Bulletin T.CXLVIII de l'Académie serbe des sciences et des arts − 2025
Classe des Sciences mathématiques et naturelles
Sciences mathématiques, № 50

MACHINE AND FEDERATED LEARNING AS A SUPPORT TO MEDICAL DECISIONS

MIRJANA IVANOVIĆ

(Accepted at the 9th Meeting, held on December 26, 2025)

A b s t r a c t. This survey on machine and federated learning as a support to medical decisions is an extended version of my inaugural lecture delivered on 3*rd meeting of Department of Mathematics, Physics and Geosciences in the Serbian Academy of Sciences and Arts under the same title on April* 25, 2025.

AMS Mathematics Subject Classification (2000): 03B42, 68Q10, 68M14, 68T99.

Key Words: Machine learning, Federated learning, Medical data, Explainable artificial intelligence.

1. *Introduction*

After the first great successes in the development of various approaches and aspects of artificial intelligence (AI) in the 1960s and 1970s, there was a silent period characterized by slow development of this field that lasted for several decades. However, in the last two decades, and especially in the last few years, artificial intelligence has experienced an incredible boom and rapid development of both theoretical aspects and practical applications. Today, it is one of the leading areas of research and a driving force of progress, primarily in science, industry, but also in business. Furthermore, the development of techniques, approaches, and services based on artificial intelligence finds its place and plays a significant role in almost all disciplines and areas of human work, creativity, and life. The field of AI encompasses a wide range of subfields. Some subfields gain more importance and use in some periods, while others lose attractiveness for a period of time, but very often return in a new and modified form and are applied in solving some existing and newly emerging problems.

One among the most significant applications of artificial intelligence techniques and approaches is in various fields of medicine and healthcare.

Applications are possible in almost all medical and healthcare domains, including assistance in diagnosing and monitoring the treatment of patients diseases; analysis of medical images as well as monitoring and analyzing dynamic processes in the body; discovery, design and testing of new drugs; personalized medicine [13]; monitoring patients and adjusting their therapy; recommendations and maintenance and/or increase of the quality of life of patients as well as a number of other activities. AI techniques and approaches are a unique mechanism for identifying patterns and detecting characteristic phenomena in medical and clinical data.

Medical and clinical data include information related to the health of patients and, in addition to basic data on the health of patients, content and data collected during patient care and additional medical examinations and research. Clinical data[1] are a subset of medical data, specifically referring to information collected during clinical trials or patient care, such as electronic health records, administrative data, and health surveys.

Artificial intelligence techniques also provide insight into data quality and increase their reliability for better patients diagnosis and treatment. The starting point of the application of AI approaches in medicine and health domains is the availability of patient data. Patient data are often complex. Generally, patients data record consists of data represented in different formats: alphanumeric, dates, categorical values, text records, time series (i.e. values of certain data measured at different points in time), multimedia objects such as images, sound, etc. Unfortunately, very often there is no uniform and standardized form and data model even for patients suffering from the same or similar diseases, especially if their data are located in different health and medical institutions. Inconsistent and incompatible data are not easy and simple to integrate into a centralized repository in order to be further used and processed efficiently. An additional problem arises when it is necessary to aggregate patient data from various heterogeneous sources represented in different formats and structures.

Contemporary, intelligent medical systems usually function in a distributed mode, i.e. data from multiple medical and healthcare institutions are integrated and used. Therefore, it is necessary to present data in a standardized form in order to enable communication and interoperability between multiple institutions and information

[1] `https://guides.lib.uw.edu/hsl/data/findclin`

systems. Data are of essential importance in medical and healthcare information systems (MIS), especially when various AI techniques and methods are applied to achieve the most reliable results. AI approaches support doctors' decisions and recommendations to patients to increase quality of their conditions and further treatment.

This paper contains results of lecture presented at the Department of Mathematics, Physics and Geosciences, SASA. Essential aspects of collecting, processing and building datasets from patients electronic health records and the application of artificial intelligence techniques to produce reliable predictive models are considered. Special emphasis is put on distributed processing of medical datasets available at several institutions in federated learning (FL) manner. Some of the results are original scientific research contributions from the author and her collaborators from the Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad.

The paper is structured as follows. The next section is dedicated to the quality of medical data and the standard representation of data models. Section 3 is focused on current research and applications of artificial intelligence techniques in medical domains and usual procedures for producing predictive models. Section 4 presents the original results of the author and her collaborators, which were achieved within the framework of a project funded by the European Commission: ASCAPE - Artificial intelligence Supporting Cancer Patients across Europe, Grant Agreement No 875351[2]. Finally, the conclusion briefly presents some of the possibilities of applying some AI techniques in improving the functionality of future smart, intelligent and personalized systems.

The text is prepared based on the author's and her team's research papers published in journals, as well as elements from invited lectures delivered at several conferences, published in their proceedings ([8], [9], [10], [11], [12]).

2. *Medical data quality and their aggregation*

Clinical data[3] used in healthcare and medical practice, as well as in scientific research, are essential and vital resources that influence the quality of medical decisions. The term clinical data refers to structured and unstructured information related to electronic patient health records, disease codes, symptoms and other clinical details that are used for analytical studies in clinical decision-making. Electronic health records (EHRs) or electronic medical records (EMRs) ensure that patient data are stored in a secure location. The data stored in the EHRs include basic patient data, all vital parameters, laboratory test results, assessments and notes made by

[2] https://www.ascape-project.eu/

[3] https://guides.lib.uw.edu/hsl/data/findclin

healthcare professionals on the patient's condition, diagnoses according to the ICD (International Classification of Diseases), as well as a range of other data.

However, due to time constraints in healthcare institutions in real-world settings, these data are rarely enriched with data from other sources, and their quality and usefulness are often questionable [21].

However, for better quality and use of patient data in modern research and implementation of smart and intelligent medical and healthcare systems and services, EMK additionally includes external sources of information such as the results of questionnaires filled out by patients, dietary habits, monitoring of daily physical activities, weather conditions and other data that help to have a better view on the patient's health status. These procedures are a necessary prerequisite for obtaining richer datasets and, accordingly, more reliable results of their processing [1].

In order to process and use such data for multiple purposes, it is necessary to organize and present them in a standard format, which are supported by standard terminologies and ontologies such as Classification of Diseases, ICD-9, ICD-10, IDC-11, LOINC, CPT, etc. [10]. In modern practice and research, SNOMED CT – Systematized Nomenclature of Medical Clinical Terms – is most often used. It represents a systematized healthcare terminology as well as a medical standard for presentation of clinical findings, symptoms, diagnosis, body structures and organs, and a number of other concepts and processes in medicine. SNOMED CT[4] is currently widely used in numerous countries in medical and healthcare information systems. Its use also contributes to a high degree of interoperability between different data sources, information systems and software-supported services in medical and healthcare institutions.

Data presented in this way can be more easily transformed into other formats that are more suitable for various types of processing and the application of various techniques, including artificial intelligence methods.

As a next step towards better organization and representation of complex medical data, it is important to produce their quality models, and in contemporary research and applications, among the frequently used systems for this purpose is HL7 FHIR Health Level 7, Fast Healthcare Interoperability Resources[5]. This highly functional system possesses mechanisms for representing the content of FHIR resources in various formats (XML, JSON, Turtle, .xls...) that are more suitable for use in other development environments. One of the common ways to produce a standard medical data model and its role in the process of applying AI/ML techniques is shown in Figure 1 (a more detailed description of this topic and project results is presented in Section 4).

The first step in producing the final data model in accordance with the HL7

[4] https://www.snomed.org/

[5] https://www.healthit.gov/topic/standards-technology/standards/fhir

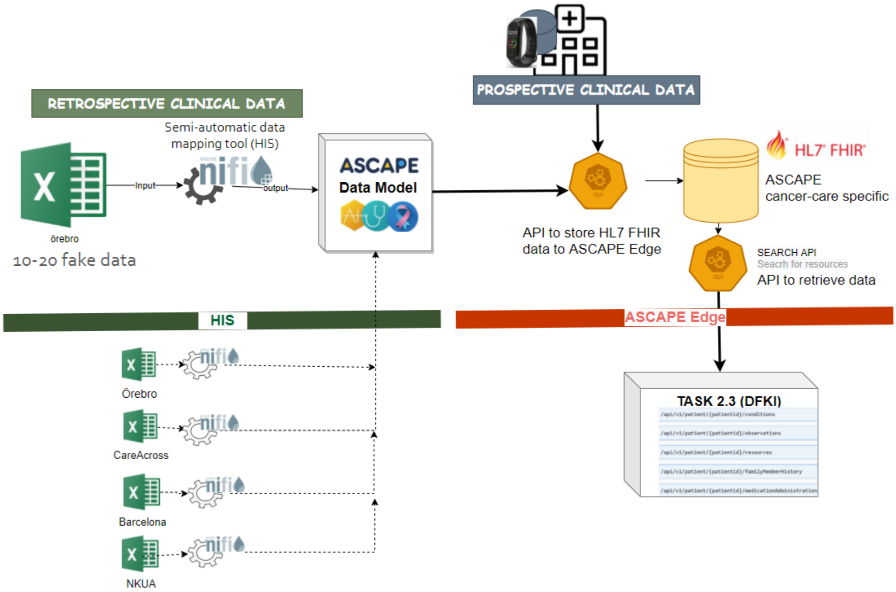

Figure 1: A typical organization of medical data and its standardized representation (https://www.ascape-project.eu/, internal documentation)

FHIR standard was to define the initial ASCAPE data model based on datasets of already available (retrospective) patients EHR in 4 healthcare institutions (left side of the figure). The data about patients collected through the project duration have been included in data model. Apart from parameters existing in retrospective data the data about new patients have been merged with additional data collected from other sources such as wearable devices, environmental parameters, dietary habits, etc. Consequently, the final, expanded, HL7 FHIR ASCAPE data model was produced (prospective data, right side of the figure).

Generally speaking, to achieve better and higher quality predictive models that better support medical decisions, a great number of parameters for patients health are needed. Additionally, richer datasets that include high number of patients are required. Furthermore, it is very important to understand the advantages of data aggregation compared to individual patient data.

Individual patient data are created by the patient's interaction with doctors and the healthcare system and helps doctors make more effective decisions about the treatment of that particular patient. On the other hand, data aggregation represents the merging of information/data from of patients suffering from similar or the same diseases from multiple systems/sources in order to achieve better quality datasets.

Better quality and higher size of datasets is a prerequisite for training more reliable predictive models using artificial intelligence approaches.

Data of many similar cases, i.e. patients suffering from the same or similar diseases (anonymized aggregated data), give doctors opportunity to have a look at data collected from a larger number of patients. They can consider data from different angles and perspectives that help them to make better and more appropriate decisions for individual patients. Aggregated data management in medical and health institutions also offers better insight that contributes to improvement of strategic planning and the creation of more reliable healthcare systems and services.

When initial/raw set of data containing records form a large number of patients is available, it is necessary to check different aspects, characteristics, and consistency of the data. If it is necessary, and usually it is, several characteristic activities/steps should be applied to improve their quality:

- Data cleaning – i.e., omitting irrelevant, redundant and inadequately measured data values, their transformation (for example categorical values, i.e., transforming date into a numerical value, also some other kinds of appropriate changes).

- Identifying influential features/properties and constructing potentially new ones that are more influential. New features are based on existing, and it is necessary to assess their power of prediction.

- Imputation of missing data in the patients EHR, as certain number of data values are often missing. It can lead to lower quality AI predictive models. Missing values are very often added/imputed with some estimated values based on existing ones during the standard workflow for training predictive models. Value estimation is performed using various techniques and approaches: the mean values of existing data for that particular feature or the most frequent values among existing or some constant values, values of existing data values estimated using for example k-NN (k nearest neighbors) approach, deep learning approach [16], etc.

Privacy protection is a crucial issue in the processing of sensitive medical data that should be taken into account [20]. Crucial demand in such situations is to make it impossible to discover the personal data of patients at any stage of data processing and further use of the achieved results. For this purpose, a number of techniques and approaches are used, starting from simple ones such as: character/record masking, anonymization, collective de-identification, pseudonymization, generalization, and etc. However, in the past two decades, many complex approaches such as K-anonymity, L-diversity, and T-Closeness have been developed, improved, and frequently used for processing data in wide range of domains.

Moreover, recently, two techniques have also been widely used and dominant in application of ML and data mining techniques of sensitive datasets: Differential Privacy (DP) and Homomorphic Encryption (HE). They play a key role in Cloud/Edge distributed medical systems for processing complex patient data [3], [11].

DP represents a minimal systematic random modification of data in order to reduce the available information about a single patient while preserving the quality of trained predictive models [23].

DP covers a set of different techniques starting from simple random mixing of input data to introducing noise over the data using, most often, Gaussian or Laplace distributions, or exponential mathematical mechanisms. Local DP ensures the privacy of the original data and helps to improve the results of processing complex and aggregated data.

HE is a technique that allows the application of arithmetic operations on encrypted texts without the need to decrypt the original data, while preserving the security and safety of the data. This technique is unavoidable in contemporary Cloud/Edge and distributed data processing, especially for use in FL federated learning settings. HE approaches perform calculations on encrypted data without decrypting them and the results obtained are identical to the results that would be obtained by processing unencrypted data.

3. *Training and using predictive models based on machine learning and federated learning*

After cleaning and preparation of purified patients datasets, if we want to use them for various processing and obtaining useful and influential new data and knowledge, application of wide range of AI techniques and approaches should be performed. This process ensures training and obtaining high-quality predictive models. Trained predictive models in medical domains provide support to medical experts in predicting the course of the disease, especially in order to avoid health risks such as adverse drug reactions, genetically determined resistance to treatment, inadequate recommendations and treatments, quality of life issues etc. The usual way to produce well trained predictive models includes 3 characteristic processes/phases:

a) application of various AI techniques on available datasets in order to discover and observe appropriate characteristic patterns of behavior in the data and to achieve new knowledge;

b) the results and conclusions produced by different predictive models are compared and validated to conclude which models provide the best results and offer adequate insight into the patient's condition for recommendation of the most appropriate further treatment procedures;

c) the last stage is application of already trained, developed and validated predictive models in real-world settings, on new datasets of new patients, or on expanded data of existing patients.

AI and ML offer a whole range of algorithms and procedures used to train predictive models. In this paper, we briefly present 2 groups of frequently used approaches, which are also selected for implementation of the architecture within the ASCAPE project, and which will be presented in more detail in Section 4.

a) **Classification models** are a type of ML model that divides data points into predefined groups called classes. Classifiers are a type of predictive modeling that learns class characteristics from input data and learns to assign possible classes to new data according to those learned characteristics [2]. Widely used approaches for this purpose are – NB (Naive Bayes) algorithm, KNN (K-nearest neighbors), DT (decision trees), CART (Classification and regression trees), CART clustering, C5.0, RF (random forests), LR (logistic regression), ANN (artificial neural network), SVM and LDA (linear discriminant analysis) and so on.

b) **Regression models** there are different definitions of regression, one of which is "Regression is a supervised learning technique that models the relationship between input features and a continuous target variable, using statistical methods to predict the target variable based on new input data. Regression models shift through large numbers of variables, identifying those with the greatest impact outcomes. Regression is foundational to ML, especially for predictive use cases."[6] These models are useful in predicting of a wide range of values such as: numerical values, specific types of values connected to the domain of application, different types of interventions, continuous values, etc. Typical approaches used for these purposes are: LINEAR (linear regression), RIDGE (ridge regression), LASSO (lasso regression), ELASTICN (elastic network regression), KRIDGE (kernel ridge regression), SVM, RF (random forests), KNN (K-nearest neighbors), ADAB (AdaBoost regression), TFNN (TensorFlow neural network) and so on.

3.1. Training predictive models in federated learning mode

Federated Learning is a rather new, modern and widely used learning technique based on the principles of AI and ML. The FL process involves two key components: Client and Edge node (EN). In such environments there can be any number of devices (ENs) varying from several to several hundred thousand. The second component is the Server or Cloud, i.e. one central node/coordinator. The Server coordinates the training process between all ENs that participate in the construction of the global predictive model.

[6] `https://www.snowflake.com/guides/types-regression-models-ml/`

In specific architectures and systems, the number of clients is known in advance. In each client sensitive data for training the model are stored. Each client receives a copy of the global training model and updates and changes it using available local data relating to the particular client. The coordinator server produces a common global model based on the local models, while the local data on which the local models are trained remains on the original client and is isolated from other clients. More precisely, the locally updated weights of the local models are sent back to the coordinator server, where a new updated global model is produced. To produce a new global model, different strategies of aggregation of the weights of the local models are used (Figure 2). When one cycle of model training is completed, all clients in the next iteration again participate in training the new global model that is forwarded to them.

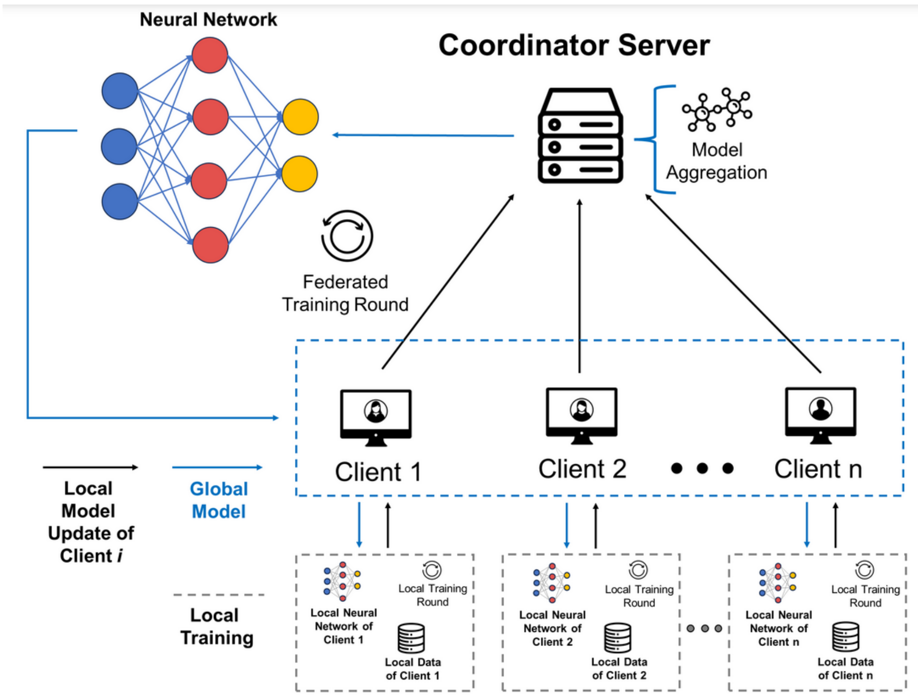

Figure 2: General approach to training a global model in the federated learning mode [18]

If the client datasets contain sensitive patient data, they remain decentralized, and FL keeps the data in health institutions locally and transmits only the updated local model to the coordinator to improve next version of global model. More precisely, clients train their own copy of the global model (which is at that moment seen as local model) on local data.

When the training is complete, i.e. when the model has achieved satisfactory results, the updated weights of that new local model are sent back to the coordinator to contribute to the creation of a new or update the existing shared global model [6], [11].

This approach is also incorporated into the ASCAPE architecture, which will be presented in more detail in Section 4.

There are several characteristic approaches to building global models in FL settings: incremental, cyclic incremental, competitive, and semi-competitive. As the ASCAPE project applied an incremental and semi-concurrent [5] approach (Figure 3), their basic characteristics are briefly presented below.

Figure 3: Incremental (a) and semi-concurrent (b) modes of federated learning

In the incremental mode, updates are performed incrementally, starting from client 1 to client n. The model is initialized by client 1 and trained over multiple epochs on the local dataset available at client 1, after which the model is forwarded to the coordinator. In addition, this local model is also forwarded to client 2, which accepts it and further updates it using the local data located at client 2. The entire process is propagated further to the remaining clients and ends when all local models are sent to the coordinator, which aggregates them and produces a new updated global FL model (Figure 3 (a)).

In the semi-concurrent mode, a kind of strategy is applied that supports free switching between incremental and concurrent FL (in the concurrent mode, all clients simultaneously send their local models to the coordinator for aggregation and production of the next new and probably improved global model). Training can be initiated by a single (any) client, after which an arbitrary number of other clients can join during the training process and participate in construction of the new global model (Figure 3 (b)).

Deep learning models are commonly used in FL because they naturally fit into this type of environment, as weight updates using stochastic gradient descent (SGD) can be easily parallelized. Deep learning is a very popular ML approach, and it is most often based on neural networks with multiple hidden layers. Deep learning sets basic parameters on the data and trains the network to recognize characteristic patterns of behavior in the data, learning on its own using many hidden processing layers.

3.2 Interoperability of Trained Models – Explainable Artificial Intelligence

Explainable Artificial Intelligence (XAI) is a set of processes and methods that enable non-AI users to understand and trust the results and outputs produced by ML algorithms and predictive models. XAI methods are used to describe the model, its expected impact, and potential biases. They help experts in other domains to assess the accuracy of the model and the transparency of the outcomes in making decisions based on ML models. In addition, such approaches can provide insight into how the algorithm achieves its results.

4. *ASCAPE - An example of a use of the considered artificial intelligence techniques*

In this section we will briefly present the results of the application of ML and FL techniques in creating predictive models, which were implemented by the author of this paper and her collaborators within a large and significant European project. In addition, some examples of the use of XAI as added value to the results of predictive models will be considered.

4.1. Medical Datasets, their Processing and Training of Predictive Models

Within the ASCAPE project [7] – Artificial Intelligence in Support of Cancer Patients in the European Region, an innovative and unique open architecture based on AI/ML/FL methods for personalized treatment of breast and prostate cancer patients has been developed. This architecture encompasses all activities related to data management, training and evaluation of AI/ML/FL predictive models, visualization and explainability of the obtained results/outcomes of predictive models (Figure 4). The ASCAPE architecture is organized according to the requirements for the implementation of modern distributed complex systems. It is based on Cloud/Edge principles that support the application of AI/ML/FL techniques and consists of 4 local nodes/healthcare institutions (Edge nodes). Local models are trained in the

[7] `https://www.ascape-project.eu/`

local nodes and the flow of activities that are iteratively repeated in each of them is shown in Figure 4(a).

The centralized coordinator (Cloud) is responsible for training global models (Figure 4(b)) based on the unification and synchronization of the produced local models.

At the very beginning of the project implementation, we were faced with large differences between the available datasets in individual institutions because they contained different features/properties, i.e. predictive and target variables that were necessary for training and later use of the model. As a logical solution to the problem, widespread and accepted standards for representing medical data were used. In this way, the integration of data from different sources was achieved, as well as the further process of their use and processing. The above-mentioned SNOMED CT and HL7 FHIR standards were selected for creating the data model (Figure 1).

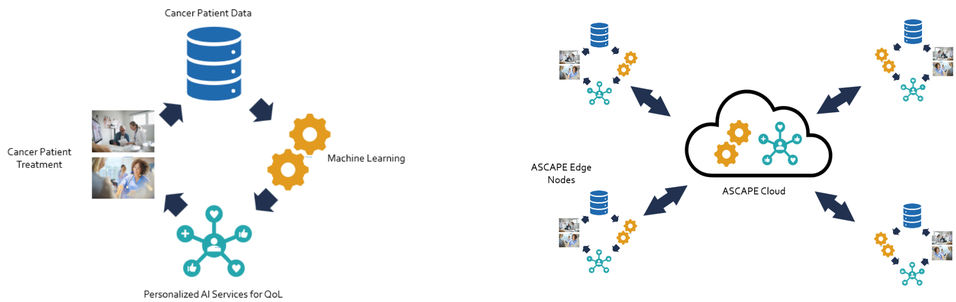

Figure 4: Iterative process of continuous data collection and model training on extended datasets (a) and global Cloud/Edge structure (figures are part of the project's internal material, [7])

The open architecture developed within the ASCAPE project (Figure 5) consists of three subsystems/components with the following functionalities: 1) Data management subsystem; 2) Intelligent data processing subsystem; 3) Intelligent/smart interface for displaying produced results and relations between patient data and results achieved from intelligent data processing subsystems using XAI. Each of these subsystems consists of a number of individual components that support specific functionalities. More details about the architecture and individual components can be found in [10], [11].

ASCAPE architecture is organized according to the requirements for the implementation of modern distributed complex systems. It is based on Cloud/Edge principles that support the application of AI/ML/FL techniques while preserving privacy (differential privacy and homomorphic encryption of data from Edge nodes). AS-

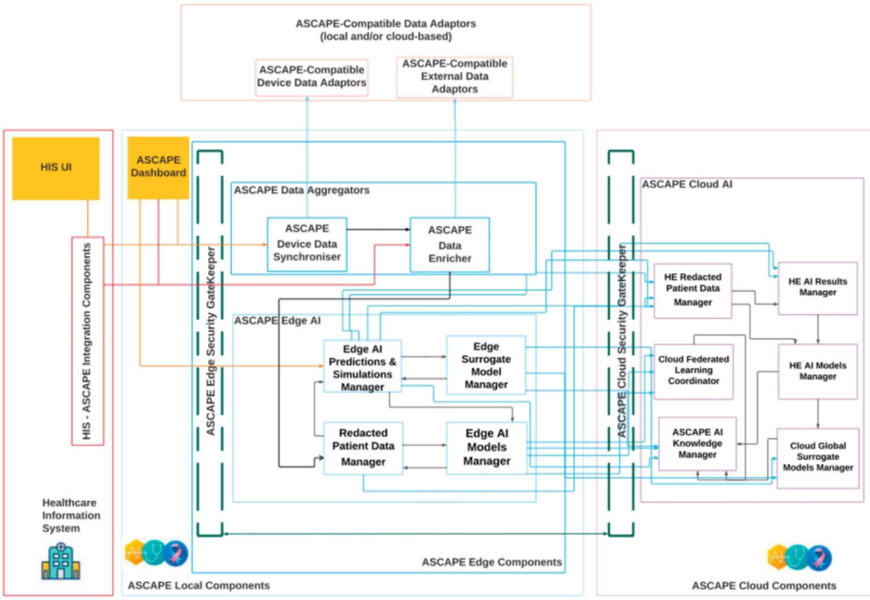

Figure 5: Complex and complete ASCAPE architecture [7]

CAPE Cloud coordinates the training of a global federated model based on deep learning and participation of individual local models from Edge nodes.

For training predictive models, basically two separate datasets were prepared, for patients with breast cancer and patients with prostate cancer. The dataset of patients with breast cancer contained 57 parameters/features (predictive variables) that represent the patient's condition, 15 features of the patients' quality of life and 21 interventions that the predictive model can suggest for the continuation of the patient's treatment (which represent target variables). Similarly, the prostate cancer dataset contained 63 features representing the patient's condition, 12 patient quality of life features, and 23 possible interventions.

The following procedures and approaches were implemented within the ASCAPE architecture. In the analysis of regression/classification predictive models for patients quality of life features, 5 classification algorithms were trained (of which NB showed the best results), 8 regression algorithms (of which Lasso showed the best results). To achieve the best quality of predictive models we reorganized data from two existing datasets into new datasets. From two large datasets of patients with breast cancer and patients with prostate cancer, 10 different new datasets were constructed. For training the FL simulator, neural networks were used for a larger number of Edge nodes, although in this case there were only 4 health institutions.

Additional testing, on more nodes (8, 16 and 32) was performed in order to achieve greater flexibility in the work of the architecture and greater reliability of work in the case of joining new institutions to the existing architecture. Testing was performed using two approaches: incremental and semi-concurrent federated learning.

Given that the datasets contained a significant number of patient records with missing values, it was necessary to test and apply various techniques for missing values imputation. Two approaches were selected: simple imputation and iterative imputation. Regardless of which imputation approach was used, the results achieved and provided by the predictive models showed no significant difference in performance.

An analysis of the data features was also performed in order to determine the most influential among them. Experiments with various combinations and selections of features additionally helped in discovering the most influential features that contribute to higher quality and reliability of models.

To preserve data privacy, the differential privacy method for adding noise to the data was used based on the Laplace distribution. The effect of adding noise for different values of the ϵ parameter was adjusted and its optimal value was determined. For the process of adding Laplace noise to determine the most favorable value, the MAE metric (mean absolute error) was used, and the obtained values were observed and analyzed for the applied regression algorithms. MAE thus helped in determining the best value of the ϵ parameter while preserving a high degree of data privacy.

Regarding the preservation of patient data privacy in a distributed mode, several open-source libraries for Homomorphic Encryption were tested - (PiSift, PiFHEL, TenSEAL, nuFHE, PHE).

Federated learning within the ASCAPE project was based on ML approaches that preserve a high degree of privacy and data security:

Federated Deep Learning – produce and constantly improves global models existing in the ASCAPE Cloud. These models are built on knowledge derived from trained local models at all Edge nodes without transferring local data to the Cloud.

Homomorphic Deep Learning – trains models that exist in the Cloud using homomorphically encrypted data transfer from Edge nodes. When such models are used to infer knowledge from homomorphically encrypted input data, they produce homomorphically encrypted results that can be homomorphically decrypted at the Edge nodes and further used in patient treatment procedures. The advantage of this approach is that the ASCAPE Cloud can be used to train and subsequently use models in real-world environments without access to unencrypted data.

4.2 Role of XAI for better explanation of predictive models outputs

XAI techniques were used in the project (developed by colleagues from Deutsches Forschungszentrum for Knstliche Intelligenz GmbH, DFKI, Bremen) for better interpretation of the predictive models results. For these purposes, the SHAP [8] framework was used because it was identified as the best option.

The formed surrogate models showed excellent approximation of the selected target models. To propose the best treatments for specific patients, a simulation method was adopted. The negative/positive effects of the predictor variables on the possible risk values for each quality of life issue were observed. Personalized effects of the interventions suggested by the predictive models for a specific patient were calculated and analyzed.

An example of the expected change in the risk of pain based on the probability determination using the NB – Bayesian classifier is shown in Figure 6. The recommended treatment for a specific patient is "Adjuvant chemotherapy with CMF" because it is expected to have the greatest impact on reducing the risk of pain and represent the best obtained result.

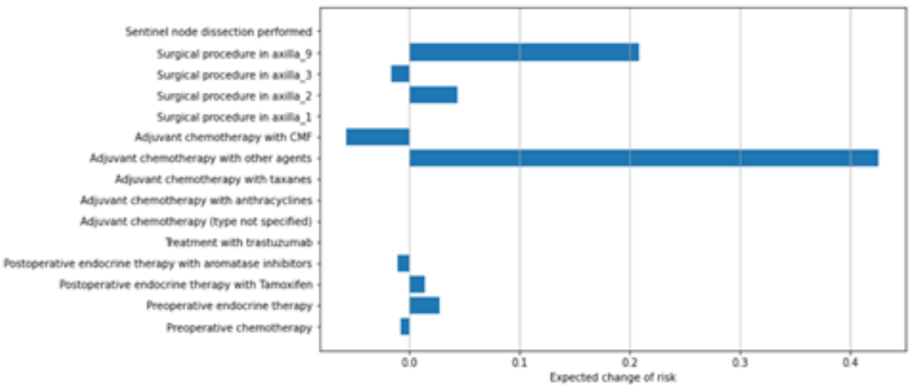

Figure 6: Expected changes in pain risk for a specific patient [22]

Another example is a spider chart (Figure 7) that shows two sets of values. First set of values represent the current values for a patient for certain issues of quality of life. Another set represents their corresponding predicted values for the patient for the next short-term period that ASCAPE considers relevant (3 months). The figure provides personalized visualizations on the ASCAPE dashboard that is displayed to

[8] `https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html`

the physician for a patient with breast cancer. Current values are shown in blue and predicted values in green.

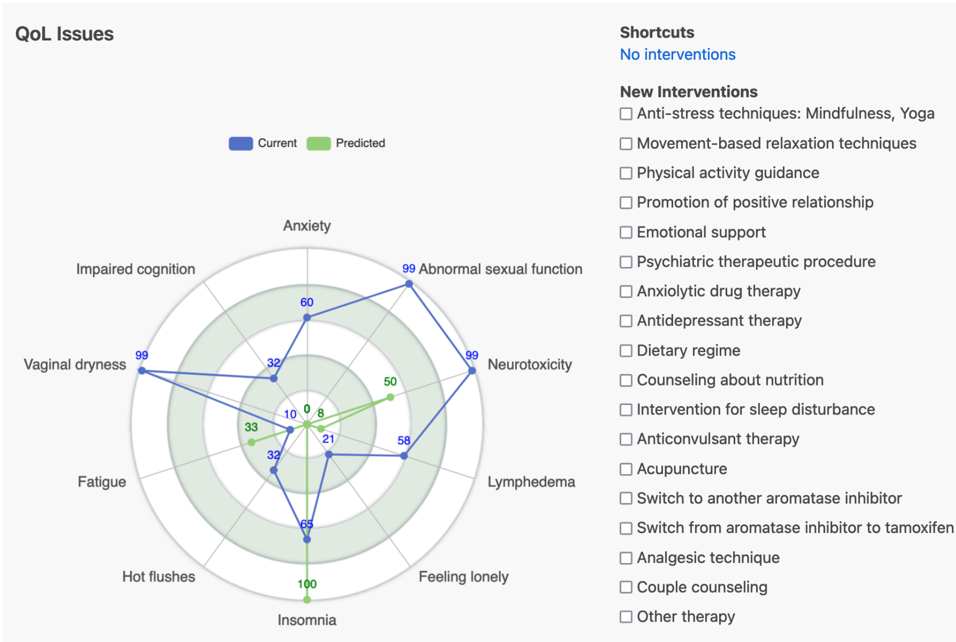

Figure 7: Graph representing quality of life parameters for a breast cancer patient

5. *Conclusion*

The possibilities of applying AI techniques in medical domains are enormous and achievements of higher quality predictive models and various intelligent services are incredible in providing more reliable results. ASCAPE architecture was implemented as a complex and comprehensive prototype. Numerous experiments are conducted and obtain very encouraging results using AI/ML/FL/XAI techniques. Positive ASCAPE results and achievements will contribute to further medical research and practice in very prominent directions [15]: powerful health analytics and predictive modeling, data visualization techniques supported by advanced XAI approaches for communication between different users (patients, doctors, medical actors) [4], personalized therapies, recommendations and interventions.

Recent concepts in ICT such as avatars, metaverse [15], holographic constructions [14], [15], generative AI have great potential and can significantly influence the future development of holistic, sophisticated, intelligent medical systems [17],

[19]. Therefore, further development of systems such as ASCAPE could go in the direction of extending existing architectures with these new approaches, especially in enriching the smart user interface and dashboard.

Acknowledgement. The research related to the development of the ASCAPE architecture was result of teamwork. The author was leader of local team consisting of younger colleagues from the Department of Mathematics and Informatics. The local team has been in charge of producing data models, processing datasets using machine and federated learning techniques. The author is thankful to the colleagues: Duan Jakovetić (who was the main representative of the institution in the preparation of the project and continuous monitoring) as well as to colleagues who participated in the implementation of appropriate system components: Vladimir Kurbalija, Milo Savić, Brankica Bratić, Mihailo Ilić, Marko Otlokan. In addition the author would especially highlight the cooperation with the research group led by Prof. Dr. Serge Autexier from DFKI, Bremen, Germany, as well as Dr. Antonios Valachis from Örebro Hospital in Sweden.

REFERENCES

[1] S. Autexier, C. Lüth, R. Drechsler, *Chapter the Bremen ambient assisted living Lab and beyond smart environments, smart services and artificial intelligence in medicine for humans*, In: Pfannstiel MA, editor. Springer Fachmedien Verlag Wiesbaden, 2022.

[2] C. Drummond, *Classification, Encyclopedia of Machine Learning and Data Mining*, Springer, 2017.

[3] J. Ficek, W. Wang, H. Chen, G. Dagne, E. Daley, *Differential privacy in health research: A scoping review*, Journal of the American Medical Informatics Association, 28(10), pp. 2269–2276, 2021.

[4] A. Holzinger, A. Saranti, C. Molnar, P. Biecek, & W. Samek, *Explainable ai methods-a brief overview*, In: *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*, pp. 13–38, Springer, 2020.

[5] M. Ilic, M. Ivanovic, V. Kurbalija, A. Valachis, *Towards optimal learning: Investigating the impact of different model updating strategies in federated learning*, Expert Syst. Appl. 249, 2024.

[6] M. Ilić, M. Ivanović, D. Jakovetić, V. Kurbalija, M. Otlokan, M. Savić, M. Vujnović-Sedlar, *ASCAPE – An intelligent approach to support cancer patients*, Proceedings of WorldCIST 2023 – 11st World Conference on Information Systems and Technologies, Pisa, Italy, 4–6 April, 2023.

[7] M. Ilić, M. Ivanović, D. Jakovetić, V. Kurbalija, M. Otlokan, M. Savić, and N. Vujnović-Sedlar, *The role of federated learning in processing cancer patients data*, In: Device-Edge-Cloud Continuum: Paradigms, Architectures and Applications, pp. 49–68, Springer, 2023.

[8] M. Ivanovic, *How Artificial Intelligent Approaches Support Medical Decisions and Patients' Wellbeing*, Novel & Intelligent Digital Systems: Proceedings of the 2nd International Conference, NiDS 2022, Athens, Greece, 29–30 September 2022, pp. 181–189, Springe, 2022.

[9] M. Ivanovic, *Influence of Federated Learning on Contemporary Research and Applications*, International Conference on INnovations in Intelligent SysTems and Applications, INISTA 2024, Craiova, Romania, September 4–6, pp. 1–6, 2024.

[10] M. Ivanovic, S. Autexier, M. Kokkonidis, *Rust: Quality medical data management within an open AI architecture – cancer patients case*, Connect. Sci. **35**(1), 2023.

[11] M. Ivanovic, S, Autexier, M. Kokkonidis, *AI approaches in processing and using data in personalized medicine* In: Chiusano S, Cerquitelli T, Wrembel R. editors, Advances in Databases and Information Systems. ADBIS 2022. Lecture Notes in Computer Science: Vol. 13389, Springer, pp. 11–24, 2022.

[12] M. Ivanovic, *Applications of Artificial Intelligence Methods in Medicine*, International Conference on Information Technology, ICIT 2023, Amman, Jordan, August 9–10, pp. 349–354, 2023.

[13] M. Ivanović, S. Ninković, *Personalized HealthCare and Agent Technologies*, 1th KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications, Vilamoura, Portugal, 21–23 June, Springer, pp. 3–11, 2017.

[14] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, et al., *Advances and open problems in federated learning*, Foundations and Trends in Machine Learning, **14**, no. 1-2, pp. 1–210, 2021.

[15] L. H. Lee, T. Braud, P. Zhou, L. Wang, D. Xu, Z. Lin, A. Kumar, C. Bermejo, and P. Hui, *All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda*, `arXiv preprint arXiv:2110.05352`, 2021.

[16] M. Liu, S. Li, H. Yuan, M. E. H. Ong, Y. Ning, F. Xie, et al., *Handling missing values in healthcare data: A systematic review of deep learning-based imputation techniques*, Artificial Intelligence in Medicine, vol. 142, Art. no. 102587, 2023.

[17] J. Lloret, A. Canovas, S. Sendra, and L. Parra, *A smart communication architecture for ambient assisted living*, IEEE Communications Magazine, **53**, no. 1, pp. 26–33, 2015.

[18] P. Riedel, L. Schick, R. von Schwerin, M. Reichert, D. Schaudt, and A. Hafner, *Comparative analysis of open-source federated learning frameworks: A literature-based survey and review*, International Journal of Machine Learning and Cybernetics, **15**, no. 11, pp. 5257–5278, 2024.

[19] A. Salih and A. Abraham, *Ambient Intelligence Assisted Healthcare Monitoring*, LAP Lambert Academic Publishing, 2016.

[20] M. Siddique, M. A. Mirza, M. Ahmad, J. Chaudhry, and R. Islam, *A survey of big data security solutions in healthcare*, In: Proc. Int. Conf. Security and Privacy in Communication Systems, Cham, Switzerland: Springer, pp. 391–406, 2018.

[21] R. Thirunavukarasu, R. Gnanasambandan, M. Gopikrishnan, and V. Palanisamy, *Towards computational solutions for precision medicine based big data healthcare system using deep learning models: A review*, Computers in Biology and Medicine, **149**, Art. no. 106020, 2022.

[22] *Artificial Intelligence Supporting Cancer Patients Across Europe (ASCAPE): ML/DL Training and Evaluation Report, Deliverable D2.4*, European Union Horizon 2020 Research and Innovation Programme, Grant Agreement no. 875351, 2023. [Online]. Available: `https://ascape-project.eu/sites/ascape/files/public/content-files/deliverables/ASCAPE%20D2.4%20ML_DL%20Training%20and%20Evaluation\%20Report-V1.0_0.pdf`

[23] T. Namlı and A. A. Sınacı, *Health Data Preparation for Artificial Intelligence*, *AICCELERATE*, Feb. 22, 2024. [Online]. Available: `https://aiccelerate.eu/health-data-preparation-for-artificial-intelligence-ai/`.

Department of Mathematics and Informatics
Faculty of Sciences
University of Novi Sad, Serbia
e-mail: mira@dmi.uns.ac.rs