International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

IJIEMR Transactions, online available on 12th September 2024. Link

https://ijiemr.org/downloads.php?vol=Volume-13&issue=issue09

DOI: http://doi.org/10.36893/IJIEMR/2024/V13/I09/7

Title: " A STOCK PRICE PREDICTION MODEL BASED ON INVESTOR SENTIMENT AND OPTIMIZED DEEP LEARNING "

Volume 13, ISSUE 09, Pages: 52- 61

Paper Authors

Mr. Vijay Kumar L,P. Usha Sri,P. Laxmi Sanjana,L. Harshini

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

A STOCK PRICE PREDICTION MODEL BASED ON INVESTOR SENTIMENT AND OPTIMIZED DEEP LEARNING

[1] Mr. Vijay Kumar L,[2] P. Usha Sri,[3] P. Laxmi Sanjana,[4] L. Harshini
[1]Assistant Professor,[234]Students
Department Of CSE
Malla Reddy Engineering College for Women

ABSTRACT

Reducing investing risks and increasing rewards are possible with an accurate stock price forecast. This study builds the MS-SSA-LSTM model by combining the multi-source data influencing stock prices and using deep learning, swarm intelligence algorithms, and sentiment analysis. In order to create a distinct sentiment dictionary and determine the sentiment index, we first scan the East Money forum postings. Next, the hyperparameters of the Long and Short-Term Memory networks (LSTM) are optimized using the Sparrow Search Algorithm (SSA). Lastly, LSTM is utilized to predict future stock prices by integrating the sentiment index with fundamental trading data. Tests show that the MS-SSA-LSTM model performs better than the others and has a high degree of generalizability. The average improvement in R2 between MS-SSA-LSTM and regular LSTM is 10.74%. We discovered that: 1) The prediction performance of the model may be improved by including the sentiment index. 2) SSA is used to adjust the hyperparameters of the LSTM, improving the prediction impact and providing an objective explanation of the model parameter values. 3) The Chinese financial market's extreme volatility makes it more suited for short-term forecasting.

I. INTRODUCTION

The rapid growth of internet banking and the improvement of China's stock market have attracted an increasing number of people to careers in the financial industry. However, the stock market is notoriously volatile and recognized for its massive data sets. Many individual investors need better data-mining skills to make a return. As a result, accurate stock price forecasting may help companies and investors reduce investment risks and maximize investment returns.

The first researchers used statistical methods to construct a linear model that matched the pattern of stock prices over time. Example examples of traditional methods include GARCH, ARIMA, ARMA, and many more. The ARMA was created in order to do a stock analysis using time series data. The ARIMA model, which is based on the ARMA, was developed to predict the general direction of stock price fluctuations [2]. It is also possible to use wavelet analysis to the ARIMA model to improve the fitting accuracy of the Shanghai Composite Index [3]. In order to forecast stock time series over an extended period of time, the GARCH model provides new ideas [4]. At the same time, several authors provided theoretical support for volumetric price analysis of multivariate stocks by creating a new prediction model that combined ARMA and GARCH [5]. Data that is regular and ordered is often the only thing that these older methods capture.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Traditional forecasting methods, however, rely on assumptions that are just not feasible. Consequently, nonlinear financial data is challenging to define using statistical approaches.

Consequently, a lot of researchers attempt to forecast stock prices using machine learning methods such as Neural Networks and Support Vector Machines (SVM). Machine learning is based on the idea that computers can learn from data, analyze it, and use it to make predictions about fresh data. Due to its unique strengths in dealing with low-sample-size data, high-dimensional data, and nonlinear situations, the SVM is widely used by academics for stock forecasting. When comparing SVM methods to statistical ones, Hossain and Nasser [6] found that the SVM method produced more accurate stock predictions. After proposing a hybrid SVM model to forecast HS300 index fluctuations, Chai et al. [7] found that a combination of least squares SVM and Genetic Algorithm (GA) produced the best results. When used to large-scale training sets, SVM would need a lot of memory and processing resources, which might limit its capacity to anticipate a lot of stock data. Next, we use ANN and multi-layer ANN to solve financial time series issues. Experimental evidence suggests that artificial neural networks (ANNs) have a number of benefits, including rapid convergence and great accuracy [8, 9, 10]. Moghaddam and Esfandyari [11] tested a number of feed forward artificial neural networks to determine their effect on stock price prediction in the market. Liu and Hou [12] improved the BP (Back Propagation) neural network by using the Bayesian regularization approach. But the following areas might need some development using the tried-and-true neural network method. Overfitting and local optimization quickly follow from poor generalizability. Since a significant number of samples needs to be educated to overcome these challenges, improved models are necessary.

In order to forecast stock prices, this study presents a new model (MS-SSA-LSTM) that uses the Sparrow Search Algorithm to combine LSTM neural networks with data from many sources. Traders and investors may use the MS-SSA-LSTM stock price prediction model to help them make more educated investment decisions by predicting the stock price in advance. Investors and traders use the MS-SSA-LSTM model to input data on particular stocks, such as shareholder feedback and historical transaction data. In addition to automatically creating a stock price trend chart, the algorithm also forecasts the stock price for the next day.

This is the source from which the ideas for this paper were derived.

When sentiment indicators are included into the model's parameters, the prediction accuracy will improve. However, success in the financial sector will not be achieved by using a wide vocabulary. Consequently, we need to develop a stock-specific emotion lexicon.

Excessive noise, nonlinearity, and high time-variability are three of the complicated aspects of stock price series. Long short-term memory is capable of handling large time series data sets. The LSTM network is a deep learning method that is used for this purpose.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Thirdly, changing the hyperparameters has an effect on how well the LSTM network predicts. Choosing the correct hyperparameters for the network model artificially will be costly. The 2020 Sparrow Search Algorithm might therefore be used to enhance LSTM.

The main contributions of this study are these.

(1) In order to impact stock prices, the SSA-LSTM model incorporates multi-source data, such as historical trade data and comments from stock forums. When compared to the model that relies on a single data source, the MS-SSA-LSTM model produces more accurate predictions.

Read the stock forum's textual comments (2). Our next step is to compile trustworthy financial investment dictionaries into a sentiment language that can be used for sentiment indicator calculation and individual stock sentiment research.

The Sparrow Search Algorithm enhances LSTM models, allowing them to outperform single LSTM networks in terms of hyper parameter values and stock price prediction.

II. LITERATURE SURVEY

The Efficient Market Hypothesis (EMH) and the Random Walk were the foundations of early research on stock market prediction. Several research, including those by Gallagher, Kavussanos, and Butler, demonstrate that stock market values are somewhat predictable and do not move in a random manner [3][7][10]. A further hypothesis that is being investigated is if it is possible to forecast changes in economic and commercial indicators using early indicators that are taken from internet sources (blogs, twitter feeds, etc.). Analysis of the same has been done in different disciplines of inquiry, for e.g. Book sales and online discussion activity were shown to be correlated by Gruhl et al [8]. Mishne, Glance, et al. have used blog sentiment analysis to forecast box office receipts [15]. The relationship between breaking financial news and fluctuations in stock prices was examined by Schumaker et al [18]. Bollen, Mao, et al. (2011) conducted one of the most significant studies in the subject of stock prediction, looking at the relationship between the Dow Jones Industrial Index and public sentiment. Using Twitter feeds, public emotions (happy, calm, and anxious) were determined [2]. By looking at and categorizing the Twitter feeds, Chen and Lazer were able to generate investing strategies [4]. After analyzing the Twitter stream, Bing et al. came to the conclusion that stock prediction was industry-specific [1]. Negative social network emotions and the DJIA index showed a strong negative association, according to Zhang et al [20]. In their research, Pagolu et al. found a significant relationship between the public sentiment expressed on Twitter and the increase or decrease in a company's stock price. Their study concentrated on creating a sentiment analyzer to classify tweets into three categories: Positive, Negative, and Neutral, as opposed to utilizing a typical word embedding approach. [16]. In their work, Mittal et al. attempted to use Twitter sentiment analysis to create a tool for managing portfolios. Using DJIA, they examined and evaluated their model. A day ahead of time, the greedy strategy-based model was able to forecast the Buy/Sell choices for the DJIA holdings based on sentiment research of social media. on their research, Chen et al. compared the LSTM model with the Random estimation model, validating the LSTM model's greater accuracy. The LSTM model was built on the basis of the algorithm, and it was used to forecast the direction of stocks on the Chinese Stock Exchange [4]. A study by Tekin et al. used a variety of forecasting algorithms to analyze data from 25 of the top corporations [19]. Their research demonstrated the greater significance of the Random Forest approach. Malandri et al. use a random forest classifier and an LSTM multi-layer perceptron (MLP) in their portfolio allocation model. According to an analysis of NYSE data, LSTM produces superior experimental outcomes [13]. For the Turkish Stock Exchange (BIST 100), Kilimci et al. presented their research on creating an Effective Word Embedding and Deep Learning Based Model to Forecast the Direction of Stock Exchange Market Using Twitter and Financial News Sites [11]. Their research included a number of word embedding and deep learning models to find the combination that might forecast stocks with the best accuracy. To categorize the information as good or bad, they use data labeling. Following that, the data is sent to word embedding models such as Word2Vec, FastText, and GloVe, which create several word embedding models to be evaluated using the three distinct deep learning algorithms, namely CNN, RNN, and LSTM. Using Twitter data as the foundation, the Word2Vec embedding model and LSTM combo produced the greatest average accuracy among the nine stocks under evaluation.

III. SYSTEM ANALYSIS

EXISTING SYSTEM

Furthermore, in a short-term financial market, Yan et al. [16] created a high-precision prediction model based on the LSTM deep neural network. The BP neural network and ordinary RNN are less accurate in predicting outcomes than the LSTM neural network, which is capable of accurately predicting market values. Ten technical indicators were chosen by Nabipour et al. [17] to be the input for the prediction model. According to the experiment, LSTM fared better at fitting models than other algorithms, such as ANN, RNN, XGBoost, Adaboost, decision tree, and random forest. A CNN-based model was proposed by Aksehir and Kilic [18] to forecast the trading activity of Dow Jones 30 Index stocks the next day. Data on oil prices, gold prices, and technical indicators are input into the algorithm. 3-22% more accurate than previous CNN-based algorithms are the findings.

However, researchers often have to intentionally set the very subjective LSTM network hyperparameter values. The network model's topology is directly controlled by the hyperparameter setting, which has an impact on the prediction performance of the model. A significant amount of computational and human resources will be needed for the artificial selection of the network model's ideal hyperparameters. Researchers attempt to tune neural network hyperparameters using Swarm Intelligence (SI) algorithms in order to address the aforementioned issues. To a certain degree, SI can search worldwide and determine the estimated value of the best answer. The LSTM optimized using the Improved Particle Swarm Optimization model (IPSO) is better to the LSTM model, as Ji et al. [19] showed. By using an Adaptive Genetic Algorithm (AGA) to optimize LSTM, Zeng et al. [20] increased prediction accuracy. Neural network optimization also often makes use of other sophisticated techniques. The sparrows' ability to fend off predators and engage in foraging led to the development of the Sparrow Search Algorithm (SSA) [21]. The SSA method is very new and has a strong optimization ability in price prediction when compared to the conventional particle swarm and gray wolf optimization algorithms. Furthermore, the method takes into account the stochastic nature of people throughout the search process, hence exhibiting strong global search capabilities. In addition, the method is capable of adaptively modifying the search parameters to suit the requirements of various issues.

Compared to other data sources, the information gleaned from the stock forum's real-time content updates is more accurate in reflecting the emotion of investors. But it will have a big impact on investors' investing inclinations, which will then have a significant effect on changes in stock prices. For this reason, it is essential that we examine the textual data from stock forums for our study. When it comes to sentiment categorization techniques, scholars mostly use machine learning and semantic analysis approaches. The machine learning approach has a high classification accuracy but increases time expenses since it requires finance staff to manually categorize training sets. While using a normal dictionary to an economic environment might be difficult, the semantic analysis technique is more user-friendly when used to financial and economic analysis [34]. The secret is to create a distinct financial lexicon.

Disadvantages

• Data complexity: in order to identify stock price prediction, most current machine learning algorithms need to effectively understand big and complicated datasets.

• Access to data: In order for machine learning models to provide reliable predictions, they often need massive datasets. The reliability of the model could be compromised if there is a lack of data in enough amounts.

• Mislabeled data: Current ML models can only learn as much as the data used to train them. The accuracy of the model's predictions is dependent on the accuracy of the data labels.

PROPOSED SYSTEM

In order to forecast stock prices, this study presents a new model (MS-SSA-LSTM) that uses the Sparrow Search Algorithm to combine LSTM neural networks with data from many sources. Traders and investors may use the MS-SSA-LSTM stock price prediction model to help them make more educated investment decisions by predicting the stock price in advance. Investors and traders use the MS-SSA-LSTM model to

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

input data on particular stocks, such as shareholder feedback and historical transaction data. In addition to automatically creating a stock price trend chart, the algorithm also forecasts the stock price for the next day. This is the source from which the ideas for this paper were derived.

When sentiment indicators are included into the model's parameters, the prediction accuracy will improve. However, success in the financial sector will not be achieved by using a wide vocabulary. Consequently, we need to develop a stock-specific emotion lexicon.

Excessive noise, nonlinearity, and high time-variability are three of the complicated aspects of stock price series. Long short-term memory is capable of handling large time series data sets. The LSTM network is a deep learning method that is used for this purpose. (3) Changing the LSTM network's hyperparameters has a direct effect on the model's prediction accuracy. Choosing the correct hyperparameters for the network model artificially will be costly. As a result, LSTM might benefit from the 2020 Sparrow Search Algorithm..

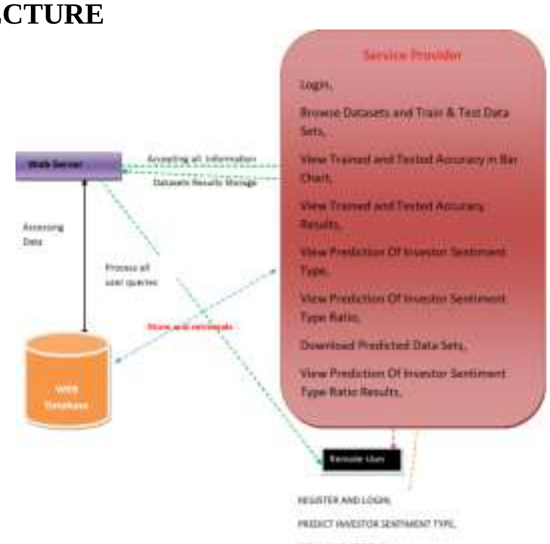
Advantages

(1) Previous transaction data and comments made on stock forums are only two examples of the multi-source data that the SSA-LSTM model utilizes to influence stock prices. When compared to the model that relies on a single data source, the MS-SSA-LSTM model achieves better prediction accuracy.

(2) Review the textual material found in the stock forum comments. Next, we build a sentiment dictionary that is appropriate for calculating sentiment indicators and analyzing particular stocks based on reputable dictionaries in the area of financial investments.

(3) Compared to a single LSTM network, an LSTM model that has been tuned using the Sparrow Search Algorithm may acquire better hyper parameter values and provide more accurate stock price forecasts.

IV. SYSTEM ARCHITECTURE



V. SYSTEM IMPLEMENTATION

MODULES

Service Provider

A valid username and password are required for the Service Provider to access this module. He would be able to do actions like browsing datasets and running tests and training on them when he successfully logs in. Check out the Bar Chart for Trained and Tested Accuracy, Check Out the Results for Trained and Tested Accuracy, View All Remote Users, Download Predicted Data Sets, View Results for Investor Sentiment Type Prediction, and View Investor Sentiment Type Ratio Prediction.

View and Authorize Users

The admin can get a complete rundown of all registered users in this section. Here, the administrator may see the user's information (name, email, and address) and grant them access.

Remote User

All all, there are n users in this module. Registration is required prior to performing any operations. Details will be entered into the database after a user registers. He will need to log in using the permitted username and password when registration is completed. After logging in, users will be able to do things like see their profile, predict the sentiment type of investors, and register and log in.

VI. CONCLUSION

Emotions among shareholders impact the stock price, and the hyper parameters in the LSTM network are often selected on the basis of subjective experience. As a result, the MS-SSA-LSTM stock price prediction model is presented in this work. The model is trained and tested using six sample data sets of individual equities from the Chinese financial market. In addition, four evaluation indicators are used to evaluate the prediction performance of the model. By use of comparison analysis, the following conclusions may be made.

The MS-SSA-LSTM model takes into account a variety of data sources first. The features of the historical transaction data, such as the previous closing price, opening price, closing price, etc., are the data source on the one hand. Conversely, the market's shareholders' attitude serves as the data source. When a mood indicator is used in addition to basic stock trading indicators as input, the model performs better in terms of prediction. As such, the stock bar may serve as a platform for directing investor emotion. The platform managers have the ability to regulate public opinion, provide early warnings, advise shareholders to make reasonable investments, and take proactive actions to prevent panic or rioting. In the end, we need to create a positive and well-organized stock market atmosphere.

Second, getting the best prediction results is difficult as the LSTM parameters must be purposely changed. As a result, the SSA is selected to maximize the hyperparameters of the LSTM model. This method enhances the model's flexibility and forecasting skills in addition to providing an objective explanation of the model's network topology and parameter settings. The LSTM model designed by SSA can deliver high-precision price prediction, reduce investor risk, and rapidly and accurately understand data features in a demanding financial market.

Thirdly, the excellent accuracy, dependability, and stock market adaptability of the MS-SSA-LSTM model are confirmed by the comparative tests conducted on six distinct stocks in various models. The experiment shows that 5–10 time steps are necessary for the model's prediction impact to reach its ideal level, suggesting that short-term predictions are better suited given China's financial market's extreme volatility. Other time series issues may be solved using this approach in the meanwhile.

The MS-SSA-LSTM model is ubiquitous, to sum up. The model is appropriate for international stock markets, even though we only tested it on the Chinese financial market. The technical stock data and the shareholder sentiment index are two of the several sources of data that the model uses as input values. These multi-source data are available in international and Chinese financial markets. Diverse and accessible social media channels are available for debating stock pricing. We can compute the sentiment index and utilize it as one of the model's input features as long as we can locate the stock comments. Additionally, setting up a stock market commenting platform makes it easier and faster for us to get feedback.

Our multi-source data still have constraints in the MS-SSA-LSTM model. This study solely distinguishes between good and negative emotions in terms of sentimental analysis. In addition, additional factors that might affect the stock price projection include changes in policy and macroeconomic circumstances. Sentiment analysis needs to be improved in the future. Different emotional indications, such as grief, fear, rage, and disgust, should be included into our study. In the meanwhile, more data sources for estimating market mood may be added, such official WeChat and Weibo accounts. We should also use mining methods to identify other variables that influence stock price predictions.

REFERENCES

[1] M. M. Rounaghi and F. N. Zadeh, ''Investigation of market efficiency and financial stability between S&P 500 and London stock exchange: Monthly and yearly forecasting of time series stock returns using ARMA model,'' *Phys. A, Stat. Mech. Appl.*, vol. 456, pp. 10–21, Aug. 2016, doi: 10.1016/j.physa.2016.03.006.

[2] G. Bandyopadhyay, ''Gold price forecasting using ARIMA model,'' *J. Adv. Manage. Sci.*, vol. 4, no. 2, pp. 117–121, 2016, doi: 10.12720/joams.4.2.117-121.

[3] H. Shi, Z. You, and Z. Chen, ''Analysis and prediction of Shanghai composite index by ARIMA model based on wavelet analysis,'' *J. Math. Pract. Theory*, vol. 44, no. 23, pp. 66–72, 2014.

[4] H. Herwartz, ''Stock return prediction under GARCH—An empirical assessment,'' *Int. J. Forecasting*, vol. 33, no. 3, pp. 569–580, Jul. 2017, doi: 10.1016/j.ijforecast.2017.01.002.

[5] H. Mohammadi and L. Su, ''International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models,'' *Energy Econ.*, vol. 32, no. 5, pp. 1001–1008, Sep. 2010, doi: 10.1016/j.eneco.2010.04.009.

[6] A. Hossain and M. Nasser, ''Recurrent support and relevance vector machines based model with application to forecasting volatility of financial returns,'' *J. Intell. Learn. Syst. Appl.*, vol. 3, no. 4, pp. 230–241, 2011, doi: 10.4236/jilsa.2011.34026.

[7] J. Chai, J. Du, K. K. Lai, and Y. P. Lee, ''A hybrid least square support vector machine model with parameters optimization for stock forecasting,'' *Math. Problems Eng.*, vol. 2015, pp. 1–7, Jan. 2015, doi: 10.1155/2015/231394.

[8] A. Murkute and T. Sarode, ''Forecasting market price of stock using artificial neural network,'' *Int. J. Comput. Appl.*, vol. 124, no. 12, pp. 11–15, Aug. 2015, doi: 10.5120/ijca2015905681.

[9] D. Banjade, ''Forecasting Bitcoin price using artificial neural network,'' Jan. 2020, doi: 10.2139/ssrn.3515702.

[10] J. Zahedi and M. M. Rounaghi, ''Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran stock exchange,'' *Phys. A, Stat. Mech. Appl.*, vol. 438, pp. 178–187, Nov. 2015, doi: 10.1016/j.physa.2015.06.033.

[11] A. H. Moghaddam, M. H. Moghaddam, and M. Esfandyari, ''Stock market index prediction using artificial neural network,'' *J. Econ., Finance Administ. Sci.*, vol. 21, no. 41, pp. 89–93, Dec. 2016, doi: 10.1016/j.jefas.2016.07.002.

[12] H. Liu and Y. Hou, ''Application of Bayesian neural network in prediction of stock time series,'' *Comput. Eng. Appl.*, vol. 55, no. 12, pp. 225–229,2019.

[13] A. M. Rather, A. Agarwal, andV. N. Sastry, ''Recurrent neural network and a hybrid model for prediction of stock returns,'' *Expert Syst. Appl.*, vol. 42,no. 6, pp. 3234–3241, Apr. 2015, doi: 10.1016/j.eswa.2014.12.003.

[14] A. Sherstinsky, ''Fundamentals of recurrent neural network(RNN) and long short-term memory (LSTM) network,'' *Phys. D,Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306, doi:10.1016/j.physd.2019.132306.

[15] G. Ding and L. Qin, ''Study on the prediction of stock price based on the associated network model of LSTM,'' *Int. J. Mach. Learn. Cybern.*, vol. 11,no. 6, pp. 1307–1317, Nov. 2019, doi: 10.1007/s13042-019-01041-1.

[16] X. Yan, W. Weihan, and M. Chang, ‘‘Research on financial assets transaction prediction model based on LSTM neural network,’’ *Neural Comput. Appl.*, vol. 33, no. 1, pp. 257–270, May 2020, doi: 10.1007/s00521-020-04992-7.

[17] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, and S. Shahab, ‘‘Deep learning for stock market prediction,’’ *Entropy*, vol. 22, no. 8, p. 840, Jul. 2020, doi: 10.3390/e22080840.

[18] Z. D. Aksehir and E. Kiliç, ‘‘How to handle data imbalance and feature selection problems in CNN-based stock price forecasting,’’ *IEEE Access*, vol. 10, pp. 31297–31305, 2022, doi: 10.1109/ACCESS.2022. 3160797.

[19] Y. Ji, A. W. Liew, and L. Yang, ‘‘A novel improved particle swarm optimization with long-short term memory hybrid model for stock indices forecast,’’ *IEEE Access*, vol. 9, pp. 23660–23671, 2021, doi: 10.1109/ACCESS.2021.3056713.

[20] X. Zeng, J. Cai, C. Liang, and C. Yuan, ‘‘A hybrid model integrating long short-term memory with adaptive genetic algorithm based on individual ranking for stock index prediction,’’ *PLoS ONE*, vol. 17, no. 8, Aug. 2022, Art. no. e0272637, doi: 10.1371/journal.pone.0272637.