of VaR and ES

Time Horizon: 5 Days			**Time Horizon: 60 Days**		
Distribution	**VaR**	**ES**	Distribution	**VaR**	**ES**

Normal	-3.267886	-4.280729	Normal	-11.32029	-14.828879
t-distribution	-0.1975	0.84917	t-distribution	-0.684159	2.94161
Laplace	-0.959501	-29.839531	Laplace	-3.323808	-103.367169
Skew Normal	-3.444719	-102.034307	Skew Normal	-11.932857	-353.457207
Generalised Hyperbolic	-0.220735	-24.173343	Generalised Hyperbolic	-0.764649	-83.738915
Time Horizon: 30 Days			**Time Horizon: 90 Days**		
Distribution	**VaR**	**ES**	**Distribution**	**VaR**	**ES**
Normal	-8.004654	-10.485601	Normal	-13.864467	-18.161594
t-distribution	-0.483773	2.080032	t-distribution	-0.83792	3.602722
Laplace	-2.350287	-73.091626	Laplace	-4.070817	-126.59841
Skew Normal	-8.437804	-249.931988	Skew Normal	-14.614706	-432.894902
Generalised Hyperbolic	-0.540689	-59.212355	Generalised Hyperbolic	-0.9365	-102.558807

Table 5 95% Confidence level in-sample Forecasting of VaR and ES

Time Horizon: 5 Days			**Time Horizon: 60 Days**		
Distribution	**VaR**	**ES**	**Distribution**	**VaR**	**ES**
Normal	-4.045121	-4.939074	Normal	-14.012711	-17.109454
t-Distribution	-0.460702	1.701476	t-Distribution	-1.59592	5.894086
Laplace	-1.372068	-119.35348	Laplace	-4.752984	-413.452583
Skew Normal	-4.17578	-280.304258	Skew Normal	-14.465325	-971.002432
Generalised Hyperbolic	-0.343686	-313.338245	Generalised Hyperbolic	-1.190564	-1085.435521
Time Horizon: 30 Days			**Time Horizon: 90 Days**		
Distribution	**VaR**	**ES**	**Distribution**	**VaR**	**ES**
Normal	-9.908483	-12.098211	Normal	-17.161996	-20.954716
t-Distribution	-1.128486	4.167748	t-Distribution	-1.954594	7.218751
Laplace	-3.360867	-292.355125	Laplace	-5.821192	-506.373931
Skew Normal	-10.228529	-686.602404	Skew Normal	-17.716333	-1189.230248
Generalised Hyperbolic	-0.841856	-767.518817	Generalised Hyperbolic	-1.458137	-1329.381588

Table 6 99% Confidence level in-sample Forecasting of VaR and ES

Time Horizon: 5 Days			**Time Horizon: 60 Days**		
Distribution	**VaR**	**ES**	**Distribution**	**VaR**	**ES**
Normal	-5.503085	-6.228043	Normal	-19.063246	-21.574574
t-distribution	-3.156278	9.330764	t-distribution	-10.933666	32.322714
Laplace	-2.33002	-2983.79984	Laplace	-8.071427	-10336.185845

Skew Normal	-5.604596	-2597.643358	Skew Normal	-19.41489	-8998.50055
Generalised Hyperbolic	-0.820702	-3299012.823968	Generalised Hyperbolic	-2.842994	-11428115.651869
Time Horizon: 30 Days			**Time Horizon: 90 Days**		
Distribution	**VaR**	**ES**	**Distribution**	**VaR**	**ES**
Normal	-13.47975	-15.255528	Normal	-23.347612	-26.423349
t-Distribution	-7.731269	22.85561	t-Distribution	-13.390951	39.587078
Laplace	-5.707361	-7308.787103	Laplace	-9.885439	-12659.190603
Skew Normal	-13.7284	-6362.90076	Skew Normal	-23.778286	-11020.867399
Generalised Hyperbolic	-2.0103	-8080898.073621	Generalised Hyperbolic	-3.481942	-13996526.034296

The backtesting results for Value-at-Risk in Table 7 reveal that confidence levels of 90%, 95%, and 99% demonstrate varying degrees of model performance across different distributions. At the 90% level, the normal and skew-normal distributions exhibit significant underestimation of risk, with only 20 violations compared to the expected 50.7, resulting in low Kupiec test p-values (0.0394), indicating poor model performance. Conversely, the t-distribution and generalised hyperbolic models perform relatively well, with observed violations (49 and 48, respectively) closely matching expectations and p-values exceeding 10%, signifying an acceptable fit. At the 95% level, the normal and skew-normal models continue to show under-estimation of risk (20 violations versus 25.35), whereas the generalised hyperbolic distribution produces excessive violations (40), leading to the lowest p-value (0.0789). At the 99% confidence level, most distributions align reasonably well with expected violations, as reflected in higher p-values, except the t-distribution, which registers an excessive 22 violations compared to the expected 5.07, suggesting a tendency to overestimate tail risk at extreme quantiles. The same interpretation can be induced for the Expected Shortfall results in Table 8. Overall, while no single distribution performs optimally across all confidence levels, the generalised hyperbolic distribution offers the most consistent performance, particularly at higher confidence thresholds, making it a strong candidate for risk modelling in extreme market conditions.

Table 7 Backtesting of Value-at-Risk

Confidence Level	Distribution	Violations	Expected Violations	Kupiec Test p-value
90	Normal	20	50.7	0.0394
90	t-distribution	49	50.7	0.1066
90	Laplace	40	50.7	0.1789
90	Skew Normal	20	50.7	0.0394
90	Generalised Hyperbolic	48	50.7	0.1947
95	Normal	20	25.35	0.0394
95	t-distribution	27	25.35	0.1789
95	Laplace	26	25.35	0.2789
95	Skew Normal	20	25.35	0.1394
95	Generalised Hyperbolic	24	25.35	0.7789
99	Normal	3	5.07	0.4394

99	t-Distribution	22	5.07	0.7434
99	Laplace	4	5.07	0.7789
99	Skew Normal	2	5.07	0.5394
99	Generalised Hyperbolic	6	5.07	0.6789

Table 8 Backtesting of Expected Shortfall

Distribution	Confidence Level	ES Violations	Expected ES Violations	Kupiec ES Test p-value
Normal	90%	66	99.9	0.0061
t-Distribution	90%	79	99.9	0.0971
Laplace	90%	89	99.9	0.7580
Skew Normal	90%	90	99.9	0.7850
Generalised Hyperbolic	90%	98	99.9	0.7377
Normal	95%	26	49.95	0.0087
t-distribution	95%	34	49.95	0.0980
Laplace	95%	51	49.95	0.8971
Skew Normal	95%	53	49.95	0.887
Generalised Hyperbolic	95%	48	49.95	0.9277
Normal	99%	4	9.99	0.006
t-distribution	99%	5	9.99	0.011
Laplace	99%	7	9.99	0.128
Skew Normal	99%	13	9.99	0.127
Generalised Hyperbolic	99%	9.5	9.99	0.9207

Comparative Analysis

The purpose of this section is to determine the distribution which best mimics the data and also produces fewer forecasts. This helps in assisting the maximum dispatching of South Africa's mining industry. The four error metrics discussed in the previous section are used to measure the performance of each distribution and the results are summarised in Table 9. Some tentative conclusions are drawn from this table, which indicates that the generalised hyperbolic distribution dominates the other distributions. This distribution has the smallest values for all the error metrics used. What is more, the conclusion is made and found that in general, the generalised hyperbolic distribution is the one with fewer forecasting errors in forecasting the FTSE/JSE-Mining index in SA.

Table 9 Performance Model Selection Criteria

Value-at-Risk					
Metric	**Normal**	**t-Distribution**	**Laplace**	**Skew-Normal**	**Generalised Hyperbolic**
MSE	8.1591	2.2451	3.2616	8.4717	2.1387
MAE	2.8352	1.3836	1.7527	2.8881	1.3362
MASE	2.8564	1.4984	1.806	2.9106	1.4624
WAPE	271.9984	119.589	158.3403	277.554	114.6135
Expected Shortfall					
Metric	**Normal**	**t-Distribution**	**Laplace**	**Skew-Normal**	**Generalised Hyperbolic**
MSE	10.4347	2.5106	4.1047	10.7459	1.2209
MAE	3.1972	1.4926	1.9937	3.2431	0.4821
MASE	3.2303	1.5845	2.026	3.2781	1.1049
WAPE	310.0092	131.0258	183.6486	314.8266	112.1411

Out-of-Sample Forecasting with LSTM-Generalised hyperbolic distribution

After selecting the best distribution that best fits the data, out-of-sample forecasts are produced. The expected forecasts are presented in Table 10. Looking at both VaR and ES estimates, there is a more conservative and extreme downside risk as the forecasting horizons increase. This implies that the South African mining sector for the next 5 to 90 days after the prediction interval of 2025-02-04 will still experience more downside risk. The generalised hyperbolic distribution, which best fits the data, suggests that losses could reach 4.38% over 5 days and escalate to 19.13% over 90 days. Additionally, the ES values show that extreme losses could be even more severe, emphasizing the sector's vulnerability to adverse market movements. This trend implies that investors and industry participants should brace for increased volatility, particularly in response to external shocks such as fluctuations in commodity prices, regulatory changes, and global economic uncertainties.

Table 10 VaR and ES out-of-sample Forecasting Results

Distribution	Forecast Horizon	VaR	ES
Generalised Hyperbolic	5 days	-4.381064842414823	-1.683638614470457
	30 days	-10.972103487737249	-4.364785610473207
	60 days	-15.585689816296927	-6.241531263359911
	90 days	-19.12581911757154	-7.681608852392074

The estimated volatility also signals a shift in market sentiment, where cautious trading behaviour and potential sell-offs in mining stocks could amplify downward pressure on the index. Delle-Monache, De Polis and Petrella (2023).did not use an out-of-sample forecast but found a substantial increase in downside risk to U.S. economic growth emerged over the last 30 years, associated with the long-run growth slowdown started in the early 2000s. These results of sustained increase of downside risk are also conveyed by Figure 4.

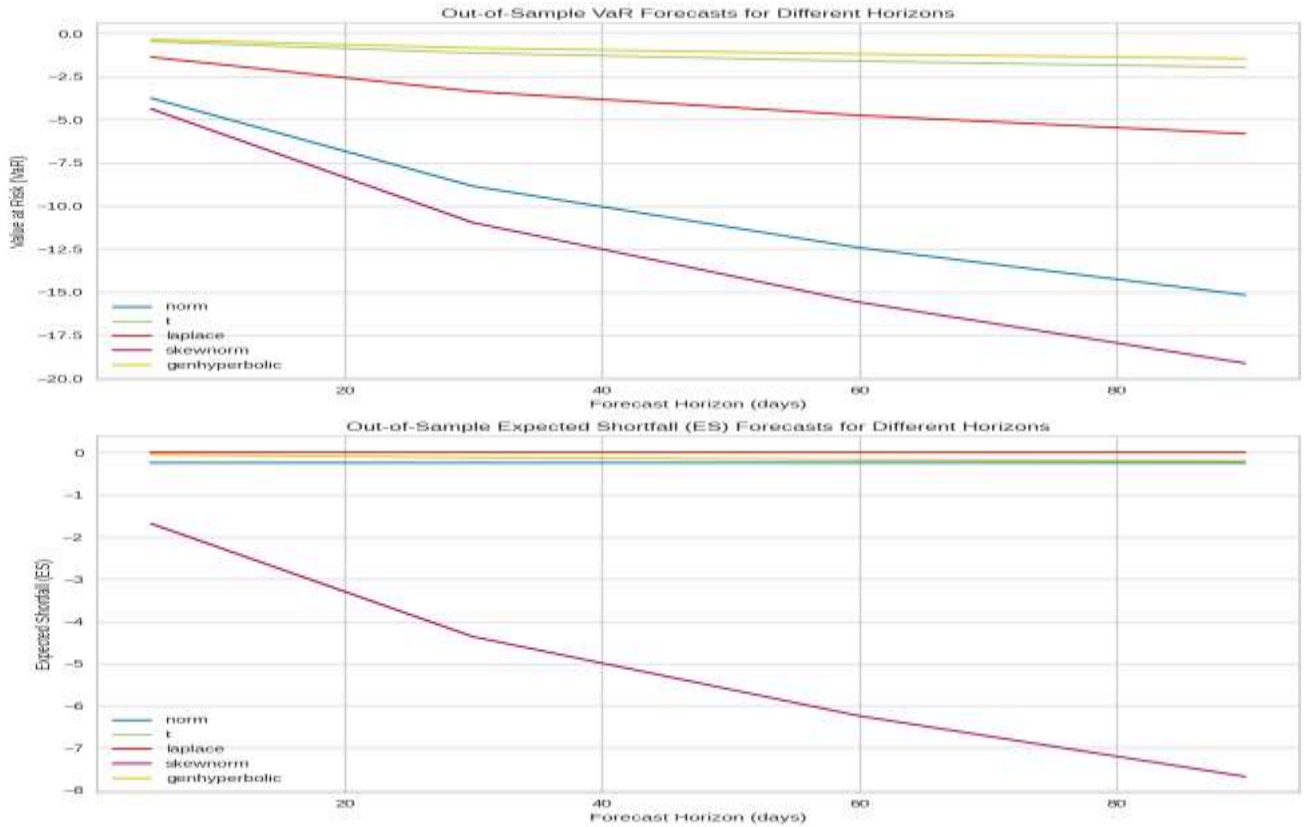

Figure 4 Forecasting of VaR and ES at different Horizons

Testing for the Stability of LSTM-generalised hyperbolic distribution

Evaluating model stability is essential to determine whether the model has accurately predicted extreme market fluctuations and prevent the underestimation of potential losses. The mean prediction interval width (MPIW) of 1.0277423, in conjunction with the 99% prediction interval depicted in Figure 5, offers valuable insights into the predictive accuracy of the model and the uncertainty surrounding its forecasts. The relatively narrow MPIW indicates that the model is capable of generating precise predictions while accounting for inherent uncertainties. The true values, represented by the blue line, remain predominantly within the 99% prediction interval (shaded grey region), demonstrating that the model effectively captures the underlying downside tail risk. This suggests that our combined model is robust in its ability to provide reliable out-of-sample forecasts, a crucial feature for economic and market applications.

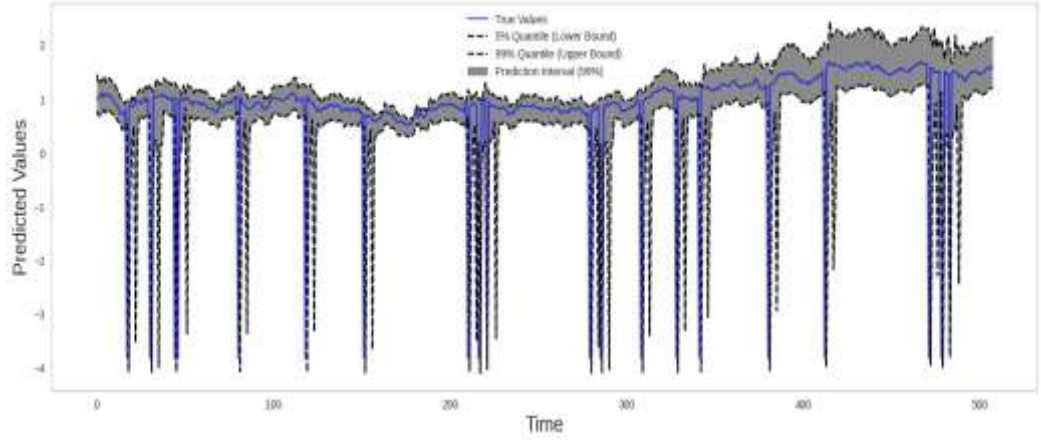

Figure 5 Assessment of Model Stability

Conclusion and Recommendations

By combining probability distributions with deep learning methods, this study provides a comprehensive analysis of patterns and forecasts related to the downside risk of the FTSE/JSE Mining Index. The feature importance analysis, based on Shapley Additive Explanations values, highlights the critical role of macroeconomic and geopolitical factors in predicting stock prices. Key influences, including the Russia-Ukraine war, fluctuations in the South African rand to the United States dollar exchange rate, and the COVID-19 crisis, have significantly shaped market sentiment and stability. The downside risk assessment, using value at risk and expected shortfall, highlights the inadequacy of normality assumptions in capturing extreme losses. Alternative distributions, such as the generalised hyperbolic distribution, offer more realistic estimates, particularly at higher confidence levels and over longer time horizons. Backtesting results reveal that while some distributions align reasonably well with expected violations, others, particularly the normal and skew-normal distributions, tend to underestimate risk. These findings emphasise the necessity of including external factors and non-normal distributions in financial forecasting and risk management. When properly trained and validated, the long short-term memory model effectively captures complex dependencies and enhances predictive accuracy. However, excessive reliance on normality assumptions has led to a systematic underestimation of downside risk, potentially resulting in inadequate capital buffers and flawed risk mitigation strategies.

One key limitation of this study is the sensitivity of the long short-term memory network, to hyperparameter selection. The performance and predictive accuracy of this architecture is found to be highly dependent on appropriate tuning, and suboptimal choices that lead to biased or inefficient forecasts. Additionally, the amalgamation of probability distributions with LSTM in this study significantly increases computational complexity, making real-time risk estimation challenging. The high computational demands also limit the practical applicability of methodology, particularly in environments with constrained resources.

To address these challenges, future research should explore the use of adaptive modelling techniques, such as regime-switching models or reinforcement learning, to account for the evolving impact of macroeconomic and geopolitical factors. Automated hyperparameter tuning methods, such as Bayesian optimisation, could enhance model performance by systematically identifying optimal configurations. Furthermore, leveraging distributed computing frameworks or cloud-based solutions may improve computational efficiency, enabling more scalable and real-time applications. Finally, conducting rigorous stress testing through simulated crises can provide deeper insights into model robustness under extreme conditions, ensuring more effective risk management strategies.

Acknowledgements

The authors thank numerous people for their helpful comments on this paper.

Competing interests

The authors declare that they have no competing interests.

Author contributions

All authors contributed equally to the conception and design of this manuscript. K.M. has developed the conducted data analysis of this research. C.S. has written the introduction,

abstract and discussion of results, conclusion and recommendations. While N.M. has proof read the final draft of the paper and work on reducing the similarity index of the paper. All in all, the authors have read and agreed to the published version of the manuscript.

References

Abad, P., and Benito, S. (2013). A Detailed Comparison of Value-at-Risk Estimates. *Mathematics and Computers in Simulation, 94*, 258-276. https://doi.org/10.1016/j.matcom.2012.05.011

Bao, W., Yue, J., and Rao, Y. (2017). A Deep Learning Framework for Financial Time Series using Stacked Autoencoders and Long-Short-Term Memory. PloS one, 12(7): e0180944. https://doi.org/10.1371/journal.pone.

Bayer, J. S. (2015). Learning Sequence Representations (Doctoral Dissertation), Technische Universit¨at M¨unchen. Available at https://mediatum.ub.tum.de/1256381. (Accessed on 2025 March, 25).

Borovykh,A., Bohte, S., and Oosterlee, C.W. (2017). Conditional Time Series Forecasting with Convolutional Neural Networks. Available at https://doi.org/10.48550/arXiv.1703.04691 (Accessed on 2025 March 16)

Box, G. (2013). Box Jenkins: Time Series Analysis, Forecasting, and Control. InA Very British Affair: Six Britons and the Development of Time Series Analysis During the 20th Century, 161-215: Springer. https://doi.org/10.1057/9781137291264 6

Çelik, T. B., İcan, Ö, and Bulut, E. (2023). Extending machine learning prediction capabilities by explainable AI in financial time series prediction. Applied Soft Computing, 132, 109876. https://doi.org/10.1016/j.asoc.2022.109876

Chinhamu, K., Huang, C. K., Huang, C. S., and Hammujuddy, J. (2015a). Empirical Analyses of Extreme Value Models for the South African Mining Index. *South African Journal of Economics*, 83(1), 41-55. https://doi.org/10.1111/saje.12051

Chinhamu, K., Huang, C. K., Huang, C. S., and Chikobvu, D. (2015b). Extreme risk, value-at-risk and expected shortfall in the gold market. *The International Business and Economics Research Journal (Online)*, 14(1), 107. https://www.proquest.com/docview/1646153038?pq-origsite=gscholar&fromopenview=true&sourcetype=Scholarly%20Journals

Delle-Monache, D., De Polis, A., and Petrella, I. (2023). Modeling and Forecasting Macroeconomic Downside Risk. Journal of Business and Economic Statistics, 42(3), 1010–1025. https://doi.org/10.1080/07350015.2023.2277171

Dixon, M. (2018). Sequence Classification of the Limit Order Book using Recurrent Neural Networks. Journal of computational science, 24:277-286. https://doi.org/10.1016/j.jocs.2017.08.018

Du, Z., Wang, M., and Xu, Z. (2019). On Estimation of Value-at-Risk with Recurrent Neural Network. In 2019 Second International Conference on Artificial Intelligence for Industries (AI4I) (pp. 103-106). IEEE. https://doi.org/ 10.1109/AI4I46381.2019.00034

Hamilton, J. D. (2020). Time Series Analysis: Princeton University Press. https://muse.jhu.edu/pub/267/monograph/book/77442

Hochreiter, S., and Schmidhuber, J. J. N. (1997). Long Short-term Memory. Neural Computation MIT-Press, 9(8):1735-1780. https://doi.org/ 10.1162/neco.1997.9.8.1735

Jacobo, A.D., Marengo, A., (2020). Are the Business Cycles of Argentina and Brazil Different? New Features and Stylised Facts. Paradigma econ´omico. Revista de econom´ıa regional y sectorial 12(2):5–38. https://doi.org/10.36677/paradigmaeconomico.v12i2.14028

Jabeur, S.B., Mefteh-Wali, S. and Viviani, JL. (2024) Forecasting gold price with the XGBoost algorithm and SHAP interaction values. Ann Oper Res 334, 679–699. https://doi.org/10.1007/s10479-021-04187-w

Kakade, K., Jain, I., and Mishra, A. K. (2022). Value-at-Risk Forecasting: A Hybrid Ensemble Learning GARCH-LSTM Based Approach. Resources Policy, 78, Article 102903. https://doi.org/10.1016/j.resourpol.2022.102903

Lundberg, S.M.; Lee, S. (2017). A Unified Approach to Interpreting Model Predictions. In Proceedings of the Annual Conference on NIPS 2017, Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9. https://dl.acm.org/doi/pdf/10.5555/3295222.3295230

Makatjane, K., Shoko, C., and Sigauke, C. (2025). An end-to-end Combined Forecasting Architecture: Forecasting Stock Price Data. Journal of Statistics Applications and Probability An International Journal 14(1), 59-75. http://dx.doi.org/10.18576/jsap/140105

Makatjane, K., Moroke, N., Munapo, E. (2021). Predicting the Tail Behavior of Financial Times Stock Exchange/Johannesburg Stock Exchange (FTSE/JSE) Closing Banking Indices: Extreme Value Theory Approach. In: Adıgüzel Mercangöz, B. (eds) Handbook of Research on Emerging Theories, Models, and Applications of Financial Econometrics. Springer, Cham. https://doi.org/10.1007/978-3-030-54108-8_2

Pearce, A. Brintrup, M. Zaki, and A. Neely (2018). High-quality prediction intervals for deep learning: A distribution-free, ensembled approach. In Proceedings of the 35th International Conference on Machine Learning, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., 80: 4075–4084. Available: https://proceedings.mlr.press/v80/pearce18a.html

Shoko, C., Moroke, N. D., and Makatjane, K. (2024). A Deep Learning Framework for Modeling Temporal Dependencies and Hierarchies in Hourly Electricity Demand Load. In Machine Learning and Computer Vision for Renewable Energy (pp. 42-65). IGI Global. https://doi.org/10.4018/979-8-3693-2355-7.ch00

Sirignano, J., and Cont, R. (2021). Universal Features of Price Formation in Financial Markets: Perspectives from Deep Learning. In Machine Learning and AI in Finance, 5-15: Routledge. https://doi.org/10.1080/14697688.2019.1622295

Youness, G.; Aalah, A. An Explainable Artificial Intelligence Approach for Remaining Useful Life Prediction. *Aerospace* **2023**, *10*, 474. https://doi.org/10.3390/aerospace10050474