Deep Learning-based VVIX Forecasting with Time Series Image Encoding and Hybrid ResNet-LSTM Model

Accepted for publication in Computational Economics.

https://doi.org/10.1007/s10614-025-11122-9

DEEP LEARNING-BASED VVIX FORECASTING WITH TIME SERIES IMAGE ENCODING AND HYBRID RESNET-LSTM MODEL

Ahoora Rostamian
Centre for Computational Finance and Economic Agents
University of Essex
Colchester, UK
ahoora.rostamian@gmail.com

John O'Hara
School of Mathematics and Acutarial Sciences
University of Essex
Colchester, UK
johara@essex.ac.uk

Delaram Jarchi
School of Computer Science and Electronic Engineering
University of Essex
Colchester, UK
delaram.jarchi@essex.ac.uk

ABSTRACT

This paper introduces an innovative approach to predict VVIX (Volatility of VIX) values using a combined image and recurrent pathway model. The dataset spans 2006 to 2023, providing insights into market volatility expectations for S&P 500 options. VVIX is crucial for risk management, portfolio allocation, and trading strategies. The model integrates spatial patterns and temporal dependencies through two pathways. The image pathway converts VVIX data into spatial images using Gramian Angular Fields and Markov Transition Fields and is processed through pre-trained ResNet-18 and convolutional layers. These transformations are suitable for VVIX forecasting as they effectively capture nonlinear temporal dependencies and transition dynamics in volatility data, enabling robust feature extraction for deep learning models. The recurrent pathway captures temporal trends with recurrent layers. Data is preprocessed with varying sliding windows for short-term, mid-term, and long-term sequences. The model is optimised with MSE loss and Adam optimiser, employing a decaying learning rate. Results show mid-term predictions yield balanced accuracy and training time. The proposed ResNet-LSTM model achieves a high coefficient of determination R^2 of 0.93, demonstrating robust accuracy in predicting VVIX. Further research should explore diverse model architectures, representations, and optimisation strategies, and assess generalisability to varying market conditions and external factors. In conclusion, the proposed model enhances predictive analytics for financial markets, aiding risk management and decision-making with improved VVIX forecasts.

Keywords VIX Volatility, Financial Forecasting, Volatility Forecasting, Transfer Learning

1 Introduction

Volatility, due to its non-observable nature, is often described as latent [15]. Challenges in quantifying volatility have directed researchers to propose multiple definitions [14]. Volatility in financial markets is defined as the degree of variation in the prices of financial assets, including stocks, bonds, currencies, and commodities. It aims to measure the associated risk of an investment, which is critical to market professionals. It affects all aspects of investment returns, portfolio construction, risk management, and trading strategies. High levels of financial volatility can complicate price prediction and therefore increase the investment risk. Different events, such as market sentiment and economic and political turmoil, have the power to adversely impact companies' financial performances [10].

VIX, or the Chicago Board Options Exchange Volatility Index, is a recognised measure of expected volatility. It represents the expected level of uncertainty in the market. Technically, the VIX index quantifies the 30-day expected implied volatility of options in S&P 500 index (SPX). Due to its tendency to rise during market downturns, it is interpreted as the market "fear gauge" reflecting heightened investor anxiety, and it presents valuable information about the market's expectations of future volatility. By utilising this metric, investors and traders can make more informed financial decisions. VVIX, or the volatility of VIX, on the other hand, measures the implied volatility of the VIX index. It is calculated using the prices of a range of options on the VIX with different strike prices and expiration dates. Therefore, it shows the expectations of VIX volatility. VVIX has multiple applications in the financial industry. It can be applied in portfolio risk management, emphasising the need to adjust positions. It can also assist in determining potential trading opportunities. For instance, a significant rise in VVIX may signal mispricing in options and provide an opportunity to capitalise on market inefficiencies. Another application is helping traders to take positions that offset losses caused by market volatility.

There are several methods to measure volatility. Statistical measures, option pricing models, and implied volatility are the main approaches. The standard deviation of returns, which measures the dispersion of returns around the mean, a widely used measure, is in the category of statistical methods. Historical volatility is another metric, which is calculated as the standard deviation of previous returns. The Black-Scholes model is an option pricing model that utilises implied volatility that is implied by the market price of an option. Generalised Autoregressive Conditional Heteroscedasticity (GARCH) is one of the earliest statistical models for volatility forecasting [1]. The GARCH model is a time series model that captures the conditional heteroscedasticity of financial returns. This term refers to the time-varying characteristic of the returns' volatility. In a study by Brailsford and Faff [2], a GARCH model was used to forecast realised volatilities based on daily prices.

Machine and deep learning models have recently gained popularity, parting ways from traditional techniques in quantitative finance [4]. Machine and deep learning capacities have tackled many of the problems in this domain, including return forecasting, risk modelling, and portfolio construction [12]. These innovative approaches are capable of looking into large amounts of data, extracting complex patterns, and delivering accurate predictions. Long Short-Term Memory (LSTM), a type of recurrent neural network, has been broadly used to predict price and volatility. Fischer and Krauss [6] employed the S&P 500 index daily prices over 23 years with a 240-day sequence to predict market movement using the LSTM network. They concluded that LSTM provides a more accurate prediction compared to Random Forest and can be used to create profitable trading strategies. In another study, Sagheer and Kotb [19], utilised a deep LSTM to model time series effectively. The proposed model was compared with ARIMA (AutoRegressive Integrated Moving Average), demonstrating its strength. Zang et al. [28] introduced a double-jump stochastic volatility model based on the general parametric affine model. They then derived a linear relationship between the stochastic volatility factor and the VVIX index.

This study addresses a significant gap in the existing literature on VVIX forecasting by utilising deep learning techniques, particularly pre-trained vision and recurrent networks, to develop a forecasting model. VVIX plays an important role in market volatility and investment decision-making, yet few studies have explored the applications of deep learning models in VVIX forecasting. This research introduces a novel hybrid model that combines a ResNet model for image-based spatial feature extraction with LSTMs as temporal modelling. It is designed to create a reliable tool for forecasting volatility by integrating recent advancements in computer vision and recurrent neural networks, supporting better risk management and decision-making.

2 Related Works

Market volatility is a key concern in financial markets, as it can significantly impact investment decisions. In the literature, there are a vast number of models to predict volatility. GARCH-based models are considered to be the most extensively used models for volatility forecasting [16]. Kroner et al. [11] explored the applicability of the combination of the GARCH-ISD (Implied Standard Deviation) model to forecast the price volatility of different commodities. Trück and Liang [22] investigated the GARCH, TARCH, TGARCH, and ARMA model's predictive capabilities in gold volatility forecasting. They concluded that the TARCH model produced the best results. Stochastic volatility models are another way to express volatility, as they assume that volatility has a stochastic process.

The aforementioned models are used in parallel with the Black-Scholes formula for derivatives pricing. The model used in [8] is the most important process, assuming that volatility follows a Cox-Ingersoll-Ross process and stock return is a Brownian motion. Machine learning and deep learning approaches represent another group of solutions to predicting volatility. Paiva et al. [13] applied Support Vector Machine and Hidden Markov Models to forecast financial time series. Deep learning approaches iteratively learn the non-linear internal correlation between data and can make precise predictions for complex systems like the stock market [23].

In another study [5], authors utilised deep neural networks to forecast the price changes and futures of multiple commodities within a 5-minute window. In their study, a 42% accuracy was the result of using the Back-propagation

(BP) algorithm. Xiong et al. [27] predicted the S&P 500 volatility using an LSTM-based model, concentrating on the effect of the input set on the forecasting results. A 60% increase in accuracy was reported and compared to other baseline models. Rostamian and O'Hara [18, 17] utilised one-dimensional CNN-LSTM to predict directional changes for four major currency pairs and reported high prediction accuracy of their proposed model. Selvin et al. [20] employed RNN, LSTM, and CNN to predict minute-by-minute prices. They determined the contribution of the CNN algorithm to the performance.

GARCH-based models stand out in capturing volatility clustering during stable market conditions; however, they are dependent on parametric assumptions and struggle with nonlinear patterns in dynamic markets, with studies like [22] reporting mean squared errors around 0.05 for gold volatility forecasting under volatile conditions. Similarly, CNN and LSTM models, as explored in [5, 6, 20], were effective in capturing short-term patterns or temporal dependencies; however, they failed to integrate both spatial and sequential features fully. For instance, [5] achieved only 42% accuracy in commodity price forecasting, while [6] reported daily returns of 0.46% for S&P 500 predictions, limited by the lack of spatial feature integration. In contrast, the proposed hybrid ResNet-LSTM model achieves an R^2 up to 0.93 (see Table 3), leveraging pre-trained ResNet-18 for robust spatial feature extraction and LSTM for temporal modeling to address these limitations and capture the complex, nonlinear dynamics of VVIX.

Recent studies emphasise the expanding role of advanced modelling in the financial domain and support the imperative for innovative approaches in VVIX forecasting. For instance, Zhang et al. [30] developed an interpretable deep learning framework to model complex interactions in time series data. They stressed the capacity of deep learning approaches to capture nonlinear dynamics in volatile systems. Similarly, Zeng et al. [29] employed quartile-based techniques to explore tail risk spillovers in financial markets, highlighting the need for models to address extreme market behaviours essential for volatility forecasting. In addition, Wu et al. [26] utilised statistical techniques to evaluate policy impacts. They illustrated the importance of integrating diverse data representations to enhance predictive accuracy. These studies support the motivation rationale for proposing the hybrid ResNet-LSTM model to effectively capture the intricate dynamics of VVIX.

3 Methodology

The proposed methodology uses supervised learning to predict VVIX values. It merges image-based and recurrent pathways. Time series data are transformed into Gramian Angular Fields and Markov Transition Fields as described in subsections 3.2 and 3.3 to detect nonlinear temporal patterns suitable for convolutional processing. The ResNet-18 is the model for the image pathway architecture due to the balance of depth and computational efficiency. It leverages pre-trained weights from ImageNet to extract spatial features while avoiding the complexity of deeper models such as ResNet-50. This hybrid strategy incorporates spatial feature extraction with temporal modelling to enhance predictive accuracy.

Deep learning techniques have deeply impacted many areas of artificial intelligence, such as natural language processing, speech recognition, and computer vision. One of Computer Vision's successful deep learning architectures is convolutional neural networks. CNN architecture takes advantage of translational invariance by extracting features through receptive fields and learning with weight sharing. Convolutional neural networks have become the go-to approach in various computer vision and image recognition problems. Accomplishments of supervised and unsupervised learning techniques in computer vision inspired [25] to consider the problem of encoding time series as images to enable machines to recognise and learn patterns and structures visually. Redefining time-series features as visual clues have raised attention in computer science and physics.

Researchers have built different network structures from time series for visual inspection or designing distance measures. Silva et al. [21] utilised the compression distance to expand the recurrence plot paradigm for time series classification. Another approach to construct a weighted adjacency matrix is extracting transition dynamics from the first-order Markov matrix; see [3]. Despite demonstrating distinct topological properties amongst different time series, these maps are unclear about how they relate to the original time series since they lack exact inverse operations.

3.1 VVIX Index

The CBOE VVIX functions as an index that assesses the volatility of volatility. Its core objective is to gauge the anticipated volatility of the VIX's 30-day forward price. The forward price, in this context, alludes to the hypothetical value of a VIX futures contract, which would expire in 30 days. The VVIX, while not precisely identical to the expected volatility of the VIX itself, shares a close relationship with the latter due to the tracking of nearby VIX futures to the VIX. The oscillation of VVIX between 2019 and 2023, as shown in Figure 1, illustrates its dynamic behaviour, with peaks corresponding to periods of heightened market uncertainty, such as 2020 pandemic. The methodology used to

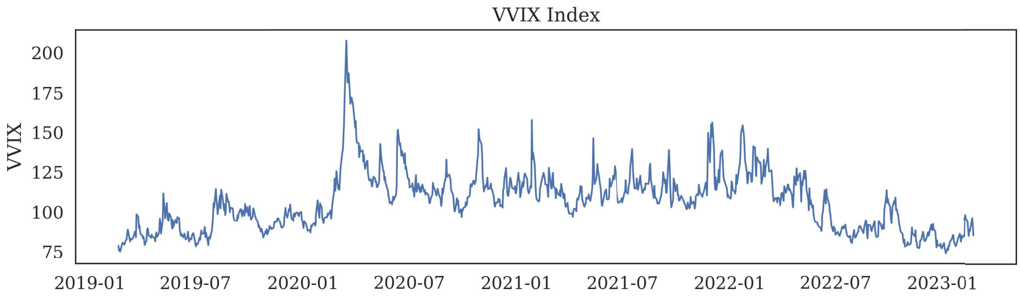

Figure 1: Volatility of Volatility Index

compute VVIX is analogous to that of VIX. It is obtained from the price of a portfolio of liquid at and out-of-money VIX options.

This portfolio can be utilised to manage volatility risk linked to exposures to VIX and to exploit the risk premium between the expected and realised volatility of VIX forward prices. Although the availability of VIX options with a 30-day expiration is generally limited, the computation of VVIX-like values can be achieved by utilising VIX options that expire at two distinct dates within a 30-day time frame. The determination of VVIX is subsequently performed through interpolation founded on said values. Following is the calculation of VVIX from VIX options prices at each expiration using the VIX formula. The VVIX is computed using Equation (1), which adapts the VIX formula to VIX options prices.

$$VVIX = \frac{2}{T}\sum \frac{\Delta K_i}{K_i^2} e^{RT}\, Q(k_i) - \frac{1}{T}\left[\frac{F}{K_0} - 1\right]^2 \tag{1}$$

where:

T Time to expiration

F Forward index level

K_0 First strike below the forward index level F

K_i Strike price of i^{th}out-of-money option

ΔK_i Interval between strike prices

R Risk-free rate to expiration

$Q(k_i)$ Spread midpoint for option with strike k_i

3.2 Time Series Image Encoding

Gramian Angular Fields (GAF) imaging proposed by [25, 24] is a technique to encode time series into images. As detailed in their paper, the goal of this transformation is to generate an image that will be part of the time series. Images will be constructed for each time period t as an RGB image where each channel is a feature matrix. To generate GAF images, time series observations require to be re-scaled. Let $X = \{x_1, x_2, \cdots, x_n\}$ be the considered time-series with n observations, re-scaling to the $[-1, 1]$ interval is done using the following equation:

$$\tilde{x_i} = \frac{(x_i - max(X)) + (x_i - min(X))}{max(X) - min(X)} \tag{2}$$

Hence, the re-scaled series is denoted by $\tilde{X}$. The re-scaled time series is transformed into a polar coordinates system by calculating the angular cosine of every single component of the scaled time series:

$$\tilde{X} = \{\tilde{x_1}, \tilde{x_2}, \cdots, \tilde{x_n}\} \tag{3}$$

$$\begin{cases} \phi_i = \arccos(\tilde{x}_i), & -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X} \\ r_i = \dfrac{i}{N}, & i \in N \end{cases} \tag{4}$$

Here, i is the time stamp and N is a constant factor to regularise the span of the polar coordinate system. r represents the angular relationship between data points in the time series. Contrary to the Cartesian coordinates, polar coordinates preserve temporal relations. In the GAF transformation, the time series $X = \{x_1, x_2, \ldots, x_n\}$ is first normalised to the interval $[-1, 1]$, resulting in $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n\}$, using Equation (2). The normalised values are then converted to polar coordinates, where each element is mapped to an angle $\phi_i = \arccos(\tilde{x}_i)$. In this representation, $\tilde{x}_i$ corresponds to $\cos(\phi_i)$, and $\sin(\phi_i) = \sqrt{1 - \tilde{x}_i^2}$ is computed element-wise. These trigonometric terms are used in the Gramian Angular Summation Field (GASF) and Gramian Angular Difference Field (GADF) as shown in Equations (5) and (6). The resulting matrices ingrain both the magnitude and relative temporal order of the original time series, thereby preserving temporal dependencies in the image representation.

Gramian Angular Summation Field (GASF) and Gramian Angular Difference Field (GADF) can be obtained by calculating the sum or difference between the points of the time series. I in Equations 5 and 6 is the unit row vector $[1, 1, \cdots, 1]$.

$$GASF = [\cos(\phi_i - \phi_j)] = \tilde{X^{\top}} \cdot \tilde{X} - \sqrt{I - \tilde{X^2}}^{\top} \cdot \sqrt{I - \tilde{X^2}} \tag{5}$$

$$GADF = [\sin(\phi_i - \phi_j)] = \sqrt{I - \tilde{X^2}}^{\top} \cdot \tilde{X} - \tilde{X}^{\top} \cdot \sqrt{I - \tilde{X^2}} \tag{6}$$

3.3 Markov Transition Field

Quantile bins Q of a given time series X are identified, and then each is assigned x_i to its corresponding bin q_j ($j \in [1, Q]$). Therefore, a $Q \times Q$ weighted adjacency matrix W is constructed by counting transitions between quantile bins in the same manner as the first-order Markov chain along the time axis. $\omega_{i,j}$ is provided by the frequency with which a point in quantile q_j is followed by a point in q_i. Subsequent to normalisation $\Sigma j \omega_{ij} = 1, \forall i \in [1, Q]$. Given W as the Markov transition matrix, the Markov Transition Field is defined as:

$$\begin{aligned} M &= \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ \vdots & \ddots & \vdots & \\ M_{n1} & M_{n2} & \cdots & M_{nn} \end{bmatrix} \\ &= \begin{bmatrix} w_{ij|x_1 \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_1 \in q_i, x_n \in q_j} \\ w_{ij|x_2 \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_2 \in q_i, x_n \in q_j} \\ \vdots & \ddots & \vdots \\ w_{ij|x_n \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_n \in q_i, x_n \in q_j} \end{bmatrix} \end{aligned} \tag{7}$$

The M_{ij} value in the transition probability matrix represents the probability of transitioning from state q_i to state q_j in one time step.

3.4 Long Short-Term Memory

The Long Short-Term Memory (LSTM) neural network [9] is an extension of Recurrent Neural Networks (RNNs) designed to address the vanishing gradient problem. It achieves this by utilising interconnected nodes within the same hidden layer. At each time step, the input to the hidden layer comprises not only the output from the previous layer but also the output of the same hidden layer from the preceding time step. LSTM models consist of four key components: memory units, input gate, output gate, and forget gate. The memory unit is responsible for storing the temporal state of the network. The input gate modulates the input activations of the cell, while the output gate performs a similar function for the cell's output. Finally, the forget gate adaptively resets the cell's memory.

LSTM has emerged as a widely used tool across various domains within the finance industry, including price prediction and trend analysis. Its ability to effectively capture long-term dependencies in sequential data renders it particularly valuable for these applications. By utilising historical financial data, LSTM models can analyse patterns and trends to make predictions about future prices and market movements. These predictions assist traders, investors, and financial institutions in making informed decisions regarding buying, selling, and timing their investments. Furthermore, LSTM's capability to process and analyse large volumes of data makes it well-suited for handling the complex and dynamic nature of financial markets. Consequently, LSTM has become an essential tool in financial forecasting, risk management,

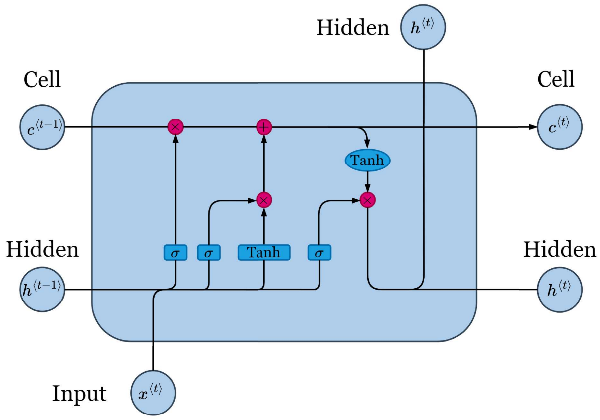

Figure 2: LSTM Memory Block Architecture

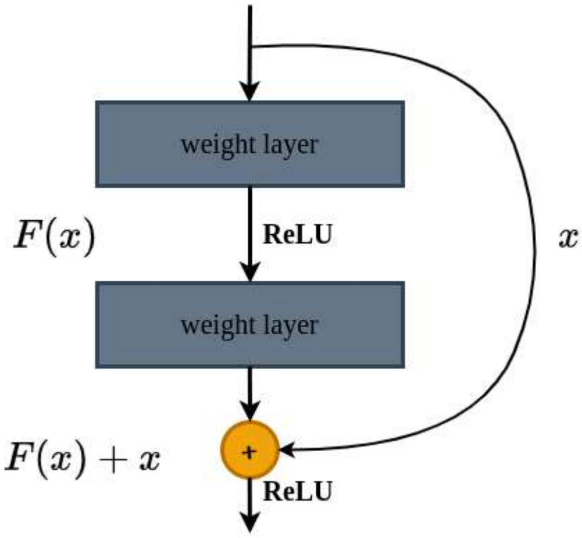

Figure 3: Residual Learning Block

and decision-making processes. In the following equations, i_j, f_j, and o_j represent the input gate, forget gate, and output gate activations, respectively. $\hat{c}_j$ denotes the memory cell input activation, c_j is the cell state, and h_j is the hidden state output at time step j. The design of LSTM is based on the following equations:

$$
\begin{aligned}
i_j &= \sigma(W_i \cdot [x_j, h_{j-1}] + b_i) \\
f_j &= \sigma(W_f \cdot [x_j, h_{j-1}] + b_f) \\
o_j &= \sigma(W_o \cdot [x_j, h_{j-1}] + b_0) \\
\hat{c}_j &= \tanh(W_c \cdot [x_j, h_{j-1}] + b_c) \\
c_j &= f_j \odot c_{j-1} + i_j \odot \hat{c}_j \\
h_j &= o_j \odot \tanh(c_j)
\end{aligned} \tag{8}
$$

where x_j, i_j, f_j, o_j, and $\hat{c}_j$ represent the input vector, input gate's activation, forget gate's activation, output gate's activation, and memory cell input activation, respectively. c_j is the cell state vector at step j, while h_j is the LSTM unit output vector. The symbol $\odot$ denotes an element-wise dot product. W and b are the weight and bias matrices to be optimised, and σ is the sigmoid function.

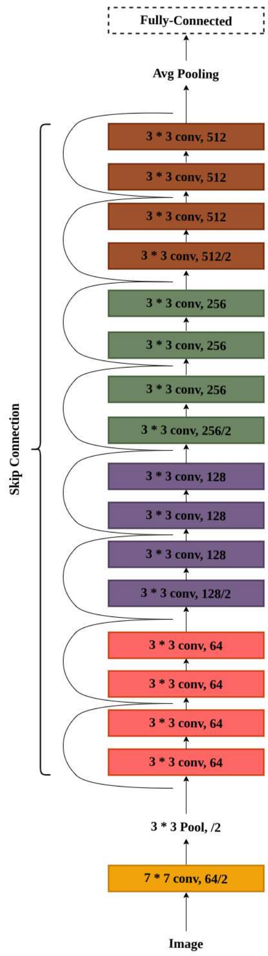

Figure 4: ResNet-18 Architecture

3.5 Residual Networks

Residual networks have been introduced by [7] to address the challenge of vanishing gradients in deep neural networks. ResNets contain residual blocks, which enable the formation of deeper networks by introducing skip or shortcut connections. Residual blocks illustrated in Figure 3 are the building blocks of ResNets. Within each block, there exist a series of convolutional layers followed by an element-wise operation called a skip connection. This connection bypasses one or more convolutional layers. The sole purpose of the skip connection is to propagate the input to the block's output, creating a residual path. The residual path contributes to learning the residual mappings, thus capturing the difference between the input and output. Identity mapping is a common form of residual block although various configurations are available. The skip connection is crucial to the training of deep neural networks. It allows the gradients to easily flow from later layers to earlier ones, making it easier for the network to learn hence facilitating the training of deeper architectures without performance degradation. ResNet-18 is a specific variant of residual networks, designed for image classification tasks which consist of 18 layers.The ResNet-18 architecture comprises multiple stages and layers.

Initially, the input of a fixed-sized image, typically 224×224 pixels undergoes a convolutional layer with 64 filters with a spatial size of 7×7, following a batch normalisation layer and Rectified Linear Unit (ReLU) activation resulting in dimension reduction of the input image. Subsequently, a 3×3 max-pooling layer is employed which reduces the spatial dimension by a factor of two. From here onwards there exist four sets of layers each containing two residual blocks with convolutional layers with respective 64, 128, 256, and 512 filters.

Following a global average pooling layer is applied to aggregate spatial information across the entire feature map leading to a fixed-size representation. Ultimately, the output of the global average pooling layer is passed to a fully connected layer. As for ResNet-18, the fully connected layer consists of 1000 units corresponding to the number of classes in the ImageNet dataset. The final classification probabilities are obtained through the application of a soft-max activation function at the exit of this layer.

4 Experiments and Results

4.1 Dataset Description

The dataset used in this study comprises VVIX time series data spanning from 2006 to 2023, obtained from the Chicago Board Options Exchange website. The volatility of VIX (VVIX) measures the expected volatility for the S&P 500 options. It is a crucial indicator for market participants, including investors and traders, as it reflects the market's expectations of future volatility. High VVIX values signify increased market uncertainty and potential price fluctuations, while low VVIX values suggest market stability and reduced volatility. Accurate VVIX predictions are of great importance in making informed decisions related to risk management, portfolio allocation, and trading strategies. To ensure data quality, we handled outliers and missing values. Outliers were defined as values that exceeded three standard deviations from the mean. These outliers were capped at the 99^{th} percentile to reduce their impact on model training. CBOE is continuous in nature; thus, missing values were minimal. However, we used linear interpolation for imputation. This method helped maintain temporal continuity and ensured robustness in the time series analysis.

4.2 Proposed Hybrid Model

The proposed ResNet-LSTM model combines an image-based pathway for spatial feature extraction with a recurrent pathway to capture temporal dependencies (Figure 5).

- **Image Pathway**: Converts time series data into three distinct visual representations utilising the Gramian Angular Difference Field (GADF), the Gramian Angular Summation Field (GASF), and the Markov Transition Field (MTF). These visualizations elucidate the intrinsic temporal structures and interdependencies, thereby converting the data into spatial formats amenable to efficient processing by convolutional neural networks (CNNs). In our study, we utilise a pre-trained ResNet-18 model to extract high-level features, subsequently incorporating three additional convolutional layers to enhance the spatial representations. Conventional approaches, such as feature engineering within long short-term memory networks (LSTMs), necessitate specialized domain knowledge to construct meaningful features. Conversely, our methodology capitalizes on deep convolutional neural networks to autonomously acquire hierarchical representations of features. In comparison to Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, which are predicated on parametric assumptions regarding the dynamics of volatility, our image-based pathway possesses the capability to capture intricate, nonlinear interdependencies and latent patterns.
- **Recurrent Pathway**: Employs two Long Short-Term Memory (LSTM) layers to analyse sequences of VVIX values, thereby effectively capturing temporal dependencies and emerging trends. We conduct experiments utilizing three distinct window lengths—10, 21, and 41 days—to appropriately model short-term, mid-term, and long-term dynamics. Although the LSTM architecture is capable of modelling sequential dependencies independently, our hybrid methodology enhances this temporal information with spatial patterns that are derived from the image-based pathway, ultimately facilitating a more comprehensive understanding of volatility behaviour.

Table 1: Image Pathway kernel dimensions.

Image Pathway	Kernel
Convolutional Layer 1	128
Convolutional Layer 2	64
Convolutional Layer 3	32

4.3 Model Training

The model's parameters were optimised using the mean squared error (MSE) loss function and the Adam optimiser. To enhance the model's performance during training and avoid convergence to sub-optimal solutions, a decaying learning rate was applied to reduce the initial learning rate 0.01 if the loss plateaued after three consecutive epochs. Experimented with sizes of 16, 32, and 64 a batch size of 32 provided a balance between convergence speed and model stability. We

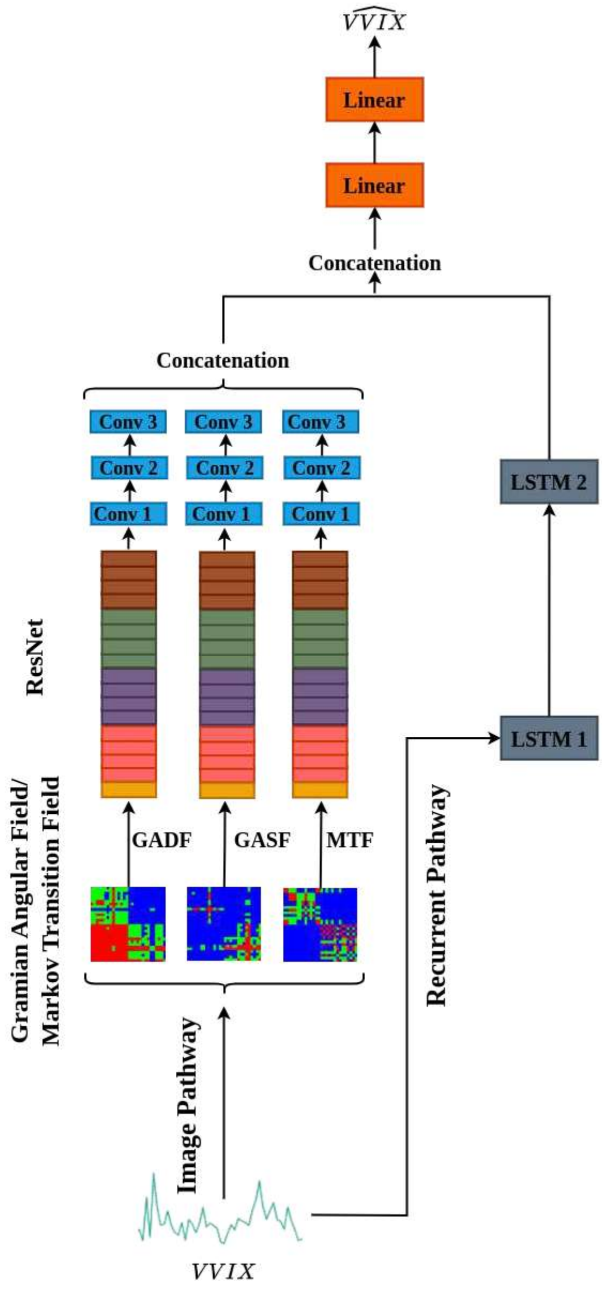

Figure 5: Proposed ResNet-LSTM Model

Table 2: Recurrent Pathway LSTM dimensions.

Recurrent Pathway	Input Dim	Hidden Dim
LSTM Layer 1	1	32
LSTM Layer 2	32	32

implemented early stopping based on validation loss to prevent over-fitting. Compared to simpler baselines, the hybrid model required more computational time but achieved substantial performance gains. The CNN model trained fastest but often failed to capture long-term dependencies, while the GARCH model trained efficiently but struggled to adapt to complex, non-linear patterns.

4.4 Baseline Models and Comparative Analysis

We formulated three foundational models—LSTM, CNN, and GARCH—to assess the efficacy of our proposed hybrid methodology in predicting VVIX. The Long Short-Term Memory (LSTM) model represents a category of recurrent neural networks that proficiently addresses temporal dependencies by preserving pertinent historical data through memory cells. This attribute is imperative for time series prediction, as it facilitates the identification of patterns across diverse temporal intervals.

The Convolutional Neural Network (CNN), commonly used for image processing, was applied to the VVIX time series data to capture localised patterns and short-term dependencies. By employing convolutional layers, the CNN architecture acquired characteristics that are conventionally feature-engineered in conventional approaches, offering an effective methodology to retrieve information. Nonetheless, due to the fact that CNNs predominantly emphasise local interactions, they may overlook longer-term dependencies that are pivotal in financial time series forecasting, constraining their predictive efficacy in comparison to more intricate architectures.

The GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model, a well-known econometric approach for volatility modelling, served as our classical baseline. GARCH is particularly effective in capturing volatility clustering and adapting to changes in variance over time. Its interpretability and computational efficiency make it a popular choice for financial applications. However, GARCH models operate under strict parametric assumptions and often struggle to adapt to nonlinear, complex volatility patterns present in financial markets. Thus, while GARCH models excel in stable market conditions, they may underperform in more dynamic environments.

We used paired t-tests to compare model performance across 100 test predictions for each forecasting horizon (short-term, mid-term, long-term). These predictions came from an 80/20 split of the VVIX dataset into training and testing sets. Before running the t-tests, we used the Shapiro–Wilk test to check that the differences in prediction errors were normally distributed, confirming that the t-test was suitable.

5 Discussion of Findings

The comparative results in 3 highlight how well the suggested ResNet-LSTM hybrid model performs when compared to baseline methods, proving its capacity to accurately forecast VVIX over a range of time horizons. Metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and coefficient of determination (R^2) show that the hybrid model outperforms the GARCH, CNN, and LSTM models in terms of explanatory power and error rates. The statistical significance of these improvements is confirmed by its consistently lower p-values, which are obtained from paired t-tests, especially in the short- and mid-term horizons. The paired t-tests employed a sample size of approximately 4,000 daily observations (2006–2023), and standard statistical techniques were employed to validate the assumption of normality.

Although the forecasting errors of the ResNet-LSTM model are lower than those of baseline models, sudden market shocks or macroeconomic events may still have an effect. High VVIX values, for instance, which usually coincide with shifts in interest rates, financial crises, or world events, can introduce noise that reduces the accuracy of the model, particularly for long-term horizons. In their investigation of volatility spillovers, Zeng et al. [29] noted that these errors could be a result of the model's vulnerability to abrupt changes in market sentiment or external factors not present in the input data, suggesting the potential benefit of including macroeconomic indicators to increase robustness.

The observed drop in performance for longer time horizons for all models is consistent with the inherent difficulties of long-term financial volatility forecasting. Long-term forecasts are more vulnerable to external shocks and market uncertainties, which increase noise and decrease predictability. This problem is somewhat mitigated by the hybrid

Table 3: Performance Comparison by MSE, MAE, and R^2 Across Time Horizons

(a) Mean Squared Error (MSE)

Model	MSE Short	MSE Mid	MSE Long	p-value Short	p-value Mid	p-value Long	95% CI Short	95% CI Mid	95% CI Long
GARCH	0.045	0.055	0.087	0.015	0.020	0.035	[0.010, 0.030]	[0.015, 0.035]	[0.030, 0.060]
CNN	0.039	0.047	0.060	0.012	0.018	0.025	[0.008, 0.025]	[0.012, 0.030]	[0.015, 0.040]
LSTM	0.033	0.035	0.050	0.010	0.015	0.025	[0.005, 0.020]	[0.008, 0.025]	[0.010, 0.035]
ResNet-LSTM	**0.023**	**0.020**	**0.042**	**0.005**	**0.008**	**0.015**	**[0.003, 0.015]**	**[0.004, 0.016]**	**[0.008, 0.030]**

(b) Mean Absolute Error (MAE)

Model	MAE Short	MAE Mid	MAE Long	p-value Short	p-value Mid	p-value Long	95% CI Short	95% CI Mid	95% CI Long
GARCH	0.110	0.130	0.160	0.020	0.025	0.035	[0.020, 0.050]	[0.030, 0.060]	[0.040, 0.080]
CNN	0.102	0.115	0.125	0.020	0.030	0.040	[0.015, 0.040]	[0.020, 0.050]	[0.025, 0.060]
LSTM	0.095	0.100	0.110	0.018	0.025	0.035	[0.010, 0.035]	[0.015, 0.040]	[0.020, 0.050]
ResNet-LSTM	**0.075**	**0.070**	**0.090**	**0.008**	**0.010**	**0.020**	**[0.005, 0.025]**	**[0.008, 0.030]**	**[0.015, 0.040]**

(c) Coefficient of Determination (R^2)

Model	R^2 Short	R^2 Mid	R^2 Long	p-value Short	p-value Mid	p-value Long
GARCH	0.75	0.70	0.65	0.010	0.025	0.035
CNN	0.78	0.76	0.72	0.020	0.025	0.030
LSTM	0.82	0.81	0.78	0.015	0.020	0.025
ResNet-LSTM	**0.91**	**0.93**	**0.85**	**0.003**	**0.004**	**0.008**

model's adaptability across different window lengths (10, 21, and 41 days), which allows it to maintain predictive accuracy above the baselines. This capability emphasises how crucial it is to combine temporal dynamics and spatially rich representations in order to handle the complexity of financial markets.

5.1 Economic Implications of VVIX Predictions

The predictive power of the ResNet-LSTM model for VVIX has significant economic implications for regulators and market participants alike. Accurate VVIX forecasts can show shifts in market participants' expectations for future volatility, enabling investors to proactively modify their portfolios and potentially minimising losses during erratic periods. For instance, a rising VVIX may indicate anticipated market stress, prompting traders to hedge positions or exploit options mispricing, according to Zeng et al. [29]. By providing early warning indicators of systemic risks, such as those caused by extreme market conditions, the model's ability to detect nonlinear patterns may also assist financial regulators in creating proactive stabilisation policies. Additionally, Wu et al. [26] demonstrated how external factors, such as green finance policies, can affect market dynamics, indicating that adding such factors could improve the model's relevance to actual investment and regulatory decisions.

Although computational trade-offs were not specifically examined, the hybrid model's more complex architecture suggests that it will require more time and resources to train than stand-alone LSTMs or more straightforward models like GARCH. Nonetheless, the proven performance improvements make this extra computational work worthwhile, especially for applications where accuracy is crucial. Future research might concentrate on optimising the hybrid approach's computational effectiveness for real-time deployment.

Lastly, it is unknown how resilient the suggested model would be in the face of severe market circumstances, like financial crises or spikes in volatility brought on by pandemics. Further testing under stress scenarios is required to ensure generalisability, even though its capacity to capture nonlinear patterns and temporal dependencies suggests potential resilience. Notwithstanding these drawbacks, the hybrid model sets a new standard for VVIX prediction and provides a strong, statistically supported framework for dealing with volatility forecasting difficulties.

6 Conclusion

The importance of this study lies in its objective to fill the gap in the literature regarding VVIX forecasting, building on previous research in volatility modelling [15, 6] and the need for accurate volatility predictions to help with risk management as well as support informed investment decisions in financial markets. In this study, we proposed and evaluated a hybrid ResNet-LSTM model for predicting VVIX values by integrating spatial feature extraction through ResNet with temporal dependency modelling via LSTMs. The VVIX dataset, spanning from 2006 to 2023, served as the foundation for our investigation into accurately forecasting market volatility. This research addresses the critical need for reliable VVIX predictions, enabling market participants, such as investors and traders, to make informed decisions regarding risk management, portfolio allocation, and trading strategies. The proposed model leverages the strengths of both image and recurrent pathways, capturing spatial patterns and temporal dependencies present in the VVIX time series data.

In the image pathway, spatial features were extracted by converting raw VVIX data into images using techniques such as Gramian Angular Difference Field, Gramian Angular Summation Field, and Markov Transition Field. These representations were processed through a pre-trained ResNet-18 to harness valuable feature representations while optimizing computational efficiency. Meanwhile, the recurrent pathway utilized LSTMs to model sequential information, discerning underlying trends and temporal dynamics in the VVIX time series. By combining these pathways, the hybrid model demonstrated a robust capability to capture complex, non-linear patterns, resulting in improved predictive performance over baseline approaches such as GARCH, CNN, and standalone LSTM models.

The experimental results highlighted superior predictive accuracy across short-term, mid-term, and long-term horizons, as evidenced by better MSE, MAE, and R^2 metrics compared to baseline methods. Notably, the model exhibited statistically significant improvements in performance, confirmed through paired t-tests, which underscored its ability to integrate spatial and temporal information effectively. However, the model's declining accuracy for long-term predictions reflects the inherent challenges of increased market uncertainty and noise in financial time series forecasting.

Overall, the hybrid ResNet-LSTM model showcased potential advantages for accurate VVIX value predictions, offering practical applications in risk management, portfolio allocation, and trading strategies. Its ability to adapt to diverse time-window lengths emphasizes its suitability for varying prediction horizons, although computational trade-offs and robustness under extreme market conditions warrant further exploration. The limitations of our study include the lack of explicit evaluation under stress scenarios and the need for optimisation strategies to enhance real-time deployment. To further advance the predictive capability of the proposed hybrid model, several directions for future work can be considered:

- Investigating the model's performance under extreme market conditions, such as financial crises or sudden volatility spikes, to assess its robustness and reliability during periods of heightened uncertainty.
- Incorporating additional data sources, such as macroeconomic indicators, sentiment analysis of financial news, or options market data, to enrich the model's input features and further enhance predictive accuracy.
- Evaluating the model's transferability to other volatility indices or asset classes and examining its generalisability across different market regimes and economic cycles.

In conclusion, this study contributes to the growing body of knowledge in financial market forecasting, specifically in the context of VVIX value prediction. The proposed hybrid model represents a significant step forward in volatility forecasting methodologies, providing meaningful insights for market participants and opening new avenues for future research. As the financial landscape evolves, the ResNet-LSTM hybrid model offers a robust framework for advancing predictive analytics, helping stakeholders navigate the complexities of modern financial markets.

References

[1] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[2] Timothy J. Brailsford and Robert W. Faff. An evaluation of volatility forecasting techniques. *Journal of Banking & Finance*, 20(3):419–438, 1996. doi: https://doi.org/10.1016/0378-4266(95)00015-1.

[3] Andriana S. L. O. Campanharo, M. Irmak Sirer, R. Dean Malmgren, Fernando Manuel Ramos, and Luis A. Nunes Amaral. Duality between time series and networks. *PLoS ONE*, 6, 2011.

[4] Rodolfo Carneiro Cavalcante, Rodrigo C. Brasileiro, Victor L. F. Souza, Jarley Palmeira Nóbrega, and Adriano Oliveira. Computational intelligence and financial markets: A survey and future directions. *Expert Syst. Appl.*, 55: 194–211, 2016.

[5] Matthew F. Dixon, Diego Klabjan, and Jin Hoon Bang. Classification-based financial markets prediction using deep neural networks. *Big Data & Innovative Financial Technologies Research Paper Series*, 2016.

[6] Thomas G. Fischer and Christopher Krauss. Deep learning with long short-term memory networks for financial market predictions. *Eur. J. Oper. Res.*, 270:654–669, 2017.

[7] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2015.

[8] Steven Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

[9] S. Hochreiter and J. Schmidhuber. LSTM can solve hard long time lag problems. In *NIPS*, 1996.

[10] P. O. J. Kelliher, D Wilmot, Jivesh Vij, and Paul Klumpes. A common risk classification system for the actuarial profession. *British Actuarial Journal*, 18:91–121, 2013.

[11] Kenneth F. Kroner, Kevin P. Kneafsey, and Stijn Claessens. Forecasting volatility in commodity markets. *Journal of Forecasting*, 14(2):77–95, 1995. doi: https://doi.org/10.1002/for.3980140202.

[12] Akib Mashrur, Wei Luo, Nayyar A. Zaidi, and Antonio Robles-Kelly. Machine learning for financial risk management: A survey. *IEEE Access*, 8:203203–203223, 2020. doi: 10.1109/ACCESS.2020.3036322.

[13] Felipe D. Paiva, Rodrigo Tomás Nogueira Cardoso, Gustavo P. Hanaoka, and Wendel Moreira Duarte. Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Syst. Appl.*, 115:635–655, 2019.

[14] Andrew J. Patton. Volatility forecast comparison using imperfect volatility proxies. *ERN: Time-Series Models (Single) (Topic)*, 2006.

[15] Ser-Huang Poon and Clive William John Granger. Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41:478–539, 2003.

[16] Eduardo Ramos-Pérez, Pablo J. Alonso-González, and José Javier Núñez-Velázquez. Multi-transformer: A new neural network-based architecture for forecasting s&p volatility. *ArXiv*, abs/2109.12621, 2021.

[17] Ahoora Rostamian. *Applications of Deep Learning Models in Financial Forecasting*. PhD thesis, University of Essex, 2024.

[18] Ahoora Rostamian and John G. O'Hara. Event prediction within directional change framework using a CNN-LSTM model. *Neural Computing and Applications*, 34:17193 – 17205, 2022.

[19] Alaa El. Sagheer and Mostafa Kotb. Time series forecasting of petroleum production using deep lstm recurrent networks. *Neurocomputing*, 323:203–213, 2019.

[20] Sreelekshmy Selvin, R. Vinayakumar, E. A. Gopalakrishnan, Vijay Krishna Menon, and K. P. Soman. Stock price prediction using lstm, rnn and cnn-sliding window model. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1643–1647, 2017.

[21] Diego Furtado Silva, Vinicius M. A. Souza, and Gustavo E. A. P. A. Batista. Time series classification using compression distance of recurrence plots. *2013 IEEE 13th International Conference on Data Mining*, pages 687–696, 2013.

[22] Stefan Trück and Kevin Liang. Modelling and forecasting volatility in the gold market. *The International Journal of Banking and Finance*, 9:48–80, 2012.

[23] Haishuai Wang, Zhicheng Cui, Yixin Chen, Michael S. Avidan, Arbi Ben Abdallah, and Alex Kronzer. Predicting hospital readmission via cost-sensitive deep learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15:1968–1978, 2018.

[24] Zhiguang Wang and Tim Oates. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. 2014. URL `https://api.semanticscholar.org/CorpusID:16409971`.

[25] Zhiguang Wang and Tim Oates. Imaging time-series to improve classification and imputation. *ArXiv*, abs/1506.00327, 2015.

[26] Yongjie Wu, Liuru Hou, Yue Yuan, Shenglin Ma, and Hongjun Zeng. Green credit policy's influence on construction firm esg performance: a difference in differences estimation. *Journal of Asian Architecture and Building Engineering*, 0(0):1–13, 2025. doi: 10.1080/13467581.2025.2480752.

[27] Ruoxuan Xiong, Eric P. Nichols, and Yuan Shen. Deep learning stock volatility with google domestic trends. *arXiv: Computational Finance*, 2015.

[28] Xin Zang, Jun Ni, Jing-Zhi Huang, and Lan Wu. Double-jump diffusion model for vix: evidence from vvix. *Quantitative Finance*, 17(2):227–240, 2017. doi: 10.1080/14697688.2016.1159318.

[29] Hongjun Zeng, Mohammad Zoynul Abedin, Brian Lucey, and Shenglin Ma. Tail risk contagion and multiscale spillovers in the green finance index and large us technology stocks. *International Review of Financial Analysis*, 97:103865, 2025. ISSN 1057-5219. doi: https://doi.org/10.1016/j.irfa.2024.103865.

[30] Gongquan Zhang, Shenglin Ma, Mingxing Zheng, Cheng Li, Fangrong Chang, and Fangbing Zhang. Impact of digitization and artificial intelligence on carbon emissions considering variable interaction and heterogeneity: An interpretable deep learning modeling framework. *Sustainable Cities and Society*, 125:106333, 2025. ISSN 2210-6707. doi: https://doi.org/10.1016/j.scs.2025.106333.