Detection Framework for Financial Risk with R-VAE

Ricatti Luca SM3800083
Tavano Matteo SM3800057

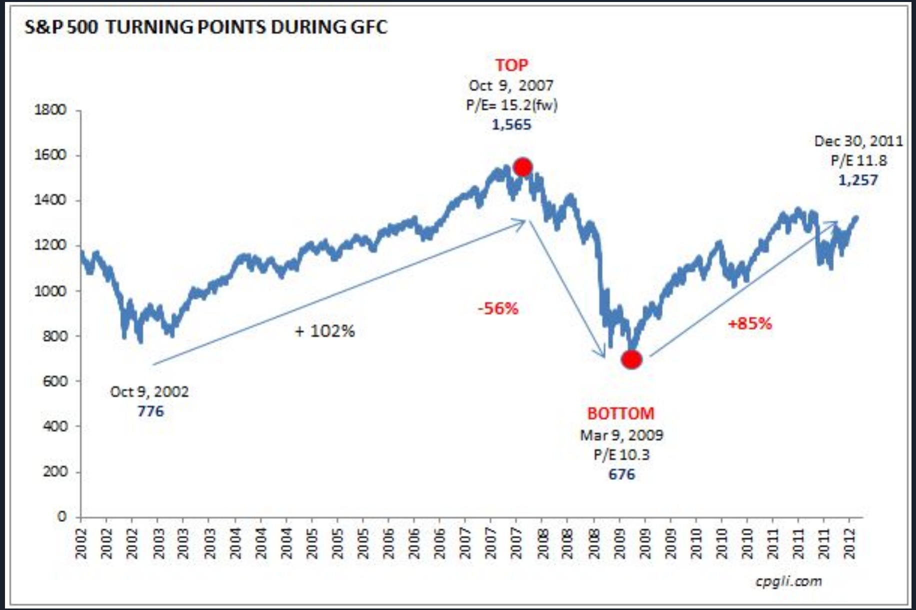
The Tulip Mania: The First Recorded Financial Bubble

- Tulip bulbs in the Netherlands reached exorbitant prices in the 1630s: One bulb of the Semper Augustus variety could cost as much as a luxury house in Amsterdam.
- Prices soared based on speculative futures contracts, unregulated promises to buy bulbs at a later date.
- The bubble burst in 1637, causing prices to collapse by over 90%, leaving many investors bankrupt and triggering a financial panic across the Dutch economy.
- Early example of a speculative bubble driven by irrational exuberance and new and unregulated financial instruments.

No More Crises Thanks to Our Models

“I would like to say to Milton [Friedman] and Anna [Schwartz]: Regarding the Great Depression [1929]. You're right, we did it. We're very sorry. But thanks to you, ***we won't do it again****.”* (Ben Bernanke, 2002)

- Friedman explained the causes of 1929 Great Depression, influencing economic thought for decades.
- By the 2000s, central banks relied on advanced mathematical models (like DSGE) that seemed powerful enough to predict and prevent financial crises by adjusting interest rates and using monetary policy tools.



What DSGE Models Missed

Dynamic Stochastic General Equilibrium Models:

- **Strong priors and simplified structure**: Assumed rational expectations and representative agents → limited model capacity and poor generalization.
- **Linearization**: Approximated dynamics near equilibrium → unable to model nonlinear transitions like crashes or regime shifts.
- **Low-dimensional latent space**: Ignored complex interactions in the financial sector (e.g. leverage, credit contagion).
- **Bayesian estimation with restrictive priors** limits posterior flexibility → underfit real-world uncertainty and rare events.
- **No learning from <u>high-frequency data</u>**: Operated on aggregated, low-frequency macro data: <u>financial market data (e.g. stock prices, credit spreads) were largely ignored</u>, making it impossible to capture early warning signals of systemic risk.

Data: Yahoo Finance

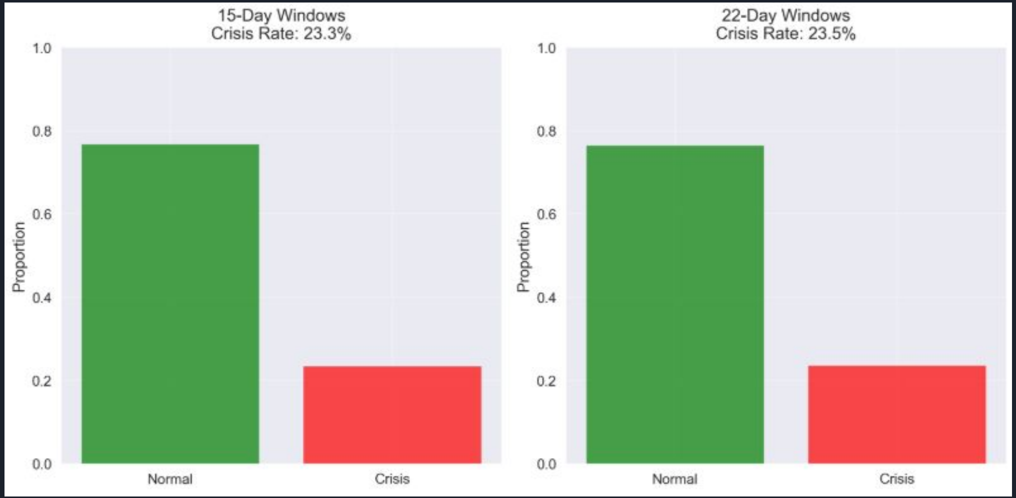
- Comprehensive set of financial time-series data, sourced from **Yahoo Finance** via **yfinance** Python library.
- The Dataset consists of <u>**daily closing**</u> prices for wide array of financial instruments, spanning a long historical period (from **1990-01-01**), to cover multiple economic cycles.
- Selection is designed to provide a **holistic view** of global markets dynamics and includes:
- **Major Market Indices**: *S&P 500*, *NASDAQ*, *Dow Jones.*
- **Volatility Measures**: *VIX.*
- **Sector specific ETFs**: covers all major sectors such as *Technology*, *Financials* and *Energy*.
- **Fixed Income**: *US Treasury bond ETFs.*
- **Commodities and Currencies**: *Key assets as Gold*, *Oil* and *US Dollar*.

Data Processing and Feature Engineering

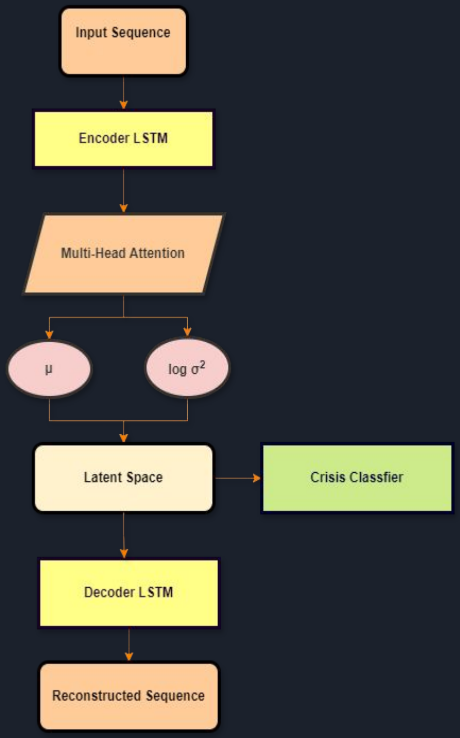
- **Crisis Labelling**: Major crisis has been labeled. In particular, we covered **6** periods of global crisis: **Gulf War Recession, Asian Financial Crisis, Dot-Com Crash, Global Financial Crisis, COVID-19 Crisis and Inflation Bear Market**.
- Preprocessing of the sequences to train and evaluate or **R-VAE model**.
- Calculating the **daily return** and engineering additional features such as **rolling volatility**, **sector divergence**, and **yield curve differentials**.
- **Data Cleaning**: clapping extreme values and keeping the **20 most informative features**, selected based on **historical volatility**.
- **Feature scaling** using **Robust Scaler normalization** (reduce sensitivity to outliers).
- Building a f**ixed-length sliding windows** and aligns each sequence with a binary crisis label by checking temporal overlap.

Sequences and Crisis Distributions

- Sequences labeled as “Crisis” vs “Normal” sequences distribution.
- This indicates that longer temporal windows tend to **capture slightly more crisis periods**, with crisis events that represents less than **25%** of total sequences.
- **Crucial consideration** when designing and training predictive models.



R-VAE Architecture



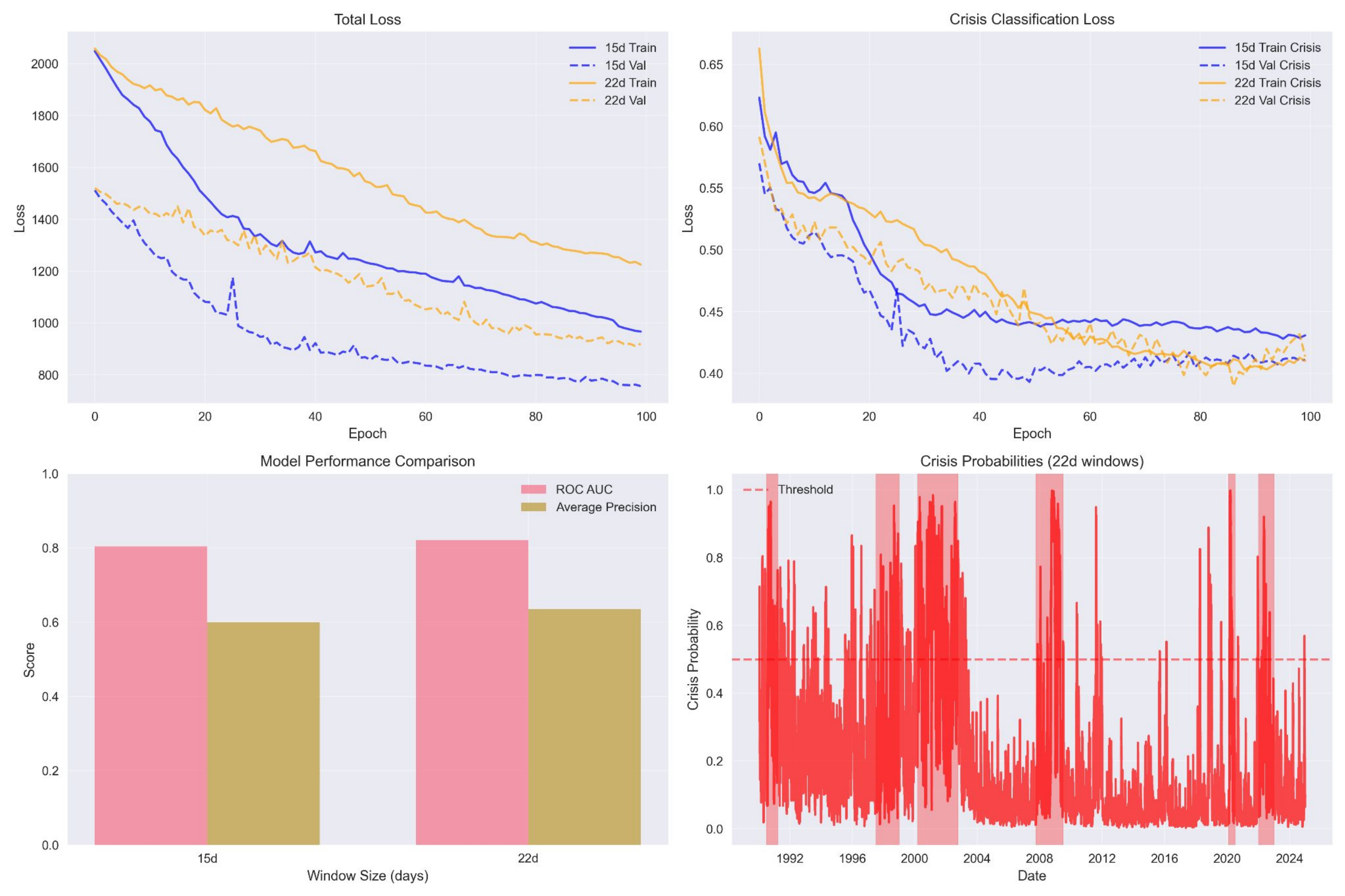
- **Recurrent Variational Autoencoder** for financial time series.
- **Encoder**: bi-directional **LSTM** with an additional **self-attention** layer to capture both temporal and contextual dependencies.
- **Mean-pooling** operation and mapping to the **latent space** via **μ** and **log**$\boldsymbol{\sigma}^2$.
- **Reparameterization trick**: latent vector **z** is sampled using this trick. A parallel **MLP head** takes this latent vector and predicts probability of a crisis.
- **Decoder**: Latent vector is projected to a hidden representation and expanded, then passed through a **unidirectional LSTM** to reconstruct the input sequence.
- **Multi-objective Loss function**:

$$\mathcal{L}_{total} = MSE + \beta * KL + \alpha * BCE$$

- Architecture suited for **crisis prediction**, **financial anomaly detection** and **sequence reconstruction**.

Training & Evaluation

- **Data Processing**: 80% training and 20% validation.
- **Training Loop** with **batch processing**, forward and backward passes with **gradient clipping** and multi-component loss calculation.
- **Hyper-parameters**: **100** epochs, **15** and **22** window sizes, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.
- **Validation**: model performance without gradient updates.
- **Optimizations**: **AdamW**, **LR scheduling** based on validation loss **plateau detection** and **early stopping**.
- **Evaluation metrics**: **ROC AUC** (discrimination ability), **Average Precision** (AP), classification report (**precision**, **recall**, **f1-score**) for classification task.
- **Results** stored in python dictionaries (anomaly scores, crisis probabilities, detection threshold, performance metrics).


Total Loss
15d Train
15d Val
22d Train
22d Val
Loss
Epoch
Crisis Classification Loss
15d Train Crisis
15d Val Crisis
22d Train Crisis
22d Val Crisis
Model Performance Comparison
ROC AUC
Average Precision
Score
Window Size (days)
15d
22d
Crisis Probabilities (22d windows)
Threshold
Crisis Probability
Date

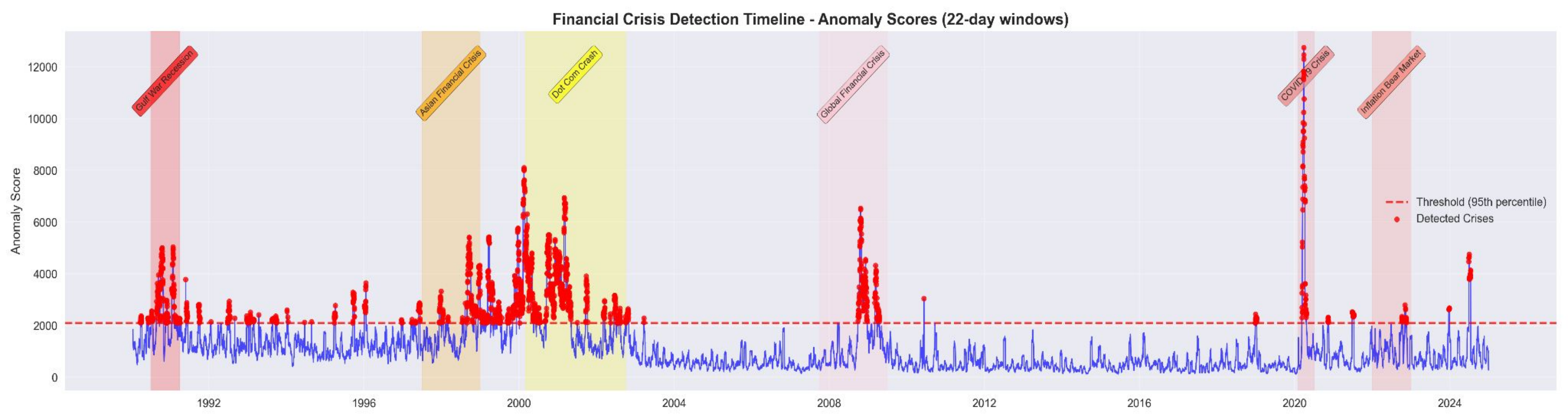
Results

Final **result summary** for the classification task of our models.
Both models achieved high discriminative power, with the best performing **22-days model**.

	ROC AUC	Average Precision
15-days model	0.804	0.600
22-days model	0.821	0.634

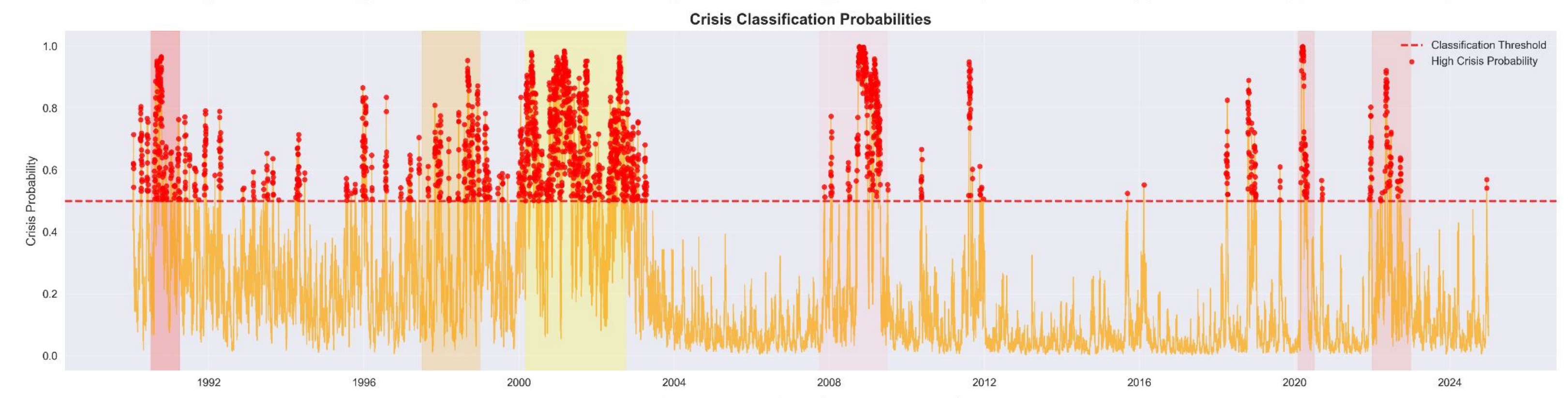
Anomaly Scores (Unsupervised Detection)

- **Raw reconstruction error** from the autoencoder.
- How much a given market sequence deviates from the “normal” patterns the model has learned.
- Significative spikes during major crises, red dots indicates where anomaly score exceeded the **95th percentile threshold**.



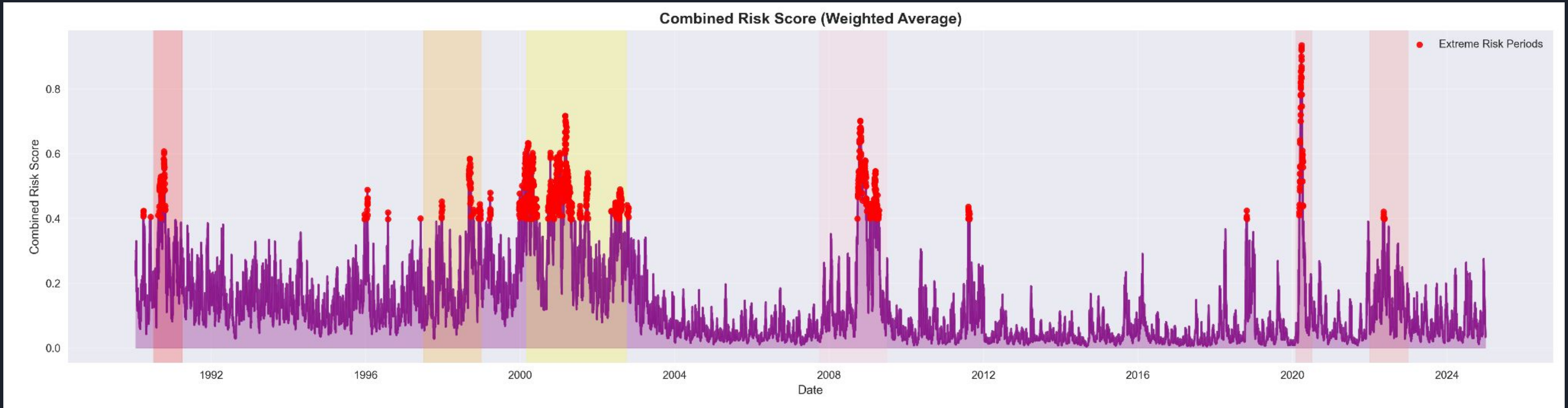
Crisis Classification Probabilities (Supervised Detection)

- Output from the model's **supervised classification head**.
- Direct prediction of probability that a given period is part of historical crisis.
- Much **cleaner** than anomaly score. Crisis probability are close to 1.0 during major crisis and remain low during calm periods.



Combined Risk Score (Hybrid View)

- **Weighted average** of both worlds: less noisy than anomaly score but still sensitive enough to capture emerging stress signals. Peaks align **almost perfectly** with all labeled historical crisis.
- **Normalized_anomaly** obtained using **min-max normalization** and applying:
- **Combined Risk** = 0.6 * Normalized_anomaly + 0.4 * crisis_probability



Conclusions - Qualitative analysis

PROs:

- High Reliability.
- Effective Hybrid Approach.
- Early warning capability.

CONs:

- Sensitivity to volatility (anomaly score).
- Potential for Minor False Positives.
- Variable Signal Intensity.

Thanks for listening!!! :)