PhD in Information Technology and Electrical Engineering
Università degli Studi di Napoli Federico II

PhD Student: Vincenzo Lanzetta

Cycle: XXXVII

Training and Research Activities Report

Year: First

Vincenzo Lanzetta

Tutor: Prof. Roberto Prevete Roberto Prevete

Co-Tutor:

Date: October 31, 2022

1. Information:

- **PhD student: Vincenzo Lanzetta**
- **DR number:**
- **Date of birth: 25/08/74**
- **Master Science degree: Chemistry University: Federico II of Naples**
- **Doctoral Cycle: XXXVII**
- **Scholarship type:** ***no scholarship***
- **Tutor: Prof. Roberto Prevete**
- **Co-tutor:**

2. Study and training activities:

Activity	Type[1]	Hours	Credits	Dates	Organizer	Certificate[2]
Operational Research: Mathematical Modelling, Methods and Software Tools for Optimization Problems	Course	10	4	from September to October, 2022	Prof. Adriano Masone	Y
Sustainable ship for the energy transition of maritime transport	Course	10	4	from September to October, 2022	Prof. Tommaso Coppola	Y
Machine Learning for Science and Engineering Research	Course	20	5	from june to July, 2022	Professors Corazza, Prevete, Isgrò, Sansone, Pezzulo	N
Imprenditorialità accademica	Course	12	4	from May to July, 2022	Prof. Pierluigi Rippa	Y
Statistical data analysis for science and engineering research	Course	12	4	from march to april, 2022	Prof. R. Pietrantuono	Y
Bench to Bytes to Bedside: Converting genomic data intohealthcare tools	Seminar	1	0.2	March 4, 2022	Serena Nik-Zainal	Y
Explainable Natural Language Inference	Seminar	1.5	0.3	April 13, 2022	Dr. Marco Valentino	Y
Using Delays for Control	Seminar	1	0.2	April 21, 2022	Prof. Emilia	Y

					Fridman	
Symbiotic Control of Wearable Soft Suits for human motion assistance and augmentation	**Seminar**	**2**	**0.4**	May 20, 2022	Prof. Lorenzo Masia	**Y**
AR for remote use of measurement instrumentation	**Seminar**	**2**	**0.4**	May 24, 2022	Prof.ssa Annalisa Liccardo and Dr. Francesco Bonavolonta'	**Y**
Artificial Intelligence and 5G combined with holographic technology: a new perspective for remote health monitoring	**Seminar**	**2**	**0.4**	May 26, 2022	Dr. Pietro Ferraro and Dr. Pasquale Memmolo	**Y**
Towards substainable IT	**Seminar**	**2**	**0.4**	May 27, 2022	Dr. Giorgia Sepe	**Y**
History of fusion	**Seminar**	**1**	**0.2**	July 1, 2022	Prof. Pietro Martin	**Y**
Introduction to Intellectual Property Management	**Seminar**	**2**	**0.4**	July 19, 2022	Prof. Alessandro Marroni	**Y**
Tutorial on “statistical Learning for sensory and consumer science	**Tutorial**	**8**	**1.6**	September 14 from 9:00 to 13.00 and September, 15 2022 from 9:00 to 13:00	Prof. Naes, Prof. Tomic, Prof. Romano	**Y**
Privacy-preserving machine learning	**Seminar**	**2**	**0.4**	October 14, 2022	Prof. R. Natella, Prof. S. P. Romano	**Y**
Image processing for medical applications	**Seminar**	**1.5**	**0.3**	October 20, 2022	Evans Cosgrove	**Y**

1) Courses, Seminar, Doctoral School, Research, Tutorship
2) Choose: Y or N

2.1. Study and training activities - credits earned

	Courses	Seminars	Research	Tutorship	Total
Bimonth 1			**3**		**3**
Bimonth 2			**6**		**6**

Bimonth 3		0.7	9		9.7
Bimonth 4		1.6	6		7.6
Bimonth 5	4	0.6	3		7.6
Bimonth 6	17	2.3	6		25.3
Total	21	5.2	33	0	59.2
Expected	20 - 40	5 - 10	10 - 35	0 – 1.6	

3. Research activity:

Topic: deep learning methods for analysis and prediction of financial data

Abstract: current literature on deep learning methods for analysis and prediction of financial data is characterized by studies on CNNs-based transfer learning methodologies constituted by a pre-training phase on specific financial datasets and by a fine-tuning phase on a different financial dataset, and in which the financial technical indicators values, or the security prices, are used as input features [22, 23]; but there is lack of studies related to the above CNNs-based transfer learning methodologies in which the pixel values of financial candlestick images are used as input features; thus, we intend to develop a transfer learning methodology based on a CNN architecture trained on financial candlestick images, for the fine-tuning stock class prediction; our proposed methodology will let to extract the general pattern, underlying financial candlestick images, from different securities and will let to predict a specific security class, according to a defined threshold, in a fine-tuning step. This research idea could be able to give interesting insight from the practical stock trading point of view, as the very common use of financial candlestick images - among financial practitioners - for the security price prediction

Methodology: we propose a novel methodology that uses a CNN-based architecture on financial candlestick images - transfer learning based - for the binary label classification of financial securities; each “i” candlestick image, of the dataset, will be constituted by a defined number of candles, such as 50 candles for example (obviously, this number will be defined by a specific literature analysis to be conducted); in case of candlestick images constituted by 50 candles, for each “i” candlestick image the related candles will be the ones from “i-49” timestep to the one till “i” time-step. The time-step will be the daily one.

The label will be associated to a “buy” or “ not buy” class, depending on whether the security price percentage change is above a specific percentage threshold, or not; the percentage change will be calculated by using the following formula:

$$c(i) = [p(i+1) - p(i)]/p(i)$$

where

c(i) = percentage change with respect to the “i” time

p(i+1) = security price of the “i+1” time

p(i) = security price of the “i” time

Each “i” candlestick image will be associated to the “i” label calculated by comparing the output of the above formula and the defined percentage threshold.

With the aim to use a cross-validation approach which is able to let the model to learn from data by considering the correct temporal sequence within the usual neural network data splitting methodology (training, validation and test set), we will split the dataset according to a sliding window approach such as in [12].

Obviously, we will test several subset numbers, in which the original dataset could be split for the cross-validation aim, accordingly to literature suggestions on the optimal subset size with respect to the size of the original dataset.

As our transfer learning methodology is constituted by a pre-training phase (on an “x” financial dataset), and by one (or more) fine-tuning phases on one (or more) “y” financial dataset, we will save the final weights of each phase – excepting the last ones - in order to use them as starting weights of the following phase; obviously, the starting weights of the very first phase (i.e., the pre-training one) will be random weights.

Performance measurement description: Literature have highlighted that financial trading systems has to be evaluated not only with respect to the accuracy, but also with respect to the profitability, because costs of the financial trading activity are able to have a very critical impact from the practical point of view [8]; a possible metric, aimed at measuring the trading algorithms performance, could be the cumulative return over an extensive period of time, or the average return per trade [8]; on this regard, from the practical point of view it is important to evaluate the above financial return metrics with respect to the “buy and hold” strategy (i.e., with respect to the usual strategy of passive investors).

In order to evaluate the real practicability of a trading algorithm, it is also important to evaluate risk management aspects by understanding if there are stable and consistent returns; with this aim, some possible metrics could be the “maximum drawdown” [8] or standard deviations of the returns [12]; on this regard, the adequate metrics has to be defined by going deeper on the risk management literature.

Statistical methodology for the validation of the results. scholars have highlighted that <<statistical tests should be incorporated as best practices in the field of machine learning financial time series prediction for significance and robustness.>>[1]; on this regard, literature highlighted that, with a lot of test to be conducted, the statistical evaluation - for the validation of the results - has to be developed with respect to a possible statistical bias related to an exhaustive search on combination of variables, because <<the probability that a result arose by chance grows with the number of combinations tested (White 2000)>>[13]. Thus, with respect to the optimization of our transfer learning model, if we need to find the ideal combination of conditions which are able to produce the better result, we need to experiment a lot of test, on the same data;

When we need to make a lot of pair-wise comparisons on the same data, we have to deal with the multiple testing issue which happens if we don’t reduce the usual threshold ($\alpha = 0.05$) of the rejecting critical region; in fact, if we use $\alpha = 0.05$ without considering that we deal with a battery of pair-wise tests on the same data, we will reject – by chance – the null hypothesis in 5% of the cases; with a battery of pair-wise tests constituted by 100 tests on the same data, for example, if we use $\alpha = 0.05$ in each test, we will get 5 times the rejection of the null hypothesis within the evaluation of the 100 hypothesis tests; in other words, for 5 times (on the 100 pair-wise comparisons of the current example) we will decide to choose 1 model, wrongly; more in depth, we get the above 5 wrong decisions because the more tests we make, the larger the probability that a random fluctuation is mis-classified as a meaningful result, notwithstanding the chance of noise affecting one result is small.

In order to deal with the multiple testing issue by correctly reducing the threshold of the rejecting critical region, we have to develop the proposed literature corrections, such as the so-called "Bonferroni correction" or the so-called "Benjamin & Hochberg correction", or other ones [41, 42]

Results:

- Developing different methods to build financial images from financial time series
- Developing and experimenting a transfer learning model, with a pretraining phase and a fine tuning one, to perform stock market prediction
- Developing and experimenting a pre-trained deep learning model (VGG-16) for the classification of financial images
- Developing a method aimed at using financial data within a CNN-LSTM architecture
- Draft of a review paper on deep learning approaches for the financial market predictions and statistical methodologies for the validation of the results

Some references:

[1] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Literature review: Machine learning techniques applied to financial market prediction," Expert Syst. Appl., vol. 124, pp. 226–251, 2019.

[8] E. A. Gerlein, M. McGinnity, A. Belatreche, and S. Coleman, "Evaluating machine learning classification for financial trading: An empirical approach," Expert Syst. Appl., vol. 54, pp. 193–207, 2016.

[12] O. B. Sezer and A. M. Ozbayoglu, "Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach," Appl. Soft Comput. J., vol. 70, pp. 525–538, 2018.

[13] G. Sermpinis, A. Karathanasopoulos, R. Rosillo, and D. de la Fuente, "Neural networks in financial trading," Ann. Oper. Res., 2019.

[22] Hoseinzade, E., Haratizadeh, S., & Khoeini, A. (2019). U-cnnpred: A universal cnn-based predictor for stock markets. arXiv preprint arXiv:1911.12540.

[23] Q.-Q. He, P. C.-I. Pang, and Y.-W. Si, ''Transfer learning for financial time series forecasting,'' in Proc. Pacific Rim Int. Conf. Artif. Intell. Cham, Switzerland: Springer, 2019, pp. 24–36.

[41] S. Lee, D.K. Lee, What is the proper way to apply the multiple comparison test? Korean J. Anesthesiol., 71 (5) (2018), pp. 353-360

[42] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (Method). 1995; 57:289–300.

4. Research products:

5. Conferences and seminars attended

- Tutorial on "statistical Learning for sensory and consumer science (September 14 from 9:00 to 13.00 and September, 15 2022 from 9:00 to 13:00) – 1.6 credits – Lecturers: Prof. Naes, Prof. Tomic, Prof. Romano

5. Activity abroad:

None. 0 months spent abroad.

7. Tutorship