Lecture Notes in Networks and Systems 872

Miroslav Trajanovic
Nenad Filipovic
Milan Zdravkovic *Editors*

Disruptive Information Technologies for a Smart Society

Proceedings of the 13th International Conference on Information Society and Technology (ICIST)

Springer

Lecture Notes in Networks and Systems 872

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Miroslav Trajanovic · Nenad Filipovic ·
Milan Zdravkovic
Editors

Disruptive Information Technologies for a Smart Society

Proceedings of the 13th International Conference on Information Society and Technology (ICIST)

Springer

Editors
Miroslav Trajanovic
Faculty of Mechanical Engineering
University of Nis
Nis, Serbia

Nenad Filipovic
Faculty of Engineering
University of Kragujevac
Kragujevac, Serbia

Milan Zdravkovic
Faculty of Mechanical Engineering
University of Nis
Nis, Serbia

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-031-50754-0 ISBN 978-3-031-50755-7 (eBook)
https://doi.org/10.1007/978-3-031-50755-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

This book includes selected high-quality peer-reviewed research papers presented at 13th International Conference on Information Society and Technology held on Kopaonik Mountain, Serbia, on Mar 12–15, 2023. In an era where technology disrupts many facets of our lives, the papers included in this issue exemplify the remarkable ways in which information technologies are reshaping our world, driving innovation, and paving the path toward a smarter, more efficient society.

The selected papers represent a diverse range of topics, all connected by the common commitment to explore information technologies as tools for positive transformation. From e-government requirements specification to advanced machine learning algorithms for river water quality management and from the application of artificial intelligence in healthcare to the analysis of financial markets using social media data, these contributions collectively illuminate the profound impact of information technologies on various domains.

One of the several focal points of this special issue is the intersection of artificial intelligence and public administration. Some papers delve into this area, addressing topics such as the role of AI in public administration and business sectors and the adoption of e-contracting and smart contracts for legally enforceable conformance checking in collaborative production. These papers underscore the potential of information technologies to enhance governance, streamline processes, and promote transparency in the public sector.

Another significant theme explored in this issue is the application of disruptive technologies in healthcare. Whether it's the prediction of coronary plaque progression using data mining and artificial neural networks or the risk stratification of patients with hypertrophic cardiomyopathy through genetic and clinical data features, these studies demonstrate how advanced information technologies are contributing to the diagnosis, treatment, and overall well-being of individuals. The use of technology in healthcare is becoming increasingly indispensable, and the papers presented here showcase the latest trends in this field.

Furthermore, this special issue highlights the importance of sustainability and environmental consciousness in our technologically driven society. From estimating solar power potential for rooftops to optimizing wind production forecasting and analyzing hydropower system resilience, these papers underscore the critical role of information technologies in promoting eco-friendly practices and renewable energy solutions.

The paper review process, organized as a single blind, had two stages. In the first stage, the papers were reviewed to be accepted for presentation at the 13th International Conference on Information Society and Technology. A total of 80 papers were accepted for presentation at the conference. The authors had the opportunity to prepare an improved version of the manuscript for these proceedings. After the second stage of review, 48 papers were accepted for publication in the proceedings.

In conclusion, "Disruptive Information Technologies for a Smart Society" represents a collective effort to explore the transformative power of information technologies in diverse domains. We extend our heartfelt gratitude to all the authors who have contributed their valuable research. Also, we would like to thank the reviewers who, with their expertise and comments, contributed to significantly improving the quality of the selected papers. We believe that the insights and findings presented in these papers will not only advance our understanding of the potential impact digitalization may have but also inspire further research and innovation in the pursuit of a smarter, more connected, and sustainable society.

November 2023

Miroslav Trajanovic
Nenad Filipovic
Milan Zdravkovic

Organization

General Chair

Miodrag Ivkovic	Information Society of Serbia, Serbia

Program Committee Chairs

Miroslav Trajanovic	Ministry of Science, TD and Innovation, Serbia
Milan Zdravkovic	University of Nis, Serbia
Nenad Filipovic	University of Kragujevac, Serbia
Zora Konjovic	Singidunum University, Serbia

Steering Committee

Hervé Panetto	Université de Lorraine, Nancy, France
Georg Weichhart	Primetals Technologies, Austria
Anton Kos	University of Ljubljana, Slovenia
Osiris Canciglieri Jr.	Pontifical Catholic University of Parana, Brazil
Nikola Milivojevic	Jaroslav Cerni Institute, Serbia
Boban Stojanovic	University of Kragujevac, Serbia
Milan Stojkovic	The Institute for Artificial Intelligence R&D of Serbia

Program Committee

Thomas Ahmad	University of Minho, Portugal
Alexis Aubry	University of Lorraine, France
Marija Boban	University of Split, Croatia
Zorica Bogdanovic	University of Belgrade, Serbia
Miloš Bogdanović	University of Nis, Serbia
Ivan Ćirić	University of Nis, Serbia
Zarko Cojbasic	University of Nis, Serbia
Dubravko Culibrk	Institute for Artificial Intelligence R&D of Serbia
Michele Dassisti	Politecnico di Bari, Italy
Igor Dejanovic	University of Novi Sad, Serbia

Additional Reviewers

Ivan Ilic
Milos Ivanovic
Lidija Korunovic
Nikola Korunovic
Srdjan Kostic
Brankica Majkic-Dursun
Jelena Mitic
Vladimir Mitrovic
Jelena Rakocevic
Mirko Stojiljkovic
Dejan Tanikic
Vesna Tripkovic
Nikola Vitkovic

Contents

Advanced Information Systems

ICT for Health, Well-Being and Sports

Prediction of Coronary Plaque Progression Using Data Mining and Artificial Neural Networks

Lemana Spahić[1(✉)], Leo Benolić[1], Safi Ur Rehman Qamar[1], Vladimir Simic[1,2], Bogdan Miličević[1,3], Miljan Milošević[1,2,4], Tijana Geroski[1,3], and Nenad Filipović[1,3]

[1] Research and Development Center for Bioengineering, BioIRC, Prvoslava Stojanovica, 34000 Kragujevac, Serbia
lemanaspahic@gmail.com
[2] Institute for Information Technologies, Jovana Cvijića, 34000 Kragujevac, Serbia
[3] Faculty of Engineering, University of Kragujevac, Sestre Janjić, 34000 Kragujevac, Serbia
[4] Metropolitan University, Tadeuša Košćuška 63, 11158 Belgrade, Serbia

Abstract. Coronary artery disease represents one of the most significant health burdens worldwide. As the onset of the disease is multifactorial in nature, physicians are often struggling with determining the rate of progression of the arterial narrowing caused by buildup of plaque. Computational models have brought upon a significant shift in the paradigm and the advent of Big Data and machine learning has enabled far better understanding of disease dynamics. This study is based on a cohort of patients recruited through SMARTool project for whom an extensive monitoring system was set up. In order to select for the most influential parameters on the progression of coronary atherosclerosis, different feature selection algorithms were used. The dataset used for development of the system for prediction of coronary plaque progression consisted of demographic data, data considering comorbidities and different blood cholesterol parameters. The developed artificial neural network showed significant strength in diagnosing progression of coronary arterial plaque, and features selected within this study indicate the high potential of machine learning to be used in clinical practice as well as that specific types of cholesterol are important markers impacting plaque progression.

Keywords: coronary artery disease · coronary atherosclerosis progression · machine learning · prediction · feature selection

1 Introduction

Coronary artery disease (CAD), more specifically coronary atherosclerosis (CATS), is one of the leading causes of death worldwide, accounting for approximately 17.9 million deaths annually (Bhatt et al. 2010). Coronary atherosclerosis is a condition marked by the accumulation of plaque on the artery wall, which is made up of fat, cholesterol, calcium, and other components. This causes arteries to gradually narrow, eventually occluding

M. Trajanovic et al. (Eds.): ICIST 2023, LNNS 872, pp. 3–13, 2024.
https://doi.org/10.1007/978-3-031-50755-7_1

and preventing blood flow (Libby 2021, Weber and Noels 2011) The most prevalent signs and symptoms of CAD are chest pain and discomfort, which are medically known as angina (Ong et al. 2018). Excessive plaque buildup in the arteries, which obstructs blood flow to the heart and the rest of the body, causes angina. Reduced oxygen and nutrition delivery as a result of this insufficient blood flow runs the risk of causing tissue damage and, in extreme circumstances, even death (Taqueti and Di Carli 2018).

Obesity, physical inactivity, an unhealthy diet, smoking, a family history of CAD or heart disease, and comorbidities such as diabetes, high blood pressure, and elevated blood cholesterol levels are all risk factors contributing to coronary artery disease (Yusuf et al. 2020, Khera and Kathiresan 2017). The significance of early detection and prevention techniques is emphasized by the fact that many of these characteristics can be altered by alterations in lifestyle and medical treatment (Arnett et al., 2019).

Aside from causing partial or total blockage of arteries, plaque can separate from the artery wall and flow into the bloodstream, resulting in an acute thrombotic event (Falk et al. 2013). This can lead to a heart attack or a stroke, which both have high morbidity and death rates (Benjamin et al. 2018).

It is essential to comprehend the relevance of factors influencing the evolution of atherosclerotic lesions in order to properly treat and prevent future cardiac events. Inflammation, endothelial dysfunction, and oxidative stress are a few of the mechanisms that have been linked to the development of atherosclerosis in studies (Hansson 2005). It has been demonstrated that pharmaceutical therapies that target these processes, such as statins and antihypertensive drugs, lower the incidence of CAD-related events (Trialists 2012). In addition, crucial elements of CAD management and prevention include stress management, regular physical activity, a heart-healthy diet, and quitting smoking (Eckel et al. 2014). These adjustments can enhance cardiovascular health overall, lower the risk of future cardiac events, and slow the development of atherosclerosis. Successful treatment and prevention of coronary artery disease depend on an understanding of the variables influencing the development of atherosclerotic plaques. It is possible to lessen the overall burden of CAD and enhance patient outcomes by focusing on modifiable risk factors and the underlying processes of atherosclerosis.

It is well known that atherosclerosis occurs because of an interplay of a variety of factors. The correlations of these factors to atherosclerosis is explored computationally in order to aid physicians in treating the exact cause of CATS, however research has found that most commonly several factors influence characteristics and hence optimal treatment strategy in the case of arterial plaque (Obermeyer and Emanuel 2016). For this reason, it is crucial to apply a multiscale approach to analysis of risk factors leading to CATS, starting from cells that make up the coronary arteries, through tissues to the entire organism and its environment (Tran et al. 2017). Pinpointing the most significant combination of risk factors for CATS development and treatment prognosis would enable physicians to target the disease with optimal treatment strategy and enable better patient outcomes.

Numerous computational methods, from basic statistics to more complex machine learning (ML) techniques, have been used to analyze the mechanism of coronary atherosclerosis (CATS) progression (Krittanawong et al. 2017). Data analytics has increased in capability with the development of big data and artificial intelligence (AI), allowing machines to learn from data without human input. Machine learning, a subfield of artificial intelligence, refers to a collection of techniques that may learn from data, i.e., find hidden patterns in databases with the intention of using the learned information to forecast future system outputs (LeCun et al. 2015). For descriptive or prescriptive purposes, a variety of mathematical models are used to approximate complex relationships in data (Hastie et al. 2009). As a result, ML has emerged as a valuable approach for various domains where collections of relevant real data could not be explored deterministically due to the presence of uncertainty or noise and nonlinear dependencies among features (Motwani et al. 2017).

Machine learning has been used to examine enormous datasets of patient medical records, genetic data, and imaging data in the context of CATS progression with the goal of revealing fresh insights and creating predictive models (Ambale-Venkatesh et al. 2017, Kwon et al. 2019). For instance, ML algorithms have been applied to discover novel biomarkers linked to CATS progression, forecast patient outcomes, and evaluate the effectiveness of treatments (Coelho-Filho et al. 2011).

These methods have the potential to advance clinical decision-making and increase understanding of CATS. Researchers and clinicians can create individualized treatment plans for patients at risk by utilizing the capacity of ML and AI to better understand the intricate interplay of factors that contribute to CATS progression (Al'Aref et al. 2019, Johnson et al. 2018). Improved patient care and management have resulted from the use of machine learning in the study of CATS progression, which has permitted the creation of more precise predictive models and sophisticated risk stratification tools (Shameer et al. 2018). As an illustration, ML algorithms have been used to forecast the probability of plaque rupture, a major contributor to acute coronary syndromes (Narula et al. 2013).

In order to better characterize the composition and form of atherosclerotic plaques, machine learning has also been used to analyze medical imaging, including coronary computed tomography angiography (CCTA) and intravascular ultrasound (IVUS) (Kolossváry et al. 2017, Lee et al. 2016). ML-guided imaging makes it possible to more precisely identify high-risk plaques, which helps in selecting the best therapies and tracking treatment response.

In the past few years, there has been a substantial increase in the use of machine learning (ML) approaches in the study of coronary artery disease (CAD) and coronary atherosclerosis (CATS), which has had a positive impact on risk assessment, diagnosis, prognosis, treatment, and management. The growing body of writing on the subject has reflected this tendency. The 2017 review "Artificial Intelligence in Precision Cardiovascular Medicine" by Krittanawong et al., is a significant instance. The review, examines the potential impact of machine learning and artificial intelligence on cardiovascular medicine, particularly in the areas of precision medicine, risk prediction, and treatment optimization (Krittanawong et al. 2017). A study by Motwani et al., entitled "Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis," was released in the same year. In it, the authors showed how clinical, demographic, and computed tomography angiography (CTA) data can be used to predict all-cause mortality in patients with suspected CAD (Motwani et al. 2017). The Multi-Ethnic Study of Atherosclerosis (MESA) cohort is used to predict cardiovascular events in a second compelling research article titled “Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis,” which was published in Circulation Research (Ambale-Venkatesh et al. 2017). Another outstanding review, “Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging,” provides an in-depth analysis of ML applications in cardiovascular disease with a focus on how ML can enhance cardiac imaging’s diagnostic precision, risk assessment, and treatment planning (Al’Aref et al. 2019). The potential uses and difficulties of AI in cardiology were examined in a study of “Artificial Intelligence in Cardiology,” which also examined ML methods for CAD and CATS as well as other cardiovascular illnesses (Johnson et al. 2018).

2 Methodology

The dataset used in this study is derived from SMARTool project. It is comprised of patient’s data from two-time moments, at the beginning of their monitoring and after certain timepoints with various data domains carefully collected to obtain the comprehensive description of disease and its mechanisms. Data about pharmacological treatment in combination with other parameters, represented by this dataset, could reveal significant medication factors that affect physician’s prescription decision and a plenty of research initiatives in the recent period is directed towards the application of data science approach for this purpose.

There is a standardized process to perform classification tasks and discover the potential the database holds, a concept known as Knowledge Discovery in Databases (KDD) (Frawley et al. 1992). Prior to starting with the application of ANN it is necessary to perform in depth analysis of the dataset and to mitigate all potential sources of error arising from ambiguities in the dataset. In general, the methodological approach taken in this research is:

– Data elucidation – defining the clinical significance to prognosis of CATS of individual parameters
– Database preparation – raw data preprocessing by dealing with missing values, different feature types, unbalanced dataset in terms of classes etc.
– Data division into training and testing dataset for ANN
– Training of ANN-based prediction model using a defined subset of data (training dataset)
– Validation of prediction model using a defined subset of data (testing dataset)
– Evaluation of the results

This is a generalized approach used in most studies concerning application of ML on real-world data. However, due to the complexity of the obtained dataset, an additional step was applied prior to implementation of ANN.

Missing values problem must be mitigated at first. The main justification for addressing this issue is the fact that ANNs cannot manage data with missing values. Additionally, if a feature value is absent for some patients, the algorithm may underestimate or exaggerate its significance for the discrimination issue. In some circumstances, it is possible to generate accurate guesses for the missing values by utilizing a feature’s unique characteristics and its correlation with other features. There are solutions that resolve this issue effectively:

- completely remove the missing value-related features. However, in case of high significance of a certain feature, this cannot be done.
- substitute particular values for missing ones. These values can either be the mean or median of the current numerical values, or for a nominal characteristic, the mode, which is the nominal value that occurs most frequently (this can also be used when the numerical values are discontinuous and typically few in number).

For further analysis, the features with more than 10% missing values were disregarded. The latter method was used for the remaining features.

Secondly, class imbalance must be addressed. Imbalanced class distribution can have a detrimental impact on ML models, and misclassifying the minor class can result in lower decision-making capacities. Performance measures that consider class distribution should be used for model evaluation. The majority of solutions rely on ensemble methods, while cost-sensitive learning is based on algorithm tweaks, and resampling techniques for data transformation. Resampling techniques are the most popular way for handling uneven data, according to (Haixiang et al. 2017). When a data collection has a limited number of minority cases, the SMOTE algorithm (Chawla et al. 2002) is demonstrated to be a good option. This method is based on the k-NN methodology, and it oversamples instances from minority classes by producing artificial minority class examples that are comparable to the ones that are already available. The key result is the identification of more focused decision areas for the minority class in the feature space, without encroaching on the region for the majority class.

Feature selection represents identification of subsets of significant features. This step is important for decreasing the dimensionality of the data and increasing the overall quality of data by disregarding irrelevant features. For the purpose of this study, the following feature selection methods were used: ReliefF, MRMR and wrapper technique with genetic algorithm. The relief (Chikhi and Benhammada 2009) estimates the contextual usefulness of the data in accordance to differences and similarities between the values for neigboring instances belonging to the different or same classes respectively. According to literature, the algorithm is able to correctly estimate the quality of attributes in problems with strong intra-attribute dependencies. mRMR (minimum redundancy-maximum relevance) feature selection algorithm (Parmar et al. 2007)selects the subset of features with the highest relevance in terms of their correlation to the output and least correlation to other input features. Lastly, wrapper algorithm in conjunction with genetic algorithm (Huang et al. 2007) is a feature selection algorithm evaluates the selected subsets of features by estimating the "area under the curve" (AUC) parameters using -fold cross validation.

As the aim of this study is to determine the risk and pace of progression of CATS, lipid-species, anti-thrombotic drugs, clinical data, risk factors and general biomarkers were selected as parameters of interest from the database. As the database consisted of 242 samples in total and an overall 112 parameters, with a significant class imbalance towards the minority class (patients with no progression or insignificant progression of CATS) it was important to decrease the number of used parameters to optimize the performance of the neural network. Upon applying the reliefF feature extraction algorithm, the following parameters were selected for use: No of plaques, CAD score, Max stenosis class, Gender, Hypertension, Dyslipidemia, Uric Acid, Statins, LDL, TG(52:2)

TG(18:1/18:1/16:0), CE(16:0), TG(52:3) TG(18:1/18:1/16:1), CE(20:3), PC(38:4), CE(18:2), CE(22,6), CE(18:1), PC(38:6), PC(40:6), TG(50:2)TG(16:0/16:1/18:1). Main statistical characteristics of the selected parameters in terms of mean, median and standard deviation are presented in Table 1. Along with correlation and p-value evaluation for each parameter.

Table 1. Summary statistics for the selected parameters.

Variable	Mean class 0	Median class 0	StdDev class 0	Mean class 1	Median class 1	StdDev class 1	Correlation	p-value
Gender	0.257	0.000	0.437	0.343	0.000	0.475	−0.150	0.391
Plaque count	3.886	2.000	3.115	5.314	5.000	2.806	0.125	0.474
CAD score	11.767	10.000	8.223	15.472	14.800	6.797	0.026	0.880
Max stenosis classes scan 2 (0 = no stenosis, 1 = < 30%, 2 = 30–50%, 3 = 50–70%, 4 = > 70%)	2.057	2.000	1.120	2.171	2.000	1.000	−0.034	0.845
Dyslipidemia	0.800	1.000	0.400	0.743	1.000	0.437	−0.294	0.086
Hypertension	0.857	1.000	0.350	0.743	1.000	0.437	0.133	0.445
Uric acid	5.874	5.800	1.241	5.494	5.200	1.061	0.049	0.778
LDL	89.789	84.000	31.172	72.859	68.200	31.361	−0.046	0.794
TG(52:2) TG(18:1/18:1/16:0) M	502.277	209.993	505.725	541.737	224.298	411.222	−0.024	0.890
TG(52:3) TG(18:1/18:1/16:1) M	111.813	107.451	24.585	113.250	114.270	24.804	0.151	0.388
TG(50:2)TG(16:0/16:1/18:1) M	104.865	99.764	38.872	126.073	119.885	55.284	−0.100	0.569
CE(16:0)	501.783	465.631	216.459	397.497	352.312	246.851	−0.126	0.469
CE(18:2)	1932.305	1897.609	491.412	1757.615	1665.873	373.051	−0.079	0.654

When observing the individual correlations, taking into account that 1 is the highest correlation while 0 is the lowest, it can be concluded that none of the parameters individually have significant impact on the overall classification. In order to account the difference between the classes, p-value estimation was calculated. The cut-off p-value, indicating statistical significance of intergroup variability is $p < 0.05$. As it can be concluded from Table 1, there is no statistical significance observed for individual parameters.

Upon feature selection and data preprocessing, a total of 204 samples remained in the database with 79% of them belonging to the majority class (patients with significant progression of CATS). The dataset was divided in an 80%–20% manner for training and testing respectively.

3 Results and Discussion

As artificial neural networks are the most robust algorithms for classification tasks and most prominently used when complex datasets with different types of data are used, they were chosen as the classification algorithm for prediction of CATS progression. During the development of the ANN, several different combinations of activation functions and the number of neurons in the hidden layers were tested in order to improve the overall performance of the system. The measure of performance for this binary classification task was mean squared error.

In the field of machine learning, hyperparameters are parameters that are defined before the learning process begins, which distinguishes them from other parameters learned during model training. These can include things like the learning rate, the number of hidden layers in the neural network, the number of neurons in each layer, and the type of optimizer used, among other things. The goal of hyperparameter optimization is to search the hyperparameter space for an optimal set of hyperparameters that will produce a model with the best performance, usually evaluated through a defined metric. Predetermined, such as precision, mean error, or a combination of several metrics. This is an important step in ANN development because the selected hyperparameters can significantly affect the performance of the model. For example, a high learning rate may cause the model to converge rapidly but miss the global minimum, while a low learning rate may cause slow convergence or even stagnation at a suboptimal point.. Likewise, the number of hidden layers and neurons in each layer can significantly affect the model's ability to capture complex patterns in the data, potentially leading to missing or overfitting pages. Level. In the present study, we used the Keras Tuner library to optimize the hyperparameters for our ANN model. We used a variety of techniques, including stochastic search and Bayesian optimization, to tune the model's hyperparameters, such as the learning rate, the number of hidden layers, and the number of neurons in each layer.. Through rigorous optimization, we can significantly improve the performance of the model, highlighting the important role of hyperparameter optimization in ANN development.

A closer examination of the training process revealed a discrepancy between training and validation metrics (Fig. 1). Specifically, while the Mean Squared Error (MSE) and accuracy on the training set showed a consistent improvement throughout the training iterations, the validation MSE and accuracy did not follow the same trend. Instead, they fluctuated and did not converge with the training metrics, indicating a potential issue with overfitting.

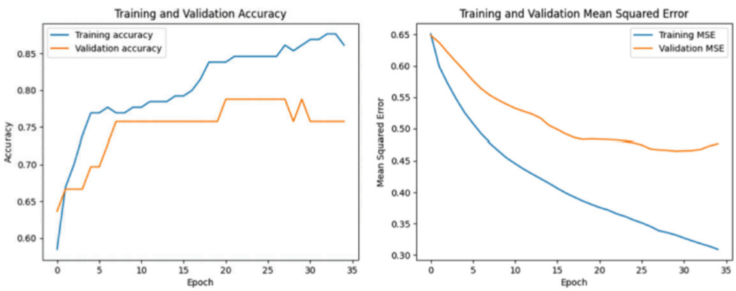

Fig. 1. Training and validation accuracy and MSE before regularization

Overfitting is a common problem in machine learning and occurs when a model learns to perform exceptionally well on training data but fails to generalize to unseen data. In our case, the continuous improvement of the training indices suggests that our ANN may have become too specialized in capturing the nuances and noise of the training set, which is not necessarily the case. It is necessary to apply to the authentication training set or to the general public. Several factors can contribute to this overfitting phenomenon. One possibility is the complexity of our ANN. The model, equipped with many layers and a large number of neurons, may have too much capacity, thus learning the training data too well. Another factor could be the lack of regularization techniques, such as dropout or L1/L2 regularization, which are often used to avoid overlearning by adding constraints to the learning process. To solve these problems and improve the performance of ANNs, we have studied several strategies. First, we tried to simplify the model by reducing the number of layers or neurons. Second, we explored the use of regularization techniques to limit model learning and avoid overfitting. For this purpose, dropout layers were used. After many iterations of the system, the overall performance has improved significantly (Fig. 2), eventhough the overfitting problem persisted.

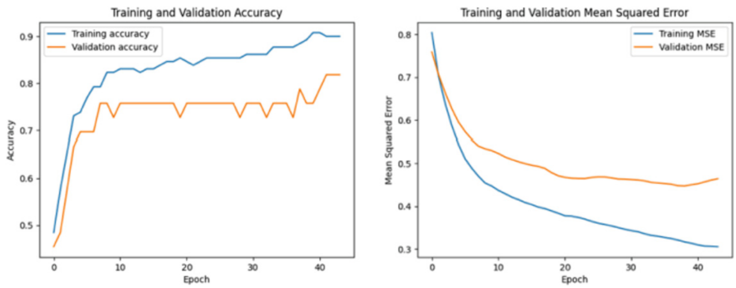

Fig. 2. Training and validation accuracy after regularization

In spite of overfitting during training, the final validation of the system performed on an independent testing set showed an accuracy of 81.81% for the developed system (Fig. 3).

The specificity of 37.5% indicates that the ANN does not generalize well for the insignificant plaque progression samples. This problem persisted across all iterations of the ANN. Considering significant class imbalance of the dataset where only 22% of the data corresponded to the minority class.

Future perspectives of this research should include expanding the dataset with more samples of minority class to decrease the class bias as this would significantly improve the overall accuracy of the system and potentially render it suitable for practical application in form of a decision support system for CATS diagnosis.

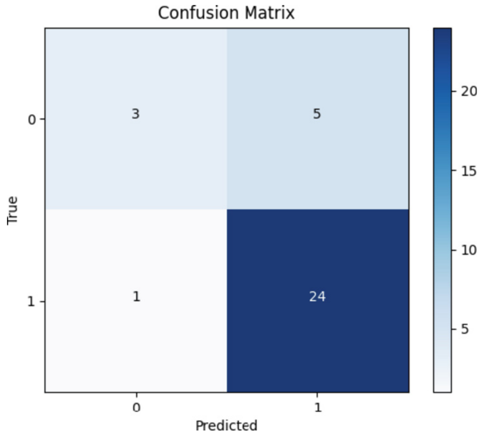

Fig. 3. Confusion matrix for the developed system (accuracy 81.81%, sensitivity 96% and specificity 37.5%)

4 Conclusion

This study demonstrates the significant potential of machine learning, especially artificial neural networks (ANNs), in the study of coronary artery disease (CAD) and coronary atherosclerosis (CATS). The application of ANNs in this study highlights the ability to process complex multidimensional data and create models suitable for risk assessment, diagnosis and prognosis.

In the course of this work, an ANN model with 19 input parameters representing binary classifiers of two output classes was developed. Application of hyperparameter optimization techniques, by means of the Keras Tuner library, was key to fine-tuning the model's architecture to improve performance.

By employing an early-stopping regularization strategy, the overfitting problem was effectively resolved, adding an additional layer of complexity to the training process. These strategies help prevent the model from learning noise in the training data, resulting in a better ability to generalize to unseen data. The final model achieved an accuracy of 81.81%, a significant improvement over the original model. This improved model demonstrates the importance of careful hyperparameter tuning and good regularization techniques in developing robust and reliable ANN models.

However, the model had a specificity of 37.5% and a sensitivity of 96%. High sensitivity indicates that the model is very good at identifying truly positive outcomes. Conversely, lower specificity suggests that the model may have more difficulty identifying true negatives. This disparity underscores the need for balanced datasets in training and highlights the importance of considering various performance metrics beyond mere accuracy when evaluating a model's performance.

Despite the challenges faced during model development, this study highlights the potential of ANNs and machine learning in biomedical research. As the understanding and application of these advanced computational techniques continue to advance, they will undoubtedly play a key role in advancing precision medicine, improving patient outcomes, and expanding the understanding of complex diseases such as CAD and CATS.

It would be interesting to explore other machine learning techniques, and even different model combinations, to see if specificity can be further improved without significantly compromising sensitivity in the future. Moreover, applying these methods to larger and more diverse datasets will lead to a more comprehensive and more informed understanding of the potential in this area.

Overall, the results of this study suggest that ANNs, if properly tuned and regulated, have great potential in analyzing CAD and CATS. It is hoped that these findings will contribute to the ongoing debate in this field and stimulate further research to realize the full potential of machine learning in cardiovascular medicine.

Acknowledgements. This paper is supported by the DECODE project (www.decodeitn.eu) that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 956470. This article reflects only the author's view. The Commission is not responsible for any use that may be made of the information it contains.

References

Al'aref, S.J., et al.: Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. Eur. Heart J. **40**, 1975–1986 (2019)

Ambale-Venkatesh, B., et al.: Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. Circ. Res. **121**, 1092–1101 (2017)

Arnett, D.K., et al.: 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation **140**, e596–e646 (2019)

Benjamin, E.J., et al.: Heart disease and stroke statistics—2018 update: a report from the American Heart Association. Circulation **137**, e67–e492 (2018)

Bhatt, D.L., et al.: Comparative determinants of 4-year cardiovascular event rates in stable outpatients at risk of or with atherothrombosis. JAMA **304**, 1350–1357 (2010)

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)

Chikhi, S., Benhammada, S.: ReliefMSS: a variation on a feature ranking ReliefF algorithm. Int. J. Bus. Intell. Data Min. **4**, 375–390 (2009)

Coelho-Filho, O.R., et al.: Stress myocardial perfusion imaging by CMR provides strong prognostic value to cardiac events regardless of patient's sex. JACC: Cardiovascular Imaging **4**, 850–861 (2011)

Eckel, R.H., et al.: 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation **129**, S76–S99 (2014)

Falk, E., Nakano, M., Bentzon, J.F., Finn, A.V., Virmani, R.: Update on acute coronary syndromes: the pathologists' view. Eur. Heart J. **34**, 719–728 (2013)

Frawley, W.J., Piatetsky-Shapiro, G., Matheus, C.J.: Knowledge discovery in databases: An overview. AI Mag. **13**, 57 (1992)

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G.: Learning from class-imbalanced data: review of methods and applications. Expert Syst. Appl. **73**, 220–239 (2017)

Hansson, G.K.: Inflammation, atherosclerosis, and coronary artery disease. N. Engl. J. Med. **352**, 1685–1695 (2005)

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York (2009)

Huang, J., Cai, Y., Xu, X.: A hybrid genetic algorithm for feature selection wrapper based on mutual information. Pattern Recogn. Lett. **28**, 1825–1844 (2007)

Johnson, K.W., et al.: Artificial intelligence in cardiology. J. Am. Coll. Cardiol. **71**, 2668–2679 (2018)

Khera, A.V., Kathiresan, S.: Genetics of coronary artery disease: discovery, biology and clinical translation. Nat. Rev. Genet. **18**, 331–344 (2017)

Kolossváry, M., et al.: Radiomic features are superior to conventional quantitative computed tomographic metrics to identify coronary plaques with napkin-ring sign. Circul. Cardiovascular Imaging **10**, e006843 (2017)

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., Kitai, T.: Artificial intelligence in precision cardiovascular medicine. J. Am. Coll. Cardiol. **69**, 2657–2664 (2017)

Kwon, J.-M., et al.: Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. PLoS ONE **14**, e0219302 (2019)

Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)

Lee, S.-E., et al.: Rationale and design of the Progression of AtheRosclerotic PlAque DetermIned by Computed TomoGraphic Angiography IMaging (PARADIGM) registry: a comprehensive exploration of plaque progression and its impact on clinical outcomes from a multicenter serial coronary computed tomographic angiography study. Am. Heart J. **182**, 72–79 (2016)

Libby, P.: The biology of atherosclerosis comes full circle: lessons for conquering cardiovascular disease. Nat. Rev. Cardiol. **18**, 683–684 (2021)

Motwani, M., et al.: Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. Eur. Heart J. **38**, 500–507 (2017)

Narula, J., et al.: Histopathologic characteristics of atherosclerotic coronary disease and implications of the findings for the invasive and noninvasive detection of vulnerable plaques. J. Am. Coll. Cardiol. **61**, 1041–1051 (2013)

Obermeyer, Z., Emanuel, E.J.: Predicting the future—big data, machine learning, and clinical medicine. N. Engl. J. Med. **375**, 1216 (2016)

Ong, P., et al.: International standardization of diagnostic criteria for microvascular angina. Int. J. Cardiol. **250**, 16–20 (2018)

Parmar, D., Wu, T., Blackhurst, J.: MMR: an algorithm for clustering categorical data using rough set theory. Data Knowl. Eng. **63**, 879–893 (2007)

Shameer, K., Johnson, K.W., Glicksberg, B.S., Dudley, J.T., Sengupta, P.P.: Machine learning in cardiovascular medicine: are we there yet? Heart **104**, 1156–1164 (2018)

Taqueti, V.R., Di Carli, M.F.: Coronary microvascular disease pathogenic mechanisms and therapeutic options: JACC state-of-the-art review. J. Am. Coll. Cardiol. **72**, 2625–2641 (2018)

Tran, J.S., Schiavazzi, D.E., Ramachandra, A.B., Kahn, A.M., Marsden, A.L.: Automated tuning for parameter identification and uncertainty quantification in multi-scale coronary simulations. Comput. Fluids **142**, 128–138 (2017)

Trialists, C.T.: The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. Lancet **380**, 581–590 (2012)

Yusuf, S., et al.: Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. The Lancet **395**, 795–808 (2020)

Rule-Based System for Pregnancy Monitoring

Dragana Filipović(✉), Anđela Trajković, and Siniša Nikolić

Faculty of Technical Sciences, The University of Novi Sad, Novi Sad, Serbia
{dragana.filipovic,trajkovic.andjela,sinisa_nikolic}@uns.ac.rs

Abstract. During the pregnancy period, it is important to monitor the baby's and the mother's condition. Many tests are necessary during pregnancy, from screen tests for chromosomal abnormalities detection in a fetus to disease diagnosis tests for the mother. The results of those tests should be taken into consideration when defining prenatal care policies. Inadequate prenatal care is associated with an increased risk of preterm delivery and low birth weight newborns. This paper proposes a software solution for improving prenatal care by allowing cooperation between doctors and expert systems. Integrating expert system knowledge and experience within the rule-based system to guide doctor decisions can produce more adequate and safer prenatal care and reduce the infant mortality rate. The proposed system detects diseases based on symptoms and chromosomal abnormalities based on different parameters. Also, it monitors cardiotocography and calculates its results. This system is an expert system based on rules. During runtime, rules can be modified and changed without affecting the application workings. The proposed system is not a replacement for doctors, it can be seen as a support system for medical personalities.

Keywords: Expert system · knowledge-based systems · Drools · prenatal care · chromosomal abnormalities · cardiotocography · an indication of the disease of a pregnant woman

1 Introduction

During pregnancy, it is vital to monitor the mother's and baby's condition and the baby's development. This is done through regular visits to the doctor, where various tests and analyses are performed on the mother and baby. These visits represent prenatal care, which is essential for the health and well-being of both the mother and the baby. Inadequate prenatal care is associated with an increased risk of preterm delivery and low birth weight of newborns, which results in a higher infant mortality rate. If the mother is getting adequate prenatal care, it can help prevent and address health problems in both mother and baby, and in that way, it can help save and improve the lives of both mother and baby [1]

In the first and the second trimester, double, triple, and quadruple screens, as well as amniocentesis, are used to check whether the baby has anomalies. Fetal anomalies are physical or genetic defects of the fetus and these screens can identify major chromosomal abnormalities such as Down, Patau, and Edwards syndrome.

M. Trajanovic et al. (Eds.): ICIST 2023, LNNS 872, pp. 14–22, 2024.
https://doi.org/10.1007/978-3-031-50755-7_2

Cardiotocography (CTG) is a machine that monitors fetal heart rate and uterine contractions [2]. It can detect problems with fetal heart rate and can report them to the doctor. In the third trimester, CTG is done.

Besides tests that indicate problems with the baby, tests are also run on the mother to monitor her condition. Her weight, blood pressure, and different symptoms that are manifested are being tracked. These symptoms indicate different illnesses. Some of the illnesses that a mother can have during pregnancy are preeclampsia, eclampsia, gestational diabetes, hypertension (high blood pressure), kidney disease, and others.

In the knowledge-based system [3], we can group and store key information about babies' and mothers' conditions in one place. The system is primarily intended for doctors and pregnant women as a web application. Rules in the system are written in Drools [4]. By relying on the application, the mother can be diagnosed with illness depending on manifested and reported symptoms, chromosomal anomalies can be detected in the fetus by calculating risk from the results of the double, triple, and quadruple test, as well as from results of amniocentesis procedure. The doctor can be alarmed when a critical condition for the baby is detected during CTG, where the results of CTG are calculated by the system. The doctor only needs to act as a supervisor of the whole process.

The following section defines the research questions and methodology. The third section provides an overview of related works. The analysis of existing solutions is described in the fourth chapter. The software solution design is in the fifth chapter. The sixth chapter states the implementation specifics. The last chapter discusses systems improvements and possible future work.

2 Research Questions and Methodology

The research in this paper will try to answer to 2 questions:

- Is it possible to make a support system for pregnancy monitoring by relying on the patients' and fetuses' symptoms and calculating CTG results based on values detected with the CTG machine?
- If so, can pregnancy monitoring be done by using a rule-based system where domain knowledge is stored in the form of some formally defined rules?

The methodology consists of the following:

- Review of related work concerning expert systems that are designed to solve problems mimicking the human experts in the field of prenatal care, tests, CTG, and medicine. There are many references to how formally defined rules have been used to represent expert knowledge.
- Analysis of some of the existing popular applications that are used for pregnancy monitoring.
- Collection of the domain knowledge from the field of prenatal care, tests, diagnosis, and CTG that will include consultation with field experts and consulting books.
- Choosing an appropriate rule-based system

3 Related Work

Pregnancy often requires doctors to conduct a large number of tests and track the patient's symptoms, which can be a demanding and exhausting process. The results of these tests must also be calculated, adding to the time and effort required.

In these situations, expert systems [4] and systems based on rules can become relevant and helpful because they can store large amounts of knowledge, which may be accessed quickly and manipulated easily.

Paper [5] describes an expert system for the diagnosis of disorders during pregnancy using the forward chaining method. It can diagnose different disorders such as preeclampsia, eclampsia, abortus, and others depending on symptoms such as excess headaches, proteinuria, blurred vision, vaginal bleeding, and others.

Article [6] presents a web-based forward-chaining expert system for maternal care. In Ethiopia maternal mortality and morbidity rate is high because patient to doctor ratio is 1 doctor to 1000 patients. That is why this system is made, to help diagnose diseases depending on symptoms and to reduce maternal mortality and morbidity rate.

A rule-based algorithm for intrapartum cardiotocography pattern features extraction and classification using MATLAB is described in the paper [7].

Most of the listed works are focused on one specific domain. The first and the second paper are focused on the diagnosis of diseases depending on the symptoms, while the third one is focused on calculating CTG results.

At the moment of writing this paper, not one paper was found that focuses on detecting chromosomal abnormalities and calculating screen test results.

The software solution proposed in this paper follows on the previously mentioned research and will focus on: chromosomal abnormalities detection via double, triple, and quadruple tests and amniocentesis which calculate results based on multiple different parameter values, disease diagnosis based on symptoms and monitoring CTG, calculating its results in real-time and alarm system for the doctor when critical fetus condition is determined.

4 Analysis of Existing Solutions

A large number of applications for pregnancy monitoring have been developed in recent years, they enable patients to track their appointments and tests on the calendar. They give users information on their baby's size and tips for healthy foods.

A few advanced solutions will be described in this chapter. These solutions are CTG Home monitoring [8] and MedTel Remote Pregnancy Monitoring [9].

1. CTG Home monitoring

CTG Home monitoring enables pregnant women and midwives to make a CTG at home or in the clinic. These data are sent directly to a portal in the cloud. The hospital can view these data in real-time or later.

2. MedTel Remote Pregnancy Monitoring