

Article

Does ESG Predict Systemic Banking Crises? A Computational Economics Model of Early Warning Systems with Interpretable Multi-Variable LSTM based on Mixture Attention

Shu-Ling Lin and Xiao Jin * 

Department of Information and Finance Management, College of Management,
National Taipei University of Technology, Taipei 10608, Taiwan

* Correspondence: t107749406@ntut.edu.tw

Abstract: Systemic banking crises can be very damaging to economic development, and environmental, social, and governance (ESG) can also damage national finances, but there is no research on whether ESG affects systemic banking crises, and we fill this gap. We first employ Fisher scores (FS) to select features and then use an interpretable multivariate long-short-term memory (IMV-LSTM) model with focal loss (FL) to account for class imbalance to model an early warning system (EWS) that can predict up to one year in advance. This study finds that ESG influences the occurrence of systemic banking crises, with our early warning system predicting each crisis a year in advance.

Keywords: systemic banking crisis; early warning system; environmental, social, and governance (ESG); deep learning; LSTM

MSC: 62P05; 62P20; 68T07



Citation: Lin, S.-L.; Jin, X. Does ESG Predict Systemic Banking Crises? A Computational Economics Model of Early Warning Systems with Interpretable Multi-Variable LSTM based on Mixture Attention. *Mathematics* **2023**, *11*, 410. <https://doi.org/10.3390/math11020410>

Academic Editors: Yi-Hsien Wang and Chien-Ming Huang

Received: 26 December 2022

Revised: 9 January 2023

Accepted: 10 January 2023

Published: 12 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Systemic banking crises are very destructive. Whether it is an emerging country or a developed country, when faced with a systemic banking crisis, government departments are required to actively respond [1]. Failure to issue early warning signals in a timely and accurate manner and activate macro-prudential policy tools may bring huge losses to a country's finance, the economy, and even society as a whole [2,3]. The 2022 Nobel Prize in Economics was awarded to Ben S. Bernanke, Douglas W. Diamond, and Philip H. Dybvig for their outstanding contributions to the study of banking and financial crises [4]. Therefore, an early warning system that can accurately predict systemic banking crises is crucial.

This study adopts the systemic banking crises database II in Laeven and Valencia [3], and following them we defined a banking crisis: "Signs of a serious financial crisis in the banking system, and important banking policy intervention measures are taken in response to major losses in the banking system." This study focuses on systemic banking crises because sovereign debt and currency crises usually coincide with or follow bank crises [3]. The crisis time of the database is 1970–2017, so the last prediction time of this study is 2017.

The Group of Twenty (G20) members represent around 85% of the world's economic output, two-thirds of the global population, and 75% of international trade [5,6]. Considering the complexity of the European Union member states, this study does not include the European Union in the scope of the study. The following G20 only represents the 19 countries after the European Union is excluded.

Combining the direction of sustainable development, this study considers Environmental, Social, and Governance (ESG) indicators. ESG refers to how companies and investors integrate environmental, social, and governance issues into their business models [7]. A healthy ESG score can reduce a business' risk [8–10].

ESG can be extended from the individual economic level to the national macroeconomic level. In 2015, the United Nations identified Sustainable Development Goals to promote the increase in ESG [11]. There is a strong relationship between ESG and financial risk, and increased ESG risk leads to increased credit risk and sovereign debt risk [12–15]. To our knowledge, there are no studies on the relationship between ESG and systemic banking crises, and this study fills the gap.

Systematic bank risk warning models have gradually shifted from statistical to deep learning models. Considering neural networks have internal relationships that are difficult to explain, Huang et al. [16] selected the financial index that affects systemic banking crises in early warning with logistic regression. Beutel, List, and von Schweinitz [2] found that the logit (logistic regression) model was better than machine learning models (ANN, artificial neural networks; SVM, support vector machines; KNN, k-nearest-neighbors; DT, decision trees) in recursive out-of-sample because the performance of cross-validation and hyperparameter selection to solve the overfitting problem was overestimated. Tölö [17] argues that even with panel data, predictions in Logit and machine learning models are almost entirely based on cross-sections of variables. He found LSTM in Recurrent Neural Network (RNN) produces coherent predictions that outperform logit, at the same time as solving overfitting with a country-by-country cross-validation, and proved that hyperparameter is not important. However, he did not consider the impact of feature selection with different timesteps on the crisis.

There are three main differences between this study and previous studies: firstly, we use a Fisher score (FS) [18] for feature (variable) selection, which can reduce the computational complexity and enhance the model training speed and prediction performance; secondly, we use interpretable multivariate long-short-term memory (IMV-LSTM) [19] to capture the dynamic changes of different variables in the time series, which enhances the prediction performance and interpretability of the model; finally, we use focal loss (FL) [20] as the loss function to solve the class imbalance problem of the crisis.

This study uses the FS-FL-IMV-LSTM hybrid model to establish an early warning system that can predict systemic banking crises one year in advance. This study takes the period from 1993 to 2017 (including the 1997 Asian financial crisis, the 2007–2008 global financial crisis, etc.). The objectives of this study are as follows:

1. This study explores the impact of ESG on systemic banking crises.
2. This research uses IMV-LSTM to build an early warning model that can predict systemic banking crises one year in advance and can explain the importance of different features in different years.
3. We added the focal loss commonly used in machine vision to solve the problem of class imbalance and difficult classification, and proved the feasibility.

As a theoretical contribution, we explore the impact of ESG and systemic banking crises. We found that for unbalanced data, such as systemic banking crises, the combination of Fisher score, Focal loss, and IMV-LSTM can achieve very good results.

In terms of application value, we have established an early warning system that can predict crises one year in advance. In the test set, our model can accurately identify the occurrence of each crisis, government departments can use it to regulate macroeconomic policies, and financial institutions and investors can predict risk-taking.

The paper is organized as follows. Section 2 reviews the literature on systemic banking crises. Section 3 presents the feature selection and model architecture. Section 4 presents experimental results on the importance of different features for systemic banking crises at different timesteps. Finally, Section 5 presents the conclusions and shortcomings of this paper.

2. Literature Review

Financial crises are divided into four types: currency crises, sudden stops, foreign and domestic debt crises, and banking crises [1]. Sovereign debt and currency crises usually coincide with or follow bank crises [3]. This study follows the same definition adopted in

Laeven and Valencia [3] for systemic banking crises, reproduced below for convenience: “Signs of a serious financial crisis in the banking system, and important banking policy intervention measures are taken in response to major losses in the banking system.”

Prior literature studies early warning systems for systemic banking crises, which are mainly divided into two categories: statistical methods (signal extraction approach, probit, logit, etc.) and machine learning (ANN, SVM, KNN, etc.). Demirguc-Kunt and Detragiache [21] proposed multinomial logit and Davis and Karim [22] considered that multinomial logit was more applicable to multiple countries than the signal extraction approach, and many studies used this method [23–26]. Nakatani [27] used the probit model. The recent literature considered that machine learning is more accurate than traditionally employed methods, Dabrowski et al. [28] considered dynamic Bayesian networks to be better than the signal extraction approach and logit, Samitas et al. [29] found SVM was better than logistic regression, DT, KNN, discriminant analysis, and ensemble classifiers, Wang et al. [30] considered random forest experts voting was the best in machine learning and better than logit. However, Beutel, List, and von Schweinitz [2] found logit was better than machine learning (ANN, SVM, KNN, and DT) in recursive out-of-sample because the performance of cross-validation and hyperparameter selection to solve the overfitting problem was overestimated. Tölö [17] considered that although traditionally employed methods and machine learning use panel data, the forecast is almost entirely based on variable cross-sections, and he found that LSTM generated half fewer false alarms than logit, and proved that hyperparameter is not important.

Considering that a systemic banking crisis is a rare event (also known as skewed data or class imbalance data), we use the Fisher score and focal loss to address it. The main reason for the class imbalance problem is that the difference in the number of categories is too large, which affects the model. At present, the main solutions are data-driven and algorithm-driven [31,32]. Data-driven includes over-sampling (increasing the minority category), under-sampling (reducing the majority category), and dynamic sampling (combining the former two) [31]. Data-driven will change the data’s original structure, and the other is algorithm-driven, including a new loss function, cost-sensitive learning, threshold method, etc. [32]. Focal loss [20] is a modified version of the cross-entropy function for class-imbalanced data, originally used in computer vision, but also in other fields. Focal loss can not only solve the class imbalance problem but also solve the hard classification problem. Feature selection is an approach often overlooked in dealing with class imbalance problems [31]. Tölö [17] had tried permuting each feature, which would have increased the computational cost substantially. For class imbalance problems, feature selection is also one of the commonly used methods [31]. Fisher score can filter irrelevant features, select relevant features, save computation, and improve model performance. Maldonado et al. [33] used the combination of the Fisher score and support vector machine to deal with class imbalance data and achieved better results. Therefore, we added feature selection and chose the Fisher score.

We extend ESG from the individual economic level to the national macroeconomic level. ESG is a method for evaluating a company’s operating policies and actions, focusing on their impact on the environment, society, and corporate governance. [7]. National-level ESG is an assessment of a country’s environmental, social, and governance status, which can help to understand the country’s sustainable development level. Good ESG ratings can reduce sovereign borrowing costs [12] and mitigate sovereign credit risk [13].

We use the ESG risk score from countryrisk.io. This database includes the ESG risk scores of countries around the world from 1990 to the 2023, and even estimates ESG risk scores for the next few years. The ESG risk score is based on the sovereign risk score, adding risk factors that measure the health of a sovereign’s environment, social system strength, and quality of governance [34]. The risk sections consist of these 12 parts: economic growth prospects; institutions and governance; monetary stability; fiscal solvency and public debt; sovereign liquidity; external debt sustainability; private sector strength; climate change and renewable energy; biodiversity; education; health, food insecurity, and poverty;

and the labor market, social safety nets, and equality. Dan and Tiron-Tudor [35], Daniel Francois [36], and Hassan and Meyer [37] have used this database for other studies.

Based on the above literature, the research hypotheses are as follows:

Hypothesis 1. *The higher the ESG risk score, the greater the likelihood of a systemic banking crisis.*

3. Methodology

How to determine whether systemic banking crises will occur is very important. We chose the IMF Economic Review systemic banking crises database II [3] for 2020. This database draws on 151 systemic banking crises episodes around the globe from 1970 to 2017. This study establishes an early warning system for systemic banking crises., the research object is G20, and the research period is from 1993 to 2017.

The research object is G20 because it has an important influence on the global economy. The G20 members represent around 85% of the world's economic output, two-thirds of the global population, and 75% of international trade [5]. The G20 comprises 19 countries (Argentina, Australia, Brazil, Canada, China, France, Germany, India, Indonesia, Italy, Japan, Republic of Korea, Mexico, Russia, Saudi Arabia, South Africa, Turkey, UK, and United States) and the European Union. Considering the complexity of the European Union member states, this study does not include the European Union in the scope of the study.

This study takes the period from 1993 to 2017, including the 1997 Asian financial crisis, the 2007–2008 global financial crisis, etc. Since our model uses the systemic banking crises database II [3], which includes systemic banking crises from 1970–2017, our research period is 2017 at the latest. Due to the difficulty of collecting old data, especially in developing countries, we ended up collecting data after 1993.

The research architecture is shown in Figure 1, there are three main characteristics. Firstly, feature selection can improve the training speed and prediction accuracy of the model, and we choose the Fisher score because of its advantages of small computational effort and high accuracy. Secondly, IMV-LSTM model is a dynamic model that can calculate multiple time series simultaneously, which is more accurate and explanatory compared with LSTM. Finally, considering that crises are rare events, we solve the class imbalance problem by focal loss without changing the data structure.

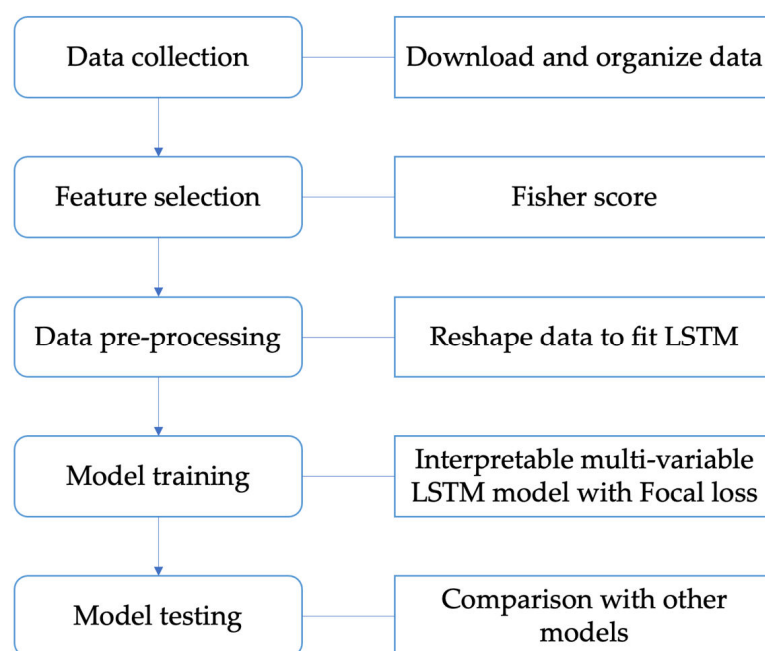


Figure 1. This research architecture.

3.1. Crisis Indicator

This study uses the systemic banking crises database II [3]. In this study, the occurrence of systemic banking crises is defined as 1, and the absence of occurrence is defined as 0. To better predict, we lead $Y_{i,t}$ by one year and mark it as $LY_{i,t}$ [17,30]. The year of the systemic banking crisis is shown in Table 1, and its formula is as follows:

$$Y_{i,t} = \begin{cases} 1, & \text{the occurrence of systemic banking crises} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$$LY_{i,t} = Y_{i,t-1} \quad (2)$$

where i represents different countries and t represents different years if a systemic banking crisis occurs, $Y_{i,t}$ takes 1, otherwise 0. In other words, we use the independent variable of the 4 years before year t to predict the dependent variable of the t year.

Table 1. Year of the systemic banking crisis.

Country	Year of Systemic Banking Crisis	LY
Argentina	1995, 2001	1994, 2000
Australia		
Brazil	1994	1993
Canada		
China	1998	1997
France	2008	2007
Germany	2008	2007
India	1993	1992
Indonesia	1997	1996
Italy	2008	2007
Japan	1997	1996
Republic of Korea	1997	1996
Mexico	1994	1993
Russia	1998, 2008	1997, 2007
Saudi Arabia		
South Africa		
Turkey	2000	1999
UK	2007	2006
United States	2007	2006

Note: Considering the complexity of the European Union member states, this study does not include the European Union in the scope of the study. Therefore, there are only 19 countries in the G20.

3.2. Early Warning Features

Macroeconomic variables can measure the overall economic trend of a country. This study uses ESG risk scores (ESG) where higher scores indicate greater ESG risk, real GDP annual growth rate (GDPGR), consumer price inflation (CPI), unemployment rate (UR), house price indicator (HP) which was 100 in 2010, current account balance divided by GDP (CABGDP), gross capital formation divided by GDP (GCFGDP), net foreign direct investment inflows divided by GDP (FDIGDP), and gross domestic savings divided by GDP (GDSGDP) are suitable features for predicting systemic banking crises [2,17,21,23,38]. The specific definitions and sources of early warning features are shown in Table 2.

ESG is the risk score of environmental, social, and governance, where higher scores indicate greater ESG risk. The ESG risk score is based on the sovereign risk score, adding risk factors that measure the health of a sovereign's environment, social system strength, and quality of governance. We use the indicators from the countryrisk.io database [34], consisting of the following 12 determinants: (1) economic growth prospects, (2) institutions and governance, (3) monetary stability, (4) fiscal solvency and public debt, (5) sovereign liquidity, (6) external debt sustainability, (7) private sector strength, (8) climate change and renewable energy, (9) biodiversity, (10) education, (11) health, food insecurity, and poverty, and (12) labor market, social safety nets, and equality. To our knowledge, no one has used

ESG as a dependent variable to explore its relationship with systemic banking crises, Dan and Tiron-Tudor [35] have used this database for other studies.

Table 2. Early warning features.

Variable	Definition	Sources
ESG	ESG risk score	A
GDPGR	Real GDP annual growth rate	B
CPI	Consumer price inflation	B
UR	Unemployment rate	B
HP	Index of housing prices	C
CABGDP	Current account balance divided by GDP	D
GCFGDP	Gross capital formation divided by GDP	D
FDIGDP	Net foreign direct investment inflows divided by GDP	D
GDSGDP	Gross domestic savings divided by GDP	D

Source: A: <https://countryrisk.io>; B: <https://www.marketline.com>; C: <https://www.euromonitor.com>; D: <https://www.worldbank.org>.

GDPGR is the annual growth rate of real gross domestic product (GDP), which reflects the growth rate of a country's economic size and wealth and is an important macroeconomic indicator. GDPGR is often used as an indicator to predict systemic banking crises [2,17,23,26].

CPI is the consumer price index, which reflects the change in the price level of consumer goods and services related to people's life. CPI is an important indicator of inflation and an important macroeconomic indicator. CPI is often used as an indicator to predict systemic banking crisis [2,23].

UR is the unemployment rate, which is the number of people in the labor force who are still not working out of the total number of employed people who meet the conditions of employment and measures the amount of idle labor capacity [26].

HP is the index of housing prices, which refers to the market value of occupied land during the time period. The average property price for each country in 2010 is 100, and the price of housing at other times is calculated on that scale. HP is often used as an indicator to predict systemic banking crises [2,17].

CABGDP, GCFGDP, FDIGDP, and GDSGDP are the values of Current Account Balance, Gross Capital Formation, Net Foreign Direct Investment Inflows, and Gross Domestic Savings divided by GDP, respectively. This eliminates the influence of different countries' background factors in this indicator and is often used as a predictor of systemic banking crises [2,17,21,23,38].

3.3. Fisher Score

The purpose of feature selection is to reduce the number of features needed to train the model, thus reducing the computational complexity and increasing the model's training speed and prediction accuracy.

A Fisher score is a feature selection algorithm for supervised learning, which can retain highly relevant features and filter redundant features. The Fisher score (FS) principle is that the characteristics of samples of the same category are similar, and its formula is as follows [39]:

$$FS(f_i) = \frac{\sum_{j=1}^c n_j (\mu_{ij} - \mu_i)^2}{\sum_{j=1}^c n_j \sigma_{ij}^2} \quad (3)$$

where i represents feature and j represents class, n_j means number in class j , μ_{ij} and σ_{ij} represent the mean and standard deviation of the j class of the i feature, respectively, and μ_i is the mean of samples corresponding to the class j .

3.4. Interpretable Multi-Variable Long Short-Term Memory

Interpretable Multi-Variable Long Short-Term Memory (IMV-LSTM) is a dynamic model based on the attention mechanism, which can be computed for time series data by combining multiple time periods, which enhances the prediction performance and

interpretability of the model. Tölö [17] considered that statistical models and machine learning models both have drawbacks in predictions based on cross-sectional data, and LSTM is a dynamic model that can really take the effect of data time series into account. IMV-LSTM adds an attention mechanism to the LSTM to improve the accuracy of the model and the interpretability of the effect of different timestep variables on the results [19].

Recurrent Neural Networks (RNNs) are very suitable for processing time series data and can combine multiple temporal features for dynamic crisis prediction [17,40]. However, when the number of time steps of the recurrent neural network is small or large, there will be problems of gradient disappearance and gradient explosion. The Long Short-Term Memory (LSTM) proposed by Hochreiter and Schmidhuber [41] can solve its gradient disappearance and gradient explosion problems well by adding input gate, forget gate, and output gate to RNN.

Interpretable Multivariate Long Short-Term Memory (IMV-LSTM) has better interpretability compared with LSTM [19]. Compared with Tölö [17], which calculates the impact of features in LSTM on the crisis through Shapley value, IMV-LSTM can not only calculate the feature contribution, but also analyze the driving factors for the crisis at different times. IMV-LSTM is a probabilistic mixture-based training method with an attention mechanism, which can learn the importance of parameters, features, and time at the same time. Compared with LSTM, the IMV-LSTM formulation is updated as follows:

$$\tilde{J}_t = \tanh(b_j + W_j \otimes \tilde{h}_{t-1} + U_j \otimes x_t), \quad (4)$$

$$\begin{bmatrix} i_t \\ f_t \\ o_t \end{bmatrix} = \sigma(b + W_j \otimes \tilde{h}_{t-1} + U_j \otimes x_t), \quad (5)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{J}_t, \quad (6)$$

$$\tilde{h}_t = o_t \odot \tanh(c_t), \quad (7)$$

where $\tilde{J}_t$ means hidden state update, $\tilde{h}_{t-1}$ is the hidden state matrix of the previous time step, U_j is input-to-hidden transition, $\otimes$ represents the product of two tensors, $\odot$ is Hadamard (element-wise) product, $W_j \otimes \tilde{h}_{t-1}$ and $U_j \otimes x_t$ are updated according to the hidden state of the previous step and the new input, respectively.

3.5. Focal Loss

Focal loss (FL) is an algorithm that can solve the class imbalance problem without changing the data. The class imbalance problem was solved by over-sampling or under-sampling the data, which by its nature changes the data structure. Lin, Goyal, Girshick, He, and Dollár [20] proposed focal loss in the field of computational vision, which can solve this problem without changing the data, has good generalization, and has been used in several fields.

The cross entropy (CE) is a commonly used binary classification loss function, and its formula is as follows [20]:

$$CE(p_t) = -\log(p_t), \quad (8)$$

where p_t represents the estimated probability for class $y = 1$.

CE loss is difficult to deal with the class imbalance problem, and Balanced Cross Entropy (BCE) loss is a commonly used solution, the formula is as follows:

$$BCE(p_t) = -\alpha_t \log(p_t), \quad (9)$$

which introduces a weighting factor α for class $y = 1$ and a weighting factor $1 - \alpha$ for the other class, α can be set as a hyperparameter.

Although a solves the class imbalance problem, it cannot distinguish between easy-to-classify and hard-to-classify samples for training. Focal loss down weights easy examples to focus training on hard-to-classify examples, the formula is as follows [20]:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t), \quad (10)$$

where α and γ are hyperparameters, α deals with class imbalanced samples, and γ deals with hard-to-classify samples.

4. Results

4.1. Data Split and Hyperparameter Setting

This research is done using python 3.7.10 based on PyTorch [42] and scikit-learn [43] framework, and the code refers to GitHub [39,44–47] and scikit-learn API [48].

Considering that cross-validation will overestimate the performance of machine learning models [2,49], We split the data between the training dataset (including the Asian financial crises) and the test dataset (including the subprime mortgage crisis) with a ratio of 5:5.

The final hyperparameters of FL-IMV-LSTM are set as follows: time steps = 4, batch size = 1, neuron = 16, epochs = 50, optimizer is Adam, learning rate = 0.0001; the loss function is focal loss, where gamma is 2 and alpha is 0.97 (is approximately equal to the probability of not having a crisis).

4.2. Descriptive Statistics

We lead y (LY) by one year for forecasting. Therefore, FY is for 1994–2017 and the other variables are for 1993–2016, for a total of 24 years for 19 countries.

Table 3 indicates Descriptive statistics. The mean value of FY is 0.0351, indicating that the probability of a systemic banking crisis in 24 years in 19 countries is about 3.51%, which is a rare event (also known as class imbalance problem).

Table 3. Descriptive statistics.

Variable	Obs	Mean	Std.Dev.	Min	Max
FY	456	0.0351	0.1842	0	1
ESG	456	0.3088	0.1479	0.0134	0.7706
GDPGR	456	0.0321	0.0375	−0.1312	0.1423
CPI	456	0.1843	1.3956	−0.0140	20.7588
UR	456	0.0806	0.0526	0.0160	0.3030
HP	456	0.8118	0.3156	0.0020	2.4120
CABGDP	456	0.0022	0.0487	−0.1297	0.2740
GCFGDP	456	0.2423	0.0657	0.1085	0.4666
FDIGDP	456	0.0204	0.0188	−0.0361	0.1276
GDSGDP	456	0.2581	0.0815	0.1106	0.5542

A larger ESG indicates a higher ESG risk, with a mean ESG of 0.3088 and a standard deviation of 0.1479, indicating that the average ESG risk is low across 19 countries over 24 years and varies widely across countries.

The CPI has a mean value of 0.1843 and a standard deviation of 1.3956. This indicates that inflation is generally large across countries and varies widely across countries. The mean value of HP is 0.8118 and the standard deviation is 0.3156, indicating that there is also a large variation in house prices across countries. The mean value of CABGDP is 0.0022 and the standard deviation is 0.0487, indicating that the current account balance as a percentage of GDP does not vary much between countries. The mean value of UR is 0.0806 and the standard deviation is 0.0526, indicating an average unemployment rate of about 8.06% for the 19 countries over 24 years.

Figure 2 shows the ESG risk score of G20 countries in 2022, arranged in descending order (Due to the war in Russia, Russia's ESG risk score in 2022 has not been updated, and

Russia's ESG risk score in 2021 is shown in the figure). The greater the ESG risk scores, the greater the ESG risk of the country in 2022. The ESG risks of Brazil, India, South Africa, and Argentina are all very high in 2022. Germany has the lowest ESG risk, followed by Australia.



Figure 2. 2022 ESG Risk Score. Note: Source: <https://countryrisk.io>.

Table 4 represents the ESG risk score before a systemic banking crisis, t represents the year of the crisis, and $t - 1$ represents one year before the crisis. Since our model predicts whether a systemic banking crisis will occur in the fifth year based on the data of the previous four years, the number of crises actually predicted will be less than that in Table 1.

Table 4. ESG Risk Score Before Systemic Banking Crisis.

Country	$t-4$	$t-3$	$t-2$	$t-1$	Crisis Year (t)
Argentina	0.3426	0.4047	0.4665	0.4024	2001
China	0.5940	0.3940	0.3857	0.3495	1998
France	0.1790	0.1629	0.1684	0.1843	2008
Germany	0.1491	0.1491	0.1257	0.1459	2008
Indonesia	0.5473	0.5680	0.6143	0.6438	1997
Italy	0.3302	0.3169	0.3168	0.3045	2008
Japan	0.1484	0.2148	0.2726	0.2297	1997
Republic of Korea	0.2790	0.3162	0.3274	0.3815	1997
Russia	0.5644	0.5043	0.6333	0.5397	1998
Russia	0.3032	0.2802	0.2471	0.2581	2008
Turkey	0.5515	0.5785	0.4359	0.5066	2000
UK	0.1333	0.1237	0.1108	0.1124	2007
United States	0.0404	0.0705	0.0479	0.0279	2007

Source: Laeven and Valencia [3], <https://countryrisk.io> [34].

4.3. Feature Selection

We use the Fisher score for feature selection. The principle is that features of the same category are similar and features of different categories are different. Finally, we choose a score greater than 0.001 as the model feature variable. Finally, we select CPI, HP, ESG, CABGDP, and UR; exclude FDIGDP, GDPGR, GCFGDP, and GDSGDP, as in Table 5.

Table 5. Fisher Score.

Feature Selection	Variable	Fisher Score
select	CPI	0.0232
	HP	0.0189
	ESG	0.0021
	CABGDP	0.0016
	UR	0.0016
exclude	FDIGDP	0.0009
	GDPGR	0.0005
	GCFGDP	0.0003
	GDSGDP	0.0002

4.4. Prediction Performance

4.4.1. Confusion Matrix Metric Description

Classification tasks are often evaluated by confusion matrices, including true positive (TP), true negative (TN), false positive (FP), and false negative (FN), as Table 6. TP and TN indicate correct predictions, where TP indicates that both the prediction and the actual are positive, and TN indicates that both the prediction and the actual are negative. FP and FN denote incorrect predictions, where FP indicates that the prediction is positive but actually negative, and FN indicates that the prediction is negative but actually positive.

Table 6. Confusion Matrix.

	Negative (Predict)	Positive (Predict)
Negative (Actual)	TN (125)	FN (69)
Positive (Actual)	FP (0)	TP (6)

Notes: In parentheses are the values of our proposed FS-FL-IMV-LSTM in the confusion matrix.

The class imbalance problem should not only focus on Accuracy but also on Recall, Precision, and F1-score, formulas and examples are shown in Table 7.

Table 7. Confusion Matrix Performance Metrics.

Metric	Formula	Example
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Percentage of correct predictions
Precision	$\frac{TP}{TP+FP}$	Percentage of spam that is really spam
Recall	$\frac{TP}{TP+FN}$	Percentage of nucleic acid tests found positive
F1	$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	Takes into account both precision and recall

Area Under the Curve (AUC) is a common metric that can be used to evaluate the performance of different classification models. The higher the AUC, the better the predictive power of the model.

4.4.2. Comparison of Model Prediction Performance

Table 5 shows that among the 200 samples in the test set, we successfully predicted the occurrence of six systemic banking crises, 125 did not occur, and no crisis was missed.

Our proposed FS-FL-IMV-LSTM can successfully predict every systemic banking crisis one year in advance, as shown in Table 7. The accuracy of this model is 0.6550, indicating an overall accuracy rate of 65.5%, which is in the middle to upper range. However, systemic banking crises are rare events that can bring huge losses if they are not warned in advance, so it is more important to focus on recall. The recall of this model is 1, indicating the ability to provide 100% early warning of systemic banking crises one year in advance. The AUC of this model is 0.8153, which also gets a substantial improvement compared with other models. Both Precision and F1 also outperformed the other models.

Table 8 represents the performance evaluation of each model prediction. The other models (The first six machine learning models) have high accuracy but 0 recall, even if the class weight is set to balance, this means that it is impossible to accurately predict the crisis. FL-IMV-LSTM, although recall reaches 100%, can accurately predict crises. However, the accuracy rate is too low, only reaching 3%, and almost all the data are predicted to have a crisis. Figure 3 represents that FS-FL-IMV-LSTM's AUC is 0.8149, significantly higher than the chance level.

Table 8. Prediction Performance.

Model	Accuracy	Precision	Recall	F1	Auc
Logistic	0.9605	0.0000	0.0000	0.0000	0.4932
Gaussian Naive Bayes	0.9693	0.0000	0.0000	0.0000	0.4977
Decision Tree	0.9737	0.0000	0.0000	0.0000	0.5000
Support Vector Machine	0.9254	0.0000	0.0000	0.0000	0.4752
Random Forest	0.9737	0.0000	0.0000	0.0000	0.5000
Multi-Layer Perceptron	0.9737	0.0000	0.0000	0.0000	0.5000
FL-IMV-LSTM	0.0700	0.0213	0.6667	0.0412	0.2637
FS-FL-IMV-LSTM	0.6550	0.0800	1.0000	0.1481	0.8153

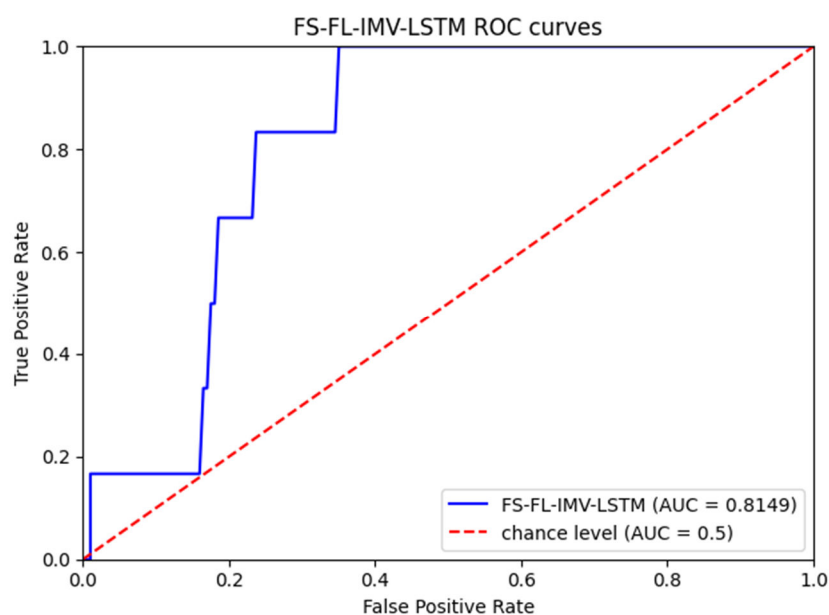


Figure 3. FS-FL-IMV-LSTM ROC curves.

Our proposed FS-FL-IMV-LSTM outperforms other models in terms of overall performance and practical use.

4.5. Interpretation

Figure 4 represents the importance of variables in crisis forecasting. Where HP is 0.2148, the largest share, indicating the greatest importance of house prices for predicting the crisis. The importance of the other variables ranged from 0.1940 to 0.1982, and there was no particularly large difference.

It is worth mentioning that the importance of ESG reaches 0.1942, indicating that, among these five variables, ESG has an impact of 19.42% on predicting systemic banking crises. Therefore, we believe that the country-level Environment, Social, and Governance variables can predict systemic banking crises, which is consistent with **Hypothesis 1**. *The higher the ESG risk score, the greater the likelihood of a systemic banking crisis.*

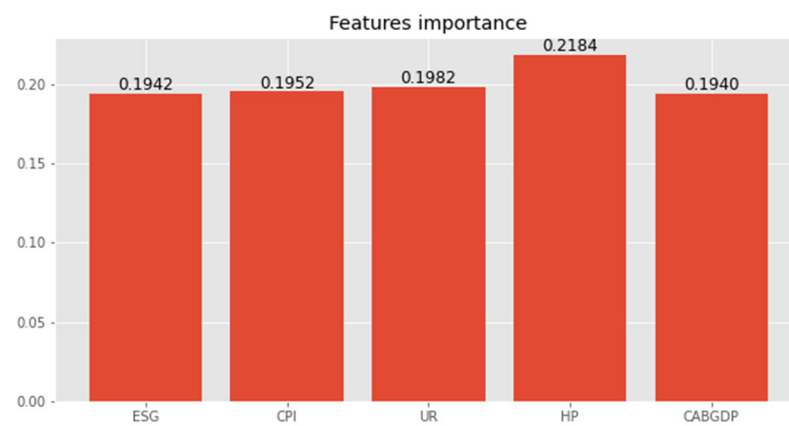


Figure 4. Feature importance.

Since the ESG risk score is based on the sovereign risk score and incorporates risk factors such as environmental health status, social system strength, and governance quality, we believe that this indicator reflects the country's macroeconomic situation to a certain extent, so the ESG risk score is important for the future and has significant implications whether or not a systemic banking crisis occurs.

Figure 5 represents the importance of each variable at different timesteps, with colors closer to yellow indicating a more significant contribution to the prediction.

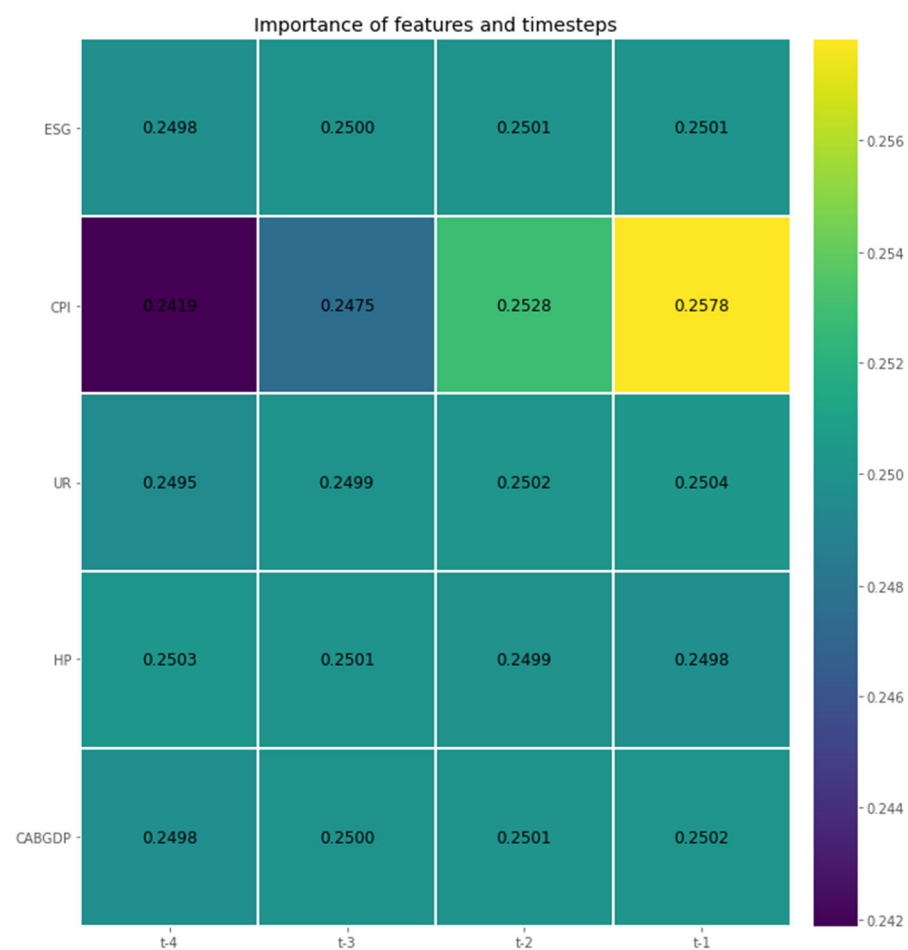


Figure 5. Importance of features and timesteps.

The importance of each timestep of ESG is about 0.25, indicating that whether a systemic banking crisis will occur in the fifth year, the contribution of ESG in the first four years is equal.

Regarding whether there will be a systemic banking crisis in the fifth year, the importance of CPI in the timesteps of year $t - 4$ is 0.2419, year $t - 3$ is 0.2475, year $t - 2$ is 0.2528, and year $t - 1$ is 0.2578, indicating that for the first four years the impact of China's Consumer Price Index on the systemic banking crisis is increasing year by year.

The importance of UR, HP, and CABGDP at each time step is almost the same, all around 25%, indicating whether there will be a systemic banking crisis in the fifth year, and the contribution of UR, HP, and CABGDP in the first four years is equal.

Compared with Tölö's [17] use of the Shapley value to analyze the importance of variables in LSTM, IMV-LSTM can analyze the impact of different variables on each timestep in the model.

5. Conclusions

This study collates data from 1993 to 2017 for G20 countries, uses the FS-FL-IMV-LSTM model to establish an early warning system that can successfully predict every systemic banking crisis one year in advance, and explores the relationship between ESG and systemic banking risks.

The theoretical contribution of this study is to fill in the blank of whether ESG affects the systemic banking crisis and the importance of different variables to the crisis at different timesteps. We find that ESG has a significant impact on systemic banking crises, consistent with **Hypothesis 1**. *The higher the ESG risk score, the greater the likelihood of a systemic banking crisis*. We find that ESG, the inflation rate, the unemployment rate, house price, and the ratio of the current account balance to GDP all affect the occurrence of systemic banking crises. It is worth noting that inflation in the year before the crisis had a greater impact than in the four years preceding the crisis. We believe that the ESG risk score is a sovereign risk score that incorporates ESG, which can reflect the country's macroeconomics to a certain extent and can predict systemic banking crises. We also found that when dealing with class imbalance, adding feature selection and focal loss to the deep learning model can achieve better results.

The practical contribution of this study is that it achieved a 100% recall rate on the test set, which means that every crisis was predicted by our early warning system (FS-FL-IMV-LSTM). We hope that this early warning system, which can be predicted one year in advance, can be used as a reference for government departments, financial industry regulators, and researchers to reduce the major losses caused by systemic banking crises.

However, this study has not obtained how much impact the changes of different variables at different times will have on the probability of crisis occurrence. This study fails to divide ESG into three parts, which are studied from the aspects of the environment, society, and management.

The occurrence of a systemic banking crisis has a significant impact on the economy and society. We hope this research can make some contributions to theoretical and practical research.

Author Contributions: Conceptualization, S.-L.L.; Methodology, X.J.; Software, X.J.; Validation, S.-L.L.; Formal analysis, S.-L.L.; Investigation, X.J.; Data curation, X.J.; Writing—original draft, X.J.; Writing—review & editing, S.-L.L.; Visualization, S.-L.L. and X.J.; Supervision, S.-L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available in sources below Tables 1 and 2.

Acknowledgments: We thank the editorial board and the reviewers for their useful feedback that improved this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Claessens, S.; Kose, M.A. *Financial Crises Explanations, Types, and Implications*; IMF Working paper; IMF: Washington, DC, USA, 2013.
2. Beutel, J.; List, S.; von Schweinitz, G. Does machine learning help us predict banking crises? *J. Financ. Stab.* **2019**, *45*, 100693. [CrossRef]
3. Laeven, L.; Valencia, F. Systemic Banking Crises Database II. *IMF Econ. Rev.* **2020**, *68*, 307–361. [CrossRef]
4. NobelPrize. The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2022. Available online: <https://www.nobelprize.org/prizes/economic-sciences/2022/summary/> (accessed on 17 October 2022).
5. G20. About the G20. Available online: <https://g20.org/> (accessed on 17 December 2022).
6. Song, Q.; Liu, J.; Sriboonchitta, S. Risk Measurement of Stock Markets in BRICS, G7, and G20: Vine Copulas versus Factor Copulas. *Mathematics* **2019**, *7*, 274. [CrossRef]
7. Gillan, S.L.; Koch, A.; Starks, L.T. Firms and social responsibility: A review of ESG and CSR research in corporate finance. *J. Corp. Financ.* **2021**, *66*, 101889. [CrossRef]
8. Cohen, G. ESG risks and corporate survival. *Environ. Syst. Decis.* **2022**, 1–6. [CrossRef]
9. Broadstock, D.C.; Chan, K.; Cheng, L.T.W.; Wang, X. The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China. *Financ. Res. Lett.* **2021**, *38*, 101716. [CrossRef]
10. Gil, C. What can we learn from the financial market about sustainability? *Environ. Syst. Decis.* **2022**, *42*, 1–7. [CrossRef]
11. Nations United. *Transforming Our World: The 2030 Agenda for Sustainable Development*; United Nations, Department of Economic and Social Affairs: New York, NY, USA, 2015.
12. Crifo, P.; Diaye, M.-A.; Oueghlissi, R. The effect of countries' ESG ratings on their sovereign borrowing costs. *Q. Rev. Econ. Financ.* **2017**, *66*, 13–20. [CrossRef]
13. Hübel, B. Do markets value ESG risks in sovereign credit curves? *Q. Rev. Econ. Financ.* **2020**, *85*, 134–148. [CrossRef]
14. Zhang, J.; De Spiegeleer, J.; Schoutens, W. Implied Tail Risk and ESG Ratings. *Mathematics* **2021**, *9*, 1611. [CrossRef]
15. Batrancea, I.; Batrancea, L.; Maran Rathnaswamy, M.; Tulai, H.; Fatacean, G.; Rus, M.-I. Greening the Financial System in USA, Canada and Brazil: A Panel Data Analysis. *Mathematics* **2020**, *8*, 2217. [CrossRef]
16. Huang, A.; Qiu, L.; Li, Z. Applying deep learning method in TVP-VAR model under systematic financial risk monitoring and early warning. *J. Comput. Appl. Math.* **2021**, *382*, 113065. [CrossRef]
17. Tölö, E. Predicting systemic financial crises with recurrent neural networks. *J. Financ. Stab.* **2020**, *49*, 100746. [CrossRef]
18. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
19. Guo, T.; Lin, T.; Antulov-Fantulin, N. Exploring interpretable LSTM neural networks over multi-variable data. In *Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research, Long Beach, CA, USA, 9–15 June 2019*; pp. 2494–2504.
20. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef]
21. Demircuc-Kunt, A.; Detragiache, E. The determinants of banking crises in developing and developed countries. *Int. Monet. Fund Staff Pap.* **1998**, *45*, 81–109. [CrossRef]
22. Davis, E.P.; Karim, D. Comparing early warning systems for banking crises. *J. Financ. Stab.* **2008**, *4*, 89–120. [CrossRef]
23. Qin, X.; Luo, C. Capital account openness and early warning system for banking crises in G20 countries. *Econ. Model.* **2014**, *39*, 190–194. [CrossRef]
24. Caggiano, G.; Calice, P.; Leonida, L. Early warning systems and systemic banking crises in low income countries: A multinomial logit approach. *J. Bank. Financ.* **2014**, *47*, 258–269. [CrossRef]
25. Hamdaoui, M. Are systemic banking crises in developed and developing countries predictable? *J. Multinatl. Financ. Manag.* **2016**, *37–38*, 114–138. [CrossRef]
26. Filippopoulou, C.; Galarotis, E.; Spyrou, S. An early warning system for predicting systemic banking crises in the Eurozone: A logit regression approach. *J. Econ. Behav. Organ.* **2020**, *172*, 344–363. [CrossRef]
27. Nakatani, R. Macroprudential policy and the probability of a banking crisis. *J. Policy Model* **2020**, *42*, 1169–1186. [CrossRef]
28. Dabrowski, J.J.; Beyers, C.; de Villiers, J.P. Systemic banking crisis early warning systems using dynamic Bayesian networks. *Expert Syst. Appl.* **2016**, *62*, 225–242. [CrossRef]
29. Samitas, A.; Kampouris, E.; Kenourgios, D. Machine learning as an early warning system to predict financial crisis. *Int. Rev. Financ. Anal.* **2020**, *71*, 101507. [CrossRef]
30. Wang, T.; Zhao, S.; Zhu, G.; Zheng, H. A machine learning-based early warning system for systemic banking crises. *Appl. Econ.* **2021**, *53*, 2974–2992. [CrossRef]
31. Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from class-imbalanced data: Review of methods and applications. *Expert Syst. Appl.* **2017**, *73*, 220–239. [CrossRef]
32. Johnson, J.M.; Khoshgoftaar, T.M. Survey on deep learning with class imbalance. *J. Big Data* **2019**, *6*, 27. [CrossRef]
33. Maldonado, S.; Weber, R.; Famili, F. Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines. *Inf. Sci.* **2014**, *286*, 228–246. [CrossRef]
34. Countryrisk.io. ESG Risk Score. Available online: <https://countryrisk.io/> (accessed on 27 February 2021).
35. Dan, A.; Tiron-Tudor, A. The determinants of green bond issuance in the European Union. *J. Risk Financ. Manag.* **2021**, *14*, 446. [CrossRef]

36. Daniel Francois, M. An assessment of the interrelations between country risk, economic growth and good governance: The case of the Visegrád four. *J. East. Eur. Cent. Asian Res.* **2021**, *8*, 610–627. [[CrossRef](#)]
37. Hassan, A.S.; Meyer, D.F. Does countries' environmental, social and governance (ESG) risk rating influence international tourism demand? A case of the Visegrád Four. *J. Tour. Futures* **2022**, 1–20. [[CrossRef](#)]
38. Roy, S. What drives the systemic banking crises in advanced economies? *Glob. Financ. J.* **2022**, *54*, 100746. [[CrossRef](#)]
39. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature Selection: A Data Perspective. *ACM Comput. Surv.* **2017**, *50*, 94. [[CrossRef](#)]
40. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
41. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
42. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. Pytorch: An imperative style, high-performance deep learning library. In Proceedings of the Advances in Neural Information Processing Systems 32, Vancouver, BC, Canada, 8–14 December 2019.
43. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
44. Azure, M. Deep Learning for Predictive Maintenance. Available online: https://github.com/Azure/lstms_for_predictive_maintenance/blob/master/Deep%20Learning%20Basics%20for%20Predictive%20Maintenance.ipynb (accessed on 18 August 2021).
45. Cheung, C. Pytorch-Loss. Available online: https://github.com/CoinCheung/pytorch-loss/blob/master/focal_loss.py (accessed on 18 June 2022).
46. Kurochkin, A. IMV_LSTM. Available online: https://github.com/KurochkinAlexey/IMV_LSTM (accessed on 12 March 2022).
47. Licheng0794. IMV-LSTM-Classification. Available online: <https://github.com/licheng0794/IMV-LSTM-Classification/> (accessed on 12 March 2022).
48. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J. API design for machine learning software: Experiences from the scikit-learn project. *arXiv preprint* **2013**, arXiv:1309.0238.
49. Neunhoeffer, M.; Sternberg, S. How cross-validation can go wrong and what to do about it. *Political Anal.* **2019**, *27*, 101–106. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.