Further Evolution of Mortality Prediction with Ensemble-based Models on Hungarian Myocardial Infarction Registry

Péter Piros[1], Rita Fleiner[1], András Jánosi[2], Levente Kovács[1]

[1] Óbuda University, Bécsi út 96/b, 1034 Budapest, Hungary,
piros.peter@uni-obuda.hu, fleiner.rita@nik.uni-obuda.hu,
kovacs.levente@nik.uni-obuda.hu

[2] Gottsegen National Cardiovascular Center, Haller u 29, 1096 Budapest, Hungary,
andras.janosi@gokvi.hu

Abstract: In the current study, we present a new approach to predict 30-day and 1-year mortality of patients hospitalized with acute myocardial infarction. The dataset of this research is originated from Hungarian Myocardial Infarction Registry, a full, real-world, unfiltered database of myocardial infarctions from year 2014 to 2016 (n = 47,391). The new approach is based on ensembling and uses the prediction capability of different (already ensembled, in some cases) models like Random Forest, General Boosting Machine, Neural Network and Generalized Linear Model. We previously presented more different modelling techniques with the same target on the same dataset, and this new ensemble-based way of prediction proved to be the best among all the others. By numbers, this means 0.856 ROC AUC (area under the receiver operating characteristic curve) for the 30-day, and 0.839 ROC AUC for the 1-year mortality, both measured on validation datasets. We came to the conclusion that the combination of machine learning algorithms and regression models results the best performance in mortality prediction on the dataset of HUMIR.

Keywords: Myocardial infarction; Mortality prediction; Machine learning; Ensemble classifier

1 Introduction

In *Heart Disease and Stroke Statistics*, American Heart Association annually reports, that approximately every 40 seconds, an American will have an myocardial infarction (MI) - they did the same in the recent statistics titled *2022 Update* [1]. The estimated annual incidence of MI is 605,000 new attacks and 200,000 recurrent attacks, they reported. The overall prevalence for MI is 3.1% in US adults (>= 20

years of age). Males have a higher prevalence of MI than females for all age groups except 20 to 39 years of age. MI prevalence is 4.3% for males and 2.1% for females.

In addition, *Heart disease* (which can lead to myocardial infarction) is still at the first position of the ten leading cause of death, followed by cancer, unintentional injuries, chronic lower respiratory diseases, stroke and Alzheimer disease, respectively.

In the area with numbers like these, mortality prediction can and should play a very important role in the hand of physicians: with validated models, it becomes possible to select patients with high-risk of death and use this information in the process of treatment. Using new, real-life datasets to extract hidden information can lead to more effective treatment and prevention.

Reliable, high-quality datasets are mandatory to build and train any type of predictive model. Hungarian Myocardial Infarction Register (HUMIR) project was introduced in 2010, initially collected AMI information from five districts of Budapest and the county of Szabolcs-Szatmár-Bereg. In 2014, the Hungarian government selected it as the official myocardial database and obligated all hospitals in the territory of Hungary to report all MI-cases to HUMIR. In the recent years, around 15,000 new patients got registered per year and until December 2022, the 94 participating hospitals reported 157,724 cases in 142,439 patients.

The dataset of this research is originated from HUMIR, it is a full, real-world, unfiltered database of myocardial infarctions from year 2014 to 2016 (n = 47,391). The features of the dataset consists of three attribute groups: General information about the patient (Group 1), Previously reported diseases (Group 2) and Information about the pre- and in-hospital treatment (Group 3). All relevant features of each group will be discussed later in *Section 3*.

In previous researches we developed several machine-learning models based on Decision Tree, Neural Networks (NN), Logistic Regression, Random Forest (RF), and Generalized Boosted Models algorithms to predict 30-day and 1-year mortality on the same dataset. The results achieved with these methods were published in several conferences and papers.

The aim of the current paper is to provide an ensemble-based modelling technique in the field of mortality prediction which is trained on the introduced HUMIR-dataset and combines the predictive power of the constituent learning algorithms. The idea behind the approach that we are working on the same dataset both here and in related researches is the following: we are trying to establish an order in the list of different modelling techniques by keeping the dataset fixed and trying to maximize the prediction capability of each of our models. Here, we present the details and results of the ensemble-based technique, and, as a result, we can compare the prediction power of the different learning algorithms.

The rest of the paper is organized as follows. In Section 2, we summarize our previous efforts with different modelling algorithms, and quote some related research results from other authors. Section 3 discusses the tools and methods involved in the current study, especially the background of ensemble modelling. Then, in Section

4, we present the new results and compare them with all our formers'. Finally, we conclude with a summary.

2 Related Work

2.1 Previous Results

In one of the previous works, we overviewed [2] the health care registries in Europe and found that although several databases store information about patients and diseases, only a few exist that focus directly on myocardial events and treatments. Three European projects were investigated: Myocardial Ischaemia National Audit Project (MINAP) in England, Swedish Web-system for Enhancement and Development of Evidence-based care in Heart disease Evaluated According to Recommended Therapies (SWEDEHEART) in Sweden and National Registry of Acute Myocardial Infarction in Switzerland (AMIS Plus) in Switzerland. Then, we stated that from 2014 HUMIR operated as the official hungarian myocardial register, and we examined the changes of completeness and validity of data stored in HUMIR. Completeness was calculated with the official numbers of National Health Insurance Fund of Hungary (the central official organ of health insurance, supervised by the Government of Hungary; Hungarian acronym: OEP). As a result, by 2016, the completeness reached around 84%. As the ensemble-based research is based on the same dataset, this value of completeness applies to the current research as well.

After receiving the dataset from HUMIR, we compared the relative performance of our three initial models, which were based on decision tree, neural network, and logistic regression techniques [3]. Area under the receiver operating characteristic curve (ROC AUC) was used in all cases for evaluating performance. For 30-day mortality, we achieved an average of 0.788 for decision tree models, 0.837 for neural net models and 0.836 for regression models on training set (on validation sets: 0.774, 0.835 and 0.834, respectively). In the case of 1-year mortality, these averages were 0.754 for decision tree models, 0.8194 for neural net models and 0.8191 for regression models (on validation sets: 0.743, 0.8179 and 0.8176, respectively). So, differences were non-significant between our neural network and regression, but both signicantly outperformed our decision trees.

In the next study [4], we investigated if an order could be declared between different tuning approaches on decision tree models. 1-year mortality was selected as target variable and K-fold cross validation, repeated cross validation and bootstrap were used to find the optimal parameters for each model on the dataset of HUMIR. The differences were measured in 10 different cases with increasing, randomly selected number (starting from n = 300, until n = 18,000) of records. It was found that a relatively small difference exists, but K-fold cross validation proved to be the best before repeated cross validation and bootstrap.

Then, we involved a new technique to the research: the predictive power of our newly developed Random Forest models were compared with the result of our de-

cision tree [5]. ROC AUC values of Random Forest models for predicting 30-day mortality were 0.843 and 0.847 (training and validation set), while for the 1-year models these were 0.835 and 0.836, respectively. These meant to be a significant difference: Random Forest models were at least 5% better than the decision tree models, but in some cases the improvement is above 9%. This amount of difference between these two tree-based solution could lead to serious relevance in clinical environment.

Generalized Boosted Model (GBM) was the next learning technique which we used to train models [6]. The ROC AUC values of the new GBM-models for 30-day mortality were 0.847 and 0.839 (training and validation set), while for the 1-year models these were 0.828 and 0.821, respectively. The numbers represented a strong and stable learner: the standard deviation of ROC AUC values between models of different imputations were 0.0035 for the 30-day models, and 0.0038 for the 1-year models, both calculated on the validation datasets.

The difference of predictive performance (measured in ROC AUC values) between our RF and GBM models were between 0.5% and 0.9%, except in the case of 1-year model on the validation dataset: it was 1.7% compared to the RF results. Our conclusion said that GBM almost reached the performance of the RF models.

2.2 Ensembled Modelling

Ensembled modelling is one of the most promising area of machine learning-based predicting. In different domains researchers try to combine the advantages of individual classifiers to produce a strong learner. In the current subsection we summarize the results of some of the most-related articles.

Latha et al. [14] used ensembled modelling on the Cleveland Heart Disease Database to improve the accuracy of prediction of heart disease risk. They used weak classifiers like decision tree (C4.5), Bayesian network, Naïve Bayes, Random forest and neural networks to combine them in different ensembled-based modelling techniques like Boosting, Bagging, Stacking and Majority vote. This comparative analytical approach was done to determine how the ensemble technique can be applied for improving prediction accuracy in heart disease. As a result, a comparison of the various ensembling strategies revealed that the accuracy of the weak classifiers could be increased by a maximum of 7.26%.

Austin el al. found [15] that improvements in the misclassification rate using boosted classification trees were at best minor compared to when conventional classification trees were used. They analysed short-term (30-day) mortality in two cohorts of patients hospitalized with either acute myocardial infarction (N = 16,230) or congestive heart failure (N = 15,848). They observed minor to modest improvements to sensitivity, with only a negligible reduction in specificity.

In another study [16] on the same datasets, Austin el al. evaluated the improvement that is achieved by using ensemble-based methods, including bootstrap aggregation (bagging) of regression trees, random forests, and boosted regression trees. They

found that ensemble methods offered substantial improvement in predicting cardiovascular mortality compared to conventional regression trees; but conventional logistic regression models that incorporated restricted cubic smoothing splines had even better performance. An example of ROC AUC values from their study: on the "EFFECT Follow-up" database, their models achieved the following results by ROC AUC: regression tree: 0.767, bagged trees: 0.820, random forest: 0.843, Boosted trees (depth four): 0.852, Logistic regression: 0.852, Logistic regression—Splines: 0.858, Logistic regression—GRACE score: 0.826.

A neural network ensemble method was proposed [17] by Das et al.. Three independent neural networks models were used (Levenberg–Marquardt, scaled conjugate gradient and Pola–Ribiere conjugate gradient algorithms) as primary learners, and the final, ensembling layer combined their results with averaging. The investigated database contained 303 complete samples. Although they didn't published the predictive performance of the individual models, the final model gained 89.01% classification accuracy, 80.95% sensitivity and 95.91% specificity values on the validation dataset.

Subramanian et al. were also focused on heart failure mortality and used partial patient data from the dataset of Vesnarinone Evaluation of Survival Trial [18]. On the data of 963 patients, they established three logistic regression models to predict survival and an ensemble model learned by boosting. On of the major finding of the study is that their ensemble model performed significantly better than the standard approach of logistic regression. As authors discuss, the reason for this significant increase in predictive accuracy is that "an ensemble of models adjusts better for the biological variability inherent in clinical studies that are derived from patient data."

Although the previous examples were focusing on heart failure and mortality prediction, researchers gain advantages of ensemble modelling in various fields: Bagging, Random Forests and Extra Trees were used by [19] Petkovic et al. when they addressed the task of feature ranking for hierarchical multi-label classification. Extra Tree is similar to RF, with two main differences: instead of using bootstrap replicas, Extra Trees use the whole original sample; and the selection of cut points is random and not an optimum split, like in RF [20]. Three feature ranking scores like Symbolic, Genie3 and the Random Forest Score were investigated and authors found the first two scores yield relevant feature ranking. In the domain of medical image processing, Tóth et al. [21] described an efficient 3D visualization framework in connection with an ensemble-based decision support system.

3 Materials and Methods

3.1 Fundamentals of Ensembled Modelling

Ensembled modelling as a strategy based on the idea that if we combine the predictive performance of different classifiers, it can produce a stronger learner. *Bagging* also known as *Bootstrap aggregation*, *Boosting* and *Stacking* are the main classes of

ensemble learning methods.

1. In *Bagging*, from the original dataset new datasets (called *bootstrap samples* or *bootstrap replicates*) are selected with replacement; we train the models on each of them; and finally the outcome is calculated with averaging (in case of regression) or majority vote (in case of classification).
2. In *Boosting*, simple, 2-3 level depth trees are used and we build the models trying to predict based on the prediction error of the previous tree. The two types are: ADA Boosting and Gradient Boosting.
3. In *Stacking* different types of individual machine learning models are applied (*1-st level learner*) and trained on the same, original dataset, then we combine the prediction results of them in an upper level (*meta-learner* or *second-level learner*).

In the list of previously published models, our RF models is similar to *Bagging* category (although there are some differences between RF and bagged models); and our GBM models belong to *Boosting* category: we were using a given number of decision tree to construct a final, better learner.

In this paper, we are focusing on *Stacking*, as we are using different types of first-levels learners then we try to exploit the common predictive power of them in an upper level.

The schematic overview of Stacking is depicted on Figure 1. As it shows, there can be any number of 1st-level learners, they are trained on the full, original dataset and produce their "local predictions". These different predictions serve as inputs for the meta-learner who attempts to combine these predictions to have the best possible final outcome. As can be seen, the 1-st level learners have to be fully trained and the local predictions have to be made *before* the Meta-learner starts to operate.

In the current study, our 1-st level learners are RF, GBM and NN, while the Meta-learner is Generalized Linear Model, so the ensembled model is a combination of machine learning algorithms and regression models. The modelling structure of the current research is explained and visualized in details in *Chapter 3.4*.

3.2 Patient Record

As we are working on the same dataset to have comparable results, the structure of a patient record is the same as it was in our previous researches and consists of the following groups and fields (the following categorisation is made by the authors of this research to make referencing easier):

1. Group 1: General information about the patient (*Event ID*, *Patient ID*, *If the patient alives*, *Date of death*, *Gender*, *Date of birth*, *ZIP code*)
2. Group 2: Previously reported diseases (*Myocardial infarction*, *Heart failure*, *Hypertension*, *Stroke*, *Diabetes*, *Peripheral vascular disease*, *Hyperlipidaemia*, *Smoking*)

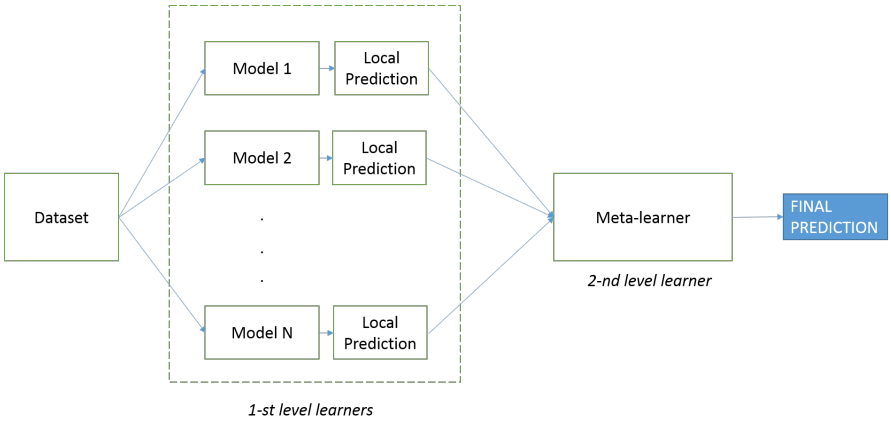

Figure 1
Schematic overview of Stacking

3. Group 3: Information about the pre- and in-hospital treatment (*Prehospital reanimation*, *Cardiogenic shock*, *Percutaneous Coronary Intervention*, *Level of creatinine*, *Diagnosis*, *Treatment ID*, *Date of admission*, *Creatinine*)

We had two fields without holding any relevant information in the given context: *Event ID*, *Patient ID*, they were both eliminated. *Level of creatinine* was almost an empty field (was filled only in 2.8% of the rows), so it was eliminated as well. Finally, a total of 21 fields were involved as the features of our dataset.

3.2.1 Target Variables

As mortality usually examined in short and long run by physicians, our dependent variables (target variables) were the 30-day and the 1-year mortality. Since these values were not initially present in the dataset, the following simple technique was used both here and in our previous works: they were calculated as the date range between two of our fields, the date of hospital admission and the date of the death.

3.3 Missing Values and Imputation

The presence of missing values proved to be essential in our researches. These values in percentage for each field are shown in Table 1 (table contains only the attributes where at least one missing value is present):

To handle the issue with missing data, multiple imputation using Fully Conditional Specification and Bayesian linear regression was applied with 5 imputations and 5 iterations, leaving the final, prepared dataset size at $n = 47{,}391$.

As a result, 5 different sub-datasets were created, and on each we performed the full process of modelling for both the 30-day and 1-year mortality, as it can be seen in the *Modelling structure* subsection.

Table 1
Rates of missing values of attributes

Attribute name	Missing value rate (%)
Hypertension	2.5
Myocardial infarction	4.3
Diabetes mellitus	4.4
Stroke	5.3
Prehospital reanimation	5.8
Creatinine	6.0
Heart failure	7.4
Cardiogenic shock	8.3
Peripheral vascular disease	9.9
Hyperlipidaemia	18.4
Smoking	39.4

3.4 Modelling Structure

After generating the imputations, training and validation sets were created on each imputations with maintaining the original distribution of the target variables. The trainings were used as the input data of the models (on these, the algorithm performed boosting to find the optimal hyperparameters for the given model); while the validation datasets were used to manually measure the prediction performance in ROC AUC.

ROC AUC was applied to select the optimal parameters using the largest value. For each parameter-combination, a bootstrap based validation with 10 resampling iterations were used on the training set to obtain a reliable estimate of model performance.

The Modelling structure can be visualized in three figures: on Figure 2, the full modelling structure is visualized, while the next two figures focus on the separate sections in a more detailed way.

Figure 3 depicts the first step: the connection between the original dataset, the imputations, the target variables and the models as inputs of the ensembled models. It contains only one case (RF model) out of the three, but the same processes were performed for GBM and NN as well.

After we finally had all the 5 (number of imputations) * 2 (number of target variables) * 3 (number of model types) = 30 models, we could go on with the ensembling phase. Figure 4 depicts the connection between the initial models and the ensembed ones.

3.5 Software Environment

As in the connected previous researches, we used R as an open-source software environment for statistical computing and graphics [7].

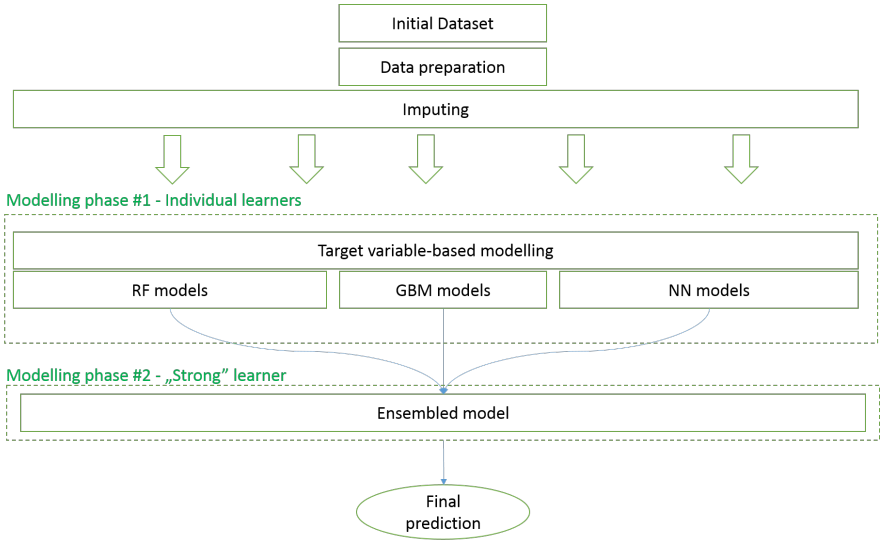

Figure 2
Modelling structure - Overview

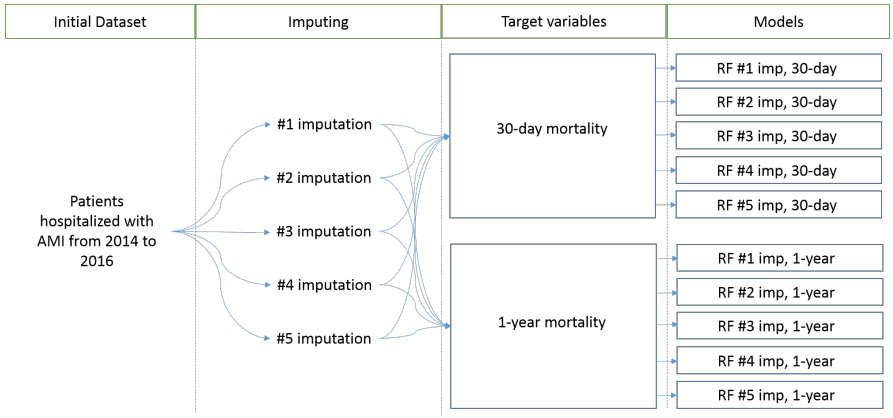

Figure 3
Modelling structure - Step 1

For the three initial models the following packages were applied: randomForest [8], GBM [9] and rms [10] (RF, GBM, LR, respectively), while the ensembling methods were handled by caretEnsemble [11] package. For resampling and training the models, we used Caret [12] package. To deal with missing data and imputation, mice [13] package was used.

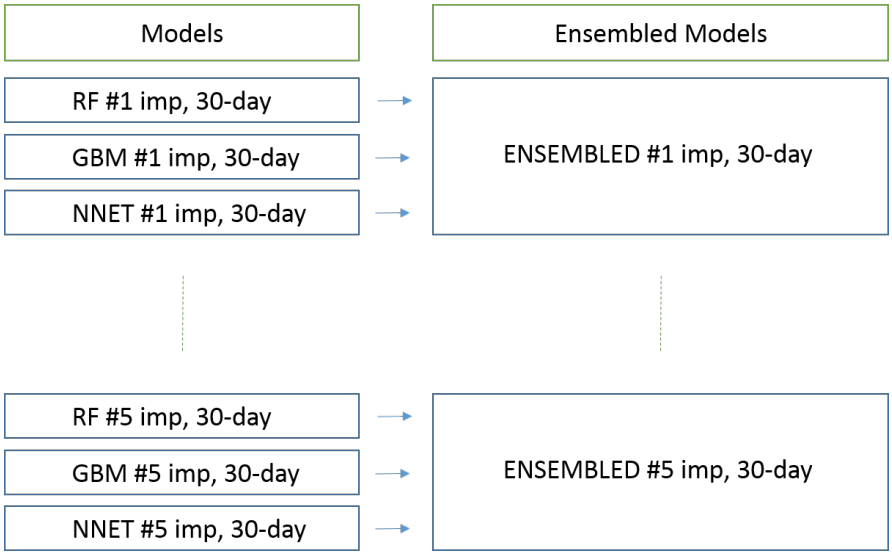

Figure 4
Modelling structure - Step 2

3.6 Hardware Environment

During the investigation, we experienced that the *usual* hardware environment was not suitable for Ensembled modelling on this size of dataset. Since with an *average* configuration *(Intel Core i3 processor, 12 GB memory)*, the training times were around 5 hours with neural networks, we chose *Amazon Web Services* with *EC2* instances with the following parameters: 16 vCPU, 70 ECU, 64 GB memory (*m5.4xlarge* configuration). On this configuration, the average training time for a given model was below 25 minutes.

This hardware configuration differs from the previously published RF and GBM models', but this fact doesn't have any affect on the prediction power.

4 Results and Discussion

4.1 Prediction Capability

In Table 2 and Table 3 we summarized the ROC AUC values of the individual and ensembled models for 30-day and 1-year mortality. All values were calculated on the corresponding validation datasets.

Figure 5 depicts the performance of all the four models in a ROC curve while numerical differences between the methods with 99.2% confidence intervals are shown on Figure 6, both for a randomly selected case (30-day mortality as target variable and

Table 2
ROC AUC values of the 30-days models, validation set.

	Imp. #1	Imp. #2	Imp. #3	Imp. #4	Imp. #5	Avg
GBM	0.8411	0.8381	0.8375	0.8443	0.8346	0.8391
RF	0.8499	0.8472	0.8416	0.8528	0.8436	0.8470
NN	0.8358	0.8334	0.8353	0.8398	0.8326	0.8354
Ensembled	0.8592	0.8542	0.8517	0.8602	0.8522	0.8555

Table 3
ROC AUC values of the 1-year models, validation set.

	Imp. #1	Imp. #2	Imp. #3	Imp. #4	Imp. #5	Avg
GBM	0.8169	0.8202	0.8251	0.8178	0.8246	0.8209
RF	0.8323	0.8332	0.8392	0.8312	0.8384	0.8349
NN	0.8134	0.8166	0.8234	0.8140	0.8224	0.8180
Ensembled	0.8358	0.8371	0.8439	0.8349	0.8432	0.8390

the first imputation was selected). Table 4 reports the standard deviation between the ROC AUC values of the separate models trained on the different imputations.

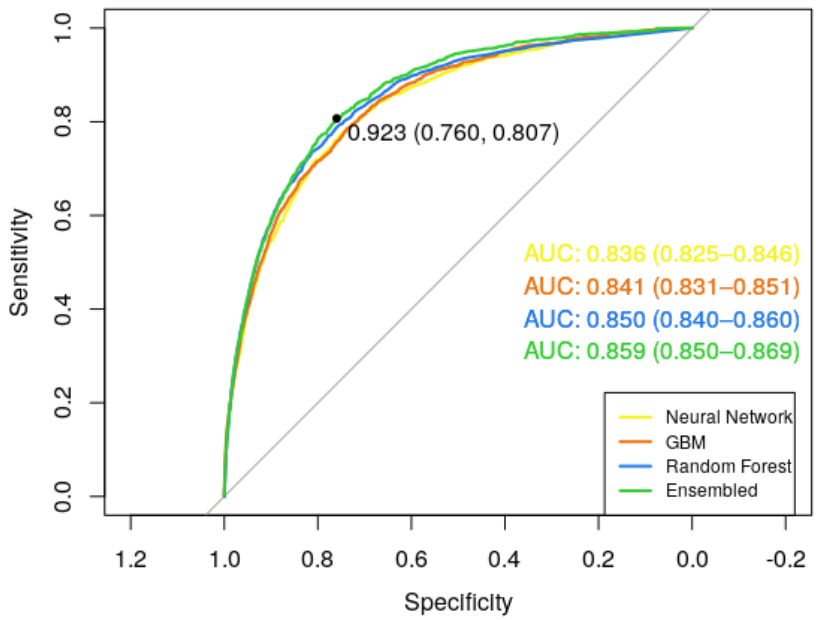

Figure 5
Performance of our Neural Network, Random Forest, Generalized Boosted and Ensembled models.
Target: 30-day mortality, dataset: Imputation #1, validation set.

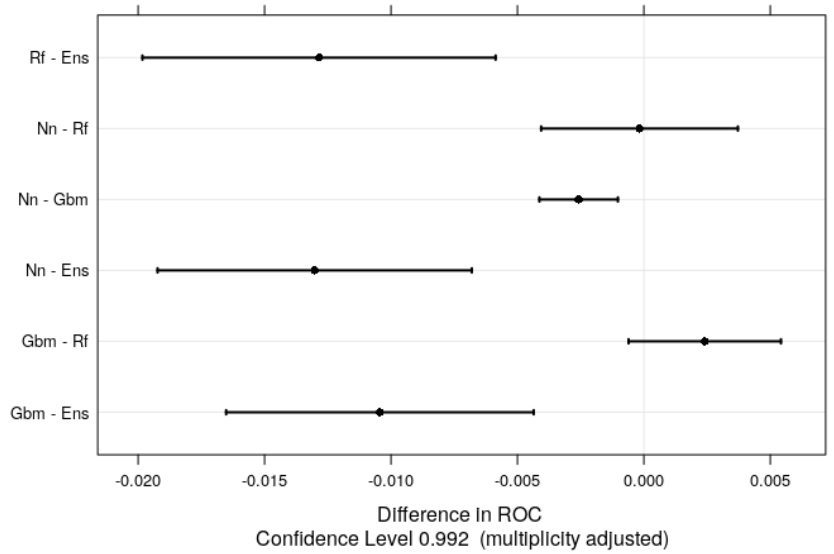

Figure 6
Numerical differences between our Neural Network, Random Forest, Generalized Boosted and Ensembled models. Target: 30-day mortality, dataset: Imputation #1, validation set.

Table 4
Standard deviation of the ROC AUC values of imputations per model type and target variable.

	GBM	RF	NN	Ensembled
30-day models	0.0037	0.0045	0.0028	0.0040
1-year models	0.0038	0.0037	0.0047	0.0043

4.2 Variable Importance

Variable importance, in general, refers to a measure of how much a model uses a given variable to make accurate predictions. In this subsection, we are dealing with the variable importance values for each individual and the ensembled model.

Since the definitions and the methods of calculating the variable importance in separate model types differ, instead of listing the exact feature importance values for each model type, we deal with relative importance: the position of the given feature on the list of the most important fields. With using relative importance it becomes possible to compare the most important features of different models types, i.e. we can make a global order between the variables over the different models.

As we used multiple imputations, we aggregated variable importance values in the imputations: summed up all the relative importance values for each field for a given target variable and for a given model type, then divide this value by the number of imputations. The resulted value represents the relative importance of the given

feature, and in this number, all the imputations added their effects.

The aggregated and relative values of feature importance in descending order for the 30-day models are the following:

1. GBM: Cardiogenic shock (36.3), Age (21.1), Abnormal level of creatinine (10.4), Percutaneous Coronary Intervention (6.7), Prehospital reanimation (6.6)
2. Random Forest: Age (31.1), Cardiogenic shock (14.2), Smoking = never (13.5), Smoking = quit (13.3), Hyperlipidaemia (6.6)
3. Neural net: Age (19.8), Cardiogenic shock (15.2), Percutaneous Coronary Intervention (9.6), Abnormal level of creatinine (9.1), Prehospital reanimation (7.4)
4. Ensembled: Age (26.1), Cardiogenic shock (15.7), Smoking = never (8.6), Smoking = quit (8.4), Abnormal level of creatinine (7.4)

The aggregated and relative values of feature importance in descending order for the 1-year models are the following:

1. GBM: Age (34.1), Cardiogenic shock (16.9), Abnormal level of creatinine (11.9), Percutaneous Coronary Intervention (10), Heart failure (7.8)
2. Random Forest: Age (36.6), Smoking = never (12.4), Smoking = quit (12), Cardiogenic shock (8.2), Abnormal level of creatinine (6.5),
3. Neural net: Age (23.6), Cardiogenic shock (10.9), Percutaneous Coronary Intervention (10.7), Abnormal level of creatinine (9.1), Prehospital reanimation (7)
4. Ensembled: Age (30.8), Cardiogenic shock (10.6), Abnormal level of creatinine (8.3), Percutaneous Coronary Intervention (7.3), Smoking = never (7.3), Smoking = quit (8.4), Abnormal level of creatinine (7.4)

Conclusions

As can be seen in *Section 4.1*, for both target variables, the ensembling technique proved to be the best among GBM, RF and NN models. In addition, it outperforms not only the Decision Tree but the regression models published by the current authors in previous papers and conferences.

In case of the 30-day models, the improvements were 1.64% 0.85% and 2.01% (compared to GBM, RF and NN, respectively), while in case of the 1-year models, these were 1.81%, 0.41% and 2.10% (compared to GBM, RF and NN, respectively), all calculated and compared on the validation datasets.

Although the improvement is typically 1-2%, in the area of cardiovascular diseases, this difference can play a significant role in the hand of physicians when they try to select patients with high-risk of death.

As mentioned in *Section 3*, in the data preparation process we applied multiple imputation using Fully Conditional Specification and Bayesian linear regression. We can state, that the reported standard deviation between the imputations verifies the operation of the imputations. At 5% significance level, the differences between the results of the imputations were non-significant. To assess the differences, pairwise differences of performance measures were calculated and checked if they're equal with zero in expected value using a Welch corrected t-test with Bonferroni adjustment for multiplicity. This means, that next to our RF and GBM models, the Ensembled models also represent a stable learner.

Comparing the most important features from the view of Ensembled model, there were two common items in the list of top 5 features between the ensembled and the constituent models, namely *Age* and *Cardiogenic shock*; *Abnormal level of creatinin* is present in GBM's, NN's and Ensembled's list, while derived fields of *Smoking* appear also in RF and Ensembled.

According to the general results of several papers, that an ensembled model can boost up the predictive performance of the individual constituent models, we can confirm this finding in the case of our models on the dataset of HUMIR. This result is in line with referenced researches in *Section 2*. As a final conclusion, we have found that the combination of machine learning algorithms and regression models results the best performance in mortality prediction on the dataset of HUMIR.

References

[1] Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Alonso, A., Beaton, A. Z., Bittencourt, M. S., ... & American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. (2022). Heart disease and stroke statistics—2022 update: a report from the American Heart Association. Circulation, 145(8), e153-e639.

[2] Fujita, H. (2017, September). An overview of myocardial infarction registries and results from the hungarian myocardial infarction registry. In IOS Press (Vol. 297, p. 312).

[3] Piros, P., Ferenci, T., Fleiner, R., Andréka, P., Fujita, H., Főző, L., ... & Jánosi, A. (2019). Comparing machine learning and regression models for mortality prediction based on the Hungarian Myocardial Infarction Registry. Knowledge-Based Systems, 179, 1-7.

[4] Piros, P., Fleiner, R., Ferenci, T., Kovács, L., & Jánosi, A. (2019, May). Comparing the predictive power of decision tree models with different tuning approaches on Hungarian Myocardial Infarction Registry. In 2019 IEEE 13th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 326-331). IEEE.

[5] Piros, P., Fleiner, R., & Kovács, L. (2020, June). Random Forest-based predictive modelling on Hungarian Myocardial Infarction Registry. In 2020 IEEE 15th International Conference of System of Systems Engineering (SoSE) (pp. 525-530). IEEE.

[6] Piros, P., Fleiner, R., & Kovács, L. (2020, November). Finding improved predictive models with Generalized Boosted Models on Hungarian Myocardial Infarction Registry. In 2020 IEEE 20th International Symposium on Computational Intelligence and Informatics (CINTI) (pp. 179-184). IEEE.

[7] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[8] A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18–22.

[9] Brandon Greenwell, Bradley Boehmke, Jay Cunningham and GBM Developers (2019). gbm: Generalized Boosted Regression Models. R package version 2.1.5. https://CRAN.R-project.org/package=gbm

[10] Frank E Harrell Jr (2018). rms: Regression Modeling Strategies. R package version 5.1-2. https://CRAN.R-project.org/package=rms

[11] Zachary A. Deane-Mayer and Jared E. Knowles (2019). caretEnsemble: Ensembles of Caret Models. R package version 2.0.1. https://CRAN.R-project.org/package=caretEnsemble

[12] Max Kuhn. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan and Tyler Hunt. (2017). caret: Classification and Regression Training. R package version 6.0-77. https://CRAN.R-project.org/package=caret

[13] Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. URL http://www.jstatsoft.org/v45/i03/.

[14] Latha, C. B. C., & Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked, 16, 100203.

[15] Austin, P. C., & Lee, D. S. (2011). Boosted classification trees result in minor to modest improvement in the accuracy in classifying cardiovascular outcomes compared to conventional classification trees. American journal of cardiovascular disease, 1(1), 1.

[16] Austin, P. C., Lee, D. S., Steyerberg, E. W., & Tu, J. V. (2012). Regression trees for predicting mortality in patients with cardiovascular disease: What improvement is achieved by using ensemble-based methods?. Biometrical journal, 54(5), 657-673.

[17] Das, R., Turkoglu, I., & Sengur, A. (2009). Effective diagnosis of heart disease through neural networks ensembles. Expert systems with applications, 36(4), 7675-7680.

[18] Subramanian, D., Subramanian, V., Deswal, A., & Mann, D. L. (2011). New predictive models of heart failure mortality using time-series measurements and ensemble models. Circulation: Heart Failure, 4(4), 456-462.

[19] Petkovic, M., Dzeroski, S., & Kocev, D. (2020). Feature ranking for hierarchical multi-label classification with tree ensemble methods. Acta Polytechnica Hungarica, 17(10), 129-148.

[20] Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine learning, 63(1), 3-42.

[21] Tóth, J., Tornai, R., Labancz, I., & Hajdu, A. (2019). Efficient visualization for an ensemble-based system. Acta Polytechnica Hungarica, 16(2), 59-75.