# Machine Learning Models for Statistical Analysis

Marko Grebovic
Faculty for Information Systems and Technologies,
University of Donja Gorica
Podgorica, Montenegro

Luka Filipovic
Faculty for Information Systems and Technologies,
University of Donja Gorica
Podgorica, Montenegro

Ivana Katnic
Faculty of International Economics, Finances and Business,
University of Donja Gorica
Podgorica, Montenegro

Milica Vukotic
Faculty for Information Systems and Technologies,
University of Donja Gorica
Podgorica, Montenegro

Tomo Popovic
Faculty for Information Systems and Technologies,
University of Donja Gorica
Podgorica, Montenegro

**Abstract:** *Compared to traditional statistical models, Machine Learning (ML) algorithms provide the ability to interpret, understand and summarize patterns and regularities in observed data for making predictions in an advanced and more sophisticated way. The main reasons for the advantage of ML methods in making predictions are a small number of significant predictors of the statistical models, which means limited informative capability, and pseudo-correct regular statistical patterns, used without previous understanding of the used data causality. Also, some ML methods, like Artificial Neural Networks, use non-linear algorithms, considering links and associations between parameters. On the other hand, statistical models use one-step-ahead linear processes to improve only short-term prediction accuracy by minimizing a cost function. Although designing an optimal ML model can be a very complex process, it can be used as a potential solution for making improved prediction models compared to statistical ones. However, ML models will not automatically improve prediction accuracy, so it is necessary to evaluate and analyze several statistical and ML methods, including some artificial neural networks, through accuracy measures for prediction purposes in various fields of applications. A couple of techniques for improving suggested ML methods and artificial neural networks are proposed to get better accuracy results.*

**Keywords:** *Machine learning, artificial neural networks, statistical models, accuracy measures.*

*Received April 3, 2023; accepted May 10, 2023*
*https://doi.org/10.34028/iajit/20/3A/8*

## 1. Introduction

Statistical models are basic tools that sum up patterns of the used data [50]. They are used for discovering causality, making predictions, and describing different events [41]. Statistical models are often recognized as tools for uncovering causality in different scientific areas, but achieving this goal is a serious challenge [35]. On the other hand, prediction and description are more practical reasons. Regardless of the statistical model's task, links and associations between variables in complex systems are ignored [6].

For prediction, researchers sometimes use statistical models that provide a regular pattern that seems to hold statistically without a previous understanding of causal mechanisms in the observed data [50]. Prediction models based on statistical methods like regression modeling only include a few essential predictors, so they have limited informative capability [15].

Traditional statistical models have a low scientific value in predicting the observed data. Gained prediction results are only an overview of statistical information in data of interest which should be only interpreted but not devalued. Descriptive statistical models are involved in all causal interpretations because they cannot conclude causality by themselves [50]. Descriptive results are observed information disguised in a mathematical form that cannot be wrong. However, descriptive interpretation is not the primary goal, but making predictions and understanding causality should be the ultimate one [39]. Achieving this with the usually available dataset takes much work.

Artificial Intelligence (AI) systems are based on algorithms that provide learning by example and errors, improving their performance over time [32]. Machine Learning (ML), as a particular class of AI, allows data interpretation and understanding more sophisticatedly. ML methods have been becoming more important with rising interest in AI and can be exploited to develop prediction models based on experience and to improve existing time series predictions [31]. ML methods have become very important over the last years through many applications like autonomous traffic intelligent systems [42], power consumption prediction [49], credit-card fraud detection [3], behavior and facial expression recognition [16, 27, 28], image recognition systems for diagnosis [4, 7, 8, 36, 37], crop disease prediction [40], and sign language recognition [1, 17].

ML approach provides unbiased robust prediction

models and cooperation between parameters affecting the outcomes consequentially [6]. As the alternative to traditional statistics, ML algorithms propose new complex models, which also consider links and associations between parameters suggesting methodological advances and accuracy improvements [14, 44, 52]. For example, artificial Neural Networks (NN) are ML models that imitate the learning process in the brain. Namely, the neural cells have a hierarchical structure where each input node receives multiple inputs, giving each of them a weight, resulting in outputs in the shape of a decision determined by the weighted data gained sum.

ML and statistical methods aim to improve prediction accuracy by minimizing cost functions like the root of medium square error [19]. The main difference is in the minimization methods: ML methods offer some solutions like NN that use non-linear algorithms, but on the other hand, statistical methods use linear processes. It makes ML methods computationally complex and more dependent on computer science for implementation. Also, statistical predictions use a single or few time series, making the results' importance and generalization unreliable. Statistical methods are estimated one step ahead for short-term predictions without considering medium and long-term predictions [32].

However, designing an ML model, especially a NN, can be complex. Determining parameters like the number of hidden layers and nodes in them is complex and requires much time. Sometimes, artificial neural networks must improve the interpretability of variable weights obtained in the model-developing process. Conversely, statistical models allow the interpretation of individual coefficients (parametric assumptions), which is crucial for making conclusions in prediction problems [38].

It is also necessary to emphasize that ML methods will not automatically improve prediction accuracy because they can generate implausible solutions. Before claiming that ML, and especially NN, bring improvement in terms of making predictions compared to traditional statistics, it is necessary to compare several methods, using different performance measures, on a couple of datasets. In other words, it is necessary to answer whether ML methods can be trained to make more efficient and accurate predictions, as opposed to statistical ones, by using more information about the future rather than past events.

## 2. Datasets

For this paper, three different datasets are used. The first is from PREDISE, a web-based study investigating how different factors are associated with healthy eating habits like Vegetable and Fruit (VF) intake among adults [9]. There are two possible outcomes (classes) of VF intake: adequate VF consumption, corresponding to 5 or more servings per day, and inadequate VF consumption, corresponding to less than five servings per day. The study is based on data about 1147 male and female participants between 18 and 65, containing information regarding individual, social, and environmental factors, three 24 h dietary recalls and food intake frequency, anthropometric measurements, and blood sampling. Of all the participants, 1083 completed all three, 34 completed two, and 30 completed only one recall [29]. VF intake in servings per day was calculated by averaging intakes from all recalls available. Anthropometric measurements and blood sampling data from the clinical assessment include serum cholesterol, triglycerides, HDL-cholesterol, fasting blood glucose and insulin concentrations, systolic and diastolic blood pressures, measured height, measured weight, body mass index, body fat percentage, and waist circumference.

The second dataset used is based on one of the studies assisting clinicians when offering couples personalized treatment options for undergoing In Vitro Fertilization (IVF). It contains fertilization data about a population of 1136 participating couples/patients, including several fertilization outcomes leaning on factors such as clinical features, age, and Body Mass Index (BMI): number of oocytes retrieved, number of mature oocytes, number of fertilized oocytes, number of top-quality embryos, positive beta-hCG, clinical pregnancies, and live births [12, 43, 45].

The third dataset includes 1045 out of 3003 time series from the M3 Competition forecast study [2]. The original dataset includes various types of time series data (micro, industry, macro, finance, demographic, and other) and different time intervals between successive observations (yearly, quarterly, monthly, and other). A minimum number of observations is set for each data type to enable the development of an adequate forecasting model. All the time series data have positive values because, in the case of a negative, it was substituted by zero to avoid any problem with the advanced performance measures.

Before analyzing forecast prediction accuracy measures, it is recommended to perform data preprocessing to provide stable processes and optimal results. In this case, the forecast data preprocessing includes three steps: Seasonal adjustments, power transformations, and trend removal. Firstly, the multiplicative decomposition removes seasonal determinants from the data [33]. After that, obtained forecasts are reversed to seasonal form to make the final predictions. Exceptions are cases when methods include seasonal models and their complexity, defined through the information criteria and comparative tests. The Box-Cox power transformation is applied to the original forecast dataset to accomplish variance stationarity. Eliminating the trend in data series is essential in cases of bouncing activation function, making it more stable [10]. Determining the most suitable trend elimination

method in forecast data and achieving mean stationarity requires more empirical methods like the Cox-Stuart test. It is executed to establish if a deterministic linear trend or first differencing should be used [32]. Combining described preprocessing techniques is recommended for getting even better results.

## 3. Methods and models

Traditional statistical models for prediction in classification tasks on the first dataset are Logistic Regression (LR) and penalized regression (Lasso). The LR model calculates the probability of belonging to one of two classes by computing the logit function of weighted input features, which are estimated using maximum-likelihood estimation [23]. Lasso model uses feature selection and shrinkage to reduce the number of features for classification purposes [48]. Also, four commonly known supervised ML classification algorithms were applied: Decision Tree (DT), Random Forest (RF), Support Vector Machines (SVM) with linear and polynomial kernels, and K-Nearest Neighbors (KNN). DT algorithm has a flowchart-like structure that makes predictions by learning decision rules where each node represents an input feature which is compared through each branch (decision rule) until a leaf node (prediction) is reached [46]. RF algorithm generates a large ensemble of decision trees where the predicted class is determined by averaging the estimated outcome variable of each decision tree [51]. The SVM algorithm attempts to sort the data between two classes with a hyperplane which can either be a linear or a polynomial function determined using only the points closest to the hyperplane [24]. KNN algorithm assumes that close data points are similar, so the class of a new data point is determined according to the shared characteristics of a pre-determined number of closest points [11]. In the second dataset, the NN model is applied instead of the suggested ML methods to get improved results [21].

Research work in forecasting, like the M3 Competition study, is based on utilizing ML, especially NN methods, on making time-series predictions. The idea is to compare the performances of NN and statistical methods to determine if NN methods can improve the accuracy of Statistical Methods (SM) and issues of improving forecast accuracy in general [32]. Several advanced NN and SM are used for more precise and reliable analysis. Method Simple Exponential Smoothing (SES) aims to predict trendless series, while Holt and Damped Exponential Smoothing are adequate for trend time series [18]. The next model, Theta, achieves very good overall sMAPE [5], and finally, for Exponential Smoothing (ETS) model, which provides substantial accuracy [25].

On the other hand, NN methods used are Multi-Layer Perceptron (MLP), Bayesian Neural Networks (BNN), Generalized Regression Neural Networks (GRNN), Recurrent Neural Networks (RNN), and Long Short Term Memory neural networks (LSTM). Multi-Layer Perceptron (MLP) consists of constructing a single hidden NN layer, defining the optimal number of input nodes N and hidden nodes to 2N+1, using a k-fold validation process [30]. Defining optimal weights using the Scaled Conjugate Gradient instead of Standard Backpropagation is recommended because it performs better in weight optimization tasks [34]. It is recommended to use a hidden layers' logistic activation function and an output nodes' linear function to maximize the method's flexibility [32]. The main reason is that a logistic output activation function is bounded for optimizing time series with trends and can easily flop. In the case of nonlinear activation functions, it is recommended to scale the data values within the interval from 0 to 1. The main reasons for data scaling are: keeping away from problems in performing computations, providing quicker network learning, and meeting the algorithm's requirements. Bayesian Neural Network (BNN) has many similarities with the MLP method. The only difference is that process of optimizing network parameters is based on the Bayesian concept, where the weights are evaluated with distributions of errors suspected in advance. It is designed and built through: the Nguyen and Widrow algorithm for assigning initial weights, the Gauss-Newton algorithm for the optimization, the k-fold validation process for determining the optimal number of input nodes N and the hidden nodes, which is defined with 2N+1, while an overall number of considered iterations is linearly scaled [13]. Generalized Regression Neural Networks (GRNN) is a nonparametric method where predictions are obtained by calculating a mean value for all differences between each training data point's target output and the respective observed values. For this purpose, it is also necessary to calculate the parameter sigma, which represents the fit smoothness. Finally, the number of inputs N in the k-fold validation should be determined [47]. Recurrent Neural Network (RNN) is similar to the MLP. However, the main difference is that for each output, RNN provides feedback connections for checking earlier states, which are utilized together with each contemporary input. Feedback connections are created by copying previous values in the recurrent nodes layer. RNN usually consists of a hidden layer with recurrent nodes and an output layer with linear node(s). Besides using k-fold validation, in terms of defining the optimal NN structure, it is possible to use a couple of input and recurrent nodes within the hidden layer for all available time series. Several input nodes and recurrent units are selected using results from a random time series sample with the best parameterization performance [32]. LSTM neural network is sequential and contains hidden and output layers like RNN, but with the additional task which includes avoiding the dependency on the long-term scale. Because of the complex architecture, another advantage is the ability to keep information longer [22].

## 4. Performance Measures

In the first example, it is necessary to compare the used statistical and ML models' performance in making predictions, using predictive metrics based on simple rates of accurate and inaccurate predictions like accuracy, precision, recall, and F1 score. Accuracy measures the rate of accurate predictions on the entire population. The F1 score represents the balance between precision and recall and is measured like their harmonic mean, which assigns more weight to lower, compared to the regular mean that treats all values equally. Precision represents the ratio of true positive to the sum of true positive and false positive predictions (accuracy of positive predictions). At the same time, recall is the ratio of true positives to the sum of true positive and false negative predictions (rate of correctly detected instances) [19]. For the second dataset, it is sufficient to compare the accuracy and F1 scores of the suggested methods.

Considering that many forecast tasks are about predicting numerical values, sometimes it is more suitable to evaluate predictive methods by calculating the difference between observed and predicted values, as in the third example. One example is Root Mean Square Error (RMSE), which represents system predictions error. It does not treat all errors equally but gives higher weight to large ones. Also, in some contexts, it is preferred to use the Mean Absolute Error (MAE), also called Average Absolute Deviation [19]. However, for this paper, it is suitable to use more precise measures. One of them is the symmetric Mean Absolute Percentage Error (sMAPE), which is defined with the following equation [20]:

$$sMAPE = \frac{2}{k}\sum_{i=1}^{k}\frac{|Y_i-\hat{Y}_i|}{|Y_i|+|\hat{Y}_i|} \quad (1)$$

Where $k$ counts instances in the observed data, $Y_i$ is the vector of actual observations, for instance $i$, and $\hat{Y}_i$ is the model's prediction vector, for instance $i$. As the opposite of sMAPE, which does not consider positive and negative errors equally but favors large positive ones, the Mean Absolute Scaled Error (MASE) is introduced to complement the former [25]:

$$MASE = \frac{1}{k}\frac{\sum_{i=1}^{k}|Y_i-\hat{Y}_i|}{\frac{1}{n-m}\sum_{i=m+1}^{n}|Y_i-Y_{i-m}|} \quad (2)$$

Where $n$ counts available observations and $m$ measures time series incidence. Also, it is worth mentioning that MASE is data scale independent. For representative prediction accuracy estimation, $\hat{Y}_i$ should be computed at least ten times. Additionally, the produced errors average should be utilized to avoid issues induced by choosing specific initial values for ML methods parameters. It is also important to measure the precision of Model Fitting (MF) into the observed data:

$$MF = \frac{n\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{\left(\sum_{i=1}^{n}Y_i\right)^2} \quad (3)$$

MF represents the Mean Squared Error of $n-k$ model fit forecasts, normalized by the examined time series mean [32].

## 5. Results and Discussion

ML methods do not necessarily perform better in predistion tasks, compared to traditional statistics. For example, comparing accuracy measures of these two groups of prediction methods give similar results in PREDISE study. As the proposed ML methods do not give better results than the statistical ones, the idea is to implement more advanced models like neural networks and try them on a data set of a similar size but with more labels.

In the second example, considering couples' personalized treatment options for undergoing In Vitro Fertilization, NN methods show a clear advantage compared to traditional statistics. Table 2 displays accuracy rates and F1 scores of NN and LR methods for In Vitro Fertilization (IVF) outcomes on the previously described population. Figure 2 graphically represents data from Table 2 and clearly shows the superiority of NN methods: Their accuracy outcomes are between 0.69 and 0.90, while LR accuracies are between 0.34 and 0.74. NN F1 scores are also higher than LR, ranging between 0.69-0.89 and 0.35-0.74 for NN and LR, respectively.

Table 1. ML and statistical accuracy measures for VF intake habits.

| Accuracy measures | Statistical methods | | Machine Learning | | | | |
|---|---|---|---|---|---|---|---|
| | *LR* | *LASSO* | *DT* | *RF* | *SVM* | | *KNN* |
| | | | | | *Linear* | *Polyn.* | |
| Accuracy | 0.64 | 0.64 | 0.62 | 0.64 | 0.55 | 0.64 | 0.58 |
| Precision | 0.65 | 0.65 | 0.63 | 0.63 | 0.57 | 0.64 | 0.58 |
| Recall | 0.68 | 0.68 | 0.66 | 0.73 | 0.58 | 0.72 | 0.69 |
| F1 Score | 0.66 | 0.66 | 0.64 | 0.68 | 0.58 | 0.67 | 0.63 |

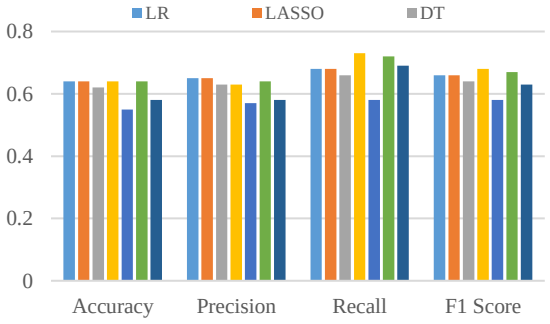


Figure 1. Performance measures of ML and statistical methods for VF intake habits.

One of the main reasons for the apparent superiority of NN as an ML algorithm over LR as a classic statistical method is taking complex and non-linear links among various parameters into consideration and their better utilization, involving parameters that are not connected with the outcomes opposed to classical statistical

models. Also, the IVF outcomes study has several strengths, which are crucial for excellent performances achieved by NN: It emphasizes the importance of modifiable variables, the dataset is collected from a homogeneous population undergoing the same IVF protocol is used, subjective factors which might affect IVF outcomes are excluded, enabling noise reduction and developing high-performing robust models [6]. Based on this, an approach where models are developed based on the training set and tested on various sample sets provides reliable results. It also considers the models' performances and is known as the training-test approach.

Regardless of mentioned strengths, the IVF outcomes study has several limitations, including a relatively limited sample of participants and a need for more external testing and validation data. In addition, much better rates of prediction accuracy for NN compared to LR could indicate that NN methods offer improved performances for making predictions or that LR is far from good in this task. Also, in many studies, the validity of assumptions in traditional statistical methods, which is fundamental for the quality and performance of predictions for used datasets, is only confirmed after application [38]. Before claiming that ML, especially NN, can overcome the shortcomings of traditional statistics, it is necessary to compare several NN algorithms with advanced statistical methods using sophisticated performance measures.

Table 2. NN and LR accuracy measures for IVF outcomes.

| **IVF outcome** | **Accuracy measures** | | | |
|---|---|---|---|---|
| | ***Accuracy*** | | ***F1 Score*** | |
| | ***NN*** | ***LR*** | ***NN*** | ***LR*** |
| Retrieved oocytes | 0.69 | 0.34 | 0.69 | 0.35 |
| Mature oocytes | 0.88 | 0.74 | 0.87 | 0.74 |
| Fertilized oocytes | 0.78 | 0.55 | 0.77 | 0.56 |
| Top-quality embryos | 0.86 | 0.61 | 0.85 | 0.60 |
| Positive beta-hCG | 0.85 | 0.53 | 0.84 | 0.43 |
| Clinical pregnancy | 0.90 | 0.58 | 0.89 | 0.46 |
| Live birth | 0.87 | 0.55 | 0.86 | 0.36 |

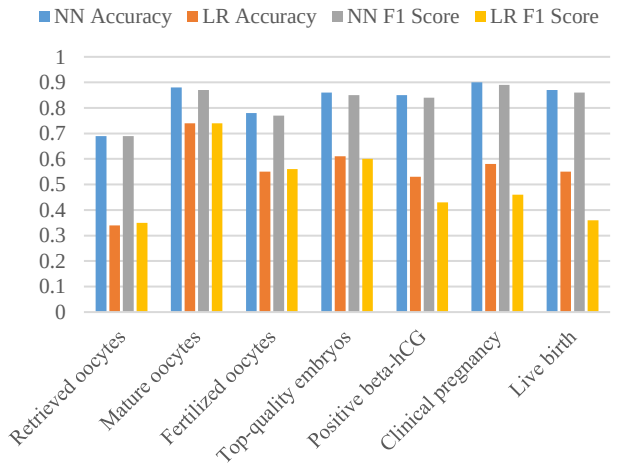


Figure 2. Accuracy rate and F1 scores of NN and LR for IVF outcomes.

sMAPE, MASE, and MF forecast prediction accuracy measures, calculated on the M3 Competition dataset for NN and statistical methods, are shown in Table 3 and 4, respectively. Gained results, represented by the scatter diagram in Figure 3, are not encouraging for NN methods: in two (sMAPE and MASE) out of three categories, excluding only Model Fitting, SM gives better results, as the goal is to minimize all three cost functions. It is necessary to determine the causes for poor NN performance and improve their accuracy side by side with introducing new NN methods [2].

Identifying the cause of NN methods underperforming in forecast tasks includes comparing their accuracy with SM, one series at a time, and explaining the observed. An even more important issue is to answer the question if NN methods can be adjusted to learn more efficiently using additional information about the future, including unknown errors. In this context, learning means finding a solution to an optimization problem. The goal is to choose a set of parameters to minimize a cost function, which is usually based on the sum of square errors. It requires an approach to future information in order to minimize future errors. One way to achieve this is dividing the data set into two parts: one containing the 1/3 of the data set used for developing a much simpler model and the second one with the remaining 2/3 for training to learn predicting remaining observations with expanding each set for one observation in every iteration until all available observations are used. Another way is to provide alternative forecasts and instruct NN methods to learn to select the most accurate one for each data series by minimizing errors. Sometimes, it is required to divide the data into different categories or types of series and deploy adequate models for each category [32].

Table 3. Performances of selected NN models.

| **Method** | **ID** | **Performance measure** | | |
|---|---|---|---|---|
| | | ***sMAPE*** | ***MASE*** | ***MF*** |
| MLP | 1 | 8.39 | 0.55 | 2.11 |
| BNN | 2 | 8.17 | 0.53 | 2.11 |
| GRNN | 3 | 9.49 | 0.67 | 1.80 |
| RNN | 4 | 9.48 | 0.54 | 1.98 |
| LSTM | 5 | 11.67 | 0.72 | 1.84 |

Table 4. Performances of selected statistical models.

| **Method** | **ID** | **Performance measure** | | |
|---|---|---|---|---|
| | | ***sMAPE*** | ***MASE*** | ***MF*** |
| SES | 1 | 7.36 | 0.49 | 2.37 |
| HES | 2 | 7.41 | 0.48 | 2.35 |
| DES | 3 | 7.30 | 0.48 | 2.34 |
| Theta | 4 | 7.31 | 0.48 | 2.34 |
| ETS | 5 | 7.19 | 0.47 | 2.28 |

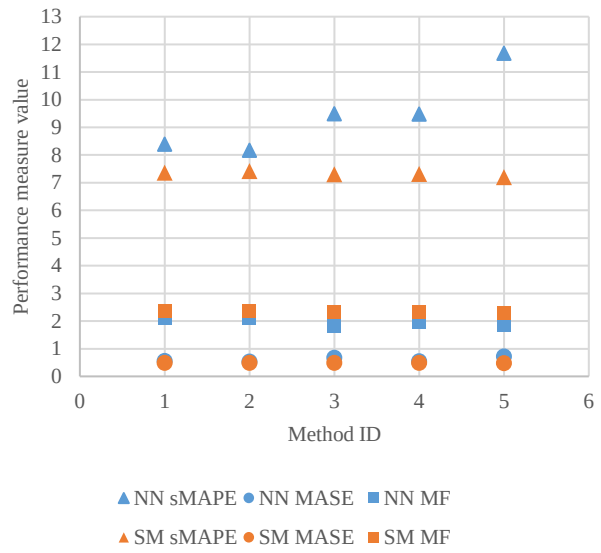


Figure 3. Performance measures for M3 Competition time series.

Proposed techniques for obtaining better future information and minimizing future errors can lead the NN model to excessive fitting into observed data. In addition to the fact that NN methods have nonlinear nature and dynamical training, randomness extent in time series and the capability to differentiate the useful patterns from the noise can also cause model overfitting. In opposition, the overfitting of linear statistical methods can be controlled by setting the number of parameters utilized or information criteria.

## 6. Conclusions

One of the essential ML features, including NN, is the ability to estimate different types of non-linear functions. It is valuable in cases of unknown or complex relationships between variables. On the other hand, traditional statistical prediction models do not include enough significant predictors, so they have limited informative capability.

However, it is necessary to remember that designing an optimal ML model, especially an artificial neural network, is a very complex process when it comes to defining the number of hidden layers and nodes in each of them. Compared to traditional statistical methods, which allow the interpretation of individual coefficients, ML methods need help interpreting variable weights gained through the model development.

Putting aside that designing an optimal ML model, and especially an artificial neural network is a very complex process and that they lack in interpretability of variables weights, their use is justified in examples where rates of accuracy measures are much better in favor of NN methods, like in IVF predictions analysis. Although NN methods offer improved performance for making IVF predictions, it is necessary to consider that LR, as the traditional statistical method, could be better in this task. That study has a relatively limited sample of training and validation data.

Before claiming that NN methods can overcome shortcomings and improve traditional statistical methods' prediction accuracy, it is necessary to compare several NN algorithms with more advanced statistical methods. For this purpose, several NN and statistical methods are evaluated and analyzed for prediction purposes in making forecasts. Comparisons of accuracy measures for several NN and statistical methods used in forecasting indicate that this can easily be the case. Statistical methods give better results for data preprocessed using seasonal adjustments, power transformations, and detrending in two out of three categories.

Discovering the causes for weak NN performance includes comparing their accuracy with those of statistical methods, one series at a time, and describing the observed distinctions. Improving their accuracy means finding ways to improve learning efficiency with additional future information, including unknown errors. One is to iteratively divide the data set into parts to develop a simpler model (1/3). The second is for training to learn predicting remaining observations (2/3) until all available observations are used. Another way is to provide alternative forecasts and learn to select the most accurate one for each data series by minimizing errors, clustering the data into different categories or time series types, and building different models for each. The proposed techniques can lead to overfitting because of the time-series random nature, the capability of differentiating the valuable data patterns and the noise, and the fact that NN has non-linear nature and dynamic training.

The summative conclusion is that ML and especially NN methods can be seriously considered a potential improvement for traditional statistical methods, despite their high design complexity. However, instead of their uncritical use, it is necessary to ensure that the used dataset is optimal in terms of size and availability of external data to validate the results and to carry out a detailed analysis of several ML and statistical methods using adequate performance measures.

## Acknowledgment

This research was supported in part by the EUROCC project, European High-Performance Computing Joint Undertaking (JU) under grant agreement No. 951732. The JU receives support from the EU's Horizon 2020 research and innovation programme and EUROCC project participating institutions.

## References

[1] Agrawal S., Jalal A., and Tripathi R., "A Survey on Manual and Non-Manual Sign Language Recognition for Isolated and Continuous Sign," *International Journal of Applied Pattern*

*Recognition*, vol. 3, no. 2, pp. 99-134, 2016. https://doi.org/10.1504/IJAPR.2016.079048

[2] Ahmed N., Atiya A., El Gayar N., and El-Shishiny H., "An Empirical Comparison of Machine Learning Models for Time Series Forecasting," *Econometric Reviews*, vol. 29, no. 5-6, pp. 594-621, 2010. https://doi.org/10.1080/07474938.2010.481556.

[3] Al Balawi S. and Aljohani N., "Credit-card Fraud Detection System using Neural Networks," *The International Arab Journal of Information Technology*, vol. 20, no. 2, pp. 234-241, 2023. https://doi.org/10.34028/iajit/20/2/10.

[4] Alzu'bi A., Najadat H., Eyadat W., Al-Mohtaseb A., and Haddad H., "A New Approach for Detecting Eosinophils in the Gastrointestinal Tract and Diagnosing Eosinophilic Colitis," *The International Arab Journal of Information Technology*, vol. 18, no. 4, pp. 596-603, 2021. https://doi.org/10.34028/18/4/12.

[5] Assimakopoulos V. and Nikolopoulos K., "The Theta Model: A Decomposition Approach To Forecasting," *International Journal of Forecasting*, vol. 16, no. 4, pp. 521-530, 2000. https://doi.org/10.1016/S0169-2070(00)00066-2

[6] Barnett-Itzhaki Z., Elbaz M., Butterman R., Amar D., Amitay M., *et al.* "Machine Learning Vs. Classic Statistics for the Prediction of IVF Outcomes," *Journal of Assisted Reproduction and Genetics*, vol. 37, no. 10, pp. 2405-2412, 2020.

[7] Berin Jones C. and Murugamani C., "Malaria Parasite Detection on Microscopic Blood Smear Images with Integrated Deep Learning Algorithms," *The International Arab Journal of Information Technology*, vol. 20, no. 2, pp. 170-179, 2023. https://doi.org/10.34028/iajit/20/2/3.

[8] Chandra Bose A. and Ramesh V., "Highly Accurate Grey Neural Network Classifier for an Abdominal Aortic Aneurysm Classification Based on Image Processing Approach," *The International Arab Journal of Information Technology*, vol. 20, no. 2, pp. 215-223, 2023. https://doi.org/10.34028/iajit/20/2/8.

[9] Côté M., Osseni M., Brassard D., Carbonneau E., Robitaille J., et al., "Are Machine Learning Algorithms More Accurate in Predicting Vegetable and Fruit Consumption Than Traditional Statistical Models? An Exploratory Analysis," *Frontiers in Nutrition*, vol. 9, 2022. doi: 10.3389/fnut.2022.740898.

[10] Cottrell M., Girard B., Girard Y., Mangeas M., and Muller C., "Neural Modeling for Time Series: A Statistical Stepwise Method for Weight Elimination," *IEEE Transactions on Neural Networks*, vol. 6, no. 6, pp. 1355-1364, 1995. DOI: 10.1109/72.471372.

[11] Cunningham P. and Delany S., "k-Nearest Neighbour Classifiers 2nd Edition (with python examples)," *arXiv arXiv.*, 2020. https://doi.org/10.48550/arXiv.2004.04523

[12] Curchoe C. and Bormann C., "Artificial Intelligence and Machine Learning for Human Reproduction and Embryology Presented at ASRM and ESHRE 2018," *Journal of Assisted Reproduction and Genetics*, vol. 36, no. 4, pp. 591-600, 2019. DOI: 10.1007/s10815-019-01408-x.

[13] Dan Foresee F. and Hagan M., "Gauss-Newton Approximation to Bayesian Learning," *IEEE International Conference on Neural Networks*, vol. 3, pp. 1930-1935, 1997. DOI: 10.1109/ICNN.1997.614194.

[14] Deng L., "A Tutorial Survey of Architectures, Algorithms, and Applications for Deep Learning-ERRATUM," *APSIPA Transactions on Signal and Information Processing*, vol. 3, 2014. DOI: https://doi.org/10.1017/ATSIP.2014.4.

[15] Desai R., et al., "Comparison of Machine Learning Methods with Traditional Models for Use of Administrative Claims with Electronic Medical Records to Predict Heart Failure Outcomes," *JAMA Network Open*, vol. 3, no. 1, pp. 1-15, 2020. doi:10.1001/jamanetworkopen.2019.18962

[16] Elgendy F., Alshewimy M., and Sarhan A., "Pain Detection/Classification Framework including Face Recognition based on the Analysis of Facial Expressions for E-Health Systems," *The International Arab Journal of Information Technology*, vol. 18, no. 1, pp. 125-132, 2021. https://doi.org/10.34028/iajit/18/1/14.

[17] Fakhfakh S. and Ben Jemaa Y., "Deep Learning Shape Trajectories for Isolated Word Sign Language Recognition," *The International Arab Journal of Information Technology*, vol. 19, no. 4, pp. 660-666, 2022. https://doi.org/10.34028/iajit/19/4/10.

[18] Gardner E.S., "Exponential Smoothing: The State of the art-Part II," *International Journal of Forecasting*, vol. 22, no. 4, pp. 637-666, 2006. https://doi.org/10.1016/j.ijforecast.2006.03.005.

[19] Géron A., Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, *O'Reilly Media*, 2019.

[20] Goodwin P. and Lawton R., "On the Asymmetry of the Symmetric MAPE," *International Journal of Forecasting*, vol. 15, no. 4, pp. 405-408, 1999. https://doi.org/10.1016/S0169-2070(99)00007-2.

[21] Grebovic M., Filipovic L., Katnic I., Vukotic M., and Popovic T., "Overcoming Limitations of Statistical Methods with Artificial Neural Networks," *in Proceedings of the 23th International Arab Conference on Information Technology*, Abu Dhabi, pp. 1-6, 2022. DOI: 10.1109/ACIT57182.2022.9994218.

[22] Hochreiter S. and Schmidhuber J., "Long Short-

Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.

[23] Hosmer D., Lemeshow S., and Sturdivant R., *Applied Logistic Regression*, John Wiley and Sons, 2013. DOI:10.1002/9781118548387.

[24] Howley T. and Madden M.G., "The Genetic Kernel Support Vector Machine: Description And Evaluation," *Artificial Intelligence Review*, vol. 24, pp. 379-395, 2005.

[25] Hyndman R. and Koehler A., "Another Look at Measures of Forecast Accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679-688, 2006. https://doi.org/10.1016/j.ijforecast.2006.03.001.

[26] Hyndman R., Koehler A., Snyder R., and Grose S., "A State Space Framework For Automatic Forecasting Using Exponential Smoothing Methods," *International Journal of Forecasting*, vol. 18, no. 3, pp. 439-454, 2002. https://doi.org/10.1016/S0169-2070(01)00110-8.

[27] Jaouedi N., Boujnah N., and Bouhlel M., "A Novel Recurrent Neural Networks Architecture for Behavior Analysis," *The International Arab Journal of Information Technology*, vol. 18, no. 2, pp. 133-139, 2021. https://doi.org/10.34028/iajit/18/2/1.

[28] Kumar Y., Verma S., and Sharma S., "Multi-Pose Facial Expression Recognition Using Hybrid Deep Learning Model with Improved Variant of Gravitational Search Algorithm," *The International Arab Journal of Information Technology*, vol. 19, no. 2, pp. 281-287, 2022. https://doi.org/10.34028/iajit/19/2/15.

[29] Lafrenière J., Lamarche B., Laramée ., Robitaille J., and Lemieux S., "Validation of a Newly Automated Web-Based 24-Hour Dietary Recall Using Fully Controlled Feeding Studies," *BMC Nutrition*, vol. 3, pp. 1-10, 2017. https://doi.org/10.1186/s40795-017-0153-3.

[30] Lippmann R., "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine*, vol. 4, no. 2, pp. 4-22, 1987. DOI: 10.1109/MASSP.1987.1165576.

[31] Makridakis S., "The Forthcoming Artificial Intelligence (AI) Revolution: its Impact on Society And Firms," *Futures*, vol. 90, pp. 46-60, 2017. https://doi.org/10.1016/j.futures.2017.03.006.

[32] Makridakis S., Spiliotis E., and Assimakopoulos V., "Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward," *PLoS ONE*, vol. 13, no. 3, pp. 1-26, 2018. https://doi.org/10.1371/journal.pone.0194889.

[33] Makridakis S., Wheelwright S., and Hyndman R., *Forecasting: Methods and Applications*, John Wiley and Sons, 2008.

[34] Møller M., "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning," *Neural networks*, vol. 6, no. 4, pp. 525-533, 1993. https://doi.org/10.1016/S0893-6080(05)80056-5.

[35] Morgan S. and Winship C., *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, Cambridge University Press, 2007.

[36] Narayanan L., Krishnan S., and Robinson H., "A Hybrid Deep Learning Based Assist System for Detection and Classification of Breast Cancer from Mammogram Images," *The International Arab Journal of Information Technology*, vol. 19, no. 6, pp. 965-974, 2022. https://doi.org/10.34028/iajit/19/6/15.

[37] Nawaz M., Nazir T., and Masood M., "Glaucoma Detection using Tetragonal Local Octa Patterns and SVM from Retinal Images," *The International Arab Journal of Information Technology*, vol. 18, no. 5, pp. 686-693, 2021.

[38] Paliwal M. and Kumar U., "Neural Networks And Statistical Techniques: a Review of Applications," *Expert Systems with Applications*, vol. 36, no. 1, pp. 2-17, 2009. https://doi.org/10.1016/j.eswa.2007.10.005.

[39] Pearl J., *Causality: Models, Reasoning, and Inference*, Cambridge University Press, 2009.

[40] Pokkuluri K. and Nedunuri U., "Crop Disease Prediction with Convolution Neural Network (CNN) Augmented With Cellular Automata," *The International Arab Journal of Information Technology*, vol. 19, no. 5, pp. 765-773, 2022. 0 https://doi.org/10.34028/iajit/19/5/8.

[41] Powers D. and Xie Y., *Statistical Methods for Categorical Data Analysis*, Emerald Publishing, 2008.

[42] Ramakrishnan D. and Radhakrishnan K., "Applying Deep Convolutional Neural Network (DCNN) Algorithm in the Cloud Autonomous Vehicles Traffic Model," *The International Arab Journal of Information Technology*, vol. 19, no. 2, pp. 186-194, 2022. https://doi.org/10.34028/iajit/19/2/5.

[43] Ratna M., Bhattacharya S., Abdulrahim B., and McLernon D., "A Systematic Review of the Quality of Clinical Prediction Models in Vitro Fertilisation," *Human Reproduction*, vol. 35, no. 1, pp. 100-116, 2020. https://doi.org/10.1093/humrep/dez258.

[44] Salaken S.M., Khosravi A., Nguyen T., and Nahavandi S., "Extreme Learning Machine Based Transfer Learning Algorithms: a Survey," *Neurocomputing*, vol. 267, pp. 516-524, 2017. https://doi.org/10.1016/j.neucom.2017.06.037.

[45] Siristatidis C., Vogiatzi P., Pouliakis A., Trivella M., Papantoniou N., and Bettocchi S., "Predicting Ivf Outcome: A Proposed Web-Based System

Using Artificial Intelligence," *in Vivo*, vol. 30, no. 4, pp. 507-512, 2016.

[46] Song Y. and Lu Y., "Decision Tree Methods: Applications for Classification and Prediction," *Shanghai Archives of Psychiatry*, vol. 27, no. 2, pp. 130-135, 2015. 10.11919/j.issn.1002-0829.215044.

[47] Specht D., "A General Regression Neural Network," *IEEE Transactions on Neural Networks*, vol. 2, no. 6, pp. 568-576, 1991.

[48] Tibshirani R., "Regression Shrinkage And Selection Via The Lasso: A Retrospective," *Journal of the Royal Statistical Society: Series B*, vol. 73, part 3, pp 273-282, 2011.

[49] Tiwari S., Jain A., Yadav K., and Ramadan R., "Machine Learning-Based Model for Prediction of Power Consumption in Smart Grid," *The International Arab Journal of Information Technology*, vol. 19, no. 3, pp. 323-329, 2022. https://doi.org/10.34028/iajit/19/3/5.

[50] Xie Y., "Values and Limitations of Statistical Models," *Research in Social Stratification and Mobility*, vol. 29, no. 3, pp. 343-349, 2011. https://doi.org/10.1016/j.rssm.2011.04.001.

[51] Zhang C. and Ma Y., *Ensemble Machine Learning: Methods and Applications*, Springer, 2012. https://doi.org/10.1007/978-1-4419-9326-7.

[52] Zhang L. and Suganthan P., "A Survey of Randomized Algorithms for Training Neural Networks," *Information Sciences*, vol. 364-365, pp. 146-155, 2016. https://doi.org/10.1016/j.ins.2016.01.039.

**Marko Grebovic**, he holds undergraduate degree in Computer Science (Faculty of Science and Mathematics, University of Montenegro) and postgraduate degree in IT and Management (Faculty of Management, Adriatic University). Currently, he is PhD candidate at the University of Donja Gorica, Faculty for Information Systems and Technologies, where he works as teaching assistant in AI and ML courses. Previously, he worked as teaching assistant at Adriatic University, Faculty of Business Economics and Law in applied mathematics and information systems courses. He participated in several IEEE conferences and training projects concerning AI/ML. His main areas of scientific interest are machine learning and data science.

**Luka Filipovic**, he holds PhD degree in Technical sciences (Faculty of Electrical Engineering, University of Montenegro). He works at the University of Donja Gorica as lecturer in computer science and information systems courses and research scientist at EuroCC project (NCC Montenegro). He has experience in various national and European research projects based on information systems, distributed computing and IT development. His main areas of scientific interest are distributed and parallel computing, data science and machine learning.

**Ivana Katnic**, she finished undergraduate and postgraduate studies "Entrepreneurial economy" at Faculty of economics, University of Montenegro and defended her PhD at Faculty for International Economics, Finance and Bussines, University of Donja Gorica, where she works as Associate Professor, teaching courses: Microeconomics, Statistics, Econometrics and Academic writing. She is working at the Institute for Strategic Studies and Prognoses as researcher and analyst. She was member of the working group for the accession negotiations EU with Montenegro for the Chapter 18 - Statistics and also National Contact Point for the Spreading excellence and Widening participation measures for Montenegro within Horizon 2020. She was member of Higher Education Reform Experts (HEREs) for Montenegro, appointed by Education, Culture and Audio visual Executive Agency (EACEA). She is member of Association of Montenegrin economists, Montenegrin society of statisticians and demographers, MENSA Montenegro, and MENSA international. She participated in many workshops and conferences in country and abroad and published number of papers.

**Milica Vukotic**, she graduated from the Department of Economics at the University of Belgrade, received MA in International Economics and Finance from Brandeis University (Boston) as fellow of US Department of State, and received her PhD in Economics at the University of Montenegro. At the moment, she is holding positions of dean and professor in areas of statistics and microeconomics at the University of Donja Gorica. Before that, Milica was Assistant Professor in the Department of Economics at the University of Montenegro. She is/was leading or coordinating around forty projects, organizing both domestic and international conferences. Taught at the course for the preparation of the CFA exam, organized by USAID. Completed Training of Trainers Program at the US Department of Labour, Washington DC. She is also a senior analyst at the Institute for Strategic Studies and Projections and a member of Montenegrin Society of Economists and a member of the prestigious Mont Pelerin Society, founded by F.E. Hayek.

**Tomo Popović**, he holds a PhD degree in Technical sciences (Faculty of Electrical Engineering, University of Montenegro), Master of Engineering (Electrical and Computer Engineering, Texas A&M University, USA), and B.Sc in Electrical Engineering (Faculty of Technical Sciences, University of Novi Sad). He works at the University of Donja Gorica, where he teaches computer science and information systems courses. He has 25 years of international experience in both industrial and academic environment (Serbia, USA, Montenegro). His newer projects include the development of of smart solutions based on Internet of Things and blockchain for agri-food sector, as well as to apply AI/ML in agriculture, medicine, and energy sectors. Tomo Popović has over 100 papers published in scientific journals or conferences. He is a Senior member of the IEEE and member of ACM and CIGRE.