Process Mining Organization (PMO) Based on Machine Learning Decision Making for Prevention of Chronic Diseases

Angelo Rosa [1,2,*,†] and Alessandro Massaro [1,2,3,†]

1 Department of Management, Finance and Technology, LUM, Libera Università Mediterranea "Giuseppe Degennaro", S.S. 100-Km.18, Parco il Baricentro, 70010 Bari, Italy; massaro@lum.it
2 Department of Engeneering, LUM, Libera Università Mediterranea "Giuseppe Degennaro", S.S. 100-Km.18, Parco il Baricentro, 70010 Bari, Italy
3 LUM Enterprise S.r.l., S.S. 100-Km.18, Parco il Baricentro, 70010 Bari, Italy
* Correspondence: rosa@lum.it
† These authors contributed equally to this work.

Abstract: This paper discusses a methodology to improve the prevention processes of chronic diseases such as diabetes and strokes. The research motivation is to find a new methodological approach to design advanced Diagnostic and Therapeutic Care Pathways (PDTAs) based on the prediction of chronic disease using telemedicine technologies and machine learning (ML) data processing techniques. The aim is to decrease health risk and avoid hospitalizations through prevention. The proposed method defines a Process Mining Organization (PMO) model, managing risks using a PDTA structured to prevent chronic risk. Specifically, the data analysis is focused on stroke risk. First, we applied and compared the Random Forest (RF) and Gradient Boosted Trees (GBT) supervised algorithms to predict stroke risk, and then, the Fuzzy c-Means unsupervised algorithm to cluster information on the predicted results. The application of the proposed approach is able to increase the efficiency of healthcare human resources and drastically decrease care costs.

Keywords: prevention of chronic disease; process mining; Process Mining Organization (PMO); machine learning; decision making; telemedicine

1. Introduction

Process mining (PM) is an important approach suitable for designing processes based on machine learning (ML) decision-making engines. PM has been applied to improve industrial processes [1,2] and subsequently to design healthcare processes [3,4] regarding the cost optimization of healthcare services [5], telemedicine [6], and patient fall risk management [7]. The application of PM is important for the design of organizational models based on workflows integrating ML algorithms and supporting decisions about human resource (HR) allocation or engagement [6]. The ML-HR decision-making engine upgrades the PM model to a Process Mining Organization (PMO) model. A method suitable for representing and sketching processes is the Business Process Modeling and Notation (BPMN) approach [8]. BPMN is an international standard [9,10], providing graphical elements to map processes. The BPMN is useful to design healthcare processes such as Diagnostic and Therapeutic Care Pathways (PDTAs). An example of a PDTA mapped by BPMN is illustrated in Figure 1, representing the 'AS IS' care path of diabetics [11]. As observed in Figure 1, the Italian diabetic PDTA is exhaustive for the chronic pathology, but no details are provided for primary prevention highlighted by the green box (prevention task). The goal of the proposed paper is therefore precisely the optimization of the prevention task using telemedicine and ML facilities. Specifically, we have proposed a basic organizational model and technological facilities that can be used to implement a prevention PDTA with the goal of eliminating the risk of chronic degeneration, and, consequently, avoiding the

execution of the whole PDTA process of chronic care, which requires high resource costs. Figure 2 presents the sketch of a diagram summarizing the goal of the paper.
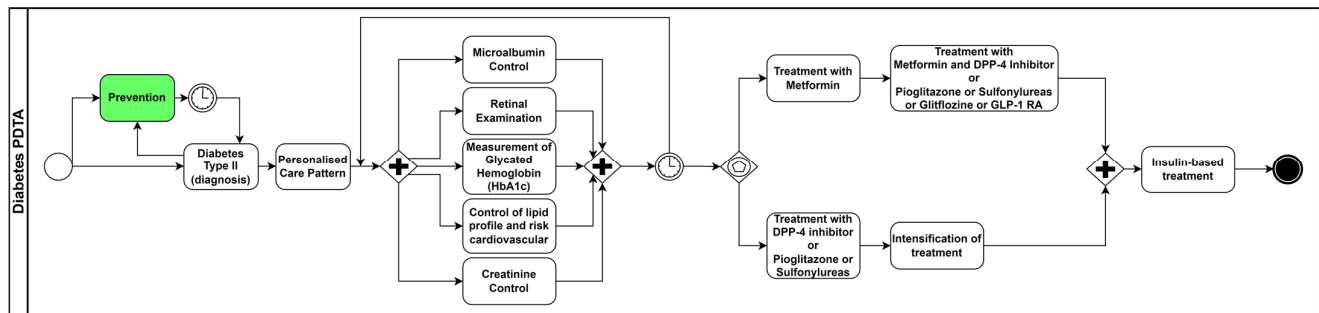

Figure 1. 'AS IS' PDTA process of type 2 diabetes (translation of the procedure indicated in [11]). The green box highlights the prevention task goal of the proposed work. No stroke PDTA is available in Italy.
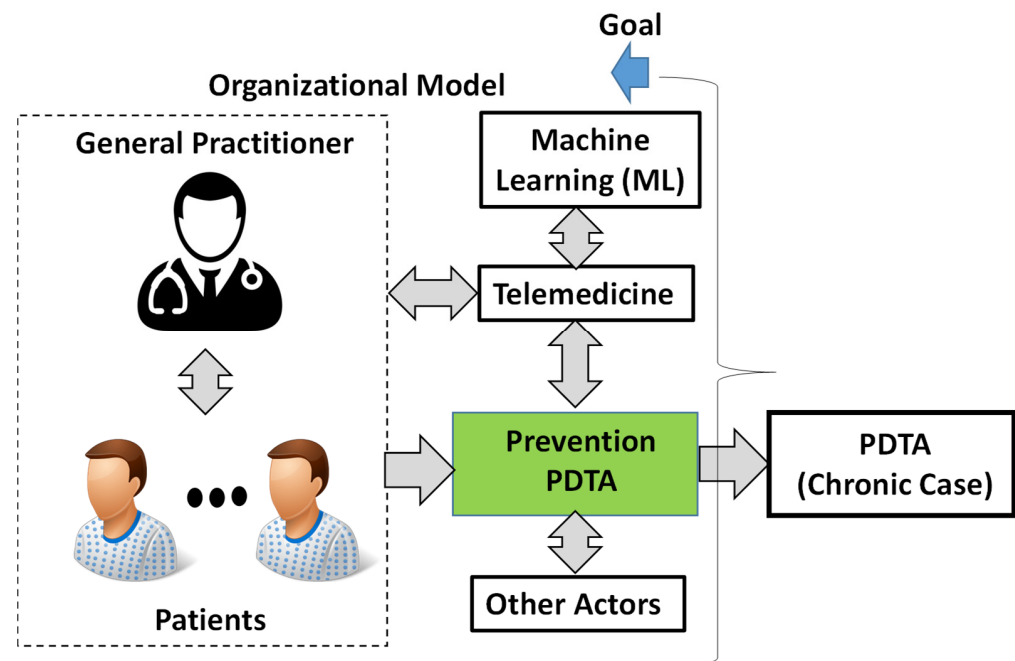

Figure 2. Basic organizational model and technological facilities for implementing a prevention PDTA with the goal of avoiding chronic risk.

In Italy, high care costs are estimated for chronic diabetic patients; by analyzing the socio-economic impact of diabetes, it is noted that in only the Italian Puglia region, approximately 5% of the adult population aged 18–69 years is affected by diabetes [12]; furthermore, the average annual cost per diabetic patient is EUR 2792 [13], and for the Puglia region alone, there are an estimated 232,000 diabetic people (Source: Istat 2020), corresponding to a total annual cost for the region of EUR 647,744,000. Concerning the stroke cost, in Italy, there are 100 thousand new cases a year, corresponding to an estimated cost of EUR 16 billion for the whole National Healthcare Service (source: 'Sanità 24', 2018). These initial analyses highlight the importance of finding a solution capable of reducing, from a predictive perspective, the onset of chronic diseases in various possible forms.

The prevention of diabetes could also have implications in the prediction of the risk of stroke. In particular, cases of hypoglycemia or hyperglycemia represent stroke risks [14,15]. Other elements concerning risks which are correlated with strokes are hypertension [16] and heart disease [17]. These risk elements are analyzed in the paper.

Regarding prevention, telemedicine could improve healthcare organization by optimizing care processes [6]. For this reason, the goal of the proposed work is to design a prevention process matching diabetes and stroke risks by highlighting ML data processing aspects in decision-making procedures and in organizational aspects.

The purpose of the study is to provide ML tools and process mapping methods to actuate the new PDTA to prevent chronic diseases. The PDTA prevention process is modeled in this paper through BPMN, thus suggesting an innovative workflow to follow the patient monitoring care patterns through telemedicine tools. Furthermore, telemedicine provides digital data useful for executing the preventive decision-making procedures.

The paper is structured as follows:

- In Section 2, we provide information about the materials and methods, discussing the analyzed dataset and the ML data processing workflow.
- In Section 3, we propose a PDTA BPMN workflow management prevention process and care service organization, including telemedicine monitoring tools to decrease the risks of diabetes and stroke.
- In Section 3, we apply supervised and unsupervised ML algorithms, improving the initial decision making about possible risks and focusing the data analysis on stroke risk.
- In Section 4 and in the appendices, we provide information about organizational aspects supported by telemedicine facilities and improved by other possible healthcare actors.
- In Section 4 and in the appendices, we also discuss advantages, disadvantages, limitations, and perspectives of the proposed approaches and technologies.

2. Materials and Methods

In this section, we discuss the ML tools and the dataset processed by the ML algorithms.

2.1. Dataset Testing Machine Learning Algorithms to Predict Stroke Risk

The open dataset [18] is used for the testing of the ML algorithms, focusing attention on stroke prediction. The dataset is available in the *Kaggle* dataset repository [18] as a .csv file containing 5110 observations with the following 12 attributes:

- *id*: unique identifier of the patient;
- *gender*: 'Male' or 'Female';
- *age*: age of the patient;
- *hypertension*: 0 for patients without hypertension, 1 for patients with hypertension;
- *heart_disease*: 0 for patients without any heart diseases, 1 for patients affected by heart disease;
- *ever_married*: 'No' (never married) or 'Yes' (married);
- *work_type*: 'children' (not a worker), 'Govt_jov' (public worker), 'Never_worked' (unemployed), 'Private' (worker of a private company), or 'Self-employed' (professionals or managers);
- *Residence_type*: 'Rural' or 'Urban';
- *avg_glucose_level*: average glucose level measured in blood;
- *bmi:* body mass index;
- *smoking_status:* 'formerly smoked' (smoker in the past), 'never smoked' (not a smoker), 'smokes' (smoker), or 'Unknown' (no information is available);
- *stroke:* 1 if the patient had a stroke, or 0 if not.

In Figure 3, a screenshot is shown of the dataset imported in the local repository (local memory) of the personal computer (11th Gen Intel(R) Core(TM) i5-1135G7, 2.42 GHz) used for the data processing. The figure shows all the attributes listed above as digital records locally imported.

Row ID	S gender	D age	I hypertension	I heart_disease	S ever_married	S work_type	S Residence_type	D avg_glucose_level	D bmi	S smoking_status	S stroke
Row90	Female	79	0	1	Yes	Private	Urban	226.98	29.8	never smoked	1
Row91	Male	81	0	0	Yes	Private	Urban	72.81	26.3	never smoked	1
Row92	Female	57	0	0	No	Self-employed	Urban	68.02	37.5	never smoked	1
Row93	Female	80	1	0	Yes	Self-employed	Urban	68.56	26.2	Unknown	1
Row94	Male	45	0	0	Yes	Private	Rural	64.14	29.4	never smoked	1
Row95	Female	78	0	0	Yes	Private	Rural	235.63	32.3	never smoked	1
Row96	Female	70	0	0	Yes	Self-employed	Rural	76.34	24.4	formerly smoked	1
Row97	Male	58	0	1	Yes	Private	Rural	240.59	31.4	smokes	1
Row98	Male	57	1	0	Yes	Govt_job	Urban	78.92	27.7	formerly smoked	1
Row99	Female	69	0	0	Yes	Govt_job	Urban	82.81	28	never smoked	1
Row100	Male	64	0	1	Yes	Govt_job	Urban	74.1	28.8	Unknown	1
Row101	Female	77	0	0	Yes	Govt_job	Rural	190.32	31.4	never smoked	1
Row102	Female	74	0	0	Yes	Private	Rural	231.61	34.6	formerly smoked	1
Row103	Female	81	0	1	Yes	Self-employed	Rural	78.7	19.4	Unknown	1
Row104	Female	57	0	0	Yes	Govt_job	Urban	110.52	28.5	Unknown	1

Figure 3. Dataset extraction (data imported in the local repository), indicating the attributes to be processed by the ML algorithms.

2.2. Machine Learning Algorithms

The ML data processing model is structured in the following three data processing steps:

(1) *Stage I*: data pre-processing, preparing the input dataset;
(2) *Stage II:* supervised ML algorithm data processing, predicting stroke risks;
(3) *Stage III:* unsupervised ML algorithm data processing, supporting data interpretation and clustering of the results.

Random Forest (RF) and Gradient Boosted Trees (GBT) are applied as ML supervised algorithms. The choice of these algorithms is due to their ability to efficiently process heterogeneous attributes [19–22], including numerical and categorical ones. Specifically, RF is an ensemble ML algorithm consisting of the use of a chosen number of decision trees by combining the outputs of the decision trees into a single result. RF has many further advantages, such as the ability to also process qualitative data, high performance with a good classification accuracy, good robustness regarding numerical entropy introduced by missing values or wrong information, and good ability to analyze complex attribute interdependencies.

On the other hand, the GBT algorithm adopts very shallow regression trees and a special form of boosting to build an ensemble of trees [23]. The used base learner for this ensemble method is a simple regression tree, as for the RF algorithm. For each iteration step, the parameters are adjusted to minimize the loss function, indicating the difference between the classified/predicted and actual values. The gradient represents the incremental parameter adjustment and the boost is the method to accelerate the improvement of the accuracy.

The supervised RF and GBT algorithms are applied in this work to classify the patients to be controlled. These patients are initially characterized with no stroke risk; some patients initially with no stroke risk are classified by the RF and GBT algorithms as patients with a possible stroke risk (patients to be monitored).

The output of the ML supervised algorithms is successively clustered to facilitate data reading and data interpretation by simultaneously analyzing more clustered attributes. The adopted ML unsupervised algorithm is the Fuzzy c-Means [24]; fuzzy clustering allows each data point to fit in different clusters, defining a degree of membership to each cluster.

The tool adopted for ML data processing is the Konstanz Information Miner (KNIME). KNIME is an open-source tool [25] providing a large set of ML algorithms. Its versatility is in the use of interconnected graphical block (or nodes) behaving as interfaces, allowing data processing parameters to be changed. All the blocks are linked to structure the workflow. In Figure 4, the KNIME workflow executed in this work is illustrated. The workflow is structured into the three stages:

- *Stage I:* Containing blocks suitable for data pre-processing operations, such as importing the .csv input dataset into the local repository environment ('CSV Reader'), attribute conversion ('String to Number'), attribute section filtering more significant

attributes ('Column Filter'), and data partition (partition by the 'Partitioning' block of the dataset into a training and a testing dataset).

- *Stage II:* Implementing blocks to execute the RF and GBT supervised algorithms, including learning models processing the training dataset ('Random Forest Learner' and 'Gradient Boosted Trees Learner'), prediction models processing the testing dataset ('Random Forest Predictor' and 'Gradient Boosted Trees Predictor'), visual representation of results ('Scatter Plot', 'Color Manager' and 'Statistic'), and algorithm performance score ('Scorer' and 'Numeric Scorer').
- *Stage III:* Implementing blocks to run the unsupervised Fuzzy c-Means algorithm ('Fuzzy c-Means') and visual representation blocks.
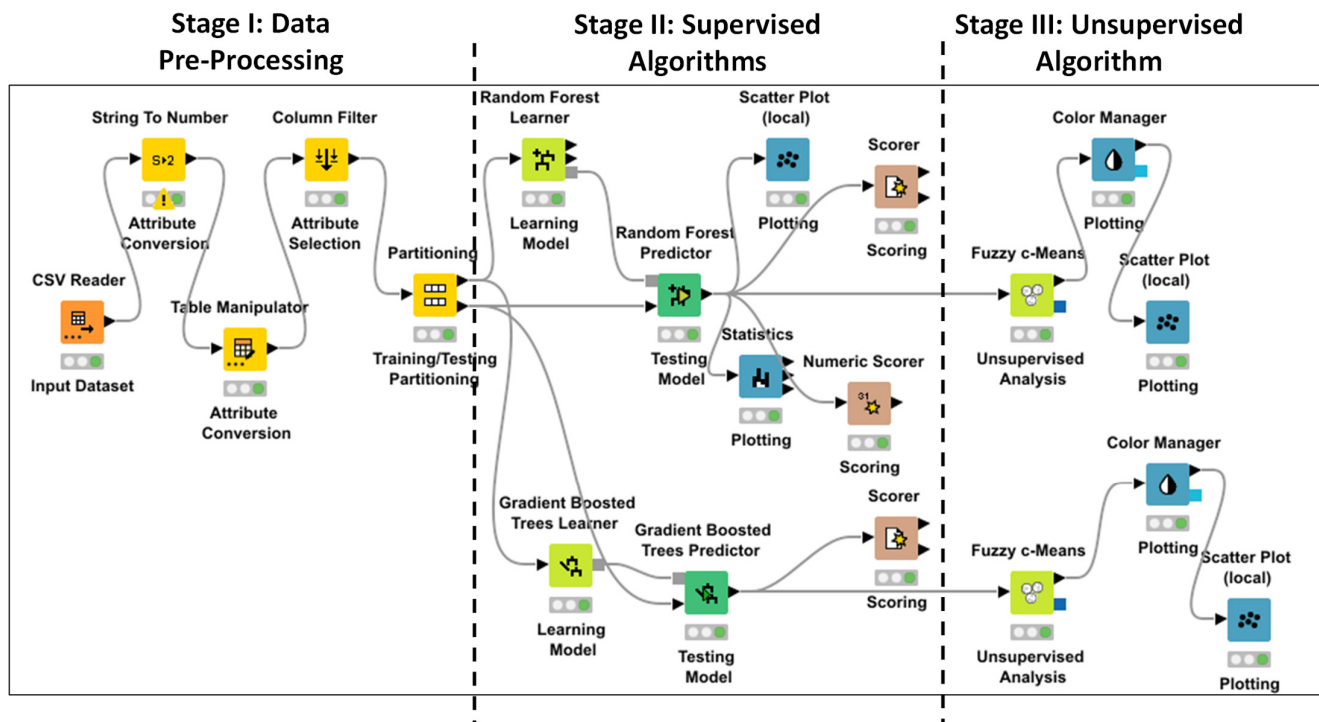

Figure 4. KNIME workflow implementing ML supervised and unsupervised algorithms, following from left to right the three stages of data processing (Stage I, Stage II, Stage III).

For both the FR and GBT algorithms, the original dataset is partitioned into training (2146 records containing stroke conditions important for training the ML models) and testing datasets (last 2964 records). The 'stroke' attribute (see Section 2.1) is chosen as a class to predict. The hyper parameters of the used GBT algorithm are as follows: limit the number of levels, indicating that the tree depth equals 4, learning rate equals 0, 1, and 100 is the number of models (number of decision trees to learn). Concerning RF, the used split criterion is the *Information Gain Ratio*, which is able to normalize the standard information gain by split entropy to overcome any unfair preference for nominal splits with many child nodes.

3. Results

The first result is obtained by discussing the ML application with medical staff (general practitioners and specialists validating the methodology) and by evaluating the feasibility of deploying a telemedicine platform by adopting certified medical kits (standard CE: 93/42/CEE, 2017/745/UE, class 2a) for diabetes and stroke prevention.

The BPMN workflow of Figure 5 is the result of the validated design, describing the platform monitoring patients during the prevention phase. The workflow examines the combined risk of patients to be affected by diabetes or to be injured by a stroke, thus defining a PDTA of prevention. The same workflow is also suitable for the prevention of

the hypertension risk by adopting specific sensors, mainly measuring heart disease. The diagram in Figure 5 is structured into three pools, indicating the processes of the three main actors involved in the system:

- patient to be monitored for diabetes risk (pool named 'Diabetes Prevention Process');
- patient to be monitored for stroke risk (pool named 'Stroke Prevention Process');
- general practitioner deciding the medical kit to assign and analyze data to decide on possible exams or drug assignments after the detection of digital alerting conditions (alerting thresholds overcoming critical values of physiological parameters, or alerting predicted results).

The model is designed by considering real-time monitoring of the patient's physiological parameters, and an automatic alerting condition enabling the decision of the general practitioner.
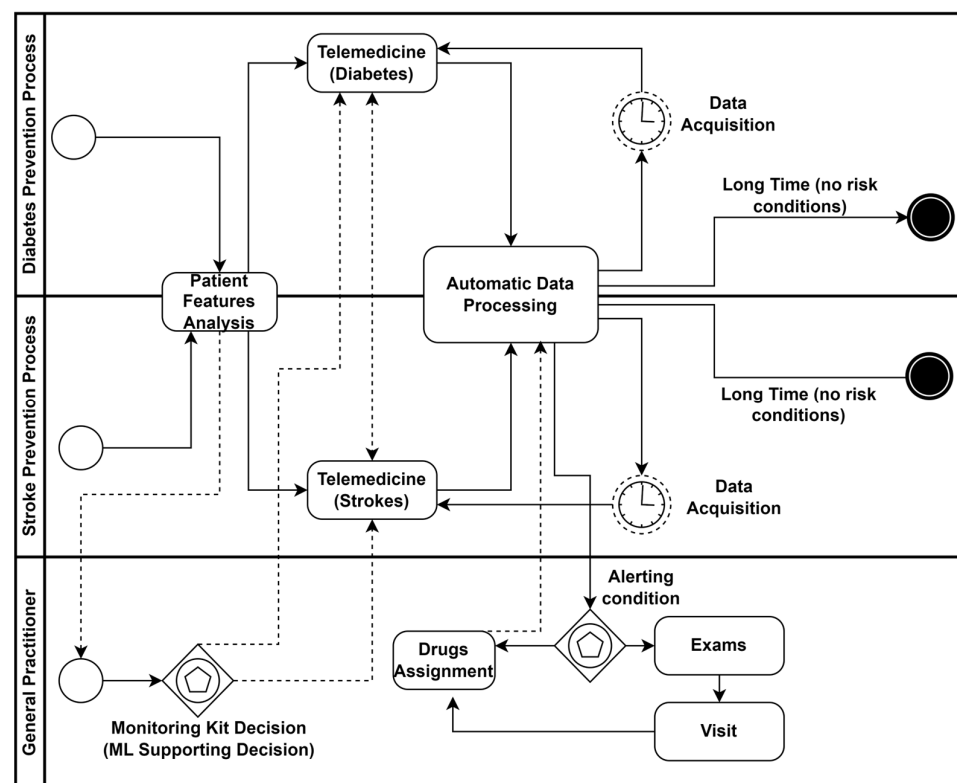

Figure 5. BPMN PDTA prevention process matching stroke and diabetes risk. The automatic data processing activates the general practitioner decision making.

Data Result Interpretation and Decision-Making Processs Preventing Stroke Risk

The dataset [18] is processed by the two algorithms RF and GBT. In Figure 6a,b, two screenshots are shown, indicating the stroke prediction of the same records by executing both the ML RF and GBT algorithms; the algorithms provide the same alerting condition with a weakly different confidence (0.7 in the case of RF and 0.776 for GBT). The algorithms provide similar prediction results and good performances (see performance parameters of Table 1). In Figure 7 can be seen the Receiver Operating Characteristic Curves (ROCs) of both the approaches, providing the values of the Area Under the ROC Curve (AUC) of Table 1, furthermore confirming the high algorithm performance. A further estimated performance index indicated in Table 1 is the F-measure (or F-score), measuring the predictive performance. The F-measure is typically adopted for statistical analysis of binary classification and information retrieval systems.

(a)

Row ID	S gender	D age	I hypertension	I heart_disease	S ever_married	S work_type	S Residence_type	D avg_glucose_level	D bmi	S smoking_sta...	S stroke	S Prediction (stroke)	D Prediction (stroke) (Confidence)
Row2542	Male	14	0	0	No	children	Urban	101.87	20.3	never smoked	0	0	1
Row2543	Female	13	0	0	No	children	Rural	77.55	21.3	Unknown	0	0	1
Row2544	Male	71	0	1	Yes	Private	Urban	215.72	39.2	smokes	0	1 → Alerting	0.7
Row2545	Female	66	0	0	Yes	Private	Rural	87.84	52.8	Unknown	0	0	0.88
Row2546	Male	32	0	0	No	Govt_job	Rural	91.93	30.2	never smoked	0	0	0.98

(b)

Row ID	S gender	D age	I hypertension	I heart_disease	S ever_married	S work_type	S Residence_type	D avg_glucose_level	D bmi	S smoking_status	S stroke	S Prediction (stroke)	D Prediction (stroke) (Confidence)
Row2542	Male	14	0	0	No	children	Urban	101.87	20.3	never smoked	0	0	0.995
Row2543	Female	13	0	0	No	children	Rural	77.55	21.3	Unknown	0	0	0.995
Row2544	Male	71	0	1	Yes	Private	Urban	215.72	39.2	smokes	0	1 → Alerting	0.958
Row2545	Female	66	0	0	Yes	Private	Rural	87.84	52.8	Unknown	0	0	0.958
Row2546	Male	32	0	0	No	Govt_job	Rural	91.93	30.2	never smoked	0	0	0.993

Figure 6. Prediction of the alerting case: (**a**) samples processed by RF algorithm; (**b**) same samples processed by GBT algorithm.

Table 1. Estimated accuracy, AUC, and F-measure of the adopted ML algorithms. All of the indexes confirm the good performance of the adopted RF and GBT algorithms.

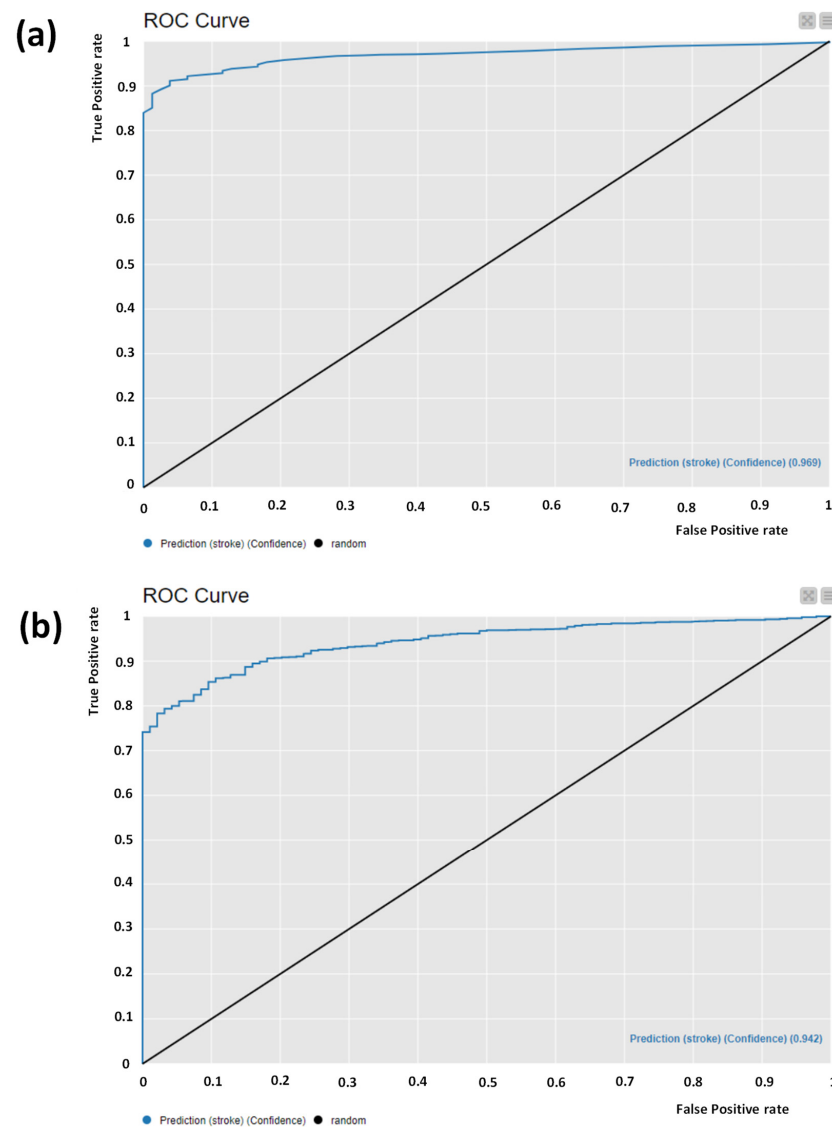| ML Algorithm	Accuracy	AUC	F-Measure
Random Forest (RF)	0.974	0.969	0.987
Gradient Boosted Trees (GBT)	0.968	0.942	0.984

(a)

(b)

Figure 7. (**a**) RF ROC curve. (**b**) GBT ROC curve (implementation of the KNIME 'ROC Curve' node).

We observe that the results to pay attention to are those where the risk of stroke is predicted even though there has not been an alert condition in the past. In order to group the results, the RF and GBT outputs are clustered by the Fuzzy c-Means algorithm, highlighting in red the predicted risk cases of stroke (the red color indicates the cluster of the predicted stroke risk). By considering three clusters, it is observed that the predicted stroke cases do not appertain to the first cluster characterized by patients having an age lower than 40 years old (this allows for excluding the preventive monitoring of these patients). In Figure 8, a comparison is presented of RF and GBT risk cases due to hyperglycemia and hypoglycemia status, by confirming the results expected in the literature [14,15]. The results of both algorithms present few differences.

Figure 8. RF (**a**,**b**) GBT risk cases matching the variables of average glucose level and patient's age. An example of result variation is indicated by the black dashed circle.

In Figure 9a,b, the RF and the GBT stroke risks are shown, matching patient age and patient work type, respectively. Also, in this case, both algorithms provide similar results by highlighting that private companies employees or managers are characterized by a high stroke risk. This result enhances the impact of the work about the stroke risk and could enable the formulation of new welfare policies in private work environments; according to the results, possible interventions could be applied regarding the optimization of working conditions, improving the quality of life, and consequently decreasing the health risk.
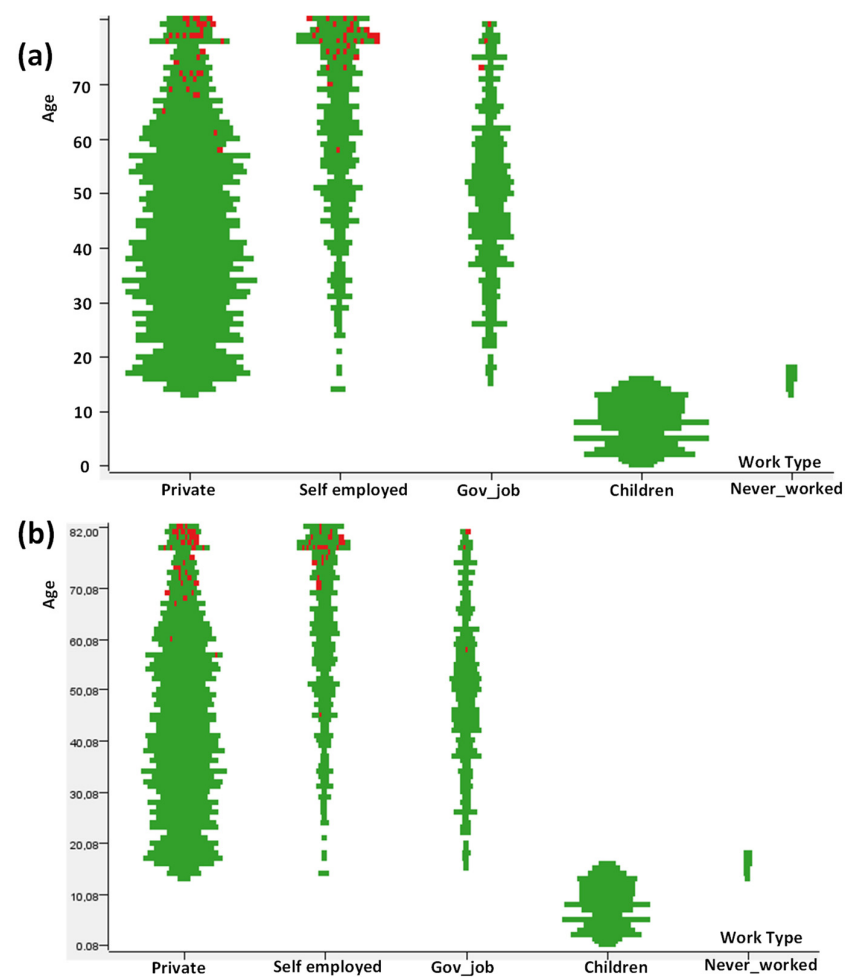
(a)

(b)

Figure 9. RF (**a**,**b**) GBT risk cases matching the variables of patient age and work type.

As expected in the literature, stroke risk happens when heart disease and hypertension cases are checked [16,17]. The results of Figures 10 and 11 confirm these scientific expectations.
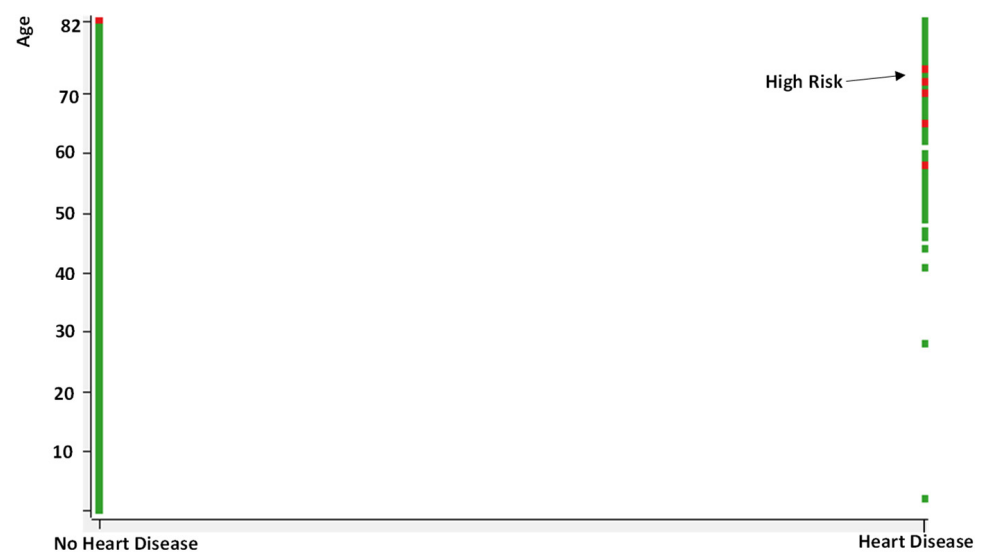

Figure 10. RF stroke risk matching variables of patient age and past verified heart disease conditions.
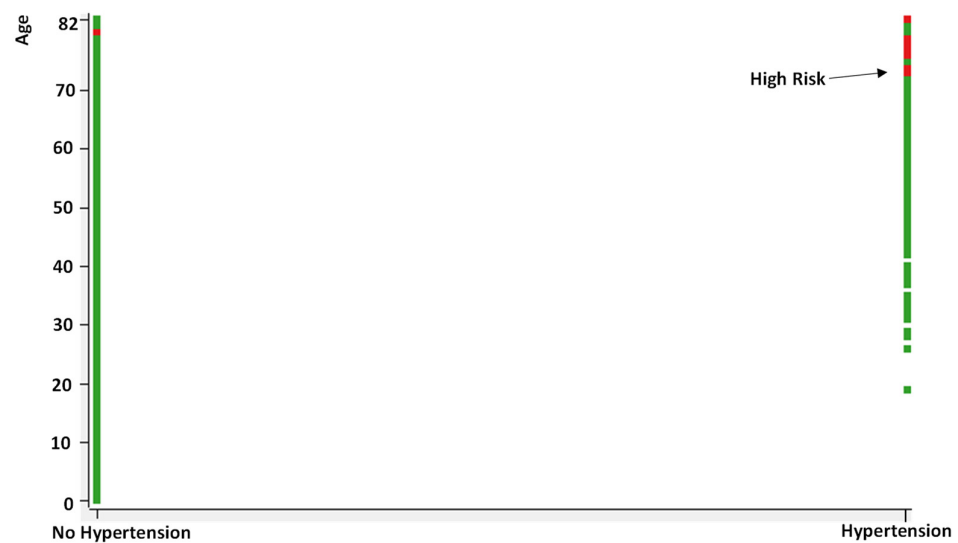
Figure 11. RF stroke risk matching variables of patient age and past verified hypertension conditions.

In addition, social conditions could also have an impact on the risk. Figure 12 demonstrates that no married patients are characterized by a possible stroke risk; this can be explained by the fact that that unmarried workers overwork themselves (work hard) and are therefore subject to a greater risk of stroke.
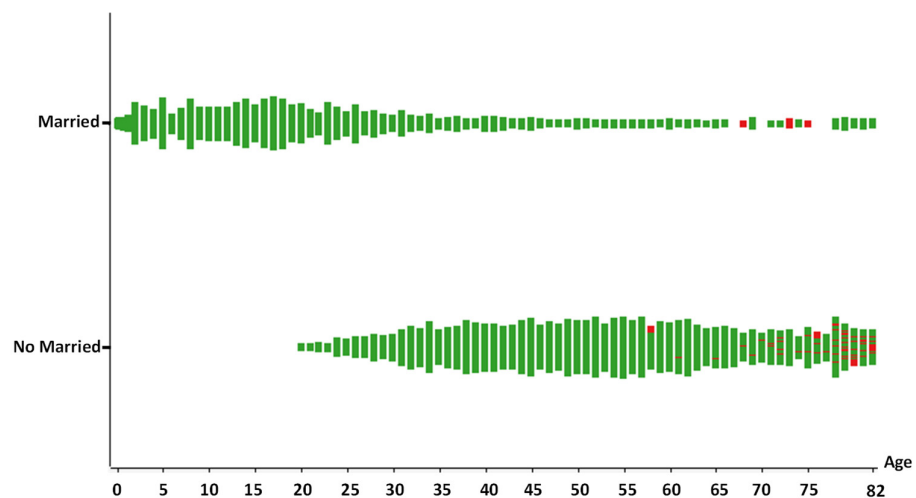

Figure 12. RF stroke risk, matching variables of marriage status and patient age.

4. Discussion: The Telemedicine Framework

The BPMN workflow of Figure 5 is a basic process involving only the patient and the general practitioner actors. More actors could be involved by further optimizing the whole prevention process. The ML results discussed in Section 3 are to be used for the decision of patients to enroll, for the assignment of the medical kit, and for the decisions of the general practitioner upon reading the alerting conditions (corresponding to the 'Exclusive Event-Based' symbols of Figure 5). For example, concerning the diabetes risk, an important parameter to predict with ML is glycemic values. In Table 2, a list of further actors is presented which could collaborate to prevent chronic cases or dangerous health status conditions. In Appendix A, we detail the whole ecosystem involving actors, companies, and research units, improving the telemedicine system and describing the interconnections (action fluxes) between the actions required to perform a preventive PDTA. The realization of a telemedicine platform suitable for chronic prevention requires investment in technology having a high Key Performance Indicator (KPI) of technology readiness (see Appendix B).

The technology readiness KPI indicates the company's capabilities in terms of technology development and organization management.

Table 2. Possible further actors for improving prevention actions and related organizational aspects.

Possible Further Actor to Involve in the Prevention Process	Role	New Organizational Aspects
Pharmacist	Support for patient enrollment in monitoring activities	Coordination with the general practitioner to select the patient to control (the monitoring kit is available in pharmacy)
Dietician	Especially for diabetic risk, a dietician could be important to prevent chronic risk (as for obesity cases [26])	Dietician is involved when it is difficult to control diet and a nutritional balance. Useful solutions for children could be enabled through mobile apps and gaming [27] (organizational aspects are addressed mainly on the management of the mobile app data)
Psychologist	If eating disorders are serious, the help of a psychologist may be required	The psychologist should collaborate with the dietician. The psychological care process should be synchronized with the prevention process, including dietician interactions
Cardiologist	Mainly for stroke cases, the reading of alerting conditions could require the intervention of a cardiologist to provide a second opinion	The prevention process should be synchronized with the cardiologist monitoring process
Private Nurse	Private nurses are required for non-self-sufficient patients	The general practitioner could be relieved of the workload through the work of the nurse
Educational Trainer	Educational trainers to facilitate technology transfer	Organizational aspects are mainly in the planning of the training courses about the use of digital technologies
Public Teachers (schools and universities)	Educational interventions about correct diet and lifestyles in schools and universities could have important effects on risk prevention	Organizational aspects are mainly in the planning of the training courses about lifestyles and nutrition
Politicians (welfare policies)	Important interventions in the social environment require political laws supporting a sustainable prevention platform	Politicians should have a list of available actors or health structures to hypothesize possible collaborative frameworks

The advantages of the use of a telemedicine platform can be estimated by KPIs. For example, in Table 3, we list and comment on some qualitative and quantitative KPIs associated with a telediabetology platform.

Table 3. Possible qualitative and quantitative KPI in telediabetology.

KPI	Description
Qualitative KPI	• Level of functioning of the new PDTA processes of diabetics with the use of the telediabetology platform; • Decrease in cases of hypo- and hyperglycemia (for monitored patients); • Amount of data recorded in the backend system (useful for georeferenced monitoring of diabetics); • Socio-economic impact deriving from the implemented system; • Level of satisfaction of actors (general practitioners, pharmacists, clinical staff, patients, etc.);
Quantitative KPI	• Number of enrolled patients; • Number of hospitalizations and percentage of incidence of monitored diabetic patients; • Number of comorbidities (percentage annual incidence of monitored patients); • Number of electronic health records connected to the project platform and used to predict chronicity; • Number of PDTA processes implemented with integration of the telediabetology facilities; • Risk stratification of hyper- and hypoglycemia; • Percentage reduction in episodes of hypo- and hyperglycemia classified as high risk; • Level of reliability of ML predictive and classification algorithms (evaluation of some parameters such as accuracy, recall, Mean Squared Error—MSE, Root Mean Squared Error—RMSE, ROC curve, etc.); • Scoring of TO BE process monitoring sheets; • Quality of Life (QoL) of monitored diabetic patients.

Table 4 indicates limitations and perspectives of telemedicine technology. The listed limits are common for all telemedicine platforms.

Table 4. Limitations and perspectives of telemedicine technologies.

Technological Limits	Technology Description	Technological Perspectives
Limited availability in the market of certified medical kits (standard CE: 93/42/CEE, 2017/745/UE, class 2a) for diabetes and stroke diagnoses	A certified medical kit groups different sensors in terms of function of the pathology to control. For example, the kit for diabetes includes the following certified devices: • glucometer; • profile analyzer lipid; • balance; • blood pressure meter; • glycated hemoglobin meter; • saliva tester for nutrigenic test. The kit for stroke monitoring includes the following certified sensors: • heart rate sensor; • inertial motion sensor; • pulse oximeter.	The certified devices will require a protocol validating the detected measurements (control room activity)
Digital solution integrated in PDTA	• The integration of digital solutions (electronic health records, sensor data, telemedicine platform databases, etc.) requires data flow integration into a digital PDTA; • the policies of privacy could block data integration in the backend system; • the Software Development Kit (SDK) [28] is necessary to integrate and manipulate data detected by medical sensors.	Future prevention PDTA could include ML decision- making engine and big data analytics tools [29]
Dataset availability	Supervised ML algorithms require a large amount of digital data and clean dataset to optimize the training model	Big data, data fusion techniques, and augmented data are to be considered to improve ML performance

In Appendix B, we discuss an example of the KPI technology readiness model structured by the Ishikawa diagram, typically adopted to model organizational aspects in healthcare [30,31] and production management processes [32]. For the stroke telemonitoring platform, different KPIs should be considered, including neurological assessment, nutritional assessment, hyperthermia management, lipid management [33], stroke education, and screening actions [34]. Another approach useful for mapping processes is the Unified Modeling Language (UML) [35,36], used in Appendix A, where we detail a complete framework of a telemedicine platform. In Appendix C, we list some PMO aspects associated with actors listed in Table 2, and possible advantages and disadvantages following the corrective actions.

The limitations about the adoption of the new PDTA based on ML data processing are mainly in the availability of enough clinical data to learn an ML model, and in the execution of new organizational models capable of ensuring the correct functioning of the PDTA. In this direction, future developments are in the design of new structured process workflows able to efficiently manage new human resources having new roles.

The results proposed in this paper regarding stroke analysis highlight that there are many aspects to consider for risk assessment decision processes in order to prevent dangerous cases. For example, the risk could be reduced over time, improving the social quality of life or optimizing the work conditions, as well as suggesting the inclusion of corrective actions or lifestyles, including the choice of a specific diet. The limitations mainly involve the deployment of an organizational model suitable for directing patients to the correct health path. The organizational model implies new human resources and a synergic

collaboration between them by executing efficient processes. The future direction of the research is to integrate as much as possible into the PDTA the automatisms of AI decision making and the synergies between all the actors that serve to implement risk prevention.

Other alternative ML algorithms adopted in the literature for stroke classification or prediction are Artificial Neural Networks (ANNs) and Support Vector Machine (SVM) [37–40]. In Table 5, we compare the performance of the ANN and SVM methods found in the literature with the FR and GBT algorithms applied in this work; the ANN and the SVM algorithms exhibit a performance lower or slightly lower than that of the FR and GBT algorithms.

Table 5. ML comparison of the accuracy and AUC parameters.

ML Algorithm	Accuracy	AUC
RF (this work)	0.974	0.969
GBT (this work)	0.968	0.942
ANN ([37])	0.91	0.90
ANN ([38])	0.97	0.6587
ANN ([39])	0.875	0.914
ANN ([40])	0.902	No data
SVM ([37])	0.84	0.93
SVM ([38])	0.90	0.5622

Future works will apply the experimentation of the proposed PDTA of Figure 5 and of the use of medical kits, which will be assigned to the hospitalization units characterized by many confirmed cases, as for the Italian district units represented in Figure 13 and Table 6 (Unit 1, Unit 6, and Unit 7).
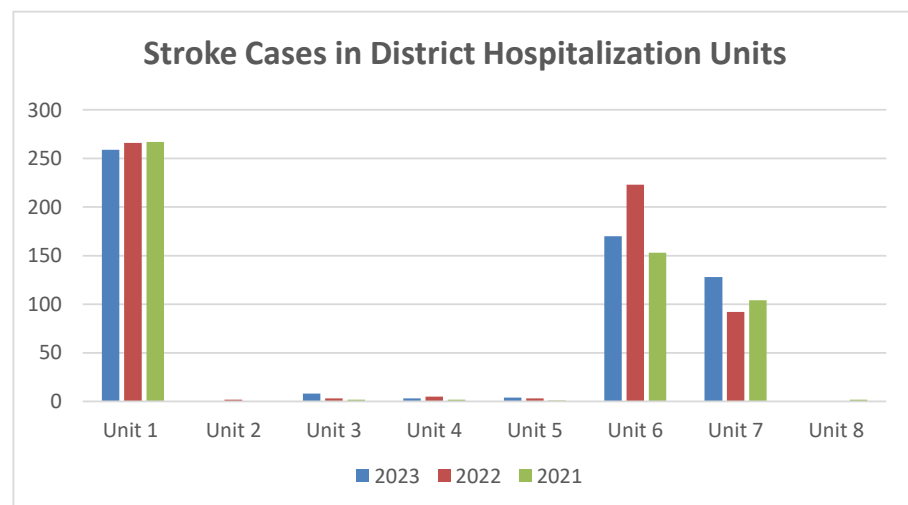

Figure 13. Confirmed stroke cases of different hospitalization units appertaining to a local healthcare district.

Table 6. Stroke cases of hospitalization units of a local regional healthcare district.

Hospitalization Unit	Year 2021	Year 2022	Year 2023
Unit 1	267	266	259
Unit 2	0	2	0
Unit 3	2	3	8
Unit 4	2	5	3
Unit 5	1	3	4
Unit 6	153	223	170
Unit 7	104	92	128
Unit 8	2	9	0

The proposed framework will be able to integrate decision-making procedures suggested by research topics, such as the use of efficient dietary components [41], sentiment analysis [42] matching with the psychological profile of the patient, the metabolism balancing approach [43], the improvement of psychological processes of consumer purchase decision making [44], and the detailed analyses of etiological aspects [45].

5. Conclusions

The proposed paper introduces a methodology to apply ML in order to improve a telemedicine prevention platform for diabetes and stroke, and discusses the related organizational models. Specifically, alerting conditions could be predicted by means of data processing of supervised and unsupervised algorithms to enable a preventive control process of patients, thus avoiding the risk of injuries or becoming chronic cases. The study is focused on the design of a prevention PDTA based on telemedicine platforms adopting a PMO approach. The ML algorithms are applied using an open dataset with the goal of explaining the data processing methodology and how it could interact with the decisions to be made. Important aspects about organization management and technology development are highlighted. The study has been developed within the framework of projects in collaboration with hospitals and companies working in telemedicine. The presented approach is suitable for the design of different prevention healthcare platforms for other chronic risks or comorbidities, allowing the processing of new digital data useful for medical and clinical advances. Furthermore, the discussed methodology allows the researchers to write telemedicine research projects based on ML data processing. The paper is mainly addressed to provide a new PMO methodology to consider PDTA based on prevention processes and optimized by telemedicine tools and ML. Actually, there is an active collaboration with local hospitalization units to collect data about stroke cases in order to define an operative plan financed by projects, including the software and hardware of telemedicine platforms. Today, few data related to the more significant aspects of the post-COVID 19 situation have been collected. Future works will process local clinical datasets to validate the proposed PDTA prevention process, which may be subject to possible further revisions.

Appendix A

A further graphical approach to designing telemedicine processes is based on the standard Unified Modelling Language (UML). In Figure A1, we illustrate the UML Use Case Diagram (UCD) involving different actors and related activities. The telemedicine ecosystem is composed of the following main actors:

- Patients;
- Specialists;
- General Practitioners;
- Industries developing telemedicine tools;
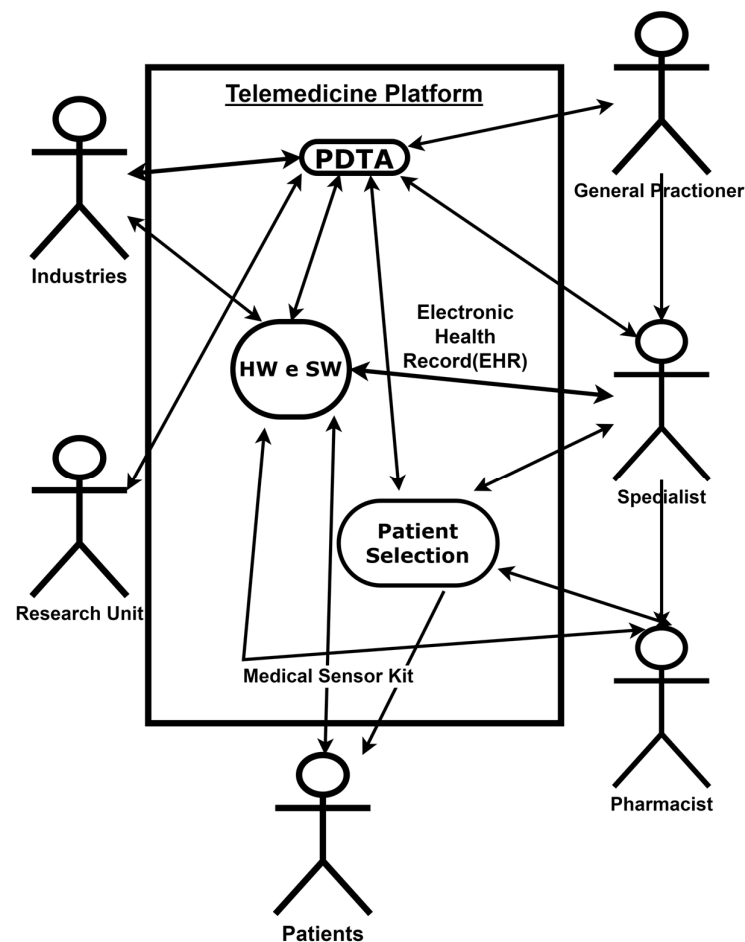- Pharmacists;
- Research units.

Figure A1. UML UCD of a telemedicine ecosystem for the prevention of chronic disease.

Appendix B

The Ishikawa diagram turns out to be the most effective and efficient tool for problem solving. According to a study, 95% of problems in processes can be solved by using the Ishikawa "Quality Control" (QC) tool, exploring their validity in the healthcare sector.

The Ishikawa diagram of Figure A2 illustrates a methodology to formulate the KPI of technology readiness for a company oriented towards the development of telemedicine tools.

The diagram is structured into the following two main groups of variables:

- Upper part (variables of technological aspects): The orange color indicates the readiness aspects for the development of hardware, software, and the whole embedded telemedicine system.

-    Lower part (variables correlated with management and organizational aspects): The green color represents the ability to manage the project developments, the suppliers, the licenses, and the patients.

The diagram of Figure A2 is representative of the hi-tech industrial and social ecosystem of the telemedicine platforms.
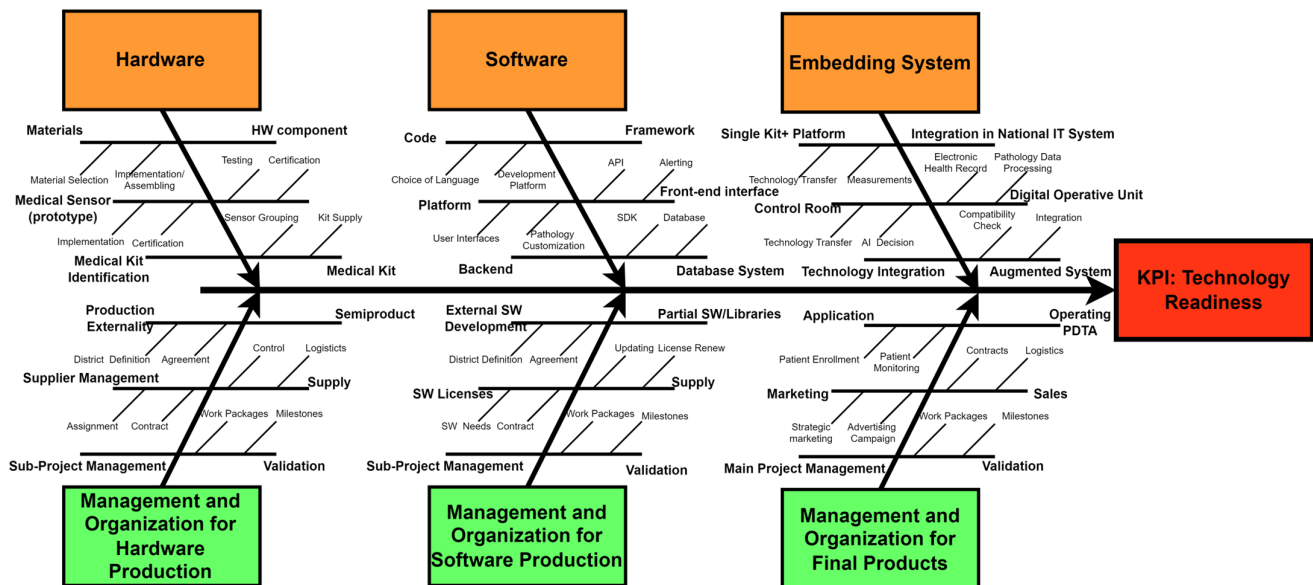

Figure A2. Ishikawa diagram defining the KPI of technology readiness (orange color: technological aspects; green color: organizational and management aspects).

Appendix C

In Table A1, we list some PMO aspects correlated to the management of the medical staff, including the management of the activities of actors listed in Table 2. The table enhances solutions of human resource (HR) interventions in telemedicine, improving the prevention process by specifying possible negative impacts and action guidelines.

Table A1. Framework and action guidelines to optimize the prevention process using a telemedicine platform.

HR Intervention Typology	Goal	Possible Correlated Negative Impact	Action Guidelines
Training of the health HR staff	Decrease in the health risk about chronic pathologies and of irreversible cases (mainly for stroke cases)	In cases of limited staff, the training could generate inefficiencies for PDTA processes, as for the process of Figure 1	Reskilling and upskilling of HR about new technologies enabling telemedicine
Further actor allocation and displacement	• Improvement of prevention processes allocating new actors; • HR allocation according to priorities (patients with high risk); • decrease in the visit delay due to the real-time monitoring of patients.	Imbalances of working times between traditional PDTA and new prevention processes	HR management synchronizing prevention processes and tracing patients which become chronic
HR recruitment	Recruitment of new medical staff skilled in telemedicine (biomedical engineers, specialists, etc.)	Increase in costs due to the engagement of new HR	Recruitment is executed according to the HR necessary to realize an operating prevention platform

Table A1. *Cont.*

HR Intervention Typology	Goal	Possible Correlated Negative Impact	Action Guidelines
Definition of new roles in prevention processes	Formulation of new protocols by designating new staff roles	Possible confusion in the initial organization and HR management due to the interventions of the new actors. This stage requires a patient enrollment phase which could take a long time	Formulation of the new procedures to be performed by the new actors (possibly using BPMN workflows)
HR for control room of telemedicine platforms	Designation of a part of the medical staff to control patients remotely at home (homecare assistance)	Convince patients to use wearable sensors	Guidelines about data privacy and information procedures of patients to be monitored by a telemedicine platform

In Table A2, we describe possible corrective actions to optimize a telemedicine framework by showing related advantages and disadvantages following the corrective actions.

Table A2. Main advantages and disadvantages of the PMO related to the telemedicine framework.

HR Corrective Action	Advantages	Disadvantages
Training on the use of medical kits	• Decrease in the chronic risk; • decrease in injuries (strokes and heart disease).	The training requires a specific plan: different training courses should cover all the HR skills. The training is also to apply for reskill and upskill operations about technology transfer in telemedicine (more complex planning)
Increase in the medical staff or HR allocation/displacement	• Formulation of new prevention care protocols to be integrated into more efficient PDTA; • creation of new operation units for prevention based on telemedicine platforms.	Increase in the HR management impact due to new organization, with a part of the staff operating in the prevention process and in the new integrated PDTA
Monitoring of the prevention risk (control room action)	The patient traceability allows for the estimation of the prevention efficacy	Possible confusions in the reconstruction of the care pattern of the monitored patients

References

1. Massaro, A. Advanced Control Systems in Industry 5.0 Enabling Process Mining. *Sensors* **2022**, *22*, 8677. [CrossRef] [PubMed]
2. Massaro, A. Process Mining in Production Management, Intelligent Control, and Advanced KPI for Dynamic Process Optimization: Industry 5.0 Production Processes. In *Advances in Systems Analysis, Software Engineering, and High Performance Computing*; IGI Global: Hershey, PA, USA, 2023; pp. 1–17.
3. Martin, N.; Wittig, N.; Munoz-Gama, J. Using Process Mining in Healthcare. In *Lecture Notes in Business Information Processing*; Springer International Publishing: Cham, Switzerland, 2022; pp. 416–444.
4. Munoz-Gama, J.; Martin, N.; Fernandez-Llatas, C.; Johnson, O.A.; Sepúlveda, M.; Helm, E.; Galvez-Yanjari, V.; Rojas, E.; Martinez-Millana, A.; Aloini, D.; et al. Process Mining for Healthcare: Characteristics and Challenges. *J. Biomed. Inform.* **2022**, *127*, 103994. [CrossRef] [PubMed]
5. Zerbino, P.; Stefanini, A.; Aloini, D. Process Science in Action: A Literature Review on Process Mining in Business Management. *Technol. Forecast. Soc. Chang.* **2021**, *172*, 121021. [CrossRef]
6. Rosa, A.; Massaro, A. Process Mining Organization (PMO) Modeling and Healthcare Processes. *Knowledge* **2023**, *3*, 662–678. [CrossRef]
7. Rosa, A.; Massaro, A.; McDermott, O. Process Mining Applied to Lean Management Model Improving Decision Making in Healthcare Organizations. In Proceedings of the 18th International Forum on Knowledge Asset Dynamics, Matera, Italy, 7–9 June 2023; pp. 222–231, ISBN 978-88-96687-16-1. Available online: https://aran.library.nuigalway.ie/handle/10379/17801 (accessed on 11 December 2023).
8. Pufahl, L.; Zerbato, F.; Weber, B.; Weber, I. BPMN in Healthcare: Challenges and Best Practices. *Inf. Syst.* **2022**, *107*, 102013. [CrossRef]

9. *ISO/IEC 19150:2013*; Information Technology—Object Management Group Business Process Model and Notation. ISO: London, UK, 2013. Available online: https://www.iso.org/standard/62652.html (accessed on 11 December 2023).

10. Object Management Group Business Process Model and Notation. Available online: https://www.bpmn.org/ (accessed on 11 December 2023).

11. PDTA Diabete. Available online: https://www.salute.gov.it/portale/lea/documenti/pdta/Risultati_2017_PDTA_Diabete.pdf (accessed on 11 December 2023).

12. Epicentro. I Dati PASSI sul Diabete. Available online: https://www.regione.puglia.it/web/ufficio-statistico/-/epicentro.-i-dati-passi-sul-diabete (accessed on 11 December 2023).

13. Il Costo della Malattia Diabetica. Available online: https://www.sanita24.ilsole24ore.com/art/focus-diabete/2018-06-27/il-costo-malattia-diabetica-105318.php?uuid=AE02hEDF (accessed on 11 December 2023).

14. Smith, L.; Chakraborty, D.; Bhattacharya, P.; Sarmah, D.; Koch, S.; Dave, K.R. Exposure to Hypoglycemia and Risk of Stroke. *Ann. N. Y. Acad. Sci.* **2018**, *1431*, 25–34. [CrossRef] [PubMed]

15. Jingqi Yan, Z.Z. Hyperglycemia as a Risk Factor of Ischemic Stroke. *J. Drug Metab. Toxicol.* **2013**, *4*, 4. [CrossRef]

16. Wajngarten, M.; Silva, G.S. Hypertension and Stroke: Update on Treatment. *Eur. Cardiol.* **2019**, *14*, 111–115. [CrossRef]

17. Tsao, C.W.; Aday, A.W.; Almarzooq, Z.I.; Alonso, A.; Beaton, A.Z.; Bittencourt, M.S.; Boehme, A.K.; Buxton, A.E.; Carson, A.P.; Commodore-Mensah, Y.; et al. Heart Disease and Stroke Statistics—2022 Update: A Report from the American Heart Association. *Circulation* **2022**, *145*, e153–e639. [CrossRef]

18. Stroke Prediction Dataset. Available online: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset (accessed on 11 December 2023).

19. Fan, Z.; Yuan, C.; Xin, L.; Wang, X.; Jiang, J.; Wang, Q. HSRF: Community Detection Based on Heterogeneous Attributes and Semi-Supervised Random Forest. In Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Dalian, China, 5–7 May 2021; pp. 1141–1147. [CrossRef]

20. Yao, D.; Zhang, T.; Zhan, X.; Zhang, S.; Zhan, X.; Zhang, C. Geometric Complement Heterogeneous Information and Random Forest for Predicting lncRNA-Disease Associations. *Front. Genet.* **2022**, *13*, 995532. [CrossRef]

21. Luetto, S.; Garuti, F.; Sangineto, E.; Forni, L.; Cucchiara, R. One Transformer for All Time Series: Representing and Training with Time-Dependent Heterogeneous Tabular Data. 2023. Available online: https://www.arxiv-vanity.com/papers/2302.06375/ (accessed on 11 December 2023).

22. Khan, N.U.; Wan, W.; Riaz, R.; Jiang, S.; Wang, X. Prediction and Classification of User Activities Using Machine Learning Models from Location-Based Social Network Data. *Appl. Sci.* **2023**, *13*, 3517. [CrossRef]

23. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

24. Hashemi, S.E.; Gholian-Jouybari, F.; Hajiaghaei-Keshteli, M. A Fuzzy C-Means Algorithm for Optimizing Data Clustering. *Expert Syst. Appl.* **2023**, *227*, 120377. [CrossRef]

25. KNIME. Available online: https://www.knime.com/ (accessed on 11 December 2023).

26. Siopis, G.; Colagiuri, S.; Allman-Farinelli, M. Effectiveness of Dietetic Intervention for People with Type 2 Diabetes: A Meta-Analysis. *Clin. Nutr.* **2021**, *40*, 3114–3122. [CrossRef] [PubMed]

27. Álvarez, R.; Torres, J.; Artola, G.; Epelde, G.; Arranz, S.; Marrugat, G. OBINTER: A Holistic Approach to Catalyse the Self-Management of Chronic Obesity. *Sensors* **2020**, *20*, 5060. [CrossRef] [PubMed]

28. Henriksen, A.; Haugen Mikalsen, M.; Woldaregay, A.Z.; Muzny, M.; Hartvigsen, G.; Hopstock, L.A.; Grimsgaard, S. Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. *J. Med. Internet Res.* **2018**, *20*, e110. [CrossRef] [PubMed]

29. Massaro, A.; Ricci, G.; Selicato, S.; Raminelli, S.; Galiano, A. Decisional Support System with Artificial Intelligence oriented on Health Prediction using a Wearable Device and Big Data. In Proceedings of the 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, Roma, Italy, 3–5 June 2020; pp. 718–723. [CrossRef]

30. Chang, H. Evaluation Framework for Telemedicine Using the Logical Framework Approach and a Fishbone Diagram. *Healthc. Inform. Res.* **2015**, *21*, 230. [CrossRef] [PubMed]

31. McDermott, O.; Antony, J.; Sony, M.; Rosa, A.; Hickey, M.; Grant, T.A. A study on Ishikawa's original basic tools of quality control in healthcare. *TQM J.* **2023**, *35*, 1686–1705. [CrossRef]

32. Botezatu, C.; Condrea, I.; Oroian, B.; Hriţuc, A.; Eţcu, M.; Slătineanu, L. Use of the Ishikawa Diagram in the Investigation of Some Industrial Processes. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *682*, 012012. [CrossRef]

33. Urimubenshi, G.; Langhorne, P.; Cadilhac, D.A.; Kagwiza, J.N.; Wu, O. Association between Patient Outcomes and Key Performance Indicators of Stroke Care Quality: A Systematic Review and Meta-Analysis. *Eur. Stroke J.* **2017**, *2*, 287–307. [CrossRef]

34. Mohammed, M.; Zainal, H.; Tangiisuran, B.; Harun, S.N.; Ghadzi, S.M.; Looi, I.; Sidek, N.N.; Yee, K.L.; Aziz, Z.A. Impact of Adherence to Key Performance Indicators on Mortality among Patients Managed for Ischemic Stroke. *Pharm. Pract.* **2020**, *18*, 1760. [CrossRef]

35. *ISO/IEC 19505-1/2:2012*; Information Technology—Object Management Group Unified Modeling Language (OMG UML). ISO: London, UK, 2012. Available online: https://www.iso.org/obp/ui/#iso:std:iso-iec:19505:-2:ed-1:v1:en (accessed on 31 December 2023).

36. Suriya, S.; Nivetha, S. Design of UML Diagrams for WEBMED—Healthcare Service System Services. *ICST Trans. e-Educ. e-Learn.* **2023**, *8*, e5. [CrossRef]

37. Sohn, J.; Jung, I.-Y.; Ku, Y.; Kim, Y. Machine-Learning-Based Rehabilitation Prognosis Prediction in Patients with Ischemic Stroke Using Brainstem Auditory Evoked Potential. *Diagnostics* **2021**, *11*, 673. [CrossRef] [PubMed]

38. Srinivasu, P.N.; Sirisha, U.; Sandeep, K.; Praveen, S.P.; Maguluri, L.P.; Bikku, T. An Interpretable Approach with Explainable AI for Heart Stroke Prediction. *Diagnostics* **2024**, *14*, 128. [CrossRef] [PubMed]

39. Iosa, M.; Paolucci, S.; Antonucci, G.; Ciancarelli, I.; Morone, G. Application of an Artificial Neural Network to Identify the Factors Influencing Neurorehabilitation Outcomes of Patients with Ischemic Stroke Treated with Thrombolysis. *Biomolecules* **2023**, *13*, 334. [CrossRef] [PubMed]

40. Usama, N.; Niazi, I.K.; Dremstrup, K.; Jochumsen, M. Detection of Error-Related Potentials in Stroke Patients from EEG Using an Artificial Neural Network. *Sensors* **2021**, *21*, 6274. [CrossRef] [PubMed]

41. Hill, C.R.; Shafaei, A.; Balmer, L.; Lewis, J.R.; Hodgson, J.M.; Millar, A.H.; Blekkenhorst, L.C. Sulfur Compounds: From Plants to Humans and Their Role in Chronic Disease Prevention. *Crit. Rev. Food Sci. Nutr.* **2023**, *63*, 8616–8638. [CrossRef] [PubMed]

42. Cai, Y.; Ke, W.; Cui, E.; Yu, F. A Deep Recommendation Model of Cross-Grained Sentiments of User Reviews and Ratings. *Inf. Process. Manag.* **2022**, *59*, 102842. [CrossRef]

43. He, W.-J.; Lv, C.-H.; Chen, Z.; Shi, M.; Zeng, C.-X.; Hou, D.-X.; Qin, S. The Regulatory Effect of Phytochemicals on Chronic Diseases by Targeting Nrf2-ARE Signaling Pathway. *Antioxidants* **2023**, *12*, 236. [CrossRef]

44. Zhu, P.; Miao, C.; Wang, Z.; Li, X. Informational Cascade, Regulatory Focus and Purchase Intention in Online Flash Shopping. *Electron. Commer. Res. Appl.* **2023**, *62*, 101343. [CrossRef]

45. Darnton-Hill, I.; Nishida, C.; James, W.P.T. A Life Course Approach to Diet, Nutrition and the Prevention of Chronic Diseases. *Public Health Nutr.* **2004**, *7*, 101–121. [CrossRef]